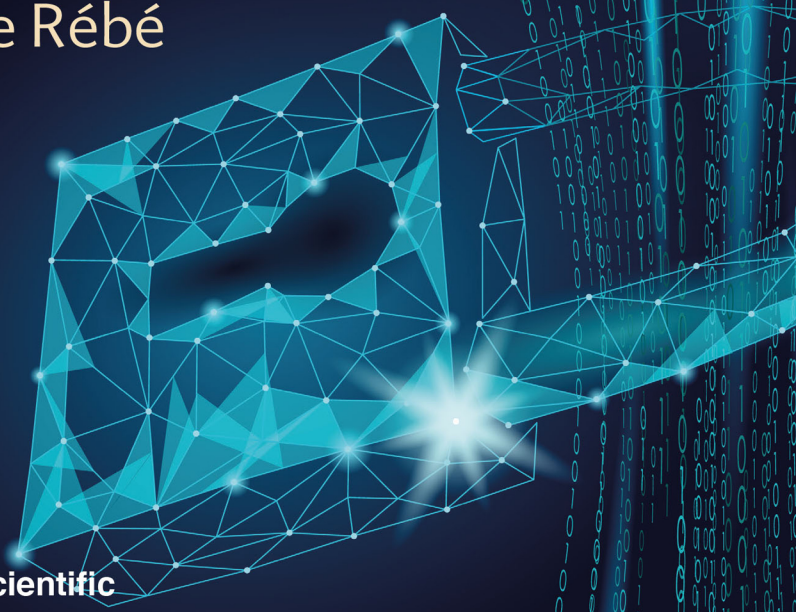
Regulating Cyber Technologies

Privacy vs Security

Editor

Nathalie Rébé



World Scientific

Regulating Cyber Technologies

Privacy vs Security

This page intentionally left blank

Regulating Cyber Technologies

Privacy vs Security

Editor

Nathalie Rébé

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI • TOKYO

Published by

World Scientific Publishing Europe Ltd.

57 Shelton Street, Covent Garden, London WC2H 9HE

Head office: 5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

Library of Congress Cataloging-in-Publication Data
Names: Rébé, Nathalie, editor.
Title: Regulating cyber technologies : privacy vs security / editor Nathalie Rébé.
Description: New Jersey : World Scientific, [2023] | Includes bibliographical references and index.
Identifiers: LCCN 2022026286 | ISBN 9781800612853 (hardcover) |
ISBN 9781800612860 (ebook) | ISBN 9781800612877 (ebook other)
Subjects: LCSH: Information technology--Management. | Computer security. |
Database security. | Computer networks--Security measures.
Classification: LCC HD30.2 .R45 2023 | DDC 005.8--dc23/eng/20220614
LC record available at https://lccn.loc.gov/2022026286

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

For any available supplementary material, please visit
https://www.worldscientific.com/worldscibooks/10.1142/Q0379#t=suppl

Desk Editors: Soundararajan Raghuraman/Adam Binnie/Shi Ying Koe

Typeset by Stallion Press
Email: enquiries@stallionpress.com

Printed in Singapore

https://doi.org/10.1142/9781800612860_fmatter

Preface

Regulating cyber matters is a complex task, as cyberspace is an intricate world full of new threats related to a person's identity, finance, and private information. Algorithm manipulation, hate crimes, cyber-laundering, and data theft are strong menaces in the cyber world. New technologies are generating both privacy and security issues, involving anonymity, cross-border transactions, virtual communications, and assets, among others.

This book is a collection of works by experts on cyber matters and legal considerations that need addressing in a timely manner. It comprises cross-disciplinary knowledge that is pooled to this end. Risk mitigation tools, including cyber risk management, data protection regulations, as well as ethical practice guidelines, are reviewed in detail.

The regulatory issues associated with new technologies along with emergent challenges in the field of cybersecurity that require improved regulatory frameworks are considered. We probe ethical, material, and enforcement threats, thus revealing the inadequacy of current legal practices. To address these shortcomings, we propose new regulatory privacy and security guidelines that can be implemented to deal with the new technologies and cyber matters.

This page intentionally left blank

https://doi.org/10.1142/9781800612860_fmatter

Editorial Note

I dedicate this important work to my esteemed colleagues and dear friends who contributed to this edited volume on regulating cyber technologies. I want to warmly thank these amazing experts, lawyers, judges, and academics who believed in me and in my project and joined the team to build this valuable collection of essays. Together, their works on privacy and security represent the landscape of such a complex hot topic which deserves more consideration at the international level.

This page intentionally left blank

https://doi.org/10.1142/9781800612860_fmatter

About the Editor

Dr. Nathalie Rébé is a Financial Crime and AML Consultant in Luxembourg. Dr. Rébé holds a Doctorate in Business Administration (DBA) from Paris School of Business, and a Doctorate in Juridical Science (JSD) on Financial Crimes from Thomas Jefferson School of Law (USA). She has participated in various international conferences as an academic author, and taught on both International Criminal Law, and Business Administration university-level courses. With a Postgraduate Diploma in Cyber Law from the University of Montpellier, Dr. Rébé's research and publications have been focused on new technologies, security, privacy, and regulatory matters.

This page intentionally left blank

https://doi.org/10.1142/9781800612860_fmatter

About the Contributors

Dr. Costel Ciuchi is a Senior Expert in the Information Technology and Digitalization Directorate, General Secretariat of the Romanian Government with responsibilities in developing government apps and infrastructure, security of IT services (INFOSEC), and coordinating Gov.ro Domain Registry. As an Associate Professor at University Politehnica of Bucharest, he conducts research activities in decision-making, cybersecurity, and security risk area.

Dr. Dragos Nicolae Costescu is a Lecturer in EU Internet law and also the vice-dean at the University of Bucharest, Law Faculty. He is also a public notary in Bucharest (for 20 years now). In addition, his area of interest and practice is focused on Internet law and emergent technologies.

Hon. Dr. Fausto Martin De Sanctis is a Federal Appeals Judge at the Federal Court of Appeals for the 3rd Region, in Sao Paulo, Brazil. Previously, he was a São Paulo State Judge (1990–1991), Public Prosecutor of the Municipality of São Paulo, and Public Prosecutor of the State of São Paulo, in the area of the Public Defender's Office. He was a Professor at São Judas Tadeu University for 12 years. He holds a PhD in Criminal Law from São Paulo University (USP), and a Postgraduate Diploma in Civil Procedure from Brasília University (UnB). He has 39 legal works published in Brazil and abroad, in addition to articles specialized in Civil Procedure.

Prof. Dr. Steven Furnell is a Professor of Cybersecurity at the University of Nottingham. He is also an Honorary Professor with Nelson Mandela

University in South Africa and an Adjunct Professor with Edith Cowan University in Western Australia. His research interests include usability of security and privacy, security management and culture, and technologies for user authentication and intrusion detection. He has authored over 340 papers in refereed international journals and conference proceedings, as well as various books and book chapters. Prof. Furnell is the Chair of Technical Committee 11 (security and privacy) within the International Federation for Information Processing and a Board Member of the Chartered Institute of Information Security.

Jorge Alan García Bazán serves as the Cybersecurity Services Director and Co-Founder of Nemotek. With more than 15 years of experience in data protection and defensive cybersecurity, including business development and leadership roles, Jorge Alan is responsible for the strategic planning, operations, and cybersecurity services for actual and future needs of Nemotek. Jorge Alan is very passionate and enjoys using his skills and knowledge to contribute to the technological advances in cybersecurity and fight against cybercrime. He holds certifications in Security Engineering, Disaster Recovery, and Information Systems Security, a diploma in Security Operations from Monterrey Institute of Technology and Higher Education, and a bachelor's degree in Information Technology from the University of Monterrey.

Dr. Elena Lazar is a Lecturer in Public International Law and EU Internet Law at the University of Bucharest Law Faculty, as well as a lawyer at the Bucharest Bar. Her area of practice as a lawyer includes human rights, data protection, and data privacy. She also holds a post-doctorate in AI and international law from Paris II University-Pantheon Assas.

Hon. Margaret M. McKeown is a Judge on the United States Court of Appeals for the Ninth Circuit, appointed in 1998. She is a member of the American Academy of Arts and Sciences and the Council of the American Law Institute, where she has been an advisor on multiple international and intellectual property projects. She teaches at the University of San Diego and Northwestern University Law Schools and has lectured worldwide on judicial independence, ethics, rule of law, and intellectual property.

Dr. Ioan-Cosmin Mihai is a Researcher, Professor, Trainer, and Conference Speaker, with an experience of more than 15 years in

cybercrime and cybersecurity. He is Seconded National Expert at the European Union Agency for Law Enforcement Training (CEPOL), Associate Professor at "Al. I. Cuza" Police Academy, Visiting Professor at the University Politehnica of Bucharest, and Vice President of the Romanian Association for Information Security Assurance (RAISA).

Dr. Sérgio Nunes is an Assistant Professor at ISEG, Universidade de Lisboa, Portugal. He has a PhD in Management from the same university and a Master in Information Technology, Information Security, from Carnegie Mellon University, USA, and another Master in Information Security from the Faculty of Science of Universidade de Lisboa, Portugal. He has worked as a cybersecurity manager at several auditing and consulting firms and holds several certifications in that domain, namely, CISA, CISM, and CISSP.

Dr. Tal Pavel is the Head of Cybersecurity Studies in the Information Systems school, at the Academic College of Tel Aviv Yaffo and the founder and director of the privately owned Institute for Cyber Policy Studies. Dr. Pavel is an academic lecturer, researcher, and speaker specializing in cyber threats and cyber policy; has served as a keynote speaker at international conferences; and has been interviewed as a cyber expert by major media outlets. Dr. Pavel holds a PhD in Middle Eastern Studies from Bar-Ilan University, Israel (Dissertation: "Changes in Governmental Restrictions over the Use of Internet in Syria, Egypt, Saudi Arabia, and the United Arab Emirates between the Years 2002–2005").

Dr. Gabriel Petrică has an extensive experience acquired in over 25 years of work in ICT field. With a PhD in Electronics, his area of interest includes the dependability of systems, Web programming, and information security. Currently, he performs teaching and research activities within the Faculty of Electronics, Telecommunications, and Information Technology at the University Politehnica of Bucharest.

Prof. Dr. Andy Phippen is a Professor of Digital Rights at Bournemouth University, UK. Andy has spent over 20 years researching and writing about ethical practices in the IT sector, with a particular focus on social aspects of technology. A computer scientist by background, Andy's current work focuses on the intersection of policy, law, and technology with matters related to privacy and safeguarding. He publishes frequently on these issues and is a regular contributor to policy debates and the media.

Dr. Vignesh Ram is an Assistant Professor at the Department of Geopolitics and International Relations at the Manipal Academy of Higher Education in Manipal, India.

Prof. Mikhail Reider-Gordon is on Faculty at the International Anti-Corruption Academy (IACA) in Austria. She is a lawyer and also the Managing Director of Institutional Ethics & Integrity at Affiliated Monitors, Inc. Prof. Gordon's practice areas of expertise include anti-corruption, anti-money laundering, technology and privacy compliance, and international law. She holds leadership positions with the American Bar Association, including Operations Officer for the International Law Section. She is the past co-chair of the Section's International Anti-Corruption and Anti-Money Laundering Committees; co-chaired the Working Group on Beneficial Ownership Transparency; chair of the greater ABA's Task Force on Gatekeeper Regulation and the Profession; and served on the Association's Standing Committee on Technology and Information Systems. She is a noted and widely published expert on issues related to AML, corruption, and cybercrime. She currently has a podcast series, *Lies, Spies & Corporate Crime: Wirecard, the Saga*, that explores the Wirecard matter from a transnational corporate crime perspective.

Dan Shefet is a French lawyer, born in Denmark. He is the author of the individual specialist report to UNESCO on "Online Radicalization," a member of the American Law Institute, and an Expert with the Council of Europe on the Internet Ombudsman as well as the President of AAID. Dan specializes in European law and Human Rights as they apply to tech.

Dr. Georg Thomas is a cybersecurity and risk expert based in Melbourne, Australia. He has knowledge and experience in hacking, cyberwarfare, cyber terrorism, cyber defense, and blockchain technologies. He currently lectures on cyberwarfare and terrorism and hacking countermeasures at Charles Sturt University. Georg holds a number of industry certifications including CCISO, CDPSE, CEH, CISM, and CISSP and is on the Australian Computer Society Ethics Committee and a former Board Director of ISACA New York Metropolitan Chapter.

https://doi.org/10.1142/9781800612860_fmatter

Contents

Preface v

Editorial Note vii

About the Editor ix

About the Contributors xi

List of Figures xvii

List of Tables xix

List of Abbreviations xxi

Introduction xxvii
Nathalie Rébé

Chapter 1 The Latest Challenges in the Cybersecurity Field 1
Ioan-Cosmin Mihai, Costel Ciuchi, Gabriel Petrică

Chapter 2 Governance of Cyberspace — El Dorado of States and Private Actors 19
Dragos Nicolae Costescu

Chapter 3 Defining Cyber Risk Management Objectives 37
Sérgio Nunes

Chapter 4	Data Protection Concerns in Emerging Technologies *Jorge Alan García Bazán*	61
Chapter 5	Biometric Technology and User Identity *Steven Furnell*	81
Chapter 6	National Cyber Policies Attitude Toward Digital Privacy *Tal Pavel*	111
Chapter 7	Too Much Information: OSINT in Criminal Investigations and the Erosion of Privacy *Mikhail Reider-Gordon*	145
Chapter 8	The Balance of Opinion in Social Media Regulation — Regime Stability and Risk in Democratic and Non-Democratic Nation-States *Vignesh Ram*	177
Chapter 9	Children, Data Collection, and Privacy — Is the Safeguarding Fallacy a Justification for Excessive Regulation and an Erosion of Human Rights? *Andy Phippen*	203
Chapter 10	Privacy and Security of Health Data — What's at Stake? *Elena Lazar*	235
Chapter 11	Hate Speech: A Comparative Analysis of the United States and Europe *Margaret M. McKeown and Dan Shefet*	257
Chapter 12	Cyber Risks, Dark Web, and Money Laundering *Fausto Martin De Sanctis*	283
Chapter 13	Discussing Regulation for Ethical Hackers *Georg Thomas*	315
Index		335

https://doi.org/10.1142/9781800612860_fmatter

List of Figures

Chapter 1

Figure 1 Cyber kill chain intrusion model 2
Figure 2 ENISA top threats 2019–2020 4
Figure 3 The risk management process 7

Chapter 5

Figure 1 Biometric enrollment and verification processes 90
Figure 2 Relationship between false acceptance and false rejection errors 91
Figure 3 Fingerprint readers on laptops with (a) a separate sensor and (b) a sensor integrated within the power button 96
Figure 4 Concern and confidence around use of biometrics 100

Chapter 6

Figure 1 Number of cyber policy documents published by the top 20 cyber countries during 2015–2021 133

This page intentionally left blank

https://doi.org/10.1142/9781800612860_fmatter

List of Tables

Chapter 3

Table 1 Means and fundamental objectives for cyber risk management 42

Chapter 5

Table 1 Examples of physiological and behavioral characteristics that can be used for biometrics 83

Table 2 Comparing the characteristics of different modes of user authentication 87

Chapter 6

Table 1 List of websites and indices that rank countries in cybersecurity 118

Table 2 Mapping the cybersecurity of world's leading countries 119

Table 3 Mapping the number of countries according to the number of occurrences 122

Table 4 Ranking of the 20 leading countries in cybersecurity according to the number of occurrences 122

Table 5 Reference to cyberspace privacy as part of national cyber policy documents 123

Table 6 How cyberspace is addressed as part of national cyber policy documents 128

Table 7 Summary of how cyber privacy is addressed as part of the national cyber policy documents 129
Table 8 Reference to the COVID-19 pandemic in national cyber policy documents 129
Table 9 The rate of the cyber leading countries among each continent's countries 132
Table 10 Mapping the factors that publish the national cyber policy documents 133

https://doi.org/10.1142/9781800612860_fmatter

List of Abbreviations

3D	Three Dimensional
AAID	Android Advertiser ID
ABA	American Bar Association
ACS	Australian Computer Society
AI	Artificial Intelligence
ANPD	Autoridade Nacional de Proteção de Dados
API	Application Programming Interface
ABNT	Associação Brasileira de Normas Técnicas
APP	Australian Privacy Principles
APT	Advanced Persistent Threat
ARPANET	Advanced Research Projects Agency Network
AUD	Australian Dollar
BATX	Baidu, Alibaba, Tencent and Xiaomi
BBC	British Broadcasting Corporation
BCSC	British Columbia Supreme Court
BJP	Bhartiya Janta Party
BLE	Bluetooth Low Energy
CBP	Customs and Border Protection
CCISO	Certified Chief Information Security Officer
CCPA	California Consumer Privacy Act
CDA	Communications Decency Act
CDO	Chief Data Officer
CDPSE	Certified Data Privacy Solutions Engineer
CEA	Committee on Economic Affairs
CEH	Certified Ethical Hacker

CEO	Chief Executive Officer
CIA	Central Intelligence Agency
CIO	Chief Information Officer
CIS	Center for Internet Security
CISA	Certified Information Systems Auditor
CISM	Certified Information Security Manager
CISO	Chief Information Security Officer
CISSP	Certified Information Systems Security Professional
CJUE	Cour de Justice de l'Union Européenne
COPPA	Children's Online Privacy Protection Act
COVID-19	Coronavirus Disease-2019
COVINTS	Covert Overt Intelligence Solutions
CPENT	Certified Penetration Testing Professional
CPO	Chief Protection Officer
CRC	Convention on the Right of the Child
CSRF	Cross-Site Request Forgery
CSA	Conseil Superieur de l'Audiovisuel
CSIRT	Computer Security Incident Response Team
CSIS	Center for Strategic & International Studies
CTR	Clinical Trial Regulation
DCMS	Department for Digital, Culture, Media and Sport
DDoS	Distributed Denial of Service
DEA	Drug Enforcement Administration
DHS	Department of Homeland Security
DLP	Data Leak Protection
DNS	Domain Name System
DOJ	Department of Justice (US)
DoS	Denial of Service
DPA	Data Protection Authority
DPO	Data Protection Officer
DSA	Digital Services Act
EAR	Equal Acceptance Rate
EC	European Council
ECHR	European Convention for the Protection of Human Rights and Fundamental Freedoms
EDA	European Defense Agency
EDPB	European Data Protection Board
EEA	European Economic Area
EER	Equal Error Rate

EEAS	European External Action Service
EFF	Electronic Frontier Foundation
EHF	Ethical Hacking Framework
EMEA	Europe, Middle East, and Africa
ENISA	European Network and Information Security Agency
EPIC	Electronic Privacy Information Center
ERM	Enterprise Risk Management
EU	European Union
EUR	Euro
FA	False Acceptance
FAR	False Acceptance Rate
FAR	False Alarm Rate
FBI	Federal Bureau of Investigation
FMR	False Match Rate
FNC	Federal Networking Council
FR	False Rejection
FRA	France
FRR	False Rejection Rate
FTA	Failure To Acquire
FTE	Failure To Enroll
G20	Group of 20
G7	Group of 7
GAFAM	Google, Apple, Facebook, Amazon, Microsoft
GDPR	General Data Protection Regulation
GGE	Group of Government Experts
GIAC	Global Information Assurance Certification
GIATOC	The Global Initiative Against Transnational Organized Crime
GPEN	GIAN Penetration Tester
GPS	Global Positioning System
GWAPT	GIAC Web Application Penetration Tester
GXPN	GIAC Exploit Researcher and Advanced Penetration Tester
HIPAA	Health Insurance Portability and Accountability Act
HP	Hewlett Packard
IaaS	Infrastructure as a Service
IACA	International Anti-Corruption Academy
IAPP	International Association of Privacy Professionals

IBM	International Business Machines
ICANN	Internet Corporation for Assigned Names and Numbers
ICCPR	International Covenant on Civil and Political Rights
ICJ	International Court of Justice
ICO	Information Commissioner's Office
iCOP	Internet Covert Operations Program
ICT	Information and Communication Technology
ID	Identification
IDAC	International Digital Accountability Council
IEC	International Electrotechnical Commission
IGF	Internet Governance Forum
INC	Indian National Congress
IoT	Internet of Things
IPR	Impostor Pass Rate
ISACA	Information Systems Audit and Control Association
ISC	International Information Systems Security Certification Consortium
ISIS	Islamic State in Iraq and Syria
ISO	International Standards Organization
IT	Information Technology
KPI	Key Performance Indicators
KPMG	Klynveld Peat Marwick Goerdeler
LED	Law Enforcement Directive
LGPD	Lei Geral de Proteção de Dados Pessoais
LPT	Licensed Penetration Tester
MIT	Massachusetts Institute of Technology
MitM	Man in the Middle
MMS	Media Monitoring Services
NBC	National Broadcasting Company
NCA	National Crime Agency
NCSC	National Cybersecurity Centre
NDB	Notifiable Data Breaches
NetzDG	Network Enforcement Act
NFC	Near Field Communication
NHS	National Health System (British)
NIS	Network and Information Security (Directive)
NIST	National Institute for Standards and Technology
NDA	Non-Disclosure Agreement

NPCC	National Police Chief Council (UK)
NYCRR500	NYDFS Cybersecurity Regulation
NYDFS	New York Department of Financial Services
NYPD	New York City Police Department
OAIC	Office of Australian Information Commissioner
OECD	Organization for Economic Co-operation and Development
OEWG	Open-Ended Working Group
OHCHR	Office of the High Commissioner for Human Rights (United Nations)
OM	Operation Manual
OS	Open-Source
OSCE	Offensive Security Certified Professional
OSCP	Offensive Security Certified Expert
OSI	Open-Source Information
OSINT	Open-Source Intelligence
PACT	Platform Accountability and Consumer Transparency Act
PC	Personal Computer
PCI-DSS	Payment Card Industry-Data Security Standard
PDA	Personal Digital Assistant
PDF	Portable Document Format
PDPA	Personal Data Protection Act
PDPC	Personal Data Privacy Commissioner
PIA	Privacy Impact Assessment
PIN	Personal Identification Number
PIPEDA	Personal Information Protection and Electronic Document Act
PRISM	Political Risk and Intelligence Services Management
PSDB-ES	Partido da Social Democracia Brasileira-Espírito Santo
QR CODE	Quick Response Code
RAISA	Romanian Association for Information Security Assurance
R.A.V	Robert A. Viktora
RAND	Research ANd Development
RCMP	Royal Canadian Mounted Police
RFID	Radio Frequency IDentification
ROE	Rules of Engagement

RQ	Research Question
RUSI	Royal United Services Institute
SaaS	Software as a Service
SEAP	SEcretariat of Penitentiary Administration
SIENA	Secure Information Exchange Network Application
SOCMINT	Social Media Intelligence
SOX	Sarbanes-Oxley Act
SPA	Swedish Polish Authority
SQLi	Structured Query Language injection
TCP	Transmission Control Protocol
TFEU	Treaty on the Functioning of the European Union
TIOS	Tactical Internet Operational Support
TOR	The Onion Router
UCLA	University of California, Los Angeles
UIC	Union for International Communications
UK	United Kingdom
UN	United Nations
UNCLOS	United Nations Convention on the Law of the Sea
UNCTAD	United Nations Conference on Trade and Development
UNODC	United Nations Office on Drugs and Crime
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNGGE	United Nations Group of Governmental Experts
UPI	Undercover Policing Inquiry
US	United States
USB	Universal Serial Bus
USCYBERCOM	United States Cyber Command
USD	United States Dollar
USP	University of São Paulo
USPS	United States Postal Service
VOIP	Voice over Internet Protocol
VPN	Virtual Private Network
WHO	World Health Organization
WITI	Why Is This Important
XSS	Cross-Site Scripting

https://doi.org/10.1142/9781800612860_0001

Introduction

Nathalie Rébé

This book brings together valuable works on online privacy and security threats. It is written by leading scholars, lawyers, and judges and offers a significant contribution to how we understand cyber matters. This introduction describes the essays included and provides the general background to the content.

New technologies are generating new threats and legal challenges that are not easy to counteract with basic regulatory frameworks, as the cyber world crosses physical borders and allows for anonymity. In this virtual world, data, identity, and freedom of expression are not safe. Imposing strict data protection and privacy guidelines to online service providers is paramount for users of the Internet. International associations, governments, and enterprises should collaborate to share best practices in order to fight cybercrime.

However, as regulations might prove inefficient, better risk management is a must. Users need to be educated and provided guidance on the dangers of sharing information online. Since, nowadays, it is nigh on impossible to secure information completely, striking the right balance in terms of what we are willing to share with the world is crucial.

The Structuring of the Book

In this book, we discuss a range of examples of cybersecurity and privacy issues that can be encountered in the online world. While analyzing cyber technologies' ethical, material, and enforcement threats, the authors reveal

the inadequacy of current legal practices. The chapters are now introduced before concluding with some remarks about their selection for this book.

First, in Chapter 1, Ioan-Cosmin Mihai, Costel Ciuchi, and Gabriel Petrică start by setting out *The Latest Challenges in the Cybersecurity Field*, which must be overcome to build a safer and more secure Internet that ensures a resilient infrastructure and protection of critical services. The second theme engaged with is the *Governance of Cyberspace — El Dorado of States and Private Actors*. In Chapter 2, Dragos Nicolae Costescu advocates the need to have better cooperation among not only the public and private sectors but also civil society. States should also collaborate in order to establish norms for governing cyberspace, as well as better framing and definition of the concept of sovereignty.

The human element plays a critical role in cybersecurity, sometimes providing solutions, while potentially also causing problems. In Chapter 3, Sérgio Nunes considers *Defining Cyber Risk Management Objectives.* This chapter points out that security awareness must be continuously enforced and audits must be undertaken with employees in order to identify abnormal or fraudulent behaviors that could compromise the company. It also recommends mandating top management to engage in cyber risk management as part of the enterprise risk management strategy and for it to be a criminal offense if leaders in companies fail to do so. Specifically, Nunes argues that organizations should adopt a cybersecurity strategy that enforces processes, procedures, education, and compliance as part of the overall business strategy.

As many usages of personal data regarding our private lives and personal opinions can be endangered online, Jorge Alan García Bazán next addresses *Data Protection Concerns in Emerging Technologies* in Chapter 4. The author further discusses important subjects related to privacy and new technologies, including Data Discovery and Classification, Privacy Concerns in the Cloud, Big Data, as well as Privacy and Ethical Issues in the Internet of Things.

To further this argument, Steven Furnell discusses the need for appropriate security and protection to be applied to the processing, storage, and any transmission of biometric data in Chapter 5, on *Biometric Technology and User Identity*. The focus of this chapter is how we represent — and more particularly how we prove — our identity in IT and online contexts. Issues for regulatory consideration regarding biometric data are also proposed.

As part of cybersecurity legislation, countries outside of the EU must not only strengthen data protection laws to be a minimum consistent with the GDPR but also seek to enact more robust personal privacy and personality rights laws to address specifically the collection and use of PII, biometric, and other sensitive data open-sourced from social media platforms, public cameras, IoT devices, and software that collects consumer information. In order to understand *National Cyber Policies Attitude Toward Digital Privacy*, Tal Pavel examines the importance of the issue of online privacy in Chapter 6, by assessing 20 countries' national cyber policy documents relating to this issue. The author examines the degree to which these countries ranked as world leaders in cyberspace are handling the issue of online privacy.

To comprehend the exposure of data, Mikhail Reider-Gordon explains the role of *OSINT in Criminal Investigations and the Erosion of Privacy* in Chapter 7. She proposes that regulatory oversight of state agencies that collect, process, and utilize OSINT should be codified into national laws in order to ensure transparency, accountability, and effective oversight.

There is immense power in the hands of global corporations acting as "digital gatekeepers" of social media. Technology flows may also be used to suppress citizen actions or dissent in all types of regimes. Checks and balances need to be maintained so that technology may not be used for the abuse of citizens. In Chapter 8, Vignesh Ram examines the issues of *The Balance of Opinion in Social Media Regulation* pertaining to *Regime Stability and Risk in Democratic and Non-Democratic Nation States*. Ram emphasizes that citizens should have transparent knowledge of information accessed by states for any reason. The author also argues that the use of social media for politics needs to be regulated by a three-party consensus including citizens, companies, and government.

Another key issue regarding technology is its accessibility. Children have access to technology at a very young age. However, they do not yet understand fully the implication of their actions regarding safety, privacy, speech, and data. Therefore, regarding the regulation of cyber matters, it is essential to consider the rights of the child as a specific category of end user. In Chapter 9, Andy Phippen analyzes *Children, Data Collection, and Privacy* concerns. This chapter, thus, explores the need to keep children safe online and concludes that excess data collection and erosion of their privacy rights may be necessary for doing so.

Companies are seeking to collect more customer information than ever before, particularly data regarding health matters, such as for insurance purposes. As cyberattacks are becoming more common nowadays, there is a need for a clear framework to address how such private information can be protected. Accordingly, in Chapter 10, Elena Lazar considers the stakes of *Privacy and Security of Health Data.* Since sensitive data exposes companies to many risks, they should review their industry regulations regarding data protection to ensure that proper security measures are taken to protect data wherever it resides. Parties who process or consume personal data must comply with data regulations. For this purpose, Lazar stresses the necessity to have an EU regulation that addresses the processing of sensitive data. The author states the importance of having public education programs in order to inform patients of their rights related to data protection and privacy, as well as clear internal procedures for hospitals to follow when targeted by cyberattacks. In addition, the author emphasizes the need to conduct training in order to inform medical personnel of the risks posed by cyberattacks.

With pending and sweeping legislation in the offing, coupled with public pressure and the sensitivity of technology companies to the demands of their users, the regulation of hate speech remains front and center in the public debate. In Chapter 11, the Honorable Margaret M. McKeown and Dan Shefet explain *Hate Speech* and provide *A Comparative Analysis of the United States and Europe*. For the authors, balancing the important principle of free expression with the acknowledged need to curb hateful speech presents a critical public policy challenge for the coming decade.

Other main privacy and security concerns including *Cyber Risks, Dark Web, and Money Laundering* are discussed by the Honorable Fausto Martin De Sanctis in Chapter 12. Cyberspace must be an environment in which people's rights, especially their personal data, can be effectively protected against all kinds of criminal savagery and to avoid money laundering. Nowadays, there is no guarantee an organization is secured from attackers, especially where the Dark Web is concerned. According to the author, governments must take into account the uncertainty of technological transformation, thereby ensuring that legislation provides adequate social protection and ensuring that criminals are prevented from hiding in cyberspace and laundering money on an ongoing basis. Cybersecurity risks must become a leading priority for organizations, guided by a

well-designed and rigorous international regulatory environment, which will require the updating of the corresponding criminal legislation.

Last, in Chapter 13, Georg Thomas is *Discussing Regulation for Ethical Hackers* and offers an important argument in support of creating guidelines in this field. These would include codes of ethics and compliance along with experience and certification requirements, among other things, for ethical hackers.

Conclusion

A final comment should be made about how these essays were selected for publication. We received submissions from all over the world, with the authors being chosen according to their recognized expertise in technology, privacy, security, and legal matters regarding their selected topic.

Rather than being treated as a single narrative, the essays can be read in any number or order, according to reader preference. They present a range of important perspectives and insights on the topic of cybersecurity that will provide stimulating insights regarding the central debates and ideas in this field. Moreover, our intention is that readers interested in cyber technologies will learn more on a variety of topics involving privacy and security issues, thus being made aware of the inadequacy of current legal practices.

I hope that you enjoy reading the essays in this book as much as I have. Finally, my most sincere thanks must be reserved for the authors' contributions, for sharing their knowledge and expertise in a world where privacy and security are at stake every day.

This page intentionally left blank

https://doi.org/10.1142/9781800612860_0002

Chapter 1

The Latest Challenges in the Cybersecurity Field

Ioan-Cosmin Mihai, Costel Ciuchi, Gabriel Petrică

The high-speed evolution of our cyber environment prompts not only development opportunities for computerized society but also perils for its stability. The most poignant concern for all involved parties is the proliferation of information system vulnerabilities that can be exploited by malicious actors. Cyber threats have matured into global outbreaks due to their achieving increased complexity and detection avoidance capabilities (Mihai *et al*., 2018). Cybersecurity awareness and training are crucial, since attacks may impact critical infrastructure, and, due to their transnational, interconnected nature, vulnerabilities exploited in Member States can inflict losses throughout the entire European Union. Consequently, concerted action must be taken at the national and European levels to realize the full potential of prominent cybersecurity.

Cyberattacks' Analysis and Risk Management

Common cyberattack breakdown

The quintessence of a breach is that attackers must devise a means through which the target's cyber defenses can be penetrated, granting persistent access to the environment and allowing actors to act upon the system's equipment, data, or applications, thereby tampering with their

Figure 1. Cyber kill chain intrusion model.

Source: Mihai *et al*. (2019).

confidentiality, integrity, and availability. The *Cyber Kill Chain* model (Lockheed, 2021), as defined by Lockheed Martin researchers, characterizes this intrusion's structure (see Figure 1).

An intrusion model, essentially a well-defined itemized checklist, can be exemplified by this American military operational chain:

- *Target acquisition*: Identifying suitable adversaries.
- *Localization*: Establishing the target's geo-positioning.
- *Tracking*: Monitoring and scrutinizing their activities.
- *Preparation*: Assembling an appropriate toolkit for the mission.
- *Capture*: Engaging and taking the opponent into custody.
- *Evaluation*: Debriefing and quantifying the operation's effects.

This integrated procedure constitutes a "chain" (Varonis, 2021) because setbacks at any stage will disrupt the entire process. The steps of the cyberattack's intrusion model are outlined in the following schema:

The phases (Mihai *et al*., 2018) of an attack on cyberinfrastructure are typically comprised of the following:

- *Reconnaissance* — The research, identification, and selection of targets can combine elements of email address lookups, sifting through social media data, acquiring details on specific technologies in use, as well as exploring publicly available information from various online sources.
- *Weaponization* — Crafting malware or other kinds of exploits that, when inserted into the target's system, may allow remote access or otherwise disrupt security. Commonly used organizational documents, such as PDFs and Office files, can be employed to gain an advantage or further insight into the target system.
- *Delivery* — Implanting the weaponized software into the targeted environment. Primary avenues are electronic mail attachments, malicious website scripts, and detachable USB drives.
- *Exploitation* — Once delivered, exploitation occurs when the infected attachment is accessed by the user or the malicious link is viewed in a browser.

- *Installation* — Gaining access to the victim's machine ensures the attacker's continued presence and allows lateral movement to other high-value systems.
- *Command and Control* — Typically, the infection of one host establishes a command-and-control channel between the victim and the attacker's outside network. Once two-way communication has been achieved, the attacker can extend their activities within the target environment and can launch commands to exploit other systems.
- *Actions on objectives* — Once the other phases are concluded, attackers can proceed with collecting information from compromised systems, corrupting data, or performing attacks that undermine the availability of applications and services, while the infected machine can be used as a departure point to broaden their control over the target.

This type of attack is categorized as an Advanced Persistent Threat (APT), which represents one of the main types of cyberattacks in use today, alongside denial of service (DoS/DDoS) on critical infrastructure, such as electronic mail and web applications, adding compromised systems to bot networks, and ransomware attacks via malware infection.

Malware is cyberspace's most common threat, according to the latest statistics from the European Union Agency for Cybersecurity (ENISA) (see Figure 2).

The main forms of malware (ENISA, 2020a) can be categorized as follows:

- *Computer viruses* are software with predominantly destructive behavior designed to damage a computer system. They self-multiply and embed themselves in legitimate applications on the infected machine.
- *Trojans* are applications that trick the user into running them for a genuine purpose, while they attempt to exploit system vulnerabilities or open ports to allow remote access.
- *Computer worms* are released to infect and then propagate further through the network often slowing down or disabling afflicted systems.
- *Adware* is a category of applications which are generally free but aggressively display ads to the user and may attempt to trick them into installing additional adware.
- *Spyware* secretly records and transmits various types of data about the machine, the user's activity, and their personal information.
- *Ransomware* is a type of malware that encrypts user data, attempts to replicate onto other systems, may restrict access to them entirely, and

Top Threats 2019-2020		Assessed Trends	Change in Ranking
1	Malware	---	---
2	Web-based Attacks	---	↗
3	Phishing	↗	↗
4	Web application attacks	---	↙
5	Spam	↙	↗
6	Denial of service	↙	↙
7	Identity theft	↗	↗
8	Data breaches	---	---
9	Insider threat	↗	---
10	Botnets	↙	↙
11	Physical manipulation, damage, theft and loss	---	↙
12	Information leakage	↗	↙
13	Ransomware	↗	↗
14	Cyberespionage	↙	↗
15	Crytojacking	↙	↙

Figure 2. ENISA top threats 2019–2020.

Source: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2020-list-of-top-15-threats.

requires a non-traceable payment to remove its effects. Completing the payment does not guarantee removal, although most ransomware developers will unblock access and decrypt in order to generate further payments from other victims, via security research articles or even word of mouth.

- *Rogueware* is a type of software which deceives users into thinking they have no choice but to pay in order to eradicate a fake infection of their machine. Most often, this type of application claims to remove malware but, in fact, installs other undesirable software.
- *Scareware* is an application that induces users' fear, uncertainty, and doubt with the goal of proliferating malicious software.

A *(Distributed) Denial of Service (DoS/DDoS)* attack (CloudFlare, 2021) aims to compromise the operation of specific Internet-facing services. One of the most common types of DDoS attacks is the packet flood

(SRI, 2021), whereby an inordinate number of packets are sent to the target system with the goal of consuming all open connections and overloading network usage, leading to the crippling or cessation of services running on that system.

Attacks involving email have increased exponentially in recent years (ENISA, 2020b). These attacks can be classified as follows according to the goal of the malicious actor:

- *Email bombing* consists of sending multiple emails with large attachments to specific email addresses, leading to the user's quota being reached, which renders the account inaccessible.
- *Email spoofing* implies sending emails with a modified sender address. This is used to obscure the true identity of the sender in order to steal information or deceive the victim into performing certain actions.
- *Email spamming* is the process of sending unsolicited emails with commercial content, with the purpose of tricking recipients into accessing certain sites, subscribing to newsletters, purchasing products or services of questionable quality, or simply confirming that the email address is actively used and monitored.
- *Email phishing* is a flourishing attack in which messages are sent to defraud recipients of information on passwords, credit cards, bank accounts, or other personal details.

Attacks on web applications (ENISA, 2020c) have been thriving due to the remarkable development of web technologies that enabled the design of dynamic, interactive content with consistently high user engagement. Such platforms often contain vulnerabilities that can be exploited, allowing cybercriminals to bypass security measures and gain unauthorized access to their data. The most common attack varieties are the following:

- *SQLi*: Structured Query Language injection allows an actor to modify the logic of the query that is transmitted to the database, granting them the ability to avoid authentication mechanisms or expose data they do not normally have access to.
- *Cross-Site Scripting (XSS)*: The attacker inserts or alters scripts that are executed in the victim's browser every time they visit an infected site.
- *Cross-Site Request Forgery (CSRF)*: The attacker abuses established trust relationships between authenticated users and web applications in

order to take control of a victim's session, granting complete access to the user's account.
- *Man in the Middle (MitM)*: The attacker intercepts communication between a user and a web server, with a view toward capturing unencrypted data.

Advanced persistent threats represent complex, lengthy cyberattacks, often over months or years, that target specific organizations (Imperva, 2021) in order to compromise their systems and retrieve information. Government installations, military infrastructure, companies, and individuals are all valid targets. Such endeavors are generally undertaken by terrorist organizations or nations with significant technological capabilities and considerable financial resources required for a cyber operation (Hutchins *et al*., 2010) of this magnitude.

Cybersecurity risk management

In the specialty literature, risk management is defined as "the identification process of vulnerabilities and threats within an organization and the development of measures to minimize their impact on information resources." Most of the institutions and companies focus on physical protection and fail to determine the effects on the most important resources (Mihai *et al*., 2018). Risk management is a process that permits the management level to secure the balance between the costs, the financial resources for implementing the measures of protection, and the achievement of resource protection objectives which support the activity.

Risk management is "the process that allows IT managers to balance operational and financial costs of protection measures to achieve a gain in relation to the capability of protection of computer systems and data which are instrumental for the mission of the organization," according to the National Institute of Standards and Technology (NIST). Risk management (NIST, 2021) planning is the process of deciding how to administrate the activities of risks. A list of activities and a coordination of potential risks in risk-free, low-, and high-risk activities is needed to be done.

The risk management process consists of the following steps:

- *Risk assessment* — Identification and analysis of risks that can affect the institutions.
- *Coordinating the decision-making process* — Finding and assessing the control measures considering the cost–benefit ratio.

- *Implementing controls* — Implementing and running control measures to reduce the risks.
- *Measuring the performance of the program* — Analyzing the performance of the control measures and verification of the amount of protection.

Risk analysis is one of the most important aspects of cybersecurity (Logsign, 2021). The institutions and companies need to address the following steps:

- Identify and analyze the information.
- Identify and analyze the threats.
- Evaluate the vulnerabilities.
- Evaluate the cybersecurity risks.

The process of analyzing the risks involves identifying and classifying the security risks, finding the magnitude of the risks, and identifying areas with high risks. Risk assessment is the result of a risk analysis process. Risk analysis is a method used for evaluating the impact of risk on possible decisions in each situation. The purpose of this approach is to guide the decision makers to solve the issues of a certain level of uncertainty (Mihai *et al*., 2015) (see Figure 3).

Adopting cybersecurity controls to protect the computer systems without an adequate evaluation of cybersecurity risks generates overprotection

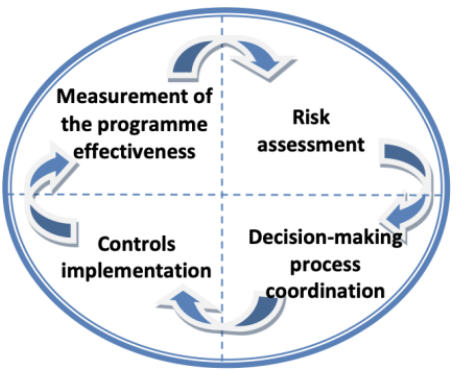


Figure 3. The risk management process.

Source: Mihai *et al*. (2015).

of resources, making cybersecurity an obstacle to the unfolding of operational processes or inadequate safety that will expose the resources of the institutions and companies to different threats.

The European Framework in the Field of Cybersecurity

Protecting critical infrastructure at the national level has been steadily elevated to a major concern that transcends technical boundaries, as the plethora of breaches in recent years firmly entrenched this threat on the public agenda of governments.

Consequently, the European Union has already pursued a series of steps to bolster resilience and preparation where cybersecurity is concerned. EU's Cybersecurity Strategy (European Council, 2021), unveiled in 2013, establishes practical measures and strategic objectives that underpin a reduction in cybercrime, an increase in the development of cyber-defense capabilities, and a concrete international policy regarding cyberspace. The adoption of the Directive on Security of Network and Information Systems (NIS) and the second mandate of the European Union Agency for Networks and Information Security were other measures which aimed to address this escalating issue.

Additionally, the Communication on "Strengthening Europe's cyber resilience and encouraging a competitive and innovating cybersecurity sector," adopted in 2016 by the European Commission, brought new measures to reinforce cooperation and knowledge transfer between Member States. This set forth a plan to establish a security certification framework for Information and Communications Technology (ICT) products and services, in order to strengthen the security of the digital single market and the confidence of those who are utilizing it. Particular importance is placed on this certification by the rising number of technologies requiring a commensurate degree of cybersecurity, such as automated industrial control systems, smart electric vehicles, or electronic health equipment.

The Cybersecurity Package (Digital Strategy, 2021a), proposed in October of 2017 by the European Parliament and the Council of the European Union, contains the ENISA Regulation, "EU Agency for Cybersecurity," the repeal of EC Regulation 526/2013, and the cybersecurity certification for ICT. The Regulation's proposal offers the following

comprehensive set of actions grounded in past experience that highlight specific measures which synergize with each other:

- Enhancing the training level and countermeasures of Member States and state companies.
- Facilitating coordination and cooperation between EU bodies and the various agencies and institutions of Member States.
- Bolstering EU-wide capabilities of complementing Member States' actions in the event of a major cross-border cyber threat.
- Raising awareness with individual citizens and the private sector regarding cybersecurity.
- Augmenting transparency on cybersecurity measures and certification of ICT products and services, thereby boosting confidence in digital innovation and the digital single market.
- Preventing fragmentation of certification systems and related security controls in the EU, as well as of evaluation criteria in all sectors of Member States (European Parliament, 2013).

The draft regulation reviews the current ENISA mandate and designates a renewed set of functions and goals aimed at ensuring the safety of cyberspace as an ongoing effort of Member States, EU institutions, and other stakeholders.

Therefore, the newly proposed mandate bestows the agency a more prominent position, particularly in bringing aid to Member States as they implement the NIS Directive, deploying active initiatives against specific threats, and elevating it to a center of expertise in the utilization of the cybersecurity certification by Member States and the EU Commission.

The fourth industrial revolution requires the implementation of the Cybersecurity Strategy in order to meet and exceed targets involving cybersecurity, as well as a review of the existing cybersecurity approach to the digital single market.

The European Cybersecurity Strategy

As its strategic objectives for 2016–2020 were derived from relevant contributions of Member States and communities, European regulations, and private sector input, ENISA proposes the following preeminent directions of development (European Parliament, 2016):

- *Expertise* — Collect, analyze, and disseminate data on the key aspects of the NIS Directive with potential impact on the EU.
- *Policy* — Promote network and information security by aiding EU and Member State institutions in drawing up and implementing relevant legislation on NIS as a priority of EU policy.
- *Competence* — Update and diversify security capabilities and coordination already in place at the European level in order to upgrade safeguards and incident response.
- *Communities* — Endorse EU's IT community by consolidating EU cooperation between Member States, relevant EU bodies, and the private sector.
- *Engagement* — Enhance ENISA's institutional coordination by optimizing resource management with stakeholders, including international ones.

The NIS and NIS2 Directives

EU Directive 1148/2016 (NIS), adopted by the European Parliament and the Council of the European Union on 6, July 2016, is comprised of policy and actions for elevated security of networks and information systems. The first pan-European legislation on cybersecurity (European Parliamentary Research Service, 2021), the Directive on the Security of Network and Information Systems, focuses on augmenting cyber authorities at the national level, increasing coordination among them, and specifying security requirements for key industry sectors.

The aim of this Directive is to ensure a high common level of security of networks and information systems in the EU and ask operators, respectively, digital services providers, to adopt adequate measures to prevent cyberattacks and for risk management and to report serious security incidents to competent national authorities (European Parliament, 2013).

In order to ensure the adoption of this common, high-security framework by digital service providers, to warrant proper risk management and the reporting of severe security incidents to the proper authorities, the NIS Directive:

- Sets forth an obligation to adopt a national strategy on NIS security.
- Creates a group meant to foster strategic cooperation and information exchange between Member States while boosting confidence in each other.

- Develops a network of rapid response teams to promote quick and effective operational capacity during cybersecurity incidents.
- Establishes notification and security requirements for essential service operators and digital service providers.
- Highlights the Member States' obligation to designate proper CSIRT authorities (Computer Security Incident Response Team) at the national level, acting as a single point of contact with powers and competence related to network and information systems' security.

Member States must ensure that essential service operators (KPMG, 2019):

- Implement adequate technical and organizational controls to manage risks to the information networks and systems they employ.
- Take proper measures to prevent and minimize the impact of incidents affecting NIS security.
- Involve CSIRT entities and proper authorities in incidents with a significant impact on the continuity of critical hosted services.

Additionally, the Directive underpins the following measures which must be implemented by each Member State:

- Adoption of a national strategy on NIS security, which sets strategic objectives and adequate political and regulatory procedures.
- Designation of one or more competent national authorities on network and information systems' security.
- Appointment of a single contact point on NIS security.
- Pledge that CSIRT points of contact and other competent authorities receive incident notifications which are aligned with the present directive.

As a response to emerging threats posed by the surge in cyberattacks and digitalization of key services, on the 16th of December 2020, the Commission has submitted a revised NIS Directive proposal to address the security of supply chains, streamline reporting obligations, strengthen security controls, and introduce more stringent supervisory measures and stricter enforcement requirements, which include harmonized sanctions across the EU. NIS2's proposed expansion of scope would raise the

long-term level of cybersecurity in Europe by effectively forcing more sectors and entities to take appropriate steps (European Parliament, 2021).

The General Data Protection Regulation (GDPR)

With its adoption on the 27th of April 2016, EU Regulation 2016/679 (General Data Protection Regulation — GDPR) on the protection of individuals with regard to the processing of personal data and the free movement of such data, alongside the repeal of Directive 95/46/EC, the European Parliament and the Council codified the protection of data with personal character as an important aspect of security at the EU level. GDPR entered into force on 25th, May 2016, and its provisions are applicable in all EU Member States, reaching legally binding status as of 25th, May 2018.

Both Chapter II, Article 8 of the EU Charter of Fundamental Rights and Article 16 of the EU Treaty enshrine data protection for people in the community as basic elements of the fundamental rights of a person.

The GDPR Regulation introduces several significant changes by expanding upon the guaranteed rights of people whose data are processed and streamlining administrative formalities for operators who process personal data. It also broadens the scope to include data operators located outside the Union, should they process the personal data of Community citizens, and grants operators the ability to interact with a single supervisory authority in the State in which they are incorporated.

To wit, binding requirements (EUR-Lex, 2016) have been outlined with regard to the following:

- The ability to obtain comprehensive information on the purpose and legal basis of the data processing.
- The data storage period and associated rights.
- The *right to be forgotten*, applicable online (except where it is necessary to ensure freedom of expression and the right to information, for compliance with a legal obligation to perform a task of public interest).
- The operator's obligation to demonstrate consent for personal data processing.
- Portability of data refers to the option of requesting the transfer of data to a different operator and the current one's mandate to automatically transfer data in a structured, machine-processable format.

Another novel aspect is the inclusion of active cooperation between supervising authorities in different states, when the data processing under scrutiny involves citizens from multiple EU countries, empowering the designated authority in the affiliated state to communicate with their counterparts, to oversee that information is being handled correctly under the established Regulation. Also, it imposes better accountability (ICO Org, 2021) on data operators based on the impact study conducted on the associated risks of personal data processing and the category it belongs to.

An adequate risk management and a fair estimate of the impact it may have on the person, as well as on the data warehouse and its operator, will serve as the foundation for the technical and organizational action plan required to avoid and address incidents.

The risk assessment on data protection implies the following:

- A description and purpose of the processed data.
- An evaluation of the proportionality and necessity of the performed data processing.
- An appraisal of potential risks to the rights and freedoms of the data subjects.
- A plan to address risks and ensure compliance with the GDPR provisions.

The impact assessment on data protection references the following:

- Achieving proper observation of private life through personal data processing.
- Considering the impact on the private life of targeted individuals.
- Testifying to the proper observation of the fundamental principles of the Regulation.

Privacy by design and *privacy by default*, two new concepts, were developed in this regard. The Confidentiality principle, *data protection by design*, relies on implementing confidentiality and data protection in the initial stage of the system design process and is a strategy preferable to an attempt to adapt an existing product or service later in its lifecycle.

The main obligations derived from GDPR pertaining to data operators are (ICO Org, 2021) as follows:

- *Designation of a Data Protection Officer* (legally binding since 25, May 2018, pursuant to articles 37–39 of the General Data Protection Regulation relating to public authority).

- *Mapping of personal data processing* (clarifying purpose and legal basis for the data processing, data type classification, and operational record-keeping).
- *Prioritization of the actions to be undertaken* (based on the identified risks to the rights and the freedoms of the data subjects).
- *Risk management* (classification of processing activities, accounting for the nature of the data, the scope, the context and purposes of processing, and the technologies involved).
- *Organization of internal procedures* (outlining procedures for gathering consent and guaranteeing that data protection is upheld throughout the process, with careful consideration of all events that may arise as a result).
- *Privacy by design* (mandatory features addressing data protection included in the design phase).
- *Privacy by default* (enforcing adequate measures which safeguard that only the data necessary for each specific processing goal are collected, with a plan of preparing dissemination of information with parties involved in the processing of personal data).
- *Safeguarding secure processing and confidentiality* through the adoption of proper organizational actions including the following, among others:
    - Encryption and anonymization of personal data (storing collected data in a manner that no longer allows one to uniquely identify an individual).
    - Continuously offer assurances of confidentiality, integrity, availability, and the resilience of processing services and mechanisms.
    - The ability to restore, within a reasonable time frame, the availability of personal information and access to it in the event of an undesirable physical or technical event.
    - A process for periodically gauging the effectiveness of the technical and organizational measures meant to ensure secure data processing.

The EU Cybersecurity Act

This Act empowers the EU Agency for Cybersecurity (ENISA) and outlines a cybersecurity certification framework for ICT products and services.

The EU Cybersecurity Act confers the agency a permanent mandate and grants it additional tasks and resources. ENISA will hold a key position in setting up and maintaining the European cybersecurity certification framework by preparing the technical groundwork for specialized certification schemes. It will be responsible for the public awareness of these certification schemes and issued certificates via its dedicated website. ENISA is mandated to increase operational cooperation at the EU level, assisting Member States who desire aid in handling their cybersecurity incidents and supporting EU coordination in the event of large-scale, transnational cyber threats or crises (Digital Strategy, 2021b).

Companies operating in the EU will benefit from performing a single, standardized certification of their ICT products, processes, and services, recognized EU-wide by means of the cybersecurity certification framework.

Conclusion

Threats to information systems are characterized by their increasingly prevalent global character and dynamic aspects, which compounds their difficulty to be identified and counteracted. Within the last few years, cyberattacks have experienced an explosive diversification, some having been categorized as global epidemics due to their rapid propagation in the online environment.

Due to their nature as a primarily human issue, cybersecurity assurance cannot be achieved solely through technical means. While protective measures are plentiful and of varying efficacy, security incidents are often caused by inadequate configuration or deployment of security policies rather than by security system failure. Consequently, it is imperative to develop and implement strategies and campaigns on cybersecurity by defining policies which prevent and combat cybercrime at the national level.

Public policy must make research and education in the field of cybersecurity its priorities. Facilitating information security research, reforming education, and promoting a trained workforce are crucial to reaching the cybersecurity policy objectives. Research and education policies will only be effective so long as they include the multifaceted and multidisciplinary nature of cybersecurity as a ubiquitous, fundamental element in culture, techniques, systems, processes, and technical infrastructure.

International cooperation is of high importance in this field, as cybersecurity challenges transcend geographic boundaries, affecting globally interconnected systems. Whether referring to government institutions, private companies, research centers, or educational establishments, cooperation and collaboration between European and International entities are vital (Mihai *et al*., 2018), as efficient communication between organizations, institutions, and the cybersecurity community may prove momentous in finding and resolving security exposures. The coordinated disclosure of vulnerabilities is a mechanism with a demonstrable track record in this regard.

Adopting coherent public policies at the Member State level with respect to coordinated vulnerability disclosure and joint trans-sectoral actions and cooperation mechanisms will offer the appropriate ecosystem to establish good security practices. Facilitating communication channels, creating working groups, and encouraging public consultation involving civil society and public–private partnerships are key directions that public policy should concentrate on.

Subsequently, producing comprehensive, reviewed, and updated cybersecurity legislation to assist the development of national defense capabilities must be a priority of each Member State. Promoting a secure cyberspace is the responsibility of both the state and the proper authorities, the private sector, and civil society. For the proliferation of adequate cybersecurity culture, the most effective levers are research and education, cooperation mechanisms at the European level, and public–private partnerships.

References

CloudFlare (2021). What is a DDoS attack? *CloudFlare*. https://www.cloudflare.com/learning/ddos/what-is-a-ddos-attack/.

Digital Strategy (2021a). Cybersecurity package ‘resilience, deterrence and defence: Building strong cybersecurity for the EU’. https://digital-strategy.ec.europa.eu/en/library/cybersecurity-package-resilience-deterrence-and-defence-building-strong-cybersecurity-eu.

Digital Strategy (2021b). EU Cybersecurity Act. https://digital-strategy.ec.europa.eu/en/policies/cybersecurity-act.

ENISA (2020a). ENISA threat landscape 2020 — Malware, European Union agency for cybersecurity. https://www.enisa.europa.eu/publications/malware/.

ENISA (2020b). ENISA threat landscape 2020 — Phishing, European Union agency for cybersecurity. https://www.enisa.europa.eu/publications/phishing/.

ENISA (2020c). ENISA threat landscape 2020 — Web application attacks. European Union agency for cybersecurity. https://www.enisa.europa.eu/publications/web-application-attacks/.

ENISA (2020d). ENISA threat landscape 2020 — Spam. European Union agency for cybersecurity. https://www.enisa.europa.eu/publications/spam/.

EUR-Lex (2016). General data protection regulation. https://eur-lex.europa.eu/eli/reg/2016/679/oj.

European Council (2021). Cybersecurity: Council adopts conclusions on the EU's cybersecurity strategy. shorturl.at/dBHW5.

European Parliament (2013). Regulation of the European Parliament and of the Council on ENISA, the "EU Cybersecurity Agency," and repealing Regulation (EU) 526/2013, and on Information and Communication Technology cybersecurity certification ("Cybersecurity Act").

European Parliament (2016). Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the union, Official Journal of the European Union.

European Parliament (2021). The NIS2 directive: A high common level of cybersecurity in the EU. https://europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI(2021)689333.

European Parliamentary Research Service (2021). The NIS2 directive. https://epthinktank.eu/2021/02/22/the-nis2-directive-a-high-common-level-of-cybersecurity-in-the-eu-eu-legislation-in-progress/.

Hutchins, E. M., Cloppert, M. J. and Amin, R. M. (2010). Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains. Lockheed Martin Corporation, US.

ICO Org (2021). Guide to the general data protection regulation. https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf.

Imperva (2021). What is an advanced persistent threat. https://www.imperva.com/learn/application-security/apt-advanced-persistent-threat/.

KPMG (2019). Complying with the European network information service (NIS) directive. https://assets.kpmg/content/dam/kpmg/nl/pdf/2019/advisory/complying-with-the-eu-nis-directive.pdf.

Lockheed (2021). Cyber kill chain. https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html.

Logsign (2021). The main elements of a security risk analysis. https://www.logsign.com/blog/the-main-elements-of-a-security-risk-analysis-report/.

Mihai, I. C., Petrică, G., Ciuchi, C. and Giurea L. (2015). Cybersecurity challenges and strategies. Ed. Sitech.

Mihai, I. C., Ciuchi, C. and Petrică, G. (2018). Current challenges in the field of cybersecurity — The impact and Romania's contribution to the field. Ed. Sitech.

Mihai, I. C., Prună, Ș. and Barbu, I. D. (2019). Cyber kill chain analysis. *International Journal of Information Security and Cybercrime*, 3(2), 37–42, https://www.ijisc.com.

NIST (2021). Risk management framework. https://www.nist.gov/cyberframework/risk-management-framework/.

SRI (2021). Hybrid warfare and cyber-attacks. *Intelligence Journal*. http://intelligence.sri.ro/razboi-hibrid-si-atacuri-cibernetice/.

Varonis (2021). What is the cyber kill chain and how to use it effectively. https://www.varonis.com/blog/cyber-kill-chain/.

https://doi.org/10.1142/9781800612860_0003

Chapter 2

Governance of Cyberspace — El Dorado of States and Private Actors

Dragos Nicolae Costescu

Before delving into this research, we feel the need to explain our choice for the expression "El Dorado." According to Collins Dictionary (Collins Dictionary Website, 2021), El Dorado means any place of great riches or fabulous opportunities and this is exactly what cyberspace represents to state and non-state actors. Each and every one wants a piece of this "cake."

This chapter aims to highlight the challenges regarding cyberspace governance. The first part analyses the particularities of cyberspace, the second part addresses the issue related to the regulatory framework and governance of cyberspace governance also in relation to state sovereignty, and finally, the third part stresses the existent struggle for power between state and non-state actors in cyberspace. The cases of distributed governance, multilateral governance, and multistakeholderism vividly demonstrate the concerns related to the use and control of cyberspace that states face.

With the technological evolution, cyberspace has become a new strategic area of interaction between various actors, mainly States in terms of sovereignty. The growing importance of such a space, for both state and non-state actors, makes its regulation by international law definitely necessary but yet hard to achieve for the reasons we will develop later on.

Particularities of Cyberspace in Terms of Governance

The concept of cyberspace does not correspond to the classic definition of a territory geographically speaking, but it is the representation of a new space, which can vary according to the actors within. There is actually no universal definition of cyberspace. On the contrary, there are dozens of them which struggle to grasp all of the dimensions that make it so unique. Cyberspace is first and foremost an information space generated by the global interconnection of information and communication systems, in which data are created, stored, and shared. The term designates both the physical infrastructure which is at the source of this environment, namely, the various elements that make up the Internet, such as cables, servers, routers, satellites, and all connected devices that are anchored in the physical and political geographic territory, and the intangible space in which data, information, and ideas circulate the space in which interactions occur between individuals behind their screens everywhere in the world at almost instantaneous speed (Emerson, 2016).

However, cyberspace, unlike other *global commons*[1] (for example, outer space or the high seas) which are tangible, is entirely *artificial*, non-tangible, and non-space limited, created and conceived by mankind. It is also *versatile*, since computer programs, tools, software, and users often change location, function, and even identity, thus generating anonymity, a characteristic which proves of utmost importance for cyberspace activities. Despite this, the specificity of cyberspace does not justify its exemption from the law, taking into account that whenever man has been able to expand his ability to conquer new spaces (airspace or the Antarctic), international law has both found application and an explanation (ESIL-SEDI Website, 2021).

As previously stated, cyberspace, which we sometimes imagine as virtual, is actually based on massive and expensive infrastructure that supports networks (Lessig, 2006), of which the Internet user has little awareness since all these data circulate in optical fibers and switch at a speed close to that of the light. The physical elements, i.e., satellites, computers, cables, networks, and routers, are owned by individuals, governments, organizations, and Internet service providers.

[1]According to Lexico by Oxford dictionary, global commons can be defined as *the earth's unowned natural resources, such as the oceans, the atmosphere, and space.* '*financial speculators and other abusers of our global commons*'.

It is of utmost importance in our view to distinguish between two notions we have previously made use of: "Internet" and "cyberspace." The Federal Networking Council (FNC) issued in 1995 the following definition of the term "Internet": "*Internet*" *refers to the global information system that (i) is logically linked together by a globally unique address space based on the Internet Protocol (IP) or its subsequent extensions/follow-ons; (ii) is able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP) suite or its subsequent extensions/follow-ons, and/or other IP-compatible protocols; and (iii) provides, uses, or makes accessible, either publicly or privately, high-level services layered on the communications and related infrastructure described herein* (It-law.at website). The American Supreme Court classified the Internet as *a unique and wholly new medium of worldwide communication. [...] Taken together, these tools [emails, mailing list servers, newsgroups, chat rooms, and World Wide Web] constitute a unique new medium — known to its users as* "*cyberspace*" *— located in no particular geographical location but available to anyone, anywhere in the world, with access to the Internet* (Menthe, 1998).

In addition, the physical network "Internet" — consisting of its cables and satellite mains connected to backbones and local networks — is subdivided by topography and political borders, whereas cyberspace seems to be borderless, independent of physical location, with no territorially based boundaries. The location within cyberspace consists only of the IP addresses of the computers in the physical network (Murray, 2012).

As cyberspace forms an international space for information and data distribution, its governance seems extremely important. *Governance is the sum of the many ways individuals and institutions, public and private, manage their common affairs. It is the continuing process through which conflicting or diverse interests may be accommodated and cooperative action may be taken. It includes formal institutions and regimes empowered to enforce compliance, as well as informal arrangements that people and institutions have agreed to or perceive to be their interest* (Our Global Neighborhood, 1995). States may thus be examining the option of creating "national cyberspaces," building trans-oceanic cables and store Internet data and information on servers within their national territories, with the view of ensuring their protection and integrity. Nevertheless, we have to consider the utility of data localization (Chang, 2017). Storing data and information on national territories does not necessarily make them invisible to foreign hackers, for example. It is not geography but

mainly technological advancement that defines security in this virtual world.

However, "national cyberspaces" are hard to create taking into account as previously stated the ownership of the physical infrastructure that makes the Internet. While some satellites may serve military and government functions, others serve commercial purposes solely and are owned by private companies (non-state actors). Of course, countries can grant licenses to companies to install cables, and, depending on the government, they may take steps to regulate, monitor, and set standards for cyberinfrastructure, but this does not mean they exercise total sovereignty over them. In addition, when we talk about cyberspace governance and regulation, we should bear in mind a double-sided approach: the regulation within national borders (also double-sided approach: first, the regulations applicable to the physical infrastructure and the rules applicable to ISP or to companies owning the infrastructure and second, the regulation of acts happening in cyberspace, for example, the regulation of freedom of expression or of cybercrimes identified within the borders of a country) and the regulation of international cyberspace (from the same double-sided approach).

The governance and regulation of international cyberspace are the object of our study, and if we were to follow traditional approaches of public international law (Brownlie, 1998), the solution for Internet governance should reside in an international treaty establishing an international governmental organization. However, as we will prove later on, this is an extremely difficult task to achieve, and the role of non-state actors (companies, civil society, and academia) becomes more and more pregnant, moving toward a multilevel governance model (Proksch and Schweighofer, 2011).

The Non-Existence of a Regulatory Framework for International Cyberspace Governance

Cyberspace is not an area of lawlessness, being subject to the rules of international law, although the details have yet to be defined (Lazar, 2017). This principle was recorded in the report of the UN (United Nations) Group of Governmental Experts (UN GGE), as well as in the G20 and the G7.[2] Standards of responsible behavior of States —

[2]The Group of Seven (G7) and the Group of Twenty (G20) are **informal governance clubs** which hold annual Summits of Heads of State to discuss issues of global importance.

admittedly non-binding — and confidence-building measures have also been adopted by States to regulate this space and prevent the risk of conflict escalation.

Since 2004, the UN has established (at Russia's initiative in the 98s; Russia has been pushing for the U.N. to have a significant role in the governance of technology for decades and has been a vocal proponent of the need for an internationally binding cyber treaty) a group of government experts on cybersecurity (previously mentioned above as UN GGE) that has met five times and produced consensus documents in 2010, 2013, and 2015, articulating a set of norms for responsible state behavior in cyberspace. In 2013, these experts agreed that international law was applicable to cyberspace. In 2017, however, the round of negotiations broke down. Inter-state regulation attempts were, therefore, at a standstill, which has led some in wanting to officially integrate non-state actors into Internet governance. The failure of the GGE in 2017 paved the way for a different approach, which resulted in two resolutions being passed by the U.N. in late 2018. One was a renewal of the GGE, and the other was the creation of the Open-Ended Working Group, which also focused *on developments in the field of information and telecommunications in the context of international security*.

We may notice two opposing approaches to ensuring stability and security in cyberspace. The first, which prevailed during UN GGE, being also supported by the United States, is based on political commitments. Its working hypothesis is that the existing international law is sufficient to regulate behavior, and its objective is to "translate the expectations of the international community" by adopting standards of responsible behavior between States. However, the limits of such an approach are quickly reached because the standards are non-binding and are endorsed as a unilateral commitment. The second is legalistic (Liaropoulos, 2017), the objective being to create new rights and obligations in order to adapt international law to the realities of cyberspace, based on the principle that the existing international law cannot regulate and govern state behavior in cyberspace due to its particularities previously discussed (Kremer and Müller, 2014). This second path is being endorsed, for example, by the Member States of the Shanghai Cooperation Organization. It is these two approaches that led to the creation in 2019 (A/RES/73/27 UN Resolution) of Open-Ended Working Group (OEWG) and the renewed mandate of the GGE (A/RES/73/266 UN Resolution).

GGE[3] brings together 25 Member States (including France, Russia, China, and the United States) while the OEWG brings Member States (including China, Russia, the United States, and France) and non-state actors, whether industrial (such as Microsoft and Kaspersky), academics (e.g., National Law University Delhi's Center for Communication Governance), or non-governmental organizations (e.g., the Internet Society). Both groups have a similar plan: to develop (or change) standards, laws, and principles; build trust between actors and increase the general level of cybersecurity; and work on the application of the existing body of international law to cyberspace.

The OEWG adopted very recently a report, that although not binding, which shows a consensus comprising all 193 of its Member States. The report outlines positive engagement in the working group on areas that everyone could agree on, with emphasis on the need for cyber capacity-building and on norms for responsible state behavior in cyberspace.[4] That includes enhancing the capability to defend against cyberattacks as well as the capacity to engage in international debates on cyber issues. This document brings health care facilities into consideration of "critical infrastructure," along with the Internet's "public core" — that is, the architectural components of the networks on which everyone depends. It also acknowledges the harms caused by election interference, emphasizing that these "malicious activities … are also a real and growing concern." What is the most significant aspect of our view is that there is consensus among all U.N. Member States in a field that has been wrought with division and contention, especially in the past years (World Politics Website, 2021).

Also, for Internet governance, a charter or code of good conduct was proposed at the 66th session of the UN General Assembly (UN Official Website, 2021) by Russia and China on 14th, September 2011. A similar proposal has been renewed a year later at the International Telecommunication Union conference in Dubai in December 2012. These proposals were rejected by the United States since their vision of cyberspace regulation and governance is different. Thus, two visions of the Internet are opposed here, reflecting a geopolitical confrontation centered on a change in the governance of cyberspace: one, led by the United States, which considers that the Internet is a space of free movement but under their control, predominantly benevolent, and the other, led by Russia and China, who

[3] Group of Governmental Experts — UNODA.

[4] A Breakthrough for Global Cyber Governance (worldpoliticsreview.com).

want more state control, in the face of external and internal influences on their populations. A new code of conduct was proposed in 2015 by the Shanghai Cooperation Organization, but the notion of information security mentioned still reflected the inadmissible desire for population control.

Turning our attention away from UN to the Council of Europe level, we have here the Budapest Convention. The Budapest Convention is the first and only international convention to encourage harmonization of cyber laws and regulations and to build cooperation among nations in controlling cybercrime, being also open to non-Member States. It is currently the most accepted convention on cybercrime, but nevertheless, most countries in Latin America, the Middle East, and Asia-Pacific, including Brazil, Russia, China, and India, are not, which clearly reduces the effectiveness of the convention as it applies to less than half of the world's Internet users. Furthermore, these countries have at times argued that a UN treaty or code would be more appropriate. In 2012,[5] a new global cybercrime treaty was proposed by China, India, South Korea, and a number of other regional countries at the 12th UN Congress on Crime Prevention and Criminal Justice in Salvador, Brazil. Although the proposal did not gain much support from Western countries, it might provide a good basis for a new, more inclusive convention. However, to our view, taking a quick look at these countries' conceptions about cyberspace, we doubt that this treaty would manage to balance the interests of law enforcement and respect for fundamental rights provided by the Budapest Convention and would instead facilitate repression and censorship.

At the EU level, we bring to attention a Motion for a European Parliament resolution on regulating cyberspace from February 2016.[6] However, this document is not binding. More recently, in May 2019, the Council established a framework which allows the EU to impose targeted sanctions to deter and respond to cyberattacks which constitute an external threat to the EU or its Member States. This framework allows the EU for the first time to impose sanctions on persons or entities that are responsible

[5]United Nations (2010). Recent developments in the use of science and technology by offenders and by competent authorities in fighting crime, including the four Cybercrime in Asia: Trends and Challenges 63 case of cybercrime'. Working paper A/CONF.213/9, UN 12th Congress on Crime Prevention and Criminal Justice, Salvador, Brazil, 12–19 April 2010–22 January 2010. Accessed on 6, July 2010, from http://www.unodc.org/documents/crime-congress/12thCrime-Congress/Documents/A_CONF.213_9/V1050382e.pdf.

[6]MOTION FOR A RESOLUTION on regulating cyberspace (europa.eu).

for cyberattacks or attempted cyberattacks, who provide financial, technical, or material support for such attacks or who are involved in other ways, thus being a binding framework. Later on, in December 2020, the European Commission and the European External Action Service (EEAS) presented a new EU cybersecurity strategy, with the aim of strengthening Europe's resilience against cyber threats and ensuring that all citizens and businesses can fully benefit from trustworthy and reliable services and digital tools. The new strategy contains concrete proposals for deploying regulatory, policy instruments. The EU is also working on two legislative proposals to address current and future online and offline risks, respectively, an updated directive to better protect network and information systems, thus the infrastructure and a new directive on the resilience of critical entities.

In terms of organizations and institutions involved in the governance of cyberspace, the EU cooperates on defense in cyberspace through the activities of the European Defense Agency (EDA), in collaboration with the EU cybersecurity agency and Europol. The EDA supports Member States in building a skilled military cyberdefense workforce and ensures the availability of proactive and reactive cyberdefense technology (Consilium EU official website).[7]

While we can notice that the demand for governance is great, the prospect of an international comprehensive cyber treaty does not seem feasible, and for the time being, effective cyberspace governance seems to occur less in formal institutions and more on regional organizations and to rely on non-binding frameworks. The argument is that it is very difficult for a UN treaty to regulate the whole range of cyberspace (both at the national level and international level)-related issues (cyberwarfare, cybercrime, other kinds of cyber threats, and protection of infrastructure), and it seems more practical to focus on particular aspects, regionally or even at a national level. We must also bear in mind that a global UN treaty might be conceived by states as limiting their sovereignty.

And while we can count a handful of international specialized conventions that can be applied internationally for the governance of spaces like the sea or outer space (the 1967 Outer Space Treaty and the 1982 UN Convention on the Law of the Sea (UNCLOS)), there is currently no multilateral convention, governing cyberspace in international law. What is more, it is also difficult to identify any practice repeated over time that

[7]Cybersecurity: how the EU tackles cyber threats — Consilium (europa.eu).

would count as customary law applicable to this space or has any specific case law emerged in the area of cyberspace governance.

However, if we were to take a look at the infrastructure previously mentioned (satellites and cables), we might draw the conclusions that the Conventions mentioned above apply indirectly also to cyberspace and we might even affirm that cyberspace has borders since it is based on this physical infrastructure located in national territories (or subject to national jurisdiction if undersea or in outer space). State sovereignty and international law clearly apply, but how they apply still remains a subject of dispute.

The Stake for Actors in Cyberspace's Governance or "the Clash of Actors"

As cyberspace has become a central domain and source for international conflict, security and the role of states have become more and more important. Cyberspace is being conceived by both state and non-state actors as "fertile soil" for espionage and coercion, as a tool of power. Furthermore, cyberspace may nowadays be considered as the fifth domain of warfare, as critical to military operations as land, sea, air, and space.

So, if we were to ask ourselves what is the stake in governing cyberspace? What is the catch that makes it so desirable? With the advancement of technologies, cyberspace will be the future in terms of power, providing major opportunities for innovation, economic progress, and access to information. While its quick development has proved hugely useful, it also brings new threats,[8] and dangerous practices are developing in cyberspace: cybercrime, information manipulation, cyberattacks, economic espionage, theft of personal information or confidential data, compromise of information, and communications systems. These attacks can come from state or non-state groups that respect no borders and are becoming more and more sophisticated and intense.

The main threat for countries in cyberspace, for example, is the preeminence of the attack over defense. That is, today, it is easier to attack than to defend yourself with cybernetic weapons. The second threat is the great difficulty in attributing computer/machine attacks and finding who the perpetrator of an attack is. This is problematic in the classic international game

[8]Paris Call for Trust and Security in Cyberspace — Paris Call.

since it hinders a state's ability to legitimately defend itself. Third, digital weapons are increasingly volatile and constantly changing; new weapons are created and developed every day, so we might state that nowadays we get to talk about cyberwarfare more often. Furthermore, the cyberspace represents a means to control the masses through the info that reaches them over the Internet, the so-called media information. As a means of example, we might consider the suppressing of voices of marginalized and oppressed communities on platforms like Facebook, the removal of online content posted by activists in countries like Syria or Iraq, or even shutting down the Internet during protests like the Togolese Government did in September 2017 (European Parliament, Digital Technology Study, 2021). It should also be noted, that for smaller states, or states facing difficulties, the use of cyberspace can serve other objectives (Glen, 2014). For instance, by applying for (state) membership to the International Union of Communications (UIC) and to the Internet Corporation for Assigned Names and Numbers (ICANN) for the attribution of a country top-level domain, Palestine might perhaps be seeking an international "recognition" from the main institutions of cyberspace.

When it comes to non-state actors, private entities, the major challenge is again that of the dissemination of power. With cyberspace, power tools are no longer just the prerogative of states. They are also of non-state actors. Indeed, both very large technology companies like GAFAM in the United States and BATX in China and also hackers and criminal groups are now inextricably involved in cyberspace. In addition, they maintain very equivocal relations with States, between competition and cooperation or even collusion. Many pirates become "proxies," a sort of "computer corsairs," in the service of States which lean on their expertise, but who, above all, use these intermediaries to launch attacks against other states. Furthermore, states demand access to a massive amount of information collected and stored by telecommunication companies, Internet service providers, or giant tech companies like Facebook, asking all these non-state actors to assist in hacking operations or to provide them a way into encrypted end-to-end communications[9] (Scottishsun Website, 2021). And it often happens for these giants of tech to be resilient to government requests, making use of the right to privacy as an excuse.

[9]Facebook will give "gifts to terrorists" & child abusers by giving them total anonymity online, says MI5 boss (thescottishsun.co.uk).

So, in this clash of powers between actors, how can states exercise their sovereignty in cyberspace? And can we talk about sovereignty in cyberspace or should we consider it a global commons? Sovereignty can be defined as the basic international legal status of a state that is not subject, within its territorial jurisdiction, to the governmental, executive, legislative, or judicial jurisdiction of a foreign state or to foreign law other than public international (Aurescu, 2018).

There are several theories related to exercising sovereignty over cyberspace (West, 2014). In the early days of Internet development, governance was limited, unorganized, and restricted within online communities (distributed governance theory), who asserted that information had to be free and not controlled (Deibert and Crete-Nishihata, 2012). However, this approach illustrates an era where online communities were rather small (Liaropoulos, 2017) and even able to self-regulate (Tusikov, 2016). In 1996, John Perry Barlow, a founding member of the Electronic Frontier Foundation (EFF), went so far as to publish a "Declaration of cyberspace independence," in which he affirmed that cyberspace had its own sovereignty and that the laws of *Governments of the physical world do not apply in this civilization of the mind. We are forming our own Social Contract. This governance will arise according to the conditions of our world, not yours*' (Barlow, 1996). This portrayal continues to animate many hacktivists, who fight any attempt to impede the free flow of information on the Internet (Chang, 2017). Nowadays, Internet users are counted in billions and cyberspace has become an integral part of modern societies (Betz and Stevens, 2011). Cyberspace has clearly reached in our view an evolutionary phase where regulations are needed (Chang, 2017). The distributed governance model, although still popular in some online communities, cannot provide efficient policy solutions that would be acceptable to the entire community of cyberspace users. The argument that state sovereignty should have a limited role in cyberspace has also been embraced by those who view cyberspace as a global commons. In sharp contrast to land, sea, air, and space, cyberspace is a human-made domain that lacks physical space and thereby borders, as earlier contended. Cyberspace comprises a global infrastructure, but according to Cornish, it is not a global commons (Cornish, 2015). Cyberspace seems borderless but is actually bounded by the physical infrastructure previously explained that facilitates the transfer of data and information. Such an infrastructure is mostly owned by the private sector and is located in the sovereign territory of states. Paul Cornish labels cyberspace as a virtual commons that is neither private

property nor sovereign territory nor global commons in the same way that the sea and the air are considered to be (Cornish, 2015).

Even with self-regulation, one can see the influence of government in the form of constructing, shaping, promoting, and/or facilitating self-regulation (Tusikov, 2016). National regulations still manage against cybercrime and hate speech and prove really efficient (Barlow, 1996; Katyal, 2003). However, as mentioned earlier, state regulatory institutions have limitations when it comes to regulating cyberspace due to the decentralization and borderless character of cyberspace (Feick and Werle, 2010).

The issue of state sovereignty is of central importance to the supporters of multilateral governance. The multilateral theory argues that states should have the power to set their own regulations, the theory being supported by Russia, China, India, Iran, and Saudi Arabia. In the aftermath of the Edward Snowden disclosure (Chang, 2017), multilateral governance has scored more and more points for states that seek to protect their data from the surveillance systems of other countries and hackers (European Parliament, Digital Technology Study, 2021), companies like Google or Facebook being conceived as a threat to digital sovereignty of States (Nocetti, 2015).

According to the multilateral approach, Internet governance should respect the Westphalian notion of sovereignty, a concept according to which all nation-states have sovereignty over their territory, with no role for external interferences in domestic structures. The protection of cyber sovereignty and information security thus represents the main priorities for states that embrace the multilateral governance model. Embracing this theory and in the context of the fight against terrorism, many states have put in place procedures for accessing, blocking, and monitoring networks. China has developed dynamic strategies for controlling access and content in "its cyberspace," which it considers an area of sovereignty. The Chinese government, meanwhile, has created with the same ambition a separate specifically Chinese space in the form of a national intranet with a technical barrier, a sort of "Great Electronic Wall" or "Great Firewall of China," to block routing IP address or filter domain names (DNS). Russia has even given a name to what it represents as its sovereign Internet, the "RuNet." Today, this RuNet is characterized by its own infrastructure, cables, means of transporting information, and crossing points with the rest of the physical layer of the Internet. It is currently over 40% "fiber" and is one of the fastest in the world. Its content is mainly for Russian-speaking with national social networks, such as Vkontakte and

Odnoklassniki, whose servers are in Russia. Russia has developed its own search engines, Yandex (60.5% of Internet searches) and Rambler, competitors of Google which does not reign supreme in the Russian market unlike the European market. Since 2011, a national operating system has been under development, based on the open Linux system accessible more easily by its intelligence services on behalf of national security. Russia also announced in 2015 that it was embarking on a program to develop a mobile OS, in partnership with the Finnish start-up Jolla, on a national version of SailFish OS with the aim of bringing back the share of OS Android and iOS in the Russian market 95–50% by 2025 (Revue de medias Official Website, 2021).

Going to the third theory, embraced by us, the multistakeholder governance theory stresses that state and non-state actors that represent the business sector and civil society (like Microsoft, Apple, Google, Yahoo, Weibo, Skype, Dropbox, Amazon, Twitter, and Facebook) should cooperate in governing the cyberspace, arguing that governments alone cannot regulate cyberspace successfully. Supported by the US, the UK, Canada, Australia, and organizations like Google and ICANN (Glen, 2014), the multistakeholder theory has been quite popular in the pre-Snowden era. In the aftermath of the Snowden disclosure, the legitimacy and credibility of this approach have been considerably weakened or even abandoned (Deibert, 2015). However, nowadays, states seem more reticent to apply this theory and to share the "cake" with non-state actors, being thus reluctant to entrust their fundamental security to private actors. Big tech companies might have significant power over how cyberspace operates, but it is finally the states who govern and might overrule them since they have the resources to do so[10] (Orfonline Website, 2021).

Despite states' position, in our view, the cooperation between state and non-state actors is essential to face the risks previously mentioned, and there is a need not only to develop strategies to protect themselves from risks but also to take advantage of the opportunities offered by cyberspace. This requires not only technical measures to protect the systems and ensure their resilience but also a good risk assessment policy, an understanding of what constitutes strategic information, the establishment of risk management procedures, staff training, and the implementation of good practices, as threats will continue to grow. An example of good practices in terms of cooperation between state and non-state actors

[10]How the cyberspace narrative shapes governance and security | ORF (orfonline.org).

in the domain of the fight against cybercrime is one of the CERTs. CERTs are prominent, non-governmental organizations that share information on malicious cyber activities, providing an incident response to victims. They not only help safeguard information security within one country but also collaborate with other CERTs at international and regional levels.

Turning back thus to state actors, on 12 November 2018, on the occasion of the meeting at UNESCO of the Internet Governance Forum (IGF), the President of the French Republic, Emmanuel Macron, launched the *Paris Call for Confidence and Security in cyberspace*. This high-level declaration in favor of the development of common principles for securing cyberspace has already received the support not only of many States but also of private companies and civil society organizations. However, the French approach has not achieved consensus. In fact, only about 79 states have signed the Paris Call. In addition, several major players in the international system have not signed it, including the United States, China, and Russia, since they perceive it as a way to limit their hegemony in this area, which would result in control over their capacities.

As for non-state actors, in February 2017, Microsoft (Microsoft Official Website, 2021) called for the signing of a digital Geneva Convention and accused states of being responsible for the cyber-arms race. Today, the Paris Call is signed by the main American technological firms: Google, Facebook, Microsoft, IBM, Oracle, Cisco, and Intel, companies which, consequently, oppose the official position of their State of "origin." Above all, they affirm, on the one hand, their rise in power and, on the other hand, the dissociation of their interests from those of the American state. This is a very important dynamic because we know that these companies are historically linked to the state apparatus and, in particular, to the military and the intelligence community. This trend indicates a form of emancipation and empowerment fueled by the gradual disengagement of states in several areas (Jayawardane *et al.*, 2015).

To conclude, it seems that States compete with each other in their attempt to create norms and institutions that will shape the future of governance, but at the same time, they have to fill in the sovereignty gap and compete with the private sector. As a result, cyberspace governance is still under construction.

Conclusion

States, giant tech companies, transnational corporations, civil society groups, non-governmental international organizations, and even associations act in cyberspace and struggle for power in this readjustment of territorial sovereignty principles. The clash of interests of all these actors makes it difficult to govern and to accept a legally binding framework, and it seems that this ungovernability of the cyberspace is leaning on leading more and more non-state actors to fill the void left by states. Furthermore, apart from the slow movement actors, it is also the technological advancement that makes this space hard to regulate. The appropriate institutional governance model configuration for cyberspace will vary over time of course, depending on the capacities and willingness of the participant actors. However, efforts made by the non-state actors may in some situations compensate for shortcomings on the part of state actors, but to our view, as previously shown, the cooperation between these two actors is the direction in which we should be heading.

References

A/RES/73/266 UN Resolution. A/RES/73/266-E-A/RES/73/266-Desktop (undocs.org).

A/RES/73/27 UN Resolution. A/RES/73/27-E-A/RES/73/27-Desktop (undocs.org).

Aurescu, B. (2018). *Drept International Public*, pp. 79–83. Bucuresti: CH. Beck.

Barlow, J. P. (1996). A declaration of the independence of cyberspace. https://www.eff.org/cyberspace-independence.

Betz, D. and Stevens, T. (2011). Cyberspace and the state. Toward a strategy for cyber-power (Adelphi Paper 424). Oxon: IISS, Routledge.

Collins Dictionary Website (2021). Collins Online Dictionary | Definitions, Thesaurus and Translations (collinsdictionary.com).

Commission on Global Governance (1995). Our global neighborhood: The report of the Commission on Global Governance. Retrieved from Oxford University Press website: http://www.gdrc.org/u-gov/global-neighbourhood/chap1.htm available also at: https://archive.org/details/ourglobalneighbo00comm/page/n3/mode/2up.

Consilium Website. Cybersecurity: How the EU tackles cyber threats — Consilium (europa.eu).

Cornish, P. (2015). Governing cyberspace through constructive ambiguity. *Survival*, 57(3), 153–176.

Council of Europe (2001). Convention on Cybercrime. (CETS No. 185), 23 November 2001, available at https://www.coe.int/en/web/conventions/-/council-of-europe-convention-on-cybercrime-ets-no-185-translations.

Crawford, J. (2019). *Brownlie's Principles of Public International Law*, 9th ed., pp. 191–243. Oxford: Oxford University Press.

Deibert, R. (2013). Bounding cyber power: Escalation and restrain in global cyberspace (Internet Governance Papers: Paper No. 6). The Centre for International Governance Innovation. https://www.cigionline.org/sites/default/files/no6_2.pdf.

Deibert, R. (2015). The geopolitics of cyberspace after Snowden. *Current History*, 114(768), 9–15.

Deibert, R. and Crete-Nishihata, M. (2012). Global governance and the spread of cyberspace controls. *Global Governance*, 18(3), 339–361.

Emerson, R. G. (2016). Limits to a cyber-threat. *Contemporary Politics*, 22(2), 178–196.

ESIL-SEDI Website (2021). https://esil-sedi.eu/post_name-1147/.

Europarl Website: MOTION FOR A RESOLUTION on regulating cyberspace (europa.eu).

European Parliament, Digital Technology Study (2021). Digital technologies as a means of repression and social control study, Directorate General for external policies of the Union, PE 653.636, April 2021, p. 44.

Feick, J. and Werle, R. (2010). Regulation of cyberspace. In Baldwin, R., Cave, M. and Lodge, M. (Eds.), *The Oxford Handbook of Regulation*, pp. 523–547. Oxford: Oxford University Press.

Glen, C. (2014). Internet governance: Territorializing cyberspace? *Politics Policy*, 635–657.

Goldsmith, J. T. and Wu, T. (2006). *Who Controls the Internet? Illusion of a Borderless World*. New York: Oxford University Press.

It-law Website: Verwaltung des Internet (it-law.at).

Jayawardane, S., Larik, J. and Jackson, E. (2015). Cyber governance: Challenges, solutions and lessons for effective global governance (Policy Brief No. 17). The Hague Institute for Global, from http://www.thehagueinstituteforglobal justice.org/wp-content/uploads/2015/12/PB17Cyber-Governance.pdf.

Katyal, N. K. (2003). Digital architecture as crime control. *Yale Law Journal*, 112(8), 2261–2289. doi.org/10.2307/3657476.

Kremer, J. F. and Müller, B. (2014). SAM: A framework to understand emerging challenges to states in an interconnected world. In Kremer, J. F. and Müller, B. (Eds.) *Cyberspace and International Relations: Theory, Prospects and Challenges*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-37481-4_3.

Lazar, E. (2017). Juridiction et cybercriminalité, revue TIC, Innovation et droit international, Pedone, 283–294.

Lennon Chang, Y. C. and Grabosky, P. (2017). *The Governance of Cyberspace*, In Drahos, P. (Ed.), *Regulatory Theory: Foundations And Applications*. pp. 533–551. Canberra, Australia: ANU Press, The Australian National University.

Lessig, L. (2006). *Code: Version 2.0*. New York: Basic Books.

Liaropoulos, A. N. (2017). Cyberspace governance and state sovereignty. In Bitros, G. and Kyriazis, N. (Eds.) *Democracy and an Open-Economy World Order*. Berlin: Springer. https://doi.org/10.1007/978-3-319-52168-8_2.

Menthe, D. (1998). Jurisdiction in cyberspace: A theory of international spaces. *Michigan Technology Law Review*, 4(3), 69–101. http://www.mttlr.org/volfour/menthe.html.

Microsoft Official Website (2021). A Digital Geneva Convention to protect cyberspace | Microsoft Cybersecurity.

Motion for a European Parliament Resolution on Regulating Cyberspace (2016). MOTION FOR A RESOLUTION on regulating cyberspace (europa.eu).

Murray, A. D. (2012). Internet regulation. In Levi-Faur, D. (Ed.), *Handbook on the Politics of Regulation*, pp. 267–822. Cheltenham, UK: Edward Elgar.

Nocetti, J. (2015). Contest and conquest: Russia and global internet governance. *International Affairs*, 111–130.

Orfonline Official Website (2021). How the cyberspace narrative shapes governance and security | ORF (orfonline.org).

Proksch, W. and Schweighofer, E. (2011). *Internet Governance and Territoriality Nationalization of Cyberspace*. BILETA. University of Edinburgh, Scotland, April 2001.

Report of the Commission on Global Governance. https://www.gdrc.org/u-gov/global-neighbourhood/chap1.htm.

Revue des medias Website (2021). Un cyberespace européen est-il possible? | la revue des médias (ina.fr).

Scottish Website: Facebook will give "gift to terrorists" & child abusers by giving them total anonymity online, says MI5 boss (thescottishsun.co.uk).

Tusikov, N. (2016). *Chokepoints: Global Private Regulation on the Internet*. California: University of California Press.

UN Official Website (2021). UN General Assembly — Resolutions.

West, S. (2014). Globalizing internet governance: Negotiating cyberspace agreements in the post- Snowden era. In *Conference Paper, TPRC 42: The 42nd Research Conference on Communication, Information and Internet Policy*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id¼2418762##.

Working paper A/CONF.213/9, UN 12th Congress on Crime Prevention and Criminal Justice, Salvador, Brazil, 12–19 April 2010–22 January 2010. http://www.unodc.org/documents/crime-congress/12thCrime-Congress/Documents/A_CONF.213_9/V1050382e.pdf.

World Politics Website (2021). https://www.worldpoliticsreview.com/.

This page intentionally left blank

https://doi.org/10.1142/9781800612860_0004

Chapter 3

Defining Cyber Risk Management Objectives

Sérgio Nunes

Introduction

Nowadays, most enterprises have requirements to protect information security from outsiders, competitors, and even between employees from different departments. Data breaches appear every day in the news and the cyber threat level does not seem to stabilize. Companies are urged to implement cyber risk management practices to limit their exposure to loss. Facing this new reality, it is important to have a baseline to implement a cyber risk management plan and have clearly defined cyber risk objectives.

The technological and regulatory environment of organizations is becoming increasingly complex. Basel II and SOX require companies to undertake periodic risk assessments. However, cyber risk assessment is a moving target, largely on account of the inherent complexity of infrastructures and technological interdependencies. Compliance with regulatory requirements usually results in a "checklist" approach to managing risks. In such cases, a predetermined list of identified risks is made, and any assessment typically checks whether certain requirements have been fulfilled or not. Such practices have typically been critiqued in the literature, and their limitations are highlighted (Dhillon and Backhouse, 2000).

Top management recognizes the need for cybersecurity, but how much cybersecurity investment is enough? (Stewart, 2004) The cybersecurity budget is seen by top management as a black hole that sucks resources and investments without any returns. What are the return benefits of cyber investments? Although the cybersecurity manager is fighting daily to align business objectives with security objectives, it is not always possible to achieve a direct alignment between them. A key factor in influencing top management to expand the cybersecurity budget is cyber risk.

Cyber risk is the common language between cybersecurity and the business, by translating the technical details into business losses and bringing top management long-term commitment to the security strategy (Baskerville, 1991; Vitale, 1986). Cybersecurity investments, most of the time, do not bring tangible added value to businesses, but they mitigate risk to an acceptable level by management by following a cost–benefit approach. The cybersecurity budget lives from those detected risks, based on the importance of the critical business assets that it protects. Critical information resides on the assets that are part of a complex technological environment and are targeted by multiple threats with a probability of being exploited that consequently results in multiple risks. Top management is flooded with these multiple cyber risks and is urged by stakeholders to decide on the best alternative as soon as possible.

It is, therefore, important to consider how cyber risks can be understood and prioritized and to take appropriate decisions. Rather than focusing on alternatives, Keeney (1992) argues the usefulness and relevance of value-focused thinking. Keeney notes that alternative-focused thinking limits decision criteria by focusing only on the alternatives rather than concentrating on companies' objectives, which are driven by values. The correct approach is that of value-focused thinking, whereby values are linked to alternatives for achieving them, thus identifying better decision-making situations, which consequently turn a reactive decision process into a proactive one (Keeney, 1996).

Research in Cyber Risk Management

Regarding cyber risk research, there are multiple models, frameworks, and methods to deal with cybersecurity risks in organizations along with different studies that summarize them (Eloff *et al*., 1993; Vorster and

Labuschagne, 2005). These studies are mainly divided into three groups: statistical probabilistic or economic risk management, maturity or standards focused risk management, and people-focused behavioral risk management. Transversal to these three groups, the research can be focused on the risk of a specific technological solution, adapting risk management to the requirements of an industry or business sector or changing an accepted risk assessment model or methodology to address specific needs.

Tiganoaia (2012) compares multiple methods used for information security risk management. The comparison evaluates the possibility of certification, the language of the documentation, and target organizations, namely, size, commercial focus, and government. He argues that there is a need to integrate information security risk management processes with enterprise risk management and adapt existing risk management methods to the requirements of specific business sectors.

Amancei (2011) discusses some practical methods for information security risk management by providing examples of criteria for risk assessment, impact, and risk acceptance. He uses questionnaires to assess the level of internal control and also evaluates existing controls with a vulnerability assessment.

Stroie and Rusu (2011) discuss approaches to information security risk management. They divide between a proactive and reactive approach. The reactive approach is composed of six steps: protect human life, damage control, damage assessment, determine the root cause, repair the damage, and review the process and update policies. The proactive approach is centered in training activities, defining and implementing formal procedures, and establishing an internal control system. The risk management process is divided in four phases: design, implement, monitor, and improve the risk management system.

Taylor (2015) argues that current risk assessment methods are flawed because management decisions regarding information security are often based on heuristics and optimistic perceptions. The author explains that decision makers are focused on satisfying, which means solving problems without worrying to maximize the outcome with the best solution.

Fenz and Ekelhart (2010) test information security risk management methods across three phases: verification, validation, and evaluation. They argue that the decision to choose a method for information security risk analysis is dependent on the trust on that method and the three phases allow us to discover if the security investments are going in the right path.

Cyber Risk Management Objectives

This study follows the value-focused thinking approach of Keeney (1992) divided mainly across three main steps: collect the detailed list of values for the decision context, rewrite those values in a common form and transform them into subobjectives, and finally classify the objectives using the WITI test into fundamental objectives and means objectives. The ultimate goal by following value-focused thinking in decision analysis should be to select the best alternative, but that is not always possible due the existence of hidden alternatives. The enumeration of values and the creation of objectives serve the principle of eliminating the bad decisions that looked good before but do not accomplish any of the proposed objectives. The unframing of the decision process should be performed as soon as possible by defining the problem at hand and removing the psychological traps that influence our clear judgement in creating new alternatives without the anchoring in the previous alternatives.

The initial list of values for cyber risk management was gathered by conducting semi-structured interviews in Europe with several security and IT professionals who represented a wide variety of job descriptions, such as the CIO, CISO, or IT Manager, for example. The interviewees were representative of multiple business sectors but were predominantly from consultancy, banking, and the telecommunication industry. Should the information gathered require further explanation that was relevant for the research, we interviewed other employees from the same organization in order to clearly identify the context of the information collected. The interview was planned for 1 hour, the smallest was 28 minutes and the longest passed beyond 2 hours of productive discussion. The interview data gathering approach is adequate according to the value-focused thinking methodology, as values should not be constrained and should be intrinsic to the individual. The interview process allows us to discuss and define the problem under analysis and clarify the objectives of the interview. This allows us to establish a common understanding of the concepts involved. It is also important to discuss face to face why the interviewee chose those values to gather and understand the context. The interview started with a general introduction to the value-focused thinking methodology, focusing specifically on the guidelines for understanding and identifying values. Interviews were conducted within borders by using general targeted topics, broad categories, and examples, but open questions were posed which allowed the respondents to reflect on their past decisions and

enabled a review of their judgmental values. The values were collected as part of a wishlist in an ideal situation. The values were then analyzed to see their advantages and disadvantages in the interviewee's context to collect real professional examples and generate decision scenarios.

These scenarios reflect the consequences of good and bad decisions and what was the impact of those decisions on the organization and on its employees. The examples allow us to understand more clearly the values expressed by the interviewee. A total of 71 interviews were performed. Some interviews had more than one interviewee and were conducted as an iterative discussion.

The process started by enumerating all collected raw values into a unique document. These raw values are transformed into a common form, especially if they can be transformed into multiple objectives, to capture each objective individually. Some participants detailed the value as a wish, others as a minimization of a problem, and the rest already described the value in the form of an objective.

There are many ways of wording the same raw value and that's why this common form is important. Duplicate values are merged and the number of times that each value is stated is preserved to capture the strength of that value across multiple respondents. Values are transformed into objectives and then the categorization phase takes place by grouping similar objectives into clusters. These clusters with similar objectives are analyzed and an objective that represents the cluster's idea is discovered. This part of the process involves discussion with multiple specialists, most of them are professors or experienced professionals in the field, to capture the essence of the data collected. Discussions were supported with qualitative analysis software and, if necessary to simplify the visualization of the objectives, these were printed into cards and arranged into groups.

These final objectives are divided into fundamental and means objectives by taking the WITI test. This classification is critical for making informed decisions although it is a subjective and interpretive process. Fundamental objectives are ultimately important and means objectives contribute to the achievement of another objective.

A total of 612 cyber risk management values were collected, and after the removal of duplicates, a total of 414 values were identified. These values were enlisted in a common form and followed the methodology of obtaining a wishlist from the interviewees.

The values in a common form were then transformed into 114 distinct subobjectives, and any duplicates were removed, which resulted in the

same goal in different words, following a correlation and consolidation procedure. This transformation into subobjectives is accomplished by applying an active verb which turns an objective into an effective action. The objectives were then sorted into 23 clusters, taking into account a shared common theme or idea.

These 23 clustered objectives were further classified into means and fundamental objectives, by using the "why is this important" (WITI) test. This structured procedure is important for enabling reflection as to what individuals care about in a cyber risk context and for seeing how these objectives relate in terms of importance. The fundamental objectives are

Table 1. Means and fundamental objectives for cyber risk management.

Overall objective: Minimize cyber risks	
Means objective	**Fundamental objective**
• Ensure properly configured IT infrastructure • Promote cyber risk performance metrics • Ensure ongoing monitoring of cyber risks • Ensure cyber risk management processes are audited • Maximize access control • Minimize cyber risks related to IT service providers • Reduce human negligence • Maximize vetting of employees for cyber risks • Ensure adequate internal communication regarding cyber risks • Ensure adequate external communication regarding cyber risks • Maximize cyber risk management for critical information • Ensure information confidentiality • Ensure information availability • Ensure information integrity • Develop cyber risk management competencies • Develop a cyber risk awareness program • Develop a training program for cyber risk management	• Ensure risk management governance • Maximize cyber risk knowledge • Ensure cybersecurity quality • Maximize responsibility and accountability for cyber risks • Maximize compliance • Maximize the protection of human life

the core values for the decision context and the means objectives enable those core values. The WITI test resulted in a total of 6 fundamental objectives and 17 means objectives as can be seen in Table 1.

Fundamental cyber risk objectives

This section discusses each of the six cyber risk management fundamental objectives, taking into account existing best practices and detailing the context in which they were structured.

Ensure risk management governance includes the adoption of IT and security best practices. Adequate risk management governance entails the nomination of a risk committee, which has the role of discussing risk at the top management level and which consults all relevant stakeholders (Westby and Allen, 2007). It ensures the alignment of the risk management function within the organization to match the business objectives. This alignment of business objectives and risk management practices is seen as a critical step in risk management (ISACA, 2007), following the consolidated approach of strategic alignment between business and IT (Henderson and Venkatraman, 1993). It establishes the virtual structural basis that guides everyday activities with responsibility boundaries and an adequate path of action. It allows to integrate cyber risk management into corporate governance responsibility and place that topic on the top management's agenda (Posthumus and von Solms, 2004). Consolidated risk management practices cannot be restrained to be controlled only inside the IT department or within a department dedicated to information security; these practices have to be raised to the top management level, integrated within enterprise risk management (ERM) (Chatzipoulidis *et al*., 2010; Fakhri *et al*., 2015; Fitzgerald, 1995; Tiganoaia, 2012). The governance framework allows for cyber risks to be included as a major slice inside operational risk management or in an autonomous category and have adequate attention by the top management.

Maximize responsibility and accountability for cyber risks deals with who does what, and who is ultimately responsible for risk mitigation measures (Lichtenstein, 1996). Most of the time, the person who is responsible for executing tasks is responsible for a specific delegated task, and the data owner is the person who is accountable for either accepting the risk or deciding whether to implement additional safeguards (Purdy, 2010). These data owners should be clearly identified in cyber risk

management and, together with their responsibilities in the risk management process, their role should be clear and objective. This identification and definition process prevents finger-pointing across the organization when a risk turns into a real situation. The responsibility and accountability are also enforced legally by binding security and risk management policies to the employee's contract. Otherwise, it may be difficult to take action against employees who violate defined mandatory policies. If employees perceive they are likely to get caught when violating policies, they tend to follow defined policies with more due care (Herath and Rao, 2009). The policies should delineate responsibility and accountability clearly, define to whom they are applicable, and consider how specific enumerated exceptions are viable (Palmer *et al*., 2001). The policies detail what is expected of each employee when dealing with organizational information resources and a user declaration of acknowledgment should be signed before having access to information and that signature renewed on an annual basis to reinforce the accountability (Höne and Eloff, 2002).

Maximize cyber risk knowledge entails the creation of an intangible capability as a risk management organizational culture, in which each stakeholder is aware of existing cyber risks and the controlling management practices. This organizational culture based on shared values of risk management will automatically direct and unify accepted activities while limiting the success of individual deviating behavior by some employees who prefer to follow their non-acceptable ideas and preferences (Furnell and Thomson, 2009; McFadzean *et al*., 2006). Knowledge sharing is an effective way of promoting employee involvement in cyber risk management (Furnell *et al*., 2007; Safa *et al*., 2016). Focusing on empowering employees rather than seeing risk management as a tool augments the flow of risk information within the organization (Thapa and Harnesk, 2014; Veiga and Eloff, 2007). The empowered employee puts his personal knowledge or intimate understanding into the nurturing organizational culture (Mintzberg, 1988). Adequate testing of procedures and know-how should be performed in order to ensure that every employee is informed about the cyber risk culture and knows what their role is in the risk management framework (Veiga and Eloff, 2010). People are always considered the weakest link in information security, being presented as the low-hanging fruit waiting for the attacker to collect (Furnell and Clarke, 2012; Reid and Niekerk, 2014). Promoting the cyber risk knowledge within the information society by governments and extending that knowledge to specific requirements of organizations helps strengthen that

weakest link (Furnell, 2008). The active participation of employees, in a collaborative approach, to form that risk management culture is critical to achieve adequate results (Karabacak and Ozkan, 2010; Spears and Barki, 2010).

Maximize compliance deals with ensuring that requirements from supervisory entities are met and the current regulations are followed. This objective impacts an organization's business directly, as sanctions are applied for lack of compliance, and, in extreme cases, this may lead to the legal prosecution of management. An organization has to adopt those proven methodologies, frameworks, and best practices that guarantee the maximization of compliance. Risk management has defined standards and best practices which should be adapted for cyber risk management in the context of every organization. The documentation of clear cyber risk policies and procedures, together with the definition of internal sanctions for their non-compliance, should be an initial step in establishing the risk management framework. These policies should be written with adequate care, taking into account multiple principles adapted to the current context and requirements of the business and not follow the common copy and paste from templates from other organizations or consultancy services (Höne and Eloff, 2002). Legal compliance issues should also be accounted for, such as, for example, data retention time frames, which differ for each country. Copyright management is also a compliance requirement which affects not only the software acquired by an organization but also, for example, a cyberattack with the intention of implanting illegal software. Several mandatory standards that impact organizations, for example, SOX, Basel, or Solvency, have specific requirements for ensuring information security risk management as part of minimizing the operational risk (ISACA, 2014; ITGI, 2007). PCI-DSS is another mandatory standard that affects payments with credit cards, having specific security requirements to protect cardholder information (DSS, PCI, 2016).

Maximize the protection of human life may seem an outside objective at first glance when dealing with cyber risks. However this objective makes complete sense, after careful examination, and when considered, for example, within the mindset of those cyber risks which affect critical infrastructures that may harm human life. Critical infrastructures are being attacked daily with advanced persistent threats (APTs), with the goal of compromising their infrastructures, for example, energy companies and water management. A real example was when a computer worm named Stuxnet was created to attack nuclear power plants (Farwell and

Rohozinski, 2011). These cyber risks should be carefully managed as most critical infrastructures, although not directly connected to the Internet, are indirectly exposed to attacks, and loss of human life may occur. When reflecting on cyberwarfare, wars happen first in the cybersphere before a physical attack occurs (Baskerville and Portougal, 2003). It is easier to attack a country that is rendered blind, through the lack of communications or power, for example. Recently, efforts are being secretly pursued by countries to enhance their cyber competitive intelligence. At the individual level with the Internet of Things phenomenon, those risks will affect the common citizen in their houses, in their jobs, and directly in their life with the adoption of e-health devices, for example. Cyberstalking and cyberbullying are also phenomena that migrated from the physical into the cyber domain with the rise of the social networks and the digital footprint of every individual (von Solms and van Niekerk, 2013). Maintaining data privacy is also a critical factor for maximizing the protection of human life, by contributing to maintaining freedom as an individual has the right to choose what private information is communicated to which entity (Son and Kim, 2008). As our day-to-day life increases in the digital world with different means of access beyond the traditional computer, the emergence of information contributes to social networks, wiki, or other Web 2.0 platforms, and so the risk to privacy grows. It is known that when information reaches the web, it never leaves it again. Trying to delete information that reaches the web by accident is a quest of endless loops, as that information may be copied easily before removal from a specific service. The rise of big data, with the capacity of computers to process enormous quantities of data, also increased privacy concerns, with previously impossible relations about entities and their online behavior being profiled in seconds. Geo-location tracking is also simplified with the use of devices with GPS or other geolocation mechanisms. RFID and NFC mechanisms also contribute to location tracking (Madlmayr *et al*., 2008; Pramatari and Theotokis, 2009). The emergence of smart toys also raises questions about the risks posed to small children, not only from the security point of view but also targeting privacy rights (Dobbins, 2015).

Ensure cybersecurity quality objective aggregates security concepts, including explicitly the information confidentiality, integrity, and availability triad. Examples of some of the fundamental concepts for assuring security quality are information authenticity, reliability, and nonrepudiation (Alcalde *et al*., 2009). The ability to counteract aggressive

actions can be ensured by robust authentication, complemented with strong auditing mechanisms and adequate identity management of multiple stakeholders across multiple platforms, and also by using applications with strong access controls.

Means cyber risk objectives

This section presents the 17 means objectives and details the context for the formation of each distinct theme among the subobjectives.

Ensure properly configured IT infrastructure to protect against attacks which exploit vulnerabilities in unpatched systems. A lack of system hardening procedures that do not remove unnecessary services and remove default credentials is another risk that has to be accounted for. The security of legacy systems with discontinued support from the vendor should not be disregarded. A fundamental principle for the safeguarding of information availability is the adoption of solutions that ensure high availability in case of failure. Contributions toward ensuring an adequate infrastructure architecture plan include developing a strategy for promoting interoperability across platforms and choosing a solution which minimizes technological dispersion in the technological environment.

The rise of the shadow IT concept, where organizational applications bypass local IT department administration and are contracted as a service directly by non-IT departments, is a growing risk in organizations (Fürstenau and Rothe, 2014; Silic and Back, 2014). Is the IT infrastructure where that application runs properly configured? Most of the time, the IT department has no information that some departments use shadow applications and does not reserve the right to audit that infrastructure (Paquette *et al*., 2010).

Ensure ongoing monitoring of cyber risks to comply with best practices for continuous improvement. Risk frameworks and standards, such as, for example, ISO 31000, follow continuous improvement methods within the classic cycle of "Plan, Do, Check, Act" (Johnson, 2002) and contribute to an ongoing monitoring of cyber risks, as they evaluate the current risk level, deploy risk mitigation measures or accept the resilient risks, and also re-evaluate whether the risk exposure remains the same after organizational changes.

Promote cyber risk performance metric is the best way to carry out quantitative evaluation if the cyber risk mitigation objectives are being met, depending on how much is completed during execution. This obliges

the implementation of an adequate system's logging level, as a means of extracting meaningful information. This measurement allows for the planning of changes when they are needed, according to the plan, as part of the ongoing process of enabling the delivery of benefits (Bodin *et al*., 2008; Gordon and Loeb, 2002). Key performance indicators (KPIs) allow management to follow the initiatives that have been implemented to improve the current risk level. Dashboards are also used which periodically analyze current metrics and identify deviations, which permits timely decisions to be made, which put an organization's risk management back on track. A good metric is clear and objective, and the data collection occurs automatically to minimize errors. The creation of metrics follows a defined process starting with scope definition and ending with the testing of the designed metric following the pragmatic principle, for example (Brotby, 2009; Brotby and Hinson, 2013).

Ensure cyber risk management processes are audited to have an independent vision to check whether risk management is being applied with adequate due care (Straub and Welke, 1998). It is always important to have another opinion regarding a critical business process and risk management is no exception, as an internal or external audit will inevitably point out recommendations for improvement (ISACA, 2007). In the case of an internal audit department, they may propose controls to strengthen the risk management processes, ensuring, for example, separation of duties (Saltzer and Schroeder, 1975). They discover and alert management to common anomalies within defined processes. In the case of an external audit, a specialist auditor will be able to benchmark current cyber risk management practices against those that are being practiced by similar organizations and thus guarantee the adoption of the current risk management benchmark. This objective leads to the implementation of an adequate process testing framework which is able to produce tangible evidence for the auditors. Software code auditing review tests and penetration tests should also be performed to discover hidden cyber risks.

Minimize cyber risks related to IT service providers is an important objective, as organizations tend to outsource some of their IT processes and information is now moving on to the cloud (Pereira *et al*., 2012). The alignment of objectives between providers and clients is critical to the success of these partnerships, as security objectives tend to be positioned in different priorities by the client and the IT service provider (Dhillon *et al*., 2016). A simple mistake caused by the IT provider can trigger multiple business losses across multiple clients (Salmela, 2008).

Conflicts of interest, for example, by selling and then managing security solutions or implementing and then auditing solutions should be minimized. Information and IT knowledge lock-in risks should be safeguarded in service level agreements. This lock-in technically occurs when the vendor does not adopt appropriate technological standards for migrating information to a different provider or when the vendor goes bankrupt and shuts down his/her services. This lock-in can also happen when an organization loses core technological competencies and becomes hostage to the provider, who knows everything on that specific subject about the organization and refuses to document or pass on that knowledge to the organizations' internal resources. Legal requirements also need to be safeguarded, taking into account where the information is physically stored when using cloud services and what the applicable law is in that country. The maintenance of intellectual property rights has to be ensured in contracts and with adequate monitoring. The security chain is as strong as the weakest link, so it's necessary that IT service providers maintain at least the same security level as required by their clients (Johnson *et al*., 2009). Access to critical information should be limited in the case of outsourcers, and non-disclosure agreements should be signed to prevent disclosure. Temporary credentials should be issued and these have to be renewed if the outsourcer still needs access after their contractual period ends.

Maximize access control protects information from unauthorized access. Access control plays a critical role in the validation of authentication, authorization, and accounting. The principle of segregation of duties should be enforced to prevent a single entity from having full access to a critical process. Critical processes should include an authorization phase, which is carried out by another entity rather than the one that is responsible for its execution (Saltzer and Schroeder, 1975). Access control allows for the maximization of access segmentation, following the network defense in-depth security principle (Lippmann *et al*., 2006). Password management policies should be enforced to maximize access control, using multifactor authentication when available, and obliging the use of complex passwords that must be periodically renewed. Multifactor authentication is used when two or more factors are used for accessing a resource. Examples of such factors are as follows: something that you know, for example, a password or a PIN, something that you have, for example, a smartcard or one-time token, and something that you are focusing in, for example, biometry (Jonvik *et al*., 2008). Other factors are something that you do with examples keystroke dynamics or form of

writing and something that identifies where you are accessing from: GPS or the type of device. Minimum privilege should be the default option for access control, giving users only the allowed access to fulfill their daily tasks (Saltzer and Schroeder, 1975).

Reduce human negligence attempts to minimize human errors that occur due to a lack of awareness or simply because human nature is negligent. Negligence applies to the failure of establishing the adequate due care of a prudent person to protect information from risks that may harm others (von Solms and von Solms, 2006). Human error is often underestimated as a business risk (Im and Baskerville, 2005). A business impact analysis should be carried out to evaluate the impact of negligence or malicious conduct regarding information (Whitman *et al*., 2013). Only after this analysis has been completed is it possible to evaluate and quantify what is critical to the business and then to implement adequate written procedures that explain the critical task step by step, thus minimizing human negligence. The use of applicational controls that immediately detect human errors along with the monitoring of anomalies to detect errors afterward (ISACA, 2014).

Maximize vetting of employees for cyber risks can be ensured by adopting best practices for hiring human resources. Employees should be subjected to criminal records checks. Security clearance practices that take into account the criticality of information should be mandatorily implemented. Non-disclosure agreements for internal employees and external consultants should also be enforced when dealing with critical information. The ethical behavior of internal employees should be mandatory to minimize the risk of internal breaches in security, as employees have additional privileges to sensitive information (Dhillon, 2001).

Ensure adequate internal communication regarding cyber risks across all the stakeholders and promote internal meetings to ensure that the correct communication paths are created and maximized. A formal risk communication policy should exist in the organization that encourages employees to discuss and report risks. Being aware of existing risks due to internal communication solidifies the responsibility of every employee. Data owners should communicate the criticality of information to custodians (Krause and Tipton, 2002; Peltier, 2013). Maximizing the involvement of all stakeholders enables a clear definition of accepted risk levels. When adequate internal communication regarding cyber risks is established, it minimizes user panic when a risk situation becomes real. Lack of organizational communication increases gray areas of

responsibility, leaving risks without treatment as some important risks remain unknown to top management. Risk reporting should not be viewed as a witch hunt, but their communication is promoted among open defined communication paths. Open organizational communication helps form a risk management culture (Baskerville, 1991).

Ensure adequate external communication regarding cyber risks to minimize the loss of reputation due to cyber risks. A spokesperson needs to be clearly identified and briefed for handling crises, such as data breaches, for example (Valackiene, 2015). This spokesperson has to have the clear objective of minimizing media pressure in cyber risk management and also acts as a facilitator for minimizing the effect of political decisions that affect cyber risk management.

Ensure identification of critical information deals with the definition and identification of the critical information within business processes and ensures the evaluation of critical information and defined service criticality levels. This definition allows for the centering of safeguards in critical information, as budgets are invariably limited. A data classification program should be established with a clear definition of who the information owners and custodians are, together with defined data criticality levels (Johnson *et al*., 2009; Krause and Tipton, 2002). Information systems can only guarantee the application of controls taking into account the defined information classification level. An information or data classification policy has to define the criteria for placing information across a number of defined classification levels and what are the minimum controls to be applied to physical or digital information which has that level (Appleyard, 2005; Peltier, 2013). The policy should include also the responsibilities for employees that deal with information of that level. The information classification process should follow the information across the full lifecycle, being applied when the information is created, traveling with the information across multiple levels in time, because information may be critical today but public or useless tomorrow and finally monitoring, if necessary, the information destruction. This prioritization of information results in the maximization of the efficiency of the incident response team, when critical operations are affected by system failures, asset compromises, or data breaches.

Ensure information confidentiality to prevent sensitive information being leaked to an unauthorized entity. Intellectual property protection is of great concern to organizations (Johnson *et al*., 2009). This can be ensured by adopting adequate encryption measures when dealing with

stored information and by encrypting laptops' hard drives or critical databases. Information transmitted can use secure network protocols that ensure that the data are encrypted. Data leak protection (DLP) mechanisms can be implemented to prevent leakage by disgruntled employees or external consultants. Implementing an information classification program is also a crucial step for defining information's value and for protecting printed documents (Peltier, 2013).

Ensure information availability when access to information is required by an entity. The risk of loss of information needs to be minimized, and adequate backup procedures and data recovery methods should be tested periodically. The transportation of backup information offsite should be evaluated in order to protect against disaster. Business continuity and disaster recovery best practices should be adopted by the organization, which should include the definition of recovery times and point objectives (Whitman *et al*., 2013). The presence of high availability mechanisms in the infrastructure that supports critical processes is an added protection measure against failures and protects against denial of service attacks.

Ensure information integrity by adopting good change management practices, which protect information from unauthorized modification (Joshi *et al*., 2001). Change management allows for the tracking of responsibilities and prevents unauthorized and unprepared changes (Cannon *et al*., 2007). Changes should be planned and a rollback plan should be available in case corruption of data occurs.

Develop cyber risk management competencies which allow employees to recognize cyber risks. The allocation of trained staff for cyber risk management should be ensured. These experts should not only be able to identify technical cyber risks but must also be able to recognize risks that arise from poorly defined business processes. These competencies are built with formal education and training and with on-the-job experience (Blakley *et al*., 2001; Furnell and Thomson, 2009). These competencies can be a source of competitive advantage following a resource-based view approach (Barney, 2001; Bharadwaj, 2000).

Develop a cyber risk awareness program which permits employees to recognize typical cyber risk scenarios and to become alert to deviant behaviors (Drevin *et al*., 2006; Peltier, 2005; Siponen, 2000, 2001). Awareness contributes significantly to the formation of a conscious care behavior (Safa *et al*., 2016). People are always the weakest link in any risk management program, and the implementation of a consistent awareness

program, which transmits and tests employees periodically about risk management best practices, is a vital key to success in risk mitigation. Awareness programs mitigate the unconscious incompetence of employees, as they are unaware of their responsibilities in information security and risk management (Thomson *et al*., 2006).

Develop a training program for cyber risk management that includes user cyber risk training and encourages users to become proficient in crisis management procedures. Such training goes beyond an awareness process and prescribes specific procedures that must be followed when dealing with cyber risks. Training moves the employee from a conscious incompetence stage, as he/she recognizes the skills gap, to a conscious competence stage, as he/she is able to deal with cyber risks. The continuous practice enables the employee to reach the unconscious competence stage, where he has absorbed the knowledge to fulfill his/her tasks (Thomson *et al*., 2006). This training approach increases the perceived usefulness and maximizes the perceived ease of use with the minimization of the learning curve sometimes associated with new technologies and change of processes (Davis, 1989). The need for a specialized license to practice for risk management professionals may also be a differentiation point in the future, as certifications in this area start to mature (Blakley *et al*., 2001). The certifications advise a code of conduct that reinforces ethical and professional obligations in due diligence.

Conclusion

This study provides cyber risk management objectives, grounded on stakeholders' values, to minimize cyber risk. Finding out what real stakeholders' value in cyber risk management is a new contribution to the existing knowledge gap regarding this uncharted topic. These objectives not only entail a technical point of view but also focus on managerial organizational issues captured into formal and informal controls. These objectives are segmented by their relationship into fundamental and means objectives and can be the basis for a decision model regarding cyber risk management. The justification of security investments to mitigate cyber risk is always a difficult battle, as these investments may seem useless without no tangible value to the business, according to some skeptical stakeholders. By using a decision model based on the cyber risk objectives, the decision maker can justify the cybersecurity investments to

stakeholders, as the basis for the investment was their elicited values in the first place. This simplifies the decision process in cyber risk management, as it tends to increase in complexity with the progress of the technology dependency in organizations.

References

Alcalde, B., Dubois, E., Mauw, S., Mayer, N. and Radomirović, S. (2009). Towards a decision model based on trust and security risk management. In *Proceedings of the Seventh Australasian Conference on Information Security*, Vol. 98, pp. 61–70.

Amancei, C. (2011). Practical methods for information security risk management. *Informatica Economica*, 15(1), 151.

Appleyard, J. (2005). Information classification: A corporate implementation guide. In Tipton, H. F. and Krause, M. (Eds.), *Information Security Management Handbook*. Auerbach Publications.

Barney, J. B. (2001). Resource-based theories of competitive advantage: A ten-year retrospective on the resource-based view. *Journal of Management*, 27(6), 643–650.

Baskerville, R. (1991). Risk analysis: An interpretive feasibility tool in justifying information systems security. *European Journal of Information Systems*, 1(2), 121–130.

Baskerville, R. L. and Portougal, V. (2003). A possibility theory framework for security evaluation in national infrastructure protection. *Journal of Database Management (JDM)*, 14(2), 1–13.

Bharadwaj, A. S. (2000). A resource-based perspective on information technology capability and firm performance: An empirical investigation. *MIS Quarterly*, 24(1), 169–196.

Blakley, B., McDermott, E. and Geer, D. (2001). Information security is information risk management. In *Proceedings of the 2001 Workshop on New Security Paradigms*, pp. 97–104.

Bodin, L. D., Gordon, L. A. and Loeb, M. P. (2008). Information security and risk management. *Communications ACM*, 51(4), 64–68.

Brotby, W. K. (2009). *Information Security Management Metrics: A Definitive Guide to Effective Security Monitoring and Measurement*. Boca Raton: CRC Press.

Brotby, W. K. and Hinson, G. (2013). *PRAGMATIC Security Metrics: Applying Metametrics to Information Security*. Boca Raton: CRC Press.

Cannon, D., Wheeldon, D., Taylor, S. and Office of Government Commerce UK. (2007). *ITIL: IT Service Management Practices*. ITIL v3 core publications Service operation. The Stationery Office.

Chatzipoulidis, A., Mavridis, I. and Kargidis, T. (2010). Developing strategic perspectives for enterprise risk management towards information assurance. In *Proceedings of the 9th European Conference on Information Warfare and Security: ECIW2010*, p. 35.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.

Dhillon, G. (2001). Violation of safeguards by trusted personnel and understanding related information security concerns. *Computers & Security*, 20(2), 165–172.

Dhillon, G. and Backhouse, J. (2000). Information system security management in the new millennium. *Communications of the ACM*, 43(7), 125–128.

Dhillon, G., Syed, R. and de Sá-Soares, F. (2016). Information security concerns in IT outsourcing: Identifying (in) congruence between clients and vendors. *Information & Management*.

Dobbins, D. L. (2015). Analysis of security concerns and privacy risks of children's smart toys. Ph.D. Dissertation. Washington University St. Louis, St. Louis, MO, USA.

Drevin, L., Kruger, H. A. and Steyn, T. (2006). Value-focused assessment of information communication and technology security awareness in an academic environment. In Fischer-Hübner, S., Rannenberg, K., Yngström, L. and Lindskog, S. (Eds.), *Security and Privacy in Dynamic Environments, Proceedings of the IFIP TC-11 21st International Information Security Conference (SEC 2006)*, 22–24 May 2006, Karlstad, Sweden, Vol. 201, pp. 448–453. IFIP. Springer.

DSS, PCI (2016). *Payment Card Industry Data Security Standard Version 3.2*.

Eloff, J., Labuschagne, L. and Badenhorst, K. (1993). A comparative framework for risk analysis methods. *Computers & Security*, 12(6), 597–603.

Fakhri, B., Fahimah, N., Ibrahim, J. *et al*. (2015). Information security aligned to enterprise management. *Middle East Journal of Business*, 10(1), 62–66.

Farwell, J. P. and Rohozinski, R. (2011). Stuxnet and the future of cyber war. *Survival*, 53(1), 23–40.

Fenz, S. and Ekelhart, A. (2010). Verification, validation, and evaluation in information security risk management. *IEEE Security & Privacy*, 9(2), 58–65.

Fitzgerald, K. J. (1995). Information security baselines. *Information Management & Computer Security*, 3(2), 8–12.

Furnell, S. (2008). End-user security culture: A lesson that will never be learnt? *Computer Fraud & Security*, 2008(4), 6–9.

Furnell, S., Bryant, P. and Phippen, A. D. (2007). Assessing the security perceptions of personal internet users. *Computers & Security*, 26(5), 410–417.

Furnell, S. and Clarke, N. (2012). Power to the people? The evolving recognition of human aspects of security. *Computers & Security*, 31(8), 983–988.

Furnell, S. and Thomson, K.-L. (2009). From culture to disobedience: Recognising the varying user acceptance of IT security. *Computer Fraud & Security*, 2009(2), 5–10.

Fürstenau, D. and Rothe, H. (2014). Shadow IT systems: Discerning the good and the evil.

Gordon, L. A. and Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security*, 5(4), 438–457.

Henderson, J. C. and Venkatraman, H. (1993). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal*, 32(1), 472–484.

Herath, T. and Rao, H. R. (2009). Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems*, 47(2), 154–165.

Höne, K. and Eloff, J. H. P. (2002). Information security policy: What do international information security standards say? *Computers & Security*, 21(5), 402–409.

Im, G. P. and Baskerville, R. L. (2005). A longitudinal study of information system threat categories: The enduring problem of human error. *ACM SIGMIS Database*, 36(4), 68–79.

ISACA (2007). Cobit framework 4.1. http://www.isaca.org.

ISACA (2014). IT Control Objectives for Sarbanes-Oxley: Using COBIT 5 in the Design and Implementation of Internal Controls Over Financial Reporting. ISACA.

ITGI (2007). *IT Control Objectives for Basel II: The Importance of Governance and Risk Management for Compliance*. ISACA.

Johnson, C. N. (2002). The benefits of PDCA. *Quality Progress*, 35(5), 120.

Johnson, M. E., Goetz, E. and Pfleeger, S. L. (2009). Security through information risk management. *IEEE Security & Privacy*, 7(3), 45–52.

Jonvik, T., Feng, B., Jorstad, I. *et al*. (2008). Simple strong authentication for internet applications using mobile phones. In *IEEE GLOBECOM 2008-2008 IEEE Global Telecommunications Conference*, pp. 1–5.

Joshi, J. B., Aref, W. G., Ghafoor, A. and Spafford, E. H. (2001). Security models for web-based applications. *Communications of the ACM*, 44(2), 38–44.

Karabacak, B. and Ozkan, S. (2010). A collaborative process based risk analysis for information security management systems. In *Proceedings of the 5th International Conference Information Warfare and Security*, p. 182.

Keeney, R. L. (1992). *Value-Focused Thinking: A Path to Creative Decisionmaking*. Cambridge: Harvard University Press.

Keeney, R. L. (1996). Value-focused thinking: Identifying decision opportunities and creating alternatives. *European Journal of Operational Research*, 92(3), 537–549.

Krause, M. and Tipton, H. F. (2002). Ownership and custody of data. In Hugh Murray, W. (Ed.) *Information Security Management Handbook*, 1st ed., Vol. 4, pp. 461–472. Boca Raton: Auerbach Publications.

Lichtenstein, S. (1996). Factors in the selection of a risk assessment method. *Information Management & Computer Security*, 4(4), 20–25.

Lippmann, R., Ingols, K., Scott, C., Piwowarski, K., Kratkiewicz, K., Artz, M. and Cunningham, R. (2006). Validating and restoring defense in depth using attack graphs. In *MILCOM 2006-2006 IEEE Military Communications Conference*, pp. 1–10.

Madlmayr, G., Langer, J., Kantner, C. and Scharinger, J. (2008). NFC devices: Security and privacy. In *3rd International Conference on Availability, Reliability and Security, ARES 08*, pp. 642–647.

McFadzean, E., Ezingeard, J.-N. and Birchall, D. (2006). Anchoring information security governance research: Sociological groundings and future directions. *Journal of Information System Security*, 2(3), 3–48.

Mintzberg, H. (1988). Crafting strategy. *The McKinsey Quarterly*, Summer, 88(3), 71–90.

Palmer, M. E., Robinson, C., Patilla, J. C. and Moser, E. P. (2001). Information security policy framework: Best practices for security policy in the e-commerce age. *Information Systems Security*, 10(2), 1–15.

Paquette, S., Jaeger, P. T. and Wilson, S. C. (2010). Identifying the security risks associated with governmental use of cloud computing. *Government Information Quarterly*, 27(3), 245–253.

Peltier, T. R. (2005). Implementing an information security awareness program. *Information Systems Security*, 14(2), 37–49.

Peltier, T. R. (2013). *Information Security Fundamentals*. Boca Raton: CRC Press.

Pereira, L., Soares, F. D. S. and Caldeira, M. (2012). Information systems security outsourcing key issues: A service providers' perspective. In *Proceedings of the European Conference on Information Systems 2012*.

Posthumus, S. and von Solms, R. (2004). A framework for the governance of information security. *Computers & Security*, 23(8), 638–646.

Pramatari, K. and Theotokis, A. (2009). Consumer acceptance of RFID-enabled services: A model of multiple attitudes, perceived system characteristics and individual traits. *European Journal of Information Systems*, 18(6), 541–552.

Purdy, G. (2010). ISO 31000: 2009 setting a new standard for risk management. *Risk Analysis*, 30(6), 881–886.

Reid, R. and Niekerk, J. V. (2014). From information security to cyber security cultures. In *Information Security for South Africa*. 2014, pp. 1–7, doi: 10.1109/ISSA.2014.6950492.

Safa, N. S., Von Solms, R. and Furnell, S. (2016). Information security policy compliance model in organizations. *Computers & Security*, 56, 70–82.

Salmela, H. (2008). Analysing business losses caused by information systems risk: A business process analysis approach. *Journal of Information Technology*, 23(3), 185–202.

Saltzer, J. H. and Schroeder, M. D. (1975). The protection of information in computer systems. *Proceedings of the IEEE*, 63(9), 1278–1308.

Silic, M. and Back, A. (2014). Shadow IT — A view from behind the curtain. *Computers & Security*, 45, 274–283.

Siponen, M. (2000). A conceptual foundation for organizational information security awareness. *Information Management & Computer Security*, 8(1), 31–41.

Siponen, M. (2001). Five dimensions of information security awareness. *Computers and Society*, 31(2), 24–29.

Son, J.-Y. and Kim, S. S. (2008). Internet users' information privacy-protective responses: A taxonomy and a nomological model. *MIS Quarterly*, 32(3), 503–529.

Spears, J. L. and Barki, H. (2010). User participation in information systems security risk management. *MIS Quarterly*, 34(3), 503–522.

Stewart, A. (2004). On risk: Perception and direction. *Computers & Security*, 23(5), 362–370.

Straub, D. W. and Welke, R. J. (1998). Coping with systems risk: Security planning models for management decision making. *MIS Quarterly*, 22(4), 441–469.

Stroie, E. R. and Rusu, A. C. (2011). Security risk management-approaches and methodology. *Informatica Economica*, 15(1), 228.

Taylor, R. G. (2015). Potential problems with information security risk assessments. *Information Security Journal: A Global Perspective*, 24(4–6), 177–184.

Thapa, D. and Harnesk, D. (2014). Rethinking the information security risk practices: A critical social theory perspective. In *47th Hawaii International Conference on System Sciences*, pp. 3207–3214.

Thomson, K.-L., von Solms, R. and Louw, L. (2006). Cultivating an organizational information security culture. *Computer Fraud & Security*, 2006(10), 7–11.

Tiganoaia, B. (2012). Comparative study regarding the methods used for security risk management. *Scientific Bulletin-Nicolae Balcescu Land Forces Academy,* 17(2), 149.

Valackiene, A. (2015). Efficient corporate communication: Decisions in crisis management. *Engineering Economics*, 66(1), 99–110.

Veiga, A. D. and Eloff, J. H. (2007). An information security governance framework. *Information Systems Management*, 24(4), 361–372.

Veiga, A. D. and Eloff, J. H. (2010). A framework and assessment instrument for information security culture. *Computers & Security*, 29(2), 196–207.

Vitale, M. R. (1986). The growing risks of information systems success. *MIS Quarterly*, 327–334.

von Solms, R. and van Niekerk, J. (2013). From information security to cyber security. *Computers & Security*, 38, 97–102.

von Solms, R. and von Solms, S. B. (2006). Information security governance: Due care. *Computers & Security*, 25(7), 494–497.

Vorster, A. and Labuschagne, L. (2005). A framework for comparing different information security risk analysis methodologies. In *Proceedings of the 2005 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries*, Republic of South Africa, South African Institute for Computer Scientists and Information Technologists, pp. 95–103.

Westby, J. R. and Allen, J. H. (2007). Governing for enterprise security (GES) implementation guide. Technical Report CMU/SEI-2007-TN-020, Carnegie Mellon University SEI.

Whitman, M. E., Mattord, H. J. and Green, A. (2013). *Principles of Incident Response and Disaster Recovery*. Boston: Cengage Learning.

This page intentionally left blank

https://doi.org/10.1142/9781800612860_0005

Chapter 4

Data Protection Concerns in Emerging Technologies

Jorge Alan García Bazán

Introduction

Cost and complexity are no longer limiting factors in our digital universe, which has led to an explosion of data of every type — structured, unstructured, privileged, confidential, and public. In order for businesses to be successful today, they need to consider the ever-increasing demands for real-time sharing and collaboration, which creates even more risk for the organization. Grappling with this growth of information is a challenge not just from a security standpoint but also from ever-expanding regulatory demands by nation-states and governments around the world demanding improved consumer and citizen privacy.

The rise of cloud computing technologies and big data complicates the struggle for enterprises as they fight to achieve compliance and deliver security while managing legacy debt in the form of technology and process. For small businesses to massive complex corporations, the continued pressure and desire to monetize data are at odds with the need and obligation to protect and secure the same data. As enterprise digital transformation programs mature, the ability to transform and share data quickly becomes critical to businesses' success. Easy access to even more data creates a perpetual loop for enterprises that are trying to secure information while facilitating the types of frictionless access and user experience

that users want and need. As companies work hard to implement secure mechanisms of protection across all the data journeys, usability, user interaction, and flexibility are compromised.

Data protection is critical, but it can sometimes be less relevant than other types of security controls (ISACA, 2016). This is because, very often, data protection failures are contributory rather than directly responsible for security failures. In other words, the root cause of a high-profile data breach may be something not related directly to data protection, but data protection often causes the impact to be significantly greater or the scope of compromise to be much broader because those controls are not in place.

Companies struggle to discover their own data, where it resides, sensibility of the information, the responsibilities of whom process, store or manage data and, therefore, protect data. Organizations have minimal security controls implemented regarding data security.

The adoption of cloud architectures, multicloud environments, and multiple actors processing and transmitting data leaves a big gap that can be easily compromised wherever it resides, whether it is at rest or in motion. This could lead the company to a potential data loss that could impact financially or its reputation.

While enterprises move quickly to evaluate the actions that they need to take to stay compliant across all of their regulatory obligations, the fundamental truth that being compliant does not equal being secure is increasing the pressure on chief information security officers (CISOs), chief data officers (CDOs), chief privacy officers (CPOs), and other enterprise leaders to deliver security and compliance. Regulations, whether national or international, are a very important matter in the adoption of security controls and the business strategies. A McKinsey research revealed that more 60% of the companies interviewed are not prepared to comply with industry regulations and standards, therefore companies implement temporary security controls and manual processes that need to be replaced, instead of implementing a solid, long-term security strategy (Anant and Donchak, 2020).

Why Is Cybersecurity Culture So Important?

Implementing a cybersecurity culture could be the biggest challenge for an organization. Creating a cybersecurity culture is everyone's responsibility. There are many facts that support this challenge: many organizations and

employees undervalue the importance of cybersecurity; the increasing adoption of cloud and emerging technologies; lack of support from the board and hard to engage the different roles on the same level of priority and huge investments on technology but don't sufficiently pay attention to the human factor, which is the most vulnerable asset in the company.

The increasing number of more sophisticated cyber threats creates a larger attack surface that leads to more opportunities to cybercriminals. Cybercriminals will execute attacks on a company using phishing emails, social engineering, social networks, and other tactics to trick and deceive users in the company to expose sensitive data from the company. Thus, employees become a vulnerable target and the first line of defense to be enforced. Users are the ones interacting with the company's resources, such as computers, networks, devices, data and systems, and are responsible for the appropriate use and purposes established by the company. This is the reason why employees play an important role in creating a security culture.

The main objective of cybersecurity is to protect an organization's most valuable assets addressing cyber threats that could compromise the information, operations, and reputation of the company. In other words, it's all about data protection. Many organizations underestimate the value of cybersecurity and lack of a cybersecurity strategy, a cybersecurity-business oriented mindset, and a budget for security-related purposes. A cybersecurity strategy will help organizations know their cybersecurity maturity posture, where they stand and what they need to achieve in order to be in a more robust position. A tangible outcome of a successful cyber-attack could easily lead into a loss of the organization's reputation, a decrease in the stock market, the closure of a business unit, or even bankruptcy. Thus, cybersecurity becomes an important concern for an organization's financial and legal risk. Organizations should collaborate, enforce, and adopt cybersecurity as part of the business committee in order to support the cybersecurity strategy. Last but not least, a budget should be allocated to the cybersecurity department to deal with current and emerging cybersecurity projects.

Data Discovery and Classification

You cannot protect what you can't see (Wilczek, 2018). Companies struggle to protect data because they do not know where their data reside,

where they are stored, how they are recollected, and who has access to them, and this is the biggest challenge when it comes to planning a data protection strategy. Companies are increasingly adopting new technologies, in heterogeneous environments, such as big data and IoT, and launching new products and services that help them grow at a rapid scale to be the leaders in their industry. With the rise of the use of these flexible and scalable technologies, the risk associated with them also increases.

Data come from a variety of sources, presented in various formats, created, processed, and stored by internal and external parties for different purposes (ISACA, 2019). From simple log files, marketing strategies, and commercial expansion plans to legal and intellectual property, data whether structured or unstructured should be protected according to its relevance. Organizations have data in heterogeneous environments, as a result of the increasing adoption of cloud technologies, hybrid architectures, and on-site. Consequently, these become a challenge for companies to protect a massive amount of data stored in different locations and unclassified. Traditionally, companies stored their data in their own data center and built huge walls to protect the perimeter. Those days are gone; the evolving threat landscape has changed and those defenses have vanished; there is no longer a perimeter, but the mission has never changed to protect the data. Data are the new perimeter.

From a cybercriminal perspective, information is the new trade currency and very valuable for threat actors willing to misuse, manipulate, or sell it. There are a number of motivations for cybercriminals to target companies of all sizes and industries: from fake news to extortions, from identity theft to hacktivism, and from malware to damage reputation. It's not about how many endpoints are encrypted by ransomware, how many databases are compromised, the number of admin credentials stolen, or if you are a potential candidate for the cybercrime. In the end, it's all about data confidentiality, integrity, and availability (CIA triad) (Jelen, 2019).

According to a recent global report conducted by the Ponemon Institute, it reveals that 47% of data breaches are related to insider threats (Ponemon Institute, 2020). An insider threat is the will of a user who has authorized access to the company's critical assets to use their access, either unintentionally or maliciously, in a way that could negatively impact the company. He/she could be a negligent employee or a malicious insider. Unintentional insider threats, though not motivated by malice and in many cases occurring out of ignorance, are more likely to be driven by the desire to help or to be efficient. This innate desire to help can be

leveraged by social engineers to infiltrate an information system with the purpose of exfiltrating data.

Unintentional data leakage occurs when confidential information leaves a corporation's boundaries without explicit approval by authorized personnel (Verizon, 2019). Companies are innovating constantly to bring new services, transmitting and storing data every day. The move to the cloud and the increasing usability of cloud services are accelerating the risk of potential data loss. End users purchase applications, such as cloud enterprise resource planning solutions, PDF readers and collaboration tools, without involvement from IT or cybersecurity teams. Without end-user recognition, these actions create shadow IT departments. Shadow IT increases the risk of data transmission, exposure, and storage outside of organizational standards and controls (Godfrey, 2019).

Moreover, employees are humans and, therefore, are prone to make mistakes. In some cases, an employee may inadvertently disclose classified data to unauthorized personnel or may mistakenly dispose of information in a way that makes it available to unauthorized persons through dumpster diving. As such, any strategy an organization adopts should encompass measures to safeguard the entire information security spectrum.

An organization's security strategy should provide protection to information in its three states: data can be at rest, such as when data are stored on a computer or server, data in motion, such as when an application is using or retrieving the data, and data in transit, such as being sent via email or downloaded from the server. Regardless of the state of the data, administrators should seek to protect data as they move from state to state.

Furthermore, the security strategy policy should encompass three key elements to be successful: people, processes, and technology. Without a combination of these core elements, any security policy will fall short of providing the desired outcome.

People are the backbone of any information security ecosystem. Thus, people are probably the most critical element, as people are both the threat and part of the security strategy (Harris and Maymi, 2016). Security begins with individual employees, as they are often the weakest link in any security program. Having well-trained employees who can recognize the suspicious behaviors will prevent users from falling on deceptions and can strengthen an organization's security posture.

People are also necessary to monitor and respond to incidents as well as to hunt for potential indicators of compromise (Hadnagy, 2011).

Without trained incident responders, the other key elements of processes and technology are meaningless. Having a qualified security team will enhance the effectiveness of controls designed for protection, detection, and response.

Processes are guided by policies, procedures, guidelines, and work instructions. These documents should provide high-level instructions regarding the organization's security policy; dictate how, when, and by whom communication takes place with external agencies in the event of an incident; and outline standard operating procedures to be followed to protect, detect, and correct incidents. The policies should also dictate what constitutes risky behavior and should seek to increase monitoring on those deemed to have a higher risk.

Having the right technology can serve as a force multiplier and boost an organization's information security program. Although organizations do not always need the latest and cutting-edge technology, they do require some tools to augment their employees and assist them with monitoring and responding to incidents. Tools assist with analysis and make managing large datasets across an entire enterprise more manageable. Without tools, it is hard for an organization to establish controls, which makes it difficult to protect information, detect when a problem arises, and correct the problem, preventing further damage.

Privacy Concerns in the Cloud: The Perfect Storm

How did you get my phone number? There are a number of players that collect data from individuals and organizations alike. Governments and private institutions collect a massive amount of data for different purposes, with and without the user's consent. Those pretty slim-designed and powerful devices called smartphones have become a need in our daily basis that we take for granted. It seems a natural action to wake up, no matter where you are located, and use it to talk to someone else or post a publication, no matter where the other person is located and who is able to see your publications. Mobiles store a huge amount of sensitive information, and if they are always online, they transmit data in real time. Additionally, there are tens or even hundreds of applications on your mobile phone that share your location, identity, user content, browsing history to sexual orientation, working habits, medical data, and even religious and political views. All this data are shared and processed to an

average of at least 10 third parties for analysis or to sell (Schneier, 2016). This information can give a huge context of a user's behavior, hobbies, and marketing profile.

Deal or no deal? People need to make and receive calls, but in exchange, they must sacrifice their privacy; the carrier is allowed to know where you are located at all times. This is not explicitly specified in the contract. So, to what extent is the consumer and cellphone company responsible for protecting the customer's data? It's a tough question. Customers often agree to these terms and conditions when acquiring a smartphone because it is simple, there is no choice.

On the other hand, there are a number of trending mobile apps that can create information for a user that could not be treated properly. Apps that make users look older, change gender, tell you with who your blood is compatible, dating sites that tell you who's your perfect match, or even how many times you work out during the week and if your heart rate is the appropriate level. All this information is very valuable and highly usable for benign and malicious purposes. In 2020, a hobby biker was using a fitness application on his smartphone to track how many miles he rode every day. After several months, Zachary McCoy received an email from Google's legal team to notify him that the local police were demanding information from his Google account. The vast amount of data collected from the fitness application made McCoy a crime suspect. A couple of months earlier, a woman was burglarized less than a mile from his daily habitual routes, and even though he had nothing to do with it, he had to press charges (Schuppe, 2020).

In 2015, a group of cybercriminals hacked an online cheating site requesting the company to shut down the portal (Krebs, 2015). Cybercriminals involved in the hack had access to all customer personal data, conversations, videos, employee data, and critical systems and made it public. A couple of weeks later, other cybercriminals and opportunists collected the information from public websites and extorted users with a ransom or otherwise they would reveal all the information to family and friends. These had several consequences from divorce to suicide. Like two sides of the same coin, information can be a very valuable element; it just depends on which hands it falls into.

Government institutions, like Police and Law Enforcement, also need access to civilian's data for investigations, censuses, and other purposes and collaborate with organizations and people alike to accomplish these tasks. Countries worldwide have different data protection and privacy

regulations and both public institutions and private companies must comply. International organizations are working hand in hand with countries to create greater security awareness and strengthen their data protection regulations.

On the other hand, as data growth has accelerated at lightning speed, enterprises face challenges when it comes to having a clear and solid data security strategy. Organizations are starting to realize that the separation of IT, cybersecurity, and privacy requirements must work and collaborate together in order to achieve optimal results in data protection and compliance. It's not enough to meet regulations and industry standards; companies must also build a flexible data protection program that anticipates and keeps pace with future needs.

Many industries and associations are working with legislatures from countries all over the world in order to craft regulations that tie both security and privacy. The escalating costs of fines, legal expenses, and recovery are in the hundreds of millions of dollars for a single data breach event at large corporations. A complication that most ventures confront in assembling their compliance prerequisites is that lawmakers and controllers universally have not continuously been careful of the specialized achievability of the commands they make. In spite of the fact that controllers have sought after citizens' right to be overlooked, controllers have shown small understanding of the specialized trouble in that endeavor and, in numerous cases, have exempted whole capacities, such as government administrations, from the specialized request, whereas holding organizations responsible for performance. The need for specialized understanding is additionally shown within the creation of controls that don't recognize the interconnectivity of information, systems, and users. Companies are left searching for best practices or making internal judgment calls about what meets the reasonableness test since there are no clear criteria specified.

With the ever-changing landscape of data growth, enterprises in every industry are creating unique digital identities for users specifically for the purpose of creating a tight knot between a user and their data. With the rise of cloud-based technologies, neither infrastructure (IaaS) nor software as a service (SaaS) has not completely replaced the need for on-premise infrastructure and business-critical application administration. Companies have not completely replaced their data centers with the cloud; the cloud has become an additional IT strategy for companies to oversee on top of other existing architectures. Cloud

technologies have many benefits and drawbacks when it comes to data security and privacy.

Cloud-based infrastructures can benefit from availability, scalability over time, compliance with industry standards, like PCI-DSS, HIPAA, GDPR, SOX, among others, and very attractive pricing models (ISACA, 2020). Since data are stored remotely, there is no restriction of time or location; this makes it ideal when you have a workforce distributed in different regions and even encourages collaboration between colleagues within the company. Organizations can scale over time according to their technical and business needs; this makes cloud computing attractive for businesses rather than small or big corporations.

However, the problem of adopting and securing cloud technologies arises in terms of the data location, management, integrity, confidentiality, operations, technology controls to protect data, and legal implications. Cloud vendors often have several data centers distributed across the world; it is very challenging to discover the exact location of the data stored. However, as data are transmitted from one country to another, the rules governing data storage change, bringing security issues and data protection laws into play, which are relevant to cloud data storage. As a cloud service provider, the service provider must notify customers about their data storage policies and the exact location of their data storage server. From a legal standpoint, this implies data retention, destruction, management, access from government entities, and compliance with state, country, and international privacy regulations. Data have its lifecycle management, and if the cycle of one data set is complete and no further processing is required, that data should be deleted from the server according to its destruction policy. Review the deletion policy from your provider and make sure that your information is programmed to be removed at a pre-specified time as mentioned in your contract.

Data access mainly refers to the data security policies. In an organization, the employees will be given access to the section of data based on the user needs and rights aligned with the appropriate security policies. The same data cannot be accessed by the other employee working in the same organization. Governments around the world have as of now passed laws forbidding encryption in their nations. This action takes off people's data vulnerable to surveillance and hacks. The issue is that governments aren't the only actors that can abuse this lack of security. Encryption is imperative since without it, utilizing the web at all can take off your individual information accessible for anybody to get. In the event that

encryption is prohibited on a national level, cybercrimes, like identity theft, fraud, and extortion, will likely increment significantly.

Various encryption techniques and key management mechanisms are used to ensure that data are shared only with the valid users. The key is distributed only to the authorized parties using various key distribution mechanisms. Keys must be rotated periodically depending on the regulatory requirements that apply to the company and industry. The adoption of risk-based multifactor authentication to grant access to the appropriate resources and verify the user's identity is a best practice. To secure the data from the unauthorized users, the data security policies must be strictly followed. Since access is given through the Internet for all cloud users, it is necessary to provide privileged user access.

The system should maintain security such that data can be only modified by the authorized personnel. In a cloud-based environment, data integrity must be maintained correctly to avoid any inherent data loss. Permissions granted to make changes to the data should be limited to valid users so that there is no widespread access to problems and preserve data integrity. In case of data transfer bottlenecks and disaster, organizations using cloud computing applications need to protect the user's data without any loss. If data are not managed properly, then there is an issue of data storage and data access. In case of disaster, the cloud providers are responsible for the loss of data.

Big Data — Big Exposure

Recently, big data has taken considerable attention from the industry, scientific and technology communities, and several government institutions. Many countries are also using big data to provide services in various fields, such as healthcare, medicine, public sector undertakings, distribution, marketing, and manufacturing. Big data is essentially an information-based technology that analyzes large amounts of data from multiple sources to extract valuable information and predict changes based on the context associated with the extracted information and is mainly defined by its 3Vs fundamental characteristics. The 3Vs include Velocity (data growth and changing in a rapid way), Variety (data come from different sources and multiple formats), and Volume (a huge amount of data are generated every second) (Simon, 2015). Many economic and political interests drive big data, especially the processes of

data integration, analytics, and data mining. An important characteristic of big data is that data from various sources have life cycles from collection to destruction, and new information can be derived through analysis, combination, and usage. Big data offers many advantages and potentials for innovation in various fields but also presents many issues and challenges in terms of insecure infrastructure, lack of a reliable source, access controls, data storage, and privacy concerns. First, information security, privacy preservation, and ethical issues are significant open difficulties in big data environments and include data management strategies, assurance of personal data, and misuse of information. Particularly, a lot of shared data, including security, can be exploited in an interconnected open ecosystem.

Untrusted software programs are used by cybercriminals in order to extract and turnout sensitive information from data sources. Insecure computation aside from causing data leakage can also corrupt data, leading to incorrect results in prediction or analysis. It can also result in a Distributed Denial of Service attack (DDoS) on your big data solution disabling the property of using a massively parallel programming language.

Big data needs to collect input from a variety of sources, therefore it is quite important and mandatory to validate the input and have a single source of truth. This involves making a decision about what kind of data are untrusted and what are the untrusted data sources. It also needs to segregate rogue or malicious data from the legitimate one. When a large volume of malicious data are inserted into the dataset, its influence on the outcome produced is massive. Signature-based data filtering is incapable of tracking down such attacks, thus individual custom algorithms need to be designed to deal with such cases.

Big data was traditionally designed for high performance and scalability with almost no native security in mind. Traditional databases have a very comprehensive table, row, and cell-level access control, and these have been really gone missing in big data environments. *Adhoc* queries pose another additional challenge to big data solutions where users can retrieve sensitive information out of the data using *adhoc* queries. Even though being provided by a big data solution, granular access control is disabled by default.

As data are stored at hundreds and even thousands of nodes, "authentication, authorization, and encryption of data at those nodes become a challenging task" (Bhathal and Singh, 2019). If any solution provides

encryption of real-time data, it may not be useful, as encryption of real-time data may have slow performance impacts on the data sets and is time consuming in a big data ecosystem.

Monetization of big data involves data mining and analytics, and sharing of those analytical results involves multiple obstacles, like the invasion of privacy, invasive marketing, and unintentional disclosure of information. For data sharing, digital ecosystems are based on multiple heterogeneous platforms. Such eco-systems aim to ensure real-time data access for many parties, such as partners, customers, service providers, contractors, and employees. They rely on multiple connections with different levels of security. Data collaboration associated with advanced analytics techniques brings multiple security threats, such as the discovery of sensitive information or illegal access to networks' traffic. In fact, by establishing relations between extracted data from different sources, it is possible to identify individuals in spite of data anonymization by using correlation attacks.

Private organizations and government institutions have to respect many security laws and industrial standards that aim to enhance the management of digital data privacy and to protect confidentiality. However, some information technologies may involve entities across many countries. So, enterprises have to deal with multiple laws and regulations. Furthermore, big data analytics may be in conflict with some privacy principles. As a result, various standardization organizations have published related standards in an effort for security and privacy-preserving of big data, and privacy protection laws, such as the General Data Protection Regulation (GDPR) (GDPR, 2018) in Europe and the California Consumer Privacy Act (CCPA) in the United States, have been enacted (CCPA, 2018). Hence, enterprises must adhere to the appropriate data protection acts in order to preserve data privacy.

Privacy and Ethical Issues in the Internet of Things

The rapidly growing number of interconnected (IoT) devices have become an indispensable part of how individuals and companies live, communicate, and do business. The IoT is quickly growing as more and more devices are attached to a global network. All around the world, manufacturing companies are developing IoT technologies for multiple applications, such as healthcare, finance, industrial IoT,

commerce, transportation, and consumer. These devices attached to the Internet are able to generate, collect, and exchange data using nodes, creating a network of uniquely identifiable objects capable to interact with themselves, their external environment, or both. Many IoT devices' data and applications are highly sensitive and should be accessible only to authorized individuals. These applications are the computer programs that use real-time conditions to ensure they do not fail, and they use consumption data to analyze and predict future behaviors and patterns. IoT security should include more than just the IoT device itself. IoT devices have minimal security and many vulnerabilities that could be easily exploited. IoT manufacturers are not considering security and privacy by design.

Manufacturing companies are adopting industrial IoT technologies to provide a better customer service through better customization of products and services to clients in shorter time frames. The foundation of better connectivity and communication between the assembly line and the organization, made possible by IoT, enables manufacturers to be closer to market demands and customize what they are producing to the needs of their customers. Thus, this close communication between smart facilities and the data center of the organization brings a huge risk making the attack surface exposition greater for potential threats.

Enterprises and research institutions in the healthcare ecosystem are connecting patients to healthcare systems for continuous medical data monitoring and improved patient care. Patients are able to book and consult with their physician with their smartphone without the need of telecommuting when minor cases arise. Smart medical devices are enabling people to have real-time data and share it with doctors, nurses, family members, and other parties, like insurance companies and pharmacies. Sensitive data are constantly traveling to different destinations to be processed for different purposes.

The goal of IoT is to provide quality of life and improve automation in everyday tasks for consumers and enterprises. An average user has around four devices (Heslop, 2019). Consumer-connected devices, such as smartwatches, wearables, speakers, and tablets, represent a high risk for the enterprises. As these devices are unmanaged, have different communication protocols, frequencies, and ranges, belong to the employee, and communicate with the internal networks if compromised they could lead cybercriminals to other assets and get access to sensitive information or even take control of enterprise systems.

All these benefits provided by IoT devices represent a big challenge for organizations as they are not produced with security in mind. IoT devices are now mainstream targets for hackers and have various security flaws at the device, hardware, software, and networking level that could compromise the organization. A good defense starts at ground level or hardware level. The hardware on which the IoT device is built-in forms the basis for a solid and secure IoT device. The hardware components in an IoT device vary depending on the application and usage. Sensors, accelerometers, gyroscopes, and radio-frequency identification (RFID) chips are examples of such components that make devices smart (Hancke and De Carvalho e Silva, 2012). Primary threats to an IoT device at the hardware level are that it can be stolen, physically modified, or cloned. Hardware vulnerability common causes include weak default passwords and counterfeit integrated circuits. Nevertheless, to address these hardware vulnerabilities, the process of manufacturing IoT devices must be regulated. The manufacturers of IoT devices need to be accountable for not adhering to the appropriate IoT regulatory standards and guidelines.

Major threats to the software or firmware on IoT devices are that the software can be modified or decompiled to extract credentials and leveraged to perform the DDoS attacks or code injection. The software includes platforms and applications that determine what data to collect, what data sources to connect to, which decision-making algorithms to use, and the application programming interface (API) to connect with other software components.

Like many networking devices, the most common IoT device threats at the network level are man-in-the-middle (MiTM) attacks, eavesdropping, and bandwidth theft. In order to protect the enterprise from these threats, devices must be discovered and inventoried as soon they connect to the network and ensure they integrate into an asset management program. Standards and baselines must be well defined in an IoT security strategy. Operational networks should be segmented from IoT networks such that in case of an attack, it would be easier to mitigate the threat and minimize risk. The implementation of two-factor authentication or key-based authentication is strongly recommended for communication between IoT devices and the user.

It is essential to create an adequate legal framework and develop the underlying technology with security and privacy in mind. Regulations will enforce manufacturers and vendors to make security a priority and provide guidelines on the expectation of IoT developers and

manufacturers. IoT regulations will give a level of transparency to consumers, or packaging can reflect the level of security of the IoT device. Compliance will force manufacturers to upgrade and secure their products. IoT applications need to have some consideration for the EU GDPR. GDPR introduced a general mandatory notification in the event of personal data breaches. Data controllers must report personal data breaches to their supervisory authorities no later than 72 hours after becoming aware of such a breach and, in some cases, are also required to report such breaches to affected individuals (GDPR, 2018). Manufacturers need to ensure that they are in a position to identify and react to security breaches in a manner that complies with the requirements of the GDPR.

Compliance: Regulations, Laws, and Standards

Businesses and their customers alike collect, store, and transmit vast amounts of information electronically, and they want to believe that this information is secure (Mikkelsen and Soller, 2019). At the customer level, the concern for data privacy has resulted in a growing number of laws and regulations that address issues including what information can be collected and maintained, how the information should be stored, how and where information can be transmitted, and the required actions in the event of a security breach. Notwithstanding the proliferation of requirements, reports of identity theft, the inadvertent release of customer and proprietary business information, and successful attempts by hackers to penetrate systems and steal information continue to threaten consumers and organizations' privacy.

Customers are demanding different services and a transparent user experience, from mobile banking, eCommerce, and eGovernment services, which require them to provide personal sensitive information across different channels of communication; organizations want to be able to collect, data mine, and share this information efficiently. Certain industries, such as financial services, insurance, and healthcare, are most sensitive when it comes to privacy and data protection because of the personal information they possess and exchange with third parties. However, all industries are affected by privacy and data protection requirements.

Protecting sensitive information and privacy is, in fact, one initiative that governments and businesses share. Many industry associations, such

as the Payment Card Industry Data Security Standard (Payment Card Industry Security Standards Council, 2006), the Health Information Trust Alliance Common Security Framework (Health Information Trust Alliance, 2007), Telecommunications Service Companies Privacy Regulation (Germany) (German Telecommunications Act, 1996), General Data Protection Regulation (GDPR, 2018), Information Commissioners Office (United Kingdom) (Information Commissioner's Office, 1984), and Privacy and Electronic Communications Regulations (United Kingdom) (Privacy and Electronic Communications Directive Regulations, 2003), have issued their own standards to supplement existing laws and regulations (Kennedy, 2019). A number of nations and states across the world have developed their own data protection laws and keep working to make them stricter and organizations adhere to them. The global data regulation landscape has become increasingly complex in recent years, and businesses trading internationally must keep track of an ever-changing patchwork of rules. The purpose of these laws is diverse, from controlling the use of personal data to data transfer nationally or internationally to the use and protection of consumer data through geopolitical implications. Inappropriate mechanisms for data protection have hampered data protection in developing countries. They have laws on data protection and privacy though not specific to the target.

Given the risks and related requirements, ensuring the privacy of customer information and protecting critical corporate data are priority concerns for management teams. Most companies have developed and implemented privacy and data protection programs, yet many of these programs fall short for a variety of reasons, including a lack of understanding of the legal and regulatory risk landscape related to information collection and processing, inadequate organizational policies, insufficient training, and unverified third-party providers, among many others.

While organizations are faced with an increasingly complex scenario of regulatory demands around the world, the good news is that most compliance requirements can be met with the same set of basic best practices. Frameworks, such as NIST Cybersecurity Framework (National Institute of Standards and Technology, 2014), CIS 20 Critical Controls (Center for Internet Security, 2008), and ISO 27001 (International Organization for Standardization, 2005), offer a number of best practices, guidelines, and footprints that could easily be adopted from small to big enterprises and from any industry to minimize the likelihood and impact of security incidents.

Security also needs to be a leading priority; governments and companies must ensure that strong security policies are in place for any data being stored, processed, or transmitted. Data protection requires concerted efforts which must involve the harmonization of new or existing legislation. These laws must have an international setting and be applicable to all states regardless of whether a country is developed or not. Harmonization implies cooperation between different countries. Cooperation could be evident when different countries' law enforcement agencies cooperate in fighting cybercrime. Finally, organizations should have a long-term vision against these potential threats, take a step forward, and help compliance teams proactively assess the organization's risks and liabilities in different regions and ensure they are compliant even if the regulatory landscape continues to shift.

Conclusion

The threat landscape is continuously moving and keeping the pace is almost impossible as hackers and threat actors are often very organized and are one step ahead of the practices, people, and technology, we use to combat cyber threats. Organizations seeking to build and maintain effective and compliant privacy and data protection program should take into consideration the following best practices.

Organizations should incorporate a cybersecurity culture as part of their DNA and see it as a value-added differentiator. Communication and security awareness are key elements in creating a conscious philosophy within the organization. Switch the cybersecurity approach from technology-based defenses to proactive steps that include processes and education. Commitment, collaboration, and leadership from all members of the organization are required to make cybersecurity a pillar of the corporate.

The human element plays a vital role in a security strategy. In order to develop a robust cybersecurity program, people with the appropriate skills and knowledge ensure security policies and best practices are enforced, and the right technology controls to be compliant and combat cyber threats should be adopted.

In order to improve compliance and reduce risk, a General Data Protection and Privacy program should be enforced. Discovering the purpose, use, and location of data inside and outside of the enterprise is critical to clearly identify the regulatory and legal implications that apply to it.

Enable continuous monitoring of the processes, solutions, and users to ensure that security and compliance are being accomplished.

The risk of an insecure IoT device is relative based on the domain in which it is operated and the jurisdiction in which it thrives. The geography of where the IoT device operates also matters because the legal and regulatory bindings can differ from one country to another. The governance of IoT devices needs to be handled separately but under the IT governance spectrum. IoT applications and manufacturers should be legally regulated in order to give transparency to consumers and preserve privacy.

Regulations and industry standards' main objectives are to enforce and achieve security for individuals and enterprises. A strong collaboration between government, law enforcement, and regulatory bodies could strengthen in fighting cybercrime. Regulatory agents should consolidate their current laws and enforce procedures on data protection and privacy. In today's digital world, success requires a shift from a reactive, compliance-centric posture to a proactive approach that acknowledges consumer and enterprises' rights and concerns.

References

Anant, V. and Donchak, L. (2020). The consumer-data opportunity and the privacy imperative. *McKinsey*. 27 April 2020. https://www.mckinsey.com/business-functions/risk/our-insights/the-consumer-data-opportunity-and-the-privacy-imperative.

Bhathal, G. S. and Singh, A. (2019). Big data: Hadoop framework vulnerabilities, security issues and attacks. *Science Direct*, 100002. April 2019. https://www.sciencedirect.com/science/article/pii/S2590005619300025.

California Consumer Privacy Act (CCPA) (2018). California consumer privacy act. *California Civil Code*. June 2018. https://oag.ca.gov/privacy/ccpa.

Center for Internet Security (2008). Center for internet security critical security controls for effective cyber defense. Center for Internet Security. https://www.cisecurity.org/controls/cis-controls-list/.

General Data Protection Regulation (GDPR) (2018). General data protection regulation. May 2018. https://gdpr-info.eu/.

German Telecommunications Act (1996). Telecommunications service companies privacy regulation. German Telecommunications Act. https://www.itu.int/ITU-D/treg/Legislation/Germany/TelecomAct.pdf.

Godfrey, J. (2019). Data discovery and classification are complicated, but critical to your data protection program. Security Intelligence. 4 November 2019.

https://securityintelligence.com/posts/data-discovery-and-classification-are-complicated-but-critical-to-your-data-protection-program/.

Hadnagy, C. (2011). *Social Engineering: The Art of Human Hacking*. USA: Wiley Publishing Inc.

Hancke, G. P. and De Carvalho e Silva, B. (2012). The role of advanced sensing in smart cities. National Center for Biotechnology Information, U.S. National Library of Medicine. 27 December 2012. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3574682/.

Harris, S. and Maymi, F. (2016). *CISSP All-in-One Exam Guide*, 7th ed. USA: McGraw Hill Education.

Health Information Trust Alliance (2007). Health information trust alliance common security framework. https://hitrustalliance.net/product-tool/hitrust-csf/.

Heslop, B. (2019). By 2030, each person will own 15 connected devices — Here's what that means for your business and content. *MarTechAdvisor*. 4 March 2019. https://www.martechadvisor.com/articles/iot/by-2030-each-person-will-own-15-connected-devices-heres-what-that-means-for-your-business-and-content/.

Information Commissioner's Office (1984). Information Commissioner's Office, UK. https://ico.org.uk/.

International Organization for Standardization (2005). International organization for standardization and the international electrotechnical commission 27001. International Organization for Standardization. https://www.iso.org/isoiec-27001-information-security.html.

ISACA (2016). Managing data protection and cybersecurity. https://www.isaca.org/resources/isaca-journal/issues/2016/managing-data-protection-and-cybersecurity.

ISACA (2019). Implementing a cybersecurity culture. https://www.isaca.org/resources/isaca-journal/issues/2019/volume-2/implementing-a-cybersecurity-culture.

ISACA (2020). Achieving data security and compliance. 27 April 2020.

Jelen, S. (2019). Cybersecurity culture: Why it matters for your business. *SecurityTrails*. 20 June 2019. https://securitytrails.com/blog/cybersecurity-culture.

Kennedy, G. E. (2019). Data privacy law: A practical guide to the GDPR. *Bowker Editorial*. May 2019.

Krebs, B. (2015). Online cheating site AshleyMadison hacked. *KrebsOnSecurity*. 19 July 2015. https://krebsonsecurity.com/2015/07/online-cheating-site-ashleymadison-hacked/.

Mikkelsen, D. and Soller, H. (2019). GDPR compliance since May 2018: A continuing challenge. *McKinsey*. 22 July 2019. https://www.mckinsey.com/business-functions/risk/our-insights/gdpr-compliance-after-may-2018-a-continuing-challenge#.

National Institute of Standards and Technology (2014). NIST cybersecurity framework. National Institute of Standards and Technology. https://www.nist.gov/cyberframework.

Payment Card Industry Security Standards Council (2006). Payment card industry data security standards, 2006. https://www.pcisecuritystandards.org/pci_security/.

Ponemon Institute (2020). Cost of insider threats: Global report. *Bank Info Security*. 28 July 2020. https://www.bankinfosecurity.com/whitepapers/2020-ponemon-cost-insider-threats-global-report-w-6022.

Privacy and Electronic Communications Directive Regulations (2003). Privacy and electronic communications directive regulations. Information Commissioner's Office, 2003. https://ico.org.uk/for-organisations/guide-to-pecr/what-are-pecr/.

Schneier, B. (2016). *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*. New York: W. W. Norton & Company. March 2015.

Schuppe, J. (2020). Google tracked his bike ride past a burglarized home. That made him a suspect. *NBC News*. 7 March 2020. https://www.nbcnews.com/news/us-news/google-tracked-his-bike-ride-past-burglarized-home-made-him-n1151761.

Simon, P. (2015). *Too Big to Ignore: The Business Case for Big Data*. Hoboken: Wiley. October 2015.

Verizon (2019). 2019 data breach investigations report. *Verizon*. 5 May 2019. https://enterprise.verizon.com/resources/executivebriefs/2019-dbir-executive-brief.pdf.

Wilczek, M. (2018). 95% of organizations have cultural issues around cybersecurity. *Dark Reading*. 16 November 2018. https://www.darkreading.com/vulnerabilities---threats/95--of-organizations-have-cultural-issues-around-cybersecurity/a/d-id/1333290.

https://doi.org/10.1142/9781800612860_0006

Chapter 5

Biometric Technology and User Identity

Steven Furnell

Introduction

The focus of this chapter is how we represent — and more particularly how we prove — our identity in IT and online contexts. Fundamentally what this requires is a reliable means to verify that an identity being claimed in the digital realm belongs to the right person in the real world. There are a variety of different technologies that can be used to achieve this, which have varying degrees of linkage to the user concerned. For example, traditional passwords are related to the user because they know the information, whereas something like an access card is associated with the user by possession of it. However, even from reading this, some potential downsides may already be apparent in that the relationship between the user and their means of proving their identity is not that strong. Indeed, there are easily recognizable cases in which the relationship could be *broken* (e.g., the user could forget the knowledge or lose the card) or *compromised* (e.g., someone else could discover the knowledge or acquire the card). As such, there is arguably a need for a means of representing identity via something that is more closely tied to the user themselves, and this is where the use of biometric technologies has a role to play. Biometrics are based upon characteristics of the user themselves and so (in theory at least) offer a means of overcoming some of the constraints of other approaches. However, this comes at the cost of several other considerations, including some potential risk, and so it is relevant to

recognize the trade-offs alongside the benefits. It is against this backdrop that this chapter is presented.

The discussion begins by explaining the basics of biometrics and related technologies, looking at the different categories of approaches and the characteristics we desire from them as a viable and reliable basis for representing identity. Having more fully introduced what biometrics are, Section 3 then places them in context alongside more traditional and long-established user authentication technologies. This not only highlights the areas in which biometrics stand to provide benefits but also begins to establish some of their potential risk factors. Looking into the nature of the technology a little further, Section 4 then provides a brief explanation of how biometric processes actually work, in order to provide a foundation for understanding some of the later challenges inherent in linking it to identity. The next section then explores the evolution of biometric adoption and use that has brought us to the current point, where it is now meaningful to think of them being the basis of our wider digital identity. Section 6 then explores the implications of this, in terms of the need to safeguard against the misuse of this identity and to protect the biometric data itself. Having considered this key concern, Section 7 then moves to consider various wider challenges that still need to be recognized in the biometric context. Taking stock of some of the implications, Section 8 then identifies some areas in which resultant regulation may be beneficial in relation to both biometric technologies and the handling of users' biometric data. Section 9 draws the discussion to a close, highlighting that while biometrics clearly pose some challenges, they also hold significant promise and appear likely to remain a key contributor to digital identity as we move forward.

The Basis of Biometrics

It is likely that most readers will readily think about biometrics based upon fingerprint, face, voice, and possibly iris recognition, as these have become the most prominent public-facing techniques — primarily thanks to their adoption and use within smartphones and other mobile devices. However, while they are arguably the most mature approaches, they are certainly not the only ones. Indeed, if we look at biometrics more broadly, they can be seen to fall into two main categories: *physiological* (based upon some physical/bodily characteristic of the user) and *behavioral* (based upon a characteristic of how the user behaves). Each of these categories then has

Table 1. Examples of physiological and behavioral characteristics that can be used for biometrics.

Physiological	Behavioral
Face	Gait
Fingerprint	Handwritten signature
Hand geometry	Keystroke dynamics (typing rhythm)
Iris	Mouse dynamics
Palm print	Touchscreen dynamics
Retina	Voice
Vascular/vein pattern	

a variety of underlying methods (also known as biometric modalities) within them, as illustrated by the examples listed in Table 1.

A full description of the concept and operation of each of the techniques is beyond the scope of this chapter, and so readers are referred to other sources for more details of the different approaches (Fairhurst, 2018; NCSC, 2019). However, in summary, the concept for any of the biometrics is for the characteristic concerned to be used as the basis for identifying or authenticating (i.e., verifying the identity) a given individual. In most cases, we are dealing with the authentication context, i.e., we know who the user is expected or claimed to be, and so the biometric is being used as a means of confirming it.[1]

Jain *et al*. identified a series of factors that can be used to assess the potential of different characteristics for use in a biometric context, which are summarized as follows (Jain *et al*., 1999):

- **Circumvention** — The ease with which an impostor may be able to duplicate or imitate the characteristic in order to gain unauthorized access;

[1]As a further aside, it can also be noted that the individual techniques may be utilized in multimodal contexts (also known as multibiometric systems), in which two or more traits are combined to provide an overall measure of identity. Biometrics can also find application alongside other authentication methods (i.e., secret knowledge and/or physical possessions) in order to achieve two-factor (2FA) or multifactor (MFA) approaches.

- **Collectability** — The ease with which a sensor is able to collect the sample;
- **Performance** — The accuracy, speed, and robustness of the technique;
- **Permanence** — The ability for the characteristic to remain consistent over time;
- **Acceptability** — The degree to which the technique is found to be acceptable by those that are expected to be using it;
- **Universality** — The ability for a technique to be applied to a whole population of users;
- **Uniqueness** — The ability to successfully discriminate between different individuals within the target population.

A few points are notable from this list. The first is that a high result is targeted for all of the factors other than Circumvention, where it is desirable for the potential to be low. Second, it can be seen that several of the descriptions refer to the user population amongst which the biometric is to be used, and the nature and size of the population will indeed have a significant impact on some of the factors. For example, the Uniqueness of a given characteristic would be expected to be higher within the closed user community of a specific company than if the same technique were to be deployed on a large-scale national level. Third, it can be noted that certain factors relate to the inherent nature of the biometric trait concerned (e.g., the Permanence of facial metrics is likely to be the same regardless of where they are used), and others are linked to the context in which the biometric is being deployed (e.g., the Acceptability and Universality of face recognition are likely to be very different in an environment where a sizeable proportion of the user population is using face coverings). Meanwhile, factors, such as Collectability, Performance, and Circumvention, will be closely linked to how the system has been implemented (e.g., the quality of sensors and where they have been placed). Relating these back to the main categories of biometrics, the physiological approaches are generally regarded as being "stronger" from a security perspective on the basis of exhibiting greater Uniqueness, Universality, and Permanence.

In reality, the extent to which each of the techniques will fulfill the different criteria will depend not only upon the natural characteristics of the approach but also upon how it has been *implemented*. For example, it is possible to implement a biometric that offers an inherently high degree of uniqueness in a poor way such that it still becomes circumventable. A key element of implementation is whether it incorporates a means of

liveness detection, to ensure that the biometric sample is genuine (i.e., coming from a real source, live at the point of capture) rather than being faked or impersonated in some way. Examples of the latter would be spoof attempts, such as trying to use a photo or 3D mask to fool facial recognition or playing back a voice recording to trick voice verification. More macabre examples would include attempting to use lifeless body parts to fool techniques, such as fingerprint or iris-based approaches. All of these can be prevented with suitable biometric technologies, but older or low-end implementations may lack the associated capabilities (e.g., facial recognition on early smartphone implementations used to be fooled by a simple photo of the legitimate user, whereas later versions incorporated more advanced sensors to incorporate 3D image recognition and artificial intelligence to ensure that the user is paying attention rather than asleep or dead).

Biometrics in Context

Before examining biometrics in further detail, it is worth stepping back to consider the broader landscape of proving user identity in technology devices and online systems and understanding how biometrics fit into this. The traditional way of classifying user authentication approaches is based on something the user *knows*, something the user *has*, or something the user *is*, which are as follows:

- **Something the user knows** — Approaches here are based upon the user having some form of secret knowledge and are most commonly represented by passwords and personal identification number (PIN) codes. However, methods can also utilize various other forms of secret, including responding to question-and-answer challenges and using a range of graphical methods (which can include recalling the correct sequence of images, identifying secret points within an image, or drawing a secret image).
- **Something the user has** — This is based upon the user having some form of physical token, the possession of which is considered to prove their legitimacy. Traditional methods here include badges, access cards, and plug-in devices, such as USB dongles. Common solutions also involve two-factor solutions, where the user has a hardware token that generates one-time access codes for login. In modern implementations,

the user's mobile device or some form of wearable technology (e.g., smartwatch) can often perform an additional role in acting as their authentication token. This can include having an app that generates one-time codes (and, therefore, acts as a software token rather than requiring the user to carry a separate physical token), as well as the device itself communicating with other technologies via near-field communications (e.g., automatically unlocking the user's computer when it detects the presence of their smartwatch).

- **Something the user is** — This is the category that reflects the use of biometrics and involves identifying or authenticating the user based upon physiological or behavioral characteristics. As with the *knows* and *has* categories, there are a range of underlying approaches, but the key difference is that any such techniques are leveraging something that the user naturally possesses. The challenge then becomes how to measure and use the chosen characteristic(s) in an effective manner.

Identity verification has traditionally been based on user ids and passwords. The identifier is essentially the claim, and the password is provided as a verification that the claimant has the right to use it. However, password-based approaches have long been recognized as having weaknesses in terms of both the technology and how we use it. Without going into the detail, common problems include selecting weak passwords (e.g., making obvious or easily guessable choices, or strings that are too short and therefore vulnerable to automated attacks), forgetting them, writing them down (so as not to forget them!), storing them in discoverable locations, sharing them with other people (and thereby undermining the security), and having difficulty in managing them at scale (i.e., with many devices and services protected by passwords, many people use the same password in multiple places, which increases their exposure if the password should be compromised in any individual case). Token-based approaches overcome some of these limitations. For example, the level of protection is not dependent upon the user making a potentially weak choice, and the token cannot be given away to others without causing the user themselves to lose access. However, they still pose challenges, such as the potential to be lost or stolen, as well as introducing direct costs for deployment and replacement. Meanwhile, biometrics overcome a number of the weaknesses of the other categories but at the cost of a potentially significant trade-off if they are compromised.

A summary comparison of the different approaches is presented in Table 2, highlighting a series of challenges and whether or not they represent an issue for each category. It should be noted whether a tick or a cross is a "good" answer depends upon the nature of the issue concerned, and so to further aid interpretation, the cells are shaded green for "good" and red for "bad."

Discussing the table further, we can initially look at the issues where biometrics compare favorably. In relation to the first point, about making weak choices, biometrics basically avoid the issue, as the user has no choice to make as the biometric is inherently part of them. This factor also underpins the next few issues as well. The user can forget a secret and they can find themselves without a physical token (e.g., having forgotten to take it with them or having misplaced or lost it entirely). By contrast, a biometric cannot be forgotten or misplaced in the same way. Having said this, the user's *ability* to use it can be temporarily impaired through injury or permanently prevented via a more severe incident, but these aspects are reflected in the later point around change and replacement. Meanwhile, in terms of sharing, the user can liberally give away passwords, PINs, and other secrets if they so choose without having any impact on their own ability to continue to use them (noting that this includes contexts in which users *elect* to share, such as with friends and colleagues, as well as where they may be *tricked* into doing so, such as with social engineering). They can also share tokens, albeit causing themselves potential inconvenience in the meantime because they no longer possess the means to prove their own identity. Biometrics cannot be given away in this manner, and so if

Table 2. Comparing the characteristics of different modes of user authentication.

	Something the user …		
Issue	**Knows**	**Has**	**Is**
User can make weak choices	✓	×	×
Can be forgotten/lost by the user	✓	✓	×
Can be shared with others by the user	✓	✓	×
Can be discovered/stolen by the attackers	✓	✓	×
Easy to use across multiple accounts	×	×	✓
Can be copied/cloned/impersonated	✓	✓	✓
Can be changed/replaced if compromised	✓	✓	×

the user wishes to permit others to gain access, then they need to use a legitimate means of delegating their rights. The fact that they cannot be willingly given away also limits the potential for them to be discovered or stolen by attackers — while it would be feasible to take someone's password (by watching them type it or logging their keystrokes) or steal their card, biometrics are not exposed to this risk in the same way. Finally, the point about usage across multiple accounts relates to the fact that we commonly use a multitude of devices and online services. If each requires a distinct password or a different token, then this becomes a practical challenge to manage. By contrast, biometric authentication allows the same characteristic to be used across multiple locations without introducing any further overhead for the user.

Looking at the two areas in which biometrics do not fare as positively, they start with the potential for things to be copied, cloned, or impersonated. All of the techniques are potentially susceptible, but the ease of doing it will depend upon the technique involved and the robustness of implementation (and the potential in the context of biometrics is examined later in the discussion). However, the potential for this to still happen links to the most significant issue with biometrics — they cannot be changed or replaced in the event of a compromise. This means that the safeguarding of biometric data becomes key to preserving the viability of our digital identity.

On balance, biometrics would appear to have many advantages. These relate not only to overcoming some of the vulnerabilities inherent in other approaches but also to improving the overall usability of the resulting solution. The importance of the "user experience" element should not be underestimated in the security context, insofar as the less the user notices, the technology or feels they are being "bothered" by it (e.g., being expected to do too much or do something that they feel takes too long), then the more chance that they will continue to make appropriate use of it. The reason people chose weak passwords is not because they are looking to compromise security; it is because they are trying to make their own life easier — having something that they can remember and/or can input easily. If they reach the point of feeling security is too complicated, too time consuming, or simply asking too much, then it increases the chances of them disengaging due to security fatigue (Furnell and Thomson, 2009). If implemented with appropriate care and consideration, biometrics can enable a frictionless approach, in which the technology becomes transparent and non-intrusive from the user's perspective.

The question of how well biometrics have been implemented leads to a further question of *what* is being implemented. As such, it is relevant to look briefly at what is going on under the surface with biometric technologies, and the process is broadly similar regardless of the specific biometric being used.

The Biometric Process

Although the discussion of biometrics and identity does not require a deep understanding of how the individual technologies work, it is still useful to have an appreciation of the general principles of how biometrics are put into practice. It is particularly relevant to recognize that there are various potential error scenarios that impose practical limits on the extent to which a biometric can be considered an effective and reliable proxy for identity.

In all cases, the starting point with biometrics is to establish a template (also known as a reference profile) for the legitimate user. This involves capturing the biometric from the user and extracting the necessary distinguishing information. This process is referred to as enrollment, and the resulting template is then stored and used as the basis for subsequent identification or authentication operations as appropriate.[2] Later interactions with the user then involve a new biometric sample being captured and compared to their template in order to see if it matches. In practice, thanks to the potential variability of biometrics being measured, and the conditions under which measurements are captured, the later samples will not be expected to offer an *exact* match to the original template. As such, the system needs to determine whether the match is close enough to be accepted, which in turn requires some form of threshold to be defined in order to set the boundary between acceptance and rejection decisions. These processes are summarized in Figure 1, which depicts the sequence of activities involved in both the enrollment and verification phases of a biometric system.

[2]It should be noted that the exact way in which the profile is stored and used is an important element of the overall security, as well as having significant implications for the user's privacy. As such, it is an issue that is returned to later in the discussion, but for now it is sufficient to understand the role that the template plays in the wider process.

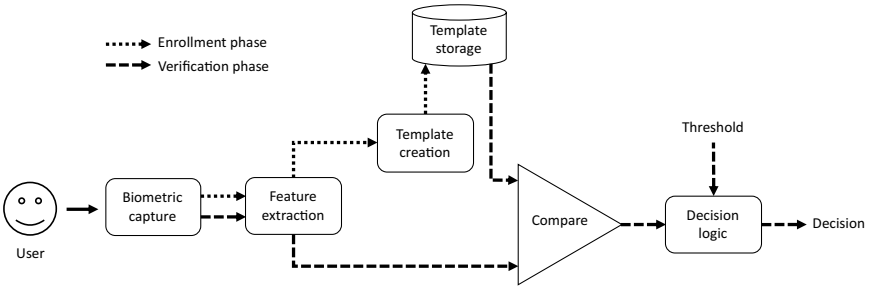


Figure 1. Biometric enrollment and verification processes.

It is impossible to discuss biometrics without also making reference to a series of associated error rates and failure scenarios that can be encountered. The most commonly highlighted relate to how the comparison process in Figure 1.

- **False rejection.** This refers to the scenario where the system fails to recognize the legitimate user and rejects them as an impostor. It is essentially a mistake by the system and occurs because the current biometric sample is not considered to be a sufficient match to the user's reference template. This type of error is also known as a *Type I error*, and its measurement is varyingly referred to as *False Rejection Rate (FRR)* and *False Alarm Rate (FAR)*.
- **False acceptance.** This refers to cases where an impostor is incorrectly believed to be the legitimate user and results from cases in which the biometric sample captured from the impostor is believed to be a close enough match to the legitimate user's template. In some cases, this could be due to a natural similarity (which could be particularly the case for biometrics that are less robust in terms of the "uniqueness" characteristic), but it could also result from deliberate attempts to deceive the system with false inputs or impersonation (e.g., fake fingerprints, voice recordings, or face photos or masks) — the susceptibility to which will depend upon the robustness of the system implementation. Following on from the terminology of the false rejection case, this scenario is also known as a *Type II error*, and the measurement of its occurrence is termed the *False Acceptance Rate (FAR)*, *False Match Rate (FMR)*, and *Impostor Pass Rate (IPR)*.

Astute readers will doubtless have spotted that these definitions leave us with two different interpretations for the acronym FAR in a biometric context, and moreover, the two variants are referring to the complete opposite of one another. As such, when encountering the FAR acronym as a measure, it is important to note whether the accompanying measure is FRR or IPR, in order to tell whether it is being used in a false acceptance or rejection context.

Adopting the FAR/FRR naming convention, the relationship between these errors is typically represented using a chart such as that shown in Figure 2, depicting a mutually exclusive situation in which improvements in one rate are typically made at the expense of the other. The basic reason for this is the decision on whether a biometric sample is considered to match the reference template is based upon the similarity threshold that one is willing to tolerate. A further element that is notable from the chart is the Equal Error Rate (EER). This is the point at which FAR and FRR coincide and is often the figure that gets quoted as a measure of the overall performance of biometric products.

In practice, techniques can be used to ensure that the situation is not necessarily as binary as this, and rather than completely permitting or denying the user, it is possible to regulate their level of access based upon the current level of confidence in their identity. Linked to this. It

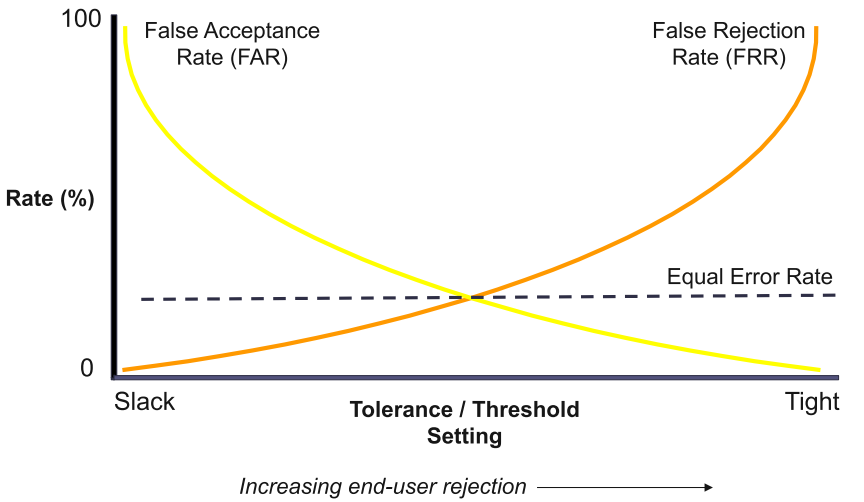


Figure 2. Relationship between false acceptance and false rejection errors.

can be observed that the two types of errors essentially reflect two distinct but desirable properties of the biometric system, namely, usability and security. False rejection ultimately concerns the former — if the legitimate user is rejected, then there is no impact upon the protection of the device or data concerned, but the users themselves are at risk of becoming frustrated or annoyed if the problem persists (with the ultimate risk that they reject the technology, which could lead to them disabling it and thereby removing the protection it would otherwise provide). Meanwhile, false acceptance is a direct compromise of the security, which is actually the main role that the technology is meant to be fulfilling.

In addition to the errors relating to the sample-to-template comparison, there are also two failure scenarios that can actually prevent the system from getting to this stage:

- **Failure to Enrol (FTE).** In this scenario, a candidate user is unable to register with the biometric system. In some cases, this may be a transient issue, perhaps because the user failed to provide enrollment samples in an appropriate manner, and so a re-enrollment attempt could meet with success. However, there are also more fundamental scenarios in which the user finds themselves unable to provide the biometric concerned (e.g., a small portion of the population does not have discernible fingerprints).
- **Failure to Acquire (FTA).** This relates to the situation in which the system is unable to capture a biometric sample from which to perform a comparison against the template. Common scenarios that readers may recognize will include a mobile device's camera not being angled properly to see the user's face or a fingerprint sensor not working properly in the rain. Such situations do not reflect a fundamental inability to work with the biometric and so can typically be rectified quickly. Nonetheless, from the user's perspective, the occurrence of an FTA will essentially *feel* the same as a false rejection and so will contribute toward their overall view of how intrusive the approach is in practice.

The FTE scenario is notable when rolling out biometrics across an overall population, as it is important to consider the proportion of users that may find themselves excluded from participating. In some cases, it is not a question of being physically unable to provide a biometric but that it is culturally incompatible to do so. A clear example here comes from

cultures in which it is commonplace for face coverings to be worn in public.[3]

Assuming that they have been able to enroll, then the user experience of biometrics in practice will be informed by the following:

- The combined False Rejection and Failure to Acquire rate (i.e., the extent to which any errors are encountered);
- The throughput of the system (i.e., whether there is any noticeable delay in operation).

A philosophical observation can be made at this point, insofar as none of these practical factors alter the fundamental fact that a given trait *belongs* to the user, and so the extent to which it represents or is part of their identity does not change. However, the practicalities of implementing and operationalizing biometrics clearly affect the ability to *use* the traits for identity purposes, and so their perceived linkage to the user is potentially weakened in practice.

From Magical to Mundane — The Evolution of Biometrics

The notion of using human characteristics to establish or verify identity goes back many years, with the most commonly recognized example being the use of fingerprinting as part of policing and criminal investigation. Of course, we have at least two examples that go back even further than this, given that people can recognize each other from their facial appearance and the sound of their voices. The key point about biometrics is that the process of capturing and comparing the characteristics is done *automatically* by the technology rather than via a process of manual analysis.

[3]At the time of writing, in the midst of the global COVID-19 pandemic, it is notable that this particular issue has suddenly become a consideration for a much wider population of users as a result of facemasks being required in a variety of public spaces (with some countries requiring coverings to be worn anywhere in public, while others restrict it to enclosed or indoor spaces such as transport and shops). As a consequence, the face recognition that many users will have become routinely accustomed to using in a fairly transparent manner on their smartphones will suddenly have become explicitly noticeable, as it fails to see their face and demands alternative authentication instead.

The last three decades or so have seen significant developments and advances in biometrics, enabling them to make the transition from being the technology of tomorrow to becoming very much the technology of today. It is interesting to track this evolution and consider how the technology progressed from something that people may have heard about to something that they regularly use. As such, the following paragraphs present a series of decade-by-decade snapshots in order to illustrate how things have changed in terms of both the availability of the technologies and the level of public awareness about them.

- The 1980s — At this point in time, the main coverage of biometrics tended to be in scientific reports and academic papers, and it was very much discussed in the guise of being the technology of tomorrow. It would be fair to say that general public awareness of biometrics was likely to be limited to what they might see in science fiction movies and the like. Indeed, the author's first recollections of biometrics were Captain Kirk's use of retinal scanning in the 1982 Star Trek movie, "The Wrath of Khan," and the "eye print" scanning used in the 1983 James Bond film "Never Say Never Again."
- The 1990s — Early commercial products. Examples included external fingerprint readers that could be purchased as add-ons for PCs.
- The 2000s — Commonly found fingerprint-based approaches being built into Personal Digital Assistant (PDA) devices, which were something of a precursor to the mobile technology that we now see within smartphones.
- The 2010s — By the start of the decade, we were witnessing the emergence of biometrics within smartphones, with both of the leading platforms (Google's Android and Apple's iOS) incorporating support for biometrics (principally face or fingerprint recognition) as a basis for user authentication. By the end of the decade, biometrics had become a default provision on many devices and a standard feature available to the public at large.
- The 2020s — We now see biometrics as being very much an expectation, and this is reflected in the technologies being sold. As an example, even at the start of the decade, Apple does not sell an iPhone, iPad, or MacBook that does not have either Face ID (facial recognition) or Touch ID (fingerprint recognition) built in as standard. As time goes on, the expectation for devices to recognize and respond to us without requiring any particular effort on our part is likely to mean that

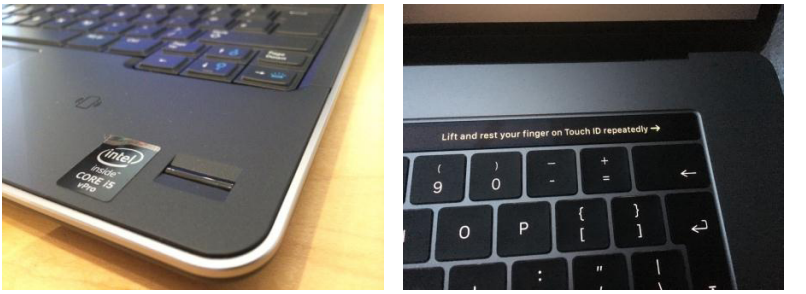
encountering technologies that demand more explicit efforts to identify or authenticate ourselves (e.g., passwords) will feel increasingly anachronistic.

The overall path toward this increased usage can be characterized by gradual steps toward making the techniques more reliable and more usable, which has widened the opportunity for deployment. While there have been advances in all areas of the technology, it would be fair to say that consumer products have tended to focus upon physiological approaches, with fingerprint, face, and iris recognition methods all having found favor in smartphones. By contrast, the only behavioral approach to have enjoyed similar attention is voice verification, but this is far less prominently featured than fingerprint and face recognition (its use has likely exceeded iris recognition purely on the basis that fewer devices currently incorporate the more specialized sensors required to support this approach). However, while the availability and use of physiological methods have grown, it is notable that smartphone-based opportunities to use voice verification have reduced over time. Most notably, Google elected to reduce the level of access granted to Android devices based on voice verification for security reasons (i.e., because the voice-based method was more susceptible to compromise via false acceptance or impersonation than other unlock routes). Prior to this, versions 5–7 of Android had enabled Voice Unlock of devices via Google Assistant (which recognized the legitimate user's voice and fully unlocked the device in response to them saying "OK Google"). However, from Android 8 onwards, this feature was limited to allowing access to a significantly restricted range of functionality, and the user had to unlock their device by other means to use the Assistant more fully (Fisher, 2019).

The successful use of biometrics does not solely depend upon the existence of suitable technology — it is also important to think about how it is integrated into the device or service that is being protected. This is exemplified by the way in which Apple introduced fingerprint recognition (Touch ID) within its mobile devices. Fingerprint readers had already been added to a range of laptops, PDAs, and smartphones by this point, but there was arguably something awkward about the way in which it had been achieved, as the sensor was typically a distinct element that had no other purpose and using it was essentially introducing a change to the users' normal behavior. For example, you could variously find sensors being included in a range of locations, including the corner of laptop

keyboards (as shown in Figure 3(a)) or on the back of smartphones, and so from the user's perspective, they had a visibly distinct "security thing" that they would only use in order to do something security related. However, Apple's approach was notably different and sought to integrate the technology in a manner that was less intrusive. When Touch ID was first introduced on the iPhone 5S, it was placed within the phone's Home button, which users would already be pressing to activate their device. Similarly, when it arrived on the MacBook, it was done by integrating it into the power button (see Figure 3(b)), which was used to wake the computer from sleep and so could authenticate the user in the process. As a result, the user experience of biometrics was more natural, and they were no longer seeing a distinct sensor being used solely for security.

It is not just the *presence* of the technology in the device but the linking of this technology to associated actions and services that require security. Again, using Apple as an example,[4] the biometric is not only used to unlock the device but can also be used for a range of other actions across the user's enabled devices, including confirming payments in Apple's online stores (e.g., iTunes and App Store), verifying purchases made using Apple Pay (online or in physical stores), auto-filling passwords for websites and services, and opening password-protected documents.



(a) (b)

Figure 3. Fingerprint readers on laptops with (a) a separate sensor and (b) a sensor integrated within the power button.

[4] Apple tends to provide a good example in these contexts because their control of both the hardware and software elements in their products means that they can offer a tighter level of integration of technologies across their platform than others may be able to achieve. This sometimes places their approach ahead of what is offered by competitors but nonetheless points toward what is likely to become the norm moving forward.

In parallel with their use on our devices, biometrics have also become commonplace in other contexts — with perhaps the most notable context being passports and identity cards (i.e., the established identity documents that we have historically used in society are now routinely incorporating biometric data). Many readers will have seen and experienced the impact of this in practice, particularly in relation to international travel, with airports now offering automated passport control gates that use facial recognition to confirm the identity of the traveler instead of the traditional manual inspection by border control staff. While the technology tends to work rather well, it also tends to offer a good context to see the challenge of integrating biometrics in practice. Even though many people sail through smoothly, you still regularly witness people failing to be recognized and having to try again — sometimes to the extent of being rejected altogether and having to go for a manual check instead. The reasons for difficulties do not necessarily relate to the biometric technology itself — some people encounter problems because they have not inserted their passport properly (i.e., the system does not have access to their template information), while for others, it is because they do not position themselves correctly in front of the sensor (leading to a "failure to acquire" situation). Some even fail simply because they have tried to use an automated gate when they do not actually have a biometric passport in the first place (in effect, experiencing an indirect version of a "failure to enroll" scenario). However, regardless of the cause, the effect may ultimately be perceived to relate to the performance of the system and undeniably has an impact on its throughput

Biometrics have essentially gone from being a wondrous notion that seemed like the stuff of science fiction to something that is essentially transparent, and that people routinely use without even thinking about it. In fact, the only time that many users will *notice* their biometrics is when they fail to work properly (e.g., if their face or fingerprint has not been properly captured and recognized when trying to access their smartphone). The techniques themselves are now better and the technologies are more widely deployed. Indeed, biometrics are now absolutely a normal and accepted part of regular IT usage and they can be found in a variety of everyday consumer-grade devices. The obvious consequence of this is that millions of people are now routinely using biometrics on a daily basis, with related personal data being captured and stored as a result.

Protecting Our Identity and Our Biometric Data

Given that they depend upon the collection of data representing some of our most personal characteristics, the use of biometrics raises questions around related protection. This concerns both how well the biometrics safeguard against the misuse of our identity as well as how well the biometric data themself are protected from compromise.

In terms of preventing the misuse of our identity, one of the most significant considerations in this context is the extent to which our identity would be perceived to match with someone else, which links back to the notion of False Acceptance errors discussed earlier. The following is how Apple describes the uniqueness properties of its Touch ID and Face ID methods (Apple, 2020a):

> *The probability that a random person in the population could unlock a user's iPhone, iPad, or Mac is 1 in 50,000 with Touch ID and 1 in 1,000,000 with Face ID. This probability increases with multiple enrolled fingerprints (up to 1 in 10,000 with five fingerprints) or appearances (up to 1 in 500,000 with two appearances). For additional protection, both Touch ID and Face ID allow only five unsuccessful match attempts before a passcode or password is required to obtain access to the user's device or account. With Face ID, the probability of a false match is different for twins and siblings who look like the user and for children under the age of 13 (because their distinct facial features may not have fully developed).*

However, it is worth considering whether the "random person" is the threat that we are most concerned about? We ought to typically be more concerned about those people regularly around us — family, friends, and colleagues — who are the most likely to have the opportunity for physical contact with the device. How resilient is the approach to *them*? For example, while your family members will be in no better position to gain false acceptance via their own fingerprints than the "random person" would, they are in a notably better position to compromise *yours*. This is well illustrated by cases of fingerprint authentication being performed while the legitimate user is asleep — several of which ended up making headlines. One example was the case of a wife who unlocked her sleeping husband's fingerprint to unlock his phone during a flight and then forced an emergency landing upon reading messages and discovering he had

been cheating on her (*Sky News*, 2017). Meanwhile a less dramatic, but still headline-grabbing incident involved a toddler who unlocked her mother's phone and made a variety of Pokémon toy purchases while she was sleeping (Ng, 2016).

Of course, both of the previous examples relate to situations in which the "attacker" had physical access to the device, and where the legitimate user was still involved (albeit passively and involuntarily) in the process. What is perhaps of greater concern is the resilience to targeted attack and/or a determined adversary, and in scenarios where the biometric data may be exposed to a more fundamental breach that undermines its potential usage in the longer term.

The protection and safeguarding of biometric data have historically been issues of concern. For example, in a survey back in 2006, the author sought to explore public attitudes and perceptions around what was then still very much an *emerging* technology (Furnell and Evangelatos, 2007). While there was a reasonable degree of awareness of biometrics, with two-thirds of the 209 respondents claiming to be aware of them (with the most recognized techniques being iris, fingerprint, voice, retina, and hand- and face-based methods), there was far less evidence of practical experience in using them. Overall, only around 10% of respondents claimed to have actually used any of the most commonly recognized methods, with fingerprint biometrics being the most commonly encountered (a factor likely motivated by their fairly established use in higher-end personal digital assistant devices, such as the Compaq/HP iPAQ). However, what was very prominent in the findings was clear evidence of concern about protecting biometric data and a lack of confidence in how it might end up being used by those collecting it. This is illustrated in Figure 4, which depicts the respondents' level of concern around the potential theft of biometric data (noting that this reflects their concern that it could happen rather than how concerned they would be if it did happen), alongside their level of confidence that private organizations and government agencies would limit the use of the data for authentication purposes. It is rather notable that increasing levels of concern are largely mirrored by decreasing levels of confidence, suggesting that at this point in time the respondents had little optimism around the safeguarding of biometric data. By contrast, when asked about how easily they perceived biometrics could be cheated, half (49%) felt that it would not be easy at all, and only 6% considered it could be very easily (noting that 21% offered a "Don't know" response, acknowledging their lack of direct familiarity with the

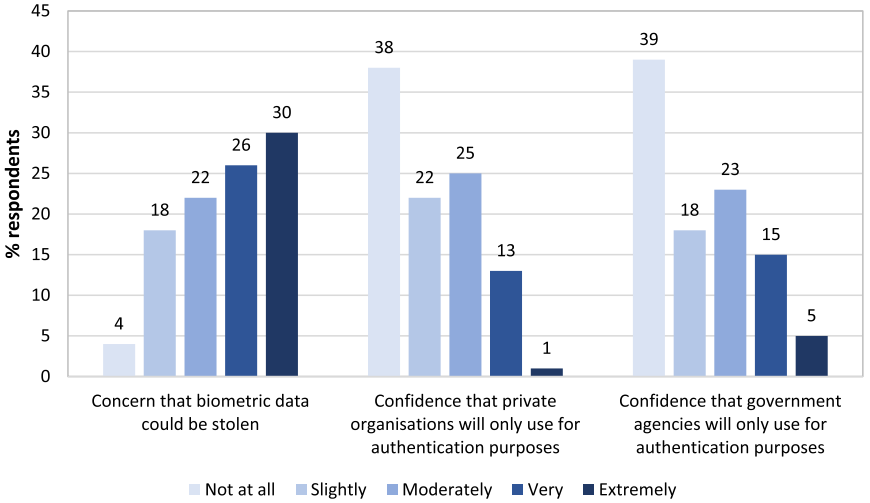


Figure 4. Concern and confidence around use of biometrics.

technologies involved). Overall, therefore, it is clear that the majority of concern was not linked to the *capability* of the technology but rather to the security and use of the resulting *data*.

Of course, these findings were essentially around instincts and expectations rather than reflecting upon personal experience. A decade and a half later, we are clearly making more use of the technology already, and so either the concerns have gone away or our use is happening in spite of them. The reality is likely closer to the latter, and this usage despite ongoing concerns can occur for the following various reasons:

- The benefits are considered to outweigh the perceived risks (e.g., the usability advantages of biometrics on mobile devices compared to entering passcodes);
- We have no real choice in the matter and the use has become non-optional (e.g., the issuance of biometric passports);
- There is an implicit sense of safety in numbers (e.g., with so many people using biometrics, we assume that any problem will affect a large community, and this will somehow lead to us being shielded from negative impact).

Overall, while the use of technology has matured and extended, the concerns and challenges have not gone away. Indeed, there are still a

variety of concerns around theft, accuracy, and potential misuse of biometric data (Wood, 2020). This in turn leads to considerations about what ought to be done to protect this information.

The fact of dealing with what could be seen as the highest level of personal data means that it needs to be handled with so much more care. It only takes one failure or weak link to compromise it, and there have been many instances of security breaches leading to large scale password exposures that have demonstrated the frequent lack of ability to safeguard traditional credentials.[5] However, it is worth considering the extent of risk that biometrics would be exposed to in a similar scenario. Assuming the data are not stored in plaintext, and what the attacker has acquired is an encrypted or hashed version of the data, then it is already a rather different situation to the one they would face with a password. With the password, it is not a case of the encryption or hashing being cracked but rather the attacker being able to find a plaintext string that hashes to the same value. If the user has chosen a weak password, then the attacker's tool will typically be able to see a pre-hashed version that they can find directly via a look-up table and then reveal the matching plaintext. If the password is not already in the list, then the tool needs to resort to a brute-force approach, progressively working through all possible combinations of plaintext characters and testing each one to see if the encrypted/hashed version matches the captured password. The same notion does not apply with biometrics — there are no "weak choices" for users to make — and so there is significantly more of a challenge for an attacker to find a match to a captured hash. Meanwhile, a brute force approach would be significantly more challenging, as the size of the biometric template would be considerably larger than a password and so would be less feasible to perform an exhaustive search for versions that encrypt or hash to the same value as the captured version.

As already highlighted earlier in the chapter, the breach of a biometric presents a fundamental problem to the user if their profile data are compromised, as it nullifies their own potential to use it legitimately in the

[5]A good example of this is provided by the HaveIBeenPwned website, which consolidates details of breached websites and allows users to check if one of their passwords has potentially been compromised as a result of such a breach. The list of "pwned" sites (i.e., those breached by attackers) is extensive and illustrates the extent to which passwords can be vulnerable even if users themselves may have followed good practice in selecting and managing them. See https://haveibeenpwned.com/PwnedWebsites for details.

future. Unlike a password or a token, they are not in a position to change or replace it (while it is acknowledged that this is not quite true for biometrics, such as fingerprint or iris recognition, where the user arguably has a fallback option available if the original is compromised, this is still not a high degree of redundancy and may systems will permit and encourage template creation for both eyes or multiple fingers anyway).

This all leads to the question of how biometric data should be treated. This includes where it is stored, how it is stored, and whether it is communicated and shared over the network. By far, the most secure approach is to ensure that biometric data are stored in a protected format and held locally with the user. This avoids the risk of it being compromised in transit over the network or through a breach of a third-party system as has been the case with passwords. Following exactly this approach, Apple has made a point of emphasizing that biometric data for user authentication on its devices are held securely and never leave the device itself (Apple, 2020b):

> *Face ID data — including mathematical representations of your face — is encrypted and protected by the Secure Enclave ... Face ID data doesn't leave your device and is never backed up to iCloud or anywhere else.*

In this context, the "secure enclave" is a secure coprocessor, isolated from the main processor and providing an extra layer of security (Apple, 2021). As a result, apps and services accessed on or through the device do not have any access to the user's biometric data itself but can call upon the biometric processes to perform authentication and return an associated pass–fail decision in order to grant or deny access. As such, the biometric techniques are available for use across the device, without any personal data being shared and potentially exposed.

The principle that this personal data stays as close to the person as possible is certainly the best way of handling the situation but is by no means guaranteed to be the manner in which all providers will approach the issue.

Wider Challenges

As the earlier discussion indicated, biometric technology and our use of it have advanced significantly in the last few decades. However, we are still far from a point at which all of the problems have been solved. A number

of notable issues remain to be faced from both the technology and user perspectives, and while some can potentially be addressed by further advances in the technology, others represent more fundamental issues around how the technology itself is perceived.

User trust

Despite the much more widespread usage, one of the fundamentals is still around ensuring trust. However, this is generally not related to the user's trust in the protection or security that the technology provides to the target device or service but rather to the risk of their biometric data getting exposed or misused (including the potential for misuse by the companies legitimately collecting it). Having seen numerous instances of passwords being exposed as a result of compromise on the server side, many users start out with an instinctive concern that the same risk could exist with a similar repository of biometric data. Moreover, if the data were to be stored in this way, they would be right to be concerned — which leads to the need for biometric data to be managed fundamentally differently from the historic practice with passwords.

Universal availability

Even now, almost a decade after they started to become a common feature in smartphones, we are still not seeing biometrics by default. They are not yet a standard feature on all the devices we use, and it is still perfectly possible to buy smartphones, computers, and tablets that do not have biometric capture capability. And even when they do have biometrics, they are not guaranteed to have the same one — with factors of cost or practicality often dictating what gets chosen. The earlier discussion used Apple as an example of successful integration of biometrics into all of their product range of mobile devices (i.e., smartphones, tablets, and laptops). However, it is notable that they are not using the *same* biometric — some use Face ID and some use Touch ID, and (at the time of writing) *none offer both*. So, even users that have stayed within the Apple ecosystem can find themselves using multiple biometrics depending on the type and generation of device(s) they are using.[6] It is important to recognize that the

[6]To illustrate the diversity and explain the reason behind it, somewhat further, Apple's original adoption of biometrics was focused on fingerprints, with the Touch ID technology.

main reason for the persistence of password-based approaches is nothing to do with their security but rather that we can rely upon them being usable because almost devices offering a means of text-based input. Specific biometrics still require an accompanying level of specific technology to be present on the device.

Reliability and user experience

While biometrics are readily deployed within mobile devices, their operation is not flawless and the opportunity to use them is not always available. There are still regular error scenarios (even if they are more likely to be caused by a failure to acquire than a false rejection) and there are still situations in which certain biometrics cannot be used (e.g., if the sensor is totally unable to take a measurement).

The recognition that biometrics cannot always be relied upon means that they cannot be the only identity mechanism on offer. Indeed, at the time of writing, all implementations to date have still incorporated a knowledge-based approach as a fallback and underlying "master" technique (i.e., while the user may well have a smartphone with a fingerprint or face recognition capability, at the end of the day, the identity verification on their device ultimately depends upon a passcode or similar).

The technologies are getting better in terms of security, reliability, and performance — they work better for the intended users (e.g., operating more quickly and more often) and are more resistant to spoofing by impostors. However, they are still not a panacea. No matter what new device or sensor you have, fingerprint recognition is not going to work through a pair of standard gloves, and face recognition will not work if you have a scarf wrapped over your face (using two examples that are likely to be familiar to readers living in cold or wintery environments!). Indeed, to use a topical example at the time of writing, the limitations of face recognition on

Following initial use in the iPhone 5S (2013), related sensors were later introduced within devices in both the iPad and MacBook product ranges. Touch ID then continued to be the primary biometric within the iPhone range until the release of the iPhone X (2017), when the desire to have a full edge-to-edge display necessitated the removal of the Home button within which Touch ID had been integrated. So, while the technique persisted within the iPhone 8 (launched at the same time), as well as within later lower-cost iPhone models that retained the Home button, the design choice relating to the screen meant that the premium price iPhone (and iPad) models moved over to using Face ID instead.

smartphones became apparent during the COVID-19 pandemic, with the mass adoption of face masks causing key facial features to be obscured and forcing users to use passcodes to unlock their devices in public places when their full faces could not be shown (Collins, 2020). Of course, users of smartphones using fingerprint-based biometrics instead of face would have experienced no change, and this helps illustrate how (at the operational level) the choice of biometric can be significant, as well as how (at the conceptual level) a single biometric is actually a long way from being a true proxy for the user's identity.

User compromise

While biometrics remove the risk of users making weak selections or sharing their credentials as they might do with passwords, they do not completely remove the potential for users to introduce complications into the process.

Consider the following scenario, based upon a real-life example of a household in which mum, dad, and their three children all share the same tablet device. One of the children is 16 years old and so is trusted to use the device without supervision, but the parents do not want the two other children (both under 10) to have the same access. The tablet does not offer any support for setting up user accounts[7] but uses fingerprint recognition to control access. The workaround that the family adopts is to register index fingerprints from all the adults so that all three of them have a means to unlock the device, while the younger children remain locked out and need to ask permission to use it. While this makes sense from the users' perspective, as they are enabling controlled access to a shared device, it is essentially compromising the biometric profile. The template no longer relates to a single user — it represents a group. And because the device is not designed for multiple users, all three users now have equal access to any other features for which the biometric is enabled in addition to unlocking the device, e.g., if the biometric is used for authorizing payments and purchases, then all users have the same ability to do this as well

[7]At the time of writing, this would be the case for a tablet, such as the Apple iPad, which is designed as a single-users device, unless deployed in a special education mode designed for schools. Android-based tablets allow multiple accounts to be set up by default and so avoid this particular scenario.

(which may not sit so comfortably with whoever of the three is ultimately responsible for the payments!).

Additionally, while biometric implementations on devices continue to use a passcode-based approach as the underlying "master" approach, there is still potential for users to introduce a fundamental weakness by selecting a weak passcode (e.g., with many devices requiring a 6-character passcode, many users may confidently elect to use codes such as 123456 on the basis that they will be using fingerprint or facial recognition to provide security). As such, people can end up in the rather incongruous situation of *believing* that their device is protected via advanced biometric protection, whereas in actual fact, all that an impostor ultimately needs to do is enter the extremely weak passcode.

Potential Issues for Regulatory Consideration

The limitations, challenges, and concerns around biometrics lead to a number of considerations in terms of what might usefully be done to control and potentially regulate the use of the technology. The rationale here would be a combination of holding providers to account and establishing expected standards in what they should deliver, alongside providing a foundation upon which citizens can base their trust and feel reassured in their use of the technology.

Perhaps unsurprisingly, the key areas in which attention should be given are related back to some of the key issues that have already been raised within the earlier discussion. Broadly speaking, they relate to issues around the assurance of biometric technologies and devices and to the appropriate safeguarding and usage of biometric data.

- **Requirement to use and comply with "Secure by Design" standards for devices incorporating biometric sensors and collecting biometric data**
  Standards for security by design already exist, with much recent attention having been targeted toward security in smart devices (DCMS, 2020). The potential role for regulation here would be to ensure that standards and guidance are appropriately utilized in the design and deployment of biometric technologies and the devices that use them. In this context, it is not just the biometric technology itself that may be of concern (e.g., the camera or fingerprint sensor) but rather the way that

it has been integrated within a device alongside other hardware and software components. The whole system needs to be realized securely in order to prevent compromise of the biometric process or data via unanticipated backdoors.

- **Ensuring the use of standards for acceptable levels of testing and performance for biometric devices**
  If we are to trust biometrics, then we need to trust the technology that implements the approaches, and this points to the desirability of standards and regulations to ensure that they are implemented and deployed to an appropriate level in terms of rigor of testing (to ensure that the technology works correctly and is free of detectable vulnerabilities) and performance (to ensure that it works to an acceptable level of accuracy in order to provide the expected security). As an example, this could include the establishment of minimum acceptable error rates to permit the use of given techniques in different contexts (e.g., personal devices and sensitive applications). There are already standards for representing the extent to which security functionality has been tested within the design, development, and implementation of hardware, firmware, and software products (e.g., the seven Evaluation Assurance Levels within the Common Criteria standard (Common Criteria, 2017), which express levels of assessment from functionality testing through to full formal verification of the design and testing). Similarly, there is good practice guidance that would inform acceptable performance, and so these aspects would not necessarily need to be regulated in a more formal manner. However, there may at least be a case for industry self-regulation, and the establishment of clear expectations that if biometrics are to be deployed in scenario X, then they ought to have been tested to at least level Y and perform to at least level Z.
- **The need for appropriate security and protection to be applied to the processing, storage, and any transmission of biometric data**
  Biometric data are already specifically recognized under some existing personal data protection legislation, such as the European General Data Protection Regulation (GDPR), where it is denoted as a "special category of personal data" under Article 9 of the regulation (European Parliament, 2016). However, while GDPR expresses the conditions under which it is acceptable to process the data, it does not express anything further in terms of the security provisions that ought to be made. While it may be impractical to regulate in terms of requiring specific technologies, it would be feasible to express the forms of

protection that ought to be applied. As an example, it may be stated that *data at rest must be held in encrypted form and protected to* "*military-grade*". This would make it clear that encryption is needed while at the same time abstracting the requirement away from a specific encryption technology (such as AES-256 — the Advanced Encryption Standard using a 256-bit key — which constitutes "military-grade" at the time of writing) and instead permitting the use of any approach that meets the requirement at the given point in time.

- **Control the permitted sharing or disclosure of biometric data**
  The more the data are shared, the more they are exposed to potential points of compromise and so there ought to be clear bounds over what can be shared with whom and under what conditions (with the default position likely being no further sharing with anyone other than in explicit, legally defined exception cases). Note that this issue is distinct from the technical controls applied to any actual transmission of biometric data and is instead aimed toward defining conditions under which such transmission (or sharing via other means) may be permissible.

Conclusion

Biometrics are an inevitable part of the direction of travel with digital identity and authentication. The ability to rely upon something that the user is, rather than requiring them to remember a secret or carry something with them, makes it inherently easier from the human perspective. However, while it is tempting to herald biometrics as salvation in terms of ease of use and making security experience more acceptable, it is important not to lose sight of the additional risks that they bring alongside the benefits. The increased personalization of the technology in the name of enhancing security is ultimately putting a lot on the line in terms of personal privacy. By contrast, approaches, such as traditional passwords, may well be poor from several perspectives, but users are giving away very little data and can change it easily.

The effective and secure use of biometrics depends upon the effective and secure design of the overall *system* in which they are working. Like cryptography, the protection afforded by biometrics can be let down by bad implementation. Unlike cryptography, however, if a weakness is found, we cannot just put things to right by using a new key or changing the algorithm.

Although it is clearly possible to identify risks and downsides, the use of biometrics demands an optimistic approach. Having taken decades to get here, with the progressive refinement and commercialization of the approaches, the wider adoption and use of the technology are inevitable. Rather than attempting to hold back the tide by resisting the use of the technologies entirely, our interests are best served by efforts to ensure that deployment is done properly and that the data are not misused.

References

Apple (2020a). About Face ID advanced technology. *Apple Support*. 2 March 2020. https://support.apple.com/en-gb/HT208108. Accessed on 7 April 2021.

Apple (2020b). Touch ID, Face ID, passcodes and passwords. *Apple Platform Security*. https://support.apple.com/en-gb/guide/security/sec9479035f1/web. Accessed on 7 April 2021.

Apple (2021). Secure enclave overview. *Apple Platform Security*. https://support.apple.com/en-gb/guide/security/sec59b0b31ff/web. Accessed on 7 April 2021.

Collins, K. (2020). iPhone 12 failed to address how Face ID is useless in the age of coronavirus. *Cnet.com*, 20 October 2020. https://www.cnet.com/news/apple-iphone-12-no-touch-id-button-face-id-useless-in-age-of-coronavirus-wearing-masks/. Accessed on 7 April 2021.

Common Criteria (2017). *Common Criteria for Information Technology Security Evaluation - Part 1: Introduction and general model*, Version 3.1, Revision 5. April 2017. https://www.commoncriteriaportal.org/files/ccfiles/CCPART1V3.1R5.pdf.

DCMS (2020). *Secure by Design*. Document Collection. Department for Digital, Culture, Media and Sport. 16 July 2020. https://www.gov.uk/government/collections/secure-by-design.

European Parliament (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*. 4 May 2016. http://data.europa.eu/eli/reg/2016/679/oj. Accessed on 7 April 2021.

Fairhurst, M. (2018). *Biometrics: A Very Short Introduction*. Oxford: Oxford University Press. DOI: 10.1093/actrade/9780198809104.001.0001.

Fisher, C. (2019). 'OK Google' will no longer fully unlock your phone. *Engadget*. 1 March 2019. https://www.engadget.com/2019-03-01-ok-google-voice-match-unlock-update.html. Accessed on 7 April 2021.

Furnell, S. and Evangelatos, K. (2007). Public awareness and perceptions of biometrics. *Computer Fraud & Security*. January 2007, 8–13.

Furnell, S. and Thomson, K. L. (2009). Recognising and addressing "security fatigue." *Computer Fraud & Security*. November 2009, 7–11.

Jain, A. K., Bolle, R. and Pankanti, S. (1999). *Biometrics: Personal Identification in Networked Society*. Amsterdam: Kluwer Academic Publications.

NCSC (2019). Biometric recognition and authentication systems. *National Cyber Security Centre*. 24 January 2019. https://www.ncsc.gov.uk/collection/biometrics/choosing-biometrics. Accessed on 7 April 2021.

Ng, A. (2016). Child uses sleeping mom's fingerprints to buy Pokemon gifts. *Cnet.com*. 27 December 2016. https://www.cnet.com/news/child-uses-sleeping-moms-fingerprints-to-buy-pokemon-gifts/. Accessed on 7 April 2021.

Sky News (2017). Flight diverted after woman unlocks husband's phone and discovers affair. 7 November 2017. https://news.sky.com/story/flight-diverted-after-woman-unlocks-husbands-phone-and-discovers-affair-11117184. Accessed on 7 April 2021.

Wood, S. (2020). Biometric authentication: The importance of transparency and trust. *ITProPortal*. 7 April 2020. https://www.itproportal.com/features/biometric-authentication-the-importance-of-transparency-and-trust/. Accessed on 7 April 2021.

https://doi.org/10.1142/9781800612860_0007

Chapter 6

National Cyber Policies Attitude Toward Digital Privacy

Tal Pavel

Arguing that you don't care about the right to privacy because you have nothing to hide is no different than saying you don't care about free speech because you have nothing to say.

— Edward Snowden

Introduction

The digital age places opportunities and innovations alongside many challenges and dangers in a variety of aspects and with an emphasis on information which is the basic part of this age based on information creation, processing, storage, and transmission. This includes personal, organizational, and state information, such as that generated or managed by private, business, and government agencies. This era is characterized by two main actors that gather information: (1) private companies that are global information giants and hold in their possession extensive, detailed, valuable information, over many years, from a variety of sources, in the knowledge of the user and especially in his lack of knowledge; along with (2) various state bodies that collect a wealth of information. All of this is usually to give us one or two in exchange for the same privacy and with a constant erosion in it: (1) services that are customized

for us more and more precisely and (2) security provided to us by various government bodies. Information in the online age and the way it is handled pose many challenges to those who collect and manage it as well as use it, as well as those on whom the information is collected. This is mainly since this information is collected, processed, preserved, transmitted, and in many cases finds its way outside the boundaries of the entity responsible for it, whether accidentally or intentionally, by internal elements as well as by various hostile elements, including hackers, cybercriminals, hacktivists, terrorists, as well as various state actors for a wide variety of reasons. To the point where our digital information is so widely available that it is now possible to purchase all the information about a person's digital life in small amounts of up to a few tens of dollars.

As a result, various experts and researchers have been raising the question for many years whether anonymity has died in the online age. And should not these ideas of privacy and anonymity be abandoned today (Brøndmo, 2004; Popkin, 2010; Pagliery, 2013; Campagnano, 2015; Lee, 2015; Lufkin, 2017; Mims, 2018; Mance, 2019; Zibuschka *et al*., 2019; Raywood, 2020; Sahota, 2020; Kind, 2021). Whereas most people today have no awareness and interest in the constant erosion of our privacy and anonymity to the point where we are completely transparent in this online age and sometimes even willing on the part of users to give up their digital privacy for a paltry sum of dollars and vice versa, their unwillingness to pay to maintain their online privacy (Bellman *et al*., 2010; Beresford *et al*., 2012; Acquisti *et al*., 2013; Christin *et al*., 2013; Prince and Wallsten, 2020).

Considering all this, there is a need for current countries to pay due attention to the issue of digital privacy security and the need to protect it so that the freedom of the Internet and the privacy of its users are not violated. Thus, the study aims to examine the importance of the issue of online privacy today among the leading countries in cyberspace, as can be seen from their national cyber policy documents. To this end, the study will examine three research questions from the field of privacy in the digital age: (RQ1) What is the degree of reference between the countries that are ranked as world leaders in cyberspace to the issue of privacy in cyberspace?; (RQ2) Is there a difference in the way cyber privacy is viewed by the various leading countries in the cyberspace?; (RQ3) Is there a reference in those national cyber policy documents to the impact of the COVID-19 era on the issue of privacy in cyberspace?

In light of these research questions, the importance of research is in drawing a picture of the current situation in several aspects: (1) mapping and analyzing countries that are considered global leaders in cyberspace based on various indices, to arrive at a uniform and clear picture as possible; (2) mapping and analyzing the national cyber policy documents of those leading countries in cyberspace and thereby learning about the extent of such documents among those countries, their level of up-to-dateness, and the entities that publish them; (3) mapping and analyzing the extent and manner of addressing the issue of privacy in these cyber policy documents and the need for government institutions to maintain it, balancing with the need to meet cyber challenges and threats posed by those various actors.

Therefore, the premise of this study is that there will be a lot of reference to the issue of digital privacy in national cyber policy documents of countries that are world leaders in cyberspace because of being cyberspace world leaders and, therefore, their high awareness of broad aspects related to cyberspace, including the privacy.

Literature Review

The explosion of information in the digital information age — The digital space is an environment in which a lot of information is collected, processed, transmitted, and stored, which is currently the most valuable resource and is a growing challenge to privacy in this digital age (Mulligan and Berman, 1999; Long and Quek, 2011; Jossen, 2017). Nowadays, this information is so vast that we cannot comprehend its scope: In May 2012, an IBM study reported that the amount of digital knowledge in the world is 2.7 Zettabytes (ZB) (Karr, 2012) and claimed several years later that we produce daily 2.5 quintillion bytes of data (Marr, 2018). Over the years, various studies have been published that try to estimate the size of the Internet, in the context of the volume of information contained in it, some of which sometimes even contradict each other (Bartley, n.d.; Mitchell, n.d.; Bergman, 2001; Pappas, 2016; Marr, 2018; Petrov, 2021; Statista, 2021). In addition, a variety of data are published on the amount of information uploaded to the digital space or created in it, as well as the amount of information that is added to the Internet every minute or day, when this amount only increases over the years (Desjardins, 2017; Schultz, 2019; DOMO, 2020; Lewis, 2020; Petrov, 2021; Vuleta, 2021), along with the

sheer amount of information available on the Deep Web (Bergman, 2001; Pagliery, 2014; Taiwo, 2015; Chikada, 2016; Carapola, 2017; Pratham, 2019), and in the dark web, with different references to the volume of information in the dark web, the number of .onion addresses and its users (Brewster, 2015; Mani *et al.*, 2018; Stone, 2019; *Onion Services — Tor Metrics*, 2021). It is worth noting that alongside the criminal activity in it, the Dark Web and the TOR are also used to maintain the anonymity and privacy of users in the online space. This is for a variety of needs, mainly to protect the privacies of individuals and families, businesses, journalists, as well as for law enforcement agencies (*Tor Project | Anonymity Online*, n.d.; Dredge, 2013; Asn, 2015; Wyciślik-Wilson, 2019). However, various studies have raised the question of how much the use of TOR does give us complete protection of our online privacy and anonymity (Taylor, 2019; Beretas, 2020).

Threats to online privacy — A wide range of entities nowadays collect online information for a variety of needs, with an emphasis on commercial and information-intensive companies as well as various state players who track and spy on various entities through a variety of espionage programs, some private companies, mostly to fight terrorism and crime in their countries or to provide improved service to their citizens (Sauer, 2021). However, the information collected is in constant danger to be stolen or leaked by a variety of hostile elements and for various purposes: government officials (Bajak, 2021), hacktivists (Bagwe, 10AD; Williams, 2021), insiders who leak information for ideological or conscientious reasons (Szoldra, 2016; Clayton, 2021), and cybercriminals who steal information for financial gain (Hope, 2021; Millman, 2021). This is in addition to cases where various human errors have caused huge leaks of information (AFP, 2021; Muncaster, 2021). All of these create daily events of information leaks, in many cases without the awareness and knowledge of the victim and the public, and sometimes even mega-events in which hundreds of millions and billions of records have been leaked (Hill and Swinhoe, 2021). In most cases, this is an outcome of both the low level of the technological infrastructure of the victim (Osborn, 2015) and of appropriate awareness, knowledge, education, training, and education to identify security breaches, threats as well as the lack of preventive measures and dealing with cyberattacks and information leaks that are manifested inter alia through the preparation of procedures, risk management, and organizational policy (Hall, 2016; Aldawood and Skinner, 2019; Zwilling *et al.*, 2020; Ho and Gross, 2021).

Threats to information privacy and users in the COVID-19 era — This new era we are experiencing from the end of 2019 poses not only a variety of international challenges in a wide range of fields and to all countries of the world but also dangers to the physical and online privacy of citizens using various technologies around the world with the purpose to monitor the activities of the citizens and their location to maintain the isolation regulations as required by the regulations and restrictions to deal with the pandemic. This means that many countries have used against innocent citizens legal measures and various technological means which are mostly used to locate and deal with criminal and terrorist entities. This is in the form of the unprecedented use of technological means, including through a variety of dedicated applications, not to fight the enemies of the state but to create a comprehensive online surveillance mechanism against civilians whose whole sin was that they were forced to stay in isolation due to this pandemic. These included means for monitoring the location and activity of the user as well as various online means for checking their health condition, all while invading their privacy without allowing any choice (Ram and Gray, 2020; The International Digital Accountability Council's (IDAC), 2020; Brown and Toh, 2021; Dwork *et al.*, 2021).

Tackling violation of online privacy — A host of these threats from private companies and government agencies around the world have led various actors and organizations to act in a variety of programs and means to strengthen online privacy and awareness of its need. It is legal, against governments that violate the privacy of their citizens as well as led human rights organizations, to take various measures, including legal ones, against governments that violate the privacy of their citizens (Doffman, 2019; Amnesty, 2020; Beens, 2020; Clark, 2021). This is through a variety of educational initiatives including for parents how to teach their children about online privacy (IAPP, n.d.; Teaching Privacy, n.d.), setting 28 January as Data Privacy Day (National Cybersecurity Alliance, n.d.), as well as a variety of activities and initiatives of various international organizations (OHCHR, n.d.; Privacy International, n.d.), headed by the General Data Protection Regulation (GDPR) of the European Union (GDPR.eu, n.d.).

These threats to the privacy and physical information of every citizen today make the protection of information and privacy of paramount importance and even critical and undoubtedly pose a challenge to organizations and governments around the world and require their involvement to, on the one hand, maintain a balance between protecting the security of

the country and its citizens and, on the other hand, maintain the privacy of the citizens. This is done, among other things, through appropriate legislation and regulation and explicit definitions of privacy, in national policy documents, and the need to protect it in the digital sphere, including the call to see privacy as one of the human rights today (Milberg *et al.*, 2000; Tsoukalas and Siozos, 2011; K. N. C., 2019; Shwartz Altshuler, 2019; Stansberry *et al*., 2019; McKeever, 2021).

These characteristics of the information age, its value, its distribution, and the dangers lurking by international information giants, as well as governments, during routine times and in times of international crisis, need to examine the extent and level of privacy in national cyber policy documents of leading cyber countries. This is to learn to what extent these countries are not only aware of the need to maintain the online privacy of their citizens but also have an actual commitment to this issue in all that is stated in their national cyber policy documents.

Methodology

To examine the extent and manner of addressing the issue of privacy in official national cyber policy documents, several steps were made for RQ1 and RQ2:

1. Examining a wide range of government policy documents in the field of national cyber strategy, to analyze the extent and type of reference in each of these documents to the issue of privacy in cyberspace. To do this, the following steps were performed:
   1.1. Map various websites that maintain metrics that rank the world's leading countries in cyberspace, as well as in the field of exposure to cyber risks and cybercrime. Among the various indices, the study selected only those that are independent indices and are not based on other indices that have already been reviewed in this study.
   1.2. Select the top 20 countries from each of the lists.
   1.3. Create a list of all the leading countries in the world in cyberspace based on the various sites and indices (Step 1.2).
   1.4. Count the number of occurrences of each of the countries in the list of countries in each of these indices (Step 1.3).
   1.5. Create a list of the number of countries according to the number of instances (Step 1.4).

1.6. Create a unified list of the 20 leading cybersecurity leading countries with the largest number of instances (Step 1.5).
1.7. For each of the selected countries (Step 1.6), analyze the existence of a reference to the issue of privacy in cyberspace, as follows:
    1.7.1. Locating the latest government policy documents in the field of national cyber strategy for each country according to UNIDIR, Cyber Policy Portal (https://unidir.org/cpp/en), and based on the report of the Center for Strategic & International Studies (CSIS), Global Cyber Strategies Index (https://www.csis.org/programs/strategic-technologies-program/cybersecurity-and-governance/global-cyber-strategies-index), and on the website the European Union Agency for Cybersecurity (ENISA) (https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map). That is, the analysis of the level of reference to the issue of privacy in cyberspace is based on policy documents that deal with cyberspace only and not those related to, for example, information security policy, military, or digital policy.
    1.7.2. A search for the term "Privacy" in these documents, in the context of each country's view of the issue of privacy in cyberspace, its attitude and commitment, and not in general references to privacy.
    1.7.3. Documentation for each country from the list of selected countries (Step 1.6) whether it has a reference to the issue of privacy in cyberspace and, if so, how the privacy is defined by each country in these documents.

For question RQ3, Step 2 was performed:

2. Based on the process carried out in Step 1.7.1, an analysis will be made as to whether these documents refer to the COVID-19 era and its implications for the privacy of each country's cyberspace. This is done by searching for the terms "Coronavirus," "COVID-19," and "Pandemic" in these documents.

Findings

As part of the mapping of various websites that maintain indices that rank cybersecurity world leaders (Step 1.1) and after cross-referencing and

Table 1. List of websites and indices that rank countries in cybersecurity.

Name	Source	Category	Number of countries	Year
National Cybersecurity Index (NCSI)	National Cybersecurity Index (n.d.)	Government	160	2021
ITU — Global Security Index (GCI) 2020	The International Telecommunication Union (ITU) (2021)	International	194	2020
Belfer Center for Science and International Affairs — National Cyber Power Index 2020	Voo *et al.* (2020)	Research	30	2020
International Institute for Strategic Studies (IISS)	International Institute for Strategic Studies (IISS) (2021)	Research	15	2021
Comparitech — Which countries have the worst (and best) cybersecurity?	Bischoff (2021)	Information Security	75	2021
NordVPN — Cyber Risk Index 2020	NordVPN (2020)	Information Security	50	2020
Password Managers — Cybersecurity Exposure Index (CEI) 2020	Frisby (2020)	Information Security	108	2020

filtering various sources, the following seven sites and indices were identified:

After mapping the seven different websites and indices that rank the cybersecurity of world's leading countries, a list was received with 53 different countries that constitute the total existing based on the websites examined (Step 1.2).

With the compilation of this comprehensive list of cybersecurity of world's leading countries, according to the various indices, we will make a summary of the number of countries according to the number of occurrences at each of the websites and indices (Step 1.5).

Table 2. Mapping the cybersecurity of world's leading countries.

		Indices						
		Gov.	International	Research		Information security		
Country	No. of appearances (out of 7)	EGA	ITU	Belfer center	IISS	Comparitech	NordVPN	Password managers
France	5	9	9	6	5	–	–	15
United States	6	16	1	1	1	–	5	7
Japan	4	–	7	9	7	–	–	6
United Kingdom	7	18	2	3	2	8	10	13
Singapore	6	15	4	18	–	12	6	16
Russia	3	–	5	4	9	–	–	–
Israel	4	–	–	11	6	15	11	–
China	2	–	–	2	8	–	–	–
Spain	4	5	4	12	–	–	–	14
Estonia	4	3	3	14	–	–	–	5
Denmark	4	12	–	–	–	1	9	2
Sweden	4	–	–	13	–	2	2	14
Ireland	3	–	–	–	–	3	7	18
Norway	4	–	17	–	–	4	4	5
Finland	4	8		–	–	5	12	1
Netherlands	6	13	16	5	–	6	17	19

(*Continued*)

Table 2. (*Continued*)

		Indices						
		Gov.	International	Research		Information security		
Country	No. of appearances (out of 7)	EGA	ITU	Belfer center	IISS	Comparitech	NordVPN	Password managers
Austria	2	–	–	–	–	7	–	8
Switzerland	5	20	–	17	–	9	19	9
Croatia	4	11	–	–	3	10	–	18
Haiti	1	–	–	–	–	11	–	
Canada	5	–	8	8	–	13	14	13
Slovakia	3	10	–	–	–	14	–	19
Ukraine	1	–	–	–	–	16	–	–
Poland	2	6	–	–	–	17	–	–
Georgia	1		–	–	–	18	–	–
Belgium	5	7	19	–	–	19	13	11
Turkey	2	–	11	–	–	20	–	–
Germany	4	14	13	7	–	–	–	17
Australia	5	–	12	10	4	–	16	4
Greece	1	1	–	–	–	–	–	–
	2	2	–	–	–	–	–	–
	2	4	6	–	–	–	–	–

Serbia	1	17	–	–	–	–	–	–
Italy	3	19	20	–	–	–	–	20
Malaysia	3	–	5	19	14	–	–	–
Luxembourg	2	–	13	–	–	–	–	3
Saudi Arabia	1	–	2	–	–	–	–	
Mauritius	2	–	17	–	–	–	–	12
Republic of Korea	4	–	4	16	–	–	20	20
Qatar	1	–		–	–	–	–	17
Iceland	1	–		–	–	–	1	–
United Arab Emirates	2	–	5	–	–	–	3	–
New Zealand	3	–	–	15	–	–	8	10
Chile	1	–	–	–	–	–	15	–
Argentina	1	–	–	–	–	–	18	–
India	2	–	10	–	12	–	–	–
Portugal	1	–	14	–		–	–	–
Latvia	1	–	15	–		–	–	–
Brazil	1	–	18	–		–	–	–
Iran	1	–	–	–	10	–	–	–
North Korea	1	–	–	–	11	–	–	–
Indonesia	1	–	–	–	13	–	–	–
Vietnam	2	–	–	20	15	–	–	–

Table 3. Mapping the number of countries according to the number of occurrences.

Number of appearances (out of 7)	Number of countries
7	1
6	3
5	5
4	11
3	6
2	11
1	16
Total	**53**

Table 4. Ranking of the 20 leading countries in cybersecurity according to the number of occurrences.

Rank	Country	Continent	Number of appearances (out of 7)
1	United Kingdom	Europe	7
2	Netherlands	Europe	6
2	Singapore	Asia	6
2	United States	North America	6
3	Australia	Oceania	5
3	Belgium	Europe	5
3	Canada	North America	5
3	France	Europe	5
3	Switzerland	Europe	5
4	Croatia	Europe	4
4	Denmark	Europe	4
4	Estonia	Europe	4
4	Finland	Europe	4
4	Germany	Europe	4
4	Israel	Asia	4
4	Japan	Asia	4
4	Norway	Europe	4
4	South Korea	Asia	4
4	Spain	Europe	4
4	Sweden	Europe	4

Table 5. Reference to cyberspace privacy as part of national cyber policy documents.

Country	Publication	Source	Definition
United Kingdom	National Cybersecurity Strategy (2016–2021)	HM Government (2017)	"Yet cyberattacks are growing more frequent, sophisticated and damaging when they succeed. So we are taking decisive action to protect both our economy and the privacy of UK citizens." "We will preserve and protect UK citizens' privacy."
Netherlands	National Cybersecurity Agenda	Ministry of Justice and Security (2018)	"Legislation aimed at protecting national security will be reviewed to what extent it provides satisfactory possibilities to promote security in the digital domain, while retaining fundamental values and privacy."
Singapore	Singapore's Cybersecurity Strategy	Government of Singapore (2016)	–
United States	National Cyber Strategy of the United States of America	The White House (2018)	"The Administration's approach to cyberspace is anchored by enduring American values, such as the belief in the power of individual liberty, free expression, free markets, and privacy." "PROTECT AND PROMOTE INTERNET FREEDOM: The United States Government conceptualizes Internet freedom as the online exercise of human rights and fundamental freedoms — such as the freedoms of expression, association, peaceful assembly, religion or belief, and privacy rights online — regardless of frontiers or medium."
Australia	Australia's Cybersecurity Strategy (2020)	Australian Government (2020)	"How do I stay secure online? Privacy — Be wary of what is shared and with whom." "This consultation will consider multiple reform options, including the role of privacy and consumer protection laws, and duties for company directors."

(*Continued*)

Table 5. (*Continued*)

Country	Publication	Source	Definition
Belgium	Belgian National Cybersecurity Strategy (2.0 2021–2025)	Centre for Cybersecurity Belgium (2021)	"The General Data Protection Regulation and "Privacy," in general, are not part of cybersecurity as such, but they are obviously a big issue in terms of the CCB's mission to detect incidents and threats. Good cooperation with the Belgian Data Protection Authority is, therefore, necessary. Similarly, while fighting online disinformation campaigns is not actually a part of cybersecurity, it is connected to it. Cooperation with the competent intelligence and security services is also indispensable in this context."
Canada	National Cybersecurity Strategy	Government of Canada (2018)	"The Government of Canada will maintain and improve cybersecurity across all federal departments and agencies to protect the privacy of Canadians' information held by the federal government and the confidentiality, integrity, and availability of critical services for Canadians."
France	The French national digital security strategy	Prime Minister (2015)	"Protecting the digital lives, privacy, and personal data of the French people. With the prospect of the European Regulation on electronic identification (Electronic Identification and Trust Services — eIDAS), France will equip itself with a clear road map for digital identity delivered by the State." "France will protect its citizens' privacy and personal data. The right to privacy and individual and collective control of personal data will be reaffirmed whenever necessary and notably during commercial negotiations between States, whether bilateral or multilateral."

Switzerland	National strategy for the protection of Switzerland against cyber risks (NCS) (2018–2022)	Federal IT Steering Unit (2018)	–
Croatia	The National Cybersecurity Strategy of the Republic of Croatia	National Security Council (2015)	"Application of basic principles as the basis of the organization of modern society in the area of cyberspace as the society's virtual dimension: 1. Application of law to protect human rights and liberties, especially privacy, ownership, and all other essential characteristics of an organized contemporary society."
Denmark	Danish Cyber and Information Security Strategy (2018–2021)	Ministry of Finance (2018)	"Citizens are used to interacting with businesses and public authorities via digital solutions and have a basic trust that exchange of data and information takes place in a responsible and secure manner which respects the privacy of the individual."
Estonia	Cybersecurity Strategy (2019–2022)	Ministry of Economic Affairs and Communications (2019)	"The development of new services and databases will follow the principles of security and privacy by design. We will relinquish outdated platforms ("no legacy" principle). For this, we will develop a central security architecture advisory capability."
Finland	Finland´s Cybersecurity Strategy	The Security Committee (2019)	–

(*Continued*)

Table 5. (*Continued*)

Country	Publication	Source	Definition
Germany	Cybersecurity Strategy for Germany	Federal Ministry of the Interior (2016)	"Secure and trustworthy electronic communication that cannot be manipulated is fundamental for individuals to be able to enjoy the rights related to communication, the right to privacy, and the right to control one's own image."
Israel	Government Resolution 2444	The Government Secretary (2015)	–
Japan	Cybersecurity Strategy	The Government of Japan (2018)	"Assurance of the Free Flow of Information. For the sustainable development of cyberspace as a place for creation and innovation, it is imperative to build and maintain a world in which transmitted information reaches the intended recipient without being unfairly censored or illegally modified en route. Consideration for privacy must also be maintained. As a basic condition for the free flow of information in cyberspace, morality and commonsense are requested not to offend rights and interests of others."
Norway	National Cybersecurity Strategy for Norway	Norwegian Ministries (2019)	"Successful digitalization also includes making sure that the solutions provided appropriately accommodate demands for the security and privacy of the individual, and that everyone can be confident that the digital services will function as they should."
South Korea	National Cybersecurity Strategy	National Security Office (2019)	"Balance individual rights with cybersecurity: strike a balance between protecting cyberspace and safeguarding the fundamental rights of the people, e.g., privacy." "Devise legal measures to guarantee confidentiality and to prevent non-purpose use, such as privacy infringement when information is shared."

Spain	National Cybersecurity Strategy	President of the Government of Spain (2019)	"Promote cybersecurity to guarantee privacy and protection for personal data within the framework of citizen's digital rights in accordance with the legal system, promoting "digital identity" protection."
Sweden	A national cybersecurity strategy	Ministry of Justice (2017)	"The Government will work to — adapt the legislation to allow it to counteract cybercrime effectively — provide the law enforcement authorities' conditions, with reference to the protection of personal privacy and legal certainty, to maintain their capability to obtain information." "Efforts to safeguard society's cybersecurity need to be conducted in a long-term and effective manner and serve the interests of fundamental societal values, such as the protection of personal privacy." "Privacy and security in the cyber area are a prerequisite for enabling individuals to exercise their rights and freedoms and to make use of the possibilities of information technology."

To create a unified list of the 20 leading countries in cybersecurity, the countries with the largest number of occurrences — 7, 6, 5, 4 occurrences — were selected from all 7 different indices. In this way, 20 countries with the most occurrences were selected (Phase 1.6). Table 4 will present the 20 selected countries and their rank according to the findings in Table 2.

At this stage, the study examines the existence of national strategy documents on cyber policy (Phase 1.7.1) and the level of reference to privacy in cyberspace in these documents (Phases 1.7.2–1.7.3), as detailed in Table 5.

Table 6. How cyberspace is addressed as part of national cyber policy documents.

Country	Type of indication to privacy
United Kingdom	Specific
Netherlands	Specific
Singapore	–
United States	Specific
Australia	Specific
Belgium	Specific
Canada	Specific
France	Specific
Switzerland	–
Croatia	Specific
Denmark	Specific
Estonia	Specific
Finland	–
Germany	Specific
Israel	–
Japan	Specific
Norway	Specific
Republic of Korea	Specific
Spain	Specific
Sweden	Specific

To summarize the findings of Table 5 and understand how privacy is treated in cyberspace as part of national cyber policy documents, Table 6 presents the findings in three different ways: (1) addressing the issue of privacy is well defined, with direct reference to the need to maintain users in cyberspace and without compromising it while protecting local cyberspace and preventing attacks; (2) a general reference to privacy in cyberspace; (3) lack of any reference to the issue of privacy in cyberspace. This is for each of the following 20 countries:

Table 7 lists the number of occurrences for each of the three options for a country's type of privacy approach in national cyber policy documents.

To analyze the extent of the COVID-19 era's impact on privacy in cyberspace around the world, Step 2 in the methodology examines the extent to which the COVID-19 era is addressed in these documents. However, 18 of the 20 documents were published before the pandemic spread (2013–2019) and only two of them were published during it (2020–2021), as detailed in Table 8.

Table 7. Summary of how cyber privacy is addressed as part of the national cyber policy documents.

Type of indication to privacy	Number of occurrences
Specific Indication	16
General Indication	0
No Indication	4

Table 8. Reference to the COVID-19 pandemic in national cyber policy documents.

Country	Source	COVID-19/Coronavirus
Belgium	Belgian National Cybersecurity Strategy 2.0 (2021–2025)	–
Australia	Australia's Cybersecurity Strategy (2020)	"The COVID-19 pandemic highlighted the evolving nature of cyber threats." "With more Australians online as a result of COVID-19, the Australian Government has invested an additional $10 million to boost eSafety's investigations and support teams so help is available to Australians when they encounter harmful content and behaviors online."

Discussion

This study is intended to analyze the extent to which the issue of privacy is addressed in the national cyber policy documents of those that are considered the top 20 countries in cyberspace based on seven different indices.

The first research question examined the level of attitudes of cybersecurity world leaders to privacy in cyberspace. To this end, Table 2 ("Mapping the cybersecurity of world's leading countries") compiled seven different indices that created a list of 53 countries (Table 3 — "Mapping the number of countries according to the number of occurrences"), a unified list was compiled with the 20 countries that had the highest number of occurrences in the seven different lists in Table 5 ("Reference to cyberspace privacy as part of national cyber policy documents") while ranking them as in Table 4 ("Ranking of the 20 leading countries in cybersecurity according to the number of occurrences").

The findings of Table 5 reveal that of the 20 countries considered to be cybersecurity world leaders, in the national cyber policy documents of four of them (Singapore, Switzerland, Finland, and Israel), 20% of all countries in this study, there is no reference to privacy in cyberspace.

The second research question examined the existence of differences in the way cyber privacy is viewed by different countries. To this end, Table 6 ("How cyberspace is addressed as part of national cyber policy documents") analyzed whether the reference to privacy issues in these policy documents is specific or general.

These findings are summarized in Table 7 ("Summary of how cyber privacy is addressed as part of national cyber policy documents") and indicate that in all 16 countries where privacy is addressed in national cyber policy documents, this reference is specific, i.e., directly related to the need to maintain and ensure the privacy of their citizens in cyberspace. So, even if the wording differs between the different countries, all of them emphasize the need to maintain the balance between protecting cyberspace and combating various hostile elements in it, and the need to maintain the privacy of users of cyberspace.

The third research question addressed the degree of reference to the COVID-19 issue and the impact of this era on privacy in cyberspace. From Table 5, it can be learned that out of the 20 countries surveyed, the policy documents of 18 of them (90%) were written before 2020, which

is the beginning of the COVID-19 era. From the other two countries, it can be learned from Table 8 ("Reference to the COVID-19 pandemic in national cyber policy documents") that Belgium's national cyber policy document has no reference to the COVID-19 pandemic, whereas in Australia, there is an extensive reference to the issue, including to protect Australian Internet users from "harmful content and behaviors online."

These findings point to several trends emerging from the mapping of the national cyber policy documents of the 20 leading cyber countries in the world:

1. **Lack of uniformity in the definition of the cyber world leaders** — Table 2 shows the great difference that exists between the various indices for defining the cyber leadership of countries. This table demonstrates that a list of 53 different countries from these seven indices is needed, to reach a unified list of the 20 countries with the most occurrences. A tangible expression of this can be found in Table 3 which has a wide range between one country mentioned in all seven indices (United Kingdom) and 16 countries mentioned only once in one index or another. This variability is the product of a variety of reasons:
   - *The issue examined* — The indices examine cyber leadership of the countries in various aspects, including exposure to cybersecurity risks.
   - *The collection* — These indices are the product of various and varied entities: government bodies, international organizations, research bodies, and companies in the field of information technology.
   - *The number of countries included in each index* — From Table 1, one can learn about the differences in the number of countries included in the various indices, from 15 countries in the IISS index to 194 countries in the ITU index.
   - *The relevance of the indices* — All seven indices were published in 2020 and 2021 so their level of up-to-dateness is high.
2. **Geographical cohesion of the leading countries in cyberspace** — Table 4 indicates, among other things, the geographical cohesion of the top 20 cyber leading countries in the world: 13 of them (65%) from Europe, 4 from Asia (20%), 2 from North America (10%), and 1 from Oceania (5%). This means that

- this list does not include any countries from South America and Africa
- the prominence of countries from Europe constitutes 65% of the top 20 countries, cyber world leaders.

This trend also deepens when examining in depth the degree of representation of these countries among all countries on the same continent, as shown in Table 9.

This means that not only are 65% of all top cyber countries from Europe but also an in-depth examination at the level of all countries on each continent reveals that these 13 countries make up 25.5% of all countries on the continent, a statistic that overemphasizes European dominance among the world's leading cyber countries.

3. **Cohesion in the body that publishes the national cyber policy documents** — Due to the importance of the issue, these documents are usually published by the government or one of its ministries. Table 10 lists the various entities behind the publication of national cyber policy documents among the 20 selected countries. Examination of the findings reveals that in all cases it was a national, official body whether the government (45%), one of its offices (30%), or a national cybersecurity agency (25%).

   This means that in all cases national cyber policy documents are published by government bodies and most often by governments themselves. This indicates the great importance that countries attach to such national policy documents as well as to cyberspace.

4. **The variance in the year of publication of the cyber national policy documents** — Figure 1 shows the number of cyber policy

Table 9. The rate of the cyber leading countries among each continent's countries.

Continent	Number of countries in continent	Number of top cyber countries in continent	Percentage of top cyber countries from number of countries in continent
Africa	52	0	0%
Asia	4	50	8%
Europe	13	51	25.5%
North America	2	23	8.7%
Oceania	1	14	7.1%
South America	0	12	0%

Table 10. Mapping the factors that publish the national cyber policy documents.

The publisher of the National Cybersecurity Strategy	Number
Government	9
Government Ministry	6
Security Agency	5

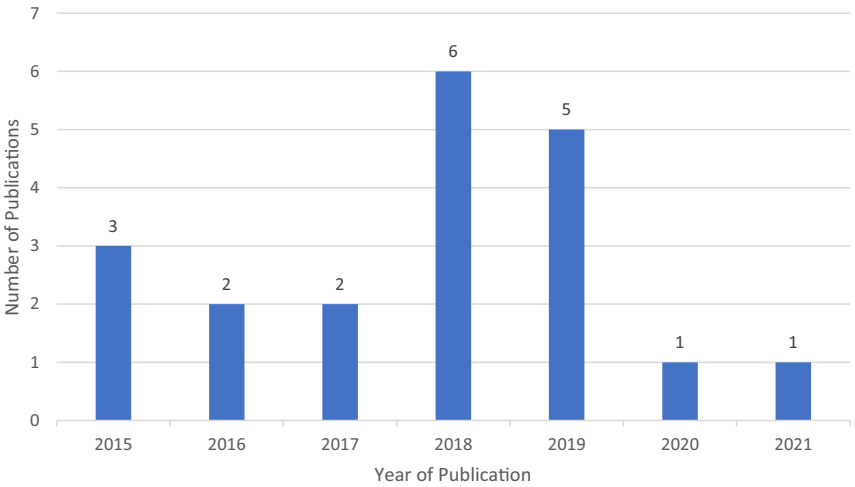


Figure 1. Number of cyber policy documents published by the top 20 cyber countries during 2015–2021.

documents published each year by the top 20 countries between the years 2015 and 2021.

This Figure 1 presents some interesting findings relating to the years in which the top 20 cyber countries published their national cyber policy documents:

- More than half of the cyber policy documents were published during 2018–2019 (11 out of 20, 55%).
- However, there are three documents (15%) that were published in 2015, therefore they have a low level of up-to-dateness and relevance.

- As noted above, two documents (10%) were published in the years 2020–2021 and are, therefore, highly up-to-date and relevant. One of them also includes references to the consequences of the COVID-19 pandemic, including in terms of online privacy.

5. **The importance of privacy when defining a national cyber policy** — Not only are 16 of the 20 (80%) leading cyber countries referring to privacy and the importance of maintaining it in cyberspace, but among all those countries, the reference is specific and explicit, in one form or another. The state needs to maintain the privacy of users in cyberspace. This means that measures will ensure the protection of the state, its institutions, and economy, as well as the privacy of its Internet users, and the existence of a free Internet. This is done, among other things, in cooperation with a variety of local, regional, and international bodies, improving information security among the various government bodies, as well as developing electronic means of communication and services on which the principles of privacy will be based. All this is to maintain a balance between the protection of individual and private rights and the protection of cyberspace, and the various assets contained therein.
6. **Reference to the COVID-19 era** — Only one of the 20 documents includes a reference to the COVID-19 era and its various effects including in cyberspace, the reference to the privacy in this document, in the context of the pandemic, in general.

Conclusion

Privacy in cyberspace is constantly changing, mainly due to a wide and evolving range of threats including information collected through tracking and monitoring online activity by various information giants and technology companies as well as a variety of government agencies; tracking and monitoring the physical location of citizens in the COVID-19 era; cyber companies providing cyber espionage; the activities of a variety of different players, including hackers, hacktivists, cybercriminals, and even a variety of state actors for leaking and stealing the information collected. At the same time, different countries are working to regulate online privacy and to protect, in as many ways as possible, the online privacy of their citizens, along with the fight against online threats. That is, to maintain a balance between protecting the national cyberspace and protecting users'

privacy while developing a variety of initiatives to raise their awareness of the need for responsible online behavior and one that maintains their privacy in cyberspace. The large number of cyber incidents around the world and the consequences of harming the information and privacy of users, as a result, require the determined and coordinated action of the cyber world's leading countries. In this context, these countries should work for a uniform approach and definition of privacy and the way to maintain it in their national cyber policy documents, update these policy documents to suit the current era and the challenges inherent in it, and include the COVID-19 era and its implications for citizens' privacy (physical and online).

At the same time, coordinated international activity is needed to promote an agreed international uniform index for ranking the cyber countries, based on a wide range of parameters and in diverse fields. There is also a need to promote the activities of countries from different continents to be able to integrate more successfully into cyberspace, including empowering those from Africa and South America in cyberspace. In addition, there is a need to update the policy documents of many countries to suit the current era and the frequently changing challenges, which, even if not directly related to cyberspace, directly affect it. All of these will lead not only to a wider range of state actors in cyberspace but also to better coping with the opportunities and challenges that exist in cyberspace including the privacy of users and keeping it on guard. This, along with initiatives taken to encourage raising awareness of online privacy and the need to safeguard it, for government officials, companies, and organizations and, of course, for the users themselves.

Limitation and Future Research

This study examined the top 20 countries in cyberspace, with 65% of them from Europe, thus completely lacking an analysis of the state of privacy reference in the cyber policy documents of countries from Africa and South America. The study also relies on the cyber policy documents published by the countries, some of which were published more than five years ago and do not necessarily reflect a picture of online privacy, the challenges, and the mitigation of them in these documents. In addition, it is not possible to learn from these documents how the COVID-19 era affected in general and cyberspace and, more importantly, its online privacy. Thus, this study provides an overview of the current state of cyber

policy documents among those considered to be cyber state leaders based on various metrics as well as the privacy issue in these policy documents.

Further research will be able to deepen and analyze the privacy in cyberspace in various aspects, including **the aspect of time** — conducting research in about a year and examining the degree of change in the identity of cyber world's leading countries, the level of updating of the national cyber policy documents of those countries, the expression they have of the cyber challenges, and the preservation of the privacy of citizens in general and in cyberspace. **The geographical aspect** — not only an examination of the cyber world's leading countries but also an analysis of each continent individually and the formation of a list of cyber leading countries in each continent. This look, which examines each continent individually, will be able to fill the existing knowledge gap regarding the cyber state in regions such as Africa and South America, which due to the nature of the study and its findings were not analyzed, as well as answering the question of whether the issue of privacy is addressed in the national cyber documents of countries that are not top cyber leaders at the global level but at the local level of each continent individually. **The thematic aspect** — further research will be able to deepen the analysis of the level of reference to current, global challenges as well as those that are directly related to cyberspace. Further research could also try to create a system of definitions related to privacy in cyberspace, on its various aspects, which could serve as a basis for different countries to address the issue of privacy in their national cyber policy documents.

References

Acquisti, A., John, L. K. and Loewenstein, G. (2013). What is privacy worth? *The Journal of Legal Studies*, 42(2), 1–27. DOI: 10.1086/671754.

AFP (2021). How a human error took down servers of Europe's major cloud computing group. https://www.ndtv.com/world-news/how-a-human-error-took-down-servers-of-europes-major-cloud-computing-group-2574397. Accessed on 17 October 2021.

Aldawood, H. and Skinner, G. (2019). Educating and raising awareness on cyber security social engineering: A literature review. In *Proceedings of 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2018*, pp. 62–68. DOI: 10.1109/TALE.2018.8615162.

Amnesty (2020). Why we're taking the UK government to court over mass spying. https://www.amnesty.org.uk/why-taking-government-court-mass-spying-gchq-nsa-tempora-prism-edward-snowden. Accessed on 19 October 2021.

Asn (2015). Some statistics about onions. *Tor Blog*. https://blog.torproject.org/some-statistics-about-onions. Accessed on 15 May 2021.

Australian Government (2020). Australia's cyber security strategy 2020. www.homeaffairs.gov.au/cyber-security-subsite/files/cyber-security-strategy-2020.pdf. Accessed on 11 October 2021.

Bagwe, M. (10AD). Anonymous leaks data from Texas GOP, GovInfoSecurity. https://www.govinfosecurity.com/anonymous-leaks-data-from-texas-gop-a-17679. Accessed on 17 October 2021.

Bajak, F. (2021). Microsoft: Russia behind 58% of detected state-backed hacks, AP. https://apnews.com/article/technology-business-china-europe-united-states-e13548edf082992a735a0af1da39b6c8. Accessed on 17 October 2021.

Bartley, K. (n.d.). Big data statistics: How much data is there in the world? *Rivery*. https://rivery.io/blog/big-data-statistics-how-much-data-is-there-in-the-world/. Accessed on 16 October 2021.

Beens, R. E. G. (2020). The state of mass surveillance. *Forbes*. https://www.forbes.com/sites/forbestechcouncil/2020/09/25/the-state-of-mass-surveillance/amp/. Accessed on 19 October 2021.

Bellman, S. *et al*. (2010). International differences in information privacy concerns: A global survey of consumers. http://dx.doi.org/10.1080/01972240490507956, 20(5), 313–324. DOI: 10.1080/01972240490507956.

Beresford, A. R., Kübler, D. and Preibusch, S. (2012). Unwillingness to pay for privacy: A field experiment. *Economics Letters*, 117(1), 25–27. DOI: 10.1016/J.ECONLET.2012.04.077.

Beretas, C. P. (2020). Mini review how really secure is TOR and the privacy it offers? *Biomedical Journal of Scientific & Technical Research*, 30(3), 1–2. DOI: 10.26717/BJSTR.2020.30.004963.

Bergman, M. K. (2001). The deep web: Surfacing hidden value. *Journal of Electronic Publishing*. Michigan Publishing, University of Michigan Library. DOI: 10.3998/3336451.0007.104.

Bischoff, P. (2021). Which countries have the worst (and best) cybersecurity? *Comparitech*. https://www.comparitech.com/blog/vpn-privacy/cybersecurity-by-country/. Accessed on 29 July 2021.

Brewster, T. (2015). Hackers scan all Tor hidden services to find weaknesses in the "Dark Web." *Forbes*. https://www.forbes.com/sites/thomasbrewster/2015/06/01/dark-web-vulnerability-scan/?sh=3181a1606d23. Accessed on 15 May 2021.

Brøndmo, H.-P. (2004). Anonymity is dead. Long live pseudonymity. *ClickZ*. https://www.clickz.com/anonymity-is-dead-long-live-pseudonymity/68075/. Accessed on 16 October 2021.

Brown, D. and Toh, A. (2021). Technology is enabling surveillance, inequality during the pandemic. *Human Rights Watch*. https://www.hrw.org/news/2021/03/04/technology-enabling-surveillance-inequality-during-pandemic. Accessed on 19 October 2021.

Campagnano, M. (2015). Online privacy and anonymity are dead. Get over it! *LinkedIn*. https://www.linkedin.com/pulse/online-privacy-anonymity-dead-get-over-mattia/?articleId=6087659560401674240. Accessed on 16 October 2021.

Carapola, S. (2017). Deep web: The 99% of the internet you can't see. https://www.amazon.com/Deep-Web-Internet-Cant-Everybody-ebook/dp/B06ZZXCMCX. Accessed on 27 April 2021.

Centre for Cybersecurity Belgium (2021). Cybersecurity strategy Belgium 2.0 2021-2025. https://ccb.belgium.be/sites/default/files/CCB_Strategie%202.0_UK_WEB.pdf. Accessed on 29 July 2021.

Chikada, A. (2016). The deep web, darknets, bitcoin and brand protection. *Law Business Research*. https://www.lexology.com/library/detail.aspx?g=effd0b0f-ddfa-400a-8a5f-6d6fb3b16dc7. Accessed on 27 April 2021.

Christin, D., Büchner, C. and Leibecke, N. (2013). What's the value of your privacy? Exploring factors that influence privacy-sensitive contributions to participatory sensing applications. In *IEEE Workshop on Privacy and Anonymity fir the Digital Economy (PADE)*, pp. 1–6.

Clark, M. (2021). NSO's Pegasus spyware: Here's what we know. *The Verge*. https://www.theverge.com/22589942/nso-group-pegasus-project-amnesty-investigation-journalists-activists-targeted. Accessed on 19 October 2021.

Clayton, J. (2021). Frances Haugen: Facebook whistleblower reveals identity. *BBC News*. https://www.bbc.com/news/technology-58784615. Accessed on 17 October 2021.

Desjardins, J. (2017). What happens in an internet minute in 2017? *World Economic Forum*. https://www.weforum.org/agenda/2017/08/what-happens-in-an-internet-minute-in-2017. Accessed on 7 May 2021.

Doffman, Z. (2019). Your social media is (probably) being watched right now, says new surveillance report. *Forbes*. https://www.forbes.com/sites/zakdoffman/2019/11/06/new-government-spy-report-your-social-media-is-probably-being-watched-right-now/?sh=65eb554d4f99. Accessed on 19 October 2021.

DOMO (2020). Data never sleeps 8.0. https://www.visualcapitalist.com/every-minute-internet-2020/. Accessed on 23 March 2021.

Dredge, S. (2013). What is Tor? A beginner's guide to the privacy tool. *The Guardian*. https://www.theguardian.com/technology/2013/nov/05/tor-beginners-guide-nsa-browser. Accessed on 17 October 2021.

Dwork, C. *et al.* (2021). On privacy in the age of COVID-19. *Journal of Privacy and Confidentiality*, 10(2), 1–6. DOI: 10.29012/jpc.749.

Federal IT Steering Unit (2018). National strategy for the protection of Switzerland against cyber risks (NCS) 2018–2022. https://www.ncsc.admin.ch/ncsc/en/home/strategie/strategie-ncss-2018-2022.html. Accessed on 29 July 2021.

Federal Ministry of the Interior (2016). Cyber security strategy for Germany 2016. https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map/strategies/cyber-security-strategy-for-germany/@@download_version/5f3c65fe954c4d33ad6a9242cd5bb448/file_en. Accessed on 29 July 2021.

Frisby, J. (2020). Cybersecurity exposure index (CEI) 2020. *PasswordManagers.co*. https://passwordmanagers.co/cybersecurity-exposure-index/#global. Accessed on 29 July 2021.

GDPR.eu (n.d.). What is GDPR, the EU's new data protection law? https://gdpr.eu/what-is-gdpr/. Accessed on 19 October 2021.

Government of Canada (2018). National cyber security strategy Canada's vision for security and prosperity in the digital age. https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/ntnl-cbr-scrt-strtg/ntnl-cbr-scrt-strtg-en.pdf. Accessed on 29 July 2021.

Government of Singapore (2016). Singapore's cybersecurity strategy. *Government of Singapore*. https://www.csa.gov.sg/news/publications/singapore-cybersecurity-strategy. Accessed on 28 July 2021.

Hall, M. (2016). Why people are key to cyber-security. *Network Security*, 2016(6), 9–10. DOI: 10.1016/S1353-4858(16)30057-5.

Hill, M. and Swinhoe, D. (2021). The 15 biggest data breaches of the 21st century. *CSO Online*. https://www.csoonline.com/article/2130877/the-biggest-data-breaches-of-the-21st-century.html. Accessed on 17 October 2021.

HM Government (2017). National cyber security strategy 2016–2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/567242/national_cyber_security_strategy_2016.pdf.

Ho, S. M. and Gross, M. (2021). Consciousness of cyber defense: A collective activity system for developing organizational cyber awareness. *Computers & Security*, 108, 102357. DOI: 10.1016/J.COSE.2021.102357.

Hope, A. (2021). Threat actor leaks login credentials of about 500,000 Fortinet VPN accounts. *CPO Magazine*. https://www.cpomagazine.com/cyber-security/threat-actor-leaks-login-credentials-of-about-500000-fortinet-vpn-accounts/. Accessed on 17 October 2021.

IAPP (n.d.). Privacy activity sheets for kids. https://iapp.org/resources/article/privacy-activity-sheets-for-kids/. Accessed on 19 October 2021.

International Institute for Strategic Studies (IISS) (2021). Cyber capabilities and national power: A net assessment. https://www.iiss.org/blogs/research-paper/2021/06/cyber-power---tier-one. Accessed on 29 July 2021.

Jossen, S. (2017). The world's most valuable resource is no longer oil, but data, RGA. https://www.rga.com/ja/futurevision/pov/the-worlds-most-valuable-resource-is-no-longer-oil-but-data-4792050. Accessed on 19 October 2021.

Karr, D. (2012). Big data brings marketing big numbers. *Martech Zone*. https://martech.zone/ibm-big-data-marketing/. Accessed on 23 March 2021.

Kind, C. (2021). Annual digital lecture 2020: The death of anonymity in the age of identity. *The National Archives*. https://media.nationalarchives.gov.uk/index.php/annual-digital-lecture-2020-the-death-of-anonymity-in-the-age-of-identity/. Accessed on 16 October 2021.

K. N. C. (2019). Open future — Surveillance is a fact of life, so make privacy a human right. *Open Future | The Economist*. https://amp.economist.com/open-future/2019/12/13/surveillance-is-a-fact-of-life-so-make-privacy-a-human-right. Accessed on 19 October 2021.

Lee, N. (2015). Anonymity is dead and other lessons from the Silk Road trial. *Engadget*. https://www.engadget.com/2015-02-08-silk-road-trial-lessons.html. Accessed on 16 October 2021.

Lewis, L. (2020). Infographic: What happens in an internet minute 2020. *Merge*. https://www.allaccess.com/merge/archive/31294/infographic-what-happens-in-an-internet-minute. Accessed on 23 March 2021.

Long, W. J. and Quek, M. P. (2011). Personal data privacy protection in an age of globalization: The US-EU safe harbor compromise. *Journal of European Public Policy*, 9(3), 325–344. DOI: 10.1080/13501760210138778.

Lufkin, B. (2017). The reasons you can't be anonymous anymore. *BBC Future*, *BBC*. https://www.bbc.com/future/article/20170529-the-reasons-you-can-never-be-anonymous-again. Accessed on 16 October 2021.

Mance, H. (2019). Is privacy dead? *The Financial Times*. https://www.ft.com/content/c4288d72-a7d0-11e9-984c-fac8325aaa04. Accessed on 16 October 2021.

Mani, A. *et al.* (2018). Understanding Tor usage with privacy-preserving measurement.

Marr, B. (2018). How much data do we create every day? The mind-blowing stats everyone should read. *Forbes*. https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/?sh=3966e50a60ba. Accessed on 23 March 2021.

McKeever, G. (2021). Why data security and privacy in the digital age are crucial. *Impreva*. https://www.imperva.com/blog/why-data-security-and-privacy-in-the-digital-age-are-crucial/. Accessed on 19 October 2021.

Milberg, S. J., Smith, H. J. and Burke, S. J. (2000). Information privacy: Corporate management and national regulation. *Organization Science*, 11(1), 35–57. DOI: 10.1287/ORSC.11.1.35.12567.

Millman, R. (2021). Cyber criminals leak one million credit cards on the dark web. *IT PRO*. https://www.itpro.com/security/cyber-crime/360534/cyber-criminals-leak-one-million-credit-cards-on-the-dark-web. Accessed on 17 October 2021.

Mims, C. (2018). Privacy is dead. Here's what comes next. *WSJ*, *The Wall Street Journal*. https://www.wsj.com/articles/privacy-is-dead-heres-what-comes-next-1525608001. Accessed on 16 October 2021.

Ministry of Economic Affairs and Communications (2019). Cyber security strategy. https://www.mkm.ee/sites/default/files/kyberturvalisuse_strateegia_2022_eng.pdf. Accessed on 28 July 2021.

Ministry of Finance (2018). Danish cyber and information security strategy. https://en.digst.dk/media/17189/danish_cyber_and_information_security_strategy_pdf.pdf. Accessed on 28 July 2021.

Ministry of Justice (2017). A national cyber security strategy.

Ministry of Justice and Security (2018). National cybersecurity agenda. https://english.ncsc.nl/topics/national-cybersecurity-agenda/documents/publications/2019/juni/01/national-cyber-security-agenda. Accessed on 29 July 2021.

Mitchell, G. (n.d.). How much data is on the internet? *Science Focus*. https://www.sciencefocus.com/future-technology/how-much-data-is-on-the-internet/. Accessed on 16 October 2021.

Mulligan, D. and Berman, J. (1999). Privacy in the digital age. *Nova Law Review*, 23(2), 551–582. https://nsuworks.nova.edu/cgi/viewcontent.cgi?article=1418&context=nlr.

Muncaster, P. (2021). Customers on alert as E-commerce player leaks 1.7+ billion records. *Infosecurity Magazine*. https://www.infosecurity-magazine.com/news/ecommerce-player-leaks-billion/. Accessed on 17 October 2021.

National Cyber Security Index (n.d.). Ranking. https://ncsi.ega.ee/ncsi-index/?order=rank. Accessed on 29 July 2021.

National Cybersecurity Alliance (n.d.). Data privacy day. *Stay Safe Online*. https://staysafeonline.org/data-privacy-day/. Accessed on 19 October 2021.

National Security Council (2015). National cyber security strategy of the Republic of Croatia. https://www.uvns.hr/UserDocsImages/en/dokumenti/Croatian%20National%20Cyber%20Security%20Strategy%20(2015).pdf. Accessed on 29 July 2021.

National Security Office (2019). National cybersecurity strategy. https://ccdcoe.org/uploads/2018/10/South-Korea_English-National-Cybersecurity-Strategy-03-April-2019_English-1.pdf. Accessed on 29 July 2021.

NordVPN (2020). Cyber risk index 2020. https://s1.nordcdn.com/nord/misc/0.13.0/vpn/brand/NordVPN-cyber-risk-index-2020.pdf. Accessed on 29 July 2021.

Norwegian Ministeries (2019). National cyber security strategy for Norway.

OHCHR (n.d.). OHCHR and privacy in the digital age. https://www.ohchr.org/en/issues/digitalage/pages/digitalageindex.aspx. Accessed on 19 October 2021.

Onion Services — Tor Metrics (2021). Tor project. https://metrics.torproject.org/hidserv-dir-onions-seen.html?start=2021-01-01&end=2021-04-26. Accessed on 26 April 2021.

Osborn, E. (2015). Business versus technology: Sources of the perceived lack of cyber security in SMEs.

Pagliery, J. (2013). Online privacy is dead. *CNN Business*. https://money.cnn.com/2013/10/17/technology/online-privacy/index.html. Accessed on 16 October 2021.

Pagliery, J. (2014). The deep web you don't know about. *CNN*. https://money.cnn.com/2014/03/10/technology/deep-web/index.html. Accessed on 27 April 2021.

Pappas, S. (2016). How big is the internet, really? *LiveScience*. https://www.livescience.com/54094-how-big-is-the-internet.html. Accessed on 16 October 2021.

Petrov, C. (2021). 25+ big data statistics — How big it actually is in 2021? *Tech Jury*. https://techjury.net/blog/big-data-statistics/. Accessed on 23 March 2021.

Popkin, H. A. S. (2010). Privacy is dead on Facebook. Get over it. *NBC News*. https://www.nbcnews.com/id/wbna34825225. Accessed on 16 October 2021.

Pratham (2019). The deep web — 99% internet that you can't through Google. *Broggl*. https://www.broggl.com/the-deep-web-99-internet-that-you-cant-through-google/. Accessed on 23 March 2021.

President of the Government of Spain (2019). Spanish national cyber security strategy.https://www.dsn.gob.es/en/documento/estrategia-nacional-ciberseguridad-2019. Accessed on 28 July 2021.

Prime Minister (2015). French national digital security strategy. https://www.ssi.gouv.fr/uploads/2015/10/strategie_nationale_securite_numerique_en.pdf. Accessed on 28 July 2021.

Prince, J. and Wallsten, S. (2020). How much is privacy worth around the world and across platforms? Washington. https://techpolicyinstitute.org/wp-content/uploads/2020/02/Prince_Wallsten_How-Much-is-Privacy-Worth-Around-the-World-and-Across-Platforms.pdf.

Privacy International (n.d.). Privacy international. https://privacyinternational.org/. Accessed on 19 October 2021.

Ram, N. and Gray, D. (2020). Mass surveillance in the age of COVID-19. *Journal of Law and the Biosciences*, 7(1), 1–17. DOI: 10.1093/JLB/LSAA023.

Raywood, D. (2020). Has the rise of identity seen the death of anonymity? *InfoSecurity Magazine*. https://www.infosecurity-magazine.com/news-features/rise-identity-anonymity/. Accessed on 16 October 2021.

Sahota, N. (2020). Privacy is dead and most people really don't care. *Forbes*. https://www.forbes.com/sites/neilsahota/2020/10/14/privacy-is-dead-and-most-people-really-dont-care/?sh=4587e8997b73. Accessed on 16 October 2021.

Sauer, P. (2021). Privacy fears as Moscow metro rolls out facial recognition pay system. *The Guardian*. https://www.theguardian.com/world/2021/oct/15/privacy-fears-moscow-metro-rolls-out-facial-recognition-pay-system. Accessed on 19 October 2021.

Schultz, J. (2019). How much data is created on the internet each day? *Micro Focus Blog*. https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/. Accessed on 16 October 2021.

Shwartz Altshuler, T. (2019). Privacy in a digital world. *TechCrunch*. https://techcrunch.com/2019/09/26/privacy-queen-of-human-rights-in-a-digital-world/. Accessed on 19 October 2021.

Stansberry, K., Anderson, J. and Rainie, L. (2019). 5. Leading concerns about the future of digital life. *Pew Research Center*. https://www.pewresearch.org/internet/2019/10/28/5-leading-concerns-about-the-future-of-digital-life/. Accessed on 16 October 2021.

Statista (2021). Total data volume worldwide 2010–2025. *Statista*. https://www.statista.com/statistics/871513/worldwide-data-created/. Accessed on 16 October 2021.

Stone, J. (2019). How many dark web marketplaces actually exist? About 100. *CyberScoop*. https://www.cyberscoop.com/dark-web-marketplaces-research-recorded-future/. Accessed on 8 May 2021.

Szoldra, P. (2016). A timeline of Edward Snowden leaks. *Insider*. https://www.businessinsider.com/snowden-leaks-timeline-2016-9. Accessed on 17 October 2021.

Taiwo, I. (2015). 90% of the internet is hidden from your browser; and it's called the deep web. *TechCabal*. https://techcabal.com/2015/11/18/90-of-the-internet-is-hidden-from-your-browser-and-its-called-the-deep-web/. Accessed on 27 April 2021.

Taylor, S. (2019). Is Tor trustworthy and safe? (Read this before using Tor). *Restore Privacy*. https://restoreprivacy.com/tor/. Accessed on 17 October 2021.

Teaching Privacy (n.d.). Teaching privacy. https://teachingprivacy.org/. Accessed on 19 October 2021.

The Government of Japan (2018). Cybersecurity strategy.

The Government Secretary (2015). Government resolution no. 2444: Advancing the national preparedness for cyber security. Jerusalem. https://ccdcoe.org/uploads/2019/06/Government-Resolution-No-2444-Advancing-the-National-Preparedness-for-Cyber-Security.pdf. Accessed on 28 July 2021.

The International Digital Accountability Council's (IDAC) (2020). Privacy in the age of COVID: An IDAC investigation of COVID-19 apps. *The International*

Digital Accountability Council's (IDAC). https://digitalwatchdog.org/privacy-in-the-age-of-covid-an-idac-investigation-of-covid-19-apps/. Accessed on 19 October 2021.

The International Telecommunication Union (ITU) (2021). Global cybersecurity index 2020. Geneva. https://www.itu.int/epublications/publication/global-cybersecurity-index-2020/en/. Accessed on 29 July 2021.

The Security Committee (2019). Finland's cyber security strategy 2019. Available at: https://turvallisuuskomitea.fi/wp-content/uploads/2019/10/Kyberturvallisuusstrategia_A4_ENG_WEB_031019.pdf. Accessed on 29 July 2021.

The White House (2018). National cyber strategy of the United States of America. https://trumpwhitehouse.archives.gov/wp-content/uploads/}2018/09/National-Cyber-Strategy.pdf. Accessed on 28 July 2021.

Tor project | Anonymity Online (n.d.). https://www.torproject.org/. Accessed on 17 October 2021.

Tsoukalas, I. A. and Siozos, P. D. (2011). Privacy and anonymity in the information society — Challenges for the European Union. *The Scientific World Journal*, 11, 458–462. DOI: 10.1100/TSW.2011.46.

Voo, J. *et al.* (2020). National cyber power index 2020 methodology and analytical considerations. www.belfercenter.org/CCPI. Accessed on 29 July 2021.

Vuleta, B. (2021). How much data is created every day? [27 powerful stats]. *SeedScientific*. https://seedscientific.com/how-much-data-is-created-every-day/. Accessed on 16 October 2021.

Williams, S. (2021). IOTW: Anonymous hacker posts salaries of "Twitchers" to 4chan. *Cyber Security Hub*. https://www.cshub.com/attacks/articles/iotw-anonymous-hacker-posts-salaries-of-twitchers-to-4chan. Accessed on 17 October 2021.

Wyciślik-Wilson, S. (2019). How to protect your privacy online with Tor browser. *TechRadar*. https://www.techradar.com/how-to/how-to-protect-your-privacy-online-with-tor-browser-improve-your-security-and-stay-anonymous. Accessed on 17 October 2021.

Zibuschka, J. *et al.* (2019). Anonymization is dead-long live privacy. Bonn, pp. 1–12. https://www.consultancy.uk/news/13487/six-privacy-principles-for-general-data-protection-regulation. Accessed on 16 October 2021.

Zwilling, M. *et al.* (2020). Cyber security awareness, knowledge and behavior: A comparative study. https://doi.org/10.1080/08874417.2020.1712269 [Preprint].

https://doi.org/10.1142/9781800612860_0008

Chapter 7

Too Much Information: OSINT in Criminal Investigations and the Erosion of Privacy

Mikhail Reider-Gordon

Introduction

The availability of publicly accessible information via social media and the Internet has grown exponentially in the past decade. Known as Open-Source Intelligence (OSINT), personal and sometimes very private details about individuals and the lives they lead have become standard pubicly available resources sought by and accessible for use by law enforcement in criminal investigations and intelligence. Commensurate with this rise in available online information, concerns around violations of privacy and human rights have been raised. Despite calls over the years for stricter regulations to govern how OSINT is used by law enforcement, and commitments, at least on paper, by some enforcement agencies to abide by certain standards to help curb abuse of OSINT, recent reports and surveys from law enforcement agencies suggest that in pursuit of cybercrime, compliance with current governance frameworks has been minimal to non-existent. In the absence of clear governance frameworks, mechanisms for oversight, training of law enforcement, and specific legislation, individual's personal privacy is now fundamentally threatened by the growth in adoption of widespread OSINT collection by enforcement bodies. Argument is made that there is a balance to be struck between

allowing unsupervised collection of open-source information to meet law enforcement needs against legal safeguards designed to help ensure individuals' privacy rights are protected.

This chapter is structured across five sections. The first three sections consider the definition of OSINT relative to law enforcement's use of it; place the historic use of OSINT in criminal investigations and intelligence activities; and analyze current privacy legal regimes applicable to OSINT collection and processing by law enforcement. Section IV investigates recent surveys and audits of law enforcement agencies' understanding and compliance with relevant data privacy requirements relative to the collection and processing of OSINT. Lastly, in Section V, recommendations are presented for addressing transparency, oversight and accountability of OSINT collection and handling by law enforcement to better safeguard individuals' privacy and bring the use of OSINT in criminal investigations and intelligence gathering into compliance with legal privacy regimes.

OSINT and Criminal Investigations

In his 1979 book, *Disturbing the Universe*, Freeman Dyson wrote, *If we had a reliable way to label our toys good and bad, it would be easy to regulate technology wisely. But we can rarely see far enough ahead to know which road leads to damnation* (Dyson, 1979). The digitalization of human lives largely played out on the Internet has revolutionized how anyone with access online obtains, collects, analyzes and disseminates information. The growth of publicly available information useful for the purposes of intelligence gathering is the result of some of our newer technological "toys." If one considers access to municipal records, accounts of commerce and the similar, open-source intelligence (OSINT) has been around for thousands of years. OSINT gained traction within the intelligence community mid-20th century. For the better part of 80 years, OSINT was a niche subject of intelligence collection largely reserved to military and intelligence analysts. Criminal investigators who knew of it were confined to locating physical records in dusty archives. It is only in the past decade as the volume of open or accessible information online has grown exponentially with social media, blogs, low-cost satellites, the Internet of Things, the deep web, the dark web, websites, chat rooms, media, publications from institutions and academia, VOIP, teleconferencing, apps, the digital exhaust left by users across the web, and the

digitization of nearly all information[1] that if one either possesses the knowledge or has access to the right tools, nearly any piece of information can be found. Combined, these disparate scraps of information build comprehensive portraits of individuals, far more granular than what people's family or physician likely knows about them. One of the most concise descriptions of what information data can provide from online sources now was said by Google's CEO, Eric Schmidt, "We know where you are. We know where you've been. We can more or less know what you're thinking about" (Thompson, 2010). OSINT collected in and of itself isn't necessarily meaningful. A single tweet doesn't inevitably draw a supportable conclusion, but the raw data once aggregated, contextualized, and analyzed become both a boon and a danger. For law enforcement investigating a myriad of crimes, OSINT can offer clues and contribute important evidence. But as Casanovas has asserted, the placement of any piece of data from online into the category of OSINT is to move it into "a no man's land, free to be grabbed and manipulated for public reasons" (2017). While OSINT may have its uses for tracking cybercriminals, it has also allowed digital surveillance to track a significant amount of individual's daily lives, fundamentally threatening personal privacy.

Multiple definitions of OSINT have been offered (see Williams and Blum, 2018; National Police Chief's Council, 2016; and US Office of the Director of National Intelligence, undated). Open-Source Intelligence (sometimes referred to as Open-Source Information) in its purest definition means only that the information is in some way publically available — from "open" data sources, be it newspapers, social media platforms, blogs, publications, commercial data, even public government data. In other words, it is not information that is classified by a government in some way that restricts access (i.e., not controlled by national security strictures such as that typically reserved for members of a military, or government employees or contractors). OSINT includes everything from media reports, both broadcast and print, to satellite imagery, maps (digital and physical), grey media (e.g., open government publications, publications from academia, think tanks, and corporations), photos, videos, commercial subscription databases, and anything and everything on the Internet. However, for this chapter's purposes, OSINT refers to Internet-related open-source information gathering. Can OSINT results be

[1]The majority of the world's data have come about in only the past two years as indicated by data growth statistics. See https://www.internetlivestats.com.

classified later after a state uses the information? Yes, absolutely. Beyond being information that is publicly available, it is information that can be used in an intelligence context. One of the significant risks of OSINT collected, analyzed and disseminated by criminal investigators is the lack of clear bifurcation between law enforcement officers in national-level agencies and intelligence organizations. It is not uncommon for the two sides to assist one another or even second officers to each other. OSINT collection and processing for criminal investigation and intelligence is very different from that of state intelligence agencies' capture of OSINT for national security purposes. Despite the misnomer of "open," the acquisition of OSINT by government can serve as a form of targeted surveillance, raising issues of civil rights, and civil liberties.

Use of OSINT

The UNODC defines OSINT as "information that is publicly available" (UNODC, 2011) and suggests its best use is to provide credibility to other information that has greater reliability. That is, information to be collected in an overt manner to help comprehend a particular concern. However, as the Internet has grown and as the Internet of Things (IoT) has begun to feed vast troves of additional data into the greater free-flowing information, a substantial volume of data have become "public" or open beyond what its original owners may have intended. A user who posts a group photo on their Facebook page does so with the intent to show a limited audience a specific activity. They may not realize that a third party can not only copy the image and redistribute it in other fora or for other uses, but the other individuals depicted in the photo likely did not give their consent for the posting or for any subsequent uses of their images. They lack agency to object to a third-party applying facial recognition software to the photo, capturing and labeling their visage and subsequently storing it for yet others downstream to use for any number of other purposes. There is no mechanism for the original user or those depicted in the group photo that has now been put in a tiny corner of the Internet to grant or withhold permission as to how that digital information is subsequently used.

Similarly, sites that post content from data leaks or hacks transform hitherto confidential or classified information into "open" intelligence. This not only raises ethical and legal challenges, but the comingling of the myriad of sources and forms data made available online now takes, in a sense placing it all on the same level, has broadened OSINT to encompass

far more than what was included in definitions created only two or three decades ago.

Existing privacy controls on platforms can shift and change, leaving users' data exposed unbeknownst to them. Demonstrations of the ease in which users' identities on social media can be de-anonymized underscore the impossibility of true privacy online without legislative intervention (e.g., mandating privacy by design; data minimization; restricting types of data collected, etc.) (*see* Wondracek *et al*., 2010; Su *et al*., 2017). Rocher *et al.* established that using an algorithm across public data in the US identified every single American, with only a .02% error rate. Pastor-Galindo *et al.* (2020) have demonstrated that available tools now make it possible to search entire social media platforms to exquisite granularity, targeting by "exact phrases, hashtags," and facilitating queries that can run across multiple platforms in pursuit of the user's online movements. Individuals can now literally be hunted across the Internet. Tools such as *NameVine*, which applies machine learning to guessing individual's precise usernames, and third-party companies now hoover up every scrap of data, applying algorithms to the voluminous raw content and transforming it into dossiers on people around the world to be repackaged again and sold on, including to law enforcement. Subscription databases provide to law enforcement access denied many in the private sector, such as credit reports, bank account information, driver's licenses, vehicle registrations, and even the footage from street and traffic cameras that capture vehicle plates, geolocations, images of drivers, and can even analyze the frequency in which that person is observed in that precise location. Particularly prevalent in the US and UK are companies who traffic in bulk data, mobile and social media information, and biometric data pulled from the Internet. Most, but not all, sell exclusively to law enforcement. As the Royal Canadian Manted Police (RCMP) has stated, *Open-source information (OSI) is a key tool that is used organisation-wide, across all business lines and within every RCMP division. OSI activities are performed by all categories of employees for a variety of purposes* (RCMP, 2021). The agency does not limit itself to one-off queries of OSINT but collects "systematically" and includes passive collection of OSINT in its standard remit for criminal intelligence and investigations (RMCP, p. 5).

Ungureanu argues OSINT dates back at least to the 15th century, citing printed works used to source intelligence by warring nations. This author traces OSINT online to the early 1980s before the formal Internet. During that time, ARPANET was still in existence and pre-Web,

information sourcing required the author to construct Boolean logic queries of digital copies of information held at repositories also connected to the network. Now, Google indexes *35 trillion* web pages,[2] but that only scratches the surface, as the same studies tell us the Internet as a whole is home to more than 17.5 *quadrillion* different pages.[3]

The conflict between law enforcement's desire to access this ever-increasing volume of information on the one hand and individual privacy rights on the other has only intensified as greater volumes of information are constantly posted to the Internet, particularly to social media platforms. The ease and relative low cost of the technologies capable of scooping up what individuals say, memorialize of their doings, their personal histories, their affiliations with religious and civic organizations, their interests, sexual orientation, who they meet, who they interact with, where they go, how they get from point A to B, and even how they pay, has transformed the types of information law enforcement has access to "publicly."[4] Where in the 1990s, OSINT was estimated to account for nearly 80% of intelligence material (RAND), 25 years on, a survey by RUSI (2015) asserted that up to 95% of all intelligence gathered by intelligence agencies now derives from OSINT. The lure of accessibility and low cost is too great to resist. But, herein lays the danger to privacy. The public and private digital landscape is now a blurred one.

As Gradecki and Curry (2017) observed, if not themselves involved in the world of technology, most users fail to understand how volumes of data and metadata are sucked up by third parties and turned over to law enforcement as a "tool" for criminal investigators. Most individuals have no concept of the quantity of information collected and how it is transformed into intelligence, thus most citizens have not pressed legislators to enact stringent privacy controls over much data that are in the public domain, such as social media. Edwards and Urquhart (2015) offered a taxonomy of social media OSINT, "SOCMINT," that seeks to delineate data about specific individuals and that of more general information. The latter constitutes personal data that can be abused. In a survey conducted by LexisNexis (2012), out of 1,200 US law enforcement agencies, 960 were already using SOCMINT for investigations. Back in 2012,

[2]https://www.internetlivestats.com/google-search-statistics/.

[3]https://firstsiteguide.com/google-search-stats/.

[4]For case of identifying "private" information. *See* Cascavilla *et al*. (2018).

Eijkman and Weggeman were calling for greater accountability of law enforcement use of OSINT. Nearly a decade on, training, oversight and legislation with respect to criminal law enforcement and intelligence use of OSINT continues to lag behind the expanding collection and exploitation of open-source data, particularly the use of SOCMINT, which is increasingly at odds with privacy expectations and protections.

Sensitive personal information becoming known can lead to discriminatory impacts, unfair targeting, and violations of user's civil rights. OSINT is largely an unregulated data market that scrapes, links and sells information collected largely covertly through online activity. "Companies gather and trade-in people's health stats, GPS location, contact lists, political leanings, purchase histories, browsing histories and numerous other data points — largely without people's knowledge" (GIATOC). The potential damage the possession of these types of data can cause is extensive. In countries with low rule of law, personal histories and identities can be exploited by corrupt police, providing fodder for coercion and extortion. With the volume of information now available online, it is possible to uniquely identify individuals amidst mass data sets and streams, and equally make decisions about these people based on broad types of information. Privacy here includes protections around identity and integrity, and non-discrimination via the misuse of information obtained from OSINT. Perhaps, the most significant challenge to privacy is that the right can be compromised without the individual being aware their personal information is being collected and analyzed. Machine learning algorithms have advanced significantly, creating tools for the processing of massive data sets, including social media. Data ethics, privacy controls specific to law enforcement use of these tools, and training of officers lags far behind. Used for the perpetual collection and analysis of social media — every swipe, tweet, "like," emoji, post, share, upload, and repost — the algorithms employed in OSINT collection do not discern between "criminal" and innocent. What separates democracies from authoritarian systems is the issue of police and intelligence oversight and safeguards against abuses of such secrecy (Colaresi in Ünver, p. 16).

Privacy Regime and OSINT

OSINT challenges the data privacy regimes that exist at national levels as with so many other aspects of the cyber world, online content is, for the

most part, borderless. Restrictions on collection in one jurisdiction do not necessarily protect the same information from exploitation in another. As Frederick (2019, p. 2) has observed, in countries such as China, the "blending of the public and private digital landscape" can mean technology companies have little recourse to deny or block government access to the data they collect. How much OSINT is scraped directly or via third-party companies from western platforms by state actors with low rule of law is not well documented. Frederick (2019, p. 3) uses China's export of AI-driven data scraping and analyzing tools as an example of how states may be exploiting OSINT that not only may pose a security threat to western countries, but also as a means of exporting legal frameworks that undermine the human right to privacy.[5] This may not, on its surface, appear to be a problem for criminal investigators, but as recent examples such as Bellingcat's ability to identify US military bases from an online app for beer drinkers evidences (Bellingcat, 2020), there is a dangerous side to OSINT that underscores why many countries should be looking to strengthen their data protection laws as part of their cybersecurity strategy.

As Drewer and Miladinova (2017) have discussed, the rights to privacy and data protection have been elevated to the level of human rights internationally in a number of hard and soft law instruments.[6] The EU has

[5]See United Nations Declaration of Human Rights (UDHR) G.A. Res. 217 (III) A (10 December 1948), Article 12: "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks;" International Covenant on Civil and Political Rights (ICCPR) 1966, Article 17: "1. No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honour or reputation. 2. Everyone has the right to the protection of the law against such interference or attacks;" Organisation for Economic Cooperation and Development (OECD), *Guidelines Governing the Protection of Privacy and Transborder Flow of Personal Data*, 23 September 1980. Additionally, over 130 countries have constitutional statements regarding the protection of privacy.

[6]European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR), 213 U.N.T.S. 221, E.T.S. No. 005 (4 November 1950), article 8; EU (2012), Charter of Fundamental Rights of the European Union, OJ 2012 C 326, arts. 7 and 8; Council of Europe, Convention for the Protection of Individuals with regard to the Automatic Processing of Personal Data, Council of Europe, CETS No. 108, 1981; Directive 95/46/EC of the European Parliament and the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (Data Protection Directive), OJ 1995 L 281, p. 31; Directive (EU) 2016/680;

evolved the most stringent of data privacy protections. Rooted in Article 8 of the European Convention on Human Rights (the right to respect for private life and family life), the European General Data Protection Act[7] restricts persistent monitoring of individuals online, as well as bulk collection of their personal data, and the retention of personal information. Drewer and Miladinova (2017), Seyyar and Geradts (2020), Akhgar and Wells (2018) have explored the impact of the EU's Law Enforcement Directive (LED)[8] on INTERPOL and national law enforcement agencies in Europe with respect to the protection of personal data. The LED shares similarities with the GDPR in establishing principles for the fair processing of information, including, but not limited to, lawful and fair processing, purpose limitation, adequate safeguards, accuracy of data, and transparency around the manner in which the data are collected, processed and stored. Importantly, the LED provides explicit protections for sensitive personal categories of data including race, ethnicity, religion, political opinions, union membership, sexual orientation, biometric data, health, and sex life information. These categories are precisely the types of information readily identified on social media. As observed in Sayyer (2020), recent ECtHR decisions[9] have emphasized the right to protection of one's personal data is not absolute; it must be balanced with the right of the state to protect society from crime and terrorism. Nonetheless, the LED is at

Regulation (EC) No. 45/2001 of the European Parliament and the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the institutions and bodies of the Community and on the free movement of such data, OJ 2001 L 8; OECD, "Recommendation of the Council Concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data," C(80)58/FINAL, revised on 11 July 2013.

[7]EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.

[8]Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA.

[9]*Volker und Markus Schecke GbR (C-92/09) and Hartmut Eifert (C-93/09) v Land Hessen* 2010).

odds with many of the OSINT tools now available to law enforcement. For instance, LED *Article 11* requires Member States shall provide for a decision based solely on automated processing, including profiling, which produces an adverse legal effect concerning the data subject or significantly affect him or her, to be prohibited unless authorized by Union or Member State law to which the controller is subject and which provides appropriate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention on the part of the controller. As discussed below, OSINT technologies, such as bulk data scrapers, and biometric identification, such as facial recognition used on photos posted on social media, are automated processes.[10] Article 27 of the LED requires agencies to undertake a privacy impact assessment (PIA), particularly on "new technologies" (at (1)) and where the "nature, scope, context, and purposes of the processing are likely to result in high risk to the rights and freedoms of natural persons…" OSINT protection process, automated and manual, when collected as Big Data are being employed many law enforcement agencies in direct contravention of many data privacy laws, chiefly because the information is deemed "open" or "public."

With the advancement of technology and user's increased use of social media, there is a heightened awareness amongst privacy advocates of the need to protect individual's privacy. As Wells and Gibson (2017) observed, both law enforcement and intelligence agencies must be mindful of public perceptions of their collection and exploitation of OSINT. If the perception is one of unmediated collection for the sake of collection, oversight minimal to non-existent and only lip-service paid to observance of existing privacy laws, trust in these institutions will further erode. Using SOCMINT to predict crime veers into the arena of big-brother panopticon of authoritarian regimes, the stuff of *Minority Report*. The collection and use of OSINT by public agencies requires strong guardrails.

Some countries, whilst beginning to institute regulations on use of OSINT by law enforcement, have inconsistent records of adherence to oversight. The UK's National Police Chief's Council (NPCC) issued guidance on the use of OSINT in investigations in 2016. In the guidance, specific reference is made to the Royal United Services Institute's (RUSI) "Ten Tests for the Intrusion of Privacy." The first rule is any intrusion into privacy must be in accordance with the law. The tests go on to specify

[10]See Balaji *et al*. (2021). Machine learning algorithms for social media analysis: A survey.

necessity, proportionality, restraint, effective oversight, recognition of necessary secrecy, minimal secrecy, transparency, legislative clarity and multilateral collaboration (NPCC, 2016). The guidance recommends officers read a RUSI report on the importance of law enforcement retaining the trust of the citizens they serve. The UK has over the past decades passed legislation designed to address intelligence collection[11] in addition to issuing the aforementioned guidance on law enforcement agencies' collection and use of OSINT. In the 1998 Human Rights Act, Article 8, privacy is referenced in the guidance with admonishments to ensure collection of data is justified, authorized, proportionate, auditable and necessary.

Canada's RCMP Operational Manual (OM), section 26.5, titled *Using the Internet for Open-Source Intelligence and Criminal Investigations* provides a framework for the collection and use of OSI and open-source intelligence based on a three-tier system (Tier 1: overt, Tier 2: discreet, and Tier 3: covert) (p. 1). The OM 26.5 defines OSI as unclassified, raw data that are derived from a primary source (e.g., the Internet) and can include any type of media in any format. The information is obtained, derived, or recovered lawfully, and is purchased or viewed from open or encrypted publicly available sources (e.g., websites, blogs, social networks, and online databases) (p. 4). The 2015 version of OM 26.5 stated that a national-level unit was "responsible for the oversight of all open-source intelligence and online investigational support activities in the RCMP." The 2019 version no longer includes oversight responsibility. It now specifies that the Federal Police's Tactical Internet Operational Support (TIOS) is responsible for: providing strategic advice and tactical OSINT operational support, conducting risk assessments of specialized OSINT tools, techniques or tradecraft, developing, coordinating and delivering advanced OSINT training, and providing advice to technical authorities on Internet network design, network security, software and desktop applications to support OSINT functions. The 2019 OM 26.5 delegated oversight responsibility from TIOS to the unit level (e.g., unit commanders and line officers), including the authorization and monitoring of forms officers are required to fill out and obtain for open-source activities (RCMP, p. 9). However, a recent audit by the RCMP found that although the national-level unit recognized that they could not fulfill their role as a national policy center, the absence of national oversight *increases the risk*

[11] See Regulation of Investigatory Powers Act (RIPA); the Management of Police Information 2010; Human Rights Act 1998; Data Protection Act 1998.

that OSINT will be gathered and used by employees across the RCMP without the appropriate authorization and without adhering to policy (RCMP, p. 10).

The US E-Government Act of 2002[12] was authored to ensure sufficient protections for the privacy of personal information as agencies implement citizen-centered electronic Government. Specifically, the Act requires a privacy impact assessment be conducted prior to "(i) developing or procuring information technology that collects, maintains, or disseminates information that is in an identifiable form; or (ii) initiating a new collection of information that — (I) will be collected, maintained, or disseminated using information technology; and (II) includes any information in an identifiable form permitting the physical or online contacting of a specific individual, if identical questions have been posed to, or identical reporting requirements imposed on, 10 or more persons, other than agencies, instrumentalities, or employees of the Federal Government."[13] PIAs must be conducted by the agency and reviewed by the Chief Information Officer or an equivalent official, and, where practical, make public the results of the Assessment.[14]

The Act also requires guidance to "(i) ensure that a privacy impact assessment is commensurate with the size of the information system being assessed, the sensitivity of information that is in an identifiable form in that system, and the risk of harm from unauthorized release of that information; and (ii) require that a privacy impact assessment addresses — (I) what information is to be collected; (II) why the information is being collected; (III) the intended use of the agency of the information; (IV) with whom the information will be shared; (V) what notice or opportunities for consent would be provided to individuals regarding what information is collected and how that information is shared; (VI) how the information will be secured; and (VII) whether a system of records is being created under section 552a of title 5, United States Code (commonly referred to as the 'Privacy Act')."[15] An agency's privacy obligations under the E-Government Act do not end with the initial publication of a Privacy Impact Assessment, rather a PIA must be revised continually "to reflect changed information collection authorities, business processes or other

[12](US) Pub. L. 107–347, 116 Stat. 2899 (codified as amended at 44 U.S.C. §3501 note).
[13]*Id.* at (1) *et seq.*
[14]*Id*. at (1)(B)(3).
[15]§208 (c)(B).

factors affecting the collection and handling of information in identifiable form."[16] The Privacy Act limits the collection, storage, and sharing of PII about US citizens, including social media.

It is without irony that some jurisdictions maintain much stricter privacy regimes around their civil servant's personal lives and data than they do of their other citizens. For instance, in the US, certain classes of individuals enjoy additional protections from private or identifying information being made public, and from harassment campaigns such as doxing. Law enforcement and court officers and personnel are covered under 18 U.S.C. §119 which provides protections for individuals performing certain official duties such as federal, state and local law enforcement officers[17] and government employees involved in federal prosecutions (including prosecutors and their staff). The protections include prohibitions against anyone who "knowingly makes restricted personal information[18] about a covered person, or a member of the immediate family of that covered person, publicly available"[19] with the intent to threaten, intimidate or incite the commission of a violent crime, or knowledge that the release of this information will prompt others too.[20] Additional protections apply to non-law enforcement federal workers.[21]

It is telling that in a recent government inquiry into police officers who abused their undercover positions (UPI, 2018), having exploited OSINT for the purposes of their investigations (Dawson and Brown, 2020) and mishandled UK citizen's personal data, including sharing it with private investigators, sought from the Chairman of the Undercover Policing Inquiry restriction orders to maintain the officers' privacy and anonymity (UPI, 2021b Ruling 2). Failing to recognize the paradox, UPI Chairman John Mitting invoked Article 8(2) of the European Convention

[16] *Id.* 2 §II.B.4.

[17] 18 U.S.C. §119(2) Protection of individuals performing certain official duties.

[18] *Id*. at (b)(1).

[19] *Id.* at (a).

[20] *Id.*

[21] See 18 U.S.C. 115 (threatening to assault, kidnap, or kill a federal official or employee, a former federal official or employee, or the family member of a current or former federal official or employee, in order to influence, impede, and retaliate against such current or former federal official or employee); and 18 U.S.C. 875(c) (transmitting in interstate or foreign commerce a threat to kidnap or injure another). Penalty: imprisonment for not more than five years.

on Human Rights, the right to privacy with respect to private and family life. Of the 16 officers who sought anonymity of their personal information, 10 were granted it. Certainly, the optics and incongruity of officers seeking privacy protections as to their private lives are not lost on British citizens. Privacy may not be something the average person thinks about until it is threatened or they understand it could be exploited in harmful ways. As Wells and Gibson (2017) observed, from "an outsider's perspective," it is not always clear just how well law enforcement agencies are applying the laws of data privacy and protection.

Law Enforcement Abuses OSINT

Recent surveys evidence law enforcement ahead in collecting and processing OSINT but frequently not in compliance with the spirit and letter of the legal privacy regime in their jurisdiction, or even with their own agency guidance. Whilst some jurisdictions have implemented laws, regulations and/or guidance specific to the collection of open-sourced information online by those engaged in criminal investigations and intelligence, *knowledge* and understanding of relevant legal frameworks including privacy and the evidentiary rules of procedure in the jurisdiction(s) in which officers and analysts are conducting investigative activities are also critical. Many countries have yet to promulgate general data protection and/or data privacy laws, let alone laws designed to provide parameters under which law enforcement may undertake OSINT collection. In those jurisdictions with frameworks specific to OSINT collection in criminal investigations, training in the rules and practical application can be seen to be failing. Given the volumes of personal information being actively collected, potentially the number of citizens wronged by current law enforcement approaches to supervising officers' OSINT practices could be in the hundred of millions. The US-based company, Anomaly Six, claims the ability to track the movements of 3 *billion* people in real time via combined mobile phone and social media surveillance (Biddle and Poulson, 2022). The company markets to law enforcement. As discussed above, North American, European and British national and local law enforcement agencies have paper policies that call for compliance with a range of rules governing the collection and processing of OSINT, including the conduct under PIAs, formal justification and approval mechanisms, mandatory training, and data minimization principles. Yet, these

articulated safeguards largely appear to be paper programs, with nominal compliance by the cybercrime investigators collecting the OSINT.

The civil society group, the Electronic Privacy Information Center (EPIC), has recently filed multiple civil suits against US agencies for failing to conduct a PIA, as required under the E-Government Act.[22] In one, *EPIC vs DHS (Media Monitoring Services),*[23] the US Department of Homeland Security was forced to admit that it had bypassed its own privacy officials, failed to conduct the legally required PIA, and had largely ignored the potential negative privacy impacts their social media monitoring services (MMS) have, including threats to individuals' First Amendment rights (EPIC, 2018). DHS was forced to suspend the MMS program and, whilst vowing to complete the required PIA, has not as yet done so. EPIC also recently filed suit against the US Postal Service,[24] also for failing to conduct a requisite PIA under the E-Government Act, when it sought to procure and employ an advanced social media surveillance tool produced by third-party Zignal Labs,[25] capable of tracking a social media "narrative" back to the individual who initiated the narrative and of identifying specific individuals as "influencers" as part of a larger surveillance system, Internet Covert Operations Program (iCOP).[26] The USPS' iCOP tracks social media posts of Americans and shares that information with other law enforcement agencies. iCOP includes multiple tools under its banner, including facial recognition software, and the controversial Clearview AI, a facial recognition software that scrapes images off public websites and connects them via biometrics to specific individuals (Winter, 2021).

[22] *Id*. at 10.

[23] In April of 2018, the DHS began seeking a contractor to develop "Media Monitoring Services" (MMS), a suite of digital tools that would have continuously tracked and analyzed media coverage and stored large volumes of personally identifiable information about journalists, bloggers, and social media users. The system would have collected and retained personal data such as "locations, contact information, employer affiliations, and past coverage." Within days of the agency's announcement, EPIC filed a (US) Freedom of Information Act request seeking the Privacy Impact Assessment DHS was required by law to produce prior to developing any online monitoring tools. When the agency failed to produce the PIA, EPIC filed suit [*EPIC v. DHS*, No. 18-1268 (D.D.C. filed 30 May 2019)].

[24] *EPIC v. US Postal Service*, No. 21-2156 (US D.D.C. filed 12 August 2021).

[25] Zingal Labs partnered with Anomaly Six, see Biddle and Poulson 2022.

[26] *Id*. at 27.

Clearview AI's database is fed by "scraping" photographs and facial pictures accessible online, in particular, those made available via social networks, for possible use by law enforcement authorities (comparing photos through facial recognition analysis against the database). The European Data Protection Board (EDPB) has ruled the use of Clearview AI's facial recognition system would be illegal within the EU, saying any use of the service by law enforcement in Europe would "likely not be consistent with the EU data protection regime" (EDPB, 2020). Clearview is currently being sued by the UK and Australian Information Commissioners and by several civil society groups.[27] Canada's Privacy Commission ruled Clearview AI violated, amongst other laws, Canada's federal privacy statute.[28] The USPS asserted they needed the iCOP program to "protect" US postal delivery workers (Winter, 2021) and for over two centuries have managed to deliver the post without reliance upon facial recognition tools and the scraping of social media. Yahoo News reported that the iCOP program appeared to be the USPS' "expanded and rebranded its 'cybercrime dark web' program," (Winter, 2021) one transformed into a broader covert operation. But for those Americans whose social media photos, profiles, and postings have been scraped, their data were not ring-fenced in the iCOP system. Rather, the information was shared and disseminated to the US Department of Homeland Security and

[27] See The Office of the Australian Information Commissioner (OAIC) and the UK's Information Commissioner's Office (ICO) have opened a joint investigation into the personal information handling practices of Clearview AI Inc., focusing on the company's use of "scraped" data and biometrics of individuals (see Press Release from the OAIC and UK's ICO open joint investigation into Clearview AI Inc., 9 July 2020, https://www.oaic.gov.au/updates/news-and-media/oaic-and-uks-ico-open-joint-investigation-into-clearview-ai-inc; Digital privacy rights groups Austria-based Noyb, UK-based Privacy International, Greece's Homo Digitalis, and Italy's Hermes Center for Transparency and Digital Human Rights filed a joint suit against Clearview AI https://fortune.com/2021/05/27/europe-clearview-ai-gdpr-complaints-privacy/; Ian Carlos Campbell, Clearview AI hit with sweeping legal complaints over controversial face scraping in Europe, The Verge (27 May 2021), https://www.theverge.com/2021/5/27/22455446/clearview-ai-legal-privacy-complaint-privacy-international-facial-recognition-eu; Office of the Privacy Commissioner of Canada (2021), Joint investigation of Clearview AI, Inc. by the Office of the Privacy Commissioner of Canada (2 February 2021), https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2021/pipeda-2021-001/#toc2.

[28] Personal Information Protection and Electronic Documents Act (PIPEDA), SC 2000, c 5.

dozens of other criminal intelligence centers. Investigative journalists identified that Clearview AI had *users at the FBI, Customs and Border Protection (CBP), Interpol, and hundreds of local police departments. In doing so, Clearview has taken a flood-the-zone approach to seeking out new clients, providing access not just to organisations, but to individuals within those organisations — sometimes with little or no oversight or awareness from their own management* (Mac *et al.*, 2020). Over 2,200 law enforcement agencies and institutions were identified as having run nearly 500,000 searches across Clearview AI's captured pool of more than 3 billion pictures from social media and "millions of websites."

Clearview has made it part of their business strategy to build market share by handing out free trial accounts to any law enforcement officer with a police department or governmental agency email address. In the US, the New York Police Department was forced to admit that it was unaware some 30 of its officers held Clearview AI accounts and that from NYPD, more than 11,000 queries of scraped social media data paired with Clearview's facial recognition tools had been run without the larger agency's knowledge, approval, or even contractual agreement with Clearview (Mac *et al*., 2020). An NYPD spokesperson told journalists that whilst the agency didn't have a contract with Clearview AI, they also hadn't instituted a written policy prohibiting culling user's personal information from OSINT, saying *Technology developments are happening rapidly and law enforcement works to keep up with this technology in real time* (Mac *et al*., 2020). Despite the technology having arrived several years before, NYPD stated it was "in the process" of updating an OSINT policy that addressed the use of facial recognition from photos pulled from user's private accounts. In other police departments, audits identified where despite the agency severing contractual relationships with third-party data scrapers such as Clearview AI, officers were found to have retained apps and continued to use OSINT despite a departmental ban (Mac *et al*., 2020).

The United States Department of Justice's guidance (DOJ, 2020) sets forth the legal constraints over the collection of OSINT. The guidance specifies that one of the principal rules is that law enforcement should not themselves become perpetrators of illegal acts. The guidance goes on to say that some of the activities it references implicate federal criminal law and may, in some instances, violate state laws or create civil liability. The guidance admonishes that passive collection will likely not constitute a federal crime, but accessing social media fora, chat sites and other online

platforms in an "unauthorized" manner could violate federal laws and some state-level invasion of privacy statutes.[29] Only the collection of communications openly posted on forums and platforms, and not engaging with users, would remain within the legal guidelines.

Clearview AI has sold or given access to its data pool of scraped SOCMINT and OSINT to law enforcement agencies in at least 26 countries outside the US including police forces in Australia, Belgium, Brazil, Canada, Denmark, Finland, France, Ireland, India, Italy, Latvia, Lithuania, Malta, the Netherlands, Norway, Portugal, Serbia, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom. In Canada, more than 30 law enforcement agencies were granted access to Clearview's software, including the Royal Canadian Mounted Police (Mac *et al*., 2020). Evidencing lack of oversight and adherence to relevant data privacy laws in their home jurisdictions, officers were identified as having accessed Clearview's software without the permission, or even knowledge, of their superior officers or agency management. When investigative reporters questioned the leaders of these various law enforcement agencies, the organizations admitted "the technology had been used without leadership oversight" (Mac *et al*., 2021).

In February 2021, the Swedish Authority for Privacy Protection (IMY) initiated an investigation against the Swedish Police Authority (SPA) for its use of Clearview AI. The SPA ultimately acknowledged that some employees had used the application without any prior authorization; unlawfully processed biometric data for facial recognition; and failed to conduct a data protection impact assessment required with the processing of the type of data Clearview AI scraped. The IMY concluded that the Police had not fulfilled its obligations as a data controller having failed to implement sufficient organizational measures to ensure and demonstrate that the processing of personal data gathered from Clearview AI had been carried out in compliance with the Criminal Data Act (IMY, 2021). IMY stated, "There are clearly defined rules and regulations on how the Police Authority may process personal data, especially for law enforcement purposes. It is the responsibility of the Police to ensure that employees are

[29]The Computer Fraud and Abuse Act (18 U.S.C. §1030); the Wiretap Act (18 U.S.C. §2511); and Access Device Fraud (18 U.S.C. §1029). Additionally, some states have created civil causes of action for online impersonation. *See*, e.g., WA ST 4.24.790 (7 June 2012) (Washington State statute for "electronic impersonation" — Action for invasion of privacy).

aware of those rules" (IMY, 2021). In addition to levying an administrative fine of SEK 2,500,000 (approximately EUR 250,000) on the SPA for infringements of the Criminal Data Act, the IMY specifically ordered the Police to conduct further training and education of its officers and analysts.

In January 2021, the RCMP published the results of an audit it conducted across its organization specific to OSINT activities in compliance with its policy. Overall, the audit determined that Internet-related open-source activities conducted across the organization were not consistent or compliant with OM 26.5. The audit identified that opportunities exist to develop a more robust governance framework and enhance national and divisional oversight of open-source activities. The audit found that many RCMP employees were not aware that an open-source policy existed or that it was applicable to the open-source activities that they specifically performed. The audit further concluded that training and information sharing are valuable investments to ensure that officers engaged in searching OSINT have access to current and up-to-date information, including relevant case law (RCMP, 2021, p. 2). The audit identified the RCMP could improve consistency and compliance with policy, training, infrastructure, and oversight to support open-source activities. The audit examined Internet-related open-source activities in support of criminal investigations and criminal intelligence gathering at the national and divisional levels from 1 April 2018, to 31 March 2019, assessing existing internal controls (e.g., policies and procedures, training and tools, monitoring and reporting mechanisms, etc.) that supported the use of Internet-related OSINT collection (RCMP, p. 7). However, many interviewees, ranging from criminal intelligence analysts to detachment commanders across divisions, stated that they were not aware of the open-source policy and/or roles and responsibilities assigned to their position. The audit found that roles and responsibilities related to OM 26.5 were not well understood by employees at all levels using OSINT. Without clearly established and communicated roles and responsibilities, there is a risk that OSINT will be inappropriately obtained and used in support of criminal investigations and criminal intelligence gathering (RCMP, 2021, p. 10).

The RCMP's own internal approval form required from senior officers prior to conducting OSINT searches was found to be largely ignored. The audit identified that only 14% (15 of 110) of employees had been appropriately authorized to conduct open-source activities. The majority of employees who were not authorized weren't even aware of the policy

requirement, or thought authorization didn't apply to them. Of the minority who did complete requests for approval, most were signed off by the wrong commanding officer, in contravention to the form's dictate. Officers in an intelligence unit didn't bother seeking approval to run OSINT queries as they felt by virtue of their role in the unit they were authorized, despite the RCMP's written policy providing no such exemption (RCMP, 2021, p. 11). The RCMP's audit also identified a failure adhere to tracking and monitoring of OSINT activities at the agencies' national and divisional levels. The audit report stated, this has resulted in the following: (*We*) *audit did not find evidence that open-source activities were being tracked or a risk that the lack of oversight and monitoring by the national policy centre could result in a lack of visibility over those who are conducting open-source activities* (RCMP, 2021, p. 12).

Europol currently provides the Secure Information Exchange Network Application (SIENA) to some 2,000 national authorities from 49 countries and 14 international partners (Europol, 2021). The SIENA platform facilitates an interconnected system that allows for the sharing and analysis of data uploaded to it, including OSINT the 2,000-plus agencies and partners (including the US, Interpol, Eurojust and others) have sourced. In 2019, 85,000 new cases were opened (Europol, 2021). As previously discussed, the EU enacted a robust data protection and privacy framework for Europol, for OSINT collection intended to ensure that Europol explicitly obtains approval under the relevant requirements, and its data collection falls within the mandate of the agency as regulated by the LED.[30]

But adherence to the data protection and privacy regulations has been inconsistent as recent disclosures have revealed. The European Union Agency for Law Enforcement Training (CEPOL), has been responsible for developing, implementing and coordinating training for law enforcement both within the EU and to non-EU countries with which the EU has overlapping security interests, e.g., the Balkans, Northern Africa and parts of the Near East. CEPOL has been providing advanced courses in OSINT gathering techniques. Training documents to European countries and

[30]Directive (EU) 2016/680 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA.

others include advice on how to use "sock puppets" — anonymous and fake social media profiles — to collect OSINT (see Privacy International, 2020). Training has even included recommended online platforms which can help manage fake accounts. These fake accounts not only violate the terms and conditions of most social media platforms, they also run counter to the EU's own Code of Practice which requires social media platforms to maintain "authenticity policies restricting impersonation and misrepresentation."[31] Instruction by CEPOL includes PowerPoint presentations that instruct how to "dork," which is the practice of identifying web pages that should have restricted access, but due to inferior or lax security practices, can be accessed using OSINT tools.

Europol, in its invitation to its fall 2021 Data Protection in Policing conference, published a statement admitting, *The human factor in data protection becomes more and more relevant in so many ways. Human intervention is an important safeguard not only when it comes to the increasing use of Artificial Intelligence including machine learning in law enforcement and beyond. Data protection will also only work on the ground if humans continue to believe in its added value. The human element is the building block of a healthy data protection culture in any organisation including in law enforcement. But sometimes things also go wrong. In the best case that is the moment when we can remind ourselves that we are all just humans, after all. In a bad scenario, humans have suffered serious impact on their fundamental right to data protection — or even worse* (Europol, 2021). As the Clearview AI case in Sweden and the CEPOL training highlight, oversight and training need to be constant.

The overarching EU privacy Directive that governs law enforcement's collection of OSINT is the LED: Art. (26) specifying that "any processing of personal data must be lawful, fair and transparent in relation to the natural persons concerned, and only processed for specific purposes laid down by law." Article 7 specifically calls for Member States to *provide for the competent authorities to take all reasonable steps to ensure that personal data which are inaccurate, incomplete or no longer up to date are not transmitted or made available. To that end, each competent authority shall, as far as practicable, verify the quality of personal data before they are transmitted or made available* (Art. 7(2)). However, the sheer volume of OSINT, particularly SOCMINT, persistently scraped by

[31] https://ec.europa.eu/digital-single-market/en/news/roadmaps-implement-code-practice-disinformation.

law enforcement agencies, ensures inaccuracies cannot be detected. Moreover, the veracity of a goodly portion of postings made to social media by individuals is questionable at the best of times.

Casavilla *et al.* (2018) demonstrated the extent of information leakage within Facebook. Utilizing two tools they built themselves, the latter known as "SocialSpy," the group was able to reveal the possibility of inferring hidden information of a user profile and retrieving private information from a user's public profile utilizing the network "surrounding" the person. The tool evaded user controls intended to keep some information "private." The friendship network of a victim showed how it is possible to infer additional private information (e.g., personal user preferences and hobbies) using the victim's friends of friends' network (a 2-hop of distance from the victim profile), and hence possibly deducing private information from the full Facebook network. The opportunities for law enforcement to ingest personal and erroneous user data in the collection of OSINT cannot be understated. The risks are significant without sufficient oversight and policies in place, including principles of data minimization and data quality checks.

Although open source is being used for all types of intelligence gathering and investigations (e.g., missing persons, hate crimes, homicides, break and enters, and drugs etc.), many of the RCMP interviewees did not realize that their open-source work required Tier 2 approval (advanced use requiring training. Only Tier 1 requires no training). The RCMP audit acknowledged that for Tier 2 users who had not been trained, there was a risk these officers and analysts might gather and use OSINT inappropriately (RCMP, p. 17). While interviewees indicated that the majority of OSI research was conducted passively (e.g., not engaging or interacting with subjects of interest), some exceptions were reported that were contrary to policy, such as joining closed Facebook groups in a proactive monitoring effort to obtain information on upcoming events such as a protest or demonstration from online discussions and using personal social media accounts to overtly try to contact a missing person (RCMP, p. 13). In some jurisdictions, monitoring protests could rise to civil rights violations. The same audit found that employees used a variety of methods to document OSI activities such as analyst work logs and notebooks, supplemental reports, detailed OSINT reports, simple or partial screen captures, e-mails, and narratives. Most analysts confirmed the use of discreet online identities and accounts to access open-source material. Some told auditors of storing gathered OSINT on personal or shared drives, or

even on USBs or other external devices that could easily be lost, misplaced. Auditors confirmed that there was widespread use of both personal and organizationally-issued mobile devices for OSINT activities by personnel. In detachments, most members stated that they use personal accounts to access OSINT (RCMP, p. 15).

The UNODC has instructed in its Criminal Intelligence Manual for Analysts (2011) that any analyst involved in handling OSINT "should receive specialist training in the subject" (UNODC, p. 12). In a 2020 global survey of cybersecurity professionals who collect and manage OSINT, fully one-third of respondents said that prior to commencing their current role collecting OSINT, they had no previous experience (Cybersecurity Insiders, 2020, p. 3). Out of more than 330 professionals, 85% had received little to no training in OSINT risks and techniques despite OSINT collection being their primary job duty. The survey asked how individuals were taught risks and regulatory requirements. Not only did most not receive training, over *half* reported that their organization provided no formal guidelines on what data should be considered a legitimate target for collection and analysis, or what laws and regulations required of them with respect to the agencies' collection and use of OSINT. Whilst more than 80% of the professionals agreed they used OSINT as part of their effort to combat cybercrime, nearly 30% of the respondents admitted their organizations had no oversight procedures in place to ensure analysts were not abusing OSINT tools; only 18% of the agencies had legal professionals providing oversight, with fully one-third reliant entirely upon a local group supervisor to provide an audit function. When asked how regularly their organization implemented any form of formal audit or review of professional's collection and use of OSINT, 34% said no audit was ever held or no regular review was standard (Cybersecurity Insiders, 2020).

Estimates are that there are over 100 "people" search sites in the US alone.[32] In 2013, KrebsOnSecurity (2013) featured a story that explored how the website Exposed.su had managed to get their hands on records and information on celebrities and US officials from a Russia-based online identity theft ring. Lack of meaningful privacy laws in the US coupled with data scrapers and brokers allowed to legally traffic in deeply

[32]Robertson A. (2017). "The long, weird history of companies that put your life on online," 21 March 2017, The Verge, https://www.theverge.com/2017/3/21/14945884/people-search-sites-history-privacy-regulation.

personal information about Americans, creates the perfect data environment for bad actors to take advantage and adds little to law enforcement's ability to source valid data. OSINT sits now at the intersection of cybersecurity and the right to privacy.

The risks associated with the unchecked use of OSINT by law enforcement in most jurisdictions are significant. How data sourced from OSINT are ultimately perceived is largely predicated on how it is processed and analyzed. Gradecki and Curry (2017) argued the context in which data are "(re)framed can influence how it is perceived." Officers and analysts lacking proper training may miss nuance, circumstance and situational factors in collected OSINT that results in them erroneously labeling it as dangerous or worse, failing to identify it as exculpatory. Unsupervised and unregulated gathering of open-source data allows for misuse or misappropriation of SOCMINT by law enforcement for the purposes of criminal evidence, breaching privacy laws and transforming it into an extension of an abuse of power.

Amnesty International has sought to answer if in highlighting or reposting videos they obtain whilst monitoring individual cases of human rights abuses via OSINT, they don't risk re-traumatizing the individuals depicted in the videos. By using content that captures ritualized abuses and drawing attention to the crimes, they ask, do they not create the risk of further or greater humiliation to the subjects, who never consented to be filmed in the first place? (Amnesty, 2020). These concerns should be part of law enforcement's considerations when scraping video and photographic information from SOCMINT, for it is not law enforcement's job to amplify abuses but rather protect from them.

Recommendations

The volumes range and of data OSINT can now provide to law enforcement are enticing. But, as Ünver (2018) has observed, the ease or swiftness of the acquisition of intelligence does not necessarily mean it is good intelligence. "Although democracies may lose time and range with their intelligence operations through the constraints set by safeguards, they more than make up for this shortcoming in two areas. First, due to intelligence safeguards and oversight mechanisms, agencies have to pass through a review system that tests the rationale, reasoning and strategic utility of surveillance practices" (Ünver, 2018, p. 17). Moreover, proactive oversight, including independent monitors or ombudsman, regular

audits and levels of authorization are more likely to catch oversteps, mistakes, or misjudgments early on, preventing or limiting damage both to citizen's privacy rights and agency reputation. Without guidance, clear roles and responsibilities, adequate monitoring and oversight, and regular training, law enforcement agencies will be unable to adequately adapt to rapidly evolving tools and technologies that will only increase the volume of information available via OSINT.

Training of officers and analysts must include teaching how to protect from unsolicited access to private information; what constitutes "open" versus available, the conduct of PIAs, accountability processes; and relevant case law and legislation. Law enforcement engaged in OSINT collection or analysis must also be annually certified to provide auditable standards.

As the RCMP audit concluded, an effective governance framework would serve to reduce the risk that OSINT is inappropriately obtained and used in support of criminal investigations and intelligence gathering, and promote the visibility of all employees conducting open-source activities organisation-wide (p. 8). The RCMP had assumed the OSINT policy it had updated in 2019 (OM, 26.5) "would provide adequate guidance for capturing, storing and retaining OSI(NT)," but came to the conclusion that having failed to widely consult in the development of the Policy, many officers and analysts did not recognize the policy as applicable to them. A judicial decision in British Columbia from 2017 (*R. v. Hamdan*, 2017, BCSC 867) highlighted improper captures of OSINT by the RCMP that resulted in an unsuccessful prosecution and has increased the need for additional prudence. The audit identified that many officers were not even aware of this case law. The RCMP in its guidance had referenced the *Hamdan* case, but national guidance was never established on how to properly capture OSINT to ensure court requirements are met under the criteria of the case established. Agencies involved in collecting and using OSINT in investigating crime must build in regular consultation with their legal counsel to learn of developing case law, receive assistance in interpreting the law, and in advancing training and processes designed to maintain evolving standards.

US DOJ guidance (2020) has specified that law enforcement agencies engaged in OSINT collection "should establish policies and protocols that have been vetted with its legal counsel to guide its employees' and contractors' activities on forums (and anywhere else)" (p. 7). The guidance refers to these policies as "rules of engagement" or compliance programs intended for employees incurring agency liability or placing it in legal

jeopardy, although no mention is made of risks to constitutional protections or user's privacy.

IBM and Amazon both announced they'll no longer provide facial recognition services to law enforcement and have called on the US Congress to increase regulation to help ensure future deployments of such software meet ethical standards (Davis, 2020). Clearview AI and other companies like them, continue to sell their services to law enforcement around the world. Democratic societies need to escalate their efforts toward implementing data privacy protections consistent with the GDPR and similar legislation. The GDPR has not throttled criminal investigators and intelligence analysts from collecting and using OSINT, but it has gone further in safeguarding citizens' data that might otherwise be exploited. Better the adoption of more stringent privacy protections than to allow authoritarian governments to determine how the data from the four petabytes of data Facebook generates daily[33] or the tens of billions of devices connected or about to connect to the Internet are used.

Opportunities exist to develop a more robust governance framework and enhance national- and local level-oversight of OSINT collection. Without clear roles and responsibilities and adequate monitoring and oversight, visibility over those who are conducting open-source activities will so wanting. Seyyrr and Geradts (2020) have shown how privacy risk assessments may feasibly be conducted on large-scale OSINT collection efforts by law enforcement. Their proposed method provides for specific privacy measures designed to comply with the LED. Taking the GDPR as the starting point, this is the time for legislators to enact both more stringent data protection and promulgate oversight mechanisms and transparency over how law enforcement is using and protecting the OSINT they do collect. Trust by the public is essential to democracy, and trust in law enforcement is a critical component. Criminal investigations can benefit from OSINT if it is used responsibly, and officers and analysts embrace ethical guidelines as part of the effort to combat cybercrime and cyber threats from malevolent actors and as part of their role in safeguarding the rule of law.

The starting point for most law enforcement agencies should be a governance framework that includes clear roles, responsibilities and oversight out in place for open-source activities in support of criminal investigations and criminal intelligence gathering. Drawing from the UK

[33] https://www.internetlivestats.com.

NPCC's (2016) 10 privacy principles, policies should be developed around the use of OSINT. These must be adequate, maintained, and clearly communicated with regular training provided. At every level, agencies, offices, and departments within agencies need OSINT oversight. There should be a central authority at the agency level that works with stakeholders, including prosecutors, privacy experts, civil society and human rights groups who develop the overarching oversight framework and mechanisms. Below that, office-level OSINT champions must feed new online developments and operational challenges back up to the central authority, whilst also working directly at the departmental level to ensure compliance with laws and guidance. Agencies must annually seek independent auditing or testing of their OSINT collection and processing programs, including review of any third-party software employed in agency online data collection, adherence to storage and destruction rules, and updated PIAs conducted.

The UK's National Crime Agency's Digital Investigation and Intelligence 2015 mandate acknowledged that *Innovation without ethics is a risk to consent-based policing. For policing the challenge is to empower personnel to be ethical actors as well as lawful ones*. (NCA, 2015). But progress in devising meaningful data ethics, particularly with respect to privacy, continues to lag behind the growth in the forms of personal data now available online.

One means of better addressing the risks to users' privacy that OSINT collection by law enforcement poses would be to both standardize the manner in which data are collected and to incorporate further the GDPR's legal requirements, particularly establishing the equivalent of national data commissioners tasked with auditing and testing criminal investigator's adherence to the law, and soft laws such as agency guidance. With respect to standards, in the early days of electronic evidence and discovery standardization, stakeholders and those in fields of mutual interest worked through the ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission), together who form the specialized system for worldwide standardization (ISO, 2021) to arrive at ISO/IEC 27050, Information Technology–Security Techniques–Electronic Discovery, intended to harmonize terminology, describe core concepts, and provide guidance. Not intended to contradict or supersede local jurisdictional laws and regulations, ISO standards have a critical role to play in harmonizing approaches across jurisdictions. Given the open-border nature of OSINT, the development of

a standard for the collection, processing and use of OSINT, for law enforcement and the private sector would help to erect guardrails against the abuse of user's privacy. The manner in which ISO standards are created includes National bodies that are members of ISO or IEC who participate in the development of the Standards through technical committees established by their respective organizations to deal with particular fields of technical activity. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. The inclusion of stakeholders would allow privacy advocates sitting alongside law enforcement and prosecutors to arrive at "reasonable" standards of collection.

The legal and ethical requirements to protect individual's privacy and the transboundary nature of OSINT require multiple steps be taken by law enforcement, legislators, regulators, civil society and privacy advocates. The platforms that host social media and other sites and the third parties who provide the OSINT tools for bulk collection cannot be relied upon to comply with privacy principles or even secure their user's data from scrapers and collection tools. Technological advances will not slow or cease.

- Law enforcement must undergo minimum training with regard to privacy, proportionality, and probative value with respect to OSINT collection.
- Transparency, accountability, and effective oversight of state agencies collecting and processing OSINT should be codified into national laws.
- The international community should develop ISO standards for collection and use of OSINT by public agencies establishing minimum good practice method and processes.
- Countries should require all public agencies conduct privacy impact statement prior to collecting, using or anticipating the use of OSINT.

References

Akhgar, B. and Wells, D. (2018). Critical success factors for OSINT driven situational awareness. *European Law Enforcement Research Bulletin*, 18, 67–74. http://shura.shu.ac.uk/22734/.

Amnesty International, Citizen Evidence Lab (2020). How OSINT helps us hold governments to account during the COVID-19 pandemic: Ethical questions.

1 May 2020. https://citizenevidence.org/2020/05/01/osint-COVID-19-pandemic/. Accessed on 13 August 2021.

Balaji, T. K., Annavarapu, C. S. R., and Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. *Computer Science Review*, 40, 100395.

Bellingcat (2020). Military and intelligence personnel can be tracked with the untapped beer App. 18 May 2020. Bellingcat. https://www.bellingcat.com/news/2020/05/18/military-and-intelligence-personnel-can-be-tracked-with-the-untappd-beer-app/.

Biddle, S. and Poulson, J. (2022). American Phone-tracking Firm Demo'd Surveillance Powers by Spying CIA and NSA. The Intercept, https://the intercept.com/2022/04/22/anomaly-six-phone-tracking-zignal.

Cascavilla, G., Beato, F., Burattin, A., Conti, M., and Mancini, L.V. (2018). OSINT — Open Source Social Network Intelligence: An efficient and effective way to uncover "private" information profiles. *Online Social Networks and Media*, 6, 58–68.

Commission nationale de l'informatique et des libertés. "Law enforcement Directive": What are we talking about? 2 June 2021. https://www.cnil.fr/en/law-enforcement-directive-what-are-we-talking-about.

Cybersecurity Insiders (2020). Cyber threat intelligence report. https://www.cybersecurity-insiders.com/portfolio/2020-cyber-threat-intelligence-report/.

Davis, R. (2020). IBM will no longer build facial recognition tech, sends letter to Congress. 9 June 2020, *IoT News*. https://iottechnews.com/news/2020/jun/09/ibm-no-longer-build-facial-recognition-tech-letter-congress/.

Dawson, J. and Brown, J. (2020). Undercover policing in England and Wales, House of Commons Briefing Paper No. 9044, 5 November 2020. https://researchbriefings.files.parliament.uk/documents/CBP-9044/CBP-9044.pdf.

Drewer, D. and Miladinova, V. (2017). The big data challenge: Impact and opportunity of large quantities of information under the Europol Regulation. *Computer Law & Security Review*, 33, 298–308.

Dyson, F. J. (1979). *Disturbing the Universe*. Alfred P. Sloan Foundation series. Harper & Row.

Edwards, L. and Urquhart, L. (2016). Privacy in public spaces: What expectations of privacy do we have in social media intelligence? *International Journal of Law and Information Technology,* Autumn, 24(3), 279–310.

Electronic Privacy Information Center (EPIC). EPIC v. DHS (Media Monitoring Services). Background. https://epic.org/foia/dhs/media-monitoring-services/.

European Data Protection Board (EDPB). Response to MEPs Sophie in 't Veld, Moritz Körner, Michal Šimečka, Fabiene Keller, Jan-Christoph Oetjen, Anna Donáth, Maite Pagazaurtundúa, Olivier Chastel, concerning the facial recognition app developed by Clearview AI, 10 June 2020. https://edpb.europa.eu/

sites/default/files/files/file1/edpb_letter_out_2020-0052_facialrecognition.pdf.

Europol (2020). Secure Information Exchange Network Application (SIENA). https://www.europol.europa.eu/activities-services/services-support/information-exchange/secure-information-exchange-network-application-siena.

Europol (2021). Invitation to human after all — Data protection in policing conference, hosted by The Europol Data Protection Experts Network (EDEN) 18–19 October 2021. https://www.europol.europa.eu/events/human-after-all-%E2%80%93-data-protection-in-policing.

Frederick, K. (2019). Hearing on "How Corporations and Big Tech Leave Our Data Exposed to Criminals, China, and Other Bad Actors." 5 November 2019, Prepared testimony before the Senate Judiciary Subcommittee on Crime and Terrorism, Center for a New American Security.

Global Initiative Against Transnational Organized Crime (2021). The global illicit economy: Trajectories of transnational organized crime, March.

Gradecki, J. and Curry, D. (2017). Crowd-sourced intelligence agency: Prototyping counterveillance. *Big Data & Society*, January–June 2017, 1–7.

International Standards Organization, ISO/IEC 27050-1: 2016, Information technology — Security techniques — Electronic discovery — Part 1: Overview and concepts. https://www.iso.org/standard/63081.html.

KrebsOnSecurity (2013). Credit Reports sold for cheap on the underweb. 13 March 2013. KrebsonSecurity.com, https://krebsonsecurity.com/2013/03/credit-reports-sold-for-cheap-in-the-underweb/.

Lewis, P. and Evans, R. Secrets and lies: Untangling the UK "spy cops" scandal. *The Guardian*, 28 October 2020. https://www.theguardian.com/uk-news/2020/oct/28/secrets-and-lies-untangling-the-uk-spy-cops-scandal.

LexisNexis (2012). *Law Enforcement Personnel Use of Social Media in Investigations: Summary of Findings*. LexisNexis Risk Solutions Government.

Mac, R., Haskins, C., and McDonald, L. (2020). Clearview's facial recognition App has been used by the justice department. ICE, Macy's, Walmart, and the NBA. *BuzzFeed News*, 27 February 2020. https://www.buzzfeednews.com/article/ryanmac/clearview-ai-fbi-ice-global-law-enforcement.

Mac, R., Haskins, C., and Pequeno, A. (2021). Police in at least 24 countries have used clearview AI. Find out which ones here. *BuzzFeedNews*, 25 August 2021. https://www.buzzfeednews.com/article/ryanmac/clearview-ai-international-search-table.

Mahmood, M. and Ungoed-Thomas, J. (2012). Tuppance a fact: The starting price for your stolen life. 18 March 2012. *The Sunday Times*, pp. 12–13.

Meyer, D. (2021). European privacy activists launch international assault on Clearview AI's facial recognition service. 27 May 2021. *Fortune.* https://fortune.com/2021/05/27/europe-clearview-ai-gdpr-complaints-privacy/.

National Crime Agency (NCA) (2015). Digital investigation and intelligence: Policing capabilities for a digital age. April 2015. https://www.uk-osint.net/documents/Digital_Investigation_and_Intelligence_Policing_capabilities_for_a%20digital_age_April_2015.pdf.

National Police Chief's Council (NPCC) (2016). NPCC guidance on open source investigation/research (Restricted version). April 2016. https://www.uk-osint.net/documents/003525-16_npcc_guidance_redacted.pdf.

Office of the Australian Information Commissioner (OAIC). (2020). Press Release 9 July 2020. OAIC and UK's ICO open joint investigation into Clearview AI Inc. https://www.oaic.gov.au/updates/news-and-media/oaic-and-uks-ico-open-joint-investigation-into-clearview-ai-inc/.

Office of the Privacy Commissioner of Canada (2021). Opinion, PIPEDA Findings #2021-001, Joint investigation of Clearview AI, Inc. by the Office of the Privacy Commissioner of Canada, the Commission d'accès à l'information du Québec, the Information and Privacy Commissioner for British Columbia, and the Information Privacy Commissioner of Alberta, 2 February 2021, https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2021/pipeda-2021-001/#toc7.

Pastor-Galindo, J., Nespoli, P., Marmol, F. G., and Perez, G. M. (2020). The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends. 16 January 2020, *IEEE Access*, Doi: 10.1109/ACCESS.2020.2965257.

Privacy International (2020). Revealed: The EU training regime teaching neighbors how to spy, 10 November 2020. https://privacyinternational.org/long-read/4289/revealed-eu-training-regime-teaching-neighbours-how-spy.

Royal Canadian Mounted Police (2021). Audit of Open Source Information, vetted report, January 2021, https://www.rcmp-grc.gc.ca/en/audit-open-source-information.

Royal United Services Institute (2015). A democratic license to operate: Report of the independent surveillance review panel. Whitehall Reports, 13 July 2015. https://rusi.org/publication/whitehall-reports/democratic-license-operate-report-independent-surveillance-review.

Seyyar, M. B. and Geradts, Z. J. M. H. (2020). Privacy impact assessments in large-scale digital forensic investigations. *Forensic Science International: Digital Investigation*, 33. https://doi.org/10.1016/j.fsidi.2020.200906.

Su, J., Goel, S., Shukla, A., and Narayanan, A. (2017). De-anonymizing web browsing data with social networks. Princeton University. https://www.cs.princeton.edu/~arvindn/publications/browsing-history-deanonymization.pdf.

Swedish Authority for Privacy Protection (2021). Press Release 12 February 2021. Swedish DPA: Police unlawfully used facial recognition app. https://

edpb.europa.eu/news/national-news/2021/swedish-dpa-police-unlawfully-used-facial-recognition-app_en.

Thompson, D. (2010). Google's CEO: "The laws are written by lobbyists." *Atlantic*, 1 October 2010. http://www.thatatlantic.com/technology/archive/2010/10/googles-ceo-the-laws-are-written-by-lobbyists/63908.

Undercover Policing Inquiry (2018). In the matter of section 19 (3) of the Inquiries Act 2005, Application for restriction orders in respect of the real and cover names of officers of the National Public Order Intelligence Unit and its predecessors/successor units, "Minded to" Note 3, 15 November 2018.

Undercover Policing Inquiry (2021a). Chairman rules on anonymity applications from 16 NPOIU undercover police officers. Press Release, 1 September 2021. https://www.ucpi.co.uk/2021/09/01/chairman-rules-on-anonymity-applications.

Undercover Policing Inquiry (2021b). In the matter of Section 19(3) of the Inquiries Act 2005, Application for restriction orders in respect of the real and cover names of officers of the National Public Order Intelligence Unit and its predecessors/successor units, Ruling 2, 1 September 2021.

Ungureanu, G. (2021). Open Source Intelligence (OSINT): The way ahead. *Journal of Defense Resources Management*, 12(1), 177–200.

United Nations Office on Drugs and Crime (2011). Criminal Intelligence Manual for Analysts, April 2011, United Nations, Vienna.

United States Department of Justice (DOJ) (2020). *Legal Considerations when Gathering Online Cyber Threat Intelligence and Purchasing Data from Illicit* Sources, v.1, February 2020, Cybersecurity Unit, Computer Crime & Intellectual Property Section, Criminal Division.

United States Office of the Director of National Intelligence (2022). What is Intelligence? https://www.dni.gov/index.php/what-we-do/what-is-intelligence.

Ünver, H. A. (2018). Digital open source intelligence and international security: A primer. Edam Center for Economics and Foreign Policy Studies.

Wells, D. and Gibson, H. (2017). OSINT from a UK perspective: Considerations from the law enforcement and military domains, XVI. In *Proceedings Estonian Academy of Security Sciences, 16: From Research to Security Union*, Estonian Academy of Security Sciences, pp. 84–113.

Williams, H. J. and Blum, I. (2018). *Defining Second Generation Open* Source *Intelligence (OSINT) for the Defense Enterprise*. RAND National Defense Research Institute.

Winter, J. (2021). Facial recognition, fake identities and digital surveillance tools: Inside the post office's covert internet operations program. *Yahoo News* (18 May 2021). https://news.yahoo.com/facial-recognition-fake-identities-and-digital-surveillance-tools-inside-the-post-offices-covert-internet-operations-program-214234762.html.

Wondracek, G., Holz, T., Kirda, E., and Kruegel, C. (2010). *A Practical Attack to De-Anonymize Social Network Users*. Santa Barbara: University of California. https://sites.cs.ucsb.edu/~chris/research/doc/oakland10_sonda.pdf.

https://doi.org/10.1142/9781800612860_0009

Chapter 8

The Balance of Opinion in Social Media Regulation — Regime Stability and Risk in Democratic and Non-Democratic Nation-States

Vignesh Ram

Introduction

The advent of social media has altered the perceptions of how information flows are controlled between state and society. While information has remained a key component of influence for governments and other similar actors to influence society, nation-states have relied on the use of information during times of crisis and peace to influence the behavior of societies. For instance, the use of information and propaganda in warfare and public diplomacy in peacetime has been a key feature of influencing populations of another state, while democratic states and non-democratic regimes use the flow of information and narrative building to influence and retain political power.

The rapid and unregulated growth of social media and its immense influence among populations have led to serious consequences for both democratic and non-democratic states. The revolutions in the Middle East, which the social media led expose in authoritarian regimes, have led to detrimental domestic and regional geopolitical changes around the world. Similarly, democratic regimes which have lost the monopoly over information and

limited or no ability to control the flow of information over the Internet have faced internal security challenges as well as threat from internal subversion to ruling political authority. While the limits of dissent are clear in authoritarian regimes, in democratic regimes, the same limits traverse through gray areas facing several legal challenges and push cultural and normative boundaries in the process. The culturally polarizing factors lead to a greater impact on politics and the development of polity.

The regulation of social media has been a vigorous debate in both democratic and authoritarian regimes which see the access to information and the ability to regulate it in a limited fashion as a challenge to maintaining political power. In democracies, the regulation (due to lack of censors) provides a challenge and restricts "free speech." At the same time, it also provides a stifling of space for dissent which can also be misused to further agendas against the national security interests of states. In authoritarian regimes, the limit of free speech is defined, and red lines are drawn toward dissent (which are mostly absent).

In the above context, the complexity presented by social media remains high for both democratic and authoritarian regimes. This complex dynamic provides a key opportunity for citizenry to strongly consider and question state authority and effect positive changes in the society. Moreover, there are challenges that present a key understanding of how cardinal principles of association between citizenry and governing authority (social contract) are altered in the day and age of social media.

In the above context, the study would explore four objectives. It would start with assessing the methods in which public opinion is shaped by media and social media and how it impacts politics and regimes in different states. The study would then move on to understand the role of social media and methods of its regulation in democratic and non-democratic regimes. In doing so, it will explore the concept of "balance of opinion" in the form of existing conceptualizations of dissents, freedom of speech and expression, privacy and classical concept of social contract, and its application in the social media era providing the paper with the ability to understand the impact of social media in the contemporary society and contrast it with the tradition media usage by states. Finally, the paper will aim to devise an understanding of regime stability and limits of balance between national security and rights infringement in the use of cyber technologies by state actors in the current context.

Social Media and Politics

In the aftermath of the Arab Spring protests about a decade back in 2011, it became amply clear that a new fusion of politics and technology was reshaping the discourse between people and government. The democratization effect of technological progress, especially in non-democratic regimes, had a potential impact not only in the states with authoritarian regimes but also possible consequences for democracy. The collective mobilization actions of social media have added a new dimension to understanding the destabilizations in society due to popular unrest and related studies. Several studies have explored the phenomenon of social media as a catalyst in social movement with political implications. Nevertheless, some studies have pointed out that an increase in the availability of social media does not necessarily lead to protest and that it should be clear from the negative correlation between the amount of media penetration and the number of protests and highlight that politics is an important factor and a catalyst for the protests (Wolfsfeld *et al*., 2013). Nevertheless, as several indicators leading to the political protests note, socioeconomic factors leading to the Arab Spring in Tunisia, for instance, highlight those conditions were ripe for such dissent and social media played a key role in amplifying these protests. Some studies have argued the need to also look at the social, political, historical, and economic factors which were important to which social media just played a secondary role (Smidi and Shahin, 2017).

However, recent studies about the nature of the protests and the role of social media have highlighted and ascertained the fact that social media had played the role of an amplifier and provided more scope for people who shared an interest in democracy to build extensive social networks and organized political action. Social media became a critical part of the toolkit for greater freedom (O'Donnell, 2011). Some reports have also suggested that social media acted as an amplifier of the fact of what was happening on the ground to the outside world rather than inside it and notes it acted more like a "megaphone" as opposed to a "rallying cry" (Pew Research, 2012). This of course was amplified later by traditional media and became a source to understand the reality of protests on the ground.

Recently, social media has started to reshape the debates in democratic setups indicating the potential power of providing transparency and accountability of the polity toward the electorate. One key instance is to

understand how elections have been transformed with the advent of social media. However, it has had both positive effects in the form of more transparency and communications and also at the same time become a major tool of influence and shaping of narratives to shape voter agenda (Bonney, 2018). In most democracies where freedom of speech and expression are engrained in the constitution, the freedom of the press was considered paramount to air and voice opinions, at several times, media which traditionally covered political discourses were able to present the views which it felt were important and was "needed" for the society at a priority.

Traditionally, gatekeepers in the traditional media space are considered "opinion shapers" for a larger population and propose how information is relayed and portrayed. Kurt Zadek Lewin is often credited with proposing the usage of the word gatekeeping in 1943. In coverage of wars and conflict over a point of time, the concept of "embedded journalists" questioned the objectivity of news and information about conflicts which were relayed to the audience. This was particularly pointed out by research studying the Iraq war where the concept of embedded journalism raised some doubts about the non-biased portrayal of news (Maguire, 2017). Comparing this instance to the previous examples of the Arab Spring as argued elsewhere in this paper, it can be very well understood that social media has not only been able to provide a key sense of accessibility to citizens and removed the gatekeepers but has also empowered citizens to directly interact and produce content. Social media was designed as a tool for the empowerment of the user and giving them broadcast power. As one viewpoint notes, social media in the process has empowered people and as a result can bring down (or disrupt) political systems but does not have anything to replace them with (Bloomberg, 2015). However, when we largely speak about the disruptive power of social media, it is also prudent to reflect on the statement provided in the preceding arguments. While governments face challenges in censoring the contents on social media, they also find quite often in charge of trying to manage the narrative, often to find opposing forces and dissenting views.

The use of social media to malign the image of political opponents and spread misinformation and disinformation campaigns has got a trend with the entrance of politics into the social media space. While in a conventional day-to-day sense this seems like a great tool of outreach for political entities to influence their populace, polarizing and often propaganda in the form of hate speech and fake news has found itself as a "normalized" part of the political discourse in such times. In these times of media transition

and the preponderance of new media systems, analyses have suggested that the flow of information and intermingling of various actors produces a hybrid system. It is further noted that *Actors in this system are articulated by complex ever-evolving relationships based upon adaptation and interdependence and concentrations and diffusions of power. Actors create, tap, or steer information flows in ways that suit their goals and in ways that modify, enable, or disable the agency of others, across and between a range of older and newer media settings. Moreover, the authors also assert that the citizens have an interest in using social media to influence politics and political decisions. However, they assert that they jointly use old and new media strategies they combine older and newer media in effective new ways* (Chadwick *et al*., 2016).

The proliferation of fake news and misinformation and disinformation campaigns have been rampant in election campaigns around the world. The exposure of the Cambridge Analytica scandal was well documented in the expose by a journalism team which revealed that the firm had used metadata from popular social networking websites and prepared a carefully orchestrated campaign including the use of fake news and fear of conflict and violence targeted at population including the young who had access to technology and were connected to social media networks (Crabtree, 2018). Correlation is also often found in the enforcement of information on social media with electoral gains by more conservative and fringe political elements in the political space. Technology that allows social media to galvanize democracy activists can be used by hate groups seeking to organize and recruit. As more and more people have moved online, experts say, individuals inclined toward racism, misogyny, or homophobia have found niches that can reinforce their views and goad them to violence (Laub, 2019).

The proliferation of political fake news is not just a phenomenon in fragile democracies that are on a path to further liberalization but have also found a more vocal position in democracies that have been quite advanced in their democratic experience. In the United States, for instance, studies have repeatedly indicated that the rampant misinformation floated on social media in the 2016 elections and following that in 2020 has increased manifold. Researchers have found that long-standing arguments that the company's algorithms fuel the spread of misinformation over more trustworthy sources are proven to be true (Dwoskin, 2021). Nevertheless, in democracies, often the reverse could also be true about fake news and misinformation campaigns. While the allegations of the

impact on those running misinformation campaigns were often delegated to domestic audiences, a new strain of accusations of election tampering by international forces or other powers became a key rallying point during elections to swing voter sentiment and opinions.

Similarly, there could also be key challenges in diverse multiparty democracies. Side-lined by political parties with absolute majorities, opposition parties are also found to thrive with the use of social media to malign and score political victories. India possesses a diverse mosaic of cultures and political actors, and with the growing number of users on most of the available social media platforms from the country, the use of the medium in communicating information and also the use of the premise for propaganda and misinformation has been a key challenge which the population has had to confront with. This is especially true during and around the elections. It was observed that Facebook conducted a campaign of weeding out multiple pages aligned toward the ruling Bharatiya Janata Party (BJP) as well as the Indian National Congress (INC) around the 2019 general elections in the country. Facebook took down over 700 assets that were posting partisan content on Indian politics ahead of the country's national elections. A cursory analysis of the pages linked to both political parties which were primarily on Facebook found inauthentic behavior and the use of covert assets to push their message across (Karan and Nimmo, 2019).

In democracies, while opinions can be subjective and can be seen through the lens of freedoms accorded by the constitution to the citizens, subjective curtailment of information from reaching citizens hits a particular gray zone in terms of having control over the narrative (especially political) in the hands of technological companies (which in international relations translate as non-state actors) control over the narrative of either part in a democratic setup. Here, the major question that arises would stem from answering questions about legitimacy in often closely polarized political debates in elections of course with the addition of social media-fueled information dissemination. Moderating and controlling the limits of how much social media can control the political narrative in politics has become an important debate in most democracies.

The decision by Facebook to ban former President Donald Trump provided an increased understanding of how powerful probably tech companies had become and the importance of social media in controlling the narrative among people. Facebook in its defense provided the rationale that the former US president had used its platform "to incite violent insurrection against a democratically elected government." This raises

important questions on what authority do social media companies authorize their ability to ban political entities especially when it comes to members of the ruling elite (Andrew, 2021). Similarly, Twitter found itself in conflict with the political establishments in India. The Indian government under its new Information Technology regulations asked major intermediaries, such as Facebook, WhatsApp, and Twitter, to comply and appoint intermediaries to resolve potential conflicts over the content and provide access to its servers in case of law enforcement related cases. While Twitter did not comply with the law, the government of India had decided to remove the safe harbor it had provided to the platform against content posted by its users. In a similar vein in its effort to "identify" possible media propaganda, Twitter started to also label media as "manipulated" akin to "fake or targeted content" which may not be true (*The Times of India*, 2021). However, the subjectivity of content and the role of intermediaries, such as Twitter, to moderate it according to their knowledge and world views have been questioned in the process.

In both of the above contexts, two basic lines of the argument need more deliberation in the course of the discussions on social media and democracy. One crucial point is to understand the concept of digital sovereignty and the other is to understand the concept of technology ethics. These two concepts have become important in the light of the growing influence of social media and the interaction thereof between the state and non-state actors, such as technology companies, which have become a key factor in access to information as well as an integral part of our lives. The debate on tech ethics, it has been argued, should focus not only on advertising revenue which in turn snowballs into algorithms working to show similar content to users but also move on to revisit user content to avoid the spread of various malpractices, such as fake news, misinformation, propaganda, or biased political agenda-setting (Dube, 2021).

While social media has become an important part of the political ecosystem, it is not surprising as to how important social media has become for political parties and not so surprising as to why they try to protect any form of loss of control toward its moderation. The question of technological companies deciding on who stays on the platform and who goes off the platform or even for a matter of fact what is reasonable or not and also for whom also excludes the purpose of an inclusive and free debate which the opening arguments in this section which this section spoke about with regard to closed regimes. In the democratic setup, a few critics of tech companies banning individuals or removing content or verified

tags from leader which it believes do not value a globalized fundamental understanding of freedom and values is quite a different perspective in this ongoing debate on ethical issues in managing social media information and its influence on users when it comes specifically on matters relating to politics. Commentators have indicated that tech companies have often followed a skewed policy on deciding who remains on the platform and who doesn't. The lack of judgment and consistency are seen as flaws in not allowing tech companies to decide on matters relating to free speech on the Internet (York, 2021). It has also been argued that even though there is no legislation that bans social media companies from removing content from their platforms, conservative voices feel that the media holds an inherent bias against them. The author provides an overview that suggests that the violence in the US capital followed by the ban on President Trump's accounts would lead to sweeping changes toward more regulation of social media platforms (Ghosh, 2021). These changes would proliferate to different parts of the world impacting ways in which information societies would function and the ensuing debate between ruling polity and citizens in a democratic setup would play out.

Sovereignty has become a contested concept in cyberspace. Scared and sacrosanct the concept of sovereignty is one of the key factors in the composition of modern nation-states. The diffusion of power in the digital space owing to its transnational nature as well as its fragmented ownership among multiple stakeholders, imagining traditional sovereignty, and the states' hold over information has had a vigorous debate on what it means to have state control on the digital environs of one's citizens. In this context, it has been argued that social media has extended the reach of sovereign nations to control the narratives and their citizens abroad (Lowe, 2021). In this context, digital sovereignty remains another key question for political entities to extend their reach to multinational tech companies controlling social media platforms and their narratives. As the digital space, in general, and social media, in particular, have become a forum and an integral part of political discourses the digital sovereignty concerns and the question of user "data" becomes important to understand. Digital sovereignty is defined as *the ability to have control over your own digital destiny — the data, hardware, and software that you rely on and create*. It has become a concern for many policymakers who feel there is too much control ceded to too few places, too little choice in the tech market, and too much power in the hands of a small number of large tech companies (Fleming, 2021). It has been argued that states have been

pushing for digital sovereignty in cyberspace due to the increasing blurring of lines between state-based traditional sovereignty and its applicability to the digital space. We should not simply equate (digital) sovereignty with the ability to defend liberal and democratic values, as is often done by policy actors in Europe. The authors argue that much more reflection and debate are needed on how sovereign powers can be held democratically accountable with regard to the digital space. It is not sufficient to propose that the power of large digital corporations could be tamed by subjecting them to democratic sovereignty, as has been suggested by many democratic governments worldwide (Julia and Thiel, 2020).

National Security, Balance of Opinion, and Social Media Regulation

National security has been described as the ability of a state to cater to the protection and defense of its citizenry (Osisanya, n.d.). In a nation's basket of national security, several tools find an important place. They are present for a state to use in the best care possible to protect territorial sovereignty and integrity. In the contemporary sense, national security as a concept has long moved beyond the confines of the military and battlefield. It has now expanded to include other forms of interests that are critical for the economic, political, diplomatic, and other allied security interests of the nation-state. One comprehensive definition highlights National security as a dynamic, fluid, and multidirectional concept. It is considered the ultimate tool for the survival of the nation-state. It embodies external security (safeguarding the nation from foreign threats) and internal security (within the state) (Abraham, 2012).

The role of the media in the preservation of freedoms of the citizenry and informed public opinion vs. the states' need to preserve secrets vis-à-vis national security has been a matter of debate in democratic societies. The abundant need to preserve information in the garb of protecting secrets puts the democratic nation-state at odds with its citizenry. In this context, the balance of opinion plays an important role in the preservation of national interest. As one analyst notes the developments of American media and its coverage of national security issues that the press and the state have often found themselves at odds due to the reason of maintaining secrecy which is paramount to national security. In choosing between the

responsibility to protect essential information (concerning national security) and in terms of self-censoring and in terms of being responsible for providing information, the media remains caught in a dilemma (Segal, 1994). The national security state has been a matter of debate since the end of the Second World War with the development of institutions that catered to threats to national security in a unified manner. The reorganization of national security with the passage of the national security act in the United States in 1947 was to serve the following three purposes: (1) it legitimated secrecy and intelligence as a necessary form of government, (2) the reorganization of the independent armed services under the Secretary of Defence, with a Joint Chiefs of Staff system, and (3) to ensure that the domestic economy would make available resources and materials for defense and national security purposes. However, the concept of national security remained open without a definition and was often defined by those military, financial, or bureaucratic elites who often framed their mandates (Raskin, 1976).

It was often found that the national security state was at odds with democratic mandates that it proposed it held up. The framing of the enemy was an important part of making national security an important part of the central debate and the control narratives especially during the Cold War (where threats were external and positions were drawn on a known consensus about the enemy). As Barnet noted (on security) that with alternative opinions the distinction between dissent and disloyalty is often blurred in such a way as to set the limits of "responsible" debate and to discredit ideas that veer too far from the orthodox consensus. Those who seek an alternative perspective in national security debates and dissented from the dominant national security vision were often charged with cowardice, disloyalty, or sinister hidden motives (Barnet, 1985). New security threats to the national security agenda have had a key role in rethinking the role of the state and the approaches toward the discourse of national security. In suggesting a resilience-based approach toward national security in the current diverse environment, nation-states should find an appropriate balance between preventive (security) and reactive (resilience) that corresponds with their particular needs, as well as the values of the society (Fjäder, 2014).

In the digital age, national security along with preceding debates in the subsequent sections of this chapter on digital sovereignty remain the key reasons distinguishing the implementation of the idea of national security in the context of digital spaces. While authoritarian and

democratic states alike use the garb of national security to increase surveillance and censor content on the Internet, the increased surveillance on personal communication through social networking platforms and applications has become a cause of concern. The challenges are two-fold. One challenge corresponds to the need to address security concerns affecting peace and stability in a state externally or internally and concerns safeguarding security. A more perverse thought process also follows the ability to use information in the garb of these purposes to target and harass any forms of dissent and free speech and expression. The latter has often been seen as subjective to the opinions of those who interpret it within the confines of limitations that the constitution provides in democratic setups.

Technological advancements have enabled mass surveillance to take place on citizens in all types of political regimes. Social media surveillance refers to the collection and processing of personal data pulled from digital communication platforms, often through automated technology that allows for real-time aggregation, organization, and analysis of large amounts of metadata and content, broader in scope than spyware, which intercepts communications by targeting specific individuals' devices (Shahbaz and Funk, 2019). Nevertheless, real-time terror threats have also been increasing over time and the use of cyberspace and social media to radicalize and even showcase terror and terror-related incidents to a helpless audience is a real-time challenge. Hence, it has been no surprise that as traditional challenges, such as geopolitical conflict and international security issues and the threat of war, give way to more hybrid forms of conflict and threats to state actors, transnational and globalized threats, such as terrorism, have remained one of the key challenges from several years. The key question democratic society should be asking is the following: Where is the red line between the use of monitoring tools for law enforcement and what constitutes otherwise to the invasion of privacy of individuals? What mandates are acceptable for scooping information and harnessing the personal information space of citizens? Is any form of social monitoring of information acceptable at all? A cost–benefit analysis of the debate would offer a better understanding of the prevailing scenario.

Terror groups, such as Al Qaeda and ISIS, have easily adapted to social media for radicalization and recruitment of individuals and other related processes helping in running their terror enterprise. Recently, horrific attacks have also found their way into online contact where terror

perpetrators have used the platforms to showcase their violent acts. Analysts have noted why social media has become an increasingly important arena for terrorist groups. Terror groups operating around the world have found a haven on the Internet as compared to the use of traditional media channels, such as the use of Al Jazeera by Osama Bin Laden for his communications against the United States and the western world. One analysis points out three reasons why terror groups are effectively able to utilize social media. First, social media channels allow terrorist organizations to be part of the mainstream. Second, social media channels are user-friendly, reliable, and free. Finally, social networking allows terrorists to reach out to their target audiences and virtually "knock on their doors" — in contrast to older models of websites in which terrorists had to wait for visitors to come to them. Social networking sites allow terrorists to use a targeting strategy known as narrowcasting (Weimann, 2015). This observation is quite true when it comes to ISIS which has by far used multiple channels to fulfill its various parts of the terrorism cycle.

One observation notes that to cover its territorial loss, ICT operations carried out by ISIS would swoop the Internet to remove negative narratives about the group from the public eye and recruit new followers by projecting only its self-proclaimed caliphate's successes (Ward, 2018). In the study of the Al Qaeda, it has been noted that the terrorist recruitment process follows a concerted model where from an initial request the engagement is built toward a final commitment to commit to a cause (refer to the model on page 31) (Guadagno *et al*., 2010). While the advent of social media has changed how groups interact with individuals transnationally, the use of social media as opposed to websites has made groups, such as ISIS, stronger. As one observer argues the better option to counter ISIS would be to use social media platforms ISIS uses to advance the states' objectives — to track the terrorist group and its operatives and to identify the at-risk populations ISIS attempts to connect with. It is also considered important to understand why young people join terror organizations by surveillance online in such forums know to be possible recruitment grounds (Blaker, 2015). The use of Twitter and Facebook as propaganda tools and the various methodologies have been well documented in literature which includes the use of videos, chat rooms, sympathetic followers, and posters and encrypted communication (Koerner, 2016; Awan, 2017). Nation-states have gone on to use the very same digital media platforms to identify and carry out digital retaliation against the enemies of the state. In what came to be known as Operation Glowing Symphony, the United States

Cyber Command launched an attack against the media and cyber portals of ISIS to take them down. The operation carried out was able to dismantle the ISIL's media and propaganda portals globally. The operation according to declassified documents set the precedent for operations in the future on this front globally (USCYBERCOM, 2016).

While the increasing attacks and counterattacks in the terrorism landscape have dominated the discourse, the relaxed protocols after the 11 September 2001 attacks. The secret program relating to mass surveillance carried out by the United States exposed a government-run surveillance program that monitored the communications records of not just criminals or potential terrorists but law-abiding citizens as well. While technology-induced automated tolls detect and thwart hate speech and crimes, it has not been able to stop either terrorism or its increasing capacity to inflict damage. The debate regarding exposure and surveillance has been two-fold. Those in favor of the governmental activities argue that the biggest concern was that foreign individuals or groups targeted for surveillance had now switched to more secure communication methods (Gjeltan, 2013). Some research also argues that society as a whole due to the rising specter of terrorism has seen acceptance of surveillance technology to be adopted and has noted that such technology interventions will only continue to increase (Haggerty and Gazso, 2005). A study conducted in 2010 argues that despite the advances in technology, it was fairly easy for terrorists to circumnavigate tech infrastructure placed to avoid such a technology in the first place. The study argues for rather following a more nuanced approach of relying on covert and deceptive human intelligence techniques to tap into terror networks rather than revealing technological tools which would give away the secrets and strategies (Maras, 2010). In hindsight, a study conducted in 2018 on the effectiveness of surveillance technology as described by intelligence officials highlights that it is extremely difficult, if not impossible, to evaluate the effectiveness of surveillance programs and despite the high number of attacks thwarted being used as an example counts of successful cases should not be a measure of effectiveness (Cayford and Pieters, 2018).

More recently, human rights advocates have argued against surveillance and the use of technology and surveillance. While it needs to acknowledge that measures are required to understand and thwart terror threats and protect the national security of a nation-state, critics have argued that choosing between targeted surveillance measures aimed at tracking potential terrorists based on reasonable suspicion, and mass

surveillance will make all of us potential suspects (Muižnieks, 2016). This point also proves the argument that surveillance technology and snooping as an intelligence-gathering feature have a level of acceptance and are here to stay. However, there needs to be a fine line that should not be drawn but respected about straying from the objective of countering terror and gathering intelligence deemed productive for national security and toward channeling and the use of information for political purposes by ruling regimes.

Closed Regimes and the Use of "National Security" as a Factor in Preserving Regime Legitimacy and Suppressing Dissent

As argued elsewhere in this chapter, there has been crucial debate in the international community on the democratizing power of social media, especially in authoritarian regimes and closed political systems. Closed political systems have often through to be characterized by a growing authoritarian control over information and communication technologies. The media at best is mostly absent and those present often reflect the worldview of the ruling elite. While in democratic setups media gatekeepers often controlled the narrative, there has been a scope for debate and reform over a point of time considering a sensitivity toward the opinions regarding the public discourse on politics. It is unfair for analysis to just say that the Internet and social media tools (which has now been proved through various arguments throughout this chapter) have only stuck to changing the outlook and affecting the discourse between politics and citizens in authoritarian setups, they have equally changed perspectives in democratic setups as well in discourses about power and politics. Nevertheless, despite the presence of the Internet and ICT technologies as well as social media around the world for some time, many states still retain their authoritarian character. There has also been a growing tendency where democratic states have also started to use freedom of expression on social media as a way of stifling any opposition to their discourse on power. The mobilization power of the Internet has not been doubted and has been agreed upon in most discourses on social media and politics (and other domains) (Bacallao-Pino, 2014; Clay, 2011; Cardoso *et al*., 2016; McKeon and Gitomer, 2019). Hence, despite this dynamic, many authoritarian states remain with sources of information communication

technology, such as the Internet. They do ban but create their ecosystem and their own rules.

One of the key effects of globalization has been the need for nation-states with differing political regimes to remain crucially linked to the global system by various means, such as trade and commerce, communication technology, as well as other forms of cooperation, such as participation in international organizations and global diplomacy. In analyzing the Internet social media as instruments of democratization or instruments of control, the author highlights that social media has a dyadic nature and can also act as vehicles of democracy and openness and at the same time as tools of scrutiny and control for states hence making it hard for providing lack of a definitive answer for the role in democratization processes (Kyriakopoulou, 2011): (1) closed regimes, for the interest of this readership, encompass multiple forms of non-democratic governmental forms which fundamentally exclude open political participation and election of the ruling disposition, (2) fundamental rights and freedom are absent or present only to a select few who favor the political dispensation, and (3) absolute control on power and delivery of justice, control of over the economic affairs, media, and information communication systems.

In the context of the above classification made by this author, several countries around the world would certainly fall under this category. Moreover, many authoritarian states provide varying degrees of access, but they do not mirror the reality as they create their ecosystems with their own rule with limited access and exposure to the global setting. In the current geopolitical context, it has become necessary to consider the role of an alternative brand of Internet governance which China has been propagating around the world. It has been suitably rooted in nationalism but designed to retain authoritarian power rooted in the hands of the communist part. China's access to the Internet has followed a carefully calibrated strategy of opening up the Internet and society to the outside world but maintaining a strong brand of authoritarian control and censorship by the creation of its ecosystem to control information flows. There are about two million censors (*BBC World News*, 2013) watching the Internet which filters words and censors' information deemed threatening to the political regime. Hence, the echo chamber that is created by this fusion of technology's careful control provides a false sense of openness which is seen as acceptable for the political system when compared to complete closure and control. The government allows the Chinese people to say whatever they like about the state, its leaders, or their policies because talk about

any subject unconnected to collective action is not censored (King *et al*., 2014).

Analysts further note that there are three fundamental ways in which critics are shut down: (1) use of fear — threats of punishment or suppression, (2) flooding — drown posts and messages with pro-government messages and propaganda, (3) friction — cost in time or money-making message slow and not work (Roberts, 2018). The strategy of the CCP and its ability to manage its ecosystem and at the same time maintain political control over its Internet have been done in two ways. While the great Internet firewall of China has been one method in which the Internet is controlled and monitoring is done, it is effective but not foolproof. Access through VPNs to the outside Internet or a backdoor across the wall has been quite possible and this too has to do much with selective and globalizing factors and the nature of the Internet. For instance, selective applications, such as LinkedIn, are accessible in China while popular social networking websites are banned. Though most of the global communication social media applications are not part of the ecosystem, their inability to access corresponds to political ambitions as well as techno nationalism.

The Chinese alternatives, such as WeChat, Weibo, and others, provide not only a sense of understanding the trends in social media communications from the outside world but also authorities with the controlling ability to plug any possible control to outside parties which it may see to disrupt politics in the country. It has been argued that the CCP's method in controlling the Internet is diversified and ranges from various tasks, such as not only censorship and blocking but also planning reporting strategies and providing accepted perspectives about the country. The country has been transitioning from banning as much unfavorable information as possible to what the officials call "dredging and blocking" (shudu jiehe) or a combination of guiding public opinion and banning news reports (Tai, 2014). Independent studies with regard to the Internet ecosystem in China and the control of information note that public opinion on the Internet is largely sustained by an extensive network of Cyberspace Affairs Commissions, Public Security Bureaus, and increasingly, content reviewers employed directly by social media platforms. Estimated nationwide spending on Internet censorship is $6.6 billion (Ryan, 2021).

The gray areas in Internet regulation in even countries, such as China, have faced stiff opposition and a challenge for the regime when it comes to political and geopolitical considerations. While censorship as a whole

stands out as an accepted concept with its occasional breakthroughs on expected commemorations, such as "Tiananmen Square episode remembrance," which remembers the 1989 crackdown by the CCP on pro-democracy protesters among other words (Hartman, 2020), Hong Kong has until recently remained the last outpost in China which enjoyed distinctive freedoms quite different from the mainland due to the "one country two systems" policy which was adopted with the British handover of Hong Kong which guaranteed these rights to the people of the territory. Social media proved to be a battleground for protestors and as one analysis notes, the use and deployment of social media became an important facet of understanding the protests in Hong Kong against China's imposition of more authority in the territory to firmly bring it under its control. One key aspect which analysts noted was that social media was used in 2014 to organize people and build support, whereas in 2019, the move was more toward the use of encrypted channels, such as Telegram, as well as local forums, such as LIHKG, an online forum similar to Reddit for more localized organizations. Instagram also served as a platform for protesters to share "visually compelling campaign posters, slogans, as well as image/video evidence of police violence." The 2019 campaign saw both the Hong Kong protesters and the state use the platform in a campaign to "win hearts and minds" (Shao, 2019).

The new national security law in China has been used as a premise to curtail free speech and dissent and precisely so in Hong Kong because for long it has remained contrary to the changes and the social setup shaping out in mainland China. In many ways, it contradicted societal development in mainland China. The "National Security law" in China is just another example of how national security is often used as a premise by states to crack down on dissent and also spread their tentacles of suppression around the world. Though China is now an economic juggernaut that cannot be sanctioned or excluded from the calculus of world trade, it is important to note that the spread of digital authoritarianism is a big challenge where technological trade and implications of dealing with countries remain a key challenge for enabling processes of freer information flows. Digital authoritarianism, also known as techno-authoritarianism, is the way that many leaders around the world wield the power of the Internet and technology to gain or solidify control over their people. The objective is also to use leaders' technology to strengthen their ruling power and attain growing influence around the world (Thacker, 2020).

Authoritarian governments not only control but also shape the behavior of their citizens via surveillance, repression, manipulation, censorship, and the provision of services to retain and expand political control. While arguing that systems and models have been used in suppressing popular dissent in Hong Kong and equally in terms of technology with regard to COVID-19 detection, the author has argued that the ensuing diplomacy that has been carried out has led to an export of this digital authoritarianism model abroad leading to a challenge for future governance and debates with regard to cyberspace and technologies, such as artificial intelligence (Khalil, 2020). While many reports argue that digital authoritarianism is a trend where non-democratic regimes, such as China play a major part, it also highlights that the challenge in countering the trends of technology proliferation solely does not rest only on China but also rests on democratic regimes which also supply such technology (Feldstein, 2020; Polyakova and Meserole, 2019).

While discussions and debates on the role of social media, democratization, and authoritarian regimes have been discussed throughout this chapter, the interplay between various concepts, such as nationalism, popular sentiment, and patriotism triumph the logic of having a zero-sum debate either in favor of people or government. The use of popular sentiment to target a hostile and geopolitically adversarial foe through concerted campaigns enforces the states' position on its citizenry. Some examples of popular protests triggered in part due to the effective use of social media have satisfied both masking the geopolitical inability to counter the foe as well as satisfy limited public apathy and as a result loss of face for the ruling elite. Nevertheless, the social media army which is now a common feature in most types of regimes is either overtly or covertly supported by the political establishments in these countries. While it acts as a positive force in distressing times to enforce political legitimacy (while diplomacy builds its slow pace to build peace), in other times, it is a genie out of the bottle which could challenge the same state apparatus for its shortcomings in providing.

Cyberwarfare and the use of trolls and cyberattacks (in the form of DDoS attacks, hacking, and malware and spyware attacks) have now become a form of indirect warfare and asymmetric conflict in several cases as observed from disputes around the world. The cyber world and the increasing connectivity of all major systems including critical infrastructure systems make a conventionally strong country vulnerable to asymmetric cyberattacks. The South China Sea dispute between China

and its maritime neighbors in the South China Sea has seen increased manifold with the increase in China's claims and assertions which have been invalidated even by the International Court of Justice (ICJ) which ruled that its claims were illegal according to international maritime law. While China has continued to use various methods to enforce its claims vis-à-vis its smaller neighbors, the use of its cyber army has been one of the methods. Similar retaliation by other countries has led to increasing deployment of cyber capacity as an acceptable dimension of operations in even conventional conflicts Manantan, 2020).

While "patriotic retaliatory responses" served a purpose in engaging and responding to coercion, it also meant that in countries such as Vietnam which publically acknowledges having a cyber army to "patrol" the Internet, the key tenet of the operations of Force 47 should be appraised in a broader frame. A closer look at how Vietnam's public opinion shapers and cyber-troops have operated offers a glimpse into the unit's goal: manipulate online discourse to enforce the Communist Party's line in a country whose leaders have been fixated on curbing anti-state content (An Luong, 2021). New regulations to Vietnam's Internet law highlight the challenges that the state faces from the Internet and particularly social media websites, such as Facebook, where concerted efforts are made by the party to shape opinions by creating pro-governmental pages and also content. Multiple riots which have happened in Vietnam against China in 2014 and 2019 have highlighted that those large mobilizations were possible with many inflammatory messages being circulated on social media. While the regime may derive a geopolitical benefit, the choice of these attacks backfiring on the government would play a key role in maintaining stability in the country. The critics are opposed to the new legislation which will enforce cyber-surveillance and have prostate propaganda leading to more stifling of critical anti-regime voices in the country. The use of negative public sentiment and social media manipulations and disinformation campaigns is a double-edged sword that could eventually backfire on the very state which has been using it for scoring geopolitical brownie points in its quest to wage asymmetric conflict.

Conclusion

Political thoughts, ideologies, and association between multiple societal stakeholders form the basis of political identifications and address the pivotal

question of what makes a balanced and strong society and what are the codes of conduct governing it. In any type of regime, people's opinions and thoughts have and will continue to matter to the ruling/governing elite despite its continued efforts to either ignore or sidestep the opinion of its citizenry. The balance of opinion which a citizen can explore in exercising their freedoms under respective political frameworks does not fundamentally change and it is this subjective limitation that citizenry and nation-states should navigate to maintain the sanctity in the societies and political constructs they live in.

Traditionally the media has played an important role in projecting showcasing the views of the society in response to political developments (albeit with its inherent political inclinations and ideological bent). Inherently dependent on a cross-sectional dependence on society and entrepreneurship and society for stories and funding, its limitations in representing actual realities free of political interference or editorial biases became too difficult to avoid. The advent of social media and Internet-based ICT communication networks revolutionized the space of public opinion and expression and provided individuals with the right to not only express their views but also directly broadcast their views to a large audience. This filterless, unstructured, and unscripted mass messaging has shaken the foundations of democratic societies and revolutionized the ways in which citizens and the sovereign see each other in political settings. The social power of social media has promoted transparency, accountability, and a culture of non-elite debate that is crucial in representing a non-gatekeeping view of on-ground realities.

Political regimes either democratic or authoritarian (in all its form) have understood that closing this system would have stronger repercussions for their image and also their stability. They continue to fight against it in many forms by enforcing restrictions, censoring data, and deploying technologies that do not have proper data use protections or intrude into personal space with the help of advancing scientific technology, such as Artificial Intelligence. While at certain stages a justification of national security protection is used to deploy these technologies, the impact on the collective psyche of the society at large is affected over time by the restrictive environment and measures. Nevertheless, the use of social media as a tool of political communication as well as influence in enforcing political legitimacy in both regime types is now an established practice. A practice that works effectively. In many democracies around the world, the use of social media to shape political opinion is an important process of the campaigning strategy which political parties employ.

However, it is nothing short of social engineering to present a view and a counterview about the topic at hand. In this sense, social media and politics are somewhere complimentary as they seem to use rather similar models of enforcing and drawing attention and interest to keep participants enamored by showing them what they like as opposed to what they should be looking out for in terms of deciding. As social media companies use social engineers to understand the psyche of the user to get them hooked on to the network so that they can earn advertisement revenues, politics very well has understood that the strategy and the key to a more impressionable effort are to use and control the narrative and speech online which are the key to winning in the digital era.

The debate on the balance of opinion in authoritarian regimes is also open to interpretation. As discussed elsewhere in this chapter, authoritarian influences on social media are as much a reality as is the debate on the democratization power which the medium produces. Authoritarian regimes find multiple ways to reason with their citizens' use of social media to dissent and show negativity toward their regime. While some try to censor and deny access, some find unique ways in which they control narratives and disrupt threatening activities. However, the most acceptable part which most scholars would tend to agree is that closed regimes which had absolute control over the traditional media find social media as a litmus test of their legitimacy and battle hard to either set the narrative or silence it. Political conditioning of citizens is the same everywhere; they yearn to accept freedoms in return for an acceptable amount of fair governance by their ruling elite. In most authoritarian regimes, the people and their opinions about the ruling elite or a supreme leader have been the most important part of shaping the narrative. What they think and how they opine about it has been important for the ruling class. It is within this framework that most authoritarian states derive their fear, i.e., internal subversion and dissent. The "insider threat" (to borrow a term from cyber terminologies) becomes the biggest challenge. Social media is seen as an anonymous tool: the ability to spread information fast and mobilize even faster. This inherently challenges authoritarian systems which due to the lack of a popular political mandate always remain fearful of loss of power.

The only possible convergence between various regime types on why states conduct surveillance on their citizens could be the premise of national security. However, we must be careful in generalizing this common aspiration to control citizens and monitor them. National security in a democratic setup emerges as seen in the preceding arguments in this

chapter out of the need to protect their citizens, sovereignty, and other constructs of the modern nation-state. While this is legitimized by a narrative of threats and the need to act, it fulfills the obligation of the sovereign in fulfilling the social contact which is rightfully owed by the state to its citizens. While democratic politics and political elites have benefited from war and the presence of a visible enemy, there could be possible safeguards and more accountability for these processes. The loss of power in the democratic due to the process of elections remains a stark reality. On the other hand, authoritarian states also long for legitimacy even more so than democratic states due to their inherent fears of a lack of legitimacy and societal and political organization subversion. Creating an enemy and using national security sugar-coated with nationalism and other such constructs help the enforcement of legitimacy. Social media has played a big spoiler in image-building regime sustenance for such states. The very ability to have narratives and counternarratives completely decentralized from state authority is probably a nightmare for regulators in closed regimes even if they severely control or restrict access to information. Social media and the Internet can bring events to the fore through targeted settings even if people do not go looking for them.

Information today is a prime piece of currency and tool to seek legitimacy from politics. As its importance grows, political players will continue to court it but at the same time restrict it to uses that would often protect their turf. Nation-states as always will find ways to tell their citizens to buy their narrative about security, that is, however, the job of politics. People are often tasked with an important job in the information age. The task for the citizenry is to construct a reality about their position in seeking legitimacy for handing power to a political elite who have agreed to protect their rights. Personal freedom and national security interest directed by the states would have to find a balance in allowing opinions to flow. It is often like a safety valve letting pressure flow out. The more states resist, the more there would be a tendency to subvert. In essence, if the nation-state has to survive the digital age, it has to rethink and reconstruct its Westphalian principles to better suit it for the modern age.

References

Abraham, R. (2012). *Media and National Security*. New Delhi: K. W. Publishers.

An Luong, D. N. (2021). How the Vietnamese state uses cyber troops to shape online discourse. *ISEAS Perspectives*, 3 March.

Andrew, M. (2021). Trump remains banned, for now, but the problem with Facebook is still Facebook. *The New Yorker*, 5 May.

Awan, I. (2017). Cyber-extremism: ISIS and the power of social media. *Social Science and Public Policy*, 54, 138–148.

Bacallao-Pino, L. M. (2014). Social media mobilisations: Articulating participatory processes or visibilizing dissent? *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 8(3).

Barnet, R. J. (1985). The ideology of the national security state. *The Massachusetts Review*, 26(4), 483–500.

BBC World News (2013). *China Employs Two Million Microblog Monitors State Media Say*. London: BBC.

Blaker, L. (2015). The Islamic State's use of online social media. *Military Cyber Affairs*, 1(1), 1–10.

Bloomberg (2015). *How Much of an Impact is Social Media Having on Politics.* s.l. New York: Bloomberg Quicktake.

Bonney, V. (2018). *How Social Media is Shaping Our Political Future*. Maine: TEDx Talks.

Cardoso, G., Lapa, T. and Fátima, B. D. (2016). People are the message? Social mobilization and social media in Brazil. *International Journal of Communication,* 10, 3909–3930.

Cayford, M. and Pieters, W. (2018). The effectiveness of surveillance technology: What intelligence officials are saying. *The Information Society*, 34(2), 88–103.

Chadwick, A., Dennis, J. and Smith, A. P. (2016). Politics in the age of hybrid media: Power, systems, and media logics. In Bruns, A., Enli, G., Skogerbø, E., Larsson, A. O. and Christensen, C. (Eds.), *The Routledge Companion to Social Media and Politics*, pp. 8, 42. New York: Taylor & Francis.

Clay, S. (2011). The political power of social media: Technology, the public sphere, and political change. *Foreign Affairs*, 90(1), 28–41.

Crabtree, J. (2018). Here's how Cambridge Analytica played a dominant role in Kenya's chaotic 2017 elections. https://www.cnbc.com/2018/03/23/cambridge-analytica-and-its-role-in-kenya-2017-elections.html. Accessed on 2 September 2021.

Dube, R. (2021). Why ethics matter for social media, silicon valley and every tech industry leader. https://www.forbes.com/sites/robdube/2021/01/14/why-ethics-matter-for-social-media-silicon-valley-and-every-tech-industry-leader/?sh=671a176b16f2. Accessed on 3 September 2021.

Dwoskin, E. (2021). *Misinformation on Facebook Got Six Times More Clicks Than Factual News During the 2020 Election, Study Says*. Washington: The Washington Post.

Feldstein, S. (2020). When it comes to digital authoritarianism, China is a challenge — But not the only challenge. https://warontherocks.com/2020/02/

when-it-comes-to-digital-authoritarianism-china-is-a-challenge-but-not-the-only-challenge/. Accessed on 2 September 2021.

Fjäder, C. (2014). The nation-state, national security and resilience in the age of globalisation. *Resilience*, 2(2), 114–129.

Fleming, S. (2021). What is digital sovereignty and why is Europe so interested in it?. https://www.weforum.org/agenda/2021/03/europe-digital-sovereignty/. Accessed on 3 September 2021.

Ghosh, D. (2021). Are we entering a new era of social media regulation? *Harvard Business Review*. 14 January. https://hbr.org/2021/01/are-we-entering-a-new-era-of-social-media-regulation.

Gjelten, T. (2013). The effects of the Snowden leaks aren't what he intended. https://www.npr.org/2013/09/20/224423159/the-effects-of-the-snowden-leaks-arent-what-he-intended. Accessed on 3 September 2021.

Guadagno, R. E. *et al*. (2010). Social influence in the online recruitment of terrorists and terrorist sympathizers: Implications for social psychology research. *Presses Universitaires de Grenoble*, 23(1), 25–56.

Haggerty, K. D. and Gazso, A. (2005). Seeing beyond the ruins: Surveillance as a response to terrorist threats. *The Canadian Journal of Sociology/Cahiers canadiens de sociologie,* 30(2), 169–187.

Hartman, L. (2020). *In China, You Can't Say These Words*. Washington: ShareAmerica.

Julia, P. and Thiel, T. (2020). Digital sovereignty. *Internet Policy Review*, 9(4). https://doi.org/10.14763/2020.4.1532.

Karan, K. and Nimmo, B. (2019). BJP and congress resorting to running deceptive social-media accounts and disinformation campaigns. *First Post*. 2 April.

Khalil, L. (2020). *Digital Authoritarianism, China and COVID*. Sydney: The Lowy Institute.

King, G., Panand, J. and Roberts, M. E. (2014). Reverse-engineering censorship in China: Randomized experimentation and participant observation. *Science*, 345(6199), 1–10.

Koerner, B. I. (2016). *Why ISIS is Winning the Social Media War*. s.l.: Wired.

Kyriakopoulou, K. (2011). Authoritarian states and internet social media: Instruments of democratisation or instruments of control? *Human Affairs*, 21(1), 18–26.

Laub, Z. (2019). Hate speech on social media: Global comparisons. https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons. Accessed on 2 September 2021.

Lowe, M. (2021). Social media is blurring the lines of national sovereignty. https://thediplomat.com/2021/04/social-media-is-blurring-the-lines-of-national-sovereignty/. Accessed on 3 September 2021.

Maguire, M. (2017). Embedding journalists shape Iraq news story. *Newspaper Research Journal,* 38(1), 8–18.

Manantan, M. B. (2020). The People's Republic of China's cyber coercion: Taiwan, Hong Kong, and the South China Sea. *Issues & Studies: A Social Science Quarterly on China, Taiwan, and East Asian Affairs*, 56(3), 2040013.

Maras, M. H. (2010). How to catch a terrorist: Is mass surveillance the answer? *Journal of Applied Security Research*, 5(1), 20–41.

McKeon, R. T. and Gitomer, D. H. (2019). Social media, political mobilization, and high-stakes testing. *Frontiers in Education*, 4. DOI: 10.3389/feduc.2019.00055.

Muižnieks, N. (2016). Human rights in Europe should not buckle under mass surveillance. *Open Democracy*. 12 February.

O'Donnell, C. (2011). New study quantifies use of social media in Arab Spring. https://www.washington.edu/news/2011/09/12/new-study-quantifies-use-of-social-media-in-arab-spring/. Accessed on 3 September 2021.

Osisanya, S. (n.d.). National security versus global security. https://www.un.org/en/chronicle/article/national-security-versus-global-security. Accessed on 30 August 2021.

Pew Research (2012). *Arab-American Media: Bringing News to a Diverse Community*. Washington: Pew Research Centre.

Polyakova, A. and Meserole, C. (2019). *Exporting Digital Authoritarianism*. Washington: Brookings Institution.

Raskin, M. G. (1976). Democracy versus the national security state. *Law and Contemporary Problems*, 40(3), 189–220.

Ryan, F. (2021). *Buying Silence: The Price of Internet Censorship in China*. s.l.: The Jamestown Foundation.

Segal, D. R. (1994). *National Security and Democracy in the United States*. New York: Sage.

Shahbaz, A. and Funk, A. (2019). *Social Media Surveillance*. Washington: Freedom House.

Shao, G. (2019). *Social Media Has Become a Battleground in Hong Kong's Protests*. Hong Kong: CNBC.

Smidi, A. and Shahin, S. (2017). Social media and social mobilisation in the middle east: A survey of research on the Arab Spring. *India Quarterly*, 73(2), 196–209.

Tai, Q. (2014). China's media censorship: A dynamic and diversified regime. *Journal of East Asian Studies,* 14(2), 185–209.

Thacker, J. (2020). *What is Digital Authoritarianism?* s.l.: ERLC.

The Times of India (2021). Feud between Twitter and government intensifies: A look at 6 flashpoints. https://timesofindia.indiatimes.com/india/govt-vs-twitter-feud-intensifies-key-developments/articleshow/83261663.cms. Accessed on 3 September 2021.

USCYBERCOM (2016). *USCYBERCOM, USCYBERCOM 30-Day Assessment of Operation Glowing Symphony, December 13 2016. Top Secret.* Washington: National Security Archive — George Washington University.

Ward, A. (2018). ISIS's use of social media still poses a threat to stability in the middle east and Africa. https://www.rand.org/blog/2018/12/isiss-use-of-social-media-still-poses-a-threat-to-stability.html. Accessed on 3 September 2021.

Weimann, G. (2015). Terrorist migration to social media. *Georgetown Journal of International Affairs*, 16(1), 180–187.

Wolfsfeld, G., Segev, E. and Sheafer, T. (2013). Social media and the Arab Spring: Politics comes first. *The International Journal of Press/Politics*, 18(2), 115–137.

York, J. C. (2021). Users, not tech executives, should decide what constitutes free speech online. *MIT Technology Review*. 9 January. https://www.technologyreview.com/2021/01/09/1015977/who-decides-free-speech-online/.

https://doi.org/10.1142/9781800612860_0010

Chapter 9

Children, Data Collection, and Privacy — Is the Safeguarding Fallacy a Justification for Excessive Regulation and an Erosion of Human Rights?

Andy Phippen

Introduction

In a book that considers the regulation of cyber matters, it is worthwhile to consider the rights of the child as a specific category of end user. Children pose particular challenges for providers in the delivery of the services in that there is a need to balance the rights to protection with the rights to privacy. They need to be mindful that online services they offer, particularly those with a social context, provide the opportunity for much positive interaction, but also, as with any social situation, there is a risk of harm from abuse or exposure to upsetting scenes. And due to these concerns, providers are also placed under pressure to ensure children are safe from governments, whose role is to ensure their citizens are safe from harm.

However, as is explored in this chapter, there can sometimes be a focus on harm reduction through technology, which has the (arguably) unforeseen fallout of excessive data collection in order to ensure the child is safe. If we take, for example, a child using a messaging platform to interact with peers, clearly there is potential for harm or abuse from either

peers or those claiming to be peers. A potential solution to this would be for an adult with caring responsibilities to also see the messages being passed between the child and their peers. That way, they can see if abuse is occurring and intervene where necessary. However, does that not impact the privacy of the child? Should adults have access to that much information about the child, and is informed consent possible in a parent/child relationship? There is a risk, in the rush to ensure children can engage with online services in a positive, harm-free, manner, that the "solutions" for their safety actually far exceed any reasonable data protection rights the child might have.

In this chapter, we explore this tension from the perspective of emerging UK legislation and the burgeoning SafetyTech industry and argue that there is a risk under the guise of safeguarding the child's right to privacy is eroded to unacceptable levels.

Safeguarding as Regulation

At the time of writing, the UK Government has just released the draft of the much anticipated Draft Online Safety Bill (2021), with the claim it will introduce the following (UK Government, 2021):

> *Landmark laws to keep children safe, stop racial hate, and protect democracy online published.*

The draft bill will now be debated in the Houses of Parliament and while not the heart of this chapter, which explores the challenges around children's privacy, data collection, and rights, it is a useful starting point to consider the regulation of digital technology to develop a more "ethical" sector. This is a widely debated area where the safeguarding of children and mitigation of risk with them being online are sometimes used to introduce regulation that, on closer inspection, might both fail to keep children "safe" and also provide a disproportionate level of power to the regulators of the technology sector. It is, therefore, worthy of exploration in the context of this chapter.

The bill, in the main, focuses on defining both illegal and "legal but harmful" activities online and specifies a nebulous "duty of care" that providers should be able to demonstrate should a user of their services be subject to online harm. There is much to unpick with this draft legislation,

but throughout the document, there is an implication that platforms should be able to implement technical approaches to prevent harm and to demonstrate "duty of care" through transparent risk assessment. It is telling that, while the bill itself is launched with a strong steer toward children's protection, a great deal of the focus lies in the regulation of technology companies and the notion that harm is done *to* an individual and can, therefore, be controlled by the platform provider. While children seem to be the justification for the legislation, its impact will be far wider than those who provide services that *might* be used by children and young people.

As an exercise on how Governments attempt to regulate the cyberspace, the Online Harms Bill is a useful starting point for our analysis. The topic of online safeguarding is broad and incorporates many disciplines and perspectives, and the UK government's attempts to regulate the technology sector in this regard can be traced back almost 10 years (Phippen, 2016). Over this time period, we have seen demands for Internet Service Providers to filter Internet access and prevent children from accessing pornography (and other "inappropriate" content), to prevent access to "inappropriate" content on public WiFi by providers, for adult sites to provide age verification if the user in is in the UK, and, in a stark example of history repeating itself, calls for Facebook not to implement end-to-end encryption into its messenger platform because pedophiles might use it to avoid detection in the exchange of indecent images of children, the subject of a recent open letter by the UK, the US, and Australia to the Facebook CEO Mark Zuckerberg (UK Government, 2019a).

In this chapter, I explore the regulation of technology and how it relates to children and young people and, more specifically whether, using a safeguarding justification, children and young people have their rights eroded as fallout from excessing data collection, monitoring, and tracking, all put in place to ensure they are safe from the risks associated with their presence online. This exploration analyzes policy developments, and legislative changes, while also drawing on a wealth of empirical data drawn from many years of work with young people and stakeholders for young people's safeguarding in the UK (for example, see Phippen, 2016).

While the UK will be the focus of this examination, due to both familiarity with legislative frameworks, policy developments and having a significant body of empirical data that have been collected over many years, there is nothing unique about the UK position. The issues attempting to be tackled by the UK government — such as how to keep children

safe online, how to control illegal activities, and how to ensure personal data are only processed if and when it is needed by a third party — are global issues, one governments must attempt to tackle within their own geographical boundaries. This has been a perpetual struggle for governments, as highlighted in John Perry Barlow's famous "Declaration of Independence for Cyberspace" (Barlow, 1996), a much-cited document that claimed governments would always fail to regulate and control the online world. The declaration was written on the day the US Telecommunications Act (1996) came into force.

> *Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather.*
>
> *We have no elected government, nor are we likely to have one, so I address you with no greater authority than that with which liberty itself always speaks. I declare the global social space we are building to be naturally independent of the tyrannies you seek to impose on us. You have no moral right to rule us nor do you possess any methods of enforcement we have true reason to fear.*
>
> *Governments derive their just powers from the consent of the governed. You have neither solicited nor received ours. We did not invite you. You do not know us, nor do you know our world. Cyberspace does not lie within your borders. Do not think that you can build it, as though it were a public construction project. You cannot. It is an act of nature and it grows itself through our collective actions.*

More specifically, Barlow was opposed to Title V of the Act, the Communications Decency Act, which sought to impose criminal liability for obscene or indecent transmitted over the Internet under certain circumstances. While most of this Act has been struck down on free speech grounds, it is of interest to note that the now-famous "section 230."

> *No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider*

remains, has been a persistent shield from liability by service providers, and is the subject of much current debate. Whether the Online Safety Bill

breaches these defenses remains to be seen. Nevertheless, the issue governments face in trying to regular behavior on a global communications platform has a rich history from which they fail to learn.

A fundamental argument presented is that governments will generally assume that because harm occurs on technologically facilitated platforms, technology can also be put in place to prevent harm from occurring. And failing to appreciate that technological intervention may have issues around privacy and the right to access information.

For example, in recent years, we have had a number of calls, such as the following:

- The Health secretary calling for algorithms to be installed on children's mobile phones to detect indecent images and prevent them from being sent (House of Commons Science and Technology Committee, 2017).
- Legislation to impose age verification technology on anyone wishing to access pornography from a UK-based device (Digital Economy Act, 2017).
- Calls to extend age verification onto social media sites to ensure no one under 13 can access these services and for social media companies to ensure children cannot access their services for more than two hours per day (Helmand and Rawnsley, 2018).
- Calls for social media companies to stop the live streaming of terrorist activities (*BBC News*, 2019).
- Calls for social media companies to prevent the posting of "anti-vax" materials (Mohdin, 2019).

A famous cybersecurity researcher, Marcus Ranum, once stated the following:

You can't solve social problems with software.

And "Ranum's law" (Cheswick, 2003) is frequently referred to by scholars of technology regulation. I find myself drawing upon Ranum's law myself when discussing both cyberlaw and safeguarding policy directions with many stakeholders. Yet, it would seem from even the most cursory glance at technology policy that there is a prevailing view that technology is the problem and, therefore, should be the solution.

If we consider the perennial issues of backdoors in encrypted communication, once again raised following Facebook's announcement about

encrypting its Messenger platform, there is clearly an argument that having backdoors in encryption allows criminals to have their communications and activities monitored and intercepted by law enforcement. However, it would be a brave or misguided government who would claim that as a result of criminals using certainty technologies, we should remove European Conventions of Human Rights Article 8 protections in order to stop criminals from taking advantage. It would be unlikely that citizens would view their communications being monitored as an acceptable fallout from the need to intercept criminal conversations online. However, as we have seen above, if governments argue regulation from a position of "protecting the most vulnerable in society," the argument becomes more compelling.

For a long time, the technology sector has talked of the "Four Horsemen of the Information Apocalypse" or Infocalypse (Schneier, 2005):

> *Beware the Four Horsemen of the Information Apocalypse: terrorists, drug dealers, kidnappers, and child pornographers. Seems like you can scare any public into allowing the government to do anything with those four.*

While the exact nature of the Four Horsemen varies in the telling of the tale (in some versions, the horsemen are organized crime, terrorists, drug dealers, and pedophiles), the observation remains the same: in order to win over public opinion about the regulation of specific aspects of technology, it is necessary to show them how one or more of the Horsemen make use of the technology. We see it once more in calls to prevent Facebook from using encryption in its Messenger platform and also the wider-ranging powers proposed within the Online Safety Bill. "We need to do this, in order to keep children safe. And you want children to be safe don't you?"

To quote from the open letter to Facebook, the authors made it clear that if Facebook were to move to end-to-end encryption, they would be helping various criminal activities:

> You stated that "we have a responsibility to work with law enforcement and to help prevent" the use of Facebook for things like **child sexual exploitation, terrorism, and extortion**.
>
> ....
>
> *Companies should not deliberately design their systems to preclude any form of access to content, even for preventing or investigating the most*

serious crimes. This puts our citizens and societies at risk by severely eroding a company's ability to detect and respond to illegal content and activity, such as child sexual exploitation and abuse, terrorism, and foreign adversaries' attempts to undermine democratic values and institutions, preventing the prosecution of offenders and safeguarding of victims.

While a political message such as "we want to undermine encryption because we want to snoop on your communications" might not be a palatable one, saying instead "paedophiles use this technology to hide their activities and we cannot do anything about it, do you think it's a good idea to ban it?" one is far more likely to win over public opinion.

However, the failing of this argument is that it considers end-to-end encryption as a new technology being applied for the first time. However, this has been a debate that has raged since the 1970s. Moreover, at the time of writing, many messaging apps, all freely available, implement end-to-end encryption, for example, Apple iMessage, WhatsApp, Viber, Telegram, and Signal. Facebook is not introducing dangerous new technology, it is catching up with the rest of the field.

The Safeguarding Fallacy

It would seem, however, that safeguarding trumps everything — if there is a safeguarding concern that might impact "the most vulnerable in society," disproportionate measures are acceptable. This is a theme that I will return to often in this discussion — the view that rights should be eroded in order to keep things "safe." The history of "exceptional access" to end-to-end encryption is a useful illustration of this fact — while few policymakers would openly say that citizens do not have a right to privacy, they might suggest that those rights are not absolute if measures to implement privacy provide opportunities for criminal activity, particularly of those measures that are technical in nature.

Until recently, the UK government made great use of the phrase "if it's unacceptable offline it should be unacceptable online" (UK Government, 2018). However, an offline comparison of this particular concern might be an expectation for a café owner to place the means for law enforcement to listen in on conversations taking place in their establishment — if the café is potentially a place where a private

conversation could include criminal collusion, surely the privacy rights of the participants should be eroded such that exceptional access can take place, perhaps via microphones placed under café tables? When this offline comparison is presented, it sounds ridiculous. However, it would seem that when communication takes place online, the argument is viewed as more acceptable.

As far back as 1999, the science fiction author Douglas Adams (1999) famously wrote an essay that made observation on the implied causation of any technology facilitated crime in a manner that would not be leveled at other ambient factors:

> *Newsreaders still feel it is worth a special and rather worrying mention if, for instance, a crime was planned by people "over the Internet." They don't bother to mention when criminals use the telephone or the M4, or discuss their dastardly plans* "*over a cup of tea*," *though each of these was new and controversial in their day.*

Yet, it would seem, with technology regulation and legislation, we fail to learn from history, particularly when there is public opinion to win over. And when it comes to child protection, there are few who would voice an opinion that children should have fewer safeguards online. However, in this rush to use children being safe online as the level with which to justify greater regulation of technology and cyberspace, and the view that technology can provide that protection, it would seem that children are viewed as passive actors in online interactions — they are subjects to which harm is done and, therefore, need protecting rather than active participants in online worlds. The prevailing, safeguarding-centric, perspective is that children have things happen to them by others via technology, and that needs to be prevented. There is less concern, it seems, with what is taken from them by technology or how in our rush to "protect" them we might be impacting the human rights.

If we reframe children and young people not as passive magnets for harm but, instead, view them as full engagement members of digital society, we can see how these prohibitive approaches, and what we might refer to as the *safeguarding fallacy*, are, arguably, as harmful as the actors' emergent regulatory frameworks propose to tackle.

A fundamental aspect of this argument is children's right to privacy. It is not really an issue up for debate — it is well defined in the United Nations Convention on the Rights of the Child (United Nations, 1989),

established in 1989 and ratified by 196 countries. However, it still seems to be poorly adopted and, in a policy space where the safeguarding of children is a useful level to justify regulatory powers, it is often forgotten from this debate. The right to privacy is fundamental and while there are some aspects of legislation that have been developed with this in mind, the view that technology might be the "solution" to online safety also causes much cause for concern. In the rush to implement technology, with pressure from legislators, do we really have the rights of the child at the heart of the debate?

Safety at the Expense of Privacy?

Excessive data collection, with a mind to children's privacy, is the often-forgotten element of online safeguarding, and this is perhaps because it has been done well on both sides of the Atlantic. However, as a safeguarding measure, it is poorly understood.

To draw from a frequent professional experience, I have, over many years of working in this area, visited many schools to both speak to children and also deliver staff training. A common narrative I am met with by staff in primary schools (where the children are aged between 4 and 11) is "We don't talk to them about social media, it's illegal for them to be on it." When I ask why they believe this to be the case, I am generally told it is for safeguarding reasons. It generally comes as something of a surprise when I tell them the reason for the "age 13 or over" is to protect young people from excessive data collection, at an age where it is viewed that they do not have the capacity to consent to these issues. When told that the "illegal" aspect of the online transaction is not the child using the platform but the platform collecting data from the child, there tends to be further surprise and perhaps a deflation of the argument that illegality is an excuse not to deliver education on the subject.

Nevertheless, the foundations of privacy legislation concerning excessive data collection are a positive aspect of the legislative canon. While the US Children's Online Privacy Protection Act (COPPA) (Children's Online Privacy Protection Act, 1998) was the instigator for the widely used but poorly understood "no children on social media until they're 13," the EU's General Data Protection Regulation (GDPR) 2018, and its national implementations (in the UK, the Data Protection Act, 2018) have brought the EU in line with this legislation and afforded

further consideration of children's data collection and how to protect them from exploitation by companies.

This is not to say that the legislation has resulted in companies never breaching these rules. In 2015, the Global Privacy Enforcement Network performed a "sweep" of 1494 commercial websites and apps (Global Privacy Enforcement Network, 2015) that were targeted at children. They discovered the following:

- 60% collected personal data;
- 50% shared data with third parties;
- 22% offered the opportunity to submit phone numbers;
- 23% offered the opportunity to submit photographs.

However, with effective legislation that places the child's right to privacy at the center of developments, we can see that it has teeth and those who do excessively collect data can be met with heavy fines. The recent Federal Trade Commission's (2019) $4.7 million fine to Tik Tok (known at the time of breach as Musical.ly) after uncovering evidence of the collection of sensitive information, such as location data, and then exposing these data on their platforms shows that legislation can be used to both safeguard children and make the technology sector think about unethical practice.

It works because it is both tangible in scope and practical in application. This can clearly be seen in some aspects of the UK's Age Appropriate Design Code (Information Commissioner's Office, 2020) which was released in final form by the UK Data Protection regulator, the Information Commissioner's Office, in 2020.

The code, aimed at "information society services likely to be accessed by children," defines a number of standards expected of those service providers when processing the data of children, that are in line with the UK's Data Protection Act 2018. The practices proposed in the code make a great deal of sense when processing data related to children and young people and bring children's privacy and data rights to the fore. Furthermore, it is a legislative approach that acknowledges young people's right to engage with online services rather than trying to prohibit them from doing so in order to ensure they are "safe."

At the heart of the code is the aforementioned UN Convention on the Rights of the Child (*Ibid.*), more specifically stating article 3 of the Convention — The Best Interests of the Child, clarified by the United Nations (2013) in that Best Interest

> *gives the child the right to have his or her best interests assessed and considered as a primary consideration in all actions or decisions that concern him or her, both in the public and private sphere.*

This seems like a very sensible, and reasonable, foundation upon which service providers may build their services. Within the code, the standard attempts to further clarify in relation to information society services such that they will

- Keep them safe from exploitation risks, including the risks of commercial or sexual exploitation and sexual abuse;
- Protect and support their health and well-being;
- Protect and support their physical, psychological, and emotional development;
- Protect and support their need to develop their own views and identity;
- Protect and support their right to freedom of association and play;
- Support the needs of children with disabilities in line with your obligations under the relevant equality legislation for England, Scotland, Wales, and Northern Ireland;
- Recognize the role of parents in protecting and promoting the best interests of the child and support them in this task; and
- Recognize the evolving capacity of the child to form their own view and give due weight to that view.

To have best interests at the core of the code, and the strong relationship to the convention, lays a strong foundation upon which other standards within the code are defined and provides a sound and tangible basis for which providers might consider to demonstrate due diligence in the processing of young people's information on their services.

The second standard within the code, the need for Data Protection Impact Assessment, reinforces the need for the provider to be able to demonstrate they have carried out a full assessment of the collection and processing of children's data for a given service with full consideration of the risks therein. The company needs to be able to demonstrate a tangible due diligence around children's data collection.

Furthermore, the code makes it clear that children and young people must be considered at different developmental stages and there is no blanket approach for any user between the age of 0 and 18. Again, this is

unusual when considering online safeguarding policy — usually "children and young people" will be referred to as a monolithic entity; this is certainly the case in the Online Safety Bill, which refers to "children" throughout the draft legislation but makes no variation of duty based upon developmental phase.

There are many other parts of the code, which details 15 standards in total, which we are encouraged to see. While, as with any regulatory framework, the detail and efficacy of the code will emerge through use, monitoring, and prosecutions arising from complaints, the code does, in general, provide a strong step forward in protecting children's privacy rights online. This is a significant difference to note — the code aims to protect children's rights rather than assuming they are passive consumers of online services who need isolation.

Within the code, there are also calls for age-appropriate information for young users of services so they can understand how their data are processed in a manner that they understand, that providers should be transparent in what and how they collect data, and that they only collect the minimal data they require in order to be able to provide the service.

Furthermore, expectation on services for young people around ensuring "high privacy" is a default setting (coupled with a challenge to those considering using "nudge" techniques to encourage young people to relinquish these privacy settings), geo-location is switched off, and profiling and data sharing are strongly discouraged are all welcome and seem proportionate in their approach and achievable for service providers. All of this is child centric and cognizant of their rights, and a far more effective and tangible model for the regulation of online services used by children and young people.

Age Verification and Excess Data Processing

However, not everything in this code is perfect, and it once again starts to lose efficacy when assumptions are made about the capability of technology to regulate technology, in contravention of Ranum's Law. There are two aspects of the code that raise significant concern, from both a technical capability and also moral perspective. The first relates to the responsibilities of the provider to apply data processing in an age-appropriate way, that raises the need for providers to adopt an age verification approach or

default to an assumption that all users are treated as children as far as the code is concerned:

> ***3. Age-appropriate application:*** *Take a risk-based approach to recognizing the age of individual users and ensure you effectively apply the standards in this code to child users. Either establish age with a level of certainty that is appropriate to the risks to the rights and freedoms of children that arise from your data processing, or apply the standards in this code to all your users instead.*

Age verification has already played a much-publicized role in a failed attempt to control access to pornography sites in the UK, as defined in Part 3 of the Digital Economy Act 2017 (*ibid*.) that was ultimately withdrawn in 2019 (Department of Culture, Media and Sport, 2019).

One of the fundamental flaws in this legislation lay in the challenge for a citizen to be able to *prove* that they were 18, given that there is no standard way with which to do this in the UK. For a child to be able to verify, for example, that they are 13, would be even more challenging without the need to provide a copy of extremely sensitive personal information, such as a birth certificate. Therefore, it was surprising to see technological approaches proposed as part of the code, particularly with an onus on the service provider to implement it in an "appropriate" manner, given that learning from the failure of section 3 of the DEA would suggest that there is no appropriate manner:

> *you should always use a method that is appropriate to the risks that arise from your data processing. Some of the methods you may wish to consider are listed below.*

And perhaps the turning point in this discussion, which has, so far focused upon the positive, rights based, aspects of the code in preventing excessive data collection and making companies responsible for data breaches, the code also details a number of approaches in a "non-exhaustive" list of techniques, alongside stating that the appropriateness of approach will be considered in the event of an investigation but would be mindful of marketplace and capabilities of service providers. The approaches they detail seem diverse with the following varying degrees of concern:

- **Self-declaration** — Which they state would only be viewed as appropriate for **low-risk processing** or when used with other techniques.
- **Artificial intelligence** — To quote the code: "It *may* be possible to make an estimate of a user's age by using artificial intelligence to analyze the way in which the user interacts with your service."
- **Third-party age verification** services — Such as those proposed during the pornography AV debates which were ultimately not put in place.
- **Account holder confirmation** — Third-party verification by an already verified user.
- **Technical measures** — To discourage false declarations of age.
- **Hard identifiers** — Such as passport, although this approach is discouraged as excessive. Moreover, aside from birth certificate, these hard identifiers are not compulsory in the UK.

Moreover, there is an assumption, as we have raised at the start of this chapter as a fundamental flaw in technology policy, that technology will be able to provide the solution, even if those specifying legislation are not sure whether the technology would work (illustrated with the point around the use of artificial intelligence).

One concern that immediately arises from this list is that a number of approaches (artificial intelligence, third-party AV, account holder confirmation, and hard identifiers) all require a higher level of data processing than that which might be required to provide the service itself. The code seems cognizant of this and states the following:

> *You may be able to collect and record personal data which provides an assurance of age yourself. If so, remember that you need to comply with data protection obligations for your collection and retention of that data, including data minimization, purpose limitation, storage limitation, and security obligations.*

This seems very much at odds with other parts of the code, such as "data minimization" and preventing "data sharing" unless absolutely necessary.

This is perhaps most at odds with the proposed use of artificial intelligence techniques, which seems to be viewed at present as the universal panacea to most technical problems. This is a poorly researched application of artificial intelligence and encourages additional processing on a platform. While there is a dearth of academic literature on this subject,

there are some commercial approaches, which propose a range of techniques, such as sharing ID documents or facial scanning — proposed techniques generally require a considerable amount of personal data to be shared (Braue, 2021).

This would immediately be in tension with the 8th standard of the code, i.e., data minimization, and require companies to process far more data than would be necessary to provide the service.

Third-party age verification would require data sharing with other parties (contradicting the 9th standard of the code), and hard identifiers, while discouraged, would require access to further personal data held in another database about the child.

Or, to put it more specifically, the code is calling on service providers to collect excessive data from children and young people in order to make them safer, without a clear appreciation of the capabilities of said data collection to achieve to aims of the regulations, which might result in the collection of personal data to verify the age of the user that far exceeds the data required by the service itself! If age verification was a simple task, it would have been put in place across online services many years ago — the problems are not technical, they are policy based. In a country with compulsory ID cards, age verification is easy. However, that is a debate long dismissed in the UK, for example, Lyon (2013).

Parental Controls, Excessive Data Collection, and the Erosion of Rights

The other standard within the code that raises cause for concern, for different reasons, is the following:

> ***11. Parental controls:*** *If you provide parental controls, give the child age-appropriate information about this. If your online service allows a parent or carer to monitor their child's online activity or track their location, provide an obvious sign to the child when they are being monitored.*

There has, in recent times, been a growing market in *SafetyTech*, the use of technology to ensure children and young people are safe.

A lot of these "solutions" have negative impacts on children's rights and do not always function as vendors claim. A code, which claims to

have the UN CRC at its heart, claiming that monitoring activity or tracking location is anything other than a breach of privacy, is somewhat contradictory. In particular, are the more controlling aspects of parental tracking and monitoring ever in the best interests of the child? Are they sufficiently respectful of the child's rights, regardless of whether they have been briefed on their use or there is an implied consent around the collection of the child's data for this purpose?

The standard's definition of parental controls is that they are

> *tools which allow parents or guardians to place limits on a child's online activity and thereby mitigate the risks that the child might be exposed to. They include things such as setting time limits or bedtimes, restricting internet access to pre-approved sites only, and restricting in-app purchases. They can also be used to monitor a child's online activity or to track their physical location.*

If we examine this statement, there are a number of issues arising. First, whether these tools actually mitigate risk. Some of my own previous work that explored the use of parental controls in the home (Phippen, 2019a) shows little evidence that a home with a high degree of parental control will result in a "safer" child. Certainly, limiting screen time and controlling device access at bedtimes will have some value for parents in ensuring effective boundaries are set in the home regarding device use, but do they do anything to mitigate risk? Even if we are to extend the definition of risk to safety to encompass risk to well-being, research by the Oxford Internet Institute (Orban and Przybylski, 2019) shows little evidence that screen time, even at bedtime, has any negative impact on well-being.

While the use of tools to restrict access to certain websites, or to use filtering to manage access to inappropriate or illegal content, does, to some degree, reduce the risk of upset for young people and there is little with which to take no issue with the use of filters for younger children, further work (Phippen, 2019b) highlights the fact that, particularly as young people get older, the diversity of content and behavior that causes upset online is significant, and we cannot simply address the risk of harm by blocking access because in encompassing all content that *might* cause harm, the filtering would restrict far beyond intended protections.

Filtering technologies have been well established in UK schools for many years. The basic approach is a simple one — prevent access to

"inappropriate" web content that might be harmful, upsetting, or offensive for young people. In schools, most of the systems use (UK schools have a statutory duty to implement "appropriate" filtering and monitoring of their online systems) keyword matching and blocking at a web address level to detect "inappropriate" content. The system looks for sexual keywords and blocks access to sites that contain them, or it checks against a list of sites already blocked, to prevent access should that website address attempt to be accessed. While the system is not perfect (many young people have told me they have experienced the filtering systems blocking innocent websites), it is accepted as a useful tool in school settings.

One of the fundamental issues with home filtering is how restrictive it can become, particularly when it makes use of keyword matching. As stated above, algorithms are poor at recognizing context, therefore they will identify the word and block, or *overblock*, regardless of the ambiguity of the use of the word.

However, filtering becomes more problematic when we move into the more ambiguous territory of "legal but harmful." In late 2020, in the UK, the opposition party (*BBC News*, 2020) has just made a call for social media platforms to prevent the spreading of "anti-vax" misinformation and the need to "stamp out" such information. They argued that emergency laws would hold platforms responsible should they fail to take down false stories about emerging COVID-19 vaccination programs. Platforms, they stated, should be held financially and criminally liable if they fail in their *duty of care* to remove such information.

However, as with any rule-based approach to content blocking, there needs to be a clear legal definition. Without a legal definition, it would be virtually impossible for an algorithm to accurately block this sort of information. Would content questioning government policy be considered "anti-vax" or a comment criticizing vaccine policy in a given country? It is very easy for someone (usually a politician) to say "this should be stopped" without actually thinking through what, technically, that would mean.

If we are to consider the UN CRC at the heart of this code, the encouragement of filters that can be ineffective does raise questions around article 17:

> *Access to information from the media: Every child has the right to reliable information from a variety of sources, and governments should encourage the media to provide information that children can*

understand. Governments must help protect children from materials that could harm them.

While one might argue that filtering contributes to the requirement for governments to help protect children from harmful materials and there is no suggestion that filters are not effective at preventing access to pornography (although, as I have been told many times by teenagers, determination will frequently bypass filters), it is not a perfect solution, and overblocking risks preventing access to reliable information on essential information related to sex and relationships education and gender issues.

There is general agreement that the year that filtering in the home became easily available was 2013, where governments pressured the four main Internet Service Providers into putting "default on" filtering tools into their home packages. Therefore, home filtering has now been widely available to subscribers for almost 10 years. However, the UK telecommunications regulator — Office of the Communications Regulator (OFCOM) — produced their annual Media Literacy report in 2019 (OFCOM, 2019) reporting a figure of 34% of parents of 5–15 year olds installing filters (this was the last time the question was posed). While we do not have much evidence on why parents do not choose to install them at home, the low numbers, after almost 10 years of being available, do raise the question: if these technologies are effective, why wouldn't parents install them in the home?

However, while filters introduce some concerns regarding children's rights, the focus of this chapter — excessive data collection from children under the guise of safeguarding and its impact on children's privacy — raises far more concerns regarding monitoring and tracking.

Monitoring is generally viewed as the more progressive, and less restrictive, bedfellow to filtering. A basic, introductory, monitoring approach is to use software to look at network traffic and raise alerts when monitoring rules are breached. The only data being collected in this scenario would be the details on the websites accessed by a child, not really cause for a rights-based concern.

While monitoring approaches will adopt similar techniques to filtering initially (for example, triggering an alert if someone generated a monitored keyword or tries to access a website on a watchlist), monitoring's toolbox can extend far beyond this. For example, message interaction and sharing, the interception, identification and redistribution of images, and elucidation of intent in communications based upon algorithmic interpretation. The monitoring class of SafetyTech has the potential for

technology to allow the parent to *See Everything, Always* (an advertising strapline of a particular SafetyTech vendor).

The central concept of any monitoring approach is simple: collect data on online access at a network or application level, and develop response strategies accordingly.

Within the school setting, the basic URL/keyword monitoring has now been superseded with other more active/pro-active platforms that can work at a far more sophisticated level, for example, being able to pro-actively monitor while a student is typing and make judgments on their intention as a result of this. There is clear guidance that, within a school setting, the technology will not be an automated solution but a tool to support staff in making safeguarding judgments, which is, arguably, the best role for technology: to collect data, raise alerts, and leave decision-making to those more capable of making informed decisions.

However, there has been significant evidence of feature creep in monitoring systems, particularly with home and app-based systems. While they used to function mainly around list-based interception and alerts, the technical capabilities of software and network systems mean that the feature suite can now be far more complex. But, with the introduction of new features, there seems to be little checking on whether, just because technology makes something possible, it *should* become part of a monitoring system. And there seems to be even less evidence of consideration of children's rights around these features which raises the question: when does a monitor become surveillance?

A good example of excessive monitoring can be seen in a famous legal case in the US: *Robbins v. Lower Merion School District* (PaceMonitor, n.d.).

This case has been subject to much discussion and is worthwhile exploring here because it does highlight the issue of technology extending moral boundaries and excessive control, and its impact on children's privacy and rights. In this case, a number of schools in the Lower Merion School District in the US adopted a policy of providing students with laptops for both in school and at home use. The expectation that the school might adopt a safeguarding approach that would use some forms of technology to monitor laptop usage is reasonable, and they needed to mitigate risk around the devices potentially being used for social or even illegal activities.

However, the software the schools decided to install far exceeded this intent. As a result of one of the schools involved in the scheme disciplining a student for what they referred to as "inappropriate" behavior at home, it

was discovered that the laptops were not only monitoring Internet access and application usage but also sending a stream of images, captured on the device's webcam, back to the school servers for analysis by staff.

As a result of suspicions raised by Blake Robbins, the student being disciplined, it was finally determined that over 66,000 images of students at his school were collected via these devices using the built-in webcams on the laptops. As well as sending images to the school directly when an online connection was available, the monitoring software was also capable of collecting images locally and uploading them at a later time. While the school argued they had valid safeguarding reasons for collecting these data, it was clear from the case that consent had not been obtained. Even if there was a safeguarding concern, the fact that the image data were subsequently used in a student disciplinary clearly demonstrated this remit had far been exceeded without fair consideration of the student's privacy or data protection rights.

Even if students *had* consented to data collected for safeguarding purposes, which would have been unlikely, the use of these data for disciplinary purposes would far exceed this remit for lawful processing. Furthermore, it was argued that given the schools took a *pro-active* decision not to inform either students or parents of the installed monitoring software or request consent, there is evidence that the intention was covert and student's privacy had further been breached.

Unsurprisingly, the case was found against the school district, and they were subject to a significant fine. This case was one of the first to highlight the potential for abuse in a monitoring system, and the temptation for excessive data collection just because the technology made this possible. It would be doubtful that, for example, the software platform used would have been advertised as "collect images of children in their home and use these data to discipline them in school."

A further aspect of SafetyTech that has more recently been added to the monitoring toolkit, and one that has become mainstream very quickly, is the measurement of screen time, something referred to in the Age Appropriate Design Code. Most smartphones will now provide screen time measures for the device users, and there are many tools that provide the means for parents to both view and control screen time for parents. As a tool for controlling access, the tools tend to work well, although we would raise the need to discuss with young people agreed screen time limits, rather than punitive imposition, and there are many reasons why limiting screen time might be a positive tool.

It is frequently suggested, such as in the Online Harms white paper (UK Government, 2019b), that parents need tools to ensure their children are not online *excessively*. However, there seems to be little understanding of what *excessive* means: it seems to be an entirely arbitrary and subjective term. I have, on occasions, asked children whether they think they spend too much time online. Many say they do. However, when I ask how long they spend online, it ranges from an hour a day to many hours. It would seem that "too much" is also subjective in the views of the young people we claim to wish to keep safe.

While in the past there seemed to have been similar concerns about how long young people watch television, again, there seemed to be little rigorous evidence to support any claims made, but there was much discourse around how watching television was a passive, negative activity, and young people would be better off playing outside. When we (as we frequently have been) are asked by parents "how long should my child be online per day?" our rather annoying response is usually "how long do you think they should be online for?" Another, equally irritating, response is "it depends." Screen time could be passive consumption on the content of platforms, such as YouTube and Tik Tok. Alternatively, they could be spending their time online collaboratively building a new extension to a Roblox game, developing technical knowledge and skills, and interacting actively with peers. Therefore, simplistic proposals to "manage" screen time are sometimes unhelpful. But, nevertheless, these proposals are made.

In an interview in *The Times* on Saturday, 10th March 2018 (*The Times*, 2018), the then Secretary of State for Digital, Culture, Media, and Sport, Matt Hancock, announced plans to bring in legislation that would restrict the amount of time children and young people could use social media platforms online in a simple soundbite:

> *There is genuine concern about the amount of screen time young people are clocking up and the negative impact it might have on their lives. It is right that we think about what more we could do in this area.*

The broader context of the suggestion proposed a legal requirement for social media providers to put effective age verification in place for anyone over the age of 13 (with the ill-informed belief that no children are on social media platforms before this age because its "illegal") and to keep track of their usage, enabling legally defined limits of access to be put in place.

Our own work with children and young people would suggest quite clearly that there is a *correlation* between the amount of time a child spends online and their exposure to risk (Phippen, 2018). We have seen from a large dataset that a child who spends a self-reported more than six hours a day online is twice as likely to have seen content or received comments that have upset them compared to someone who spends less than an hour online. It also shows that many young people who go online for over six hours a day are likely to do so because they are lonely. However, this is a correlation, not a causation, and does not show whether children are lonely because they are online, or whether they are lonely, and, therefore, go online. Equally, we can also see from our data that there are other heavy online users who are very happy (generally these would be self-disclosed gamers).

Reflecting again on our own experiences talking to children and young people, we also see many positives for screen time. For some children, for example, those in isolated communities, going online is a window to the wider world. At the time of writing, we are emerging from the third national lockdown as a result of COVID-19, and many young people have disclosed the "lifeline" that digital technology offered them.

This is not to say that young people should be free to be online for as long as they wish. However, I would take exception to the view that technology has to provide the solution to this. Surely, a more realistic approach to excessive screen time (however this is agreed upon in the home) is for a parent to manage it, through observation and house rules?

There are a number of issues arising from this list of solutions that cause concern beyond the current screen time debate. We have already discussed filtering at length, but the proposed functionality in the feature list of home monitoring solutions far outweighs proportionate response to child safeguarding concerns. It would seem that many parents, in order to reassure themselves that their children are safe, feel they need to know about every element of communication in their lives. SafetyTech providers can potentially build an effective business model on the back of a *reassurance myth* that will encourage parents to purchase their products, whether or not there is a real problem to solve. And legislators are encouraging this practice, with little concern that these platforms potentially result in excessive collection of children's personal data.

I have been told on a number of occasions by stakeholders in child safeguarding, when alerted to accusations of excessive collection of children's data, that "it's ok, safeguarding trumps data protection, it's in the GDPR," or words to that effect.

While it is true that there are safeguarding provisions within the Data Protection Act 2018 (*ibid.*) legislation:

4(1) This condition is met if
- (a) The processing is necessary for the purposes of
  - (i) Protecting an individual from neglect or physical, mental, or emotional harm or
  - (ii) Protecting the physical, mental, or emotional well-being of an individual,
- (b) The individual is
  - (i) Aged under 18 or
  - (ii) Aged 18 or over and at risk,
- (c) The processing is carried out without the consent of the data subject for one of the reasons listed in subparagraph (2), and
- (d) The processing is necessary for reasons of substantial public interest.

(2) The reasons mentioned in subparagraph (1)(c) are
- (a) In the circumstances, consent to the processing cannot be given by the data subject,
- (b) In the circumstances, the controller cannot reasonably be expected to obtain the consent of the data subject to the processing, and
- (c) The processing must be carried out without the consent of the data subject because obtaining the consent of the data subject would prejudice the provision of the protection mentioned in subparagraph (1)(a).

It would be unlikely that all monitoring interventions could be argued on the grounds of "protecting an individual from neglect or physical, mental, or emotional harm," given that a lot of the monitoring is simply collecting messages exchanged between the child and their peers. And while there is little case law to explore this argument at this stage (given most children lack the finances and legal support to sue their parents for excessive data collection), the Lower Merion case in the US does highlight covert and excessive data collection can be considered illegal even in a safeguarding content.

Additionally, there is also often an argument that the child has consented to the data collection by either agreeing to have the monitor software installed on their devices or through an agreed Acceptable Usage Policy at school.

When considering this, we can return to the Code and its calls for "young person" friendly means for them to make the consent judgment, making effort to differentiate between those viewed as capable of consenting to their data being collected, as defined in section 9 of the Data Protection Act 2018 (*ibid*.), and those younger:

> ***Under 12*** *— Provide audio or video materials for the child to explain that their parent is being told where they are and/or what they do online to help keep them safe.*
>
> ***Over 12*** *— Provide audio, video, or written materials for the child to explain how your service works and the balance between parental and child privacy rights.*

It would be interesting to consider whether this child-friendly approach to consent is implemented in all of these monitoring scenarios. Certainly, young people I have spoken to have never said that SafetyTech has been installed on their device following the discussion of their privacy rights and being talked through an Acceptable Use Policy. It is more likely to be a conversation where their parents say they are concerned for the child's safety and as a result, the child must have the technology installed.

From many conversations with parents, there are plenty who believe it is their right to see every conversation their child has online and to know exactly who they are speaking to at any time — the belief being if they can see all of the communications, they will know they are safe. In essence, they believe the erosion of the child's privacy is acceptable because the child will be safe or "safeguarding trumps data protection rights."

There is a risk that we are confusing safety with surveillance, and because technology provides the methods to achieve this, we collect a suite of tools that allow us to collect more and more data on our children — convinced with the notion that they are, in some way, safe if we have all of these data.

The concept of safety is interesting in this context — the justification for the use of increased monitoring is that it is needed to *assure* safety in the same way that overblocking is justified because it will *prevent* access to inappropriate content. Yet, do these technologies do much to actually achieve safety? Will using these tools ensure a child is safe? Or are they tools to monitor and control behavior instead, much like we saw in the Lower Merion District case?

There are some risks that can be mitigated using this level of surveillance, for example, the issues around grooming and contact from potential abusers might be mitigated by having access to contact lists and messaging. Yet, these apps will only provide access to certain messaging platforms. While access to the mobile device's own telephony (i.e., calls and SMS) is relatively straightforward, to access app-specific messaging is more problematic, which is generally why only major platforms (for example, Facebook, Instagram, Snapchat, and WhatsApp) are covered.

However, perhaps the most concerning of these is the use of tracking technology: collecting information on the location of the child or, more specifically, the location of the child's device.

In 2018, the UK celebrity Jamie Oliver (Petter, 2018) promoted the use of tracking technology, saying it was a "brilliant" way to ensure his children were safe when they were away from the family home:

> *The older girls, Jools and I are all on an app…, which means we can see exactly where everybody is and the route they've gone, so if one of the girls says, 'I'm going to Camden Town' and I can see they've gone to Reading, then we have a problem.*
>
> *He continued:*
>
> *They can check on me, too, and see how fast I'm driving. It's brilliant.*

Device tracking takes this surveillance culture/reassurance myth a step further, such that a parent can monitor, it is argued, the location of the child, and, therefore, be reassured they are safe. However, it would be more accurate to state that the tool allows a parent to see the location of the child's device and provides no means whatsoever to consider the well-being of the child or whether they are engaged in risk-taking behavior. Reassuringly, most parents I speak to do believe that access to the child's camera to see their remote interactions is still viewed as a step too far.

We have seen some scenarios where the likes of tracking technology could be used in a positive way, for example, in the case of a severely epileptic young man who was prone to many seizures per day. As he got older, the parents negotiated the use of tracking technology so that he was able to go out independently, but if he had a seizure, his parents would be able to locate him. However, these scenarios are in the minority — in our experience, many parents view the use of tracking technology as their right and, in many cases, will track their child without them knowing.

If young people are to be subject to these tools, the transparency called for in the standard is to be welcomed. However, again, my own empirical work would suggest that transparency and openness are rarely discussed in the family home. While parents sometimes justify this covert surveillance as a proactive way of ensuring the child is safe, this is problematic, controlling activity such that the tools are used more as disciplinary, than safeguarding, devices. For example, a child might tell their parent they are visiting a friend's house when they actually go elsewhere, perhaps somewhere to which the parents may disapprove. The discovery, through tracking, of this lie by the child is used as a punitive, not caring measure.

This is why the need for information around the controls, disclosed in a child-friendly manner, is to be welcomed. However, it is doubtful this will be enacted in many homes. If the premise of these tools is risk mitigation, then this potentially reduces this role further. Consider the following comment by Mr. Oliver:

> *...if one of the girls says 'I'm going to Camden Town' and I can see they've gone to Reading, then we have a problem.*

Returning to the discussions around best interest, we would question whether the legitimization of parental control, particularly the more invasive ones, really does consider the best interests of the child. If we are to take the most extreme of these controls — the GPS tracking of minors, we might argue that for young children, who will either be at school, in the home, or with parents, tracking is unnecessary. For older children, surely we should acknowledge the child's right to privacy and be encouraging intrafamilial discourse in order to determine how best to know where the child is going and with whom rather than expecting the excessive collection of sensitive location data (which, elsewhere in the code, is clearly viewed as problematic) by software to achieve this. While the code does much to acknowledge the right to privacy, as defined by the UN CRC, the discourse does little to challenge the notion that perhaps the tracking and monitoring of individuals and their online use are not a good thing.

So, on the one hand, there is legislative guidance that suggests the monitoring of others is a coercive and potentially abusive thing to do. However, within the code, there are no challenges toward the tracking of minors, as long as they are made aware it is happening and that their parents are doing it to ensure they are safe. While for minors 13 years or older

the call from the standard is to explore the balance between parental and child privacy rights, we would argue that, except in cases where the quality of life of the child is improved through the use of tracking, there are few rights-based arguments for a child to be controlled in this way. It is, therefore, disappointing to see it viewed as a normal part of parental controls in the code.

Conclusion

Of the matters facing the digital sector at the moment, greater regulation and accountability is one of the most significant changes in its brief history. In a world where digital technology, and the providers of that technology, permeates all aspects of citizens' lives, it is a reasonable expectation that there is a level of accountability to the services that are provided. However, the sector also deserves legislation and regulation that are fit for its purpose: society's darker side manifested in online channels is not a problem that providers can solve on their own, and legislation that pressures providers to solely achieve this is doomed to fail.

The challenges of protecting children online, while allowing them positive engagement, are well established but continue to be poorly regulated. In this chapter, we have explored that, through the best of intentions, the need to keep children safe online might be one of the primary reasons for excess data collection and erosions of their privacy rights. While the UK's Age Appropriate Design code adopts a rights-based approach and has much to be viewed as positive engagement with service providers, it still has the potential to encourage excess data collection and the erosion of privacy in order to implement the safeguarding fallacy.

This is not to say that governments should embrace Barlow's manifesto as to not even attempt to regulate the online world, however, it is crucial that future legislators have an appreciation of the technology they hope to regulate, and the extent to which regulation can help, rather than hinder progress. Regulators should be confident in applying legislation and holding providers not showing duty of care to account. However, they also need to be mindful that excessive technical intervention might have knock-on effects around rights, particularly where children are concerned. Children certainly have a right to best interests around their use of online services, but their best interests are rarely achieved by removing their rights to privacy or collecting excessive levels of personal data just so they can prove their age or use a service.

Many attempts to regulate online services fail to appreciate the role education plays: an informed user is undoubtedly better equipped to ensure they are safe online. A balance between education and technical tools is a far more powerful balance than regulation alone: tools can be provided, for example, to block and report abusive participants in a service, and education is there to articulate the importance of being able to do this, as well as developing citizens who are aware of the need for data protection rights and an expectation of privacy. While the Age Appropriate Design Code alludes to levels of education being provided, it does not explicitly express this. In the draft Online Safety Bill, it is even weaker: a 145-page bill mentions education only twice, in a single section, that suggests the regulator has a role to play in implementing education initiatives around media literacy.

Providers of online services undoubtedly need to play their role in the safeguarding of children, done so in a manner that is in the best interests. However, they should not be seen as the lone stakeholder in this regard, and children should not be used as a political pawn to impose excessive regulation on a sector that is doomed to fail.

References

Barlow, J. P. (1996). Declaration of independence for cyberspace. www.eff.org/cyberspace-independence. Accessed on 3 June 2021.

BBC News (2019). Christchurch shootings: Sajid Javid warns tech giants over footage. https://www.bbc.co.uk/news/uk-47593536. Accessed on 3 June 2021.

BBC News (2020). Covid-19: Stop anti-vaccination fake news online with new law says labour. https://www.bbc.co.uk/news/uk-politics-54947661. Accessed on 3 June 2021.

Cheswick, R. W. (2003). *Firewalls and Internet Security: Repelling the Wily Hacker*, p. 202. Boston: Addison-Wesley Professional.

Children's Online Privacy Protection Act (1998). Children's online privacy protection act. https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule. Accessed on 3 June 2021.

Data Protection Act (2018). Data protection act 2018. https://www.gov.uk/government/collections/data-protection-act-2018. Accessed on 3 June 2021.

Digital Economy Act (2017). The digital economy act 2017 part 3. http://www.legislation.gov.uk/ukpga/2017/30/part/3/enacted. Accessed on 3 June 2021.

Douglas Adams (1999). How to stop worrying and learn to love the internet. Available at: https://douglasadams.com/dna/19990901-00-a.html. Accessed on 3 June 2021.

Draft Online Safety Bill (2021). Draft online safety bill. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985033/Draft_Online_Safety_Bill_Bookmarked.pdf. Accessed 3 June 2021.

Federal Trade Commission (2019). Musical.ly, Inc. https://www.ftc.gov/enforcement/cases-proceedings/172-3004/musically-inc. Accessed on 3 June 2021.

General Data Protection Regulation (GDPR) (2018). General data protection regulation (GDPR) — Final text neatly arranged. https://gdpr-info.eu/. Accessed on 3 June 2021.

Global Privacy Enforcement Network (2015). GPEN sweep international report. https://www.privacy.org.nz/assets/Files/Media-Releases/2015-GPEN-Sweep-International-Report.pdf. Accessed on 3 June 2021.

Helm, T. and Rawnsley, A. (2018). Health chiefs to set social media time limits for young people. https://www.theguardian.com/media/2018/sep/29/health-chief-set-social-media-time-limits-young-people. Accessed on 3 June 2021.

House of Commons Science and Technology Committee (2017). Impact of social media and screen-use on young people's health. https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/822.pdf. Accessed on 3 June 2021.

Information Commissioner's Office (2020). Age appropriate design — A code of practice for online service providers. https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/age-appropriate-design-a-code-of-practice-for-online-services-2-1.pdf. Accessed on 3 June 2021.

Lyon, D. (2013). The border is everywhere: ID cards, surveillance and the other. In *Global Surveillance and Policing*, Zureik, E. and Salter, M. (Eds.). pp. 78–94. Routledge.

Mohdin, A. (2019). Matt Hancock 'won't rule out' compulsory vaccinations. https://www.theguardian.com/politics/2019/may/04/matt-hancock-wont-rule-out-compulsory-vaccinations. Accessed on 3 June 2021.

OFCOM (2019). Children and parents media use and attitudes: Annex 1. https://www.ofcom.org.uk/__data/assets/pdf_file/0027/134892/Children-and-Parents-Media-Use-and-Attitudes-Annex-1.pdf. Accessed on 3 June 2021.

PacerMonitor (n.d.). Robbins V lower Merion district. https://www.pacermonitor.com/view/6LZS7RA/ROBBINS_et_al_v_LOWER_MERION_SCHOOL_DISTRICT_et__paedce-10-00665__0001.0.pdf. Accessed on 3 June 2021.

Petter, O. (2018). Jamie Oliver reveals he tracks daughter's location on app — But parenting experts say it could cause future problems. https://www.

independent.co.uk/life-style/health-and-families/jamie-oliver-tracks-location-life360-kids-parenting-a8545136.html. Accessed on 3 June 2021.

Phippen, A. (2016). *Children's Online Behaviour and Safety: Policy and Rights Challenges*. Springer.

Phippen, A. (2018). Young people, internet use and wellbeing; a report series — Screentime. https://swgfl.org.uk/assets/documents/young-people-internet-use-and-wellbeing.pdf. Accessed on 3 June 2021.

Phippen, A. (2019a). Young people, internet use and wellbeing; a report series — Technology in the home. https://swgfl.org.uk/assets/documents/technology-in-the-home.pdf. Accessed on 3 June 2021.

Phippen, A. (2019b). Young people, internet use and wellbeing; a report series — What causes upset online? https://swgfl.org.uk/assets/documents/what-causes-upset-online.pdf. Accessed on 3 June 2021.

Schneier, B. (2005). Computer crime hype. https://www.schneier.com/blog/archives/2005/12/computer_crime_1.html. Accessed on 3 June 2021.

Telecommunications Act (1996). The telecommunications act 1996. https://www.fcc.gov/general/telecommunications-act-1996. Accessed 3 June 2021.

The Times (2018). Time limits for children hooked on social media. https://www.thetimes.co.uk/article/time-limits-for-children-hooked-on-social-media-3s66vwgct. Accessed on 3 June 2021.

UK Government (2018). UK government response to internet safety strategy green paper. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/708873/Government_Response_to_the_Internet_Safety_Strategy_Green_Paper_-_Final.pdf. Accessed on 3 June 2021.

UK Government (2019a). Open letter from the Home Secretary — Alongside US Attorney General Barr, Secretary of Homeland Security (Acting) McAleenan, and Australian Minister for Home Affairs Dutton — To Mark Zuckerberg. https://www.gov.uk/government/publications/open-letter-to-mark-zuckerberg/open-letter-from-the-home-secretary-alongside-us-attorney-general-barr-secretary-of-homeland-security-acting-mcaleenan-and-australian-minister-f. Accessed on 3 June 2021.

UK Government (2019b). Online harms white paper. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/973939/Online_Harms_White_Paper_V2.pdf. Accessed on 3 June 2021.

UK Government (2021). Landmark laws to keep children safe, stop racial hate and protect democracy online published. https://www.gov.uk/government/news/landmark-laws-to-keep-children-safe-stop-racial-hate-and-protect-democracy-online-published. Accessed on 3 June 2021.

United Nations (1989). United Nations convention on the rights of the child. https://downloads.unicef.org.uk/wp-content/uploads/2016/08/unicef-convention-rights-child-uncrc.pdf. Accessed on 3 June 2021.

United Nations (2013). General comment no. 14 (2013) on the right of the child to have his or her best interests taken as a primary consideration (art. 3, para.1)*. https://www2.ohchr.org/English/bodies/crc/docs/GC/CRC_C_GC_14_ENG.pdf. Accessed on 3 June 2021.

This page intentionally left blank

https://doi.org/10.1142/9781800612860_0011

Chapter 10

Privacy and Security of Health Data — What's at Stake?

Elena Lazar

Introduction

Together with the evolution of technology, the world is becoming more and more paperless and more and more digitized and, consequently, most organizations prefer nowadays to store data in the cloud or in a system rather than piling hundreds of files. But storing all these amount of data gives rise to risks and vulnerabilities in terms of privacy and security, especially when it comes to health data.

People are more tempted to use different kinds of apps or wearable devices to keep track of their health status or even to have all their data stored and secured in one place for their convenience as patients. All these systems allow continuous monitoring of individuals' psychological and health conditions by sensing and transmitting measurements, such as heart rate, body temperature, respiratory rate, chest sounds, or even blood pressure. But as technology evolves, cyberattacks also increase and health data become incredibly valuable. Furthermore, the pandemic context put an emphasis on the importance of health data and the relationship between data protection and the use of new technologies, such as tracing apps.

Before diving into the core issue of our subject, we feel the need to address the distinction between data privacy and data security. In addition, even though data security and data privacy seem to be used sometimes

interchangeably, they are distinct notions: while data security protects data from being compromised by external attackers and malicious insiders, data privacy is charged with governing how personal data are collected, shared, and used.

In order to address this subject, we will first focus on what health data represents and its particularities (I), second, we will address the processing of data for scientific research (II), then move on to tackle the impact of the pandemic on health data (III), and finally analyze the risks posed by cyberattacks (IV).

Health Data and Its Particularities

Most people consider their medical data to be among the most sensitive types of personal information, full of private and confidential details that they don't feel the need to share with other people, apart from care providers/doctors in order to be able to receive treatment. The providers and medical staff include these data in medical records, containing information on diagnoses, lab results, and treatment options and prescriptions. They might also contain information about chronic diseases or mental health counseling.

Doctor–patient privilege, also enshrined by Hippocratic Oath[1] (a pledge that many medical school students recite upon their graduation), is a long-held tradition, particularly in the light of the sensitive character of health data. It is based upon the idea that patients should be able to trust a doctor's discretion so that they will seek medical care and not withhold information during a consultation. If a doctor does not have accurate data on a patient's health, this may lead to an incorrect diagnosis and most likely an incorrect treatment.

In addition, personal data, which are, by their nature, particularly sensitive in relation to fundamental rights and freedoms, deserve specific protection taking into account that their processing could generate significant risks to the subjects' health.

[1] **Hippocratic oath**, ethical code attributed to the ancient Greek physician Hippocrates, adopted as a guide to conduct by the medical profession throughout the ages and still used in the graduation ceremonies of many medical schools, available at Hippocratic oath | Definition, Summary, & Facts | Britannica [accessed on 20 April 2021].

In order to characterize this category of data, we should look at the nature of the data: Do they relate to the health of an identified or identifiable person? and at their use and purpose: Are they used for medical purposes? Finally, if data are used for medical purposes, regardless of its nature, it can be considered health data by destination.

Health data refers as such to personal information that relates to the health status of a person (GDPR, 2016).[2] This includes both medical data, such as doctor referrals and prescriptions, medical examination reports, laboratory tests, radiographs, and CTs, and also administrative and financial information about health, for example, the calendar of medical appointments, invoices for the healthcare services provided, and medical certificates for sick leave management. Health data is considered sensitive data according to the General Data Protection Regulation (hereinafter GDPR) and is subject to strict rules and can only be processed by health professionals who are bound by the obligation of medical secrecy. According to recital 53 of the Regulation, data concerning health need to be granted higher protection, as the use of such sensitive data could have a significant impact and seriously affect data subjects. Furthermore, the processors and controllers shall take the necessary measures to ensure that the health data is protected and not subject to any unauthorized disclosure.

The notion of health data should be interpreted broadly, as it represents an autonomous notion and can be derived from different sources: Information collected by a health care provider or health care facility in a patient record (such as applied treatments or even details about pacemakers), information that becomes health data from inferences (Wachter and Mittelstadt, 2019) thus revealing the state of health or health risks (such as the assumption that a person has a higher risk of suffering from diabetes based on the amount of sugar eaten over a certain period of time or that person's weight), information from a "check" survey filled in when going to a clinic facility, for example, where data subjects answer questions related to their health (such as "do you suffer from any chronic diseases?" or "have you had any recent surgery?), or information that might become health data due to its usage in a specific context (such as information regarding a recent trip to or presence in a country affected or exposed to

[2]According to Article 4(15) GDPR, "data concerning health" means "personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status."

malaria). Also, the fact of correlating data relating to the number of daily steps taken by an individual with his age, his sex, and his eating habits can, by means of an algorithm, make it possible to deduce with a certain margin of error, the state of health of that person. Where appropriate, these data would be qualified as health data.

As a general rule, article 9(1) of the Regulation prohibits the processing of special categories of personal data, known as "sensitive data," unless one of the **exemptions** provided by this article, para. (2) applies and appropriate safeguards for data protection are put in place. As to the exemptions, health data can be processed if explicit **consent** was given by the data subject, processing is necessary to protect the **vital interests** of a person if this person is, for example, physically or legally incapable to give consent or processing is necessary in order **to provide healthcare** if the data are processed under the responsibility of a professional subject to the obligation of professional secrecy. Even in the absence of consent, letter (i) of the article 9(2) states that *it is possible to process such data, even in the absence of the data subject's consent, for* "*reasons of overriding public interest in the field of public health, such as protection against serious cross-border threats to health (...), under Union or national law*." Recitals 52[3] and 54[4] of the Preamble to the Regulation also state that a reason of public interest in the field of health might justify the processing of special categories of personal data, even without the consent of the data subject, provided that there are adequate safeguards to that effect. Suitable

[3] *(52) Derogating from the prohibition on processing special categories of personal data should also be allowed when provided for in Union or Member State law and subject to suitable safeguards, so as to protect personal data and other fundamental rights, where it is in the public interest to do so, in particular processing personal data in the field of employment law, social protection law, including pensions and for health security, monitoring and alert purposes, the prevention or control of communicable diseases, and other serious threats to health. Such a derogation may be made for health purposes, including public health and the management of health-care services, especially in order to ensure the quality and cost-effectiveness of the procedures used for settling claims for benefits and services in the health insurance system or for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes.*

[4] *(54) The processing of special categories of personal data may be necessary for reasons of public interest in the areas of public health without consent of the data subject. Such processing should be subject to suitable and specific measures so as to protect the rights and freedoms of natural persons.*

safeguards might include, for example, **pseudonymization**, anonymization, or **encryption**.[5]

The processing of health data is also possible for scientific research purposes on the legal basis of consent obtained pursuant to article 6(1)(a) and article 9(1)(h), (i), and (j) of the GDPR. In order to be deemed valid, consent must be in accordance with the conditions prescribed under the GDPR, the European Data Protection Board (hereinafter EDPB), particularly outlining that it must be *freely given, specific, informed, and unambiguous* and *made by way of statement or clear affirmative action* (EDPB Guidelines, 2020a). In order to fulfill these requirements, data subjects should not feel pressured and should understand they would suffer no disadvantages if they decided not to provide their consent. However, the processing for scientific research will be tackled in the following section.

In terms of the rights of the data subjects, they must be informed about their rights and for what purposes their health-related information is processed, must have the right to access their medical files and other health-related information to be able to verify whether it is accurate and to rectify any inaccurate or incomplete information, and must also have the possibility to withdraw consent at any time. In this context, all processing activities that were previously based on consent remain lawful, but controllers must bear in mind the need to stop the processing and delete the personal data if they can no longer rely on another lawful basis justifying the retention for further processing.

Organizations/public or private entities must make sure that information relating to health is not kept on their files for longer than necessary and clear retention periods must be established. These can vary in accordance with the reason for processing the health data. For example, if a treatment scheme for a patient diagnosed with a chronic disease is being developed and divided into steps amounting to five years, then a longer period of retention could be justified.

Also, given the sensitivity of health data, it should only be processed, like previously stated, by health professionals who are bound by

[5] *Article 32(1*) *Taking into account the state of the art, the costs of implementation, and the nature, scope, context, and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organizational measures to ensure a level of security appropriate to the risk* [..].

the obligation of medical secrecy and any other person (for example, a subcontracted processor) dealing with administrative or even financial procedures in this respect should sign a specific confidentiality declaration.

Taking into account the particular nature of health data and the limited exemptions provided for processing, we will be addressing in the following section the potential use of health data for scientific research.

Processing Health Data for Scientific Research

Article 4[6] GDPR does not entail an explicit definition of "processing for the purpose of scientific research." As indicated by Recital 159, *the term processing of personal data for scientific research purposes should be interpreted in a broad manner including, for example, technological development and demonstration, fundamental research, applied research, and privately funded research. In addition, it should take into account the Union's objective under Article 179(1) TFEU of achieving a European Research Area. Scientific research purposes should also include studies conducted in the public interest in the area of public health.*

Health data used for processing for the purpose of scientific research can be divided into two categories according to EDPB (EDPB Guidelines, 2020a): for primary use, research on health data which consists of the use of data directly collected for the purpose of scientific studies, and secondary use, research which consists of the further processing of data initially collected for another purpose, later on.

For example, when conducting a clinical trial on individuals who developed a certain form of blood cancer and health data are being collected and processed based on the consent, we are dealing with primary use.

In the situation where data subjects have used a certain treatment recommended by their doctor, if these health data recorded by him are being used later on for scientific research purposes, the processing in

[6] *Article 4(2)* "*Processing*" *means any operation or set of operations which are performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.*

question represents secondary use, taking into account that the data had been collected for another initial purpose (EDPB Guidelines, 2020a).

This distinction between primary and secondary usages of health data proves to be important when talking about the legal basis used for the processing.

It is clear from the start that the processing of health data must of course comply with the principles set out in article 5 GDPR and with one of the legal grounds and the specific derogations listed respectively in article 6 and article 9 GDPR for the lawful processing of this special category of personal data.

In general, a data subject must be individually informed of the existence of the processing operation and that personal (health) data are being processed for scientific purposes. The information delivered should contain all the elements stated in article 13 or article 14 GDPR. It has to be noted that researchers often process health data that they have not obtained directly from the data subject, for instance, using data from patient records or data from patients in other countries. Therefore, article 14 GDPR, which covers information obligations where personal data are not collected directly from the data subject, will apply.

More specifically, as regards further processing, the EDPB indicates (EDPB Guidelines, 2020a) data subjects should be provided with information *within a reasonable period of time before the implementation of the new research project*, explaining that this would create awareness with respect to the project and allow the exercise of data subjects' rights.

This processing can be performed directly on the individual, in the context of clinical trials where individuals might subscribe voluntarily (for example, to become testing subjects for certain drugs or even vaccines) because either they are paid and need money or they hope to find a treatment that works in combating a certain disease, or on the individual's samples provided voluntarily (blood sample and embryos) for the purpose of scientific research or on data collected previously for other purposes (secondary usage).

In the first situation related to clinical trials, we should also take into account the application of the Clinical Trial Regulation (hereinafter CTR) (CTR, 2014). While GDPR aims to protect individuals with regard to the processing of personal data, the CTR wishes to greater harmonize the rules for conducting clinical trials throughout the EU (EU Commission Website, 2021). According to the principle of accountability, it is the data controller (sponsor/clinic institution of the investigator according to CTR)

who is in charge of implementing the appropriate technical and organizational measures to ensure the respect of data protection rules.

And since we mentioned data protection, it is clear that both the GDPR and the CTR apply simultaneously[7] and that while the CTR contains specific data protection provisions, it must be stated from the outset that it does not allow derogation from it so that it diminishes the legal requirement to comply with the GDPR. In order to offer an adequate level of protection to data subjects, both regulations must be applied together. For example, "informed consent" provided under the CTR to participate in a clinical trial is not the same as consent to process personal data under the GDPR. The condition of informed consent under the CTR may be accomplished, while an imbalance of power between the participant and the investigator[8] may not pass as "freely given" consent as required by the GDPR. An example of that kind of imbalance would be in the situation where the individual is extremely poor and the payment received as an outcome of a clinical trial would be his only solution for a better life or if the person is hospitalized and the doctor insists on submitting that person to a clinical trial. Furthermore, the informed consent under CTR represents a fundamental condition under which a person can be included into a clinical trial and it is not conceived as an instrument for data processing compliance like in the GDPR (EDPB Opinion, 2019).

Also if the CTR allows the processing of health data for reliability and safety-related purposes to be performed on the basis of legal obligation, necessary to comply with the legal obligations to which the investigator is subject to (safety reporting, archiving of master files, and disclosure of clinical trial data), according to GDPR processing operations purely related to research activities within clinical trials cannot be based on legal obligations but on the data subject's explicit consent (subject to the GDPR's conditions around consent when processing special categories of data) (article 6(1)(a) in conjunction with article 9(2)(a)) or the legitimate interests of the controller (article 6(1)(f) in conjunction with article 9(2)(i) or (j)) or the public

[7]CTR, *Article 93* "*Member States shall apply Directive 95/46/EC [now repealed by the GDPR] to the processing of personal data carried out in the Member States pursuant to this Regulation" and that "Regulation (EC) No 45/2001 [repealed by Regulation 2018/1725] shall apply to the processing of personal data carried out by the Commission and the Agency pursuant to this Regulation*."

[8]*Article 2 (15)* "*Investigator*" *means an individual responsible for the conduct of a clinical trial at a clinical trial site*.

interest (article 6(1)(e)). However, data controllers should be mindful that their legitimate interest to process personal data in the context of a clinical trial will need to be balanced toward the interests of the individual participants. Legitimate interests cannot be relied upon if they are overridden by the interests or fundamental rights and freedoms of the individual. Whether or not the "public interest" legal basis can be relied upon will depend on whether the clinical trials fall *within the mandate, missions and tasks vested in a public or private body by national law* (EDPB Opinion, 2019).

If the processing for scientific purposes is made outside a clinical trial (EDPB Opinion, 2019), the EDPB states that it is not possible to rely only on the CTR consent to process personal data in the case of secondary use and a separate GDPR legal ground to process is required (the same ground or different one to that relied upon for the primary use). Let's take, for example, the situation in which health data were collected to conduct a clinical trial on blood diseases but later on the team of doctors conducting the study realize that they could use these results also to run a study aiming at sequencing genomes but which were not foreseen at the beginning of the first clinical trial protocol; in this case, a legally valid ground under article 6 of the GDPR would be required.

In terms of the withdrawal of consent to participate in a clinical trial under CTR, this may not necessarily affect the processing of personal data gathered in the context of that specific trial and may continue where there is an appropriate legal basis for such processing under GDPR (for example, if the patient has suffered a serious adverse reaction, the investigator has the right to process the previously obtained data by reporting the data to the national competent authorities based on the legal obligation of the controller — article 6(1)(c) of the GDPR in conjunction with article 9(2)(i). Under the GDPR, if consent is used as the lawful basis for processing (article 6(1)(a)), there must be a possibility for individuals to withdraw that consent at any time (article 7(3)), and there is no exception to this requirement for scientific research provided for under article 7. As a general rule, if consent for data processing under GDPR is withdrawn, all data processing operations that were based on consent remain lawful in accordance with the GDPR (article 7(3)).

A particularly interesting issue related to clinical trials resides in the possibility to be enrolled in one, in a situation of emergency.[9] As such,

[9] *By way of derogation from points (b) and (c) of Article 28(1), from points (a) and (b) of Article 31(1), and from points (a) and (b) of Article 32(1), informed consent to participate*

once the conditions enshrined at article 35 of the CTR are fulfilled, a subject can be enrolled in a clinical trial in the situation of emergency, exceptionally even without any prior informed consent. For example, if the patient finds himself in a coma and a clinical trial might be available in order to save or prolong his life, the consent should be sought from his or her legal representative as soon as possible in order to maintain the subject in the clinical trial. And as previously stated, since the prior informed consent of the subject only represents an additional safeguard and not the legal basis for the processing from a data protection perspective, the legal basis for the processing of personal data in the context of emergency clinical trials should be the public interest pursued as per article 6(1)(e) of the GDPR or the legitimate interest pursued provided at article 6(1)(f) of the GDPR interpreted in conjunction with article 9(2)(c). However, if a data subject dies before his consent could be confirmed, the processing of that data is no longer covered by the GDPR.

In the second situation, we envisaged when the processing is performed on individual's samples provided voluntarily for the purpose of scientific research or on data collected previously for other purposes, still one of the legal grounds provided by article 6 of the GDPR must be

---

in a clinical trial may be obtained, and information on the clinical trial may be given, after the decision to include the subject in the clinical trial, provided that this decision is taken at the time of the first intervention on the subject, in accordance with the protocol for that clinical trial" and that all of the following conditions are fulfilled: (a) due to the urgency of the situation, caused by a sudden life-threatening or other sudden serious medical condition, the subject is unable to provide prior informed consent and to receive prior information on the clinical trial; (b) there are scientific grounds to expect that participation of the subject in the clinical trial will have the potential to produce a direct clinically relevant benefit for the subject resulting in a measurable health-related improvement alleviating the suffering and/or improving the health of the subject, or in the diagnosis of its condition; (c) it is not possible within the therapeutic window to supply all prior information to and obtain prior informed consent from his or her legally designated representative; (d) the investigator certifies that he or she is not aware of any objections to participate in the clinical trial previously expressed by the subject; (e) the clinical trial relates directly to the subject's medical condition because of which it is not possible within the therapeutic window to obtain prior informed consent from the subject or from his or her legally designated representative and to supply prior information, and the clinical trial is of such a nature that it may be conducted exclusively in emergency situations; (f) the clinical trial poses a minimal risk to, and imposes a minimal burden on, the subject in comparison with the standard treatment of the subject's condition.

applicable: a task carried out in the public interest under article 6(1)(e) in conjunction with Article 9(2), (i) or (j) of the GDPR; or the legitimate interests of the controller under article 6(1)(f) in conjunction with article 9(2)(j) of the GDPR; or, when all conditions are met, data subject's explicit consent under article 6(1)(a) and 9(2)(a) of the GDPR.

Comprehensive and real-world personal health and activities' datasets (Rawassizadeh *et al*., 2015) have a very important role in data processing and analysis for scientific purposes and when health data (used in data sets) have been made available publicly by the data subjects, then the legal ground for processing is considered to have been fulfilled according to article 9(2)(e) (examples of such datasets are UbiqLog and CrowdSignals (CrowdSignal Website, 2021), which contain both the data from the smartphones and from wearable devices, such as smartwatches). Under these circumstances, no other ground under the GDPR is required.

In terms of secondary usage, it is already a common practice the use of individual monitoring systems, focused on personal data collection and analysis: tracking the users' exercise routines (footsteps), measuring activity levels (the amount of water drank per day), and pulse (Vitabile, 2019). Although the data collected have been for the sole purpose of user consumption and data subjects have given their consent only for this purpose, sharing these health data with medical healthcare facilities for scientific purposes is also common. As such, a legal ground for processing this new purpose is needed. However, the secondary use of data which is anonymized does not fall within the scope of the GDPR.

And lastly, a problematic issue related to processing data for the purpose of scientific research is represented by the transfer of health data to third countries. Taking into account that public entities and organizations at the international level have been encouraging the exchange of information also in the light of the COVID-19 pandemic in order to obtain the best version of the vaccine and such information exchange may be implemented through dialog sharing info between the public health authorities around the world, this brings us to the problem of data transfers outside EU countries (and when we use the word "problem," we bear in mind the fact that the privacy shield has been invalidated last summer by the CJUE) (CJUE Judgement, 2020). In general, when considering transfers of health data to third countries or international organizations (like WHO) (WHO Official Website, 2021), the exporters should assess the risks posed to the rights and the freedoms of data subjects and envisage solutions that could ensure data subjects the enjoyment and protection of their fundamental

rights and safeguards as regards the processing of their data, even after it has been transferred, for example, transfers to countries that have an adequate level of protection. So, in this regard, the lawful transfer of data, the EDPB (EDPB Guidelines, 2020a), outlines the potential reliance on the derogations under Article 49 of the GDPR, in lack of adequacy decisions or appropriate safeguards. While acknowledging the current exceptional health crisis, the Board suggests allowing transfers to third countries that are necessary "for important reasons of public interest" under Article 49 letter d, as well as to explicit consent under letter a.

And since we brought to attention the pandemic context, we are going to focus on his impact on health data in the following section.

The Impact of the Pandemic on Health Data

The health crisis that we are facing has led some governments to adopt measures restricting individual rights and freedoms. This crisis thus involves the interference of public or private organizations in the privacy of individuals and poses legal difficulties, particularly in terms of the protection of personal data and respect for privacy.

On March 16 and 19, the European Data Protection Board (EDPB) issued an opinion on the subject of the processing of personal data in the context of the COVID-19 epidemic (EDPB Guidelines, 2020a).

Andrea Jelinek, President of the EDPB, stated that *Data protection rules (such as GDPR) do not hinder measures taken in the fight against the coronavirus pandemic. However, I would like to stress that, even in these exceptional times, the controller must ensure the protection of the personal data of the data subjects. Therefore, a number of considerations must be taken into account to ensure the lawful processing of personal data* (EDPB Statement, 2020).

Contrary to Andrea Jelinek, who mainly focused on allowing the processing of data in the pandemic context, the UN Secretary-General put an emphasis in his policy brief on human rights and COVID-19 that *Human rights are key in shaping the pandemic response, both for the public health emergency and the broader impact on people's lives and livelihoods. Human rights put people center stage. Responses that are shaped by and respect human rights result in better outcomes in beating the pandemic, ensuring healthcare for everyone, and preserving human dignity* (EDPB Guidelines, 2020a).

For the processing of electronic communication data, such as mobile location data, the EDPB points out (EDPB Guidelines, 2020b) that additional rules apply, like Directive 2002/58/EC (hereinafter ePrivacy Directive) (ePrivacy Directive, 2002) concerning the processing of personal data and the protection of privacy in the electronic communications sector. According to this directive, this kind of processing is based on the principle that location data can be used by the operator only after anonymization or with the consent of the subscriber.[10] However, the strategy adopted by certain countries, such as South Korea, to fight against COVID-19, consisted in systematically searching for the relatives of all infected people, thus using these data in an abusive way (for example, the movements of patients, before they tested positive, were reconstructed through video surveillance images, the use of their credit card, or the demarcation of their mobile phone, then made public) (Data Guidance Website, 2021). This strategy raises obvious questions about the protection of privacy.

With a view to the possible implementation of a similar strategy in order to prevent the spread of the virus by one of the EU Member States, the EDPB specifies (EDPB Guidelines, 2020b) that when it is not possible to process only anonymous data, Article 15[11] of the ePrivacy Directive allows Member States to introduce legislative measures aimed at national security and public safety. This emergency legislation must constitute a

[10] *Article 6(1) Traffic data relating to subscribers and users processed and stored by the provider of a public communications network or publicly available electronic communications service must be erased or made anonymous when it is no longer needed for the purpose of the transmission of a communication without prejudice to paragraphs 2, 3, and 5 of this Article and Article 15(1).*

[11] *Article 15(1) Member States may adopt legislative measures to restrict the scope of the rights and obligations provided for in Article 5, Article 6, Article 8(1), (2), (3), and (4), and Article 9 of this Directive when such restriction constitutes a necessary, appropriate, and proportionate measure within a democratic society to safeguard national security (i.e., State security), defense, public security, and the prevention, investigation, detection, and prosecution of criminal offences or of unauthorized use of the electronic communication system, as referred to in Article 13(1) of Directive 95/46/EC. To this end, Member States may, inter alia, adopt legislative measures providing for the retention of data for a limited period justified on the grounds laid down in this paragraph. All the measures referred to in this paragraph shall be in accordance with the general principles of Community law, including those referred to in Article 6(1) and (2) of the Treaty on European Union.*

necessary, appropriate, and proportionate measure in a democratic society. If such measures are introduced, the Member State is required to put in place adequate guarantees, such as granting individuals the right to a judicial remedy. On the other hand, the application of such measures should be limited in time and limited by their purpose. The problem here resides in the fact that while some of these measures or systems that were put in place proved to be useful, others proved to be abusive and represent an intrusive interference with people's right to privacy.

A good example in terms of prevention is that of Google, who announced on 3rd April, 2020 the publication of reports detailing the evolution of population movements during the COVID-19 pandemic. The data come from the geolocation of Google Maps users. The aim is *to help authorities understand how social distancing measures such as telecommuting or lockdown can help flatten the curve of the pandemic*, according to the company's blog post. Reports are currently available in 131 countries. The information is presented in the form of statistics (based on "aggregated and anonymized data") and the precise number of visitors or their data are not published: *To protect individual privacy, no personally identifiable information, such as a person's location, contacts, or movements, is made available* (OECD Website, 2021).

In some European countries (for example, Germany and Italy), mobile phone operators have agreed, upon the arrival of the pandemic, to share user location data with the authorities. This practice is possible in particular thanks to telephone demarcation, which makes it possible to indicate the presence of a telephone at a given time near a relay antenna. These anonymized data are supposed to help governments to observe whether the population (especially the people infected with the virus) respects containment or to map the concentrations and movements of customers in risk areas, for example. The method is already applied in China, where, in order to circulate in certain places, citizens must present their "medical QR Code" via the Alipay Health (New York Times Website, 2021) Code application. The platform uses user data to assess, via a color code, the risk of an individual carrying the virus.

Some other applications used with the same goal, that of prevention, are TraceTogether and Pan-European Privacy-Preserving Proximity Tracing (Data Guidance Website, 2021). Developed by the Government Technology Agency of Singapore in collaboration with the Ministry of Health, TraceTogether uses Bluetooth in order to track individuals who have been exposed to the virus. This information is used to identify close

contacts based on the proximity and duration of an encounter between two users and then alerts those who came in contact with someone who has tested positive or is at high risk for carrying the coronavirus. However, once a person is confirmed or suspected to be infected, it is up to them to choose to allow hospitals, the Ministry of Health, and potential third parties to access data in the app to help identify close contacts. Pan-European Privacy-Preserving Proximity Tracing represents an open-source app which also makes use of Bluetooth technology to analyze signals between mobile phones to detect users who have been in close proximity to each other. The application temporarily stores those encrypted data locally and if the users later on test positive for COVID-19, it can alert anyone who has been in contact with the infected individual in the preceding days, while keeping users' identities protected (OECD Official Website, 2021). What is important to be noted here is that both these apps use proximity data [data generated by the exchange of Bluetooth Low Energy (BLE) and not geolocation data (EDPB Guidelines, 2020b)].

Despite their apparent useful use in the pandemic context, the problem with all these apps is that from the moment we authorize the authorities to process our health data, we open Pandora's box, in the sense that we do not know how those collected data will be used, for how long it will be stored, or even if the data are anonymized or not. The amount of health data these apps collect, process, and share could be very broad and difficult for users to understand and they might provide an uninformed consent without understanding all the technicalities of these programs (for example, the app could continue running in the background even when the device is not in use or could exchange information with other apps without the user even knowing it). We must thus bear in mind that tracking apps can embody varying degrees of privacy and data protection. The use of apps can allow data-sharing with explicit, built-in privacy and data protection and enable users to give their explicit, informed consent to the collection and sharing of their personal data (assuming the use of the app is not mandatory). For instance, Singapore's TraceTogether app has a number of privacy safeguards, including that it does not collect or use geolocation data, and data logs are stored in an encrypted form. To protect the privacy of its users, the Pan-European app encrypts data and anonymizes personal information. In addition, as two phones never exchange data directly, it is almost impossible to reveal the identity of users.

Taking a look at all these apps, it seems like we are dealing with more categories of data and of course not all of them represent health data (for

example, location or proximity data do not represent health data). However, their use might lead to obtaining or processing health data (knowing that a person entered into contact with an infected individual might show that the person might have also got infected with the virus, which indeed represents data related to health). As "location data" are protected under the ePrivacy Directive, article 5(1), 6, and 9(11), any information stored in and accessed from user's equipment is protected under article 5(3) of the ePrivacy Directive. And as we previously stated, under the same Directive (article 5), the storing of information on the user's device is allowed only if the user has consented or the storage is strictly necessary for the information society service. The personal data produced through these apps are protected under the GDPR (for example, the information that someone has been infected or the fact that they have temperature). Non-personal data (anonymized data) are not protected under the GDPR (EU Commission Communication, 2020).

In a nutshell, all these applications could prove to be useful. It depends on whether their installation is voluntary or not, whether the consent provided by the user is a "freely given," "specific," "explicit," and "informed" one, or whether the personal data are stored or not. If the data are stored only on the individual's device and not transmitted and processed, then a legal basis under the GDPR is not needed. On the contrary, if the data are being passed on, we emphasize that it should only reach the health authorities as controllers and a legal basis should exist (legal basis that was tackled in the previous section). Furthermore, health authorities should only be provided access to proximity data from the device of the infected person in order to be able to let people at risk know. These health data should be available to them only after the infected person willingly shares these data with the authorities and the identity of the infected person should not be disclosed to the persons with whom he/she has been in epidemiological contact.

Another issue, might we dare add, a trend, related to health data and the pandemic context is the frequent use of facial recognition devices or thermoscanners[12] (especially by employers) in many countries to monitor the spread of COVID-19 and to track citizens who may test positive

[12]According to article (4)(2) of the RGPD, thermoscanning may be a personal data processing activity under the RGPD [1] to the extent that personal data are recorded in a record system.

for COVID-19.[13] As we previously mentioned, body temperature constitutes sensitive data relating to people's health, subject to special protection. In addition, these kinds of data should not be recorded in an automated process or in a paper register, except where a text expressly provides for the possibility or the individual has consented. In addition, only the verification of the temperature by means of a manual thermometer (such as, for example, of the contactless infrared type) at the entrance of a site, without any trace being kept, nor that any other operation is carried out (such as readings of these temperatures or internal/external information feedback), does not fall under data protection regulations.

And since we addressed the technology connectivity aspects and impacts in relation to health data and the risks posed to data's privacy, we also find it extremely important to also stress the risks in terms of data security.

The Risks Posed by Cyberattacks on Health Data

Before delving into the subject matter, terminology should be first approached. What do we understand by Cyberattack and why are we addressing it in relation to health data?

A cyberattack represents a *disruptive cyber incident, data breach, or a disinformation operation conducted by a threat actor using a computer network or system with malicious intention to cause damage (technical, financial, reputational, or other) or steal data without consent* (Cyberpeace Institute, 2021). And since health data (information relating to patients and their diseases, information about the devices used by them, like stents or pacemakers, and results from clinical trials) nowadays are mainly stored on different devices (computers or cloud software), it seems they are an easy target for attackers and cyberattacks on health data do not represent a new phenomenon. Also, the pandemic has been a prolific context for malicious behaviors (vaccine research centers targets of cyber espionage, hospitals held to ransom), being accompanied not only by an acceleration of ransomware attacks against healthcare but also by an increase in data breaches. This increase in cyberattacks due to

[13]For example, in Poland, the government has launched a biometrics smartphone app to confirm that people who are infected with COVID-19 remain under quarantine; available at The Naked Truth About Online Privacy | Nasdaq [accessed on 13 February 2021].

COVID-19 could be perhaps also linked to the increase in remote work (work from home) and the greatly amplified value of vaccine research data.

Cyberattacks, whether in the form of ransomware or cyber espionage (the most common type of cyberattacks when it comes to health data), prove to be particularly dangerous since they put both patient care and healthcare sector in jeopardy (losing access to medical records surely affects professionals' ability to provide care and treatment), and healthcare organizations are gatekeepers of valuable and sensitive data (Cyberpeace Institute, 2021).

Each and every vulnerability or fragility of the digital infrastructure of health care facilities (such as outdated software, lack of anti-virus programs or weak anti-virus protection and outdated medical devices) leaves room for the threat actors to take the floor.

One might ask himself why would hackers want to steal your health data. Well, the answer to this question is quite simple. There is no limit to the uses they could put to those data; attackers could steal your identity using the sensitive data contained in medical records, abuse prescriptions to buy drugs or even benefit from medical interventions, or sell your information on the black market for financial benefits.

And just to provide a few examples, Medjack represents one of the latest methods of accessing a health system's network (Infosec Institute Website, 2021), targeting medical devices that integrate with applications, often through methods that are not highly protected against, creating the appearance that nothing abnormal is occurring, while data are easily stolen. UCLA Health was one of the victims of Medjack, which led to the exposure of personal data, including health data (names, birth dates, Medicare numbers, and health plan numbers) for 4.5 million patients. The facility proved to be an easy target since the patient data were not encrypted when they passed from medical devices to the electronic health record.

In May 2017, the cyberattack on the British health system (NHS) by WannaCry (NBC Official Website, 2021) has led to serious consequences: 16 health centers and 200,000 infected computers led to the cancellation of nearly 20,000 consultations. This malware also crippled more than 1,200 diagnostic equipment. The malware was able to jump from computer to computer by targeting a weakness in older versions of Windows, as well as more recent systems that hadn't been updated. Also, NCH Healthcare System in Naples reported (Naples News Website, 2021) an

email hacking incident on February 17 that exposed 63,581 patients' records and Washington University School of Medicine in St. Louis on 31 March exposing 14,795 patients' records (Becker Hospital Website, 2021). Recently, in July 2020, the Doctolib (Cyberveille-sante.gouv Website, 2021) platform has also been the victim of a cyberattack: more than 6,000 appointment data were stolen (last name, first name, gender, phone numbers, e-mails, addresses, and name of health professional).

A report drafted by the French start-up CybelAngel, specialized in the search for stolen documents on the dark web, highlights a marked interest of cybercriminals in health data. *CybelAngel analysts* (…) *note an upsurge in the search and sale of hospital data on the dark web, demonstrating an interest of malicious actors in the sector* (Yahoo Finance Official Website, 2021).

And turning to the dark web, we need to draw attention to the attack of the Vastaamo Psychotherapy, which provides an example of a potential evolution of double extortion tactics, or what could be referred to as a sort of triple extortion (Ransomware 3.0). After Vastaamo refused to pay 40 Bitcoins (est. EUR 450,000), the attacker began to both leak the health data obtained on the dark net and directly extort the data subjects, namely, the patients themselves (App News Website, 2021; Politico Website, 2021).

While there have been different types of public attacks on healthcare facilities, all resulted in similar negative consequences for both patients and the health system. However, we find that the biggest problem is that patients begin losing trust in providing their personal information to doctors and care institutions, thus increasing the likelihood that sensitive information will be withheld.

Healthcare breaches are especially serious because health data can, in some cases, mean the difference between life and death. For example, it could cause medications to become mixed up or people might fail to get the proper treatment for conditions, such as diabetes. Let us imagine for a second that a cyberattacker manages to gain access to a patient's health data records and this allows him to control his pacemaker/neurostimulator/defibrillators/drug pump. It could stop all these devices just in a few seconds. Although this scenario seems far away, in 2010, researchers demonstrated that they could gain unauthorized remote access to an insulin pump from 100 feet away (Paul *et al*., 2011). They also emphasized that some insulin pump systems also use a mobile phone to help patients monitor their glucose levels and thus an attacker who breached the

security of the mobile phone could be able to use it to change the insulin pump's settings. And if usually emergency and prompt action is needed and vital to save lives, when it comes to healthcare attacks, it is not always possible for stitches and patches to be applied.

Unfortunately, security incidents will continue to multiply in the healthcare industry taking into account that there is always a substantial financial incentive for cyberattackers.

Conclusion

The use of new technologies, the storing of data on the cloud, wearable devices, and engagement through mobile health apps represent the future of medicine now. Their role is aimed at building better health profiles, discovering new treatments, and making it easier for individuals themselves to assess their current state of health so that we can better diagnose and treat diseases.

However, the use of new technologies raises a number of privacy and security concerns, particularly when being used in the absence of specific guidance or fully informed and explicit consent. Also, the outdated systems, devices, machines, or poorly protected generate serious consequences in terms of threats posed to security.

Organizations and health care facilities commonly believe that keeping sensitive data secure from hackers means they are automatically compliant with data privacy regulations. But as we showed in the present paper, data privacy and data security are two separate issues, and both need to be taken into consideration when dealing with health data. Moreover, the technical and organizational measures implemented must ensure their security is appropriate to the level of risk. In addition, encrypting or anonymizing health data both in transition and at rest is highly recommended, noting that companies in the health system have lately become easy targets for cyberattackers.

Leveraging health data adds both benefits and challenges. Taking into account their sensitive character and their infinite value for people's lives, it is of utmost necessary, on one hand, to make sure that legal provisions are respected and human rights are not seriously affected by their storing and processing and, on the other hand, to protect these data against data breaches. And in the light of the pandemic context, it is important to stress that the outbreak does not suspend or restrict the possibility of data subjects to exercise their rights pursuant to GDPR.

References

App News Website (2021). Finland shocked by therapy center hacking, client blackmail (apnews.com); Politico Website (2021). Hacker seeks to extort Finnish mental health patients after data breach — POLITICO.

Becker Hospital Website (2021). 28 health system cyberattacks, data breaches so far in 2020 (beckershospitalreview.com).

Communication from the Commission Guidance on Apps supporting the fight against COVID 19 pandemic in relation to data protection, C(2020) 2523 final. 5_en_act_part1_v3.pdf (europa.eu). Accessed on 13 February 2021.

CrowdSignal Official Website (2021). Crowdsignal | Surveys, Polls, and Quizzes | Get the responses you need, anywhere.

Cyberpeace Institute (2021). Playing with lives: Cyberattacks on healthcare are attacks on people. Publications — CyberPeace Institute. Accessed on 24 April 2021.

Cyberveile-sante.gouv.fr. Official Website. Fuite de données concernant le service de prise de rendez-vous Doctolib | Accompagnement Cybersécurité des Structures de Santé (cyberveille-sante.gouv.fr).

Data Protection Commissioner v Facebook Ireland Ltd and Maximilian Schrems (2020). ECLI:EU:C:2020:559, (European Court of Justice).

Data Guidance Website (2021). International: Privacy implications of Coronavirus tracking mobile apps | DataGuidance.

Directive of the European Parliament and of the Council 2002/58/EC of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector, OJ L 201, pp. 37–47.

EDPB Guidelines (2020a). Guidelines of EDPB, no. 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak, adopted on 21 April 2020. edpb_guidelines_202003_healthdatascientificresearchcovid19_en.pdf (europa.eu).

EDPB Guidelines (2020b). Guidelines of the EDPB, no. 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak, adopted on 21 April 2020. Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak | European Data Protection Board (europa.eu).

EDPB Official Website. About EDPB | European Data Protection Board (europa.eu).

European Commission Official Website. Clinical trials | Public Health (europa.eu).

European Parliament and Council Regulation (EU) no. 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ L 119, pp. 1–88.

Infosec Institute Website (2021). The 5 Most Visible Cyber Attacks on Hospitals — Infosec Resources (infosecinstitute.com).

Naples News Website (2021). NCH hires forensic investigators to probe email phishing attack; warns staff (naplesnews.com). Accessed on 4 February 2021.

NBC Official Website (2021). Why "WannaCry" Malware Caused Chaos for National Health Service in U.K. (nbcnews.com).

New York Times Official Website (2021). In Coronavirus Fight, China Gives Citizens a Color Code, With Red Flags — The New York Times (nytimes.com).

OECD Official Website (2021).Tracking and tracing COVID: Protecting privacy and data while using apps and biometrics (oecd.org).

Opinion of EDPB, no. 3/2019 of 23 January 2019 on concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection regulation (GDPR). Available at Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection regulation (GDPR) | European Data Protection Board (europa.eu).

Paul, N., Kohno, T. and Klonoff, D. C. (2011). A review of the security of insulin pump infusion systems. *Journal of Diabetes Science and Technology*, 5(6), 1557–1562. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3262727.

Rawassizadeh, R., Momeni, E., Dobbins, C., Mirza-Babaei, P. and Rahnamoun, R. (2015). Lesson learned from collecting quantified self-information via mobile and wearable devices. *Journal of Sensor and Actuator Networks*, 4(4), 315–335. Available at https://doi.org/10.3390/jsan4040315.

Regulation of the European Parliament and of the Council (EU), no. 536/2014 of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC, OJ L 158, pp. 1–76.

Statement on the processing of personal data in the context of the COVID-19 outbreak, adopted on 19 March 2020. Available at EDPB (europa.eu).

Vitabile, S. (2019). Medical data processing and analysis for remote health and activities monitoring. In Kołodziej, J. and González-Vélez, H. (Eds.), *High-Performance Modelling and Simulation for Big Data Applications*. Lecture Notes in Computer Science, vol. 11400. Cham: Springer. https://link.springer.com/chapter/10.1007%2F978-3-030-16272-6_7. Medical Data Processing and Analysis for Remote Health and Activities Monitoring | SpringerLink.

Wachter, S. and Mittelstadt, B. (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review*, 2019(2), 443–493.

WHO Official Website (2021). WHO | World Health Organization.

Yahoo Finance Official Website (2021). Cybe: le darknet, supermarché des données de santé françaises (yahoo.com).

https://doi.org/10.1142/9781800612860_0012

Chapter 11

Hate Speech: A Comparative Analysis of the United States and Europe

Margaret M. McKeown and Dan Shefet

The old adage "sticks and stones may break my bones but words will never hurt [me]" seems particularly antiquated in the era of the Internet and social media (Rolfe and Schroeder, 2020, p. 3463). Hate speech in the form of racist, sexist, anti-religious, and homophobic remarks is not mere words but language meant to derogate and dehumanize. Indeed, violence related to hate speech is on the rise (Laub and Council on Foreign Relations, 2019) and hate crimes often correlate with hate speech (Relia *et al*., 2019). Despite efforts to combat hate speech, both the United States ("US") (Criminal Justices Information Services Division, 2019) and Europe (FRA) are experiencing a rise in reported incidents of hate crimes. The COVID-19 pandemic has only worsened these issues. In the US, insults and epithets have been hurled at Black, Hispanic, Asian, and Native American individuals (Akee *et al.*, 2021). In Europe, for example, in Estonia, a Malaysian girl who was wearing a mask was shouted at and blamed for bringing the coronavirus into the country (European Network Against Racism, 2020). And teenagers in Poland attacked, threw garbage at, and spit at a Vietnamese woman (European Network Against Racism, 2020).

Efforts to curb hate speech alternatively are criticized as censorship and celebrated as beneficial regulatory policy. The gulf between hate speech regimes is nowhere more evident than between the US and

Europe. Stemming from the First Amendment to the Constitution, the US venerates free speech and generally applies the principle that "the mere fact that even the vast majority of the community absolutely hates and loathes the message is not a justification for censoring it" (Oxford Academic, 2018). Government and judicial oversight of hate speech is limited. By contrast, Europe takes a more regulatory approach, outlawing and moderating certain content.

We begin this chapter by addressing the definitions of hate speech and hate crimes, followed by a discussion of the US's approach and fleshing out the European framework. We then consider where the two regimes have converged through self-regulation by technology companies. Finally, we consider the next frontier in hate speech regulation and legislation and the public policy challenge of embracing the freedom of expression while underscoring the importance of combating the proliferation of hateful and discriminatory conduct.

What Is Hate Speech?

Although hate speech is generally thought to encompass a wide range of written, verbal, and symbolic expressions that provoke hatred on the basis of race, ethnicity, religion, gender, sexual identity, gender orientation, and disability, the characterization of what is "hateful" is disputed and controversial, making it difficult to reach consensus on the meaning of term (United Nations, 2019). Perhaps unsurprisingly, there is no universal definition of hate speech under international law (United Nations, 2019).

The United Nations ("UN") (2020, p. 10) has proffered a working definition, categorizing hate speech as "[a]ny kind of communication in speech, writing, or behavior, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, color, descent, gender, or other identity factors." The UN (2010, p. 10) further specifies that this speech is "often rooted in, and generates, intolerance and hatred and, in certain contexts, can be demeaning and divisive."

Several international treaties incorporate concepts of hate speech as speech that incites racial discrimination or genocide. Among them are the Convention on the Prevention and Punishment of the Crime of Genocide (1948), the International Convention on the Elimination of All Forms of Racial Discrimination (1965), the International Covenant on Civil and

Political Rights (1966), and the Additional Protocol to the Convention on Cybercrime (2013) (Shefet, 2018, p. 3).

In addition to international protocols, the European Commission has promulgated a Code of Conduct (2016b), adopted by technology companies, for the companies' self-regulation of illegal online hate speech. Illegal hate speech under European Union ("EU") law is "the public incitement to violence or hatred on the basis of certain characteristics, including race, color, religion, descent, and national or ethnic origin" (European Commission, 2016b, p. 1). Beyond the Code of Conduct, many EU Member States have outlawed hate speech directed toward "sexual orientation, gender identity, and disability" (European Commission, 2016b, p. 1).

By contrast, across the pond, the US has neither enacted a federal hate speech law nor promulgated a legal definition of hate speech (Jawa, 2020) rather even offensive speech is broadly protected under the First Amendment of the US Constitution. Discriminating against speech based on the viewpoint expressed is presumptively unconstitutional (*R.A.V. v. City of St. Paul*, 1992), adding a significant hurdle for legislative efforts to define and regulate hate speech, which necessarily involves consideration of the ideas expressed in the speech. Even so, it is well accepted that it is permissible to prohibit some types of speech, such as "true threats," so long as the restrictions do not discriminate based on viewpoint (*R.A.V. v. City of St. Paul*, 1992). Furthermore, as will be discussed further in the following, private technology corporations have adopted — and enforced — their own definitions of hate speech.

Relationship Between Hate Speech and Hate Crimes

Although it may be difficult to overcome constitutional infirmities of hate *speech* laws, the US Congress has criminalized *acts* of hatred. Congress passed the first federal hate crimes statute in 1968 (United States Department of Justice, 2019). Protections have been expanded since. Now, federal law criminalizes certain offenses that are committed because of "actual or perceived race, color, religion, or national origin" and because of "actual or perceived religion, national origin, gender, sexual orientation, gender identity, or disability" (Hate Crime Acts). Most states, too, have their own hate crime laws (United States Department of Justice, 2021).

Hate crimes include both crimes against persons and crimes against property (Criminal Justice Information Services Division, 2019). Offenses vary widely from murder, rape, and assault to burglary and vandalism. There are some differences in how government agencies measure hate crimes, for example, the Federal Bureau of Investigation's ("FBI") Uniform Crime Reporting Program includes homicide and vandalism, while the Bureau of Justice Statistics' National Crime Victimization Survey does not (Bureau of Justice Statistics, 2017). The Criminal Justice Information Services Division of the FBI (2019) includes offenses as hate crimes when the actions were motivated by bias. In the most recent data, race/ethnicity/ancestry bias motivated the majority of hate crimes. And the majority of hate crimes were crimes against persons — with intimidation as the most common crime against persons.

Conceptually, hate crimes do not separate cleanly from hate speech, as hate crimes often include expressive elements (Lawrence, 1993). Hateful speech during the offense may help establish that the offense qualifies as a hate crime (Bureau of Justice Statistics, 2017). And it is not uncommon that offenders use hate language or leave hate signs or symbols at the scene of the offense (Bureau of Justice Statistics, 2017). Due to the close relationship, hate speech and hate crimes often arise in conjunction with each other (Relia *et al*., 2019). For this reason, economic research suggests that "raising the costs of hate speech will tend to deter hate crime[s]" (Dharmapala and McAdams, 2005, p. 132).

Despite the often-entangled relationship between hate speech and hate crimes, US courts have routinely upheld hate crime laws. In *Wisconsin v. Mitchell* (1993, pp. 484–487), the US Supreme Court considered a Wisconsin statute that penalized criminal conduct more harshly where "the victim is selected because of his race or other protected status" than the same conduct without that motivation. The Court held that the statute did not conflict with the First Amendment, as it targeted conduct, not speech.

The European Court of Human Rights generally permits restrictions on both hate crimes and hate speech (Bayer and Bárd, 2020). Indeed, it goes so far as to obligate states to take action against hate crimes (Bayer and Bárd, 2020). As in the United States, legislatures in Europe have adopted criminal provisions or penalty enhancements for crimes motivated by bias (Bayer and Bárd, 2020).

The US Approach to Hate Speech

Understanding the US approach to hate speech requires a short lesson in history. The First Amendment of the US Constitution, which was added in 1791, just shortly after the Constitution was adopted, provides that "Congress shall make no law . . . abridging the freedom of speech." The First Amendment, which also protects the free exercise of religion, freedom of the press, and the right to assemble, applies only to government action, not to private actors. Although the text has not changed since adoption, over time courts have clarified its meaning.

Early cases

Throughout the 19th century, the Free Speech Clause of the First Amendment received scant attention from the US Supreme Court (Wallmeyer, 2003). The World War I era brought the First Amendment doctrine to the fore (Wallmeyer, 2003; Wilson and Kiper, 2020). Congress passed the Espionage Act of 1917, which prohibited certain activities when the US was at war. In *Schenck v. United States* (1919), defendants convicted for mailing leaflets that criticized the draft challenged the constitutionality of the Act. Justice Holmes, writing for a unanimous court, upheld the convictions because of the wartime context but broadened First Amendment speech protections. To determine whether speech is protected, courts had to evaluate "whether the words used are used in such circumstances and are of such a nature as to create a clear and present danger that they will bring about the substantive evils that Congress has a right to prevent" (p. 52).

That consensus would not last long. The same year, again reviewing convictions under the Espionage Act, the Court fractured, with the majority in *Abrams v. United States* (1919) upholding convictions for distributing inflammatory leaflets. Justice Holmes, joined by Justice Brandeis, penned an influential dissent, arguing that the First Amendment protects the right to dissent from the government's viewpoint. He introduced the now famous marketplace of ideas framework, asserting that "the best test of truth is the power of the thought to get itself accepted in the competition of the market" (p. 630). Therefore, courts must be "vigilant against attempts to check the expression of opinions that we loathe and believe to be fraught with death, unless they so imminently threaten immediate interference with the lawful and pressing purposes of the law that an immediate check is required" (p. 630).

After World War I, the Court was faced with First Amendment challenges arising out of the "Red Scare" following the Russian revolution (Fisch, 2002). The Court's fractured perspective continued during this period. In *Gitlow v. New York* (1925) and *Whitney v. California* (1927), the Court upheld the constitutionality of prosecutions under state laws prohibiting advocacy of unlawful overthrow of the government and the use of unlawful means to achieve political change. But Justices Holmes and Brandeis wrote separately in those cases, continuing to push for more robust speech protections. Nonetheless, the Court continued to apply the "clear and present danger" test in those cases and in a series of cases that arose during the Cold War (Fisch, 2002).

At the same time, as the Court was evaluating anti-government speech and perceived threats to national security, it was also considering challenges to the breach of the peace statutes. These cases focused on speech that maligned other citizens based on their identities or beliefs.

In the seminal breach of the peace case, *Chaplinsky v. New Hampshire* (1942), the Court held that some categories of speech did not receive First Amendment protection. Chaplinsky called a city marshal "a God damned racketeer" and "a damned fascist" while being escorted away from a restless crowd (pp. 569–570). He was convicted under a state law providing that "[n]o person shall address any offensive, derisive or annoying word to any other person who is lawfully in any street or other public place, nor call him by any offensive or derisive name" (p. 569). The Court upheld the conviction, explaining that "certain well-defined and narrowly limited classes of speech," including "the lewd and obscene, the profane, the libelous, and the insulting or 'fighting' words," could be regulated despite the First Amendment because they "are no essential part of any exposition of ideas and are of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality" (pp. 571–572).

Along this vein, the Supreme Court in *Beauharnais v. Illinois* (1952) briefly entertained one manner of regulating hateful speech — group libel statutes. Based on his distribution of leaflets calling for white supremacist violence, Beauharnais was convicted of publishing an exhibition that "portrays depravity, criminality, unchastity, or lack of virtue of a class of citizens of any race, color, creed or religion which said … exhibition exposes the citizens of any race, color, creed, or religion to contempt, derision, or obloquy of which is productive of breach of the peace or riots" (p. 251). In upholding the conviction, the Court added that unprotected categories

of speech, like libel, may be prohibited without an additional determination that the speech presents a "clear and present danger" (p. 266).

While the World War I, Red Scare, and Cold War cases saw the Court rejecting various free speech challenges, the following decades reflected a marked shift. In the 1960s and 1970s, amidst civil rights marches, anti-Vietnam War protests, and widespread social upheaval, the Court expanded free speech protection, particularly for political speech — and even for hateful speech.

In *New York Times Co. v. Sullivan* (1964, p. 283), a landmark case from that era, the Court wrote that when libel actions are "brought by public officials against critics of their official conduct," a showing of "actual malice" is required. Importantly, the Court lauded broad speech protections: "debate on public issues should be uninhibited, robust, and wide open," even where that includes "vehement, caustic, and sometimes unpleasantly sharp attacks on government and public officials," given the benefits of "'unfettered interchange of ideas for the bringing about of political and social changes desired by the people'" (pp. 269–270). Quoting Justice Brandeis, the Court endorsed the view that "it is hazardous to discourage thought, hope, and imagination; that fear breeds repression; that repression breeds hate; that hate menaces stable government; that the path of safety lies in the opportunity to discuss freely supposed grievances and proposed remedies; and that the fitting remedy for evil counsels is good ones" (p. 270).

Just a few years later, the importance of open political debate came to the fore. The Court overturned the conviction of a war protester who proclaimed at a public rally that if he were ever made to carry a rifle, "the first man [he] want[ed] to get in [his] sights [was] [President Johnson]" (*Watts v. United States*, 1969, p. 706). The Court concluded that the statement was not a true threat but rather a crude expression of political opposition and, therefore, was constitutionally protected.

Later that same year, the Court continued its recognition of broad constitutional protections for political speech. In the watershed case of *Brandenburg v. Ohio* (1969), the Court shifted away from the clear and present danger test. A Ku Klux Klan leader in a recorded video warned that "if our President, our Congress, our Supreme Court, continues to suppress the white, Caucasian race, it's possible that there might have to be some revengeance taken" (p. 446). Reasoning that the First Amendment protects the "mere abstract teaching" of force and violence, which differs from actual "incitement of imminent lawless action," the Court reversed the conviction (pp. 447–449). It thus became more difficult for legislatures

to outlaw hateful speech advocating law-breaking or the use of force, "except where such advocacy is directed to inciting or producing imminent lawless action and is likely to incite or produce such action" (p. 447).

Contemporary trends

The Supreme Court's view of expansive protections for speech has continued into the modern era, even in the face of other examples of hateful speech. In *Nationalist Socialist Party of America v. Village of Skokie* (1977), the Court famously reinforced this principle in connection with the National Socialist Party's effort to demonstrate in a Chicago suburb with a large Jewish population. The state court had enjoined the group from marching, walking, or parading in uniform, displaying the swastika, or "[d]istributing pamphlets or displaying any materials which incite or promote hatred against persons of Jewish faith or ancestry or hatred against persons of any faith or ancestry, race or religion" (p. 43). In a sparse decision, the Court required the injunction be stayed, reasoning that "[i]f a State seeks to impose a restraint of this kind, it must provide strict procedural safeguards" (p. 44).

More than a decade later, in *R.A.V. v. City of St. Paul* (1992, pp. 379–380), the issue of racist speech arose again when a teenager who allegedly burned a cross in a Black family's yard was charged under a local ordinance that prohibited placing a symbol "which one knows or has reasonable grounds to know arouses anger, alarm, or resentment in others on the basis of race, color, creed, religion, or gender." The Court struck down the ordinance. Although it acknowledged that the First Amendment allows some regulation of speech, such as reasonable "time, place, or manner" restrictions, it held that this ordinance was unconstitutional because governments may not "impose special prohibitions on those speakers who express views on disfavored subjects" (p. 391).

Following *R.A.V.*, it became more difficult to restrict hate speech, as the Court moved away from the mechanical and categorical approach to a context-driven approach. Although *Beauharnais v. Illinois* (1952) was not expressly overruled, commentators generally agree that the Court cut off the avenue of regulating hate speech through group libel statutes (Smolla, 2021; Rich, 2020).

Despite this trend, the Court did not wholly foreclose the regulation of hate speech. In *Virginia v. Black* (2003), it carved out space for hate speech regulation by applying the "true threat" doctrine. States could ban

cross burning "with intent to intimidate" because "a prohibition on true threats protects individuals from the fear of violence and from the disruption that fear engenders, in addition to protecting people from the possibility that the threatened violence will occur" (p. 360). Unlike in *R.A.V.*, the statute at issue did not single out *why* the cross was burned rather it regulated the intimidating message broadly.

The Court has continued to reaffirm broad protections for speech, including hateful speech. In *Snyder v. Phelps* (2011), members of the Westboro Baptist Church, carrying offensive signs, picketed on public land near the funeral of a military veteran. The deceased soldier's family sued Church members. The Court explained that the First Amendment could shield speakers from tort liability based on the *public significance* of the speech. Considering the "content, form, and context" of the speech, the Court concluded that "[w]hile these messages may fall short of refined social or political commentary, the issues they highlight[ed] … [were] matters of public import" (pp. 453–454). This, then, was an endorsement of a hierarchical view of protected speech, where speech "at a public place on a matter of public concern … cannot be restricted simply because it is upsetting or arouses contempt" (p. 458).

The Court recently reaffirmed protections for offensive speech in *Matal v. Tam* (2017). The disparagement clause of the federal trademark law prohibited the registration of trademarks "which may disparage . . . persons, living or dead, institutions, beliefs, or national symbols, or bring them into contempt, or disrepute" (p. 1753). Under that provision, an Asian-American musical group was denied a trademark for its name — "The Slants." The lead singer explained that the use of the slur was intentional, designed to reclaim the offensive term. In striking down the provision under the First Amendment, the Court determined that "[g]iving offense is a viewpoint" (p. 1763). The Court noted that it has "said time and time again that 'the public expression of ideas may not be prohibited merely because the ideas are themselves offensive to some of their hearers'" (pp. 1763–1765).

Out of the long and winding history of free speech doctrine, a few basic principles loom large and define the contemporary approach to hate speech in the US. To begin, freedom of speech is a fundamental constitutional right and governmental restriction of that right is tightly circumscribed. Although the Constitution permits certain regulations of the "time, place, and manner" of speech, speech cannot be restricted merely because the government disagrees with the viewpoint expressed, no matter how vile or hateful. Nonetheless, the government may, in some

circumstances, restrict speech, including speech that constitutes "true threats" or "incitement to imminent lawless action."

Legislative protection of speech — The Communications Decency Act

As Internet access began its exponential expansion in the 1990s, the question of whether service providers would be held responsible under libel and other laws for content published on Internet platforms began to make its way through the legal system. After a court deemed Prodigy Services, an early Internet provider, liable for a defamatory post that it did not create but failed to delete (*Stratton Oakmont, Inc. v. Prodigy Services Co.*, 1995), debate on the issue exploded. In 1996, in an express repudiation of that case, Congress passed the Communications Decency Act ("CDA") as Title V of the Telecommunications Act of 1996 (House of Representatives, 1996).

The now-infamous Section 230 of the CDA, 47 U.S.C. § 230(c)(1), immunizes Internet platforms from being treated like publishers. Despite this safe-harbor provision, Internet platforms have endured their share of lawsuits from private citizens attempting to hold the platforms liable for posted content (*Dyroff v. Ultimate Software Group, Inc.*, 2019). Consistently, however, courts have held that Section 230 shields platforms from liability for the content that is created by a third party although it does not shield a platform that itself develops or creates the content (*Bennett v. Google, LLC*, 2018; *Force v. Facebook, Inc.*, 2019; *Chicago Lawyers' Committee for Civil Rights Under Law, Inc. v. Craigslist, Inc.*, 2008). Since platforms are not responsible for the content they host, academics (Citron and Franks, 2020) and politicians have criticized Section 230 as effectively fostering the proliferation of hate speech in those spaces — providing a "'Get Out of Jail Free' card to the largest platform companies" (Warner and Senate, 2021).

Section 230 also protects providers from liability when they take action "in good faith" against content that is "obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable." This provision thus incentivizes self-regulation by the Internet platforms (*Bennett v. Google, LLC*, 2018). Indeed, as will be discussed further in the following, many platforms have promulgated terms of use restricting the publication of hate speech (Facebook, 2021; Twitter, 2021; YouTube and Google 2021). This twist on speech regulation represented federal efforts

to protect service providers' ability to regulate as they saw fit rather than to directly intervene to regulate Internet speech.

The European Approach to Hate Speech

The regulatory tradition

Europe's tradition of speech regulation stands in contrast to that of the US. Many European countries began moderating content in earnest after World War II and the International Military Tribunal at Nuremberg. A foundational case involved Julius Streicher, a publisher and editor of the anti-Semitic newspaper, "Der Stürmer," who was convicted at Nuremberg of crimes against humanity for his incitement of persecution (*The Nurnberg Trial 1946*, 1946–1947). In its incitement cases, the International Military Tribunal considered whether the individual on trial actually exercised control over the content. In contrast with Streicher, Hans Fritzsche, a Nazi propaganda official, was acquitted by the Tribunal on the basis, among others, that he did not originate or formulate the propaganda campaigns.

The Nuremberg trials laid the groundwork for the European content regulation to come (Bayer and Bárd, 2020). The horrors of the Holocaust led the UN to promulgate the Convention on the Prevention and Punishment of the Crime of Genocide in 1948, and the Holocaust continues to inform the national legislation of many European countries (Bayer and Bárd, 2020). These foundations also persist in international law. The Streicher and Fritzsche judgments influenced *Prosecutor v. Šešelj* (2016), a case before the International Criminal Tribunal of the former Yugoslavia, where the question of causation was key in arriving at the trial judgment. Control was the key issue in the Fritzsche case, while causation was the main legal challenge in *Šešelj*.

Additionally, the laws of EU Member States are based on the International Covenant on Civil and Political Rights Article 20, which has been interpreted as creating a positive obligation on states to prevent incitement (Temperman, 2019). EU Member States also refer to the European Convention for the Protection of Human Rights and Fundamental Freedoms (1950), Article 10, which provides for the right to freedom of expression but recognizes that freedoms "may be subject to such formalities, conditions, restrictions, or penalties as are prescribed by law and are necessary for a democratic society, in the interests of

national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary." Article 11 of the Charter of Fundamental Rights of the European Union (2012) also provides for freedom of expression and information. In promulgating speech regulations, the principles of protecting the rights of others and preventing incitement are balanced with the principles of freedom of expression and freedom of information.

Today, nearly all EU Member States have passed hate speech laws, although they are not uniform (European Commission, 2020b). The laws vary on whether the list of protected grounds is closed (that is, exhaustive) or open, and on which characteristics are protected (Bayer and Bárd, 2020). Latvia, for example, has criminalized acts directed at inciting racial, national, ethnic, or religious hatred in Section 78 of its Criminal Code, but sexual orientation is not included among the protected characteristics (FRA, 2018). Similarly, Poland has a closed list of protected grounds that do not include sexual orientation (Bayer and Bárd, 2020). By comparison, in 2018, 21 Member States had expressly included sexual orientation as a protected characteristic: Austria, Belgium, Croatia, Cyprus, Denmark, Estonia, Finland, France, Greece, Hungary, Ireland, Lithuania, Luxembourg, Malta, the Netherlands, Portugal, Romania, Slovakia, Slovenia, Spain, and Sweden (FRA, 2018).

Whereas the US's ability to restrict hateful speech after *Brandenburg* turns in large part on the likelihood of violence, European domestic courts frequently treat the speech itself as actionable. The famous French hate speech law, *Loi Gayssot* (Loi 90-615), passed in 1990, is a prime example of such a law. It prohibits the denial of the Holocaust and other crimes against humanity. A French Professor convicted under the law for his statements denying extermination by gas chambers during the Holocaust challenged the law as a violation of his human rights, contending that it compromised his freedom of expression and academic freedom (*Faurrison v. France*, 1996). The Human Rights Committee considered the case and concluded that the conviction did not violate the Professor's right to freedom of expression because he was convicted for violating the rights and reputations of others and because the restriction on his right to freedom of expression was necessary to combat anti-Semitism and racism. Another example of such a

law is Section 130 of the German Criminal Code, which criminalizes incitement to hatred.

Incitement is an inchoate offense because it does not require the person accused of incitement, or anyone else, to actually commit the crime incited (Timmermann, 2006) rather the hate speech itself is an act of violence. That is, in the hate speech context, incitement to hatred is criminalized without requiring any further violence down the line. This general approach does not hold true for all states, however. In Hungary, for example, there must be a realistic possibility of the occurrence of violence, and in Italy, real danger must result (Bayer and Bárd, 2020).

The criminalization of mere speech has been challenged on the international level. The International Law Commission's 1996 report to the UN on the Draft Code of Crimes Against Peace and Security of Mankind generally requires the actual occurrence of a crime, or attempt, as a condition for criminal responsibility for incitement. Article 2(3)(f) of the draft provides that an individual shall be responsible for an international crime, such as the crime of genocide, crimes against humanity, war crimes, and crimes against UN and associated personnel, if that individual "[d]irectly and publicly incites another individual to commit such a crime which in fact occurs." The report explains that a consistent interpretation is made under the UN Convention on the Prevention and Punishment of the Crime of Genocide, Article III(c). And a similar interpretation is made by the Committee on the Elimination of Racial Discrimination (2013) in its "General Recommendation No. 35," which suggests that states consider as an element of the offense "the imminent risk or likelihood that the conduct desired or intended by the speaker will result from the speech in question." From the perspective of international law, these divergent regimes create legal uncertainty.

Moderation of online content

As a natural extension of its regulatory tradition, Europe has applied speech restrictions also to online content, including the provision for indirect liability of Internet platforms. In 2000, European legislators introduced Directive 2000/31/EC, the E-Commerce Directive, which is considered to be the European counterpart to the CDA and regulates information society services in the Member States. Mindful of the distinction between authoring speech and spreading it, the E-Commerce Directive includes an "actual knowledge" standard, whereby providers of information society services are not liable if they do not have actual

knowledge of the illegal activity or of "facts or circumstances from which the illegal activity or information is apparent." The Directive has often been "used to argue that intermediaries are exempted from liability for third-party content" (Bayer and Bárd, 2020, p. 68). The country-of-origin principle also applies, meaning that providers are subject to the law of the EU Member State in which the provider is established, and intermediaries are not liable if they meet certain conditions (European Commission, 2020a).

Caselaw and tensions between the European Court of Justice and the E-Commerce Directive have led to a lack of clarity, however, on intermediary liability. Article 15 of the E-Commerce Directive provides that "Member States shall impose neither a general obligation on providers … to monitor the information which they transmit or store nor a general obligation actively to seek facts or circumstances indicating illegal activity." But it does allow Member States to require providers to inform public authorities of alleged illegal activities. In *SABAM v. Netlog NV* (2012), the judgment interpreted the prohibition against monitoring obligations in very broad terms: "the injunction to install the contested filtering system is to be regarded as not respecting the requirement that a fair balance be struck between, on the one hand, the protection of the intellectual-property right enjoyed by copyright holders and, on the other hand, that of the freedom to conduct business enjoyed by operators, such as hosting service providers." In *Eva Glawischnig-Piesczek v. Facebook Ireland Limited* (2019), however, the European Court of Justice stated that the directive does not preclude a host provider, such as Facebook, from being ordered to remove content that is equivalent to content that had been previously declared to be illegal.

Since the issuance of the E-Commerce Directive, a few European countries have passed their own laws regulating online content (ARTICLE 19, 2018). For example, Germany in 2017 passed the Network Enforcement Act ("NetzDG") (2017). NetzDG made online platforms subject to massive fines for failing to remove unlawful content. Commentators have described the law as increasing the pressure on third-party content providers to weed out hate speech (Center for Democracy & Technology, 2017). Oversight in Germany is conducted by a relatively new body, the "Internet Complaints Office" (*Internet-beschwerdestelle.de*, 2021). Additionally, under NetzDG, the Federal Office of Justice may impose the regulatory fines.

In France, legislators in 2020 passed the *Loi Avia* (Loi 2020-766). Oversight in France is conducted by the *Conseil supérieur de l'audiovisuel* ("CSA"). The CSA is an independent public authority, created by a law dating back to 1989, whose function is to "guarantee the exercise of broadcasting freedom" (Loi 89-25). The CSA, which regulates the various electronic media in France, has broad responsibilities, including monitoring compliance with "pluralism of information," allocating frequencies to different operators, and monitoring and ensuring that the content is lawful and child appropriate (Loi 89-25). For example, during national elections, the CSA can recommend that Internet platforms remove disinformation (Loi 2018-1202). Both the German and French oversight bodies thus may make recommendations to Internet platforms, but France has chosen to treat Internet platforms as a type of media rather than creating a new entity.

The Avia Law granted the CSA the power to pass injunctions against platforms and impose financial penalties for non-compliance, but it was partially invalidated by Le Conseil Constitutionnel (the Constitutional Council) in its 18 June 2020 decision 2020-801 DC. The Constitutional Council found unconstitutional key requirements to remove manifestly illegal hate speech within 24 hours and similarly struck down the provision requiring removal of terrorist content and child pornography within one hour of a government report. Anticipating the adoption of a future EU Regulation, the Digital Services Act ("DSA"), the French Assemblée Nationale has filed Amendment 1770 to Bill 3649, *respect des principes de la République*, designating the CSA as the regulatory body in charge of ensuring the new legal obligations imposed upon platforms and granting the CSA the power to levy fines.

The European approach is currently moving toward enhanced liability for speech in general and hate speech in particular. There is growing pressure on lawmakers to hold both authors and intermediaries accountable and to moderate hate speech.

Private Regulation of Hate Speech

Although the US and European approaches to hate speech have long differed, a convergence is taking place because public pressure, coupled with the threat of government regulation, has spawned action by the technology giants. The practical reality is that Internet platforms are developing

their own protocols and self-regulation that take place largely outside of direct government regulation.

Many social media companies headquartered in the US have adopted their own hate speech policies and guidelines. Facebook's Community Standards (2021) define hate speech as "a direct attack against people" "on the basis of … race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity, and serious disease." The Community Standards define "attacks" as "violent or dehumanizing speech, harmful stereotypes, statements of inferiority, expressions of contempt, disgust or dismissal, cursing, and calls for exclusion or segregation." Facebook flags and removes content that violates these standards using a combination of automated moderation tools and human moderators (Wilson and Land, 2021). The company's recently created Oversight Board applies the Community Standards as part of its independent review of appeals of Facebook's content-moderation decisions (Klonick, 2020). For example, in one of its first cases, decision 2021-002-FB-UA, the Oversight Board upheld Facebook's decision to remove a video posted by a user in the Netherlands that included blackface, citing a hate speech violation. The Oversight Board can also issue opinions on Facebook's content-related policies (Klonick, 2020). The opinions are advisory, however, and Facebook retains full control over the substantive principles guiding the Oversight Board's review (Klonick, 2020).

In a similar vein, Google regulates hate speech on its YouTube platform. YouTube and Google (2021) has a "[h]ate speech policy" and "remove[s] content promoting violence or hatred against individuals or groups" based on age, caste, disability, ethnicity, gender identity and expression, nationality, race, immigration status, religion, sex/gender, sexual orientation, victims of a major violent event and their kin, and veteran status. YouTube may terminate an account or channel for "repeated violations," "after a single case of severe abuse," or when a channel is "dedicated to a policy violation." Thus, the companies themselves substantially restrict what speech remains on their platforms.

In the EU, the European Commission has encouraged this type of self-regulation within the parameters of agreements with the major technology companies. This effort began in 2016 with the introduction of a Code of Conduct, which has now been adopted by many technology companies. According to the European Commission (2016a), the purpose of the Code is to "ensur[e] that online platforms do not offer opportunities for illegal online hate speech to spread virally." In embracing the Code, "the

Commission and the IT companies recognize that the spread of illegal hate speech online not only negatively affects the groups or individuals that it targets but also negatively impacts those who speak out for freedom, tolerance, and non-discrimination." The Code of Conduct includes provisions on oversight, cooperation with law enforcement, and transparency. These agreements do not reflect legal undertakings, serving instead as "a self[-]regulatory commitment" (European Commission, 2016b, p. 4). Nevertheless, self-regulation is occurring. In 2019, Facebook, YouTube, Twitter, and others were removing 72% of the content flagged as illegal hate speech (European Commission, 2019).

The Next Frontier

Given the robust protections conferred by the First Amendment, efforts to regulate hate speech in the US are likely to inhere within the realm of private corporations rather than government regulation. Although the CDA permits Internet providers to remove speech, false, defamatory, and hateful speech has nonetheless exploded on the Internet, evoking ongoing concern about Section 230 immunity (Feiner, 2020; Wakabayashi, 2019).

But encouraging private regulation of hate speech also faces myriad criticisms. Private corporations may selectively remove unpopular speech and claim to be shielded by the CDA (Barr and Department of Justice 2019). Various companies' hate speech policies are more restrictive of hate speech than the US's First Amendment principles and are more akin to the European approach (Wilson and Land, 2021). Although companies strive to draw on definitions of acceptable speech that reflect universal values, that is a nearly impossible task because different communities have different norms (Klonick, 2020). And by shifting the responsibility to the private sector to define and censor speech, the public loses its voice in the process and the government avoids democratic accountability (Land, 2020). There are also related concerns about transparency: while many companies have publicized their hate speech policies, much less is known about the process for removing speech or how companies make decisions in difficult cases (Wilson and Land, 2021).

Today, Congress (PACT Act, 2021), the Federal Communications Commission (FCC) (2020), and others (Barr and Department of Justice 2019) are exploring major changes to the CDA. In the face of public tensions over hate speech, the outcome of legislative and regulatory efforts in the US remains uncertain.

By contrast, legislative efforts in Europe have seen great success. The DSA, which is which is part of an ambitious "Digital Package" that also includes the Digital Markets Act (2020) and the Data Governance Act (2020), reached a significant milestone in 2022 when the European Parliament and EU Member States reached a political agreement on the proposed legislation. The Digital Markets Act seeks to regulate competition and the Data Governance Act creates an EU space for data sharing and optimization.

The DSA seeks to harmonize existing liability standards throughout the EU and revise the E-Commerce Directive by creating a *sui generis* liability standard. Online platforms are obligated to take swift action, enhance oversight, and publish transparency reports, and the DSA adds steep penalties. Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC. The DSA distinguishes between Internet platforms primarily in terms of the size of the user base and adds additional obligations for platforms with 45 million users or more in the EU. The Act also revises the previously discussed "actual knowledge" standard from the E-Commerce Directive, which led to considerable legal uncertainty due to its vagueness. This sweeping initiative is poised to dramatically change the digital landscape.

Another European initiative is the Internet Ombudsman. The Standing Committee, acting on behalf of the Parliamentary Assembly of the Council of Europe (2020), adopted the initiative, requesting that "consideration be given to establishing an ombudsman institution (or equivalent) with the requisite independence, powers, and authority to assess whether Internet content is legal or illegal." The proposal was based on the 2018 report to the Council of Europe by Dan Shefet (one of this chapter's authors), which explains that the Internet Ombudsman would be in charge of assessing the legal or illegal nature of Internet content through a content qualification assessment procedure at the request of Internet platforms. The notion is that the Internet Ombudsman would be able to assess specific complaints about platforms and issue opinions, which would not be binding and would not limit options for legal recourse. To date, the adoption of the proposal to create an Internet Ombudsman has not led to its implementation by the Member States. Such implementation will, to a large extent, depend on the final wording of the oversight mechanisms established by the DSA.

Court proceedings, as well, are beginning to shape the online speech landscape in Europe. In France, a teenager posted a video online where she ranted against Islam (Breeden, 2021). In response, she received a

flood of harassing and threatening messages (Breeden, 2021). A 2018 revision to the French penal code (Loi 2018-703) "empowers prosecutors to seek convictions against harassers who knew they were contributing to a broader wave of abuse" against the victim (Breeden, 2021). In mid-2021, a French court convicted 11 individuals who sent her messages. The judge stated that "[s]ocial networks are the street." "What you wouldn't do in the street — don't do it on social networks" (Breeden, 2021).

With pending and sweeping legislation in the offing, coupled with public pressure and the sensitivity of technology companies to the demands of their users, the regulation of hate speech will remain front and center in the public debate. Balancing the important principle of free expression with the acknowledged need to curb hateful speech presents a critical public policy challenge for the coming decade.

Acknowledgment

The authors thank Nicole Welindt, law clerk to Judge McKeown (Stanford Law School 2019), for her research assistance.

References

Abrams v. United States (1919). 250 U.S. 616.

Akee, R., Ward, K. J. and Brookings Institution (2021). Missed opportunities to understand racism in the COVID-19 era. https://www.brookings.edu/blog/up-front/2021/05/13/missed-opportunities-to-understand-the-prevalence-of-racism-in-the-u-s-in-the-covid-19-era/. Accessed on 2 July 2021.

ARTICLE 19 (2018). Responding to "hate speech": Comparative overview of six EU countries. London: ARTICLE 19. https://www.article19.org/wp-content/uploads/2018/03/ECA-hate-speech-compilation-report_March-2018.pdf. Accessed on 5 July 2021.

Assemblée Nationale (2021). Respect des principes de la République: Amendement 1770. https://www.assemblee-nationale.fr/dyn/15/amendements/3649/CSPRINCREP/1770.pdf. Accessed on 5 July 2021.

Barr, W. P. and Department of Justice (2019). Remarks at the National Association of Attorneys General 2019 Capital Forum. 10 December 2019. https://www.justice.gov/opa/speech/attorney-general-william-p-barr-delivers-remarks-national-association-attorneys-general. Accessed on 5 July 2021.

Bayer, J. and Bárd, P. (2020). Hate speech and hate crime in the EU and the evaluation of online content regulation approaches. Brussels: European Parliament.

https://www.europarl.europa.eu/RegData/etudes/STUD/2020/655135/IPOL_STU(2020)655135_EN.pdf. Accessed on 2 July 2021.

Beauharnais v. Illinois (1952). 343 U.S. 250.

Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v. Netlog NV. (2012). 2012 ECR 85.

Bennett v. Google, LLC (2018). 882 F.3d 1163 (D.C. Cir.).

Brandenburg v. Ohio (1969). 395 U.S. 444 (per curiam).

Breeden, A. (2021). French court convicts 11 people of harassing teenager over anti-Islam rant. *The New York Times*. 7 July. https://www.nytimes.com/2021/07/07/world/europe/france-mila-online-abuse.html. Accessed on 8 July 2021.

Bureau of Justice Statistics (2017). Hate crime victimization, 2004–2015. Bureau of Justice Statistics. https://bjs.ojp.gov/content/pub/pdf/hcv0415.pdf. Accessed on 2 July 2021.

Center for Democracy & Technology (2017). Overview of the NetzDG network enforcement law. https://cdt.org/insights/overview-of-the-netzdg-network-enforcement-law/. Accessed on 5 July 2021.

Chaplinsky v. New Hampshire (1942). 315 U.S. 568.

Charter of Fundamental Rights of the European Union (2012). Official Journal. C 326. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012P/TXT&from=EN. Accessed on 8 July 2021.

Chicago Lawyers' Committee for Civil Rights Under Law, Inc. v. Craigslist, Inc. (2008). 519 F.3d 666 (7th Cir.).

Citron, D. K. and Franks, M. A. (2020). The internet as a speech machine and other myths confounding section 230 reform. *The University of Chicago Legal Forum*, 2020, pp. 45–75.

Committee on the Elimination of Racial Discrimination (2013). General recommendation no. 35, combating racist hate speech. United Nations. http://docstore.ohchr.org/SelfServices/FilesHandler.ashx?enc=6QkG1d%2FPPRiCAqhKb7yhssyNNtgI51ma08CMa6o7Bglz8iG4SuOjovEP%2Bcqr8joDoVEbW%2BQ1MoWdOTNEV99v6FZp9aSSA1nZya6gtpTo2JUBMI0%2BoOmjAwk%2B2xJW%2BC8e. Accessed on 5 July 2021.

Convention on the Prevention and Punishment of the Crime of Genocide (1948). General Assembly Resolution 260 A (III).

Criminal Justice Information Services Division, Federal Bureau of Investigation (2019). Hate crime statistics. https://ucr.fbi.gov/hate-crime/2019. Accessed on 2 July 2021.

Dharmapala, D. and McAdams, R. H. (2005). Words that kill? An economic model of the influence of speech on behavior (with particular reference to hate speech). *The Journal of Legal Studies*, 34(1), 93–136.

Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular

electronic commerce, in the Internal Market ("Directive on electronic commerce") (2000). 2000 O.J. (L 178) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32000L0031&from=EN. Accessed on 5 July 2021.

Draft Report on the proposal for a regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. (2021). Committee on the Internal Market and Consumer Protection. COD (2020) 0361. https://www.europarl.europa.eu/doceo/document/IMCO-PR-693594_EN.pdf. Accessed on 5 July 2021.

Dyroff v. Ultimate Software Group, Inc. (2019). 934 F.3d 1093 (9th Cir.).

European Commission (2016a). European Commission and IT Companies announce Code of Conduct on illegal online hate speech. 31 May 2016. https://ec.europa.eu/commission/presscorner/detail/en/IP_16_1937. Accessed on 5 July 2021.

European Commission (2016b). Code of Conduct — Illegal online hate speech: Questions and answers. European Commission. https://ec.europa.eu/info/sites/info/files/code_of_conduct_hate_speech_en.pdf. Accessed on 2 July 2021.

European Commission (2019). Code of Conduct on countering illegal hate speech online: Fourth evaluation confirms self-regulation works. European Commission. https://ec.europa.eu/info/sites/default/files/code_of_conduct_factsheet_7_web.pdf. Accessed on 5 July 2021.

European Commission (2020a). Notice to Stakeholders: Withdrawal of the United Kingdom and EU rules in the field of Electronic commerce and net neutrality. Brussels: European Commission. https://ec.europa.eu/info/sites/default/files/brexit_files/info_site/e_commerce_en.pdf. Accessed on 5 July 2021.

European Commission (2020b). October infringements package: Key decisions. https://ec.europa.eu/commission/presscorner/detail/en/inf_20_1687. Accessed on 5 July 2021.

European Convention for the Protection of Human Rights and Fundamental Freedoms (1950). 213 U.N.T.S. 221.

European Network Against Racism (2020). COVID-19 impact on racialised communities: Interactive EU-wide map. https://www.enar-eu.org/COVID-19-impact-on-racialised-communities-interactive-EU-wide-map. Accessed on 2 July 2021.

European Union Agency for Fundamental Rights (FRA) (2012). Data in Focus Report: Minorities as Victims of Crime. https://fra.europa.eu/sites/default/files/fra-2012-eumidis-dif6_0.pdf. Accessed on 12 July 2022.

European Union Agency for Fundamental Rights (FRA) (2018). Hate crime recording and data collection practice across the EU. Luxembourg: Publications Office of the European Union. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-hate-crime-recording_en.pdf. Accessed on 5 July 2021.

Eva Glawischnig-Piesczek v. Facebook Ireland Limited (2019). 2019 ECR 821.

Facebook (2021). Community standards: 12. Hate speech. https://www.facebook.com/communitystandards/hate_speech. Accessed on 2 July 2021.

Faurisson v. France (1996). UN Doc. CCPR/C/58/D/550/1993.

Federal Communications Commission (FCC) (2020). Statement of Chairman Pai on Section 230. 15 October 2020. https://docs.fcc.gov/public/attachments/DOC-367567A1.pdf. Accessed on 5 July 2021.

Feiner, L. (2020). Big Tech's favorite law is under fire. *CNBC*. 19 February. https://www.cnbc.com/2020/02/19/what-is-section-230-and-why-do-some-people-want-to-change-it.html. Accessed on 5 July 2021.

Fisch, W. B. (2002). Hate speech in the constitutional law of the United States. *The American Journal of Comparative Law*, 50(1), 463–492.

Force v. Facebook, Inc. (2019). 934 F.3d 53 (2d Cir.).

Gitlow v. New York (1925). 268 U.S. 652.

Hate Crime Acts, 18 U.S.C. § 249(a). https://uscode.house.gov/. Accessed on 2 July 2021.

House of Representatives (1996). Conference Report No. 104-458. https://www.congress.gov/congressional-report/104th-congress/house-report/458/1. Accessed on 2 July 2021.

Incitement of masses, Strafgesetzbuch [StGB] Section 130. http://www.gesetze-im-internet.de/englisch_stgb/index.html. Accessed on 5 July 2021.

International Law Commission (1996). Report of the International Law Commission on the work of its forty-eighth session, UN Doc. A/51/10. United Nations. https://legal.un.org/ilc/documentation/english/reports/a_51_10.pdf. Accessed on 5 July 2021.

Internet-beschwerdestelle.de (2021). About us. https://www.internet-beschwerdestelle.de/en/index.html. Accessed on 5 July 2021.

Jawa, M. W. (2020). Note. The "offensive" oversimplification: An argument for hate speech laws in the modern era. *First Amendment Law Review*, 19 (Fall), 130–152.

Klonick, K. (2020). The Facebook oversight board: Creating an independent institution to adjudicate online free expression. *The Yale Law Journal*, 129(8), 2418–2499.

Land, M. K. (2020). Against privatized censorship: Proposals for responsible delegation. *Virginia Journal of International Law*, 60(2), 363–432.

Laub, Z. and Council on Foreign Relations (2019). Hate speech on social media: Global comparisons. https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons. Accessed on 2 July 2021.

Lawrence, F. M. (1993). Resolving the hate crimes/hate speech paradox: Punishing bias crimes and protecting racist speech. *Notre Dame Law Review*, 68(4), 673–721.

Le Conseil Constitutionnel (2020). Decision No. 2020-801 DC du 18 juin 2020. https://www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm. Accessed on 5 July 2021.

Loi 89-25 du 17 janvier 1989 modifiant la loi 86-1067 du 30 septembre 1986 relative à la liberté de communication [Law 89-25 of January 17, 1989 modifying the Law 86-1067 of 30 September 1986 on freedom of communication]. https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000321869/. Accessed on 5 July 2021.

Loi 90-615 du 13 juillet 1990 tendant à réprimer tout acte raciste, antisémite ou xénophobe [Law 90-615 of July 13, 1990 for the suppression of racist, antisemitic, or xenophobic acts]. https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000532990/. Accessed on 5 July 2021.

Loi 2018-703 du 3 août 2018 renforçant la lute contre les violences sexuelles et sexistes [Law 2018-703 of 3 August 2018 strengthening the fight against sexual and gender-based violence]. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037284450. Accessed on 8 July 2021.

Loi 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information [Law 2018-1202 of 22 December 2018 on combatting the manipulation of information]. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037847559/. Accessed on 5 July 2021.

Loi 2020-766 du 24 juin 2020 visant à lutter contre les contenus haineux sur internet [Law 2020-76 of 24 June 2020 combatting hateful content on the internet]. https://www.assemblee-nationale.fr/15/pdf/ta/ta0419.pdf. Accessed on 8 July 2021.

Matal v. Tam (2017). 137 S. Ct. 1744.

National Socialist Party of America v. Village of Skokie (1977). 432 U.S. 43 (per curiam).

Network Enforcement Act (2017). https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.html;jsessionid=52B2A016DAD2D63C25CC1B9F968733DB.1_cid324?nn=6712350#Start. Accessed on 5 July 2021.

New York Times Co. v. Sullivan (1964). 376 U.S. 254.

Oversight Board (2021). Case decision 2021-002-FB-UA. https://oversightboard.com/decision/FB-S6NRTDAJ/. Accessed on 5 July 2021.

Oxford Academic (Oxford University Press) (2018). Does the first amendment protect hate speech? Nadine Strossen. https://www.youtube.com/watch?v=lfV-VP7dCNw. Accessed on 2 July 2021.

PACT Act (2021). S. 797. https://www.congress.gov/bill/117th-congress/senate-bill/797. Accessed on 5 July 2021.

Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive

2000/31/EC (2020). COM (2020) 825 final. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN. Accessed on 5 July 2021.

Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) (2020). COM (2020) 842 final. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:842:FIN. Accessed on 5 July 2021.

Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) (2020). COM (2020) 767 final. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:767:FIN. Accessed on 5 July 2021.

Prosecutor v. Šešelj (2016). Case No. IT-03-67-T (International Criminal Tribunal for the Former Yugoslavia).

Protection for private blocking and screening of offensive material, 47 U.S.C. § 230. https://uscode.house.gov/. Accessed on 5 July 2021.

R.A.V. v. City of St. Paul (1992). 505 U.S. 377.

Relia, K., *et al*. (2019). Race, ethnicity and national origin-based discrimination in social media and hate crimes across 100 U.S. cities. In *13th International Conference on Web and Social Media*, Munich, Germany, 11–14 June 2019.

Rich, W. J. (2020). Modern Constitutional Law, Volume 1, Section 6.16. Available through: Thomas Reuters Westlaw Edge. https://next.westlaw.com. Accessed on 2 July 2021.

Rolfe, S. M. and Schroeder, R. D. (2020). "Sticks and stones may break my bones, but words will never hurt me": Verbal sexual harassment among middle school students. *Journal of Interpersonal Violence*, 35(17–18), 3462–3486.

Schenck v. United States (1919). 249 U.S. 47.

Shefet, D. (2018). *Expert Report: On the Creation of the Position of Internet Ombudsman in Charge of Assessing the Legal or Illegal Nature of Internet Contents Through Screening Procedures* (commissioned and funded by the Parliamentary Assembly of the Council of Europe).

Smolla, R. A. (2021). Smolla & Nimmer on Freedom of Speech, Volume 1, Section 12:7. Thomas Reuters Westlaw Edge. https://next.westlaw.com. Accessed on 2 July 2021.

Snyder v. Phelps (2011). 562 U.S. 443.

Standing Committee, Parliamentary Assembly, Council of Europe (2020). Resolution 2334: Towards an internet ombudsman institution. https://pace.coe.int/en/files/28728/html. Accessed on 5 July 2021.

Stratton Oakmont, Inc. v. Prodigy Services Co. (1995). 1995 WL 323710 (N.Y. Sup. Ct.).

Telecommunications Act of 1996, Pub. L. 104-104. https://www.govinfo.gov/content/pkg/PLAW-104publ104/pdf/PLAW-104publ104.pdf. Accessed on 2 July 2021.

Temperman, J. (2019). The international covenant on civil and political rights and the "right to be protected against incitement." *Journal of Law, Religion and State*, 7(1), 89–103.

The Nurnberg Trial 1946 (1946–1947). 6 F.R.D. 69.

Timmermann, W. K. (2006). Incitement in international criminal law. *International Review of the Red Cross*, 88(864), 823–852.

Twitter (2021). Twitter rules and policies: Hateful conduct policy. https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy. Accessed on 2 July 2021.

United Nations (2019). Strategy and plan of action on hate speech. United Nations. https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf. Accessed on 2 July 2021.

United Nations (2020). United Nations strategy and plan of action on hate speech: Detailed guidance on implementation for United Nations field presences. United Nations. https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20PoA%20on%20Hate%20Speech_Guidance%20on%20Addressing%20in%20field.pdf. Accessed on 2 July 2021.

United States Department of Justice (2019). Hate crime laws. https://www.justice.gov/crt/hate-crime-laws. Accessed on 2 July 2021.

United States Department of Justice (2021). Hate crimes: Laws and policies. https://www.justice.gov/hatecrimes/laws-and-policies. Accessed on 2 July 2021.

Virginia v. Black (2003). 538 U.S. 343.

Wakabayashi, D. (2019). Legal shield for websites rattles under onslaught of hate speech. *The New York Times*, 6 August. https://www.nytimes.com/2019/08/06/technology/section-230-hate-speech.html. Accessed on 5 July 2021.

Wallmeyer, E. J. (2003). Filled milk, footnote four & the first amendment: An analysis of the preferred position of speech after the Carolene products decision. *Fordham Intellectual Property, Media and Entertainment Law Journal*, 13(4), 1019–1052.

Warner, M. R. and Senate (2021). Warner, Hirono, Klobuchar Announce the SAFE TECH Act to Reform Section 230. 5 February 2021. https://www.warner.senate.gov/public/index.cfm/2021/2/warner-hirono-klobuchar-announce-the-safe-tech-act-to-reform-section-230. Accessed on 2 July 2021.

Watts v. United States (1969). 394 U.S. 705 (per curiam).

Whitney v. California (1927). 274 U.S. 357.

Wilson, R. A. and Kiper, J. (2020). Incitement in an era of populism: Updating *Brandenburg* after Charlottesville. *University of Pennsylvania Journal of Law & Public Affairs*, 5(2), 57–121.

Wilson, R. A. and Land, M. K. (2021). Hate speech on social media: Content moderation in context. *Connecticut Law Review*, 52(3), 1029–1076.

Wisconsin v. Mitchell (1993). 508 U.S. 476.

YouTube and Google (2021). Hate speech policy. https://support.google.com/youtube/answer/2801939?hl=en. Accessed on 2 July 2021.

https://doi.org/10.1142/9781800612860_0013

Chapter 12

Cyber Risks, Dark Web, and Money Laundering

Fausto Martin De Sanctis

Introduction

The Internet was discovered by criminals as an effective and clandestine way to launder money internationally. Its attraction is the possibility of anonymity, with the use of the dark web, which facilitates the criminal onslaught for the practice of various crimes. Unfortunately, although potentially equipped with modern tools, investigators, prosecutors, judges, and regulatory agencies in most countries find difficulties, which are inherent, in accurately detecting, investigating, and prosecuting this type of criminal activity. Different types and sizes of organizations are at risk, not just financial services companies, which requires a permanent state of alert for containment. To avoid laundering large sums of money, possible factors that may remain immune to criminal law should be checked. Noting that money laundering is increasingly becoming a cyber-crime, this chapter is intended to provide insight into new ways in which money is laundered through illegal activities involving the Internet.

Personal Data Protection and Cyber Risks

In the United States, on 19 September 2012, Senator Johan D. Rockefeller IV, chairman of the Senate Committee on Commerce, Science, and

Transportation, wrote directly to the CEOs of Fortune 500 companies on cybersecurity to provide answers to eight questions pertaining to their companies' cybersecurity practices, with certain aspects of the 2012 Cybersecurity Act that did not pass the Senate (Rockefeller VI, 2012).[1] Already, the Cybersecurity Act of 2015 (The Cybersecurity, 2015) imposed mandatory security standards for owners and operators in critical sectors of this segment (infrastructure). In contrast, the law classified cybersecurity-related information under Cybersecurity Information Sharing, Federal Cybersecurity Enhancement (including enhanced federal network security), Federal Workforce Assessment, and other cyber matters, such as Security Mobile Devices.[2]

[1] 1. Has your company adopted a set of best practices to address your cybersecurity needs? 2. If so, how were these cybersecurity practices developed? 3. Were they developed exclusively by the company or were they developed outside the company? If developed outside the company, list the institution, association, or entity that provided the pertinent information. 4. When were these cybersecurity practices developed? How often have they been updated? Does your company's board of directors or the audit committee monitor the performance and implementation of these practices? 5. Has the federal government played any role, advisory or otherwise, in developing these cybersecurity practices? 6. What are your concerns, if any, with a voluntary program that allows the federal government and the private sector to develop, in coordination, cybersecurity best practices for companies to adopt as they see fit, as outlined in the Cybersecurity Act of 2012? 7. What are your concerns, if any, with the federal government conducting risk assessments in coordination with the private sector, to better understand where your country's cyber vulnerabilities lie, as outlined in the Cybersecurity Act of 2012? 8. What are your concerns, if any, with the federal government determining, in coordination with the private sector, the country's most critical cyberinfrastructure, as outlined in the Cybersecurity Act of 2012?

[2] *Sec. 205. Federal cybersecurity requirements. (a) Implementation of federal cybersecurity standards — Consistent with section 3553 of title 44, United States Code, the Secretary, in consultation with the Director, shall exercise the authority to issue binding operational directives to assist the Director in ensuring timely agency adoption of and compliance with policies and standards promulgated under section 11331 of title 40, United States Code, for securing agency information systems. (b) Cybersecurity requirements at agencies — (1) In general — Consistent with policies, standards, guidelines, and directives on information security under subchapter II of chapter 35 of title 44, United States Code, and the standards and guidelines promulgated under section 11331 of title 40, United States Code, and except as provided in paragraph (2), not later than one year after the date of the enactment of this Act, the head of each agency shall (A) identify sensitive and mission critical data stored by the agency consistent with the inventory required under the first subsection (c) (relating to the inventory of major information systems) and the second subsection (c) (relating to the inventory of information systems) of section 3505 of title 44, United States Code; (B) assess*

In turn, the Cybersecurity Act 2017 amended the National Institute of Standards and Technology Act (NIST) to require it to consider small businesses that facilitate and support the development of voluntary,

access controls to the data described in subparagraph (A), the need for readily accessible storage of the data, and individuals' need to access the data; (C) encrypt or otherwise render indecipherable to unauthorized users the data described in subparagraph (A) that is stored on or transiting agency information systems; (D) implement a single sign-on trusted identity platform for individuals accessing each public website of the agency that requires user authentication, as developed by the Administrator of General Services in collaboration with the Secretary; and (E) implement identity management consistent with section 504 of the Cybersecurity Enhancement Act of 2014 (Public Law 113–274; 15 U.S.C. 7464), including multifactor 21 authentication, for (i) remote access to an agency information system and (ii) each user account with elevated privileges on an agency information system. (2) Exception — The requirements under paragraph (1) shall not apply to an agency information system for which (A) the head of the agency has personally certified to the Director with particularity that (i) operational requirements articulated in the certification and related to the agency information system would make it excessively burdensome to implement the cybersecurity requirement; (ii) the cybersecurity requirement is not necessary to secure the agency information system or agency information stored on or transiting it; and (iii) the agency has taken all necessary steps to secure the agency information system and agency information stored on or transiting it; and (B) the head of the agency or the designee of the head of the agency has submitted the certification described in subparagraph (A) to the appropriate congressional committees and the agency's authorizing committees. (3) Construction — Nothing in this section shall be construed to alter the authority of the Secretary, the Director, or the Director of the National Institute of Standards and Technology in implementing subchapter II of chapter 35 of title 44, United States Code. Nothing in this section shall be construed to affect the National Institute of Standards and Technology standards process or the requirement under section 3553(a)(4) of such title or to discourage continued improvements and advancements in the technology, standards, policies, and guidelines used to promote Federal information security. (c) EXCEPTION — The requirements under this section shall not apply to the Department of Defense, a national security system, or an element of the intelligence community; Sec. 401. Study on mobile device security. (a) In general — Not later than one year after the date of the enactment of this Act, the Secretary of Homeland Security, in consultation with the Director of the National Institute of Standards and Technology, shall (1) complete a study on threats relating to the security of the mobile devices of the Federal Government and (2) submit an unclassified report to Congress, with a classified annex if necessary, that contains the findings of such study, the recommendations developed under paragraph (3) of subsection (b), the deficiencies, if any, identified under (4) of such subsection, and the plan developed under paragraph (5) of such subsection. (b) Matters studied — In carrying out the study

consensus-based, industry-led guidelines and procedures and that cost-effectively reduce cyberinfrastructure to critical infrastructure. NIST has now disseminated and published standard and method resources on its website so that small businesses can use them voluntarily, helping to reduce their cybersecurity risks. Features must be as follows: (1) technology neutral; (2) based on international standards, to the extent possible; (3) capable of varying according to the nature and size of the small business being implemented and the sensitivity of data collected or stored in information systems; and (4) consistent with the national security awareness and education program in accordance with the Cybersecurity Enhancement Act of 2014 (The Cybersecurity 2017).

In Europe, the General Data Protection Regulation — GDPR 2016/679 — was adopted on 14 April 2016, and, after a transition period of two years, became applicable on 25 May 2018. As the GDPR is a regulation, it is not a directive, does not require national governments to pass any legislation that permits, and is directly binding and applicable to all members of the European Union — EU, in addition to Norway, Iceland, and Liechtenstein (European Economic Area — EEA). It is a regulation on data protection and privacy for all individuals in the European Union and the European Economic Area. It also covers the export of personal data outside the EU and EEA. The GDPR's main objective is to allow citizens and residents to control their personal data and simplify the regulatory environment for international business, unifying the regulation in the EU.

---

under subsection (a)(1), the Secretary, in consultation with the Director of the National Institute of Standards and Technology, shall (1) assess the evolution of mobile security techniques from a desktop-centric approach and whether such techniques are adequate to meet current mobile security challenges; (2) assess the effect such threats may have on the cybersecurity of the information systems and networks of the Federal Government (except for national security systems or the information systems and networks of the Department of Defense and the intelligence community); (3) develop recommendations for addressing such threats based on industry standards and best practices; (4) identify any deficiencies in the current authorities of the Secretary that may inhibit the ability of the Secretary to address mobile device security throughout the Federal Government (except for national security systems and the information systems and networks of the Department of Defense and intelligence community); and (5) develop a plan for accelerated adoption of secure mobile device technology by the Department of Homeland Security. (c) Intelligence community defined — In this section, the term "intelligence community" has the meaning given such term in section 3 of the National Security Act of 1947 (50 U.S.C. 3003).

Replacing the Data Protection Directive, the regulation contains provisions and requirements regarding the processing of personally identifiable information of data subjects in the European Union. Business transactions dealing with personal data were now protected with data protection by design and by default, which meant that personal data were now stored using pseudonymization or complete anonymization and using the highest possible privacy settings by default so that the data would not be publicly available without explicit consent and could not be used to identify a subject without additional information stored separately. No personal data are processed, unless done under a legal basis specified by regulation, or when the controller or data processor has received explicit and optional consent from the data owner. The data owner now has the right to revoke this permission at any time.

It established that a personal data processor must clearly disclose any data collection. Also, it must state the legal basis and purpose of data processing, how long the data are being retained, and whether it is being shared with third parties or outside the EU. Users now have the right to request a portable copy of data collected by a processor in a common format and the right to have their data erased under certain circumstances. Companies whose core activities were focused on the regular or systematic processing of personal data and public authorities were now required to hire a data protection officer (DPO), responsible for managing compliance with the GDPR. Companies are now required to report any data breaches within 72 hours if they have an adverse effect on user privacy.

Such regulation applies whether the data controller (an organization that collects data from EU residents) or the processor (an organization that processes data on behalf of a data controller, such as cloud service providers) or in case of the person is located in the EU. Under certain circumstances, the regulation also applies to organizations based outside the EU if they collect or process personal data from individuals located within the EU.

According to the European Commission, personal data are any information related to an individual, whether relating to their private, professional, or public life. It can be anything from a name, a home address, a photo, an email address, bank details, posts on social networking sites, medical information, or a computer's IP address.

The regulation does not apply, however, to the processing of personal data for national security or EU law enforcement activities;

however, industry groups concerned about facing a potential conflict of laws questioned whether Article 48 of the GDPR could be invoked to try to prevent a data controller subject to the laws of a third country from complying with a legal order from the police, court, or security authorities to disclose the personal data of an EU person to those authorities, regardless of whether the data reside inside or outside the EU. Article 48 states that any judgment of a court of law and any decision of an administrative authority of a third country requiring a controller or processor to transfer or disclose personal data cannot be recognized or enforced in any way, unless it is based on an international agreement, such as a mutual legal assistance treaty in force between the requesting third country (outside the EU) and the EU or a Member State. The data protection reform package also includes a separate Data Protection Directive for the police and criminal justice sector, which provides rules on the exchange of personal data at national, European, and international levels.

Each Member State now has a duty to establish an independent supervisory authority to hear and investigate complaints and sanction administrative offenses. Supervisors in each Member State should cooperate with other supervisors, providing mutual assistance and organizing joint operations. If a company has multiple establishments in the EU, it will have a single supervisor as its "main authority," based on the location of its "main establishment" where the main processing activities take place. The main authority will act as a "one-stop shop" to oversee all processing activities for this business across the EU (Articles 46–55 of the GDPR). A European Data Protection Board (EDPB) should coordinate the supervisors (Articles 2(2)(a) and 88 of the GDPR).

Unless a data subject has given explicit consent to data processing for one or more purposes, personal data cannot be processed unless there is at least a legal authorization to do so. This includes the following: performing a task in the public interest or by official authority; complying with the legal obligations of a data controller; fulfilling contractual obligations with a data subject; performing tasks at the request of a data subject who is entering into a contract with the controller; protecting the vital interests of a data subject or another person; is in the legitimate interest of a data controller or a third party.

If consent is used as a legal basis for processing, it must be explicit (Article 7; defined in Article 4). Children's consent must be given by their

parents or guardians (Article 8). Data controllers must be able to prove "consent" (acceptance) and this can be withdrawn.

The GDPR consent area has several implications for companies that record calls as a practice. Typical "calls are recorded for training and security purposes" warnings will no longer be sufficient to obtain assumed consent to record calls. In addition, when recording begins, if the caller withdraws their consent, the agent receiving the call must be able to stop a previously started recording and ensure that the recording is not stored.

In order to demonstrate compliance with the GDPR, the data controller must implement measures that meet data protection principles by design and by default. Data protection by design and by default (Article 25) requires that data protection measures be designed into business development for products and services. Such measures include the pseudonym of personal data by the data controller as soon as possible (Recital 78). It is the responsibility of the data controller to implement effective measures and to be able to demonstrate compliance of the processing activities, even if the processing is carried out by a data processor on behalf of the controller (Recital 74).

When data are collected, users must be clearly informed about the extent of data collection, the legal authorization for processing personal data, how long the data are retained, whether the data are being transferred to third parties and/or outside of the EU, and disclosure of any automated decision-making made solely on the basis of algorithms. Users must receive contact details from the data controller and the designated data protection officer, where applicable. Users must also be informed of their privacy rights under the GDPR, including their right to revoke consent to data processing at any time, their right to view their personal data and have an overview of how they are being processed, their right to obtain a portable copy of stored data, their right to erase data under certain circumstances, their right to challenge any automated decision-making that was based solely on algorithms, and their right to file complaints with a Data Protection Authority — DPA.

Data protection impact assessments (Article 35) should be carried out when specific risks to the rights and freedoms of data subjects occur. Risk assessment and mitigation are required and prior approval from national data protection authorities (DPAs) is required for high risks. Data protection by design and by default (Article 25) mandates that data protection be designed into the development of business processes for products and services (design); privacy settings must also be set at a high level by

default, and technical and procedural measures must be taken by the controller to ensure that the processing, throughout its lifecycle, complies with the regulation.

The GDPR refers to pseudonymization as a necessary process when data are stored (as an alternative to the other complete data anonymization option) to transform personal data in such a way that the resulting data cannot be attributed to a specific person without the use of additional information.

An example of pseudonymization is encryption, which makes the original data unintelligible and the process cannot be reversed without access to the correct decryption key. GDPR requires that additional information (such as the decryption key) be kept separate from the pseudonymized data. Another example of pseudonymization is tokenization, which is a non-mathematical approach to protecting data-at-rest that replaces sensitive data with non-sensitive replacements, called tokens. Tokens have no extrinsic or exploitable meaning or value. Tokenization does not change the data type, which means it can be processed by legacy systems, such as size and data type sensitive databases. This requires far less computing resources to process and less storage space in databases than traditionally encrypted data. It is achieved by keeping specific data fully or partially visible for processing and analysis, while confidential information is kept hidden.

The right of access (Article 15) is a personal right to data. It gives citizens the right to access their personal data and information about how that personal data are being processed. A data controller must provide, upon request, an overview of the data categories being processed (Article 15, 1, b) as well as a copy of the actual data (Article 15, 3). In addition, the data controller must inform the data subject of details about the processing, such as the purposes of the processing (Article 15, 1, a), with whom the data are shared (Article 15, 1, c), and how you acquired the data (Article 15, 1, g).

A data subject must be able to transfer personal data from one electronic processing system to another, without being impeded by the data controller. The right to be forgotten was replaced by the more limited right to erasure in the version of the GDPR adopted by the European Parliament in March 2014. Article 17 states that the data subject has the right to request the erasure of personal data related to them for any one of several reasons, including non-compliance with Article 6, 1 (legality).

According to the GDPR, the data controller is under a legal obligation to notify the supervisory authority without undue delay, unless the

violation is unlikely to result in a risk to the rights and freedoms of individuals, no later than 72 hours after becoming aware of the data breach to file a complaint (Article 33). Furthermore, the data processor must notify the controller without undue delay after becoming aware of a personal data breach (Article 33). However, notice to data subjects is not necessary if the data controller has implemented appropriate technical and organizational protection measures that make the personal data unintelligible to anyone who is not authorized to access it, such as encryption (Article 34).

The following sanctions may be imposed: (i) written notice in cases of initial and unintentional non-compliance; periodic data protection audits; (ii) a fine of up to 10 million euros or up to 2% of the annual worldwide turnover of the previous financial year in the case of a company, whichever is greater, if there is a violation of the following provisions (Article 83, Paragraphs 5 and 6): the obligations of the controller and the processor under Articles 8, 11, 25 to 39, 42, and 43, the obligations of the certification body under Articles 42 and 43, and the obligations of the control body under Article 41 (4); (iii) a fine of up to 20 million euros or up to 4% of the annual worldwide turnover of the previous financial year in the case of a company, whichever is greater, if there is a violation of the following provisions (Article 83, Paragraph 4): the basic principles of processing, including conditions of consent, under Articles 5, 6, 7, and 9, data subjects' rights under Articles 12–22, transfers of personal data to a recipient in a third country or international organization pursuant to Articles 44–49, any obligations under the law of the Member State adopted pursuant to Chapter IX, non-compliance with an order or temporary or definitive limitation in the processing or suspension of data flows by the supervisory authority pursuant to Article 58, 2, or lack of access, violating Article 58, 1 (The General Data Protection Regulation, 2016).

In short, the GDPR grants people rights with regard to the protection and control of their personal data. This will give them a clear insight into whether their data have been used to generate ads or create profiles, or whether the companies collecting data have sold these data to third parties. The user now has the right to access, change, or delete the data that were provided to the companies. It can force the company to delete everything it owns from the person if requested by the user. Due to the transnational characteristic of the GDPR, any foreign company that has customers, suppliers, or partners located in Europe must comply with the regulation. Otherwise, it will be subject to penalties provided for by law.

It is important to say that Brazil, as stated by Rafael Mendes Loureiro and Leonardo A. F. Palhares, "lacks specific regulations on cybersecurity; although there are efforts to adopt a binding and integrated regulatory framework. Brazilian legislation on the subject is still evolving" (Loureiro and Palhares, 2018).[3]

Nicknamed Carolina Dieckmann, the Cyber Crimes Law (Law No. 12,737, of November 30, 2012) criminalizes conduct related to electronic tools, such as breaking into computers, violating user data, or "taking down" websites. The bill that gave rise to the law (PLC 35/2012) was drawn up at a time when the intimate photos of actress Carolina Dieckmann were copied from her computer and spread across the World Wide Web. The Law was claimed by the financial system, given the amount of fraud applied via the Internet (Law n. 12,737, 2012).

"Computer hacking" can be punished by imprisonment from three months to a year, plus a fine. More harmful conduct, such as "obtaining content from private electronic communications, trade or industrial secrets, and confidential information, as defined by law," can be punished by six months to two years in prison and a fine. The same occurs if the crime involves the disclosure, sale, or transmission to third parties, through the sale or free transfer, of material obtained from the invasion. The law also provides for an increase in penalties from one-sixth to one-third if the invasion causes economic loss and from one to two-thirds "if there is disclosure, sale, or transmission to a third party, in any capacity, of the data or information obtained." The penalties can also be increased from a third to a half if the crime is committed against the President of the Republic, Chief Justice of the Federal Supreme Court, presidents of the Chamber, Senate, assemblies and legislative chambers, municipal chambers, or high officials of the administration direct and indirect federal, state, local, or Federal District (Articles 154-A and 154-B, both of the Brazilian Penal Code).

The Civil Rights Framework for the Internet in Brazil (Law No. 12,965, of 23 April 2014) also considered the Brazilian Civil Rights Framework for the Internet and established that, in providing a connection with it, it is up to the respective autonomous system administrator to maintain logs of connection in a confidential and controlled environment, for a period of one (1) year, and that the Internet application provider,

[3] "The current legal framework is a patchwork of laws and regulations, as a number of flexible laws have been adopted, mainly addressing issues related to the banking sector."

established as a legal entity and carrying out this activity in an organized, professional, and economical manner, keeps the respective records of access to Internet applications, confidentially, in a controlled environment and security, for a period of six months, as per regulation.

The Brazilian Civil Rights for the Internet, in short, regulates its use in Brazil through a series of principles, rights, and duties for its users, addressing several issues, such as (1) net neutrality, (2) privacy, (3) data retention, (4) social function of the Internet, (5) freedom of expression and transmission of knowledge, and (6) obligations related to the civil liability of users and providers (Law n. 12,965, 2014).[4]

According to Rafael Mendes Loureiro and Leonardo A. F. Palhares, Brazil "has adopted international information security management policies. The Brazilian Association of Technical Standards (ABNT) developed NBR ISO/IEC 27001: 2006, which is an identical translation of ISO/IEC 27001: 2005, prepared by the Joint Committee on Information Technology (ISO/IEC/JTC 1), Subcommittee of IT Security Techniques (SC 27)" (Loureiro and Palhares, 2018, p. 17). Therefore, international best practices and standards are generally adopted by entities to protect their systems and information.

The legal framework that specifically deals with the protection of personal data, as well as its use and transfer, advanced in the Brazilian Congress on 2 July 2018. Senator Ricardo Ferraço (PSDB-ES) presented his opinion on bill no. 53/2018 to the Economic Affairs Committee (CEA), which dealt with the proposal. The senator preserved the content that was approved by the House in May and made only a few editorial

[4] *Article 10. The custody and availability of records of connection and access to Internet applications mentioned in this Law, as well as personal data and content of private communications, must take into account the preservation of the privacy, honor, and image of the parties involved, directly or indirectly. (...) § 4 The security and confidentiality measures and procedures must be informed by the person responsible for providing the services in a clear manner and in accordance with the standards defined in the regulations, respecting their right to confidentiality in relation to commercial secrets. Article 13. When providing the Internet connection, it is the responsibility of the respective autonomous system administrator to keep records of the connection, in a confidential and controlled environment, for a period of one (1) year, as per regulation. Article 15. The Internet application provider established as a legal entity and exercising this activity in an organized, professional, and economical manner will keep the respective records of access to Internet applications, confidentially, in a controlled and secure environment for a period of six (6) months, pursuant to the regulation*.

adjustments to the text to suit the House Review procedure. The project was approved and resulted in Law No. 13,709, of 14 August 2018. It is the Brazilian Personal Data Protection Law. This law amended Law No. 12,965, of 23 April 2014, which deals with the Civil Rights of Brazil for the Internet (Law n. 13,709, 2018).

The Brazilian Personal Data Protection Law is considered a fundamental step for Brazil's insertion in international forums, in addition to providing a safe business environment that increases the attractiveness and materialization of investments in the order of R$ 250 billion (around 50 billion dollars)[5] in transformation technologies digital, according to a study by Brasscom and the consultancy Frost & Sullivan (Borges, 2018).

Although the regulation, the new legislation, does not establish rules and regulations that must be followed to protect data systems or information technology against cyber threats, companies, however, can be held liable if there is data leakage (payment of 2% of the revenue, as long as it does not exceed R$50 million).

Furthermore, there is no obligation to require the industry to report data breaches to authorities. However, data breaches that significantly affect users' assets or cause moral damage are generally reported only to the data owners. Notwithstanding the absence of regulation on data breaches, based on the Brazilian Consumer Protection Law, companies must provide complete information to their consumers about their products and services, in order to guarantee their rights to security and avoid damage or loss.

According to the law, personal information, such as name, address, email, age, marital status, and financial situation, obtained by any means (paper, electronic, computer, sound, and image) is considered personal data (Article 5, I). Images captured by video surveillance, recording of phone calls, IP addresses (computer identification), and locations obtained by the GPS systems are also collected.

In the public sphere, the proposal also provides that the user is informed when the use of their data is released for the fulfillment of a legal obligation or by the administration. On the other hand, the rules do not apply if the information is used by third parties for personal purposes only or if it is used exclusively for journalistic, artistic, or academic content (Article 23, I).

Data on racial or ethnic origin, religious beliefs, political opinions, participation in unions or religious, philosophical, or political organizations,

[5] In August 2021.

data on health or sex life, and genetic or biometric data when linked to an individual must receive differentiated treatment, as they are considered sensitive data (Articles 5, II and 11–13).

In fact, this law is similar to the General Data Protection Regulation — GDPR — of the European Union, which deals with the processing of personal data. That was the outcome of a wide public debate. The objective of the law was to guarantee citizens the control and ownership of their personal information, based on the inviolability of privacy, freedom of expression, communication and opinion, informational self-determination, economic and technological development, as well as the free and free initiative, competition, and consumer protection.

Thus, the law establishes parameters and limits for the processing of personal data, including when this relationship ends. Considering the transnational nature of the flow of this information, the law covers the processing of personal data carried out in Brazil, such as that carried out abroad, but whose collection took place in Brazilian territory.

Ensuring privacy on the Internet means complying with laws regarding the collection and use of personally identifiable information. Potential misuse of personally identifiable and location information should be of concern to consumers and businesses, as well as government agencies, in order to reduce or end cybercrime or cyberattacks.

Application developers and business plans must, therefore, not only comply with current and future legislation related to privacy and security but also alleviate consumer fears and apprehensions. For this, it is important to communicate the terms and conditions of transactions to customers, to obtain their consent to these terms, authenticate them, and develop an adequate solution for electronic records retention.

An organization's risk management function needs a thorough understanding of the constantly evolving risks, as well as the practical tools and techniques available to address them. The risks and opportunities that digital technologies, devices, and media bring us are manifest.

"Cyber risks" mean any risk of financial loss, disruption, or damage to the reputation of an organization from some sort of failure of its information technology systems and or data leakage.

A proposed legislation that addresses the issues of cybersecurity standards' setting and information sharing is fully recommended.

The LGPD also gained an important change with the creation of the National Data Protection Authority (ANPD). The body will be responsible for ensuring the protection of personal data, editing rules and procedures

for its protection, implementing mechanisms by electronic means for registering complaints, and inspecting and applying sanctions when the processing of data is carried out in breach of the law.

Following the legislative trend of the obligation of civil liability and protection of the privacy of citizens' data, the insurance market must cover Data Protection and Cyber Responsibility Insurance. For people who are in the territories protected under the specific law protection (LGPD), it would be an important guarantee of protection; for companies, it means a new set of obligations that deserve care and attention. In case of non-compliance with the LGPD, companies are subject to various penalties. In order not to incur severe losses,[6] they must comply with the new legislation.

Among the obligations is the requirement for responses to incidents, with the notification of people who had their data leaked, and the prompt restoration of the system and services invaded. This means that, in the event of a cyberattack, the company must engage a number of service providers to help identify the source of the attack, restore the system, and recover lost data.

According to Negin Aminian, "one of the most common mistakes that organizations make is not having a comprehensive understanding of the inherent risk that they take on when working with these additional resources. When everyone involved knows what to look out for and what to do should an issue arise, organizations can more proactively manage and mitigate risks before they become bigger problems" (Aminian, 2021).

So, Data Protection and Cyber Liability Insurance can bring precisely that quick response that the customer will need. With a specialized team, it is relevant to be prepared to act in the identification and protection of civil liability risks provided for in the LGPD through the proper placement of insurance, with expertise to assess the exposure and limits of each company, define loss scenarios, analyze coverage gaps, and design the best program for each client, ensuring risk transfer and ongoing support.

The concern with people's privacy and the security of the information used is something commendable, but it is far from due protection when considering the field of dark web or deep web, where state law does not prevail, but the use of inappropriate use of anonymization or pseudonymization for the practice of various cybercrimes.

[6]In Brazil, such as fines of up to 2% of the billing, limited to R$ 50,000,000.00 per infringement.

Dark or Deep Web: Cybercrime and Money Laundering

Cybercrime or computer crime or e-crime or electronic crime or digital crime constitutes any offense committed using a computer, a computer network, or a networked hardware device.

Cybercrimes can be classified as pure or proper, when committed by computer and are carried out or consumed electronically, in which case information technology is the legal object protected, and also as impure or inappropriate cybercrimes, when the machine used is the instrument for carrying out illegal conduct that affects another protected legal asset.

The jurisdiction of the Brazilian Federal Trial Courts is made by exception. It will have jurisdiction in the case of cybercrimes based on Article 109, IV, of the Brazilian Constitution (to the detriment of goods, services, or interests of the Federal Government or its authorities or public companies). However, in cases of inappropriate cybercrimes, the jurisdiction of the specialized justice is based on Article 109, V, of the Magna Carta (crimes provided for in an international treaty or convention, when execution started in the country, the result has or should have occurred in the foreign or vice versa).

A question arises that, being basically transnational, the establishment of jurisdiction in the virtual world would not be so simple, demanding the invocation of prevention and the necessary International Legal Cooperation (Bechara and Flores, 2019).

Brazil has ratified some conventions regarding narcotics, indigenous populations, human trafficking, torture, racism, child pornography and pedophilia, and active corruption and influence peddling in international commercial transactions.

The 2001 Convention of the Council of Europe on Cybercrime in Budapest contemplates a series of illicit conducts that are not restricted to the invasion of computer devices, as provided for in the Carolina Dieckmann Law (Law No. 12,737, of 30 Nov. 2012) (Federal Attorneys Office or MPF, 2020). Brazil is not its signatory.

However, on 30 June 2020, the Senate approved, in a remote deliberative session, the bill to combat fake news. Bill No. 2,630/2020 intends to create the Brazilian Law on Freedom, Responsibility, and Transparency on the Internet, with rules for social networks and messaging services, such as WhatsApp and Telegram. The intention is to avoid fake news that

can cause individual or collective damage and democracy. The text went to the Chamber of Deputies (Brazilian Senate, 2020).

Money laundering can result from malicious cyber activities. Thus, preventing its realization must be a priority. The Internet was discovered by criminals as an effective and potentially clandestine way to launder money internationally.

The world is changing. Since 2020, the first year in which the COVID-19 pandemic reached alarming rates to the point of social isolation and lockdown became current practices in the four corners of the planet, technology has forcibly entered people's daily lives. Cyberspace has covered all continents.

Until recently, great schisms existed between East and West and between North and South, and even a Cold War complete with Socialism derived from Communism. As the idea of a market economy gained prevalence, even in countries without this tradition (such as China), and technological innovations advanced, there was a need for new management practices applied to companies.

Likewise, criminals have evolved over time. Despite the positive hopes brought about by the advent of globalization, cruel and destructive competitiveness has developed. With that, new and growing fears have appeared because we don't know where all of this is going.

The globalization inherent in today's world, with all its advantages and disadvantages, allows for the existence of transnational and technological criminals, whose crimes are practiced in an organized manner, in large conglomerates and unprecedented companies.

As Ronald Griffin puts it, "cyberspace technology, when placed in the wrong hands, is threatening and hostile. Commercial computers roam the scene to compile data about us. Government software spies on people to arrest lawbreakers" (Griffin, 2012).

Financial activity, whether online or not, is often justified by the simple idea that government and market rules alone cannot meet all overt business aspirations — often crossing dangerous and ethical gray areas.

The legal protection of this activity requires government intervention and social and economic regulation so that the existing customary rules are preserved from the seductive economic crime and, therefore, protected.

The object of legal protection is to enable global protection but due to the expectation of general stability fostered by rules that promote the correct and honest functioning of markets (of corporations, public and private securities, and derivative securities).

With regard to criminal law, it is necessary to fill gaps in the many definitions, largely due to the increase in criminal offenses resulting from the exponential growth of economic activity in the State and of international financial relations.

It is clear that criminal law, although it has a subsidiary and fragmented role (last resort), has, here, application. However, the fragmentation, so celebrated and indiscriminately invoked, without a rational basis, can result in a systemic lack of protection for the economic order.

Economic and financial crime, such as money laundering, is a very present issue, whether because of the magnitude of the material damage it causes or because of its ability to adapt and survive social and political changes or even because of its ability to present defenses and defeat all efforts to fight it.

Conceptualizing economic-financial crime is not a simple task as it does not lend itself to simple calculations, given the extent of the resulting damages. The classification of such an offense is based on the collective or supra-individual nature of the legal interests or assets that are to be protected.

Reducing intervention only to classically meritorious facts is as imperative as trimming the criminal responsibility hidden by excessive formalism. However, the fact remains that administrative sanctions alone are not sufficient to make market actors fulfill the basic duties that we, as citizens, are bound by common practices.

In the beginning, criminal law was concerned with protecting the basic institutions of the State and the most elementary interests of citizens. With time, however, in addition to occupying itself with the minimum standards of coexistence, it also began to lend itself to the protection of new social and economic interests. There was, in fact, a radical change in the State's intervention strategies, with laws being enacted that could combat this phenomenon, notably that which promotes the proper functioning of organized crime, money laundering, in order to prevent it from acting on the will from politicians, journalists, judges, businessmen, and so on.

Edwin Sutherland, who defined white-collar crime as that committed by an honorable person with social and professional prestige (Cavero, 2007), tried to explain this type of criminal conduct that, historically, generated little social resistance. The same can be said about cybercrimes that, in the beginning, were timidly the object of societal attention.

This phenomenon may be explained by the criminal's low perception of minimal danger in the absence of direct violence with a specific victim

or even because no small damage can be contemplated. This brings us to the idea developed by Thomas Lynch that serious crimes are those arising from ink rather than blood (Mir and Genovês, 1987). So, perhaps, it involves a certain moral neutrality.

According to José Ángel Brandariz Garcia, given their personal and socioeconomic characteristics, the arrest of financial criminals does not result from the negative social stigma normally existing for common criminals (Garcia, 2000).

There should be no common perception that cybercrime, because in a neutral environment, is less harmful to society than crimes committed by other means, given its penetration into our lives and into the social fabric.

In fact, they end up stimulating common criminality (corruption, unfair competition, and fraud) due to the ease they find, which hinders inspection efforts.

In fact, criminals are highly adaptable to the dynamics of society and do not hesitate to use the dark web in the face of a certain existing tolerance due to technical ignorance, leading to increasingly bold and dangerous criminal behavior.

In addition, criminals carry out a kind of cost–benefit analysis of the gains to be obtained from illegal conduct and possible sanctions (decisions) imposed by the legal system (Fisher, 2011). By performing a utilitarian calculation, the offender can easily conclude that getting caught involves little or no consequences, given the complexity and inefficiency of some criminal justice systems.

Cláudia Cruz Santos argues that *rational choice and situational prevention theories seem to fit them like a glove. Their assessment of the costs and benefits associated with misconduct may deter them from engaging in it if opportunities diminish and the possibility of detection and punishment increases* (Santos, 2001).

The decisive factor is not the will but the impracticality of the behavior prohibited by law. We can no longer afford to theorize about abstract risks and social harm. The categories of financial crimes have to do with increasingly complex regulatory situations and legally intolerable conduct, regardless of the criminal's intentions. These intentions would only emerge later, after the first decision to violate the rules of conduct to which we are all bound.

Financial crimes committed or not through the dark web, given their scope and potential for harm, should be an international concern and involve specialized jurisdictions. In that regard,

> *Much of this jurisdiction is applied to complex crimes, sometimes because of the suspects or defendants involved — people of great economic or political power who, as a rule, operate within a network with international ramifications — and others because of the type of financial crime involved, whether corruption, influence peddling, money laundering, etc. Its seriousness, harm to society and threat to institutions that safeguard the rule of law require a different balance between the rights of the accused and other procedural requirements and the State's duty to prosecute and punish illegal conduct* (de Oliveira, 2011).

This harmful behavior, usually under federal jurisdiction, requires the recognition of economic and financial criminal conduct as a violation of a negative legal duty, that is, to refrain from illegally harming others or public order, in addition to a positive legal duty, that one's behavior is conducive to breaking the collective well-being. This progressive view of legality is increasingly accepted. However, it requires a more complex analysis, involving the aforementioned legal duties (negative and positive) on which a specific legal and criminal assessment is based.

If we can affirm a concern of the justice system, so that it does not appear dual, treating criminals differently (powerful and not powerful), money laundering must have a peculiar approach because it serves both common criminals and socially prestigious ones, deserving, thus, the same institutional treatment.

The reintegration of criminals into society must, therefore, focus on making them rethink their behavior. If, in fact, there is some reasoning behind an illegal conduct that involves cost–benefit analysis of the results for the offender, a certain crime will be committed if, and only if, the expected penalty is not outweighed by the advantages of committing the act (Sánchez, 2004). This also applies in the case of cybercrime, where, to exhaustion, one sees complexity, anonymity, or the use of false identities.

Various transactions are carried out daily over the Internet, and criminal organizations, with their resources, launder large sums. Due to its poor traceability, notably on the dark web, more and more dirty money is used in this cyberspace.

Many criminal issues are raised, including the Chain of Custody Rules for Evidence, especially nowadays that Law No. 13,964, of 24 December 2019 (Anti-Crime Law) made this requirement in the Brazilian Code of Criminal Procedure (Articles 158-A and 158-B), which requires the prosecution to demonstrate it in court. It proves important, as Timothy

A. Vogel says, "the procedures for collecting evidence of a cyber attack" (Vogel, 2012).

The way of laundering money is undergoing changes as criminals have been optimized for e-payment mechanisms through, for example, micro-laundering with the use of electronic addresses, such as *PayPal*, *Picpay*, or *Pagseguro*, to avoid detection. This created an increasing difficulty for many oversight bodies.

Cyberspace has, in fact, changed everything and criminals are using the World Wide Web to steal even data from other devices. For example, the number of complaints from consumers who have become victims of online auction fraud increases annually. While auction e-mail addresses, according to Dara Chevlin, "should be allowed to govern themselves, claiming that their mechanisms are more effective in preventing fraud, statistics clearly show that their efforts are ineffective in curbing the growing problem of fraud at online auctions. Scammers are getting smarter and using the Internet's anonymity to their advantage. Several cases have tried to hold eBay responsible for fraudulent activities that took place on their website, because eBay hosts this activity (host), earning money at the end of every auction held on their website (fraudulent or not). What incentive would eBay have to keep physically closer to its users? Obviously, eBay wants to maintain its reputation. But until host auction sites feel the impact of fraud on their earnings, fraud prevention won't get the attention and investment of resources it deserves." The author recommends imposing stricter federal regulations on hosting auction sites (Clevlin, 2005).

The transition to Internet crime has created unique challenges for law enforcement. Speaking about Internet prostitution, Mellissa Farley, Kenneth Franzblau, and M. Alexis Kennedy say, "The prostitution negotiation includes not just the victim, the buyer, and the dealer/pimp but the most invisible partner: the online advertiser. When prostitution took place on the street, in someone's neighborhood, it was clear whose jurisdiction it was. The enforcement of a series of laws was sometimes fueled by citizens' concern about the nuisance caused in their neighborhoods rather than a concern about the exploitation of prostitution. Communities wanted prostitution out of sight and out of their neighborhoods. As online sex companies are less visible to the public, victims of sexual exploitation are left unprotected and isolated, and may be at more risk." It would, therefore, be up to "policymakers and law enforcement to enforce this and, when necessary, develop new laws and policies that will abolish online (in

addition to offline) trafficking and prostitution. While many have been recruited, sold, and trafficked into prostitution on social media sites, these can also be used against the traffickers and merchants themselves" (Farley *et al*., 2013).

Ronald Griffin mentioned the *Paradigm Alliance* case to exemplify cybercrime. *Paradigm* and *Celebritas* were part of a joint venture and each placed their business interests in the other's hands. One day, long after the relationship began, a *Celebritas* employee hacked into *Paradigm'*s computer. He stole information and paid a sum in a patent application for a new type of software (Griffin, 2012). It's true that almost all employees have been given some form of computer access and an email account. The established network at these companies was increasingly connected to the Internet and other companies, generating a new set of threats.

It is also interesting to mention here two cases decided in the USA involving the so-called dark web. Ross Ulbricht, also known as "Dread Pirate Roberts," was sentenced on 29 May 2015, in federal court in Manhattan to life in prison for his operations and ownership of *Silk Road*, a hidden website designed to allow its users to buy and sell drugs illegal and other illicit goods and services anonymously and beyond the reach of the police between January 2011 and October 2013.

Ulbricht was found guilty on 5 February 2015, on each of the seven charges he faced, following a four-week trial by the Jury. Ulbricht was a drug dealer and criminal who exploited people's addictions and contributed to the deaths of at least six young people. Ulbricht has become the face of cybercrime. Ulbricht created *Silk Road* in January 2011 and owned and operated the underground site until it was shut down by law enforcement authorities in October 2013. *Silk Road* has emerged as the most sophisticated and extensive criminal market on the Internet, serving as an expanding black market and bazaar where illegal goods and services, including illegal drugs of virtually every variety, were regularly bought and sold by e-mail users. While in operation, *Silk Road* was used by thousands of drug dealers and other illegal suppliers to distribute hundreds of pounds of illegal drugs and other illicit goods and services to more than 100,000 buyers and to launder hundreds of millions of dollars arising from these transactions illegally.

Ulbricht deliberately operated *Silk Road* as an online criminal marketplace designed to allow its users to buy and sell drugs and other illegal goods and services anonymously and beyond the reach of law enforcement. Ulbricht sought to anonymize transactions on *Silk Road* in two main

ways. First, he operated the *Silk Road* on the network known as "The Onion Router," or "TOR," a special network of computers on the Internet, distributed around the world, designed to hide the true IP addresses of computers on the network and, thus, the identities of the users of the networks. Ulbricht designed the *Silk Road* to include a Bitcoin-based payment system that served to facilitate illegal trade carried out on the site, including hiding the identities and locations of users transmitting and receiving funds through the email address.

The vast majority of items for sale on *Silk Road* were illegal drugs, publicly advertised as such on the site. As of 23 September 2013, its home page displayed nearly 14,000 controlled substance listings, listed in categories, such as "Cannabis," "Dissociatives," "Ecstasy," "Intoxicants," "Opioids," "Precursors," "Psychedelics," and "Stimulators." From November 2011 to September 2013, law enforcement officers made more than 60 individual secret purchases of controlled substances from *Silk Road* suppliers. These purchases included heroin, cocaine, ecstasy, and LSD, among other illegal drugs, and were honored by suppliers located in more than 10 different countries, including the United States, Germany, the Netherlands, Canada, the United Kingdom, Spain, Ireland, Italy, Austria, and France. Narcotics distributed have been linked to six overdose deaths worldwide. Those deaths included that of Jordan M., a 27-year-old Microsoft employee who died in front of his computer connected to *Silk Road* at the time, passing away as a result of heroin and other medications he administered.

Preston B. (from Perth, Australia) and Alejandro N. (from Camino, California), both aged 16, died as a result of using 25i-NBOMe, a powerful synthetic drug designed to mimic LSD (commonly referred to as LSD "N-Bomb"), which was purchased on *Silk Road*. Additional victims included Bryan B., age 25, of Boston, Massachusetts, and Scott W., age 36, of Australia, who died as a result of heroin acquired on the *Silk Road*, and Jacob B., age 22, of Australia, who died of health complications that were aggravated by the use of medicines also purchased on *Silk Road*.

In addition to illegal narcotics, other illicit goods and services were purchased and sold openly on this website. For example, as of 23 September 2013, there were the following: 159 listings in the "Services" category, most of which offered computer hacking services, such as a listing of a vendor that offered to hack into social media accounts of their choice; 801 listings in the "Digital Goods" category, including malicious software, hacked accounts on various online services, and pirated media content; and 169 listings in the "counterfeit" category, including offers to

produce counterfeit driver's licenses, passports, Social Security cards, credit card statements, car insurance records, and other forms of forged identification documents.

Using the online nickname "Dread Pirate Roberts," or "DPR," Ulbricht controlled and supervised all aspects of the *Silk Road* and managed a team of paid online administrators and computer programmers who assisted in the daily operation of the site. Through his ownership and operations of Silk Road, Ulbricht earned commissions worth more than $13 million generated by illicit sales made on the site. He has also demonstrated a willingness to use violence to protect his criminal enterprise and the anonymity of its users, ordering six murders for hire related to the site's operations, although there is no evidence that these murders were actually carried out.

Ulbricht (from San Francisco, California) was convicted of seven counts after a four-week jury trial: distributing narcotics, distributing narcotics through the Internet, conspiring to distribute narcotics, participating in an ongoing criminal enterprise, conspiring to commit hacks, conspiring to carry in false identity documents, and conspiring to commit money laundering. In addition to life imprisonment, he was sentenced to lose $183,961,921 (US DOJ, 2015).

On the other hand, the dark web's biggest marketplace — where hundreds of thousands of criminals have anonymously bought and sold drugs, weapons, hacking tools, stolen identities, and a host of other illegal goods and services — has been shut down as a result of one of the most popular efforts, sophisticated and coordinated to date by law enforcement agencies around the world.

In early July 2017, several computer servers used by the *AlphaBay* website were seized around the world, and the website's creator and administrator — a 25-year-old Canadian citizen living in Thailand — was arrested. *AlphaBay* has operated for over two years and had transactions in excess of $1 billion in bitcoins and other digital currencies. The site, which operated on the anonymous TOR network, was a major source of heroin and fentanyl, and sales from *AlphaBay* were linked to several overdose deaths in the United States. This case was considered a historic operation because there were several servers in different countries, hundreds of millions in cryptocurrencies, and a dark net drug trade that spread across the world.

AlphaBay was truly a global website and vendors were shipping illegal items to and from places all over the world. The site, an offshoot of

earlier obscure market e-mail addresses, such as *Silk Road* — but much larger — was launched in December 2014. It took about six months for the underground market to gain momentum, but after that, it grew exponentially. *AlphaBay* reported that it served more than 200,000 users and had approximately 40,000 providers. At the time of removal, the site had more than 250,000 ads for illegal drugs and toxic chemicals and more than 100,000 ads for stolen and fraudulent identification documents, counterfeit products, malware and other hacking tools, firearms, and fraudulent services. By comparison, the Silk Road market — the largest company of its kind before it closed in 2013 — had approximately 14,000 listings.

The operation to seize *AlphaBay's* servers was led by the FBI and involved the cooperative efforts of law enforcement agencies in Thailand, the Netherlands, Lithuania, Canada, the UK, and France, along with the agency Europol.

US authorities also worked with several foreign partners to freeze and preserve millions of dollars in cryptocurrencies, which represented the proceeds of *AlphaBay's* illegal activities. Its creator and administrator Alexandre Cazes — who was given the name *Alpha02* and *Admin online* — was arrested by Thai authorities on behalf of the US on 5 July 2017. A week later, Cazes apparently attempted to take his own life while under custody in Thailand.

Because *AlphaBay* operated on the anonymous TOR network, administrators were confident they could hide the site's server locations and user identities. The FBI and its partners used a combination of traditional investigative techniques, along with sophisticated new tools, to solve the case and dismantle *AlphaBay* (*FBI News*, 2017; US DOJ, 2017).

In turn, *Our Father* (Pai Nosso) Operation, launched on 3 March 2018, as an offshoot of *Carwash* case[7] in Rio de Janeiro, made an

[7]Car Wash (Lava Jato) in 2014 up to 2021 (involvement of Petrobrás, a mixed-capital company, Odebrecht, a big private contractor, and JBS, a large rural producer and cattle processor). It has revealed a big scheme of corruption that also allows the campaigns financing and the distortion of the most important bidding principles, like equality, transparency, and administrative probity. Arrests of those convicted demonstrated that actions were properly being undertaken by federal police, public prosecution, and the Judiciary, showing that the country is acting to correct its course. The conclusion that public funds had been deviated to supply a plot, with spurious payments to many congressmen, left clear how bold, voluptuous, and neglecting the actions of these groups were, in order to

unprecedented discovery at the time: for the first time, a money laundering scheme using bitcoin was discovered in Brazil. Confirmation of the use of the cryptocurrency was given by the Internal Revenue Service. On the occasion, a delegate and a former secretary of Sérgio Cabral, former governor of the State of Rio de Janeiro, were arrested, suspected of overcharging bread for the Secretariat of Penitentiary Administration — SEAP. The transaction was a test to circumvent the public financial control agencies, totaling R$300,000 (US$60,000) in bitcoins. According to the investigations, the suspects would have diverted at least R$73 million (US$14.6 million) from the public coffers with a scheme of overpricing and fraud in the supply of bread to state jail inmates. As for the use of digital currency, the idea was to receive money abroad using an instrument that is not regulated in most countries, through remittances abroad. The operation investigated irregularities in the provision of breakfast and snacks to detainees, under a contract (no longer in force) that involved the operation of bakeries within the Bangu complex. SEAP paid twice for the bread that was provided to prisoners (Tolotti, 2018).

In 2019 and 2020, the Federal Appeals Court for the 3rd Region unanimously decided to deny two Habeas Corpus (HC No. 5018581-89.2019.4-03.0000 on 26 September 2019 and HC No. 5007613-63.2020.4-03.0000 on 14 May 2020), relating to Singular Operation, triggered by the Federal Police which, based on surveys in the so-called dark web, a virtual environment not available in Internet browsers except for specific applications, having discovered the existence of a "Telegram" group intended for to support the sale of credit card information on surface Internet site. The group, dispersed in distant cities in Brazil (Fortaleza, in Ceará state, Sao Paulo and Praia Grande, both in Sao Paulo state, and Tapes and Santa Maria, both in Rio Grande do Sul state), would operationalize the practice of carding, a fraudulent activity of capturing bank card data for use in purchases and withdrawals (Federal Appeals Court for the Third Region (TRF3), 2020–2019)".

achieve their objectives: money laundering of R$20billion (US$4 billion), including R$10billion (US$2 billion) of kickbacks, and huge self-enrichments. In the original case, a Search Warrant in 2014 was issued to the headquarters of Mossacka Fonseca Firm, mentioned in "Panama Papers." It was the first judicial order against it. Also, from Car Wash case, other criminal activities were discovered with the involvement of state and local politicians (like in Rio de Janeiro State) thanks to, for instance, the distortion of funds of the 2014 FIFA World Cup and 2016 Olympic Games organizations.

There are some behaviors that would demonstrate the criminal practice or its preparation, for example, *phishing* (electronic fraud), improper data mining, and auction fraud using personal information to assume someone else's identity and seeking to obtain loans or credit cards or flash drives and downloading trade secrets. Often, GPS or computers are used to locate passwords.

The perpetrator of a cyberattack could be a hacker or cracker. Hackers is a term generally used for those who attack another device for fun, whereas crackers do it for profit. Timothy Vogel reveals that both may be genuinely interested in thwarting the latest computer security technologies, not to profit from it but simply to know that they can surpass even the most sophisticated security measures (Vogel, 2002).

Law No. 12,737, of 30 November 2012, brought to the Brazilian criminal legal system the new crime of "Invasion of Computer Devices," provided for in Articles 154-A and 154-B of the Brazilian Penal Code, consistent in the conduct of "invading another's computer device, connected or not to the computer network, by improperly violating a security mechanism and in order to obtain, tamper with, or destroy data or information without the express or tacit authorization of the device holder or install vulnerabilities to gain an advantage illicit." The penalty is not high (detention, from three months to one year, and fine), and it will apply to whoever produces, offers, distributes, sells, or broadcasts a device or computer program in order to allow the practice of the conduct defined in the head. Only if the victim is represented can the crime be investigated.

The invasion, as required by law, must be by an "alien" computer device and "by undue violation" of a "security mechanism" (normative elements of the crime). Anyone who entered the computer device itself is not incriminated. In addition, the violation must be "undue," that is, unauthorized and without just cause. Obviously, the IT technician who overcomes a protective mechanism to repair the switchgear does not commit a crime also because he has the express or at least tacit authorization of the client. The Police Authority does not commit the crime either, which seizes computer equipment by court order and orders its contents to be examined for criminal investigation.

It should be noted, however, that this authorization cause must exist from the beginning to the end of the agent's behavior and the agent must adhere to its strict reasonable limits. For example, if a computer technician is authorized to breach someone's system access keys for repair purposes and does so but then collects private photos stored there, he

maliciously corrupts information or data by going beyond the limits of his work without the authorization of the holder and starts to commit the criminal offense. It is important to emphasize that, as there is no guilty figure, the very common mistake in which the computer technician, when performing a repair, formats the computer and ends up destroying important content for the person without malice, but through negligence or malpractice, does not constitute a crime. There may, however, be a civil infraction subject to indemnity for moral and/or material damages (Cabette, 2013).

To avoid these irregularities, especially cybersecurity risks, it is important that guidelines are issued to effectively detect, prevent, and respond to fraud and similar behavior, essentially through written policies and through appropriate laws. Brazilian law was timid as it did not provide for several existing actions in the fraudulent practice of using the Internet and in its security.

This guidance should provide effective procedures and controls to protect against identified risks and assess the full range of risk areas related to fraud and the like, including, where applicable, market manipulation and assignment of responsibilities.

An important innovation brought about the Anti-Crime Law (Law No. 13,964, of 24 December 2019) which created the figure of Virtual Infiltration, including Article 10-A of the Law on Organized Crime No. 12,850, of 2 August 2013), which allowed admitting that agents of the State infiltrate criminal organizations for six months, renewable, not exceeding 720 days, requiring judicial authorization and being the only means of obtaining evidence.

Here, the jurisprudence will certainly have to be revisited, bearing in mind the theory of impossible crime, which was until then enshrined. According to Binding Jurisprudence Summary 145 of the Federal Supreme Court, "there is no crime when the preparation of the *flagrant delicto* by the police makes its consummation impossible," that is, there is no crime when the fact is prepared by provocation or inducement, directly or by competition, of authority, which does so in order to prepare or arrange the flagrant.

It is important to note that the concept of criminal organization encompasses the *association of 4 (four) or more people structurally ordered and characterized by the division of tasks, even informally, with the objective of obtaining, directly or indirectly, an advantage of any nature, through the practice of criminal offenses whose maximum*

penalties are greater than 4 (four) years, or which are of a transnational nature (article 1, paragraph 1, of Law No. 12,850/2013).

However, the association of at least four people would be enough to commit serious crimes or crimes of a transnational nature, as, incidentally, is characterized by cybercrime par excellence in most cases.

Conclusion

There is no doubt that digitization and artificial intelligence, big data, and the Internet of Things have impacted our lives, rapidly changing all facets of our societies and economies. The scale of technological transformation is causing uncertainty, and a clear framework approach is needed to ensure that this transition can help build more innovative and inclusive economies and prepare all sectors of society for these changes. However, that's just part of the problem.

Another part, unfortunately, is the use of this accelerated pace of digital transformation for illicit activities, forcing governments to debate the issue and consider that this important tool deserves updating for adequate social protection. The increase in the inappropriate use of the World Wide Web must be taken into account. The technology can be considered a new frontier of crime with the increasing virtualization of the world and the dramatic growth of currencies, challenging authorities to stop criminals from hiding in cyberspace.

Digital money represents the future of banking. By allowing individuals to transfer money quickly and anonymously, without tying to the generation of reports with known pertinent details, it is an ideal mechanism for money laundering. The virtual currency entity must provide for the effective investigation of fraud and other irregularities, provide effective procedures and controls to protect against identified risks, and allocate responsibilities for their handling.

On the other hand, cyberspace must be an environment in which people's rights, especially their personal data, can be effectively protected against all kinds of criminal savagery. Regulation must consider the role of users and providers in balancing the protection of freedom of expression and citizens' rights.

The tensions between free cyberspace and the need for regulation are at a historically relevant level and are especially evident in the restrictions and disputes over tax exemptions and the neutral use of the Internet.

Freedom of expression alone is not a sufficient reason to be considered tax-free or free from government oversight.

Cybercrime, as noted, given its peculiarity, challenges the State in an attempt to prevent and repress crime. At stake is the issue of territoriality principle (since the Internet is, par excellence, a transnational space), with its indispensable international legal cooperation and, regarding the means of discovery (especially when practiced in a dark or deep web environment), also virtual infiltration is necessary (now introduced by the Anti-Crime Law, Law No. 13,964, of 24 December 2019), which will require a reassessment of the Brazilian jurisprudence that established the theory of the impossible crime in the case of forged or provoked flagrant.

To avoid money laundering, one solution would be to find a way to track and monitor virtual money flows. The Internet plays an important role in ensuring and maintaining compliance with social norms through stronger individual and institutional connections. The cybercrime phenomenon intensifies the competition among security, protection, technology, and technical knowledge of the authorities, and the implementation of action plans, in addition to preserving the freedom of the Internet and the benefits of online trading. It has the task of reconciling values and public interest in their regular use.

In today's digital world, companies are collecting more customer information than ever before. These sensitive data allow organizations to optimize customer experiences and guide future decisions, but it also exposes them to many risks, especially if critical information or intellectual property is not properly protected. Organizations should review their industry regulations regarding data protection to ensure that proper security measures are taken.

As a result of a network disruption, cyber risks are typically defined by three components: *threat* (from criminal enterprises), *vulnerability* (Internet system weakness), and *consequence* (harm, damages, and money laundering). As data privacy increasingly becomes a concern for customers, more regulatory compliance standards (like GDPR) are being put into place. While these regulations are an important point of consideration that should be followed, it's important to understand that maintaining compliance with these standards does not guarantee an organization is secured from attackers (Aminian, 2021), especially when dark web is considered.

So, cybersecurity risks must become a leading priority for organizations as they embrace digital transformation and leverage advanced technology solutions to drive business growth and optimize efficiencies.

Although it is an enabling tool for everyday life, the Internet has become an intriguing and complex tool, as well as a way to keep people anonymous, making it difficult to contain international crimes. That is why cyberspace reflection and the need for a well-designed international regulatory environment are rigorous, in addition to the necessary updating of the corresponding criminal legislation.

References

Aminian, N. (2021). What is cybersecurity risk? Definition & factors to consider. https://securityscorecard.com/blog/what-is-cybersecurity-risk-factors-to-consider. Accessed on 5 August 2021.

Bechara, F. R. and Flores, D. M. (2019). Crimes Cibernéticos: Qual é o lugar do crime para fins de aplicação da pena e determinação da competência jurisdicional? *Revista de Direito Mackenzie*. http://editorarevistas.mackenzie.br/index.php/rmd/article/view/13357/10572. Accessed on 4 August 2021.

Borges, B. (2018). Lei geral de proteção de dados pessoais avança no Senado. *JOTA*. https://www.jota.info/coberturas-especiais/liberdade-de-expressao/protecao-de-dados-senado-02072018. Accessed on 4 August 2021.

Brazilian Senate (2020). *Senado aprova projeto de combate a notícias falsas; texto vai à Câmara*. https://www12.senado.leg.br/noticias/materias/2020/06/30/aprovado-projeto-de-combate-a-noticias-falsas. Accessed on 4 August 2021.

Cabette, E. L. S. (2013). O novo crime de Invasão de Dispositivo Informático. https://www.conjur.com.br/2013-fev-04/eduardo-cabette-crime-invasao-dispositivo-informati-co#:~:text=A%20Lei%2012.737%2C%20de%2030,seguran%C3%A7a%20e%20com%20fim%20de. Accessed on 5 August 2021.

Cavero, P. G. (2007). *Derecho Penal Económico — Parte General*, 2nd ed. Lima: Grijley.

Clevlin, D. (2005). Schemes and scams: Auction fraud and the culpability of host auction web sites. *Loyola Consumer Law Review*, 18, 223–255.

de Oliveira, E. P. (2011). *Direito e processo penal na justiça federal: Doutrina e jurisprudência*, p. 71. São Paulo: Atlas.

Farley, M., Franzblay, K. and Kennedy, M. A. (2013). Online prostitution and trafficking. *Albany Law Review*, 77, 1039–1094.

FBI News (2017). Darknet takedown. Authorities shutter online criminal market AlphaBay. https://www.fbi.gov/news/stories/alphabay-takedown. Accessed on 4 August 2021.

Federal Appeals Court for the Third Region (TRF3) (2020–2019). *Habeas Corpus* no. 5018581-89.2019.4-03.0000, unanimous decision on 26 September 2019

https://web.trf3.jus.br/acordaos/Acordao/BuscarDocumentoPje/90776892, and no. 5007613-63.2020.4-03.0000, unanimous decision on 14 May 2020 https://web.trf3.jus.br/acordaos/Acordao/BuscarDocumentoPje/132081706.

Garcia, J. Á. B. (2000). *El delito de defraudación a la seguridad social*, pp. 80–81. Valencia: Tirand lo Blanch.

Griffin, R. C. (2012). Cybercrime. *Journal of International Commercial Law and Technology*, 7(2), 136–153.

Law n. 12.737, of 30 November 2012. *Planalto*. http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12737.htm. Accessed on 3 August 2021.

Law n. 12.965, of 23 April 2014. *Planalto*. http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Accessed on 3 August 2021.

Law n. 13,709, of 14 August 2018. *Planalto*. http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/Lei/L13709.htm. Accessed on 3 August 2021.

Loureiro, R. M. and Palhares, L. A. F. (2018). Brazil cybersecurity 2018. Getting the deal through. *Law Business Research*, 2017, 17. https://gettingthedealthrough.com/area/72/jurisdiction/6/cybersecurity-brazil/. Accessed 3 August 2021.

Mir, J. R. S. and Genovês, V. G. (1987). *Delincuencia de cuella blanco*, p. 71. Madrid: Instituto de Estudos de Política.

MPF Site (2020). http://www.mpf.mp.br/atuacao-tematica/sci/normas-e-legislacao/legislacao/legislacoes-pertinentes-do-brasil/docs_legislacao/convencao_cibercrime.pdf. Accessed on 4 August 2021.

Rockefeller IV, J. D. (2012). Letter from Senator Johan D. Rockeffer IV (D-WV) to Virginia M. Rometty, President and Chief Executive Officer, International Business Machines. https://www.cadwalader.com/uploads/cfmemos/f6e347976c1cacb03e7d982e936e12cf.pdf. Accessed on 2 August 2021.

Sánchez, J.-M. S. (2004). Eficiência e direito penal. Coleção Estudos de Direito Penal. No. 11. São Paulo: Manole.

Santos, C. C. (2001). O crime de colarinho branco (da origem do conceito e sua relevância criminológica à questão da desigualdade na administração da justiça penal), p. 175.

The Cybersecurity Act of 2015. Office of the Director of National Intelligence. https://www.dni.gov/index.php/ic-legal-reference-book/cybersecurity-act-of-2015. Accessed on 2 August 2021.

The Cybersecurity Act of 2017, in U.S. Congress website. https://www.congress.gov/bill/115th-congress/senate-bill/770. Accessed on 2 August 2021.

The General Data Protection Regulation (GDPR) (2016). EU Official Journal Issue, L 119. http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L:2016:119:FULL&from=EN. Accessed on 2 August 2021.

The United States Department of Justice Website (US DOJ) (2015). Ross Ulbricht, A/K/A "Dread Pirate Roberts," Sentenced in Manhattan Federal

Court to Life in Prison. https://www.justice.gov/usao-sdny/pr/ross-ulbricht-aka-dread-pirate-roberts-sentenced-manhattan-federal-court-life-prison, published on 29 May 2015. Accessed on 28 June 2022.

The United States Department of Justice Website (US DOJ) (2017). Justice News. AlphBay, the Largest Online "Dark Market," Shut Down. https://www.justice.gov/opa/pr/alphabay-largest-online-dark-market-shut-down. published on 20 July 2017. Accessed on 28 June 2022.

Tolotti, R. (2018). Lava Jato descobre primeiro esquema de lavagem de dinheiro usando bitcoins em desdobramen-to da Operação. https://www.infomoney.com.br/mercados/lava-jato-descobre-primeiro-esquema-de-lavagem-de-dinheiro-usando-bitcoins-em-desdobramento-da-operacao/. Accessed on 4 August 2021.

Vogel, T. A. (2002). Dealing with cyber attacks on corporate network security. *The Practical Lawyer*, 48, 35, 35–46.

https://doi.org/10.1142/9781800612860_0014

Chapter 13

Discussing Regulation for Ethical Hackers

Georg Thomas

Introduction

Over the past decade, the importance of cybersecurity has emerged within organizations and institutions. According to Gartner (2021), worldwide spending on security and risk management is forecasted to exceed $150B USD in 2021. With such significant spending, Boards and Senior Leadership within organizations will want some confidence that the investments they are making are effective. Likewise, organizations are seeking some sort of assurance that not only their own security controls are effective but also those of their partners and suppliers. Engaging a security professional with hacking skills to validate the security posture of an organization or institution is now a common practice across the world. These professionals possess the necessary skills and experience to identify vulnerabilities within an organization's systems or processes which should be remediated before they are exploited by a threat actor (Thomas, 2020).

The increase in demand for such individuals and the relatively new nature of the field means that there is a skills shortage of cyber talent and those individuals that do possess the skills often receive high pay (Thomas *et al*., 2018) receiving salaries well into the six figures. For some, even the very idea of "being a hacker" could make such a field very attractive. This

demand and the earning potential for such a field means that many are choosing to either start a career in this field or shift into it from their existing profession, such as those in Information Communications and Technology (ICT). However, when considering the important function of such a role, an emphasis should be placed on ensuring that adequate skills, experience, and ethics are possessed. Currently, there are no official requirements an individual needs to practice. That is not to say that options for education, qualification, and codes of ethics or conduct do not exist but rather that they are voluntary, and this creates potential significant risks.

This chapter will explore what a hacker is, how such a field is necessary and aligns with the ever-growing increase in legislation and regulation related to data protection and privacy (discussed later in this chapter), potential issues, and discuss the concept of regulating the field.

Who or What Is a Hacker?

When we think of the term "hacker," thoughts of an individual sitting in a dark basement in a hoodie with computer code all over their monitors might come to mind or, perhaps, visuals of the infamous Guy Fawkes mask frequently associated with hacking group Anonymous. The term hacker is thought to have originated at MIT in the 1960s, which started with the tradition of creating attention-seeking pranks called hacks that later evolved to having the ability of invention within the context of electrical engineering (Wark, 2006).

Today, we think of a hacker as anyone who can "hack" or manipulate a computer system to get it to do something it was not designed to do. However, the modern hacker's skill set is much broader than that. Hacking has evolved to not only manipulate computer systems but often also people. This has been observed through the prevalent use of e-mail-based scams known as "phishing" which tricks the recipient into disclosing confidential or sensitive information which is then later used for nefarious purposes.

Hackers are also often divided into three categories, which reflect their motives: white hat, gray hat, and black hat. Let us look at the different categories of hackers in more depth.

White hats

Described as "the good guys," white hats are the security professionals that conduct security testing engagements on organizations and

institutions. They are known by many different names, such as white hat hackers, ethical hackers, security testers, and penetration testers.

Although it is common for the names to describe such hackers to be used interchangeably, the definition of the roles does vary. For example, penetration testers often focus on specific systems, whereas ethical hackers are broader and include the skills of a penetration tester. No matter the role or term used to describe them, white hats have a common motive, which is that they are using their skills for good and ethical purposes.

White hat hackers often utilize the same tools and techniques as used by the other categories of hackers to accomplish their goals, however, a key difference is that white hat hackers are given express permission by the target to conduct the attacks (Thomas *et al*., 2018).

Black hats

Unlike a "white hat" hacker who utilizes their skills for good and ethical purposes and with the express permission of the target organization, black hat hackers are motivated by personal gain. Black hats, also called "crackers," use their skills for malicious or illegal purposes (Graves, 2010). Black hats will attack their target, often using a variety of tools and techniques. Once access has been obtained, they will commence activities, such as disruption of access to resources, destruction or corruption of data, and information theft. Often, further attacks may be initiated, such as launching against additional targets from the victim's network, social engineering, or other fraudulent activities.

Gray hats

Gray hat hackers like black hats operate outside of the rules. Where gray hats differ is that their motives are generally not malicious. They will use the same tools and techniques to gain access to a system or network as a black or white hat, but they will not be authorized by the target and, therefore, their activities are illegal. Gray hats like white hats will identify issues and are known to offer to remediate the security weaknesses in exchange for a small fee.

Hacktivists

Often, the topic of hacktivists comes up and what category they fit into. Like black hats and gray hats, hacktivists do not obtain authorization to

attack their targets, but unlike black hats, their motives are not considered personal gain rather they are motivated by a cause whether it be social, political, ideological, or otherwise. Due to this, they are often categorized as "gray hats" because their motives and behaviors fall between that of white and black hats.

There are several examples of hacktivism over the years, for example, in 2012, #OpSaveTheArctic was carried out by hacking group Anonymous. This campaign is considered environmental hacktivism as it was against several energy companies including Exxon Mobil, Shell, and BP who were drilling in the Arctic. The attack results in the disclosure of hundreds of e-mail addresses and passwords, which were then used to support Greenpeace's Save the Arctic Campaign by signing their petition to stop drilling (#SaveTheArctic — Phase I, 2012).

Nation states

Nation state actors are those hacking groups that are affiliated with government organizations. Such groups can be either government personnel or contractors and are engaged for a variety of different purposes including but not limited to espionage, information theft, sabotage, destruction, and financial gain. Nation state groups are known to be advanced and sophisticated attackers often able to infiltrate and remain persistent within a target's infrastructure for extended periods of time. These types of attackers are referred to as Advanced Persistent Threats (APTs). Since Nation State actors are government sanctioned, their activities are not considered illegal within their jurisdiction and they are acting for the benefit of the nation they represent, and subsequently they are considered gray hats.

This chapter will focus on Ethical Hackers, which fall within the white hat category.

What Is Privacy?

The concept of privacy is difficult to define as what privacy is differs between countries, individuals, and cultures and there is no one uniform and consistent definition of what privacy is. The Office of the Australian

Information Commissioner (OAIC) identifies the following three elements to provide a general definition of privacy:

- "To be free from interference and intrusion";
- "To associate freely with whom you want"; and
- "To be able to control who can see or use information about you" (Office of the Australian Information Commissioner, 2021).

With the significant increase in technology adoption over the past few decades, it is unsurprising that there is a large association with data protection and subsequently this places an emphasis on the OAIC's point about having the ability to control who can see or use information about individuals. Although an individual's right to privacy varies across jurisdictions, privacy has been around for hundreds of years. For example, in Colonial America, there were privacy laws that would protect against eavesdropping (Solove, 2006). Data privacy is also the focus of Europe's data privacy and security law, Regulation 2016/679, better known as the European Union General Data Protection Regulation (EU GDPR).

As the use of technology and storage of personal data continues to increase, the need for appropriate laws to protect the privacy of individuals has continued to evolve. According to the United Nations Conference on Trade and Development, 66% of countries have legislation and 10% have draft legislation on data protection and privacy (UNCTAD, 2021). Here are a few examples of data privacy laws from around the world:

Australia

Privacy Act 1988 (*Cth*)

The Privacy Act was introduced in Australia in 1988 with the intention of promoting and protecting the privacy of individuals and includes 13 privacy principles (Australian Privacy Principles or APPs) that apply to some private sector and government agencies in Australia (Office of the Australian Information Commissioner, n.d.). An important addition to the Privacy Act was the introduction of the Notifiable Data Breaches scheme (NDB scheme) in 2018. The NDB scheme requires notification to the OAIC and affected individuals in the event a data breach occurs that results in serious harm to those individuals (Thomas *et al*., 2019). Penalties

for non-compliance are $420,000 (AUD) for individuals and $2.1 million for organizations.

Canada

Personal Information Protection and Electronic Documents Act (*PIPEDA*)

Canada's data privacy and security law PIPEDA includes many requirements relating to consent, collection, use, and disclosure of personal information (Office of the Privacy Commissioner of Canada, 2019). In 2018, Canada introduced a data breach notification requirement as part of the Digital Privacy Act (2015) which includes significant amendments to PIPEDA including the requirement to notify affected individuals and the Canadian Office of the Privacy Commissioner (OPC) in the event of a data breach (Thomas *et al*., 2019). PIPEDA also includes requirements for ensuring appropriate safeguards are implemented.

European Union

General Data Protection Regulation (*GDPR*)

Replacing the European Union Data Protection Directive (95/46/EC) that was enacted in 1995, the European Union General Data Protection Regulation 2016/679 referred to more commonly as the EU GDPR introduces new data protection obligations as well as reinforces previous obligations (Voigt, 2017). The EU GDPR includes requirements to notify the supervisory authority within 72 hours as well as affected individuals without undue delay in the event of a data breach (O'Brien, 2016). Organizations who are considered "data controllers" or "data processors" under the regulation must also appoint a Data Protection Officer (DPO) who is responsible for ensuring compliance with the EU GDPR and liaising with the supervisory authorities. Penalties for non-compliance include fines of 4% of annual global turnover or up to €20 million.

Philippines

Republic Act No. 10173 (*Data Privacy Act*)

The Republic Act No. 10173 also known as the Data Privacy Act in the Philippines is intended to protect all forms of information. Similar to the

EU GDPR, the Data Privacy Act requires the appointment of a Data Protection Officer (DPO) to ensure compliance with the Act (Data Privacy Philippines, n.d.). In the event of a breach, there is a requirement to promptly notify the NPC and affected individuals. Penalties for non-conformance breaching the Act include imprisonment and fines between ₱500,000 and ₱2 million.

Singapore

Personal Data Protection Act (*PDPA*)

The Personal Data Protection Act in Singapore protects personal data by establishing a baseline standard of protection (Personal Data Protection Commission Singapore, n.d.). As with other laws and regulations, the PDPA governs the collection, use, and disclosure of personal data and in 2021 implemented mandatory breach notification requirements. Both the Singapore Personal Data Privacy Commissioner (PDPC) and affected individuals must be notified within 72 hours of determining the breach is notifiable.

United Kingdom

Data Protection Act 2018

The Data Protection Act 2018 (DPA, 2018) controls the use of personal information in the United Kingdom. Prior to the implementation of the DPA 2018, organizations, businesses, and government in the United Kingdom were required to comply with the EU GDPR. The DPA 2018 is UK's implementation of GDPR (GOV.UK, n.d.). The DPA 2018 has rules on the use, accuracy, retention, and handling of personal data and includes requirements to ensure the security of information processed. The Act requires notification of the Privacy Commissioner within 72 hours and, where there are high-risk situations, affected individuals. Penalties for breaches of the DPA 2018 include 2% of global turnover or up to €10M for failure to notify and up to 4% of global turnover or €20M for more serious breaches.

United States of America

The United States of America does not generally have a federal privacy law. However, there are a variety of State-based laws and specific regulations. For example, the New York Department of Financial Services (NYDFS)

introduced Cybersecurity Regulation (23 NYCRR 500) in 2018 that applies to covered entities in banking, finance, and insurance as defined by the regulation. The regulation introduced requirements, such as the appointment of a Chief Information Security Officer (CISO) as well as the requirement for annual penetration testing services. The regulation even requires that cybersecurity personnel (including third parties) are qualified (New York State Department of Financial Services, 2020).

In 2018, California introduced the California Consumer Privacy Act of 2018 (CCPA). The CCPA is like other regulations and legislation, such as the EU GDPR, in the handling requirements of personal information. The CCPA also has notification requirements, which require notification of the Californian resident whose personal information was (or is reasonably believed) to have been acquired by unauthorized parties (State of California Department of Justice, 2021).

Rules and Regulations for Hacking

As identified, there are several data protection and privacy laws, many of which include requirements for notification of regulators, governing bodies, or individuals that have been affected because of a breach. It is crucial that organizations comply with the relevant laws and regulations as well as any other contractual requirements that they have entered. Organizations that store and process sensitive information, such as personal or health information, will likely be subject to such laws and regulations and to prevent the disclosure of such sensitive information, adequate controls must be implemented; there may even be a legal or regulatory requirement to implement such controls as seen in the NYDFS Cybersecurity Regulation.

The use of ethical hackers to validate the effectiveness of the security controls that an organization has implemented is considered a best practice approach to determining the security posture and identifying any potential risks and vulnerabilities (Thomas, 2020). However, there are some considerations that need to be considered when engaging such personnel.

One key consideration is the conduct and ethics of such security personnel. The adoption of a code of conduct or ethics is widely used by many professionals, and this is generally mandated through professional associations if the ethical hacker is a member. There are several

well-known bodies that issue ethical hacking or security testing-related certifications. These organizations have codes available for their members.

Codes of conduct/ethics

CREST

Established in the United Kingdom in 2006, CREST (crest-approved.org) is one of the most recognized and highly regarded organizations for accrediting organizations and personnel who provide penetration testing services. Both organizations and personnel are required to meet specific requirements to obtain accreditation, which is intended to provide assurance of a standard level of skills and experience are held. CREST now has chapters across the globe including in America, EMEA, Asia, and Australia. Members of CREST are required to abide by the CREST Codes of Conduct, which includes requirements and guidance on good practices, professional representation, regulations, competency, and ethics to name a few.

EC-Council

The International Council of Electronic Commerce Consultants is an organization based out in the United States that provides training and certification for cybersecurity professionals. Their most well-known certification is the Certified Ethical Hacker (CEH) credential, but there are also several additional trainings and certifications that EC-Council issues including Licensed Penetration Tester (LPT) and Certified Penetration Testing Professional (CPENT). Members of EC-Council are required to abide by the Code of Ethics which includes requirements on confidentiality, competence, and ethical conduct.

Offensive Security

Offensive Security's hacking and penetration testing certifications are well regarded within the industry. Their certifications such as the Offensive Security Certified Professional (OSCP) which is a 24-hour exam and the Offensive Security Certified Expert (OSCE) which is a

48-hour exam are both hands-on and designed to validate the practical skills of ethical hackers. Offensive Security identifies several core values to guide the way people behave.

GIAC

Global Information Assurance Certification (GIAC) is an information security certification body based out in the United States. GIAC issues several certifications in different areas including cyber defense, management and leadership, digital forensics, incident response, cloud security, and offensive operations, such as penetration testing. These offensive certifications include GIAN Penetration Tester (GPEN), GIAC Web Application Penetration Tester (GWAPT), GIAC Exploit Researcher, and Advanced Penetration Tester (GXPN), among others. GIAC members are required to abide by the Code of Ethics and to date, 122 individuals have been revoked for violations (GIAC Certifications, 2021).

There are other organizations that also require members to abide by codes including the Australian Computer Society (ACS) and International Information System Security Consortium, better known as $(ISC)^2$. In Australia, the ACS is the largest professional body and represents the Information Communications and Technology (ICT) sector. In 2018, the ACS introduced the cybersecurity specialism to its certifications, which recognized penetration testers (as well as other cybersecurity areas) (Thomas, 2020).

Current Mitigation Strategies

Many organizations and institutions have recognized that there is some risk involved with engaging third parties and even specifically ethical hacking. The last few years have seen an increase in supply chain risk management and specifically third-party providers of services. Some organizations now conduct thorough processes to vet organizations and the people they engage. Standards and frameworks, such as ISO/IEC 27001:2013 and the NIST Cybersecurity Framework, include control requirements to help address some of these risks.

Confidentiality Agreements

Confidentiality agreements and non-disclosure agreements (NDAs) are often entered into with the aim of protecting sensitive information. These

are often mutual in nature, protecting both parties, however, they are usually entered into between organizations and not the individual testing the organization. Instead, there is a reliance on the NDA between the client and the organization they have engaged and the engaged organization and their employee. Although legally binding, confidentiality agreements have significant reliance on "good faith" compliance.

Background Screening

As a minimum, conducting a police check was considered standard. Organizations are now conducting credit checks as well as employment history checks to attempt to identify any risks. One issue is that when a security testing organization is engaged by a client, there is an assumption that those checks have already been completed by the organization that has been engaged and that there are no issues.

Rules of Engagement

Often initiated by the organization conducting the assessment, the Rules of Engagement (ROE) are the rules or directives that have been established which describe the circumstances and limitations for initiation and combat with enemy forces by a country's military (Roach, 1983). Similarly, such rules are often used by security testers when engaged by an organization to validate the effectiveness of their security controls. The Rules of Engagement identify what is and is not in scope, such as IP address information, specific systems, scheduling, and who should be notified in the event of an issue or question.

Ethical Hacking Framework

The Ethical Hacking Framework (EHF) (Thomas, 2020) is a framework designed to provide some structure and process for engaging ethical hackers. The framework consists of 16 controls across the following four stages:

- Prior to engagement;
- During engagement;
- After engagement; and
- Engagement review.

The framework was developed by leveraging controls from existing frameworks and standards, such as ISO/IEC 27001:2013, NIST Cybersecurity Framework, NIST Special Publication 800-53, and the ACSC Information Security Manual as well as additional research in the area, which identified some key areas of concern including confidentiality of information, professional standards, conflicts of interest, and onboarding processes (such as validating qualifications, appropriate confidentiality agreements, and background screening).

Although some organizations may already be conducting elements of the framework, it may be too complex or difficult for many organizations despite the benefits of using it. The introduction of regulation would likely address the key controls from the EHF and those benefits would be made available to everyone.

Regulation and Uniformed Codes

Medicine, law, accounting, and teaching are some examples of regulated professions. In many countries, practicing in these professions without meeting specific criteria in education, experience, and licensing is often illegal. At present, there exists no uniform or mandatory code of ethics for ICT professionals (Thomas *et al*., 2019), and attempts to create one have failed repeatedly over the past several decades (Burmeister, 2013). ICT in comparison to the professions mentioned earlier is considered relatively new. In fact, the ACS was only admitted to Professions Australia in 2000 (Ridge, n.d.; Weckert *et al*., 2013; Thomas, 2020). Cybersecurity and the roles within it are considered even more recent professions and like ICT there is no uniform or mandatory licensing requirement to date.

A Case for Regulation

Regulation and legislation have been discussed at length earlier in this chapter and there are several requirements that must be complied with in many jurisdictions and industries. There are still inconsistencies with laws and regulations, but with the vast adoption of technology, which over the past couple of years has increased significantly due to the COVID-19 pandemic that swept the world, the importance of technology and cybersecurity cannot be understated, and such laws and regulations are expected to increase. Like the NYDFS Cybersecurity Regulation, which mandates

annual penetration testing occur, it is anticipated that the prescriptive nature of such regulation will become more commonplace. Couple these requirements with significant financial penalties and even imprisonment for compliance failures and multiple problems come to light.

Protection of Information

Some laws and regulations that explicitly require safeguards are in place to protect the information held by those the laws and regulations apply to. Although, this requirement is not prescriptive in how or what to do, often organizations and subsequently security professionals will resort to the implementation of best practice approaches which include preventative, detective, and corrective controls to protect confidentiality, integrity, and availability of information. Although not an exhaustive list, examples of such controls include the following:

- Access controls to restrict who has the ability to access the systems and information;
- Encryption controls to encode the information so that it is unreadable to unauthorized persons when stored or in transmission;
- Physical controls to prevent unauthorized access to premises or information processing facilities;
- Auditing to log access attempts and investigate incidents;
- Awareness and education to ensure personnel are aware of security risks, how to prevent them, and what to do in the event of a security event; and
- Administrative controls, such as policies, procedures, and standards.

To validate the controls are effective, vulnerability management and penetration testing activities are the standard testing approaches. Such validation is often carried out by an ethical hacker or penetration tester and such a professional is considered a trusted party. Should an ethical hacker succeed at their objectives, they will (depending on the scope of the engagement) likely obtain access to sensitive information, whether it is personal information, intellectual property, or any other private information. This places an emphasis on ensuring that the information accessed is not misused and is handled appropriately and there are numerous examples of insider threats, such as insider trading and fraud, that have involved the misuse of information.

Competence

Just as there is an importance in handling sensitive information, competence is an important factor for a few reasons. Ensuring that a test is adequately thorough is critical; the ethical hacker needs to conduct a thorough enough test to uncover vulnerabilities that need to be addressed and validate the security posture. While being able to uncover every possible vulnerability scenario is unlikely, those that are high risk and especially those that are easy to exploit should be identified so they can be remediated. The skills and tools required to be an effective ethical hacker are often complex and need specialist skills and extensive experience.

Another issue is causing accidental or inadvertent destruction, corruption, or availability issues. Not possessing the appropriate skills could result in significant adverse effects. Destruction or corruption of data or taking a system offline (denial of service) could have catastrophic impact on the operations of the client.

Ethical and Professional Conduct

Although some certification and accreditation bodies expect their credential holders to abide by their respective codes of ethics and conduct, obtaining certification and subsequently becoming a member of such an organization is not mandatory.

This means that anyone could provide ethical hacking services whether they were skilled, experienced, or not. The onus is on the client to do their due diligence as well as the organization that the individual represents. Many clients will not be suitably equipped to perform such evaluations and even a framework, such as the Ethical Hacking Framework (EHF) (Thomas, 2020) used to assist in mitigating risk when engaging ethics hackers, may be too complex for many organizations.

Insurance

Another important consideration is insurance. Should an issue arise from the engagement of an ethical hacker, which results in loss or damage to the client, what insurance does the ethical hacker carry? Although it would be considered good business practice for the ethical hacker to obtain appropriate professional indemnity and public liability insurance,

this would not be mandated or verified (except if the client were to verify this themselves). Regulation of the profession would likely see this requirement mandated, which would help protect the client and the ethical hacker in the event of loss or damage due.

Adverse Situations

Finally, regulatory bodies help address potential gaps in accountability. For example, there is nothing preventing an unqualified, unfit, or incompetent ethical hacker from continuing to provide services. Such practices would likely not be in the public's best interest and could result in substantial risks, such as those that data protection and privacy regulations and legislations are designed to prevent.

Criticisms for Regulation

It is important to consider why regulation may not be suitable. Although there are many benefits to regulating, especially when it comes to risk management and compliance with regulation and legislation related to data protection and privacy, there are some potential side effects.

Compliance

Regulation will typically add significant burden in terms of compliance requirements. To ensure up-to-date knowledge and skills, there is a requirement for continuing professional development. In addition to compliance requirements, auditing can also occur which would be an added burden.

Cost

Membership in the governing body will likely result in a cost of membership as will the requirement of continuing professional development. Costs are usually financial, time-based, or both and can be significant. This would likely result in those higher costs being passed onto the client. This may also result in dissuading people from becoming ethical hackers simply because it is cost prohibitive.

Innovation

There is also concern that regulation could be too restrictive and subsequently slow down innovation in the area, which is generally considered a very fast-paced area (Thomas, 2020). Regulators need to have a good understanding of what ethical hacking involves and understand that change can occur daily and that there is a certain level of agility that is required in such a field. Where laws and medicine change at a much slower pace, vulnerabilities are discovered all the time and techniques are continually evolving, which is reflective of the nature of both technology and how recent the industry is.

Conclusion

When one visits a medical practitioner for a diagnosis or seeks legal advice from a legal professional, there is an expectation that the professional has a minimum level of education, experience, and insurance and maintains an appropriate license with a relevant governing body. This provides some level of comfort that the professional can provide their services appropriately. Additionally, should such a profession be unqualified, incompetent, or unfit, proper regulation would prevent such an individual from practicing and subsequently protect the public's interest.

Professions, such as medicine and law, can result in significant consequences when an adverse event occurs, which highlights the importance of ensuring professionals meet minimum standards. It is important that ethical hackers are adequately skilled and experienced to ensure that they are properly and adequately validating the information systems they are assessing. In many cases, such systems contain sensitive and valuable information that is subject to data protection and privacy regulation or could otherwise result in significant impact should that system be inadequately tested and subsequently exploited by a threat actor.

It is important to ensure a minimum level of education, experience, and continuing professional development. Regulation, such as the NYDFS Cybersecurity regulation, places an emphasis on such requirements, by requiring that qualified personnel are used. Where such regulation falls short, is that it does not detail what constitutes "qualified personnel" and it is largely up to the discretion of those vetting the personnel to make that decision. Subsequently, having some uniform method of identifying what it means to be qualified would be advantageous.

Like other professions, the importance of professionalism and ethical conduct cannot be understated. To help avoid ambiguity as this can mean different things to different individuals, the use of a code of ethics or conduct can be adopted. However, to achieve consistency and wide adoption, mandating such a code should be considered. There is the possibility of misuse of information for financial gain as seen in examples of insider trading. The skill set and level of implied trust afforded to an ethical hacker put them at an advantage, and ensuring such controls are in place helps reduce that risk.

To address these requirements, regulations would provide an effective means of achieving this, however, considerations around flexibility need to be considered. In addition, regulations would need to consider the added cost and burden. Although this could likely result in an increased cost to engage an ethical hacker to test an organization, many consultancies have already become members of organizations, such as CREST, and mandatory professional membership of such an organization could be a possible solution. Many security professionals (and their clients) recognize the value of certification and, although not mandatory, have already chosen to get certified and fulfill the compliance requirements for continuing professional development.

Although unregulated, ethical hackers are already undertaking many of the requirements of a regulated profession and regulation would help ensure that only qualified, insured, and competent persons conduct engagements, this would not only further the profession but would also likely reduce the overall risk to the tester and the client.

References

#SaveTheArctic — Phase I (2012). https://pastebin.com/1ca3BR19. Accessed on 15 September 2021.

Burmeister, O. (2013). Achieving the goal of a global computing code of ethics through an international-localisation hybrid. *Ethical Space*, 10, 25–32.

Data Privacy Philippines (n.d.). Data privacy FAQ. https://www.privacy.com.ph/learn-data-privacy-compliance/data-privacy-faqs/. Accessed on 15 September 2021.

Data Protection Act (2018). Data Protection act 2018. https://www.gov.uk/government/collections/data-protection-act-2018. Accessed on 3 June 2021.

Gartner (2021). Gartner forecasts worldwide security and risk management spending to exceed $150 billion in 2021. https://www.gartner.com/en/

newsroom/press-releases/2021-05-17-gartner-forecasts-worldwide-security-and-risk-managem. Accessed on 23 September 2021.

Giac Certifications (2021). GIAC code of ethics overview. https://www.giac.org/about/ethics. Accessed on 17 September 2021.

GOV.UK (n.d.). Data protection. https://www.gov.uk/data-protection. Accessed on 17 September 2021.

Graves, K. (2010). *CEH Certified Ethical Hacker Study Guide.* Indiana: Wiley Publishing, Inc.

New York State Department of Financial Services (2020). https://www.dfs.ny.gov/industry_guidance/cybersecurity. Accessed on 15 September 2021.

O'Brien, R. (2016). Privacy and security: The new European data protection regulation and it's data breach notification requirements. *Business Information Review,* 33, 81–84.

Office of the Australian Information Commissioner (2021). What is privacy? https://www.oaic.gov.au/privacy/your-privacy-rights/what-is-privacy/. Accessed on 17 August 2021.

Office of the Australian Information Commissioner (n.d.). The privacy act. https://www.oaic.gov.au/privacy/the-privacy-act/. Accessed on 12 September 2021.

Office of the Privacy Commissioner of Canada (2019). PIPEDA in brief. https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/pipeda_brief/. Accessed on 14 September 2021.

Personal Data Protection Commission Singapore (n.d.). PDPA overview. https://www.pdpc.gov.sg/Overview-of-PDPA/The-Legislation/Personal-Data-Protection-Act.

Roach, J. A. (1983). Rules of engagement. *Naval War College Review,* 36, 46–55.

Solove, D. J. (2006). A brief history of information privacy law. 4.

State of California Department of Justice (2021). California consumer privacy act (CCPA). https://oag.ca.gov/privacy/ccpa. Accessed on 10 August 2021.

Thomas, G. (2020). *Issues of Professionalism Concerning the Ethical Hacking of Law Firms.* Doctor of Information Technology, Charles Sturt University.

Thomas, G., Burmeister, O. and Low, G. (2019). The importance of ethical conduct by penetration testers in the age of breach disclosure laws. *Australasian Journal of Information Systems,* 23, 4–5.

Thomas, G., Low, G. and Burmeister, O. (2018). "Who was that masked man?": System penetrations — Friend or foe? In Prunckun, H. (Ed.), *Cyber Weaponry: Issues and Implicatgions of Digital Arms*. Berlin: Springer.

UNCTAD (2021). Data protection and privacy legislation worldwide. https://unctad.org/page/data-protection-and-privacy-legislation-worldwide. Accessed on 15 August 2021.

Voigt, P. and Von Dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR). A Practical Guide*. Berlin: Springer International Publishing.

Wark, M. (2006). Hackers. *Theory, Culture & Society*, 23, 320–322.

This page intentionally left blank

https://doi.org/10.1142/9781800612860_bmatter

Index

3Vs, 70

A
Abrams v. United States, 261
acceptability, 84
acceptable use policy, 226
accountability, 43–44
account holder confirmation, 216
Adams, Douglas, 210
adequate risk management, 13, 43
adhoc, 71
advanced persistent threats (APTs), 3, 45, 318
adware, 3
age appropriate design code, 212, 222, 229–230
age verification, 214–217
AlphaBay, 305–306
Al Qaeda, 187–188
Amazon, 170
Aminian, N., 296
Amnesty International, 168
Anti-Crime Law, 309
anti-vax, 219
Apple, 96, 98, 102
application programming interface (API), 74
apps, 67
Arab Spring, 179
ARPANET, 149
artificial intelligence, 216
Australia, 319–320
Australian Computer Society (ACS), 324
Australian Information Commissioners, 160
Avia Law, 271

B
background screening, 325
balance of opinion, 177–198
Barlow, J. P., 29, 206
Basel, 45
Basel II, 37
BATX, 28
Bazán, J. A. G., xxviii
Beauharnais v. Illinois, 262, 264
behavioral biometrics, 82
big data, 61, 70–72, 154
biometric enrollment, 90
biometric technology, 81–109
black hats, 317
Blake Robbins, 222
Bluetooth, 248–249

Bluetooth Low Energy (BLE), 249
Brandeis, Justice, 261
Brandenburg v. Ohio, 263
Brazil, 292
Brazilian Association of Technical Standards (ABNT), 293
Brazilian Civil Rights, 293
Brazilian Code of Criminal Procedure, 301
Brazilian Consumer Protection Law, 294
Brazilian Federal Trial Courts, 297
Brazilian Personal Data Protection Law, 294
British health system, 252
Budapest Convention, 25

C
California Consumer Privacy Act (CCPA), 72, 322
Cambridge Analytica, 181
Canada, 320
CCP, 192–193
Center for Strategic & International Studies (CSIS), 117
CERT, 32
Certified Ethical Hacker (CEH), 323
Certified Penetration Testing Professional (CPENT), 323
Chaplinsky v. New Hampshire, 262
Chevlin, D., 302
chief data officers (CDOs), 62
chief information security officers (CISOs), 62, 322
chief privacy officers (CPOs), 62
Children's Online Privacy Protection Act (COPPA), 211
China, 30, 191, 193
circumvention, 83
CIS 20 Critical Controls, 76
Ciuchi, C., xxviii
Civil Rights Framework for the Internet, 292
Clearview AI, 161–162, 165, 170
Clinical Trial Regulation (CTR), 241–242, 244
closed political systems, 190
closed regimes, 190–195
cloud, 66–70
cloud-based infrastructures, 69
cloud-based technologies, 68
cloud computing technologies, 61
cloud vendors, 69
Code of Practice, 165
codes of conduct/ethics, 259, 273, 323–324
Cold War, 263, 298
collectability, 84
Communications Decency Act (CDA), 206, 266–267
Communist Party, 195
community standards, 272
competence, 328
compliance, 75–77, 329
computer hacking, 292
computer viruses, 3
computer worms, 3
confidentiality, 13–14
confidentiality agreements, 324–325
Conseil supérieur de l'audiovisuel (CSA), 271
consumer-connected devices, 73
copyright management, 45
correlation, 181
cost, 329
cost-benefit analysis, 299, 301
COVID-19, 129–131, 134–135, 224, 245–251, 257, 298, 326
CREST, 323, 331
criminal investigations, 146–148, 170
criminal law, 299
critical information, 51

critical infrastructure, 3, 8, 24, 45–46, 194, 286
critical processes, 49
Cross-Site Request Forgery (CSRF), 5–6
Cross-Site Scripting (XSS), 5
CybelAngel, 253
cyberattack, 1–6, 235, 251–254, 308
cyberbullying, 46
cybercrime, 297–311
cybercrime dark web program, 160
Cyber Crimes Law, 292
cybercriminals, 67
Cyber Kill Chain model, 2
Cyber Policy Portal, 117
cyber risk, 38, 43–44, 47, 50–51, 283–312
cyber risk awareness program, 52–53
cyber risk management, 37–54
cybersecurity, xxviii, 38, 118–122, 330
Cybersecurity Act, 284–285
cybersecurity culture, 62–63
cybersecurity field, 1–16
cybersecurity legislation, xxix
cybersecurity package, 8
cybersecurity risk management, 6–8
cybersecurity strategy, 8
cyberspace, 20–22, 27, 29–30, 128, 131–132, 134–135, 302
Cyberspace Affairs Commissions, 192
cyberspace privacy, 123–127
cyberstalking, 46
cyber-surveillance, 195
cyberwarfare, 194
cyber world leaders, 131

D

Dark Web, xxx, 283–312
dashboards, 48
data access, 69
data discovery, 63–66
data ethics, 151
data gathering approach, 40
Data Governance Act, 274
data leak protection (DLP), 52
data minimization, 216
Data Privacy Act, 320–321
data protection, 61–78
Data Protection Act 2018, 225–226, 321
Data Protection and Cyber Liability Insurance, 296
Data Protection and Cyber Responsibility Insurance, 296
Data Protection Authority (DPA), 289
Data Protection Directive, 287–288
Data Protection Impact Assessment, 213
data protection officer (DPO), 287, 320
decision-making process, 6
Deep Web, 297–310
delivery, 2
democratic nation-states, 177–198
democratization effect, 179
(distributed) denial of service (DoS/DDoS) attack, 3–4, 71
Dieckmann, Carolina, 292
digital authoritarianism, 193
Digital Economy Act, 215
Digital Goods, 304
digital information age, 113–114
Digital Investigation and Intelligence, 171
Digital Markets Act, 274
Digital Package, 274
digital privacy, 111–136
Digital Services Act (DSA), 271
digital sovereignty, 183–185
digital space, 113
Doctolib, 253
Draft Online Safety Bill, 204

Dread Pirate Roberts (DPR), 303, 305
duty of care, 204–205, 219
Dyson, F., 146

E
eBay, 302
EC-Council, 323
E-Commerce Directive, 269–270, 274
economic crime, 299
E-Government Act, 159
El Dorado, 19–33
Electronic Frontier Foundation (EFF), 29
Electronic Privacy Information Center (EPIC), 159
email bombing, 5
email phishing, 5
email spamming, 5
email spoofing, 5
embedded journalism, 180
encryption, 239
enterprise risk management (ERM), 43
EPIC vs DHS, 159
Equal Error Rate (EER), 91
erosion of privacy, 145–172
erosion of rights, 217–229
Espionage Act, 261
ethical conduct, 328
ethical hackers, 315–331
Ethical Hacking Framework (EHF), 325–326, 328
EU Agency for Cybersecurity (ENISA), 14–15
EU Cybersecurity Act, 14–15
EU GDPR, 75
EU Member States, 267–268
Europe, 257–275
European approach, 267–269
European Commission, 8, 259, 272, 287
European Court of Human Rights, 260
European Court of Justice, 270
European Cybersecurity Strategy, 9–10
European Data Protection Board (EDPB), 160, 239, 246, 288
European Defense Agency (EDA), 26
European Economic Area (EEA), 286
European External Action Service (EEAS), 26
European framework, 8–9
European General Data Protection Act, 153
European Union (EU), 1, 286–287, 320
European Union Agency for Cybersecurity (ENISA), 3–4, 8–9, 14–15, 117
European Union Agency for Law Enforcement Training (CEPOL), 164
European Union law, 259
Europol, 165
Eva Glawischnig-Piesczek v. Facebook Ireland Limited, 270
excess data processing, 214–217
excessive data collection, 217–229
exploitation, 2

F
Facebook, 28, 148, 170, 182–183, 188, 195, 207–208, 272
face recognition, 95
Failure to Acquire (FTA), 92
Failure to Enrol (FTE), 92
false acceptance, 90–92
False Acceptance Rate (FAR), 90–91
False Match Rate (FMR), 90
false rejection, 90–92

Farley, M., 302
Fawkes, G., 316
Federal Communications Commission (FCC), 273
Federal Networking Council (FNC), 21
Federal Police, 307
Federal Workforce Assessment, 284
Ferraço, R., 293
financial crime, 299–300
fingerprint, 95
First Amendment, 260
Four Horsemen, 208
France, 271, 274
Franzblau, Kenneth, 302
freely given consent, 242
French Assemblée Nationale, 271
Furnell, S., xxviii

G
G7, 22
G20, 22
Garcia, J. Á. B., 299
General Data Protection Regulation (GDPR), 12–14, 72, 107, 170, 211, 237, 243, 245, 286, 289–291, 295, 320
geographical aspect, 136
geographical cohesion, 131–132
geo-location tracking, 46
GIAC Exploit Researcher, and Advanced Penetration Tester (GXPN), 324
GIAC Exploit Researcher, 324
GIAC Web Application Penetration Tester (GWAPT), 324
GIAN Penetration Tester (GPEN), 324
Gitlow v. New York, 262
Global Cyber Strategies Index, 117
Global Information Assurance Certification (GIAC), 324
Global Privacy Enforcement Network, 212
Google, 31, 67, 95, 248, 272
governance, 21
governance of cyberspace, 19–33
Government Technology Agency of Singapore, 248
gray hats, 317
Griffin, R., 298, 303
Group of Governmental Experts (GGE), 23–24

H
hacking, 322–323
hacktivists, 317–318
hard identifiers, 216
harmonization, 77
hate crimes, 259–260
hate speech, 257–275
health data, 235–254
Hippocratic Oath, 236
Holmes, Justice, 261
Hong Kong, 193–194
human negligence, 50
Human Rights Act, 155
Human Rights Committee, 268

I
IBM, 170
ICANN, 31
Impostor Pass Rate (IPR), 90
inappropriate behavior, 221
incitement, 269
Infocalypse, 208
Information and Communications Technology (ICT), 8–9, 190, 316, 324
information availability, 52
information integrity, 52
information privacy, 115
informed consent, 242
innovation, 330

insurance, 328–329
Internal Revenue Service, 307
International Council of Electronic Commerce Consultants, 323
International Court of Justice (ICJ), 195
International Covenant on Civil and Political Rights Article 20, 267
International Electrotechnical Commission (IEC), 171
International Information System Security Consortium, 324
International Law Commission, 269
international legal cooperation, 297, 311
International Military Tribunal, 267
International Organization for Standardization (ISO), 171
international organizations, 68
International Union of Communications (IUC), 28
Internet, 21, 24, 29–30, 112, 150, 170, 178, 190, 192, 266, 283, 302
Internet Corporation for Assigned Names and Numbers (ICANN), 28
Internet Covert Operations Program (iCOP), 159
Internet Governance Forum (IGF), 32
Internet of Things (IoT), 72–75, 148
Internet Ombudsman, 274
Internet Protocol (IP), 21
Internet Service Providers, 220
INTERPOL, 153
ISIS, 187–188
ISO 27001, 76
IT infrastructure, 47
IT service provider, 48–49

J
Jelinek, Andrea, 246
Joint Committee on Information Technology, 293

K
Kennedy, M. A., 302
key performance indicators (KPIs), 48
Ku Klux Klan, 263

L
law enforcement, 154, 158–168, 170
Law Enforcement Directive (LED), 153–154
Lazar, E., xxx
Le Conseil Constitutionnel, 271
legal protection, 298
legislative protection of speech, 266–267
Lewin, K. Z., 180
Licensed Penetration Tester (LPT), 323
LIHKG, 193
Loi Gayssot, 268
Loureiro, Rafael Mendes, 292–293
low-risk processing, 216

M
MacBook, 96
machine learning algorithms, 151
malware, 3
man-in-the-middle (MiTM) attacks, 6, 74
Matal v. Tam, 265
McKeown, M. M., xxx
media information, 28
media monitoring services (MMS), 159
medical QR Code, 248
Medjack, 252
Member State, 24, 248, 270, 288
Messenger, 208
Microsoft, 32
Mihai, Ioan-Cosmin, xxviii
Ministry of Health, 248–249
mitigation strategies, 324

money laundering, 283–312
Motion for a European Parliament, 25
multifactor authentication, 49

N
NameVine, 149
National Crime Agency, 171
national cyber policies, 111–136
national cyber policy documents, 129, 132–134
national cyberspaces, 21–27
National Data Protection Authority (ANPD), 295
National Institute of Standards and Technology Act (NIST), 6, 285–286
Nationalist Socialist Party of America v. Village of Skokie, 264
national law enforcement agencies, 153
National Police Chief's Council (NPCC), 154
national security, 185–195
National Socialist Party, 264
nation-states, 198, 318
Network and Information Systems (NIS), 8, 10–12
Network Enforcement Act, 270
New York Department of Financial Services (NYDFS), 321–322
New York Times Co. v. Sullivan, 263
NIS2, 10–12
NIST Cybersecurity Framework, 76, 324, 326
non-democratic nation-states, 177–198
non-disclosure agreements (NDAs), 324–325
Notifiable Data Breaches scheme (NDB scheme), 319
NYDFS Cybersecurity Regulation, 326–327

O
Offensive Security, 323–324
Offensive Security Certified Expert (OSCE), 323
Offensive Security Certified Professional (OSCP), 323
Office of the Australian Information Commissioner (OAIC), 318–319
Office of the Communications Regulator (OFCOM), 220
Oliver, Jamie, 227–228
one country two systems policy, 193
online content, 269–271
Online Harms Bill, 205, 223
online privacy, 114–116
Online Safety Bill, 206–208, 214
Open-Ended Working Group (OEWG), 23–24
Open-Source Intelligence (OSINT), 145–172
Operation Glowing Symphony, 188
Our Father (Pai Nosso) Operation, 306
Oversight Board, 272

P
Palhares, L. A. F., 292–293
Pan-European Privacy-Preserving Proximity Tracing, 248–249
parental controls, 217–229
Paris Call, 32
Parliamentary Assembly of the Council of Europe, 274
password-based approaches, 86
passwords, 85
patriotic retaliatory responses, 195
Pavel, T., xxix
performance, 84
permanence, 84
Personal Data Protection Act (PDPA), 283–296, 321

Personal Digital Assistant (PDA) devices, 94, 99
personal identification number (PIN), 85
Personal Information Protection and Electronic Documents Act (PIPEDA), 320
Petricã, Gabriel, xxviii
Philippines, 320–321
Phippen, Andy, xxix
physiological biometrics, 82
pluralism of information, 271
Poland, 268
Police and Law Enforcement, 67
politics, 179–185
privacy, 66–70, 235–254, 318–319
Privacy Act, 156, 319–320
privacy impact assessment (PIA), 154
privacy regime, 151–158
private corporations, 273
private organizations, 72
private regulation of hate speech, 271–273
Prodigy Services, 266
professional conduct, 328
Prosecutor v. Šešelj, 267
protection of information, 327
pseudonymization, 239, 290
public policy, 15–16
Public Security Bureaus, 192

R

ransomware, 3–4
Ranum's Law, 207, 214
R.A.V. v. City of St. Paul, 264
reconnaissance, 2
Red Scare, 262–263
reference profile, 89
regime stability, 177–198
regulation and uniformed codes, 326
Reider-Gordon, Mikhail, xxix
reliability, 104–105
Republic Act No. 10173, 320–321
responsibility, 43–44
risk analysis, 7
risk management, 6–8, 45
rogueware, 4
Royal Canadian Manted Police (RCMP), 149
Royal United Services Institute's (RUSI), 154
Rules of Engagement (ROE), 169, 325
RuNet, 30

S

SABAM v. Netlog NV, 270
safeguarding, 14
safeguarding fallacy, 203–230
safety, 226
SafetyTech, 222
Santos, Cláudia Cruz, 299
scareware, 4
Schenck v. United States, 261
Schmidt, E., 147
Second World War, 186
secure enclave, 102
Secure Information Exchange Network Application (SIENA), 164
security, 235–254
security mechanism, 308
self-declaration, 216
sensitive data, 238
Shefet, D., xxx, 274
Silk Road, 303–304, 306
Singapore, 321
smartphones, 66
Snyder v. Phelps, 265
social media, 177–198
SocialSpy, 166
SOCMINT, 150–151
Solvency, 45
South China Sea, 194–195
South Korea, 247
sovereignty, 184

SOX, 37, 45
spyware, 3
Standing Committee, 274
Streicher, Julius, 267
Structured Query Language injection (SQLi), 5
subscription databases, 149
Sutherland, Edwin, 299
Swedish Authority for Privacy Protection (IMY), 162
Swedish Police Authority (SPA), 162

T
Tactical Internet Operational Support (TIOS), 155
technical measures, 216
techno-authoritarianism, 193
Telegram, 193, 297
terror groups, 187
thematic aspect, 136
The Onion Router (TOR), 303–304
third-party age verification, 216
Thomas, G., xxxi
Tik Tok, 223
Togolese Government, 28
tokenization, 290
Touch ID, 95–96
TraceTogether, 248–249
Transmission Control Protocol/Internet Protocol (TCP/IP), 21
trojans, 3
Trump, Donald, 182
Twitter, 183, 188
two-factor solutions, 85
type II error, 90–91

U
Ulbricht, Ross, 303, 306
UN Convention, 212
UN Convention on the Prevention and Punishment of the Crime of Genocide, 269
UN CRC, 219, 228
unintentional data leakage, 65
uniqueness, 84
United Kingdom (UK), 205, 321
United Nations (UN), 23, 25, 258
United Nations Group of Governmental Experts (UN GGE), 22–23
United States (US), 28, 257–275, 283, 321–322
United States Code, 156
United States Cyber Command, 188–189
universal availability, 103–104
universality, 84
UNODC, 167
untrusted software programs, 71
US approach, 261–267
US E-Government Act of 2002, 156
user authentication, 87
user compromise, 105–106
user experience, 88, 104–105
user identity, 81–109
user trust, 103
US Telecommunications Act, 206

V
verification processes, 90
Vietnam, 195
Virginia v. Black, 264
Vogel, Timothy A., 301–302
voice verification, 95

W
WannaCry, 252
weaponization, 2
WeChat, 192
Weibo, 192
WhatsApp, 183, 297
white hats, 316–317
Whitney v. California, 262

why is this important (WITI) test, 40–43
Wisconsin v. Mitchell, 260
World War I, 262–263
World Wide Web, 292, 302

Y
YouTube, 223, 272

Z
Zuckerberg, M., 205