Volume 2 Number 4 2019 ISSN 2398-1679

Journal of Data Protection & Privacy

Henry Stewart Publications
Ruskin House, 40-41 Museum Street,
London, WC1A 1LT, UK
Tel: +44 (0)20 404 3040
Website: www.henrystewartpublications.com

Henry Stewart Publications
North American Business Office
PO Box 361
Birmingham, AL 35201-0361, USA
Tel: 800 633 4931; Fax: 205 995 1588
e-mail: hsp@subscriptionoffice.com

ISSN 2398-1679

Printed in Great Britain by Latimer Trend and Company Ltd, Plymouth, UK

Aims and Scope

Journal of Data Protection and Privacy is the major professional and academic journal that publishes in-depth, peer-reviewed articles and research on all aspects of data protection and privacy practice in the wake of the new EU General Data Protection Regulation (GDPR).

The journal is guided by its Editor and an Editorial Board consisting of recognised experts in the field of data protection and privacy law and enforcement, board management, cyber security, technology, data processing, cloud computing, executive education and executive recruitment.

Each quarterly 100-page issue – published both in print and online – provides an international forum for detailed, practical and thought-provoking articles from leading professionals and academics on a wide range of regulatory, compliance, risk management, board governance issues. The journal explores innovative strategies, tools and techniques and emerging technology trends that impact on the business continuity of all private, public sector/ Government and charitable/NGOs and professional bodies in the wake of the biggest changes in data protection and privacy for over two decades.

Scope

Articles address key topics including:

- Current thinking on protection of business continuity over the GDPR transition period
- Managing customer data in accordance with the GDPR
- The data revolution and its implications for public and private sectors
- Privacy protection in the age of digital disruption
- Embedding the Internet of Things (IoT) in your enterprise and organisation
- Mobile technologies and cloud computing
- The new breed of Data Protection Officer (DPO) as 'Compliance Orchestrator'
- Powers of the European Data Protection Board
- Powers of the European Data Protection Supervisor
- Powers of supervisory authorities and regulators across the EU

Aims

Journal of Data Protection and Privacy provides a peer reviewed forum for the publication of briefings, discussion, applied research, case studies, expert comment and analysis on the key legal, regulatory and technological issues impacting data protection and privacy of all organisations impacted by the GDPR. In so doing, ***Journal of Data Protection and Privacy*** seeks to:

- Expand and disseminate the body of knowledge in data protection and privacy
- Facilitate the sharing of good practice in data protection and privacy in alignment with the GDPR across all private, public, government and voluntary sectors
- Publish new and original ideas on research, policy and management on data protection and privacy
- Facilitate cooperation and exchange of ideas between practitioners and academics in the field

SUBMISSIONS AND SUBSCRIPTIONS

The Publisher and Editorial Board welcome the submission of detailed articles, papers and reviews for publication. All articles and papers submitted will be peer-reviewed. All contributions should be submitted by e-mail in Microsoft Word. Details of the author's affiliation should be given. The correct citation for this issue is (2019) 2 JDPP 4. Submissions should be sent to Ardi Kolah, Editor, Henry Stewart Publications, Ruskin House, 40-41 Museum Street, London, WC1A 1LT, UK, Tel: +44(0)20 404 3040; e-mail: ardi@godpo.eu.

Subscriptions: The subscription for the current volume, Volume 2, comprising 4 issues, is £210 in the UK and Europe, US$295 in the USA and Canada, and £225 in the rest of the world. The price includes postage and packing. Rates and discounts for multiple subscriptions and multi-user online licences are available on request.

Subscription enquiries should be addressed to:

UK and Europe: Henry Stewart Publications, Ruskin House, 40-41 Museum Street, London, WC1A 1LT, UK; Tel: +44 (0)20 7092 3469; Fax: +44 (0)20 7404 2081; e-mail: gweny@henrystewart.co.uk

From North America: HSP Subscriptions, PO Box 361, Birmingham, AL 35201-0361, USA; Telephone: +1 800 633 4931; Fax: +1 205 995 1588; e-mail: hsp@subscriptionoffice.com

From Rest of the World: Gwen Yates, Henry Stewart Publications, 40-41 Museum Street, London, WC1A 1LT, UK; Telephone: +44 (0)207 092 3469; Fax: +44 (0)207 404 2081; e-mail: gweny@henrystewart.co.uk

Contents

Editorial

In loco parentis of the GDPR and implementing a GDPR/CCPA maturity framework

To coincide with the first anniversary of the full enforceability of the General Data Protection Regulation (GDPR), Věra Jourová, Commissioner for Justice, Consumers and Gender Equality, gave a speech in which she concluded that the GDPR was still in its infancy.[1]

On reading a transcript of her speech, I agreed with the assessment that the European Commission is *in loco parentis* for the GDPR.

It also follows that the benefits of a culture of responsible personal data use and an appropriate level of data security can only emerge with a strong commitment from European governments to provide the resources necessary for supervisory authorities to do their job.

With the changes that are happening in an increasingly politically destabilised Europe, whether the appetite of national governments to provide adequate financial support may be under threat in terms of national priorities.

But now is not the time to be defeatist or wave the white flag. Instead, it is now time to encourage EU citizens to optimise their privacy settings, argues Věra Jourová.

'NGOs active in the field of data protection have started making use of the possibility to bring representative actions before data protection authorities and courts',[2] she says.

Of course, such a statement could send a cold shiver down the spine of those companies and organisations that lack a credible GDPR maturity framework in order to comply with the higher standards of accountability, transparency and control.

> Data protection is not just a European matter. Privacy is an increasingly global issue. And we should stop to think of it as a domestic or regional one. In a world where social networks produces massive volumes of user-generated data, where cloud computing and artificial intelligence base their services on data flowing freely across countries, the intrinsic importance of personal data has never before been so clear.
>
> Europe and other countries around the world want to seize the incredible opportunities that the digital transformation of our economies and societies offer. And in doing so, we face similar challenges.
>
> Now we see new legislation adopted and hear calls around the globe for comprehensive data protection rules similar to the GDPR — from Chile to Japan, from Brazil to India, from Argentina to Indonesia, and from Tunisia to Kenya.
>
> Countries around the world are applying rules with very similar features: an overarching privacy law, with a core set of safeguards and rights, and enforced by an independent supervisory authority.
>
> And at multilateral level, the Council of Europe's Convention 108[3] — the only binding international agreement on data protection — is increasingly becoming a universal instrument. It shows that more and more countries are recognising the importance of protecting privacy, for individuals, and for society as a whole.[4]

A good example of this global move is illustrated by the California Consumer

Privacy Act (CCPA) 2018,[5] which will be enforceable from 1st July, 2020.

So, where should organisations and companies be now and what do they need to do next? I recently discussed this with Steve Wright, the founder of Privacy Culture,[6] former data protection officer (DPO) of the Bank of England and who has taught with me at Henley Business School.

According to Steve Wright, there are three phases that are part of the GDPR/CCPA journey.

Phase one is to have achieved a defensible compliance position. This involves implementing a new data governance and operating model; implementing a new data subject rights and consent framework; and implementing data deletion and security measures for high/very high-risk processing of personal data.

Phase two is to have implemented GDPR/CCPA measures to mitigate residual risks. This phase involves implementing data deletion and security measures for medium to low risk areas; an improved data governance and unstructured data discovery; improved third party management due diligence and risk management; and the implementation of improved security measures.

And phase three is about building a strategic GDPR/CCPA differentiation. The focus is on increasing customer trust by improving privacy controls and culture; potentially helping to reduce the cost of data operations through efficiencies in processing; reducing third party suppliers' risk exposure through stronger procurement processes; and ultimately leveraging personal data as strategic differentiator within the peer group for that market segment.

The type of convergence that both Věra Jourová and Steve Wright describe, based on robust laws and strong enforcement, as well as seizing the opportunity to turn data protection and privacy into a competitive advantage, will unlock the opportunity to do more, not less, with personal data.

Ardi Kolah LL.M
Founding Editor-in-Chief
Journal of Data Protection & Privacy
June 2019

References

1. European Commission (2019) 'Commissioner Jourová's intervention at the event "The General Data Protection Regulation one year on: Taking stock in the EU and beyond"', Brussels, 13th June, available at: https://ec.europa.eu/commission/presscorner/detail/en/speech_19_2999 (accessed 16th June, 2019).
2. *Ibid*.
3. Council of Europe (1981) 'Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data', Strasbourg, 28th January, available at: https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/108 (accessed 16th June, 2019).
4. European Commission, see ref. 1 above.
5. https://www.caprivacy.org/ (accessed 16th June, 2019).
6. https://www.privacyculture.com/ (accessed 16th June, 2019).

Papers

Information veracity towards a secure information posture

Received: 20th August, 2018

Clive Brindley

holds a master's degree in strategic IT management, as well as numerous industry certifications across security, governance and service management domains. A technology professional with over 25 years' experience, he has delivered IT transformation and management solutions across a wide spectrum of business segments including defence and financial services. As technology head for a financial services provider, Clive was responsible for the development, implementation and monitoring of numerous hybrid IT capabilities (including various cloud solutions). In this role, he was exposed to the real-world challenges of aligning business objectives with technology goals, improving the organisation's overall risk posture all while managing finite resources. This has allowed him to consider various practical approaches to improving information and data security, including continual compliance with various industry and country compliance requirements (Payment Card Industry Data Security Standard (PCI DSS), General Data Protection Regulation (GDPR) etc.).

Postnet Suite 556, Private Bag X153, Bryanston 2021, South Africa
E-mail: clive.brindley@accenture.com

Ben Silverstone

is a world leading research and commentator on the use of e-mail in organisations and the social issues associated with cyber security and IT. Dr Silverstone has a master's degree in management and a PhD in engineering, as well as fellowships to the World Business Institute and the Royal Society of Arts. Dr Silverstone is currently a senior teaching fellow at WMG, University of Warwick, and specialises in the design and delivery of degree apprenticeship programmes.

E-mail: Benjamin.Silverstone@warwick.ac.uk

Abstract The aim of this paper is to explore the various facets of information veracity, with the goal of unravelling the multiple permutations, methods and approaches for organisations striving to achieve their target level of compliance. Multiple sources of academic papers, commercial frameworks and related industry good practice are analysed to determine if common themes are exhibited. Through this research, four areas are consistently discussed. These areas are information and data regulation, information risk management, information and data governance, and finally information security standards and frameworks. Each of these four themes is then presented, covering the primary objectives related to information veracity. The importance of organisations having full knowledge of data regulations and laws, utilising enterprise-wide organisational knowledge to further strengthen their compliance posture, is highlighted. Information risks management requires the collaboration of numerous stakeholder groups, both business and technology, to ensure an appropriate risks posture is achieved. The role of an integrated organisational, technology and information governance operating model is emphasised as a key enabler to information veracity. Finally, the selection of appropriate, fit for purpose information security standards, frameworks and controls is discussed, with the key premise that re-use must prevail over in-house developed methods.

KEYWORDS: data protection, data regulation, security, governance

INTRODUCTION AND APPROACH

Veracity can be defined as conforming to facts, accuracy or truthfulness.[1] Ensuring information veracity is vital for business success.[2] The volume of data, information and knowledge instantiated by business processes grows each day.[3] At the same time, information owners grapple with the myriad of information and data security standards, regulations and compliance requirements.[4] Furthermore, information technology service models, be it cloud or hybrid IT, are blurring traditional information boundaries and making information veracity complex to attain.[5] The resulting change in the perception of control is forcing organisations to reassess their information security strategies all the way from the boardroom to the data centre.[6] Organisations face the reality of deploying a plenitude of information controls for the sake of adhering to a standard or applicable framework, without critically evaluating the underlying imperative. Expressed in other terms, the risk exposure to the business and related stakeholders if information is compromised needs to be carefully considered.

Developing an appropriate response to protecting information assets requires practitioners to navigate a multitude of information lifecycle management methods.[7] Selecting that which is appropriate, based on the organisations' risk appetite, operating environment and regulatory landscape, is complex and multifaceted. This paper seeks to cut through the various tenants of information veracity by conducting a literature review of industry good practice, commercial frameworks and academic research. The analysis will then be used to form the basis of any suggestions or hypothesis developed. The essential research topics of data protection and privacy, data regulations, information security, risk management, cyber security and cloud computing will be covered in this paper. Emergent from this analysis, key themes are identified for information owners, accountable executives and security practitioners to carefully consider.

These themes are listed below:

- Information and data regulations.
- Information risk management.
- Information and data governance.
- Information security frameworks and standards.

DISCUSSION

The discussion will consider work across these four themes and attempt to provide a golden thread that coherently agglutinates concepts and offer pragmatic approaches to ensuring information veracity. For the purposes of this research proposal, information veracity will focus on the securing of data and information and thereby ensuring it veracity or accuracy (void from manipulation). Effectively, this builds on the confidentiality, integrity and availability (CIA) tenants of most information security practices. Topics of data quality, lineage and validating the 'truth' based on business process execution are out of scope. Further refining the scope, the concept of cyber security is woven within the fabric of information security and is interpreted in its broader sense. This has been a deliberate action, as it is not the purpose of this paper to discuss the nuances between the two. Collectively, the objective of information security is to protect information assets across the multitude of business and technological architectures, of which cyberspace is but one. Of relevance, is the view that cyber security goes beyond the protection of information assets and covers the protection of those that function in cyberspace.[8] This '*disambiguation*' of cyber security and information security does however highlight the need to ensure a holistic, enterprise-wide view of information security and its impact on organisations, employees, customers and communities alike.

INFORMATION AND DATA REGULATIONS

Organisations strive to deliver shareholder value, be it within the private or public sector. Using trade secrets, intellectual property and other assets, organisations (including non-profit, governments, etc.) endeavour to bring differentiated offerings, product and capabilities to consumers, citizens and related stakeholders. These assets, in most cases, are digitised in various forms of technological constructs and need protection as with any physical asset. The board of directors have a duty to ensure the protection of information assets and that appropriate information technology governance structures and operating models are implemented.[9] Notwithstanding the imperative for information security practices to protect critical information assets, organisations are bound by various regulations in respect to protecting specific types of data. A cogent examination of the critical requirement of having an integrated information governance strategy is presented by Sloan.[10] Here, the information legal requirements are highlighted across laws affecting records retention, electronic recordkeeping, privacy, data security, intellectual property and litigation preservation. While predominantly focused on US law and regulation, sagacious insights are offered in terms of how information regulation can have an impact on organisations.

Nieuwesteeg[11] presents six key characteristics of data protection laws (DPL) across 71 countries. These characteristics are data collection requirements, data breach notification requirements, data protection authority, data protection officer (DPO), monetary sanctions and criminal sanctions. Organisations that traverse geographical boundaries therefore have a resulting complex legal, regulatory and compliance universe to navigate. In his paper, Nieuwesteeg explains that coding of the characteristics is primarily from the perspective of privacy control, which aims to give consumers control over their own data. In effect, DPL evolution can be attributed to market failures in information security and privacy. Furthermore, it is suggested that there is insufficient incentive for organisations to develop sociably acceptable situations in respect to privacy. Nieuwesteeg further acknowledges the existence of other characteristics, for example security guidelines, and welcomes further research into these areas. While ensuring the protection of personal information stored and processed by any organisation is indeed onerous, the six key characteristics defined earlier offers perspectives to accountable executives on what lawmakers look for in terms of data protection and privacy.

New and updated information and data protection regulations are forcing organisations to re-evaluate their information posture and critically assess their data assets. The European Union (EU) General Data Protection Regulation (GDPR) seeks to strengthen the rights of individuals in respect to personal data and special personal data that is stored and processed by organisations and governments alike.[12] Of interest, is the correlation of GDPR to the six principles presented by Nieuwesteeg. Organisations need to consider the business imperative for storing personal data, and further critique their business processes to determine '*what, why, how and where*' personal information is stored, processed and discarded. Furthermore, the data value chain will require organisations to carefully follow the 'breadcrumbs' where personal data, for which they are accountable, is processed by third party providers.[13] Organisations that have developed an information and data governance approach should be familiar with the concepts discussed and intrinsically well positioned to respond to an ever-changing data regulatory landscape.

It must be highlighted that the adherence to data protection regulations, as with the GDPR for example, should not be seen

as only a high cost and low value activity. The ethical usage of customer, employee and partner data is to be commended, and presents organisations with an opportunity to foster a transparent, open and trust-based relationship with their stakeholders. Additionally, the ability for organisations to critically assess their data processing could allow them to exploit additional business benefits through their compliance efforts.[14]

Evolving the discussion beyond pure personal data protection, there exists a plethora of industry data security standards that organisations need to adopt depending on the nature of their business. The Payment Card Industry Data Security Standard (PCI DSS) is one example of a detailed and very specific definition of minimum security standards to ensure cardholder data is protected throughout the payment value chain. While specifically developed for the credit and debit card industry, PCI DSS can be used to secure all forms of data in scope of data protection laws.[15]

In addition to data protection, it is important to recognise that various industry regulators and examiner bodies require specific application of internal controls that affect information processing. For example, the Sarbanes-Oxley Act (Section 404 specifically) defines that procedures should exist to prevent unauthorised access to data, thereby preventing its improper manipulation or deletion.[16] A co-ordinated response that utilises legal, compliance, risk, business and technology stakeholder expertise is required to ensure compliance within a changeable regulatory landscape.

This section has exposed the need for organisations to critically access their business processes and determine the '*what, why, how and where*' of information processing. The regulatory landscape has been shown to be complex, requiring senior accountable executives execute their duty to ensure appropriate protection of information assets.

INFORMATION RISK MANAGEMENT

Organisations face numerous risks in respect to information assets. Intellectual property and trade secrets, data leakage, industry compliance and adhering to regulatory requirements are just a few major areas to address. Johnson et al.[17] offer insight into the required evolution of information risk within organisations, suggesting a risk mindset is baked into every business area, function and process, and that a risk culture purveys throughout business. On a similar vein, Wheeler[18] suggests business outcomes can be improved by addressing risks within the context of value, desired business outcomes and an organisation's risk appetite. Traditional diametrically opposed risk takers (eg CEOs) and risk evaders (eg security professionals) need to pivot across their comfort areas and understand each other's world. A consensus developed risk posture is to be developed to ensure business exploits strategic opportunities while at the same time ensuring residual risk is within risk tolerance levels.

Information risk is commonly referenced as part of operational risk in enterprise risk management (ERM) frameworks. ERM[19] effectively manages all risks (market, credit, operational, strategic, reputational, etc.), which are viewed together in a co-ordinated and strategic framework. Lundqvist[20] provides a review of the most prevalent ERM frameworks and additionally highlights the challenges of reaching consensus on the core tenants of such a framework. Information risk management asks similar questions; that is, what are the critical elements for organisations to consider when developing a measured response to information risk?

Yang et al.[21] provide a detailed overview of leading information risk management methodologies, tools and practices. Here, they acknowledge the central role that risk analysis and business impact analysis (BIA) performs within every information security management system. A unified

model is offered in an attempt to guide stakeholders to effectively select and implement appropriate security controls. Lalev[22] similarly identifies BIA as a reliable and pragmatic method for identifying which information assets might need additional protection. Albeit with drawbacks (as postulated by Lalev), the BIA forces organisations to carefully consider their most critical business capabilities and supporting resources (people, technology, services, etc.). Of potential value would be the linking of BIA methods with data protection impact assessments (DPIA), as stipulated by the GDPR, for example.[23] It could be argued that an opportunity exists to review the critical nature of processing sensitive customer, employee or related stakeholder data when evaluating core business processes as part of a BIA. To support this reasoning, ISO 22313 (Guidance for Business Continuity Management Standard) suggests that 'when assessing impacts, the organisation should address those relating to its business aims and objectives and its interested parties'.[24]

Reviewing further synergies to the BIA risk assessment method, and the required treatment of residual risk by formulating a business continuity plan (BCP), is the acknowledgment that a BCP and security incident response plan (often described as a cyber incident response plan) should be tightly coupled.[25] Here, it is postulated that using standard risk assessment methodologies will aid in the unification of all risks and ensure information security residual risk is in the cross hairs of the appropriate stakeholders. Importantly, it must be recognised that the security incident detection, response and follow-up activities require a co-ordinated approach across stakeholder groups, both internal and external. Furthermore, the potential for discrepancies across security incident reporting will require careful alignment across organisational habits, processes and policy implementation.[26]

It is of interest to observe the evolution of approaches to information risk management over the years, as evidenced by the change to the prominent information security standard ISO 27001. The latest incarnation ISO 27001:2013 provides for two options when conducting a risks analysis. An asset-based or scenario-based risk methodology allows information security and risk practitioners to tailor their approach based on specific business needs. This potentially acknowledges that information risk management needs to focus more on the outcomes of the process and less on the method itself.

This section has highlighted the importance of entrenching a risk management culture across organisational constructs. Ultimately, irrespective of the methodology used, risk management needs to answer the fundamental question: are risks identified, quantified and responded to appropriately?

INFORMATION AND DATA GOVERNANCE

Information governance can be described as maximising the business benefit of information assets, while satisfying both internal and external compliance requirements, all within accepted organisationally defined risk tolerance levels.[27] While this is a useful summation, the gargantuan effort of co-ordinating activities across the 'village' of constituents is not to be underestimated. Many organisations recognise the innate value of information assets, developing solutions to exploit 'big data' as they further differentiate and grow their business. Coupled with this, large amounts of information risk and compliance requirements compel organisations to reassess information governance to ensure a holistic, business driven and enterprising response.[28]

Organisations thus need to assess their information related practices,

requirements, risks and opportunities, resulting in a defined set of objectives for information governance.[29] It then follows that organisations should implement an information governance programme to meet these defined objectives. This can be achieved by developing frameworks and controls for information (the structure), establish policies, procedures and contractual arrangements, and provide guidance and training (collectively, the direction). Furthermore, roles and responsibilities need to be defined and technology tools and systems need to be provisioned (collectively, the resources). Finally, the measurement of outcomes and development of appropriate consequences for success or failure in meeting aforementioned expectations and objectives is obligatory (accountability). These building blocks, namely assessment, structure, direction, resources and accountability, are crucial for defining and establishing an information governance programme.

Bennett[30] suggests an overall information governance framework will ensure the organisations information and data strategy is aligned, thereby supporting the attainment of business objectives. Here, an information governance framework is presented comprising of cyber security, privacy and data protection, records and information management, data governance, eDiscovery, data analytics, and risk and compliance domains. The nuanced differences and yet intrinsic integration of information and data governance, forces practitioners to ensure a holistic enterprise-wide view when developing solutions to govern effectively and manage the full quantum of enterprise information and data resources.

Data governance requires a holistic understanding of the organisation's processes and associated data,[31] integrating business and technology domains throughout the data lifecycle. Egelstaff and Wells[32] describe the evolution of data governance, and highlight the importance of ensuring the governance and management of data be pervasive across the organisation, driven by business strategy and embodied in a data governance vision. Cohn[33] presents five core elements that exhibit in the most effective data governance frameworks. The elements of leadership, adaptability, structure, standard and objectives are described. The adaptability core element highlights the need for organisations to be agile and nimble as they respond to changing external and internal environments. It can be argued that organisations that have a well-defined and implemented information and data governance culture are better prepared to respond to regulatory changes such as the GDPR mentioned earlier.[34]

Within data governance, the data access domain[35] specifically looks at safeguarding the organisation's data assets. A rigorous process of assigning value to different types of data, performing a risk assessment and identifying the required controls to protect the confidentiality, integrity and availability of data is imperative. Many of these are indeed technical controls, from edge perimeter defences through to logical access control. Raether,[36] however, highlights the point that the over-reliance on technology to protect critical data is a foolish endeavour. A holistic data governance plan is required to co-ordinate the lifecycle management of corporate assets including intellectual property and trade secrets, notwithstanding the requirements for the protection of personal information.

Developing this further , information and data governance are often seen as subsets of overall information technology and corporate governance. Tallon et al.[37] suggest that by embedding structural practices and responsibilities such as data stewardship within existing IT governance archetypes, all aspects of information governance could be addressed. This means having an integrated governance architecture that builds on the traditional 'brick and mortar' IT governance requirements of technology

investments, with the governance of often intangible and virtual information assets. Hagmann[38] offers a divergent view and suggests that IT governance itself has limited value in terms of information governance, since is prioritises architecture, application and systems management over the need for best practices in information lifecycle management. The 'orchestration' of various role players across business and technology is shown to be an important part in any information governance strategy. The development of an integrated, sustainable solution to the management of information is needed to support organisations' business objectives while maintaining risk levels within tolerance thresholds.

Goosen and Rudman[39] offer an integrated information technology governance framework that attempts to combine the best of leading industry governance models, standards and frameworks. While worthwhile in how it attempts to meld together the various control areas, the business imperatives presented are too generic and it can be easily argued that they apply to all organisations irrespective of size, complexity or industry. The framework fails to sharpen the focus for accountable executives and thereby define the most critical control areas and pinpoint where immediate management intervention is required. Furthermore, Goosen and Rudman define an information management system control area that aims to ensure the integrity, accuracy, confidentiality, availability and authenticity of organisational data. While implementation of this control area is important, the prioritisation and risk-based approach to safeguarding the most sensitive and valuable data requires further exploration.

Information Systems Audit and Control Association (ISACA) attempts to bridge the gap between information governance and IT governance by providing specific guidance on how information is to be governed in the construct of a broader information technology governance model.[40] The COBIT 5 goals cascade attempts to link enterprise's information goals with a defined set of IT-enabled processes, with clearly defined roles and responsibilities to effectively govern the lifecycle management of information. Despite the pre-eminence of COBIT 5, it is prudent to critically access the deployment of any framework and validate the benefits throughout the lifecycle of its use.[41] A knee-jerk reaction to ensuring some form of information governance is attained, could lead to significant wasted effort, cost increases (technological, contractual, human resources, etc.) and the potential to affect negatively future endeavours in information governance practices.

This section has underlined importance of having an integrated approach across information governance and broader IT and corporate governance. A common theme discussed is the requirement to ensure information governance is fully aligned to business objectives. If an information governance programme is to succeed, it requires the support of the highest decision-making bodies within the organisation.[42]

INFORMATION SECURITY FRAMEWORKS AND STANDARDS

Information security standards and frameworks are designed to improve the risk posture of organisations exploiting digital assets within their business. While developed with honourable intentions, 'checklist' standards and their adoption often leave accountable executives with a false sense of security. Acknowledging this reality, a cornucopia of information security standards and frameworks prevail for organisations to use when responding to information risks. ISACA offers a review of the most prevalent information security standards that feature controls aimed at reducing an organisation's information and data risk posture.[43] Composition of an integrated controls

framework requires the security practitioner carefully to avoid duplicating controls, create confusion by using varied terminology, and overly burden stakeholders with ineffective and bureaucratic security solutions.

Haufe et al.[44] provide a prospective information security management system core set of processes that combine ISO 27001, COBIT and Information Technology Infrastructure Library (ITIL) common processes. The work deviates from focusing on the control objectives mapping of previous and similar research, postulating that their approach is tailored to support the information security practitioner by allowing them to focus on the execution of the Information Security Management System (ISMS) processes and not grapple with identifying the myriad of controls and measures to implement. Of interest, is the appreciation that not all processes have to be implemented to full maturity, acknowledging that the limited resources available to manage an ISMS must be deployed appropriately. Furthermore, it is suggested that further research be conducted on determining the target maturity level *to be* attained, versus the determination of the *actually* attained maturity level. This is an important nuance to observe, and it can be argued that a link exists to risk management practices where the scale of the control deployed is proportional to the potential risk exposure.

It is to be conceded that cloud computing requires additional consideration in terms of information security. Many of the traditional controls, for example physical security, are relinquished by information owners and significant trust is endowed to the cloud service provider (CSP). Le and Hoang[45] offer a capability maturity model and metrics framework for cyber cloud security, where an attempt is made to consolidate numerous traditional and cloud security standards to help ascertain the security status of the organisations infrastructure and overall information risk position. Ardagna et al.[46] provide a useful cloud security and assurance taxonomy. The taxonomy covers three key areas as contributed by prevailing literature. These areas are vulnerabilities, threats and attacks; cloud security; and cloud assurance. It is recognised that the fast pace of new service offerings being introduced by CSPs and the multiple permutations of cloud architecture constructs, makes the job of security stakeholders daunting, to say the least. Of specific interest is the recognition that CSPs need to demonstrate transparency in all their internal processes, especially when they have an impact on security. This is embodied in the concept called 'outrospection' and similarly confronts the paradigm of 'security through obscurity'.

Rizvi et al.[47] offer a framework to validate the controls of a CSP, utilising the extensive work undertaken by the Cloud Security Alliance (CSA) to develop a unified controls matrix[48] to help customers and CSPs alike, identify and deploy appropriate controls to strengthen the security posture of affected organisations. Heiser[49] presents similar recommendations in evaluating CSP security, where it is advocated that organisations must guard against recreating security structures and instead make use of the various control frameworks in existence today. A number of options are presented in respect to formally assessing CSP security, further suggesting that more rigour is applied when assessing highly used cloud environments. Notwithstanding this pragmatic response, it must be noted that the risk assessment continues to present a foundational cornerstone to organisations seeking to manage information risk. It would be wrong to dismiss a low tier, less significantly used CSP and its security posture, if indeed the organisation is using it to store high value and high risk information. Utilising risk management practices will aid in the identification and treatment of information risk, be it using in-house, cloud or hybrid service deployment models.

It is important to note that the concept of evaluating service providers and ensuring their alignment with business needs has been prevalent for many years. As with any outsourcing arrangement or external service provider engagement, it is crucial that all aspects of information security are considered as part of a comprehensive information security approach. International standards such as ISO 20000[50] and ITIL[51] (sets of best practice IT service management processes), specifically address supplier management as part of a comprehensive approach to service delivery.

An important consideration presented in this section concerns the evaluation of CSP security capabilities and practices, coupled with an overarching supplier and vendor management approach. Selection of appropriate security controls should utilise established, mainstream standards, tools and methods, thereby reducing possible workload and re-invention of safeguards.

CONCLUSION

Traversing the numerous constructs of the information veracity quantum, it is clear that business and technology leaders need a comprehensive, integrated and fit-for-purpose solution to exploit the value locked in their data, while ensuring full compliance with regulatory and compliance expectations. While a deluge of techniques, frameworks, tools and approaches have been discussed, a common thread is appearing that might just be the northern star organisations need to navigate the data universe.

First, organisations need to evaluate their data posture by critically assessing the mechanics of their business. To answer the question of where customer personal data is stored for example, organisations need to unravel their business processes and 'walk the floor' through their business. No amount of technology can answer this question; organisations need to acknowledge that to know their data, they must know their business. Without doubt, this is a herculean endeavour for large geographically dispersed organisations; however, being big and complex is no excuse. Opacity of business architectures and their impact on data needs to change, and expediently so.

Knowing one's business requires a thorough review of required regulations, law and internal controls, especially as it pertains to information processing. It cannot be left to technology professionals to determine what, why and how information is to be protected; rather, a synchronised approach across stakeholder groups is recommended to ensure compliance gaps are minimised. The DPIA is but one method to hone an organisation's approach to identifying where critical personal data is stored, processed and managed.

This then pivots to the second common thread, risk management. Focusing on the 'crown jewels' and critical areas of business impact will allow organisations to eat the elephant one appropriate chunk at a time. Central to most regulations, frameworks, standards and best practice presented is the primacy of risk management. We live in an imperfect world and things will go wrong, so preparedness is paramount to business permanency and information compliance. Furthermore, the importance of having business and technology stakeholders collaboratively identify and assess information risk is paramount to architecting a measured response to a diverse threat landscape. Information risk management needs to be a pervasive, cyclical process that is supported across stakeholder groups to ensure a fluid response to an ever-changing risk landscape. The importance of having a risk culture mindset cannot be overstated, where it is persistent across the organisation, improving the identification of risks as and when they appear. For the 'defence in depth' equation to balance, the human factor of risk management is an important variable to factor in.

A third central theme, that of assembling a formal information and data governance capability, aiding in the establishment of holistic solutions (including management, operational and technical) to the lifecycle management of information assets, is presented. This cohesive framework, which supports business objectives, while meeting regulatory and compliance obligations, presents the opportunity to foster a culture of data ownership, accountability and stewardship, encapsulating the diverse needs of affected stakeholders. Organisations with a strong information and data governance culture are best positioned to respond to the fast paced and ever changing regulatory landscape. Beyond the protection of information assets, the opportunity to exploit the vast amounts of data instantiated by the execution of business processes should not be squandered, provided due ethical, legal and compliance factors have been applied.

Once the business information landscape and identified risk quantum has been carefully considered, appropriate safeguards to sensitive data and information assets need to be deployed (people, process, contractual and technology). Here we encounter the fourth key element, that of information security frameworks and standards. It has been shown that an extensive set of options prevail for the security professional to consider. While many overlapping and potentially conflicting perspectives exist, it must be acknowledged that significant attempts have been made to unify and align the various control libraries. The good news is that much of the work has been done by industry and governmental role players, it is this '20 per cent' that needs significant cognitive immersion. Ultimately, organisations cannot evade the required work to critically assess their needs and construct a viable, fit-for-purpose information security management system that balances cost and effort with business value and risk. Furthermore, the evolution of business and technology service models (cloud computing, etc.) will require additional controls to be implemented as the digital boundary blurs beyond the reaches of traditional organisational control. Strong encryption to data at rest and in transit, coupled with multifactor access control are two priority examples of controls that are suggested when considering cloud deployment models, especially when CSPs are hosting sensitive data.

Finally, the remaining key theme of incident response is discussed, one that has grown with prominence over the past years and is often considered as a 'capitulation' to an ongoing, persistent threat to information veracity. Nevertheless, this should not diminish the resolve of information security practitioners and accountable executives. Indeed, the most diligent, meticulous and prepared organisation cannot guarantee a residual risk posture impervious to successful attack. There is an acknowledgment that the perspective has changed from 'not if' an information breach will happen, but 'when'. This, however, should not diminish the resolve of organisations to do the right things. The readiness to respond when things go wrong must be a top priority for all stakeholder groups. Hiding behind closed doors is no longer accepted, and transparency with external and internal stakeholders is mandatory, with clear communications and tested responses to data breaches or unavailability events. Incident response and continuity management are tightly coupled, as they orchestrate the various activities to ensure ongoing business operations when bad things happen. Consequently, the ability to detect information breaches is a critical requirement, necessitating an orchestrated deployment of technical and procedural controls to inform key stakeholders. The advent of new regulations has further highlighted this as a mandatory obligation for all in-scope organisations, as evidenced by the GDPR.

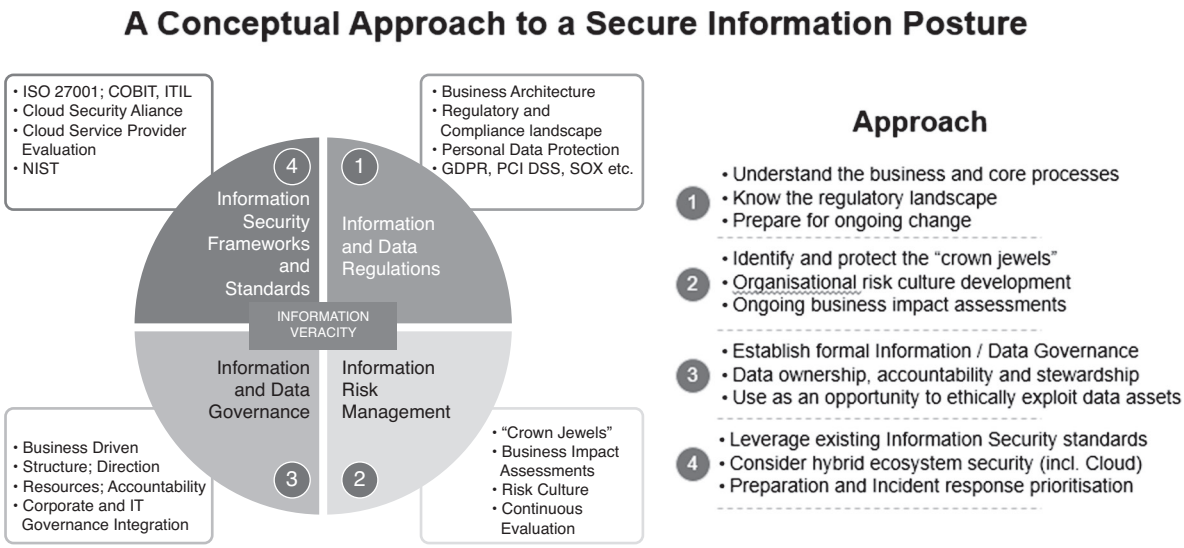


Figure 1: Summary of primary themes across information veracity

In conclusion, Figure 1 illustrates the key themes in this paper, with the most salient points summarised in each area. Organisations need to be transparent in their vision and application of effort, orchestrating activities across a range of stakeholder groups. Doing so allows for the attainment of many aspirations, in today's business world none so important perhaps, as the goal of information veracity.

References and notes

1. Veracity (2014) 'Collins English dictionary — complete and unabridged', 12th edn, available at: https://www.thefreedictionary.com/veracity (accessed 14th April, 2018).
2. Evans, N. and Price, J. (2014) 'Responsibility and accountability for information asset management (IAM) in organisations', *Electronic Journal of Information Systems Evaluation*, July, Vol. 17, No. (1), pp. 113–121. Available from: Business Source Complete
3. Zwolenski, M. and Weatherill, L. (2014) 'The digital universe: Rich data and the increasing value of the internet of things', *Australian Journal of Telecommunications and the Digital Economy*, Vol. 2, No. 3, p. 47.
4. Gikas, C. (2010) 'A general comparison of FISMA, HIPAA, ISO 27000 and PCI-DSS standards', *Information Security Journal*, Vol. 19, No. 3, pp. 132–141.
5. Kalaiprasath, R., Elankavi, R. and Udayakumar, R. (2017) 'Cloud security and compliance — A semantic approach in end to end security', *International Journal on Smart Sensing & Intelligent Systems*, pp. 10482–10494.
6. Castelli, C. (2018) 'Revitalizing privacy and trust in a data-driven world. Key findings from The Global State of Information Security Survey 2018', PwC, available at: www.pwc.com/gsiss (accessed 15th March, 2018).
7. Eroshkin, S., Kameneva, N., Kovkov, D. and Sukhorukov, A. (2017) 'Conceptual system in the modern information management', *Procedia Computer Science*, 103, XII International Symposium Intelligent Systems 2016, INTELS 2016, 5–7th October, 2016, Moscow, Russia), pp. 609–612.
8. von Solms, R. and van Niekerk, J. (2013) 'From information security to cyber security'. *Computers & Security*, 38, Cybercrime in the Digital Economy, pp. 97–102.
9. Benaroch, M. and Chernobai, A. (2017) 'Operational it failures, it value destruction, and board-level it governance changes', *MIS Quarterly*, Vol. 41, No. 3, pp. 729–A6.
10. Sloan, P. (2014) 'The compliance case for information governance', *Richmond Journal of Law & Technology*, Vol. 20, No. 2, pp. 1–46.
11. Nieuwesteeg, B. (2016) 'Quantifying key characteristics of 71 data protection laws', *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, No. 3, p. 182.
12. Goddard, M. (2017) 'The EU General Data Protection Regulation (GDPR): European regulation that has a global impact', *International Journal of Market Research*, Vol. 59, No. 6, pp. 703–705.
13. Flint, D. (2017) 'Sharing the risk: Processors and the GDPR', *Business Law Review (UK)*, No. 4, p. 171.
14. Bowman, J. and Gufflet, M. (2017) 'Meeting the challenge of a global GDPR and BCR programme',

European Data Protection Law Review (EDPL), Vol. 3, No. 2, pp. 257–261.
15. Shaw, A. (2010) 'Data breach: From notification to prevention using PCI DSS', *Columbia Journal of Law and Social Problems*, No. 4, p. 517.
16. Cook, S., Probert, D. and Martin, S. (2009) 'The lived experience of information technology workers with Sarbanes-Oxley compliance responsibilities', *Journal of Global Business Issues*, Vol. 3, No. 1, pp. 23–31.
17. Johnson, M., Goetz, E. and Pfleeger, S. (2009) 'Security through information risk management', *IEEE Security and Privacy Magazine*, No. 3, p. 45.
18. Wheeler, J. A. (2016) 'How to get your CEO to embrace digital risk management', *Gartner Research*.
19. Nocco, B. and Stulz, R. (2006) 'Enterprise risk management: Theory and practice', *Journal of Applied Corporate Finance*, No. 4, p. 8.
20. Lundqvist, S. (2014) 'An exploratory study of enterprise risk management: Pillars of ERM', *Journal of Accounting, Auditing & Finance*, Vol. 29, No. 3, pp. 393–429.
21. Yang, T., Ku, C. and Liu, M. (2016) 'An integrated system for information security management with the unified framework', *Journal of Risk Research*, Vol. 19, No. 1, pp. 21–41.
22. Lalev, A. (2017) Methods and instruments for enhancing cloud computing security in small and medium sized enterprises. *Business Management/ Biznes Upravlenie*, No. 2, p. 38.
23. Yordanov, A. (2017) 'Nature and ideal steps of the data protection impact assessment under the General Data Protection Regulation', *European Data Protection Law Review (EDPL)*, Vol. 3, No. 4, pp. 486–495.
24. Ee, H. (2014) 'Business continuity 2014: From traditional to integrated business continuity management', *Journal of Business Continuity & Emergency Planning*, Vol. 8, No. 2, pp. 102–105.
25. Putte, D. and Verhelst, M. (2013) 'Cyber crime: Can a standard risk analysis help in the challenges facing business continuity managers?', *Journal of Business Continuity & Emergency Planning*, Vol. 7, No. 2, p. 126.
26. Grispos, G., Glisson, W. B., Bourrie, D., Storer, T. and Miller, S. (2017) 'Security incident recognition and reporting (SIRR): An industrial perspective', arXiv preprint, arXiv:1706.06818, available at: http://eprints.gla.ac.uk/154923/3/154923.pdf (accessed 19th March, 2018).
27. Sloan, P. (2014) 'The compliance case for information governance', *Richmond Journal of Law & Technology*, Vol. 20, No. 2, pp. 1–3.
28. Ragan, C. R. (2013) Information governance: It's a duty and it's smart business. *Richmond Journal of Law & Technology*, Vol. 19, No. 4, p. 12.
29. Sloan, P. (2014) 'The compliance case for information governance', *Richmond Journal of Law & Technology*, Vol. 20, No. 2, pp. 21–22.
30. Bennett, S. (2017) 'What is information governance and how does it differ from data governance?', *Governance Directions*, Vol. 69, No. 8, pp. 462–467.
31. Dahlberg, T. and Nokkala, T. (2015) 'A framework for the corporate governance of data — Theoretical background and empirical evidence', *Business, Management & Education/Verslas, Vadyba Ir Studijos*, Vol. 13, No. 1, pp. 25–45.
32. Egelstaff, R. and Wells, M. (2013) 'Data governance frameworks and change management', *Studies in Health Technology and Informatics*, Vol. 193, pp. 108–119.
33. Cohn, B. (2015) 'Data governance: A quality imperative in the era of big data, open data, and beyond', *I/S: A Journal of Law & Policy for the Information Society*, Vol. 10, No. 3, pp. 811–826.
34. Al-Ruithe, M., Benkhelifa, E. and Hameed, K. (2018) 'Data governance taxonomy: Cloud versus non-cloud', *Sustainability*, Vol. 10, No. 1, p. 95.
35. Khatri, V. and Brown, C. (2010) 'Designing data governance', *Communications of the ACM* , Vol. 53, No. 1, pp. 148–152.
36. Raether, R. (2015) 'Ten years later: Data governance in the decade of the data breach', *Federal Lawyer*, Vol. 62, No. 8, pp. 40–44.
37. Tallon, P., Ramirez, R. and Short, J. (2013) 'The information artifact in IT governance: Toward a theory of information governance', *Journal of Management Information Systems*, Vol. 30, No. 3, pp. 141–178.
38. Hagmann, J. (2013) 'Information governance–beyond the buzz', *Records Management Journal*, Vol. 25/23, No. 3, pp. 228–240.
39. Goosen, R. and Rudman, R. (2013) 'The development of an integrated framework in order to address King III's IT governance principles at a strategic level', *South African Journal of Business Management*, No. 4, p. 91.
40. ISACA (2013) 'COBIT 5: Enabling information', available at: http://www.isaca.org/COBIT/Pages/COBIT-5-Enabling-Information-product-page.aspx (accessed 21st March, 2018).
41. Preittigun, A., Chantatub, W. and Vatanasakdakul, S. (2012) 'A comparison between IT governance research and concepts in COBIT 5', *International Journal of Research in Management & Technology*, Vol. 2, No. 6, pp. 581–590.
42. Evans, N. and Price, J. (2014) 'Responsibility and accountability for information asset management (IAM) in organisations', *Electronic Journal Of Information Systems Evaluation*, Vol. 17, No. 1, pp. 115–119.
43. Clements, T. (2018) 'Maintaining data protection and privacy beyond GDPR implementation', ISACA, available at: www.isaca.org/Data-Protection-Beyond-GDPR (accessed 14th March, 2018).
44. Haufe, K., Colomo-Palacios, R., Dzombeta, S., Brandis, K. and Stantchev, V. (2016) 'Security management standards: A mapping', *Procedia Computer Science*, Vol. 100, 755–761; International Conference on ENTERprise Information Systems/International Conference on Project MANagement/International Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN/HCist.
45. Le, N. and Hoang, D. (2017) 'Capability maturity model and metrics framework for cyber cloud security', *Scalable Computing: Practice & Experience*, Vol. 18, No. 4, pp. 277–290.

46. Ardagna, C., Asal, R., Damiani, E. and Quang Hieu, V. (2015) 'From security to assurance in the cloud: A survey', *ACM Computing Surveys*, Vol. 48, No. 1, pp. 2-2–50.
47. Rizvi, S., Karpinski, K., Kelly, B. and Walker, T. (2015) 'Utilizing third party auditing to manage trust in the cloud', *Procedia Computer Science*, Vol. 61, pp. 191–197; Complex Adaptive Systems, San Jose, CA, 2–4 November.
48. Cloud Security Alliance (n.d.) 'Cloud Security Alliance cloud controls matrix', available at: https://cloudsecurityalliance.org/group/cloud-controls-matrix/#_overview (accessed 16th March, 2018).
49. Heiser, J. (2017) 'How to evaluate cloud service provider security', *Gartner Research*.
50. Van Bon, J. and van Selm, L. (2008) 'ISO/IEC 20000 — An introduction', Van Haren, The Netherlands.
51. Himi, A., El Masbahi, M., Bahsani, S. and Semma, A. (2011) 'A new approach to supply chain management based on pooling ITIL and APICS principles and practices', *International Journal of Computer Science Issues*, Vol. 8, pp. 355–363.

Digital responsibility redefined in Denmark

Received: 9th March, 2019

Mads Hennelund
is a business advisor, working for the Danish consultancy, Nextwork A/S. He is an expert on digital transformation and advises companies and organisations — and in particular the financial industry — about strategy and competitive positioning, business development, branding, data ethics and organisational change in a hyper-digital world. He holds a master's degree in business administration and philosophy from Copenhagen Business School in Denmark.

Nextwork.as, Vester Farimagsgade 15, 4th Floor, 1606 København V, Denmark
E-mail: mads@nextwork.as

Abstract Denmark is venturing into new unexplored digital territory with eight political initiatives on data ethics that receive both industry and consumer support. One initiative requires large companies to include a statement about the company's data ethics policy in the annual report. Another is the establishment of a data ethics label to increase transparency about a company's data ethical standards. This paper explores these initiatives in depth and discusses the consequences. Danish companies might be able to establish higher levels of trust with customers, but with a primarily negative discourse surrounding the topic of data ethics and digital responsibility, this paper also discusses the possibility that high standards on digital responsibility may compromise the potential for further value creation and innovation through more data usage, safely and responsibly. Drawing on schools of philosophy as well as technological development, this paper proposes a framework for evaluating data ethics and privacy, not only in terms of things that happen that we do not like, but also regarding things that are not happening yet, but which ought to be, perhaps. Furthermore, the paper proposes the need for industry specific data ethical themes, and proposes five generic themes that should be used as a starting point for assessing data ethics and digital responsibility across industries.

KEYWORDS: digital responsibility, data ethics, digital transformation, assurance, consumer label, transparency, value creation, discourse analysis, utilitarianism, deontological ethics

INTRODUCTION: DENMARK TOWARDS UNCHARTED DIGITAL TERRITORY?

With eight new data ethical initiatives proposed by the Danish government, Denmark is setting course for new standards in digital responsibility. And with support from industry and consumer associations, Denmark is facing towards new unexploited digital lands. But how do we go about this, and what are the pitfalls and dangers, especially given the political understanding of digital transition and data responsibility? What story will we write as we turn the page?

Denmark is the Nordic centre from which Vikings once ventured to conquer and dominate new territories with robustly-crafted ships that enabled them to set foot in new foreign territories. But while Denmark's geographical domination may have shrunk since then, its digital mastery

has evolved and expanded markedly. Limits for digital possibilities are constantly being pushed outwards towards yet uncharted territory, for both consumers and businesses, large and small. More possibilities are emerging with the increasing amount of data from open public databases, as well as personal data points that are progressively becoming an integral part of service experience, innovation, product development, proper patient treatment, etc. Thus, we yet again venture towards new unexplored territories, and just as the well-crafted ships enabled the Vikings to fare safely to new lands, a political focus on the (re)definition of the rules applied to 'digital responsibility' might very well be the measures needed in order to master challenges and opportunities in new digital territories. Eight data ethical initiatives have recently been put forth by the current government, and the suggestions are backed heavily by leading voices representing both industry and consumers.

Denmark is very digitally advanced. It is almost a cashless society. Mobile payment is possible in most physical stores. More stores do not even accept cash anymore. Financial technology and cross-system integration is highly advanced, enabling a primarily digital interaction between customer or business owner and the banks, pension, insurance companies or the municipality/state. Citizens may only receive mail digitally from their employer and public authorities regarding paychecks, social benefits, etc. Denmark has had social security numbers registered digitally since 1968, when the CPR register (Det Centrale Personregister, founded in 1924) was established as an automatic IT-based register.[1] This has created a highly digitized public administration practice. Automated business reporting for small businesses is advanced and in 2019, the Danish Business Authority will release a platform for digital reporting of financial statements. And meanwhile, for instance, drone teams have, for some time now, been established among the large Danish auditing companies. In short, boundaries are pushed by technological development, driven by user demand for smart goods and services, whether public or private, and it is driven partly by a digitised public administration. Regulation can also sometimes have a hard time keeping up. The General Data Protection Regulation (GDPR) certainly has helped a lot, but digital responsibility and data ethics is important when new boundaries need to be set or pushed even further to enable value creation.

BACKGROUND TO THE EIGHT POLITICAL DATA ETHICS INITIATIVES FOR BUSINESS

Back in 2018, the Danish government established a data ethics expert group as part of its overall digital growth strategy. The group consisted of different people with knowledge about digital services, data usage and ethics. The expert group eventually proposed various suggestions and by the end of January 2019, the Ministry of Business Affairs announced that the government was going to implement eight initiatives that will allegedly support businesses' responsible and sustainable use of data:

(1) The establishment of an independent data ethics council. This council shall discuss the use of new technology in light of basic citizen rights, legal issues and societal values, etc.
(2) Make assurance reporting on businesses' data ethics policy a legal requirement.
(3) Establishment of a data ethics label. This label is intended to provide transparency for consumers and incentivise businesses to use data responsibly.
(4) Increase knowledge, nation-wide, on data ethics. Citizens and consumers ought to learn to set proper requirements regarding service providers, that is, digital and data literacy.

(5) Prepare a dynamic toolkit for data ethics. Companies will receive guidance and relevant tools for responsible data usage.
(6) Support Denmark as a front runner when it comes to data ethics. The Danish approach to data ethics should be promoted especially in the European Union (EU) and Organisation for Economic Co-operation and Development (OECD). Danish values ought to contribute to the formation of long-term solutions in different (presumably international) contexts.
(7) Follow the development of new innovative companies with business ideas that revolve around ethical data usage.
(8) Look into the possibility of strengthening the focus on data ethics in regards to public spending. Public buyers ought to have guidelines on how data ethical responsibility can permeate through to the purchase of digital solutions and services.[2]

INDUSTRY SUPPORT FOR DATA ETHICS INITIATIVES

Some of these eight initiatives targeted at businesses might seem somewhat invasive, especially the legal requirement of assurance regarding data ethics policy, which we will get back to. Interestingly, the Confederation of Danish Industry (DI), which represents a large part of Danish companies, is backing the political focus and the eight initiatives. In a press release,[3] the confederation's digital director said that the digital possibilities should be pursued with respect to ethical challenges:

> If we handle these challenges correctly and in a timely manner, we have the opportunity for making responsibility a strength for companies in Denmark. And DI wants to contribute to that. (author's translation) — Lars Frelle-Petersen, Director, DI

According to Lars Frelle-Petersen, both companies and end-users are key components in creating the demand for digital responsibility:[4]

> Basically, it is about making digital responsibility attractive for all parties. That is useless if we make being digitally responsible a troublesome game. Or if it becomes too technical or impenetrable for end-users and collaborators to understand whether authorities or companies are digitally responsible. This is especially important to keep in mind if imposing a legal requirement regarding assurance on companies' data ethics policy shall make sense. (author's translation) — Lars Frelle-Petersen, Director, DI

It is clear that digital responsibility is seen as a potential competitive advantage. There is a chance that this 'competitive advantage', based on responsibility, can turn out like Nietzsche's description of the origin of Christian moral values in his *On the Genealogy of Morality,* in which he describes Christianity as a religion of the angry and oppressed people.[5] New data ethics must not be the result of frustrated moralisation of actors who are primarily 'just' fed up with the current tech supremacy and their advantages. If the Danish competitive response to new entrants, such as Alibaba, Amazon and Chinese tech companies, etc., is: 'Well, we can't do what they do, so doing that is not good anyway. Therefore, choose us because we *don't* do that.'

Philosophy aside, let us take a closer look at what are possibly the two most interesting political initiatives; that is, the assurance requirement and the label, respectively.

MAKING ASSURANCE REPORTING ON BUSINESSES' DATA ETHICS POLICY A LEGAL REQUIREMENT

The initiative making assurance on data ethics policy a legal requirement, and thereby making it compulsory to include data ethics in a company's management report and financial statement, is perhaps the

most controversial suggestion. The initiative only targets large companies. In theory, this practice should increase transparency towards other business collaborators, investors and other stakeholders. But it poses a requirement on external auditors who will need to be able to provide proper assurance on sound data ethics policies. External auditors already seem to have a hard time figuring out very complex areas such as evaluating unlisted shares mark-to-model, where valuations rely on a complex set of variables and timeframes that are more open to interpretation, as opposed to market-to-market valuation where the market price determines value. The former type of shares are the ones pension companies and other large institutional investors are increasingly investing in, given the current low interest rate environment. According to the director of FSR – Danish Auditors, Tom Vile Jensen, the association of Danish Auditors greet welcome the political initiative. He stated this in a press release on the association's website:[6]

> It is an interesting recommendation wanting to make assurance on companies' data ethics policy and data usage a part of the management statement. We have seen this in the CSR field and it can be a good way to increase attention towards data ethics among companies and Auditors. (author's translation) — Tom Vile Jensen, Director, FSR, Danish Auditors

One of the toughest challenges regarding creating a standardised approach for providing assurance on data ethics policies is finding the right scope, which is also adjustable to different types of companies and industries. Data veracity is a big concept and a matter of ethics, in so far as good, healthy data sharing depends on trust. Biased data can become a liability for companies if they are advanced and have high tech and artificial intelligence (AI) powered business intelligence and customer analytics systems to support decision making. Also, tech companies that have collected vast amounts of permissions, without ensuring things such as full transparency, may be in for a wild ride if people act on their right to be forgotten or withdraw their data in a certain format.[7] These are potential financial risks connected to data ethics. For more on this, read the GDPR and privacy analyst Chiara Rustici's *Forbes* article about new data risks and valuation metrics.[8] What ethics and compliance criteria should be taken into account regarding an assurance on data ethics policy?

> If we are to take the next step with an independent assurance on the companies' data ethics policy, we will need a more detailed clarification regarding what is expected, data ethically, from the different companies. Because there will obviously be a difference as to what makes sense that a table manufacturing company declares and what a toy manufacturing company, that builds internet assistants into the products, are to declare on. It depends on the amount of data and not least the use context. The possibilities are many. (author's translation) — Tom Vile Jensen, Director, FSR, Danish Auditors[9]

The assurance should, as mentioned, be driven by both risk criteria and compliance criteria. The compliance criteria will need to refer to an established framework based on deontological and utilitarian ethics. You cannot expect external auditors to pick up Immanuel Kant and Jeremy Bentham's best works during auditing. Thus, data ethical principles will be elaborated on below, at the end of this paper.

ESTABLISHING A LABEL FOR DATA ETHICS: INCREASING TRANSPARENCY WITHOUT INCREASING NAVIGATIONAL NOISE

Less controversial is the initiative regarding a label indicating the level of digital responsibility a company has achieved. Having a number of labels can counter the

sole purpose of having these complexity-reducing signs and markers that support users' navigation and decision-making processes. In Denmark, we already have a label called 'E-mærket', which indicates trust in regard to online purchasing. A label for data ethics should thus either be incorporated into the 'E-mærket' label, or encompass more markers that do not overlap excessively with the already existing one.

Digital Director of DI, Lars Frelle-Petersen, says that the label regarding digital responsibility (political initiative number three) should encompass both IT security, privacy (the GDPR), and responsible use of big data and artificial intelligence:

> There should not be more labels than necessary. Therefore, one should be careful with making labels with narrow focus on for instance data ethics or IT-security exclusively. (author's translation) — Lars Frelle-Petersen, Director, DI[10]

The Danish consumer protection association, Tænk, also supports the establishment of a data ethics labeling initiative:

> We need to strengthen safety and transparency if we are to fully exploit the many digital opportunities. We need strong enforcement of the GDPR, but consumers should also be given tools to choose companies they can trust in order to make data protection a competition parameter. A new label for digital responsibility — or better yet, expanding one of the existing consumer labels already out there like E-mærket — is a good place to start along with digital literacy. — Anette Christoffersen, CEO, Tænk (personal communication)

The process of creating a label for data ethics in Denmark is still underway; however, in an official guide for businesses published by the official Danish Data Protection Agency (Datatilsynet), the intended framework for a future certified label has been described. In the guide, the protection agency emphasises how a label-service and a code of conduct will be a benefit for especially small companies and other organisations:

> The intent is that these schemes [code of conduct and label certification] shall help and guide responsible controllers of data complying with the GDPR, including circumstances which characterizes a specific industry ... The creation of e.g. code of conducts will be a useful tool especially for micro-, small and medium-sized companies to help ensure compliance with the GDPR. (author's translation) — Danish Data Protection Agency[11]

The emphasis is on the transparency intended for official authorities and designing relevant data ethical standards that are appropriate to the different industries and their consumers (eg children might demand different ethical adjustments than adults).

According to the GDPR, article 24, the collector and controller of personal data has a responsibility to prove that their technical and organisational methods do not violate the GDPR. A label is a certification mechanism that allows a company to provide visible proof of their compliance with the Danish and/or European regulations through design and standard set-ups.

In Denmark, the Danish Data Protection Agency and privately-owned accredited certification agencies will collaborate on overseeing and implementing certified labels. The intention is that the accredited agencies will be in charge of most of the certifications, as well as undertaking the job of renewing the certifications every third year and possibly withdrawing a label if a company no longer meets the requirements. The criteria for obtaining the label are yet to be determined, but the intention is that the accredited agencies will set out the criteria of official national and international guidelines, to secure the streamlining of practices and ensuring that companies are held to the

same high ethical standards (note that an attempt to make such guidelines has already been made by the European Network and Information Security Agency (ENISA)[12]). The certification agencies will then design the label and adjust it to the different ethical demands that suit the different industries.

A label will not, however, exempt companies from the responsibility of formal regulations, such as the GDPR, but it will show that an accredited third party has looked at and approved their data practices. Thus, it will be a 'trust mediator', so to speak. It is the official intention that the Danish Data Protection Agency will issue accreditations for certification agencies, inspect all certifications made, and possibly make injunctions or supervise the certification agencies in their certification processes. That way, private players will offer the certification service that is relevant for the business in question and the Danish Data Protection Agency will be the legal authority behind the certification.

It is important to keep in mind that a company with an approved label will still be subject to administrative fines if its practices violate national regulation or GDPR requirements, as the label cannot exempt the company from penalties. But, interestingly, in the case where a company is carrying a certified label and has followed the label policies, this could provide mitigating circumstances, which could lead to a lower fine being issued (in accordance with the GDPR, article 83.2, j+k). Similarly, if a company is not carrying a label, this could result in worsening circumstances and a bigger administrative fine. Hence, financial incentives and regulatory demands will likely drive forward the industry adoption of a data ethics label, and not the customer need for transparency and keeping up with competitors — that is, compliance with new industry threshold standards — alone.

Ensuring transparency is, as already stated, one of the biggest hurdles regarding this data ethics labelling initiative. Another hurdle is the act of ensuring that this label will not just be a GDPR compliance label. And this really goes for both the assurance initiative regarding companies' data ethics policy described in the previous section and the label initiative described in this section. Because this is where it gets really difficult and where lawyers, GDPR experts, auditors and traditional compliance staff fall short. There is a need for business data ethicists and a data ethics framework that also take cyber risk aspects, and perhaps even consumer/behavioural insights, into account.

POLITICAL DISCOURSE NEEDS CHANGE

While there seems to be support for these new initiatives regarding digital responsibility, what is the political understanding of digital responsibility and data ethics, really? The knowledge about digital transformation, technology and data possibilities, as well as risks, partly determine the horizon. And the language we use, it is argued, is important too. The discussion around data ethics is peculiar, since it often revolves around a discourse about things we do not want. The following analysis takes a closer look at the discourse existing among the IT spokespersons from the eight political parties represented in the Danish parliament in 2018. The statements about data ethics were originally published by *Prosa*,[13] a Danish journal for IT professionals, and were all answers to eight different questions about data ethics. The politicians answered eight questions individually from which statements within either a positive or negative discourse were counted. Statements that did not fit within either discourse have been counted as neutral (see Figure 1).

From Ernesto Laclau and Chantal Mouffe's tradition of discourse analysis, we have learned how a concept, or sign, exists in our culture not only within the context of its ascription of meaning, but also in the reduction of what that concept does not

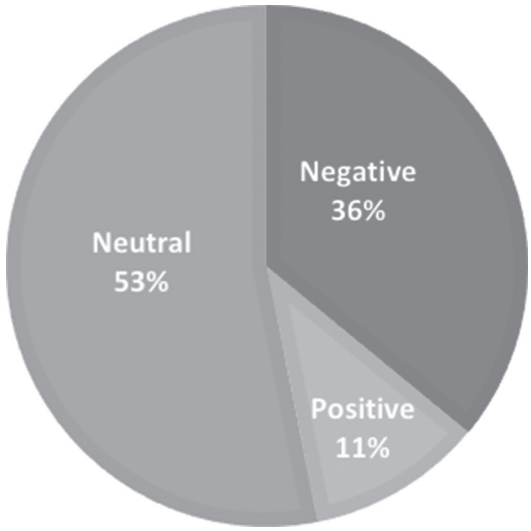


Figure 1: The negative and positive discourse in percentages within the interview

contain — what Laclau and Mouffe call *the field of discursivity*.[14] A discourse frames the sign, in this case data ethics, with a set of meanings that constitutes a uniqueness of reference within the discourse. The field of discursivity is the reservoir of meaning, which exists among alternative discourses. In general, there seems to be a (mistaken) privacy versus value dichotomy, which is the result of two different discourses that both encompass alternative conceptual understandings of data ethics.

Based on the article,[15] this brief analysis has uncovered two different discourses. One discourse emphasises the meaning of data ethics in a positive (thumbs up) light. As Figure 2 indicates, examples could be how data is an important resource to society, a solution to future problems, hopes about developing public services by increasing efficiency and freeing up resources, as well as advancing the private sector (eg pharma and medical); there is excitement and curiosity about possible future prospects in regards to utilisation of more data. The other (thumbs down) discourse associates data ethics with negativity, suspicion, surveillance and unease. Examples are concerns about the right to privacy (or risk of losing it), risk of misuse, the breaking of societal order, discrimination, killer robots, and so on. This discourse sees companies or technologies as presenting an ever-present threat of human exploitation, which induces the need for protection against and constraint from data.

Figure 2: Illustration of words used in positive and negative discourse about data ethics

This discourse denunciates the positive prospects of data collection and emphasises the negative scenarios, because data represents a threat to society and an always present potential violation of individual rights.

The analysis revealed that the negative discourse appeared in 23 out of 64 answers (ie 36 per cent), whereas the positive discourse appeared in only seven answers (ie 11 per cent).

It would appear there is a real danger that data ethics will in reality just be 'data critique' in disguise. The critical aspect is important, but Denmark could risk missing out on much more than it gains by engaging whole-heartedly in data ethics when the current discourse is so negatively charged. This will not help consumers navigate nor businesses innovate; quite the contrary, it risks further igniting a developing backlash and stifling innovation. It comes down to this: is data ethics about acknowledging things that happen (or can happen), that we do not like? Or is it about encouraging processes and conduct that perhaps are not happening yet, but which ought to be? Of course, it is not an either/or answer; however currently, the former prevails.

HOW BEST TO BUILD DIGITAL RESPONSIBILITY: EXAMPLES OF HOW TO GO FORWARD TOWARDS NEW UNCHARTED TERRITORY

In practice in the real world, data ethics are not just about what you should do less of. Patient organisations, for instance, revolve around patients sharing stories that other patients can relate to and which might help them through tough times. But if you wish to be protective of your data and avoid the risk of having unwanted parties looking at these forums where your personal stories (ie personal data) is shared, the advantages of data sharing might be limited. Therefore, breaking the privacy versus value dichotomy is at the outset a rhetorical matter; however, it is also a technological matter, ensuring the right privacy by design infrastructure, data lineage technologies and personal data stores. As Julian Ranger from Digi.me wrote in his 2019 predictions[16] earlier this year, there will be a continuing debate about the belief that privacy and data ethics are counter to innovation and value creation:

> ...fears of legislation stifling change will continue, but legacy businesses will increasingly come around to the opportunity of greater data privacy as a way to get closer to their customers or users and serve them better. Google and Facebook already made just that point in the US Senate hearings at the end of last year and we expect this counter view, of individual ownership of data actually increasing innovation, to be both proved and commonly held as a belief as the year goes on. (Julian Ranger, founder, Digi.me)

Similarly, a recent paper — 'Powering the digital economy: Regulatory approaches to securing consumer privacy, trust and security'[17] — concluded that openness and trust are key components for companies and customers. And an essential ingredient here is data usage:

> Companies that can demonstrably show they understand this and take the issue seriously will naturally move ahead of their competitors in the race for customers. As such, openness and trust should emerge in boardrooms around the world as a key opportunity to exploit, rather than seeing data protection as another regulatory hoop which organisations must jump through.

The insurance industry is a good example of this; it faces new dilemmas in the process of moving towards the collection of more personal data, especially unstructured data such as behaviour on social media, lifestyle and emotional data. Is this then a bad thing that we should inhibit? Not necessarily. This increase in data collection is helping both insurers and policyholders in regards to the

onboarding process, and it helps actuaries to be better at determining risk and price. These new data points also help insurers to be better at preventing unfortunate outcomes that can have individual as well as societal consequences. Not only can data collection help insurers identify potential bad outcomes, they can also make prevention plans that are much more personalised as a result of more precise personal data, which allows insurance companies to create more effective services. This ultimately benefits both the insurers who will have to pay less in damages, as well as policyholders, since they will be able to get help in preventing a bad health, stress and stress-related illnesses. Additionally, big data and AI can enable insurance fraud investigators to collect fewer data points on honest customers, whereas fraudsters can be closely monitored (and with a compelling reason). All the while the number of false positives — the real victims from a data privacy and ethics perspective — is reduced markedly. (Note that in practice, the Danish insurance industry is not at this stage of development yet.)

Utilising more data points for insurance is a good thing, both from a business perspective and an ethical perspective. If more data can be used to identify bad habits such as dangerous driving behaviour, then doing so and incentivising behavioural change is both legitimate and responsible in regards to the individual, the business and society. Where it gets trickier is when more data points can be used to reveal those, who have certain behavioural patterns or conditions that cannot be changed, and therefore are risky enough to be uninsurable. In this scenario, personal data is being used against the policyholder, and while this can be ethically justifiable from a utility standpoint in regard to society (ie the rest of us who would otherwise have to pay a more unfair insurance premium) and business, it is much more problematic from an ethical perspective in regards to individual rights. So, the industry either needs to self-regulate and determine how far they want to go, or collaborate with governments on what solutions should be provided for these uninsurable citizens.

When building a digitally responsible society, it is first and foremost important to realise that it is not solely about restraint, modesty and having more puritan values. Digital responsibility and data ethics are just as much about utilising the benefits of the internet, even more than we do today, by giving the individual more data-driven and valuable advice, and identifying and changing bad behaviour that can in fact be changed for the better, for both the customer and society (if the individual wants this). If, for instance, unions were allowed to use big data and personal data to make predictive models — that is, predictive unionism — and offer more service personalisation, then workers potentially could be much better prepared for when the best time is to negotiate for a pay rise and how much to ask for.

Tracking people's input into the economy — which always manifests itself as data — would be another way to increase digital responsibility and data ethics positively. A more effective value distribution and exchange is key to the digital world and the data economy, because we want more data inputs to fuel innovation and science; technology like data lineage enables this. People who contribute their personal data contribute to the economy, and data lineage is a tool for potentially tracing these value contributions back to their sources, enabling rewards to go to these productive entities, which usually is ultimately *people*. You can say that data lineage is the data ethical counterpart to privacy as anonymity and data minimisation. The social and equality enhancing potentials are, of course, immense here.

From a consumer perspective, transparency about companies' collection and utilisation of data has recently been shown to be of growing importance.

Although there has not been scientific research — in a Danish context — of how proclaiming data ethics can benefit a company's brand and attract more customers, research done in comparable countries is pointing towards an increasing critical awareness of trust when customers choose a provider for different online services. In January this year, Sitra — the independent public innovation fund in Finland — conducted a survey amongst consumers in Finland, Germany, Netherlands and France which, along with other results, showed that a total of 42 per cent of respondents agree or strongly agree that lack of trust in service providers prevents them from using their digital services.[18] The same survey also concluded the following:

> Secure service (60%), reliability of service provider (54%) as well as the fact that the purpose of the data collected is clearly and transparently reported in the service/application (43%) are the three most important features of digital applications/services.[19]

These parameters scored higher than traditional key online features such as 'it's free of charge' (39 per cent), 'it's personalised' (16 per cent) or 'easy to use' (37 per cent). Lastly, the survey showed that 66 per cent of respondents thought it was important for companies to use a label that shows the ethical use of personal data. Based on these results, the conclusion must be that ethical use of online data is highly important for the northern European consumer. This emphasises the need to consider data ethics as a key brand attribute when building an online brand in this market. Trust between consumers and companies is now more important than usability and cost, which is also backed by global research and the advisory firm, Gartner, who identified data ethics as one of the top ten strategic technology trends for 2019, stressing the fact that 'the backlash will only increase for organizations that are not proactively addressing these concerns.'[20]

Modern companies ought to consider adjusting to the higher (and industry specific) data ethical standards demanded by consumers, which demands professional ethicists or a need for third party providers to help. It can also potentially benefit smaller businesses, who are without large in-house legal departments, because it creates new barriers to competition building on customer trust relations.

ETHICS AND DATA: AN INITIAL FRAMEWORK FOR EVALUATING COMPANIES' DATA ETHICS POLICIES AND LABELLING FOR DIGITAL RESPONSIBILITY

Where do we go from here? What is ethics and what is 'good' data usage? How do you define these terms? We can as companies, industries, governments and society go in two opposite yet still ethically justifiable directions in regard to data usage. We can minimise available data and increase anonymisation in respect to the individual (security by obscurity), or we can incentivise more data proliferation with respect to the individual, thereby recognising him or her for their respective value contributions. These approaches to data availability can be regarded as ethically justifiable in terms of the utility — that is value creation, Bentham's utilitarianism — or in terms of respect for the individual — that is Kantian deontological ethics. To sum it up, these four dimensions create four data ethics viewpoints that can be used as a starting point when evaluating a company's data ethics policy (see Figure 3).

As is evident when looking at the gap between 1 and 3 in Figure 3, which accounts for the general privacy advocacy position, there as a huge difference in abstraction/complexity going from 'security' and 'privacy' to concepts such as 'data lineage', 'personal data stores' and 'privacy

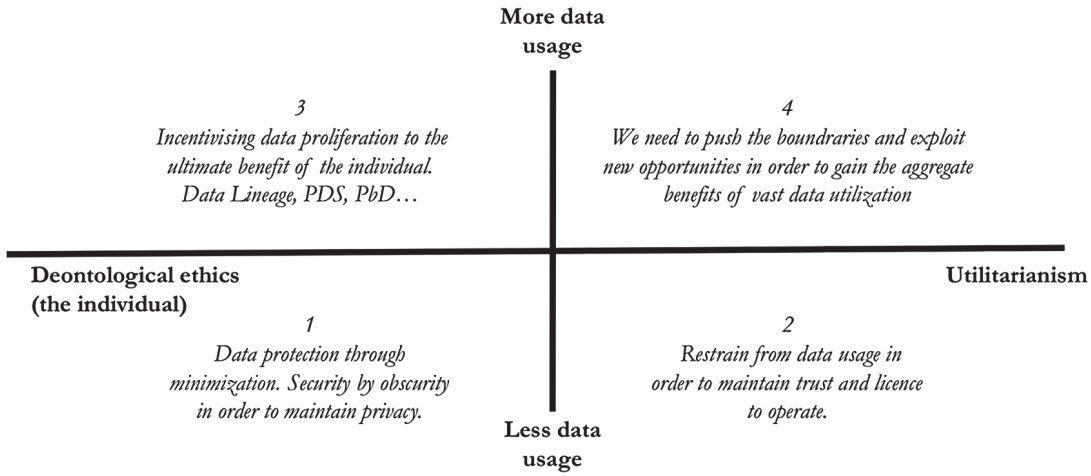


Figure 3: Matrix of the four generic data ethics positions
Note: PDS, personal data stores, are personal, digital databases, which function as autofill services with your personal information that automatically provide online business with the personal information they need to complete a purchase, for example, address, insurance number or credit card details, etc. PbD, privacy by design, is a way for companies to incorporate data protection measures in all parts of business processes from production to consumer service.

by design', which are concepts that allow for a much greater value utilisation while maintaining privacy completely. As such, technology can potentially break down the privacy–value dichotomy — that is, deontological ethics versus. utilitarianism. Therefore, there is also a learning gap in this equation as far as our understanding, horizon and common language on data ethics goes.

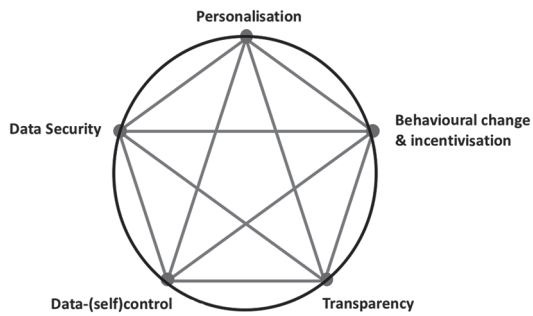


Figure 4: Model of the interconnection between digital-trust-enabling components

KEY DATA ETHICS THEMES

Earlier, the following data ethics themes that should be taken into consideration when evaluating a company's data ethics policy or establishing a label have been identified, indicating the sum of different digital-trust-enabling components (see Figure 4):[21]

- Transparency: is it transparent where my data is, how much there is, and how it is being treated and stored?
- Data security: is my data stored securely?
- Data (self)control and enablement: do I have sufficient control over my own data?
- Personalisation: is data used to treat me in a way that benefits me or is to my disadvantage?
- Behavioural change and incentivisation: is data used in a way that more or less legitimately incentivises me to change behaviour in a way that might impact me for better or for worse?

These are the generic themes. There are indeed additional industry specific themes, such as 'solidarity' in regards to insurance. If you contribute false data when reporting a claim (ie fraud), you compromise the concept

of solidarity in insurance, where people pool money together collectively for subsequent solidarity, transferring (redistribution) to the few unfortunate ones who experience an unforeseeable and tragic life event or accident. Establishing a set of industry specific benchmarks on data ethics themes across the financial industry, the pharma industry, the service industry, the consulting industry and so on, will be necessary when developing a framework for evaluating different companies' data ethics policies.

CONCLUSION

Digital responsibility is increasingly being viewed as an opportunity for gaining competitive advantage while increasing consumer protection and transparency. Being a digital frontier, it is only natural that data ethics in Denmark has received political awareness as well as industry and consumer support. Danish brands can potentially reach a level of deeper trust relations with their stakeholders compared to brands from other EU member states and the European Environment Agency (EEA). Eight concrete political initiatives now have the potential for paving the way for Denmark as a nation with clear emphasis on responsible and ethical use of data. A new assurance in regards to companies' data ethics policies in the management statement, as well as a data ethics label do, however, require building *upon* the GDPR and not just *compliance with* it. This is indeed a challenge for the different parties involved in the implementation of these two initiatives.

Meanwhile, the political discourse on data ethics is particularly negative, which may ultimately do more harm than good when it comes to building more trust and encouraging more digital business innovation. Various industries and consumer segments can benefit significantly from a more progressive data ethics approach in general, ensuring that *more* data — not less — benefits all parties involved in the economy.

Finally, four basic data ethics positions have been presented, which in different ways ensure an ethically justifiable approach to data usage. What we choose — within any given regulatory boundaries — depends partly on some degree of knowledge about technological possibilities, as well as our ethical point of view. It is crucial that industry specific data ethical themes are established — both when it comes to an assurance practice regarding large companies' data ethics policies, as well as a data ethics label/certificate. These themes will in most cases include and build upon the five generic themes presented in this paper: transparency, data security, data (self) control and enablement, personalisation/segmentation, as well as behavioural change and incentivisation.

THE JOURNEY TOWARDS NEW POSSIBILITIES CONTINUES

The journey towards yet undiscovered new territories goes on in Denmark's case. We are a small country that — as former US President Barack Obama said about us (and many other allies) — 'punches above its weight'. With high general trust and digital awareness, there is a real possibility that Denmark can become a test-hub for new innovative ideas and personal data business models, using data to try to solve complex societal issues, growing impact start-ups, nurturing impact investments, etc. In regards to digital responsibility, it is imperative that this encompasses the positive aspects alongside the fear-mongering aspects. There are large ambitions and potentials, but there is an urgent need for a new language to describe our new digital world. There is also a need for technological development and technical solutions to enable our need for privacy, further value creation and more efficient value distribution. It is to be hoped that politicians and industry leaders in Denmark continue forwards and further explore concepts and technologies of privacy

by design (PbD), personal data stores (PDS) and data lineage. 'En marche!' responsibly and sustainably when we turn the page and write the next chapter.

Acknowledgements

This paper written with assistance from my colleague, Kia Davies.

References and notes

1. Available at: https://cpr.dk/media/17545/udviklingen-paa-cpr-omraadet-frem-til-2009.pdf, pp. 27–28 (accessed 1st April, 2019).
2. The original eight initiatives are available at: Erhvervsministeriet [Ministry of Industry, Business and Financial Affairs] (2019) 'Faktaark — Dataetiske tiltag for erhvervslivet [Facts — Data ethical initiatives for financial affairs]', 29th January, https://em.dk/media/12932/faktaark_dataetiske-initiativer.pdf (accessed 1st April, 2019).
3. DI (2019) 'Krav om dataetik på vej [Requirements about data ethics are coming]', available at: https://www.danskindustri.dk/di-business/arkiv/nyheder/2019/2/krav-om-dataetik-pa-vej/ (accessed 5th March, 2019).
4. *Ibid*.
5. Nietzsche, F. (2006 [1887]) 'On the Genealogy of Morality', K. Ansell-Pearson (ed.). Cambridge University Press, Cambridge, available at: http://www.inp.uw.edu.pl/mdsie/Political_Thought/GeneologyofMorals.pdf (accessed 1st April, 2019).
6. FSR — Danish Auditors (2019) 'Dataetik på vej ind i virksomhedernes ledelsesberetning?[Data ethics making its way to companies' management reports?]', 22th November, available at: http://m.fsr.dk/Faglige_informationer/Regnskaber/Love%20og%20bekendtgoerelser/Aarsregnskabsloven/Dataetik%20paa%20vej%20ind%20i%20virksomhedernes%20ledelsesberetning (accessed 1st April, 2019).
7. GDPR §20.
8. Rustici, C. (2018) 'Facebook is not done yet: New data questions every company valuation must address', *Forbes*, 27th August, available at: https://www.forbes.com/sites/chiararustici/2018/08/27/facebook-is-not-done-yet-new-data-questions-every-company-valuation-must-address/#63ca5cf31142 (accessed 4th March, 2019).
9. FSR, see ref. 6 above.
10. DI, see ref. 3 above.
11. Datatilsynet (2018) 'Vejledning om adfærdskodekser og certificeringsordninger [Guide for code of conducts and certification schemes', p. 3, available at: https://www.datatilsynet.dk/media/6566/adfaerdskodekser-og-certificeringsordninger.pdf (accessed 5th March, 2019).
12. The European Union Agency for Network and Information Security (ENISA) (2017) 'Recommendations on European data protection certification', November, p. 24, available at: https://www.enisa.europa.eu/publications/recommendations-on-european-data-protection-certification (accessed 5th March, 2019).
13. Prosa (2018) 'It-ordførere: Der skal styr på dataetikken [IT-spokesmen: we need to control data ethics]', 2nd November, available at: https://www.prosa.dk/artikel/it-ordfoerere-der-skal-styr-paa-dataetikken/ (accessed 1st April, 2019).
14. Laclau, E. and Mouffe, C. (1985) 'Hegemony and Socialist Strategy. Towards a Radical Democratic Politics'. Verso, London, p. 111.
15. Prosa, see ref. 12 above.
16. Ranger, J. (2019) 'Personal data and privacy predictions for 2019', 8th January, Digi.me, available at: https://blog.digi.me/2019/01/08/personal-data-and-privacy-predictions-for-2019/ (accessed 5th March, 2019).
17. ITU (2018) 'Powering the digital economy: Regulatory approaches to securing consumer privacy, trust and security', ITU Publications, København, p. 18, available at: https://www.itu.int/dms_pub/itu-d/opb/pref/D-PREF-BB.POW_ECO-2018-PDF-E.pdf (accessed 1st April, 2019).
18. Sitra (2019) 'The use of digital services', January, p. 23, available at: https://media.sitra.fi/2019/01/16142451/citizen-survey-digital-services-all-countries.pdf (accessed 5th March, 2019).
19. *Ibid*., p. 46.
20. Gartner (2018) 'Gartner identifies the top 10 strategic technology trends for 2019', 15th October, available at: https://www.gartner.com/en/newsroom/press-releases/2018-10-15-gartner-identifies-the-top-10-strategic-technology-trends-for-2019 (accessed 5th March, 2019).
21. Hennelund, M. (2019) '*Developing a Data Ethics Policy for the Insurance Industry in the New Data Economy*', on Nextwork.as blog, 6th February, available at: http://nextwork.as/blog/developing-a-data-ethics-policy-for-the-insurance-industry-in-the-new-data-economy/ (accessed 1st April, 2019).

Data classification: A means to an end

Received: 29th January, 2019

David King
is Regional Information Security Officer EMEA, for Omnicom Media Group, a leading global advertising, marketing and corporate communications company. During a previous spell at the company he held the position of Director of Security and Governance. He completed a master's degree in cybersecurity (2016), specialising in data classification and digital forensics for his dissertation and he remains passionate about managing data. Between those positions, David was Head of Research & Innovation for Secon Cyber, one of the UK's leading cybersecurity service providers and re-seller. David was responsible for thought leadership and developing Secon Cyber's Advanced Managed Security Services, including their Managed Detection and Response service (MDR), which was developed in-house. He regularly represented Secon Cyber, presenting on a variety of security issues, had numerous articles published and delivered several key note speeches. David is now based in Munich, has Chartered IT Professional status with the British Computer Society, is a Certified Information Security Manager (CISM), EU GDPR Practitioner Certified, and a member of both Data Management Association (DAMA) International and ISACA.

Omnicom Media Group Germany GmbH AG Düsseldorf HRB 34168, Germany
E-mail: david.king@omnicommediagroup.com

Abstract The volume of data continues to grow at an incredible rate, with some predicting as much as a 50× growth from 2010 to 2020 and yet the tools available to manage that growth haven't changed a great deal in recent years. With changes in regulation, such as the European Union's new General Data Protection Regulation (GDPR), knowing what data you have, where it is, what it is and how it is processed and then being able to apply the appropriate controls is becoming increasingly more important. This paper puts forward reasons why data discovery and data classification are two techniques that should be used to help manage data and the additional benefits that can be realised while moving organisations closer to compliance.

KEYWORDS: data classification, data discovery, compliance, GDPR

INTRODUCTION

Most organisations today still have some form of file server, or other electronic file repository, to store their data and information. Traditional documents and spreadsheets are stored either on the premises or within cloud-based services such as Microsoft OneDrive, Sharepoint, Salesforce or other customer relationship management (CRM) systems. Many organisations are starting to collect other types of data as well, such as telemetry from building information systems, Internet of Things (IoT) devices, logs from their devices and applications, biometrics, advertising cookies and so on.[1] The list goes on. In an article by Brian Krzanich, CEO Intel, it was stated that the average car will generate 4000 GB of data per hour of driving[2] and since there has been no slowdown in car sales, and these are ever more connected, it is easy to spot how some of these projections may actually be a bit on the conservative side.

Another factor fuelling this growth in data is the fact that many businesses continue to harvest data in the expectation that they will be able to gain some form of insight from it. While this may not

be allowed under certain regulations (eg Regulation (EU) 2016/679 (ie the General Data Protection Regulation —)), it is likely to continue as organisations try to take advantage of machine learning (ML) or artificial intelligence (AI). Some statistics put this growth rate as high as 10-times what it was in 2017 to 2025[3] and the general consensus is that this trend will continue for the foreseeable future. Action to manage this growth needs to be taken sooner rather than later, else it will get out of hand.

One of the tools or processes available to help manage this is data classification. This is not new and has been around in the UK since the late 19th century and even formed part of the Official Secrets Act 1889[4] entitled 'An Act to prevent the Disclosure of Official Documents and Information'. The principle or idea is therefore 130 years old, but few organisations have implemented it or implemented it effectively. Many, or most, companies will have an asset classification or data classification policy, yet few enforce it rigidly. There will be exceptions to this, of course, but they will be in the minority. Government, banks and large financial institutions, the military and related organisations all have mature data classification programmes, but there are too many companies who have not embraced the concept or the process. With the changes in regulation and the need to be able to respond to data requests (or subject access requests), data classification is likely to come to the fore.

DRIVERS

The biggest driver for adopting data classification is *regulation*. The EU GDPR was first introduced in April 2016 with the idea of harmonising data protection regulation across all 28-member states. It became fully enforceable on 25th May, 2018 and replaced the UK Data Protection Act 1998 within the UK. With the UK set to leave the EU at the end of March 2019 the Data Protection Act 2018,[5] which received royal ascent on 23rd May, 2018, will be in place and this implements and enhances the EU GDPR. Both have the capability to impose large fines for (personal) data breaches including the misuse of data. In the GDPR this is outlined in Article 5,[6] which summarises the six key principles for the processing of data. These are:

- Lawful, fair, transparent.
- Purpose limitation.
- Data minimisation.
- Accuracy.
- Storage limitation.
- Integrity and confidentiality.

There is also seventh significant principle contained within Article 5 and that revolves around *accountability*. This relates to the need to be able to demonstrate compliance, which will invariably mean maintaining more records and thus even more data.

Another driver for adoption is the need for *efficiency*. Being able to locate and identify relevant information quickly will be important in this new more heavily regulated era. Citizens, or natural persons, will have the right to query the data stored and to have that information modified if necessary to ensure it is up to data and accurate. They will also have the right to be forgotten under Article 17. Organisations that store a lot of information about people will need robust processes in place to be able to cater for these requests; however, a solid implementation of data classification will enable this and also realise other advantages.

BENEFITS

Article 5 states that you should only collect and store what you need to provide the service, a principle known as data minimisation. That data should also only be kept for as long as is strictly necessary, the principle of storage limitation. Data classification can be used to identify different

types of data and ensure that the information needed is kept for as long as it is required. Storing less data, and organising and filing it better, any automated operation or processing (such as access or indexing) is likely to be improved and *quicker*.

Tankard talks about the amount of unstructured data that organisations store on a network that could be considered as 'ROT-ten' (Redundant, Obsolete or Trivial).[7] In a Veritas report, they go as far as saying that up to 85 per cent of the data stored could be ROT-ten.[8] Reducing storage requirements by more than half is likely to have a significant *financial* impact too, with much *lower costs* in terms of both online and offline storage (tapes or other backup media).

Another advantage of reducing storage is that backup windows will be *shorter*, backups *quicker* and, more importantly, *recovery will be faster*. Reducing what is being stored will also *reduce the risk*. If data is not being stored, then it cannot be breached or stolen.

Security awareness is likely to improve following the successful implementation of a classification scheme. As an example, a common problem is around multi-function or departmental printers, where documents are printed, users get distracted and then forget to collect the print out. People seeing a printed document labelled TOP SECRET are likely to question why it has been left where it was, why it was visible, pick it up and give it to the appropriate person or remove it pending further action. This new security behaviour will often creep into other areas as people realise the impact and become more alert.

Making sure only the right, or specifically authorised, people have access to certain data is another function of a well-designed data classification process. This links back to the sixth point of Article 5, *integrity and confidentiality*. Ensuring that only those people (or service(s), if using automation) that need access to the data, have access to it. Being able to demonstrate who has access, how it was processed, through audit and/or other controls, will give clients (be that people, businesses or auditors) assurance that the right things are being done the right way at the right time.

Finally, having identified and classified the important data, the 'appropriate technical and organisational measures' can be implemented. This is another key phrase within the EU GDPR[6] and appears 18 times, with the word 'technical' appearing 40 times. This is significant as organisations will need to be able to clearly demonstrate that they have appropriate controls, tools, systems or processes in place. Obviously, this regulation is focused on EU citizen data, but what about intellectual property within a business? How can that be protected? That data first needs to be identified and labelled and an effective and robust discovery process can be a challenge.

DISCOVERY

In recent years, there has been a significant shift away from infrastructure on the premises, with organisations preferring to move anything and everything to the cloud. There has also been a move for people to work more flexibly, be that at various offices, client sites, internet cafes or at home, and this has meant more and more people are working remotely. This has led the mobile workforce to store documents on servers, in applications, in the cloud (and in applications in the cloud), on their laptops or even on their smartphones. As stated earlier, this is compounded by the amount of data being generated and stored, including e-mails and log files.

Cloud access security brokers (CASB)[9] solutions have already started to be used by many organisations trying to police their users and control access to cloud based services. Gartner™ coined the phrase back in 2012, and since then the technology has evolved and changed. CASB was really designed to address the following four key

areas but like many tools it can be used in other ways:

- visibility;
- compliance;
- data protection; and
- threat protection.

Visibility usually relates to seeing what the end users are doing from their devices. A side effect of this, however, is that it is also possible to see what data is being sent and where, who is using (and potentially sharing) it, where it is being stored (be that on the premises, local device or in the cloud) and what that data is. Combine this with a data loss prevention (DLP) tool and that makes for a very secure data control system. CASB can also help enforce compliance by ensuring corporate policies are implemented and adhered to, that is, if data is only supposed to be stored in certain locations, it will notify on attempts to breach the policy. Compliance also allows organisations to demonstrate and report on activities to show how policies are being consistently applied.

Data protection and threat protection can be used to reinforce data-centric security policies and to reinforce good user behaviour by alerting or blocking certain actions, while stopping users or devices from accessing services that should not be accessed. This can also include the monitoring of user and entity behavioural analysis (UEBA) providing the ability to flag unusual behaviour:

> Cloud access security brokers have become an essential element of any cloud security strategy, helping organizations govern the use of cloud and protect sensitive data in the cloud.[9]

Any data classification scheme implemented needs to go beyond just looking at file types and focus on the content of the file, regardless of file type (document, presentation, spreadsheet). It should pick up personal, health and financial information and, ideally, anything that is specific to your company or industry which could be considered intellectual property or market sensitive (such as mergers and acquisitions, research plans, product launch etc).

Having identified the most important or sensitive data, it will be much easier and significantly more cost effective to build the appropriate controls around it. Whatever data discovery tool is used, it should also allow a review of the users' access and rights and any other network controls. It should provide access to detailed audit logs and permission reports showing past and current activity so that informed decisions can be made, and any issues remediated. The benefits of role-based access control have been widely discussed and National Institute of Standards and Technology (NIST) have published various documents about it.[10,11] Roles and data relate closely to each other. Certain roles should only have access to certain applications, and thus, data. People should be assigned to roles so that as they move around an organisation, they only have access to the data or information that they should have or that they need.

IMPLEMENTING DATA CLASSIFICATION

There are two possible starting points when implementing data classification. First, either work out where your data is and what it is, or decide on your classification scheme and then find and label your data. Both have their advantages and disadvantages and it is more common to decide what you want your classification scheme to be and write your policy around that. Many companies adopt a four-level classification policy using terms such as public, internal, confidential and secret, while others might have sub-sets such as technical–internal and technical–public to differentiate what can or cannot be shared. Whatever scheme is chosen it should be clear, simple and easy to follow, with little room for ambiguity. In the UK,

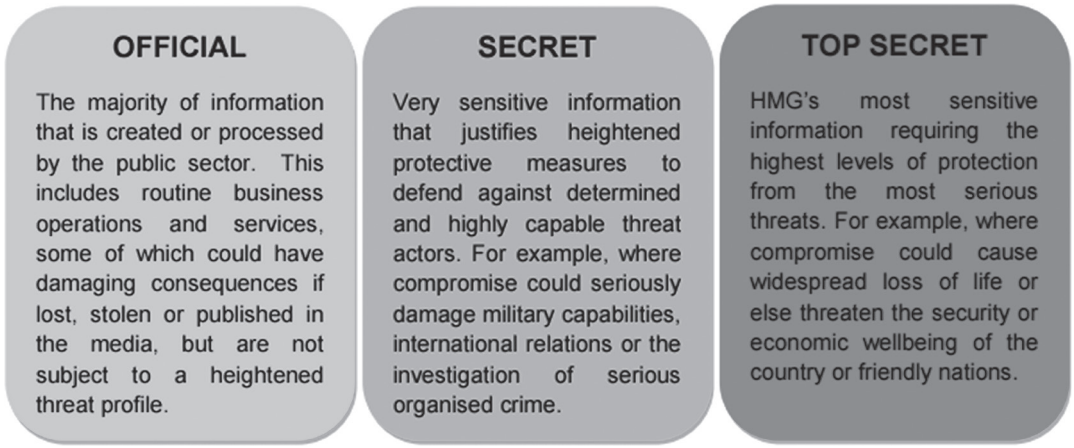


Figure 1: UK government's data classification policy
Note: HMG, Her Majesty's Government.

the government adopted three levels of classification — official, secret and top secret (Figure 1).[12]

Once the policy is designed, it needs to be communicated to everyone that deals with data and information. Everyone. There is not much point having written a policy and then expecting it to happen by osmosis. It needs to be communicated regularly and people need to be trained. One way of doing this is to install tools that prompt on creation, editing or saving of a file. That way all new electronic documents and any computer files that are touched manually are identified and/or labelled. Guidance and instructions can be provided at that point.

Different tools are likely to be required for retrospectively marking, labelling or identifying unstructured data. Equally, printed and stored information needs to be considered. Too often, existing paperwork gets ignored or forgotten. Any data classification policy needs to be both comprehensive and flexible enough to cater for all documents and information.

PLAN, DO, STUDY, ACT

First described by Shewhart[13] as *Plan, Do, Check, Act* and later modified by Deming,[14] this four-stage process perfectly captures the necessary steps for data classification. *Planning* covers much of the policy setting, tool choice, education and awareness planning. It should also cover 'how' you are going to 'do' the discovery. *Doing* the discovery, mentioned earlier, takes the longest time but leads to the greatest findings. It is highly likely that more data repositories will be uncovered using a CASB solution than initially anticipated. A study by Skyhigh networks in Q1 2015 stated that 'The average public-sector organization uses 742 cloud services'.[15] That was in 2015 and there are many more cloud-based services available today than there was then. These tools are designed to identify where files are being sent and where they are being stored. You also need to understand not just who is using it, but who has access to it and, most importantly, who is responsible for it. Knowing who the data owner is and why you have that data is important. Only the data owner can say what value the data has and how it should be classified.

Once identified (defensible) deletion of ROT-ten data should be undertaken for the reasons outlined earlier; however, you can only do that when you know 'what' that data is, that is, what it contains. As discussed, you should get rid of anything that is no longer required (ie ROT-ten

data). Defensible deletion means keeping records and this is important from a GDPR perspective. You should also keep a record of what has been deleted and why as one of the key principles within the GDPR is accountability. Again, you will need to rely on the data owner to tell you whether it needs to be kept or not.

With data reduced to the minimum, it should be possible to work with the data owners to decide which classification level should apply. This can be specific to a document or situation, or more generic. Either way, it should be clear what level classification applies to what document and why. While tools have been mentioned, more manual approaches can be taken. Headers and footers, watermarks and visible labelling, combined with a good user awareness and thorough education process can be used in place of tools and technology if scale allows. Nevertheless, tools are often capable of automating this and will frequently be part of the data discovery piece. The tools can often be programmed with the data classification scheme and/or prompt users when creating or saving documents.

STUDY

Once you have started your programme, it is important to ensure the classification scheme is both effective and being *adhered* to. It may highlight the need for additional or specific training. Having a good monitoring solution may also make it possible to spot any data breaches. Under the GDPR, any data breach will need to be notified to the relevant supervisory authority (the Information Commissioner's Office (ICO) in the UK) within 72 hours of being detected. Having a good monitoring solution will make this possible.

Some popular techniques to protect against a data breach include encryption and pseudonymisation and, if done properly, these can avoid you having to notify the end user (NB: you will still have to notify the ICO or supervisory authority).

Study can have a third meaning too. It is important not only to test the effectiveness of policies and detect breaches, but organisations also need to make sure that they are keeping up with regulative and business changes.

ACT

In another type of cycle this might be called 'review'. Under the continuous improvement cycle it is not acceptable to just plan, do and study. There needs to be a constant cycle of improvement as no process is ever 100 per cent right first time, despite all the planning. Also, *documents and data both age*. Things that were marked secret or top secret a year ago may now be public knowledge or publicly available (think mergers and acquisitions). In this case, why would it be necessary to apply overly stringent controls to protect data that is already in the public domain? Constant review and adjustment is required to ensure that the most appropriate controls are in place. You should also need to ensure the monitoring systems are still working effectively. Again, document everything.

CONCLUSION

Data classification is the simple process of organising and labelling documents, data and information. It helps organisations of all sizes understand what they have in terms of data and information and how much of that is important. It can help organisations identify data that is no longer required and can therefore be removed. This, in turn, helps improve operational efficiency and lower costs. A well-designed data classification scheme should also help companies identify data about people, in terms of the GDPR specifically EU citizens, so that the 'appropriate organisational and technical controls' can be implemented,

a key element in any GDPR project. Implementation does not have to be difficult. Plan what is required from your policy, discover what data is where and how much is important, label and mark everything, and monitor the effectiveness of the policy and implementation. Make changes where necessary.

A good data classification programme raises general security awareness, which reduces risk and improves organisational compliance. This in turn reduces the likelihood of a data breach, which can have significant financial impact.

References

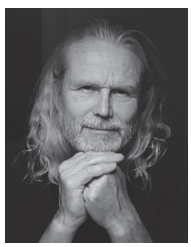
1. Preetish (2017) 'Want to ensure business growth via big data? Augment enterprise data with web data', March, available at: https://www.promptcloud.com/blog/want-to-ensure-business-growth-via-big-data-augment-enterprise-data-with-web-data/ (accessed July 2018).
2. Krzanich, B. (2016) 'Data is the new oil in the future of automated driving', November, available at: https://newsroom.intel.com/editorials/krzanich-the-future-of-automated-driving/#gs.CatF3EKn (accessed July 2018).
3. Ismail, N. (2017) 'The value of data: Forecast to grow 10-fold by 2025', April, available at: https://www.information-age.com/data-forecast-grow-10-fold-2025-123465538/ (accessed July 2018).
4. Official Secrets Act (1889) Available at: http://hansard.millbanksystems.com/commons/1888/may/10/official-secrets-bill (accessed 22nd August, 2016).
5. Data Protection Act (2018) Available at: http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted/data.htm (accessed July 2018).
6. Regulation (EU) 2016/679 of the European Parliament […] repealing Directive 95/46/EC (General Data Protection Regulation). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679 (accessed July 2018).
7. Tankard, C. (2015) 'Data classification — The foundation of information security', *Network Security*, Vol. 2015, No. 5, pp. 8–11.
8. Veritas (2016) 'Veritas global databerg report', March, available at: https://www.veritas.com/news-releases/2016-03-15-veritas-global-databerg-report-finds-85-percent-of-stored-data (accessed July 2018).
9. Riley, S. and Lawson, C. (2017) 'Magic quadrant for cloud access security brokers', November, Skyhigh Networks, available at: https://info.skyhighnetworks.com/WPGartnerMQ2018_BannerCloud-MFE.html?Source=Website&LSource=Website (accessed July 2018).
10. Ferraiolo, D. and Kuhn, R. (1992) 'Role-based access controls', Proceedings of the 15th National Computer Security Conference, Baltimore, MD, 13–16th October, pp. 554–563, available at: https://csrc.nist.gov/publications/detail/conference-paper/1992/10/13/role-based-access-controls (accessed July 2018).
11. Sandhu, R. Ferraiolo, D. and Kuhn, R. (2000) 'The NIST model for role-based access control: Towards a unified standard', Proceedings of the Fifth ACM Workshop on Role-Based Access Control, Berlin, 26–27th July, available at: https://www.nist.gov/publications/nist-model-role-based-access-control-towards-unified-standard (accessed July 2018).
12. UK Government Security Classification, May 2018, available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/715778/May-2018_Government-Security-Classifications-2.pdf (accessed July 2018).
13. McGrath, J. and Bates, B. (2013) 'The little book of big management theories ... and how to use them', Pearson, Harlow, UK.
14. Moen, R. (2009) 'Foundation and history of the PDSA cycle', September, available at: https://deming.org/uploads/paper/PDSA_History_Ron_Moen.pdf (accessed July 2018).
15. Skyhigh Networks (2015) 'Cloud adoption & risk in government report Q1 2015', available at: https://uploads.skyhighnetworks.com/2015/05/20162559/Skyhigh-Cloud-Report-Q1-2015-Government-0415.pdf (accessed July 2018).

Transparency, automated decision-making processes and personal profiling

Received: 13th January, 2019

Manuela Battaglini
is a data ethics lawyer, strategic marketer and CEO of the company Transparent Internet. She helps businesses and organisations to identify why, how and with which transparent and ethical technologies they can simultaneously increase market reach as well as client trust and benefits to society. Manuela initially worked as a lawyer for 10 years. In 2008, she engaged in the world of strategic digital marketing where she has, over the years, acquired deep knowledge about the inner working of online platforms and big data. Over recent years, she has combined these areas of expertise, as she understands the nexus of law, marketing and technology. Recently, she has founded a company, Transparent Internet, combining law, data ethics, marketing and technology to help companies and organisations develop, implement and use technology in a more transparent and ethical manner while at the same time increase their commercial edge. Manuela believes that if more organisations develop and implement technology in a more ethical and transparent manner, we will enhance personal freedom, trust, tolerance and sustainability. And this is the world she wants to live in.

Transparent Internet, Tårup Bygade 30, DK-5370 Mesinge, Denmark
E-mail: manuelabattaglini@gmail.com



Steen Rasmussen
is a professor in physics and a centre director at the University of Southern Denmark as well as an external research professor at the Santa Fe Institute in New Mexico, USA. Professor Rasmussen's research interests include the physics of living and intelligent processes, development of protocells, technology evolution, and the societal impact of living and intelligent technologies. He has co-authored 121 peer-reviewed scientific journal papers, edited 10+ books and proceedings, given 200+ invited talks and 130+ media interviews, and written 30+ consulting and internal reports. He has received many rewards for his work, starting in 1988 with P. Gorm-Petersens Mindelegat in the presence of Her Majesty the Queen, Margrethe II of Denmark, and most recently in 2018 with a Lifetime Achievement Award from the International Society for Artificial Life (ISAL). Since 2003, he has won US$39m in competitive research grants to his home institutions and international research consortia in the USA, EU and Denmark. Starting 2007, he has consulted on science and technology issues for the European Commission, the Danish Parliament, the German Reichstag, the US Congress as well as private organisations. He is co-founder of the companies BINC Technologies and Transparent Internet.

Center for Fundamental Living Technology (FLinT), University of Southern Denmark, DK-5230 Odense, Denmark
E-mail: steen@sdu.dk

Abstract Automated decision-making and profiling techniques provide tremendous opportunities to companies and organizations; however, they can also be harmful to individuals, because current laws and their interpretations neither provide data subjects with sufficient control over assessments made by automated decision-making processes nor with sufficient control over how these profiles are used. Initially, we briefly discuss how recent technological innovations led to big data analytics, which through machine learning algorithms can extract behaviours, preferences and feelings of individuals. This automatically generated knowledge can both form the basis for effective business decisions and result in discriminatory and biased perceptions of individuals' lives. We next observe how the consequences of this situation lead to lack of transparency in automated decision-making and profiling, and discuss the legal framework of this situation.

The concept of personal data in this section is crucial, as there is a conflict between the 29 Working Party and the European Court of Justice at the time to define the artificial intelligence (AI)-generated profiles and assessments as personal data. Depending on whether they are or are not personal data, individuals have the right to be notified (Articles 13–14 GDPR) or right to access (Article 15 GDPR) to inferenced data. The reality is that the data protection law does not protect data subjects from the assessments that companies make through big data and machine learning algorithms, as users lose control over their personal data and do not have any mechanism to protect themselves from this profiling owing to trade secrets and intellectual property rights. Finally, we discuss four possible solutions to lack of transparency in automated inferences. We explore the impact of a variety of approaches ranging from use of open source algorithms to only collecting anonymous data, and we show how these approaches, to varying degrees, protect individuals as well as let them control their personal data. Based on that, we conclude by outlining the requirements for a desirable governance model of our critical digital infrastructures.

KEYWORDS: machine learning, transparency, GDPR, data ethics, open source, digital infrastructure

BACKGROUND

Why are we currently in a situation where privacy and lack of transparency have become central legal issues? Obviously, it is because of rapid technological development, but it is perhaps useful for our discussion about transparency, privacy and profiling to dig a bit deeper. By understanding a bit more about *how* technology has changed our world so radically in recent years, we shall argue, we are in a better position to provide legal solutions. We shall therefore briefly review two major technological circumstances that made this transformation possible; one is about computer hardware development and the other is about the development of computer software that enables many computers to work as one. Further, we shall briefly discuss what these developments meant for organisations and business.

Moore's law

Gordon Moore is Intel's co-founder and is better known for a prediction he made in 1965 in an article he wrote for *Electronics* magazine with the title 'Cramming More Components onto Integrated Circuits.'[1] He stated:

> The complexity for minimum component costs has increased at a rate of roughly a factor of two per year. Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years.

In 1975, in an updated article, Moore adapted his prediction to 24 months.[2] Finally, the prediction was quoted as 18 months as the doubling period for general computing power.

Moore's law, in part, explains the sustained exponential growth in the big data era. It implies ever-expanding huge numbers and is explained by Ray Kurzweil through the story of the inventor of chess in India. When the inventor of chess presented his game to the emperor,[3] the emperor was very impressed by the game, and he asked the inventor to ask for any reward he

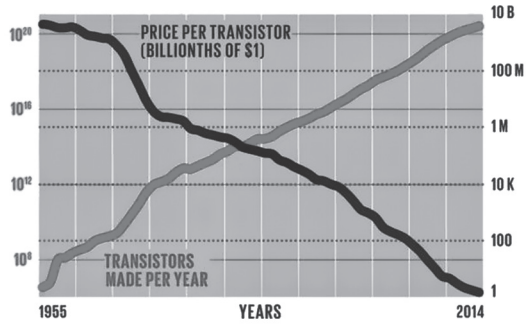


Figure 1: Moore's law shows exponential growth in transistors as a doubling approximately every 18 months, and how the price of transistors is falling every 18 months. Note the logarithmic axis. (See Hutcheson, D. (2015) 'Graphic: transistor production has reached astronomical scales', available at: https://spectrum.ieee.org/computing/hardware/transistor-production-has-reached-astronomical-scales (accessed 12th December, 2018)). Resource: VLSI Research.

wanted. The inventor only wanted rice to feed his family and he used the chessboard to show the amount of rice he would like. He put one grain of rice in the first square of the chessboard, two in the second square, four in the third one, eight in the fourth square, and repeated this process until the last square of the chessboard was filled with rice grains.

On the first part of the chessboard, the human brain can imagine the number of rice grains, but on the latter part of the chessboard the numbers become too big to imagine: trillions, quadrillions and quintillions. When this action is repeated until the last square of the chessboard, more than a quintillion grains of rice is obtained. It is more rice than has been produced in the history of the world.

Moore's law was formulated in 1965 and close to 18 months was predicted as the doubling time for transistors in use (see Figure 1). After 32 of these doublings, since 1965, we are now on the second half of the chessboard. From this point on, we were able to digitise almost everything and the immense numbers of computers enabled us to store all of these new data. But, there was a challenge: how could we access and manipulate data stored across many different computers? We needed 'the cloud' and this is the topic of the next section.

How big data analytics was created

The era of big data computing started in 2007, when it became widely possible to 'upload data to the cloud', because effective shared memory software became available, so that thousands of computers could work as one.

In 2003, Google published a paper that included a basic innovation called the Google File System (GFS).[5] This software allowed Google to access and manage a huge amount of data from thousands of computers. At this time, Google's main goal was to organise all the world's information through its search engine; however, they were not able do that without their second basic innovation, MapReduce,[6] which was published in 2004. These two innovations allowed Google to process and explore a huge quantity of data in a manageable way.

Google shared these two basic innovations with the Open Source community, so that the community could build on their insights. Even better, the community was able to improve the software and as a result Hadoop[7] was created in 2006. Hadoop is an open source software that allows hundreds of thousands of computers to work as one giant computer.

Facebook, LinkedIn and Twitter[8] already existed in 2006, and they started building on Hadoop straightaway. This is the reason why these platforms companies became global in 2007.

With Hadoop, easily accessible storage capacity for computing exploded making 'big data' available for all. Thanks to Hadoop, internet platforms could store all their data across many computers while still having access to their data. Furthermore, they could store every click of every user on every web page. This gave them a much better understanding what users were doing

over time, thus providing the basis for big data analytics.

Thanks to Hadoop, other companies were born in 2007, including Airbnb; Amazon also launched Kindle and the first iPhone was released. According to AT&T,[9] mobile data traffic on its national wireless network increased by more than 100,000 per cent from January 2007 to December 2014.

Differences between platform companies and traditional companies

The year 2007 was a crucial point in the global economy. This paved the way for the emergence of a new category of companies that reshaped how people and machines communicate, create, collaborate and think.

From 2007, we can distinguish between three different types of companies:[10]

(1) Traditional companies: companies that are owned by their shareholders and that perform most of their operations by themselves or through contractors.
(2) Distributed platform companies: companies such as Facebook, Uber, Airbnb, Google and so on that enable individuals and organisations to interact by proving a common transaction platform.
(3) Cooperative organisations: organisations such as Wikipedia or open source software projects that enable individuals to cooperate and efficiently achieve their eudemonic goals, without necessarily making a significant profit while doing so.

Since 2007, the distributed platform companies, through big data analytics, have had the opportunity to store all their data in one place and thus have a greater in-depth knowledge of the market than traditional companies. Furthermore, more customers on one platform means better service (eg social media or Airbnb), which favours larger platforms that over a few years can act as monopolies owing to their market dominance.

The main consequence for users was the benefit of a number of new services, but at the same time a total loss of control of their personal data and the possibility of being analysed and profiled thanks to big data analytics and machine learning algorithms.

This means that company decisions started to be made via automated decision-making processes through profiling of individuals and groups. In some cases, this was for advertising purposes; in other cases, these automated decisions could become life-changing owing to biased results and discrimination. In other cases, however, owing to lack of privacy and through market domination, these automated decision-making processes can distort fair markets as well as fair elections, as we shall discuss later.

In the next section, the legal approach and the lack of transparency of automated decision-making processes are examined, and the consequences for individuals in terms of loss of control over their personal data are evaluated.

LACK OF TRANSPARENCY IN AUTOMATED DECISION-MAKING AND PROFILING

Why third party predictions can harm our freedom and private life

We live in a big data analytics era where automated decision-making, machine learning and profiling techniques increasingly are able to make assessments of individual's lives based on historical data and predictions. These technological capabilities generate great opportunities and crucial challenges for our society as well as for the lives of individuals. We shall focus on the challenges that can affect our privacy, personal autonomy and freedom.

As we shall argue, automated decision-making, machine learning and profiling techniques can be harmful to

individuals, because current laws and their interpretations neither provide a data subject with sufficient control over the assessments made by automated decision-making processes nor with sufficient control over how these profiles are used.

There are many examples of assessments being made based on online automated decision-making processes. For example, Facebook can predict your political views,[11] your race, religion and sexual orientation,[12] and even predict when you are going to die.[13] Facebook can predict individual future behaviours, allowing third parties to target these individuals with advertisements that can change their decisions entirely. Facebook calls it 'improved marketing efficiency'.[14] Another example is given by Amazon's 'Alexa Hunches' feature and its capacity to predict future needs based on a user's behaviour to make suggestions,[15] and furthermore, predict a user's health status through analysing voice and coughing, which is followed by sending advertisements for sore throat products.[16] Insurance companies are also collecting data from social networks to predict how much users' healthcare could cost them.[17]

In an era of automated decision-making processes, the central question is if a data subject has real control over her personal data and whether the General Data Protection Regulation (GDPR) protects a data subject from their inherent risks. This question is examined in the next section.

What is the legal framework for automated decision-making and profiling according to the GDPR?

Article 29 of Working Party in 'Guidelines on automated individual decision-making and profiling' states that 'Automated decisions can be based on any type of data'[18] and makes a three-part differentiation:

1. Data provided directly by the individuals concerned (such as responses to a questionnaire);
2. Data observed about the individuals (such as location data collected via an application);
3. Derived or inferred data such as a profile of the individual that has already been created (e.g. a credit score).

The Article 29 Working Party Opinion 4/2007[19] on the concept of personal data has a broad definition. It states that personal information is 'any information' 'relating to' an 'identified or identifiable natural person'. 'It covers "objective" information, such as the presence of a certain substance in one's blood. It also includes "subjective" information, opinions or assessments'.

This is a very broad definition of personal data, on which almost everything is personal data, including assessments or inferred data.

What does the European Court of Justice (ECJ) say?

In an interesting paper, Watcher and Mittelstadt[20] named two law cases from the ECJ where the opinion of the court about whether inferred data are, or are not, personal data is reflected: Cases C-141/12 and C-372/12 on 17th July, 2014 and Case C-434/16 Nowak on 20th December, 2017.

Regarding cases C-141/12 and C-372/12[21] on 17th July, 2014: a person from a third country applied for a legal residence in a European country and the residence was denied. This person requested access to the assessment of the legal analysis that was written about him to find out why his residence was denied.

The court said that: 'the legal analysis contained in a minute … although it may contain personal data, it does not in itself constitute such data'[22] According to this definition, a legal analysis is not personal data, only facts can be considered personal data; that is, input data and not assessments.

With regard to the right to access, the court said: 'Regulation No 45/2001 is not designed to ensure the greatest possible

transparency of the decision-making process'.[23]

This person wanted to use his right of rectification of personal data and the court said: 'such an analysis … is not in itself liable to be the subject of a check of its accuracy by that applicant and a rectification under Article 12(b) of Directive 95/46'.[24]

In other words, the court said that the individual was not competent to decide if the assessments were accurate or not.

In this case, assessments and data used to make a profile are not personal data, and none of the rights that a data subject has under data protection law will apply.

Case C-434/16 Nowak[25] on 20th December, 2017 involved a candidate who had sat an exam and failed and they wanted to get access to the comments and assessments of the examiner, using the right to rectification. This case provides a description of what personal data is, referring to both objective and subjective information.

The court said, in general, one has the right to rectify the comments, but not the content. In this case, the right to rectification must be interpreted according to the purpose for which the data was collected.[26]

The data subject can use this right only to make sure that the exam script was complete, but the data subject has no right to assess if the reviewer's comments were accurate or not. Again, only input data are considered personal data.[27]

As a result, there is a conflict between the WP29 in its definition of personal data, because it says that all rights will apply because the scope of data protection law makes decision-making more transparent, and the European Court of Justice is unsure about if assessments and opinions are personal data.

Even if they are personal data, it does not mean you have full access to them, and is not possible to rectify any of the assessments.

Data subject rights to transparency are described in Articles 13–15 GDPR. The right to be notified (Articles 13-14 GDPR),[28] is a data controller's duty and covers data provided directly by the data subject, observed data and data from a third party. Also, the right to access (Article 15 GDPR) has to be appealed for by the data subject.

Watcher asks the following question:[29] do Articles 13 and 14 of the GDPR allow the data subject to be notified about the automated decision-making and profiles that companies have about them? And, if the answer to this question is positive, is it an *ex ante* explanation, or does it cover an *ex post* explanation too?

Article 13.2 (f) GDPR says about notification requirements when personal data is collected directly from the data subject:

> the controller shall, at the time when personal data are obtained, provide the data subject with the following further information necessary to ensure fair and transparent processing:
>
> f) the existence of automated decision-making, including profiling, referred to in Article 22 (1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

Article 14.2 (g) GDPR says about notification requirements when personal data is obtained from a third-party:

> the controller shall provide the data subject with the following information necessary to ensure fair and transparent processing in respect of the data subject:
>
> g) the existence of automated decision-making, including profiling, referred to in Article 22 (1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

It has been suggested that the notification duties outlined in these two articles grant an *ex post* explanation of 'the existence of…

logic involved…as well as the significance and the envisaged consequences' of automated decision-making.

But, this suggestion is wrong for a reason.

Only an *ex ante* explanation of system functionality is explicitly required by Articles 13.2 (f) and 14.2 (g). These notification duties precede decision-making and apply in the moment data is collected for processing and refer only to input data.[30]

Is the right to access a right to an explanation in the GDPR?

Article 15.1 (h) is identical to Articles 13.2 (f) and 14.2 (h) in the GDPR. It says that individuals have the right to access to their personal data and to the following information:

> The existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

The phrasing of Article 15.1 (h) is future-oriented and we can conclude that the data subject can ask for this information at any time, and this includes once automated decision-making has been made.

But the phrase 'envisaged consequences' suggests that the data controller has to give an explanation to the data subject about the consequences of the automated decision-making before the processing of the data. And with a lack of an explicit deadline for appealing, the right of access is limited to explanations of systems functionalities. This is, again, an *ex ante* explanation.

The phrase 'the existence of automated decision-making, including profiling' does not refer to an explanation of how the decision was made, it only informs the data subject that automated decision-making, including profiling, was used to process their data.

WP29 seems to align with Wachter when she says 'the controller should provide the data subject with information about the envisaged consequences of the processing, rather than an explanation of a particular decision' and 'The GDPR requires the controller to provide meaningful information about the logic involved, not necessarily a complex explanation of the algorithms used or disclosure of the full algorithm.'[31]

For Wacther, the reasonable question about inferences is not just linked to the GDPR. Reasonableness should be extracted from a mixture of data protection rights and the specific sectorial laws, as well as the potential risks involved; however, reasonableness has nothing to do with the inference result itself. It is not about whether the outcome is unreasonable or not. It is about having an *ex ante* justification and knowing what inferences the company or organisation are making and their purpose in doing that. It is about moving away from the output situation where the data subject can just obtain an explanation. It is about a justification before data processing begins about what the company wants to do.

Right to access is limited further by the definition of automated decision-making that Article 22.1 of the GDPR states:

> The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

Article 22.1 of the GDPR is limited to 'legal or similarly effects' produced by 'solely' automated decision-making, including profiling.

What does 'solely' mean? Its definition is crucial in the application of the rights of a data subject. Does a human intervention in the decision-making mean that it is not

automated decision-making? There is no answer to this question.

Article 29 Data Protection Working Party, 'Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679' states: 'if someone routinely applies automatically generated profiles to individuals without any actual influence on the result, this would still be a decision based solely on automated processing' and adds 'the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision'.[32]

Article 22.1 of the GDPR is limited to 'solely' automated decision-making processes. Those automated decision-making processes that do not fit in with this definition are not under the notification duties of data controllers as defined in Articles 13.2 (f) and 14.2 (g) of the GDPR and the right to access in Article 15.1 (h) of the GDPR.

What do 'legal effects' and 'similarly significant effects' mean in Article 22.1 of the GDPR? Recital 71 is more explicit and gives two examples, such as: 'automatic refusal of an online credit application or e-recruiting practices without any human intervention.'[33]

But does the data subject have the right to be hired by a company after a job interview or have the right to have a loan or a credit approved? In general, they do not, therefore these situations do not fall under a 'legal effect'.

And as for 'similarly significantly affects' — it is very hard for the data subject to prove that the processing and decisions based on automated decision-making affects them significantly.

Many automated systems now commonly make decisions both in public and in the private sector, such as criminal justice, welfare, taxation, search engine, marketing, entertainment and political opinion-making. Much concern has been raised in the legal system and by policymakers as to whether such systems create discriminatory, biased or unfair results, and it is hard to access how many of these decisions are made 'solely' by algorithms because a human intervention is almost always necessary.

Nevertheless, there is another constraining factor to the right to access and the right to an explanation by the data controller. This factor is trade secrets and intellectual property rights, established in Recital 63 of the GDPR:

> Where possible, the controller should be able to provide remote access to a secure system, which would provide the data subject with direct access to his or her personal data. That right should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software.

Further, there is a new Trade Secrets Directive[34] and its Article 2.1 explains what a trade secret is:

> (a) it is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question;
> (b) it has commercial value because it is secret;
> (c) it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret.

It means that a trade secret is everything that is not known, it is anything that has commercial value and is anything where reasonable steps are taken to keep it secret.

As a direct consequence of this definition, a data subject does not have the right to be notified, as Article 13.2 (f) and 14.2 (g) of the GDPR considers, nor the right of access

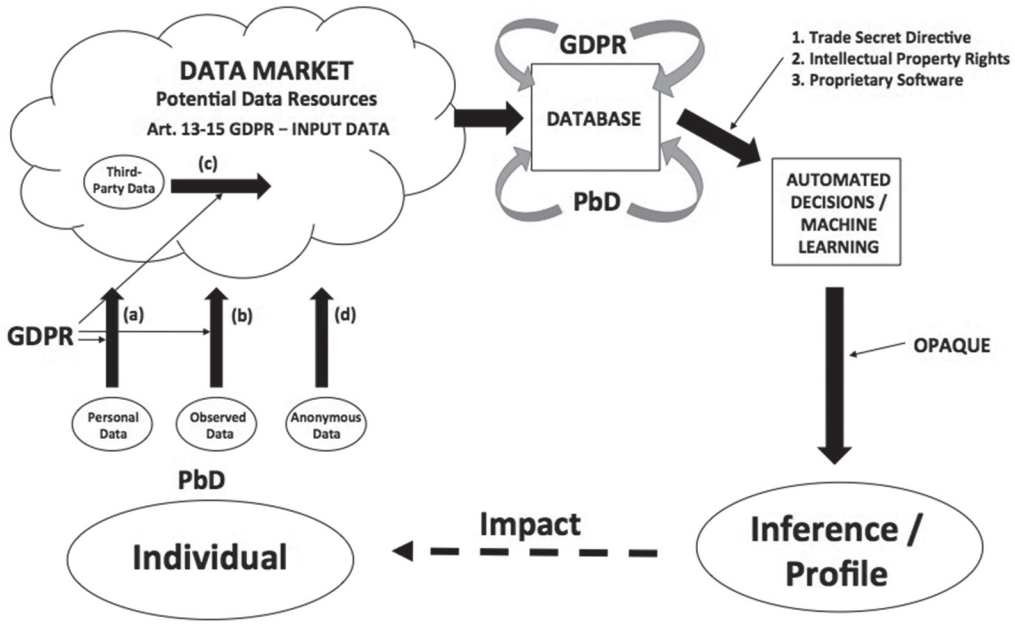


Figure 2: Lack of transparency in automated decision-making processes and profiling. The GDPR applies to a data subject in cases (a), (b) and (c), while trade secret directive, intellectual property rights (IPR), and proprietary software in practice prevent a data subject from having access to their own database information, how automated decisions are made about them as well as their profile (inference)

established in Article 15.1 (h) of the GDPR, to the profile that the companies have about them and for what purpose is used.

In other words, the GDPR does not contemplate in an explicit way a right to an explanation to the data subjected to automated decision-making, and this makes the principles of transparency and accountability impossible, as is illustrated in Figure 2.

POSSIBLE SOLUTIONS

Can open source software solve the problem of lack of transparency in automated decision-making processes and profiling?

The question of explainability to achieve transparency in how machine learning (ML) algorithms work has been debated both in the legal and in the ML community.

The technical problem is controversial, because the ML source codes are very complex and usually only the software engineers who have designed the systems understand how they work. Even that is sometimes not the case, as it is still an open scientific question, for example, as to how and why deep artificial neural networks work so well. Further, even if these ML systems could be made explainable, an explanation would most likely be neither useful nor make sense to non-expert individuals.

The legal problem is that there is a significant resistance from companies to disclose how their algorithms work because of trade secrets and intellectual property rights.

All the arguments around openness of the algorithms are understandable; however, openness about the ML algorithm codes is not enough, because of the connection between the algorithm and its training set. This means that we cannot reproduce how an ML algorithm works unless we also know how it is trained. So, the actual training set together with an algorithm determines how a final decision is made about a data subject.

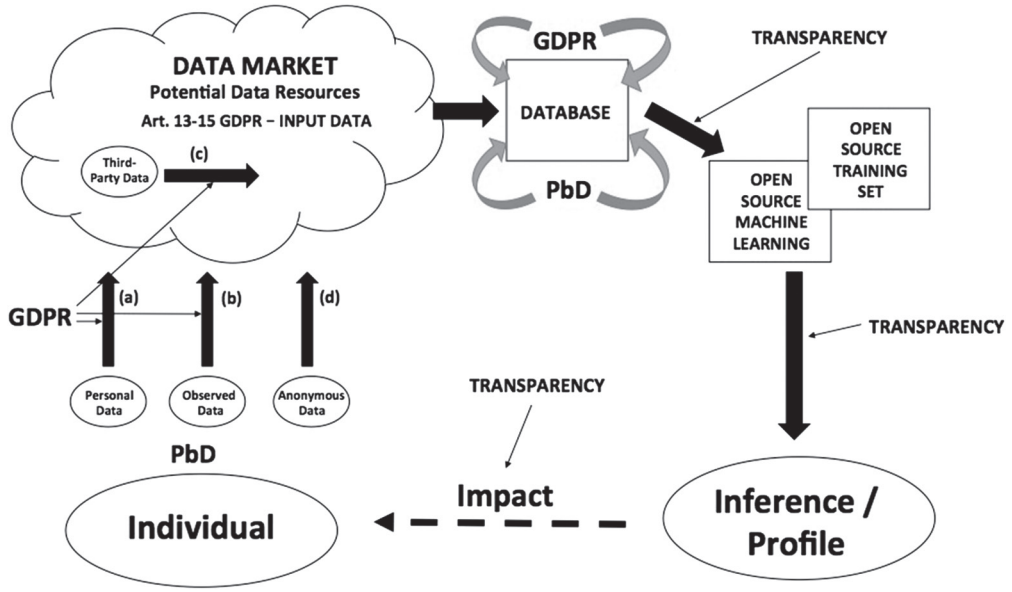


Figure 3: Transparency in automated decision-making processes through open source training sets and open source algorithms. Note that despite transparency, experts are needed to interpret both how the profiles are generated and what impacts the profiles might have.

Thus, in principle, if both the code and the training set could be open source, it should be possible for independent experts to reproduce any decision. But still that would be unlikely to be helpful for most individuals owing to the inherent complexity of the system involved.

These challenges, however, should not prevent the requirement for open source access to automated decision-making processes in many cases. For example, Epstein and Robinson,[35] showed that a monopoly or a dominating search engine could determine the outcome of a democratic election just by sorting 'good' information for one candidate to the top of a search, while sorting 'bad' information to the top of a search for other candidates.

Today, in early 2019, a dominating search engine company could make such algorithm adjustments at will, because these search algorithms are proprietary and secret. Therefore, open source is necessary as part of the solution to deal with platform dominance or monopolies in the new online economy. The fair market breakdown that is caused by this kind of platform domination is due to the growing technology-induced, so-called mandatory participation third party payer business model. This business model is used, for example, by Google, Amazon, Facebook and Airbnb;[36] recall our discussion in the first section of this paper.

In Figure 2, we showed that access to personal data stored in databases as well as algorithms and their inference could, but do not need to be, transparent. This could be made voluntarily by the companies, however, companies can choose to be opaque and decide not to be transparent by using the trade secrets and intellectual property rights as a legal defence against this argument.

Transparency is crucial if an algorithm is a dominating or central part of an individual's environment, for example, as in our basic information technology infrastructure, and if the automated decisions have a critical impact on an individual's life. An obvious way to enable democratic control over dominating global platform companies is by making their algorithms and training sets open source, as shown in Figure 3. This

enables independent experts to check them and reveal their real impact on our society in general and on individuals in particular.

It should be noted that transparency does not imply the inferred profiles make sense or are correct. Further, transparency not either means explainability. Experts are needed to interpret both how the profiles are made and what possible impacts these profiles might have. Finally, even with open source training sets and algorithms, we cannot know for what purpose these profiles will be used and with whom they will be shared. Therefore, transparency should be focused on the folloing key points: individuals need to be aware of the fact that they are being profiled in a certain way; the context within which the profile is used; the companies with whom these inferences are shared; and whether or not the profile can lead to significant decisions. This is what data protection requires companies and organisations to be transparent about.

Is data protection by design and by default, in the context of the GDPR, the solution to protect data subjects from being profiled?

The GDPR, Article 25 mentions two instances where companies have to think about data protection, and these are: 'at the time of the determination of the means for processing' and 'at the time of the processing itself'.[37]

This means that at the time where engineers are thinking about developing the products, or establishing a particular business process, companies (engineers) already must think about privacy and have a plan for how to protect it.

Privacy by design establishes a privacy that is proactive, a privacy that reduces risks. It is a privacy that is embedded in whatever an organisation is doing, both in terms of building products and in business operations.[38]

Article 25 of the GDPR continues by saying that the key requirements for organisations are to adopt privacy measures that are both organisational and technical.

In privacy by design, two techniques are explored below: pseudonymisation and anonymisation of private data.

Data pseudonymisation

Recital 26 equates data that has undergone pseudonymisation to personal data.[39] Why do we need to pseudonymise if pseudonymous data are still personal data? Answers to this question can be found in Recital 28 and Recital 29.

Recital 28 says pseudonimisation 'reduce the risks to the data subjects concerned and help controllers and processors to meet their data protection obligations'. Because pseudonomisation is a prevention measure, it helps in compliance with the GDPR.

And Recital 29 says 'In order to create incentives to apply pseudonymisation when processing personal data, measures of pseudonymisation should be used, whilst allowing general analysis'.

With pesudonymisation strategies, data subject IDs are replaced by a pseudonym (hash of the user ID). This means that companies do not know which user carried out a specific action, but a company can know that an individual did A, B, C and D. If a static hash is used, a company can know about an individual's action for an extended period of time, which make it easier for companies to make business decisions, and as is shown in Figure 4, which implies less privacy because companies can still profile users.

Data anonymisation

Anonymised datasets are more secure than pseudonymised datasets as there is no direct way to recover the identity of an individual. How can anonymised data be created and validated? There are multiple anonymisation models/schemes that are sufficient according to regulations.

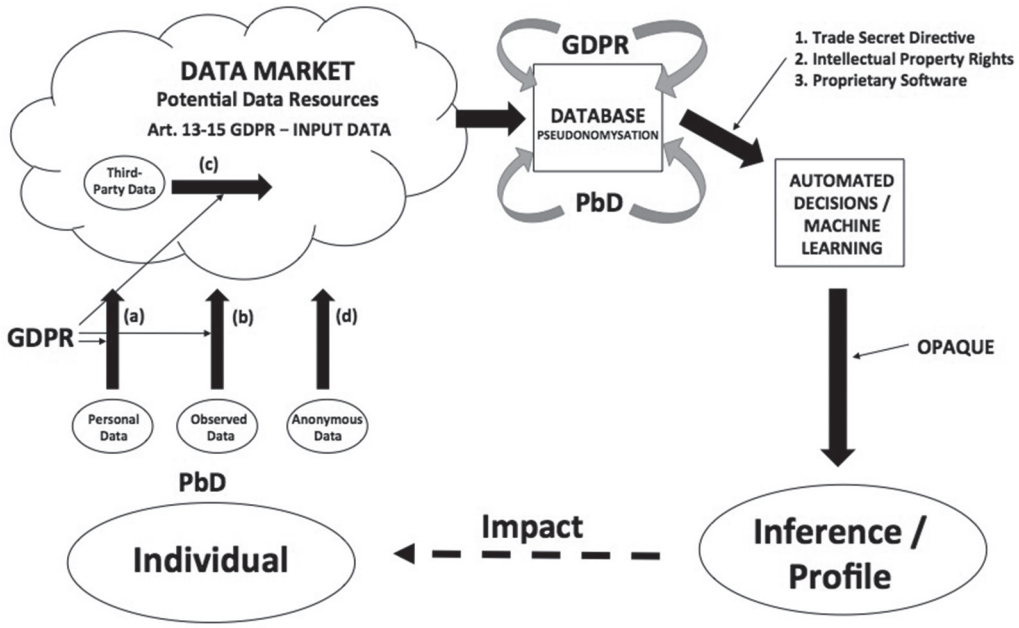


Figure 4: Pseudonymisation of personal data in a database. Some protection is provided for personal data in the database although personal profiling is still possible.

The WP29 Opinion 05/2014[40] gives two options to check whether a dataset is anonymised:

> Option 1: Your Dataset has *none* of the following properties:
> - Singling out, which corresponds to the possibility to isolate some or all records, which identify an individual in the dataset.
> - Linkability, which is the ability to link, at least, two records concerning the same data subject or a group of data subjects (either in the same database or in two different databases). If an attacker can establish (eg by means of correlation analysis) that two records are assigned to a same group of individuals but cannot single out individuals in this group, the technique provides resistance against 'singling out' but not against linkability.
> - Inference, which is the possibility to deduce, with significant probability, the value of an attribute from the values of a set of other attributes.
>
> OR
> Option 2: Perform a re-identification risk analysis.

The first option is much stronger than the second option in a sense that in the first option you have to show that you can prevent what is called attribute inferences. Companies have to show that they are not able to infer any attribute from the individuals that are part of the dataset. Figure 5 shows how the anonymisation must be done in the datasets, so that the training set is fed by anonymised data.

In the second option, companies are only concerned with re-identification, which is also called identity inferences. You have to show that you are not able to recover the identity of the individuals that are part of the anonymised dataset.

Nevertheless, re-identification is not the only problem, because many things can be learned from an anonymised dataset, even if individuals cannot be identified. Big data is currently actively being used for that.

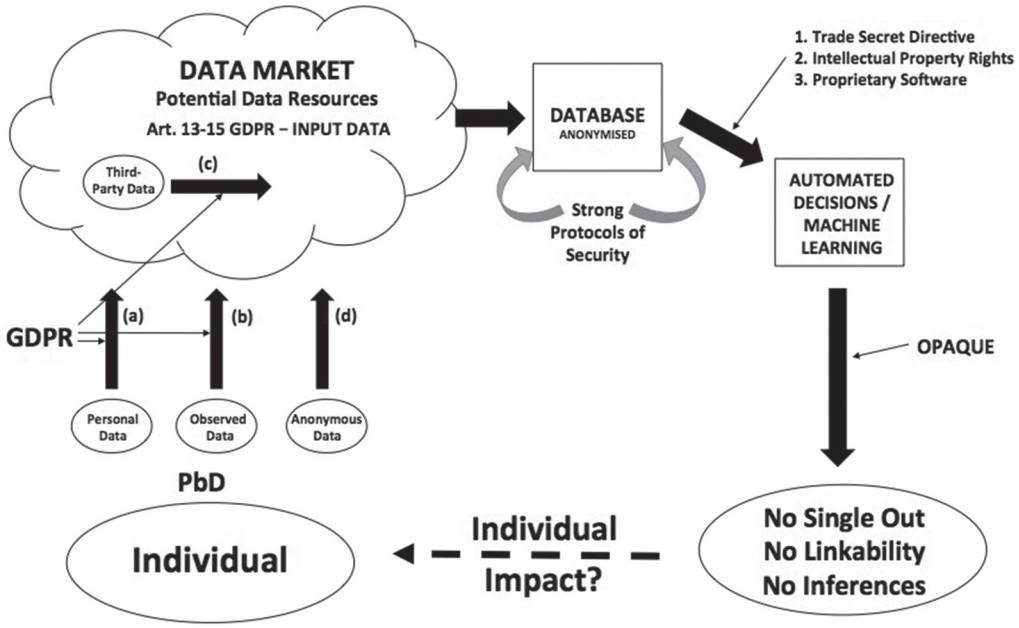


Figure 5: Anonymisation of personal data in a database. In this case, profiling is not possible, but this method still does not address the issue of lack of transparency in automated decision-making

The main issue is attribute inferences. Privacy advocates focus on unique identification as the main attack on our privacy. Data minimisation and anonymisation are strategies to avoid being singled out.

Companies often tell users that sharing their data is safe because they 'anonymise' the data by first removing or obfuscating the personal information; however, this depersonalisation leads to only *partial anonymity*, as companies still usually store and share data grouped together. This data group can be analysed, and in many cases, then linked back to the user *identity* based on its content.

Data de-anonymisation of this nature has taken place time and time again when companies release so-called 'anonymised data', even with good intentions, for example, for research purposes.[41] For instance, even though efforts were taken to anonymise data, individuals were still de-anonymised through Netflix recommendations[42] and AOL search histories.[43]

Collect only anonymous data and use only open source software

As we saw in the former section, companies and institutions can anonymise their datasets, but this is not enough in terms of privacy, transparency and to ensure control over personal data. In this section, we consider a third possibility by *collecting only anonymous data* as a way of empowering individuals and protecting their privacy and confidentiality.

Anonymous data are not connected to information that can identify an individual; however, this is not about 'just' collecting anonymous data. In order to protect data anonymity, companies and organisations must take further security measures so datasets cannot be linked and users cannot be de-anonymised.[44] Un-linkability is crucial for disabling cross-contextual aggregation of individual profiles, for example, by using credentials or attributes instead of full identification. Other security measures also have to be taken like best practices, including encryption and technical measures at the organisational level.

Combining this with open source automated decision-making processes (ADM) together with transparency about the use of the obtained ADM conclusions provide the highest degree of transparency and privacy protection.

Anonymous and purpose-defined data together with open source software

Complete anonymity is neither possible nor desirable for all types of cyberspace interactions. You want your doctor to know who you are to be able to help you with health issues. Utility companies need to know the addresses to whom they deliver power and water, and for most private person-to-person communications there is a desire in both ends to be certain about the identity of each other. Also, cybercrime is difficult to manage in an anonymous cyberspace.

Do solutions exist that address the above? A solution should use anonymous data as much as possible, minimising the use of personal data, as well as transmit personal data encrypted. Minimise the use of personal data is the intent of the GDPR and it can be obtained e.g. by creating purpose-determined data for each cyberspace process, as suggested by e.g. Engberg.[45] The content of the purpose-determined data can either be anonymous or identifiable dependent of the purpose. Thus, each new process in cyberspace would have a new identifier dependent on whether we communicate for e-commerce, health, banking, private conversations, etc. E.g. if we want to provide some of our health data or other relevant data for a research project, we could do that anonymously. In contrast, some of these same data should be identifiable if it concerns an ongoing treatment by our doctor.

Implementing software systems that enables the creation of purpose-determined data sessions on top of our current digital infrastructure is highly recommendable as it would either eliminate or reduce most of our current complex online personal data protection and security issues. It could be both developed and implemented piece-meal e.g. one sector or group of individuals at the time, and it could then grow organically to include more sectors and groups.

We discuss anonymous and purpose-defined data in Figure 6. Only collecting anonymous or purpose-determined data reduce the data privacy issues significantly. Adding strong data protection protocols (e.g. anonymised databases) for purpose-determined data (when necessary) eliminates identification of individuals as well as tracking of their presence on the internet. Also, inferences, linkability and single-out are not possible. Combining this with open source, automated decision-making processes (ADM) together with transparency about the use of the obtained ADM decisions provide the highest degree of transparency.

The solution in Figure 6 is an option for companies and organizations to pursue, if they want to demonstrate honesty, transparency and maximize privacy protection in their data collection, processing and profiling decisions.

Please note, "No solution fits all". Automated decision-making (ADM) should not be made public for certain critical infrastructures, e.g. automated coordination of metro traffic or power distribution. Complete openness could expose potential infrastructure vulnerabilities for misuse. In such situations independent experts, under democratic oversight, should have access to the ADM, while ADM details should be kept out of the public eye.

Individuals 'take home the data cloud' and stay anonymous

Currently, most transactions in cyberspace are based on identification of individuals. For example, relay-on trusted third parties,

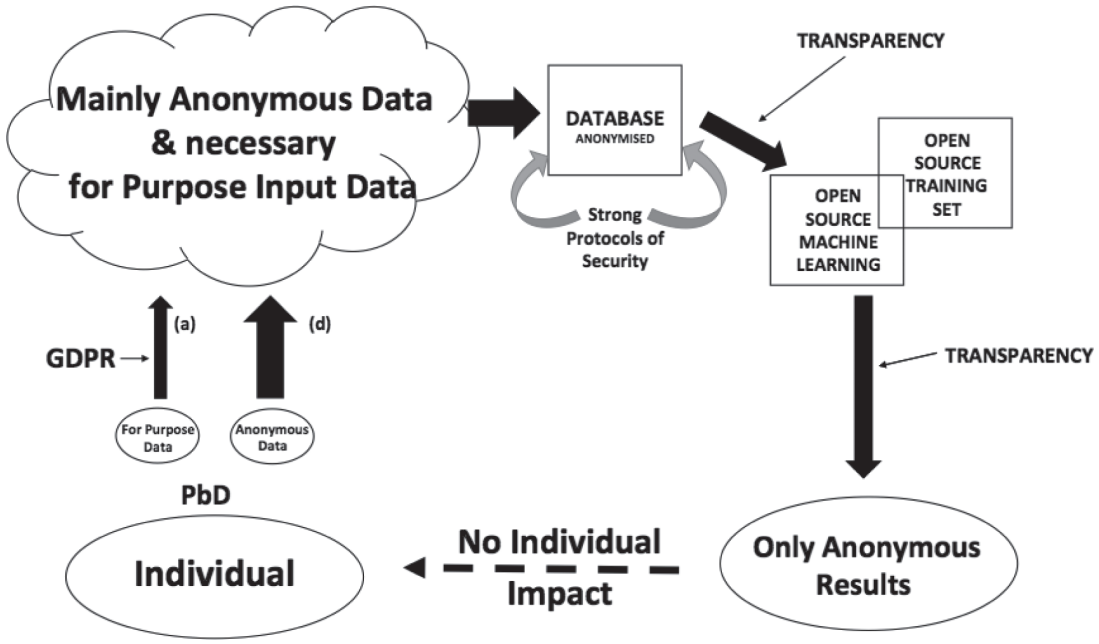


Figure 6: Collecting mainly anonymous or purpose-determined data, strong protocols of security, and enabling transparency in the automated decision-making as well as for its conclusions, provide optimal overall transparency and privacy protection. This solution also prevents profiling.

such as banks that can identify individuals via passwords for their financial transactions[46] or relay-on the state as a trusted third party, when individuals interact with a variety of social duties and services via passwords, for example, paying taxes or for legal matters.[47]

Cyberspace does not need to be organised in this way. Blockchain-style technologies could allow us to stay anonymous, for example, for all internet interactions so the individual citizen could decide which data to reveal and only do so in an anonymous manner. A state could decide to reorganise its digital infrastructure to function in this manner, as it is technically feasible today.[48] So far, no state has done this yet, although it is an attractive solution from a data privacy and a data ethics point of view and it is an ongoing discussion, for example, in Denmark.

So far we have only discussed actions and responsibilities by companies and states that use private data for a variety of reasons. Individual citizens, organisations and business could also take individual action regarding protecting their online privacy by using alternative technologies. This can be obtained by implementing a local privacy by design, internet architecture for data storage, communication and online transactions; in short a local internet architecture that 'takes home the data cloud', guarantees privacy and that reduces or eliminates the need for trusted third parties, as discussed by Monti and Rasmussen, based on a so-called RAIN-style software architecture.[49] As a default, RAIN ensures anonymity in cyberspace as well as privacy protection against cyber attacks. As needed, purpose-determined data can be created as discussed above, so a RAIN-style software architecture is also included in Figure 6 as it resides with the 'Individual'.

Longer term, we propose a RAIN-style software architecture as the backbone both for the internet of human interactions as well as the internet of things (IoT) both to ensure privacy and democratic oversight. The IoT is currently in the process of integrating all of our previously discrete critical infrastructures ranging from communication, social media,

entertainment and banking, through sensors and actuators; also affected are, transportation, digital manufacturing, energy, health, food production and distribution, water, sewage and so on. Cryptocurrency and increasing parts of governmental administration are likely to follow.

It is equally important that (1) citizens regain control of their private data and can remain anonymous for their online (and off-line) activities; and (2) that a democratic governance structure is developed and implemented for our increasingly integrated critical infrastructures, as these have an impact on almost everything in our lives. A RAIN-style software architecture would ensure this while allowing businesses to develop on top of the open source RAIN software architecture, for example, as the big data economy was able to develop on top of the open source Hadoop software architecture. Thus, the big data economy would be able to continue to develop, but with regained personal data privacy and with democratic oversight of our critical infrastructures.

Dialogue between business, law, ethics and technology

Creation of physical meeting places, events and perhaps a new kind of institution is needed such that a real debate can develop between interdisciplinary professionals related to business, law, ethics and technology. It is paramount that significantly more communication across disciplines occurs. The aim for such activities is obvious: to arrive at some level of consensus between feasible business designs and business models, legal requirements, ethically desired properties and technical feasibility when it comes to big data and automated decision-making.

Education at all levels is another crucial component. The knowledge and insight into data ethics issues of the general population, in schools as well as in the business and tech communities, need to be boosted so that our society can obtain a greater understanding of the opportunities and consequences of using data. Also competencies about the underpinning technologies and possibilities of using data and AI should become part of the law curriculum. At technical universities, students should not only learn about coding and AI, but also about the social impact of these technologies that they learn how to design and to implement.[50]

As a society, we need to engage in a continuous conversation about what constitutes a desirable level of algorithmic support and what the appropriate institutional framework might be for such support. We need to develop an intuition about what is reasonable to expect from algorithms and what reasonable means in this context. And finally, we should develop laws to support and enforce our findings. Today, there is not yet a balance between the new technologies and our laws.

CONCLUSIONS

The main goal of this paper is to outline: (1) how and why a new data driven economy emerged; (2) the consequences of this development in terms of lost control of private data; (3) the increasing impact of automated inferences and assessments that companies and the state make about individuals; and (4) how these assessments can have discriminatory effects for the individuals and/or group of individuals. We discussed the GDPR in the context of points (2), (3) and (4), and we further discuss the conflict between the intentions of the GDPR versus intellectual property and trade secret laws. (5) Finally we explored four solutions that companies, individuals and the state could develop. These four solutions reflect increasing levels of transparency and personal control over citizens' own data, and we discussed the GDPR, intellectual property and trade laws in this context.

As we document in this paper, there is currently a lack of transparency in the

automated decision-making processes and resulting profiling. Data subjects can neither be notified about when they are profiled nor have access to their profiles. The most crucial part of automated decision making-processes are assessments that companies make of individual, or groups of individuals, which have adverse and discriminatory consequences.

We propose a variety of solutions for data privacy and automated decision-making, and conclude that the long-term solution we should aim for is the implementation of an open source software architecture that ensures privacy by design from the onset. One such open source architecture is RAIN. Such an open source architecture would allow companies to build on it and further develop it in a similar manner as the open source Hadoop architecture enabled the development of a big data driven economy. Adding a RAIN-style architecture would enable the big data economy to continue to develop, but based on privacy by design.

Such a solution would recover a more levelled playing field for citizens, businesses and the state, as through regained privacy the autonomy of citizens and thus citizens' personal freedom would be recovered. As a further consequence of such a solution, we would obtain democratic oversight of our integrated digital infrastructures, recover more fair elections, as well as more fair markets.

ACKNOWLEDGMENT

We thank Stephan Engberg, Ulrik Jørgensen and Elise Lassus for critical and constructive discussions regarding Possible Solutions.

References and Notes

1. Moore, G. E. (1965) 'Cramming more components onto integrated circuits', *Electronics*, Vol. 38, No. 8. p. 117.
2. Moore, G. E. (1975) 'Progress in digital integrated electronics', Technical literature, Copyright 1975 IEEE. Reprinted, with permission. Technical digest. International Electron Devices Meeting, IEEE, 1975, pp 11–13; 'the rate of increase of complexity can be expected to change slope in the next few years […] The new slope might approximate a doubling every two years, rather than every year, by the end of the decade.' p. 13. available at: https://www.eng.auburn.edu/~agrawvd/COURSE/E7770_Spr07/READ/Gordon_Moore_1975_Speech.pdf. (accessed March 3, 2019)
3. Kurzweil, R. (1999) 'The age of spiritual machines: When computers exceed human intelligence'. Penguin Group, London.
4. Friedman, T. L. (2016) 'Thank you for being late: An optimist's guide to thriving in the age of accelerations', Penguin Group, London, p. 98.
5. Ghemawat, S., Gobioff, H. and Leung, S. T. (2003) 'The Google File System', available at: https://storage.googleapis.com/pub-tools-public-publication-data/pdf/035fc972c796d33122033a0614bc94cff1527999.pdf (accessed 19th December, 2018).
6. Dean, J. and Ghemawat, S. (2004) 'MapReduce: Simplified data processing on large clusters', available at: https://storage.googleapis.com/pub-tools-public-publication-data/pdf/16cb30b4b92fd4989b8619a61752a2387c6dd474.pdf (accessed 19th December, 2018).
7. Apache Hadoop was created by Doug Cutting and Mike Cafarella. It is an open source software framework for storage and large scale processing of datasets on clusters of commodity hardware. Hadoop is an Apache top-level project being built and used by a global community of contributors and users. It is licensed under the Apache License 2.0. http://hadoop.apache.org/.
8. Friedman, see ref. 4 above. p. 20.
9. AT&T Annual Report (2014) available at: https://www.att.com/Investor/ATT_Annual/2014/att_introduces_new_concepts_for_telecom_network.html (accessed 7th March, 2019).
10. Monti, M., Rasmussen, S., Moschettini, M. and Posani, L. (2017) 'An alternative information plan', SFI Working Paper, 2017-07-021, Santa Fe Institute; Santa Fe Institute Working Group Summary Report on 'Envisioning new modes of cultural and technological change', 31st July to 2nd August, 2017, workshop organised by D. Farmer, E. Beinhocker, J. Clippinger and S. Rasmussen; working group participants: M. Bedau, E. Beinhocker, J. Clippinger, M. Crokett, D. Farmer, N. Hanauer, N. Pinkston, L. Rasmussen, S. Rasmussen, P. Stover and J. Tainter; workshop summary written by L. Rasmussen, P. Stover, M. Bedau, E. Beinhocker, M. Crokett, D. Farmer, J. Tainter and S. Rasmussen.
11. Hepple, B. (2017) 'Can Facebook predict your political views? Yes, it can', available at: https://www.citizenme.com/public/wp/facebook-predict-political-view/ (accessed 25th November, 2018).
12. Tinker, B. (2018) 'How Facebook 'likes' predict race, religion and sexual orientation', available at: https://edition.cnn.com/2018/04/10/health/facebook-likes-psychographics/index.html (accessed 25th November, 2018).

13. Cuthbertson, A. (2018) 'Facebook patent predicts when you'll die', available at: https://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-patent-predict-die-death-prediction-algorithm-personal-data-privacy-a8417771.html (accessed 30th November, 2018).
14. Jones, R. (2018) 'Facebook reportedly wants to use AI to predict your 'future behavior' — So advertisers can change it', available at: https://gizmodo.com/facebook-reportedly-wants-to-use-ai-to-predict-your-fut-1825245517 (accessed 30th November, 2018).
15. Korosec, K. (2018) 'Amazon's Alexa can now act on "hunches" about your behavior', available at: https://techcrunch.com/2018/09/20/amazons-alexa-can-now-act-on-hunches-about-your-behavior/ (accessed 30th November, 2018).
16. Cook, J. (2018) 'Amazon patents new Alexa feature that knows when you're ill and offers you medicine', available at: https://www.telegraph.co.uk/technology/2018/10/09/amazon-patents-new-alexa-feature-knows-offers-medicine/ (accessed 30th November, 2018).
17. Allen, M. (2018) 'Health insurers are vacuuming up details about you – And it could raise your rate', available at: https://www.npr.org/sections/health-shots/2018/07/17/629441555/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates?t=1544433231164 (accessed 5th December, 2018).
18. Article 29 Data Protection Working Party, 'Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679' (2018) 17/EN WP 251rev.01, available at: http://ec.europa.eu/newsroom/article29/document.cfm?doc_id=49826, p. 8 (accessed 5th December, 2018).
19. 'Personal data shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity'. 'It covers 'objective' information, such as the presence of a certain substance in one's blood. It also includes 'subjective' information, opinions or assessments', available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf, pp. 4–6 (accessed December, 17, 2018.).
20. Wachter, S. and Mittelstadt, B. (2018) 'A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI', available at: https://www.researchgate.net/publication/327872087 (accessed 13th December, 2018).
21. YS, M and S v Minister voor Immigratie, Integratie en Asiel, Joined Cases C-141/12 and C-372/12 (European Court of Justice (Third Chamber)).
22. *Ibid*., para 39.
23. *Ibid*., para 47.
24. *Ibid*., para 45.
25. Peter Nowak v Data Protection Commissioner, Case C- 434/16 (European Court of Justice (Second Chamber)).
26. *Ibid*., para 56.
27. *Ibid*., paras 35–36.
28. Watcher, S., Floridi, L. and Mittelstadt, B. (2016) 'Why a Right to explanation of automated decision-making does not exist in the General Data Protection Regulation', available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469, p. 8 (accessed 3rd December, 2018).
29. *Ibid*., p. 7.
30. *Ibid*., p. 7.
31. *Ibid*., p. 9.
32. Article 29 Data Protection Working Party, see ref. 17 above, p. 21.
33. Recital 71 GDPR says: 'The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention'.
34. Article 2 of the Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure (2016) L 157/9.
35. Epstein, R. and Robertson, R. E. (2015) 'The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections', available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4547273/ (accessed 20th December, 2018).
36. Clemons, E. K. and Madhani, N. (2010) 'Regulation of digital business with natural monopolies or third-party payments business models: Antitrust lessons from the analysis of Google', DOI: 10.2753/MIS0742-1222270303, p. 8; https://repository.upenn.edu/fnce_papers/80/, *Journal of Management Information Systems*, Vol. 27, 2010, No. 3, Pages, 80. (accessed 2nd December, 2018).
37. Article 25 GDPR says that 'in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.' And the Article says, too 'both at the time of the determination of the means for processing', which means the moment in which an organisation is thinking about the way in which it will process information 'and at the time of the processing itself'.
38. Cavoukian, A. (2009) 'Privacy by design: The 7 foundational principles', Office of the Information and Privacy Commissioner of Ontario https://iab.org/wp-content/IAB-uploads/2011/03/fred_carter.pdf (accessed 10th November 2018).
39. Recital 26 says: 'Personal data which have undergone pseudonymisation, which could be attributed to a

natural person by the use of additional information should be considered to be information on an identifiable natural person'.
40. Article 29 Working Party, Opinion No. 05/2014 on anonymization techniques, April 2014, available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf (accessed December 10, 2018).
41. Weinberg, G. (2017) 'Privacy Mythbusting #3: Anonymized data is safe, right? (Er, no.)', available at: https://spreadprivacy.com/data-anonymization/ (accessed 18th December, 2018).
42. Schneier, B. (2017) 'Why 'anonymous' data sometimes isn't', available at: https://www.wired.com/2007/12/why-anonymous-data-sometimes-isnt/ (accessed 18th December, 2018).
43. Arrington, M. (2006) 'AOL proudly releases massive amounts of private data', available at: https://techcrunch.com/2006/08/06/aol-proudly-releases-massive-amounts-of-user-search-data/ (accessed 18th December, 2018).
44. Belu, A. M. (2018) 'Spiir's working with anonymization, security & trust', available at: https://dataethics.eu/en/interview-christian-panton-spiir/ (accessed 18th December, 2018).
45. Engberg, S., Interview by Hursh Joshi, March 3, 2019, available at: https://www.youtube.com/watch?v=HLnlte5u_s0 (accessed 13th March, 2019).
46. Such as Danske Bank in Denmark or Banco de España in Spain.
47. Digital self-service for finding information relating to public services in Denmark, see https://www.borger.dk/.
48. Monti, M. and Rasmussen, S. (2017) 'RAIN: A bio-inspired communication and data storage infrastructure', *Artificial Life*, Vol. 23, pp. 1–6, doi: 10.1162/ARTL_a_00247.
49. An extensive discussion of such possibilities can be found, for example, in Monti *et al*. (2017) (ref. 9 above) and Monti and Rasmussen (2017) (ref. 46 above). This locally rooted and bio-inspired architecture is called RAIN. It offers a file storage service that in contrast with current centralised cloud storage, has privacy by design, is open source, is more secure, is more reliable, is scalable (can grow organically), is more sustainable, has community ownership, is inexpensive and is potentially faster and more efficient.
50. Data for the benefit of the people, Recommendations from the Danish Expert Group on Data Ethics, Danish Ministry of Commerce (Erhvervsministeriet), Publication 22-11-2018, Recommendation nr. 6, available at: https://em.dk/media/12190/dataethics-v2.pdf (accessed 1st December, 2018).

Implementing a by design and by default approach

Received: 31st January, 2019

Richard Preece
is director of DA Resilience Limited. In his various roles, he provides consultancy on cyber risk, data protection and overall organisational agility and resilience. He has designed and delivers GCHQ certified training courses, via the OSP Cyber Academy and is an executive fellow on the Henley Business School GDPR Integration Programme. He was a co-opted panel member of the recent British Standards Institute (BSI) BS 31111:2018 *Cyber Risk and Resilience — Guidance for Boards and Executive Management*. He is a co-opted member of the BSI's governance panel for standards. Among various qualifications, he holds two master's degrees, including an MSc in design of information systems.

DA Resilience, SG House, 6 St Cross Road, Winchester, SO23 9HX, UK
E-mail: richard@daresilience.com

Abstract Building upon the concept of privacy by design, security and data protection by design and by default are important obligations within the General Data Protection Regulation (GDPR) and associated national legislation. This paper seeks to summarise some practical approaches to develop effective capability to deliver *by design* requirements: (1) a whole project lifecycle design approach; (2) a contextual risk-based approach; (3) the use of goals and principles approach; and (4) integration of safeguards/controls into operational use. While *by default* requires: (1) only processing that is necessary approach; and (2) not releasing data to unauthorised people.

KEYWORDS: by design, by default, risk, project management, governance, capability

INTRODUCTION

The litmus test of privacy and security is to be able to demonstrate not just consideration of privacy, but meaningful and assured capability (people, process and technology) to deliver it. Delivering the privacy of individuals and meeting the obligations of the General Data Protection Regulation (GDPR) and any associated national data protection legislation is complex, dynamic and often uncertain. Delivering the separate but overlapping security of the organisation, including information (cyber) and physical security is equally so.

Both require judgement based upon context. In particular, strategic alignment of the organisation's mission and goals with supporting privacy, data protection and security objectives to deliver effective capability. This paper intends to provide a conceptual framework, based upon the author's experience of working with multiple clients in multiple sectors.

Background

For the purposes of this paper, the European Data Protection Supervisor (EDPS) Preliminary Opinion on Privacy by Design will be taken as the paper's conceptual reference point.[1] To paraphrase the EDPS, the term 'privacy by design' (PbD) is used to designate the broad concept of technological measures for ensuring privacy. In contrast, the EDPS uses the terms *data protection by design* and *data protection by default* to designate the specific legal obligations established by the GDPR, Article 25.

The EDPS considers that a wider spectrum of approaches may be taken into account for the objective of PbD. This includes an ethical dimension, consistent with the principles and values enshrined in the EU Charter of Fundamental Rights of the EU. The EDPS also considers the security of processing covered in GDPR Article 32 and the *integrity and confidentiality* or *security* data protection principle. Equally of GDPR Article 24 and the obligation of data controllers to implement all data protection principles and the compliance with the whole of the GDPR.

Practically, this requires the ability to take a *holistic* approach on the part of those who must deliver governance, or the 'the system by which the whole organization is directed, controlled and held accountable to achieve its core purpose over the long term.'[2] This is no different to any other part of governance, but the subject matters of privacy, data protection and security are often not areas of familiarity or expertise for many at board and executive management level. This extends to the vast majority of other people who will have in some form delegated roles and responsibilities for data protection and security; arguably everyone has a personal responsibility for both regardless of role. Ultimately, this requires a shared understanding of what the privacy, data protection and security objectives are and the capability to deliver them.

Aim

The aim of this paper is to present a practical framework for delivering a capability to provide aligned privacy, data protection and security objectives, which can be applied to the context of an individual organisation. The paper will achieve this by first outlining the problem situation, described as the *cyber landscape*. Secondly, the EDPS preliminary opinion of what *by design* and *by default* requirements and some practical approaches will be outlined. Finally, some practical *by design* and *by default* assumptions, which can be applied to the context of organisations in their projects, operations and governance, including assurance will also be outlined.

To do this, the paper will draw from European Union's (EU) funded PRIPARE (PReparing Industry to Privacy-by-design by supporting its Application in Research) *Privacy and Security by Design Methodology Handbook*,[3] published in 2015, and other more recent authoritative sources, including the European Data Protection Board (EDPB)[4] and UK's National Cyber Security Centre (NCSC) guidance.[5] In particular, the PRIPARE Methodology and UK NCSC's Capability Assessment Framework (CAF) will be used as points of reference. The CAF was developed to aid organisations comply with the UK's implementation of the EU's Network Information Security (NIS) Directive, the intent being to aid decision-making across an organisation's governance oversight and *bottom-up* implementation in a complex, dynamic and uncertain cyber landscape.

The European Union Agency for Network and Information Security (ENISA)[6] mandate has recently been expanded to reflect a role more akin to the UK NCSC's at national level under the EU Cybersecurity Act.[7] The Act has reinforced ENISA's mandate and includes the creation of a framework for European Cybersecurity Certificates for products, processes and services that will be valid throughout the EU. The cybersecurity certification framework is intended to incorporate security features in the early stages of their technical design and development (security by design) and is intended to provide a level of security assurance that are independently verified. As will become apparent in this paper, this is likely to take time and will never guarantee security. But the Act, along with the NIS Directive, are clear statements of the importance of security and the regulatory *direction of travel*.

THE PROBLEM SITUATION

The problem situation is based upon an environment that has been described as the cyber landscape. This is summarised as the hyper-connected interactions between people, between people and machines, and between machines.[8,9] Figure 1 includes a visualisation of the cyber landscape, which is the basis of the problem situation that must be addressed in any by design and by default approach.

The layers of the cyber landscape, can be summarised as follows:

- *Cognitive layer*: The cognitive layer (persona, people and social) consists of the information that connects people to cyberspace and the people and groups who interact by using and operating the networks, that is, people to people (P2P) and people to machine (P2M).
- *Virtual layer*: The virtual layer (also known as the logical layer) (network and information) consists of the software/applications and connections between network nodes, that is, machine to machine (M2M).
- *Physical layer*: The physical layer (real world) consists of the physical network components (infrastructure), physical version of data (printed documents) and their associated location, that is, M2M, but also P2P and P2M.
- *Note*: It should be noted any European Cybersecurity Certificates are likely to focus only upon what are summarised as the M2M aspects of the cyber landscape; however, consideration of product development, using PRIPARE or other methodologies *may* form part of the accreditation.

This creates a hyper-connected environment or problem situation, driven by the growth-in accessibility to data from multiple sources. This growth enables deeper and new network effects, summarised by *Metcalf's Law*: the power of the network = number of nodes^2 ($P = n^2$).[10] Metcalf's Law was one of the driving forces of the early internet and world wide web. This, combined with the relative limited processing power of computer systems until recently and the general *race to market*, means that the cyber landscape has largely evolved with a priority on functionality, scalability and openness, rather than privacy and security.

The seemingly endless desire of people (personally and professionally) and organisations to connect, has led to a number of cyber landscape paradoxes,[11] which are summarised below:

- Unprecedented powers, but making users less secure.

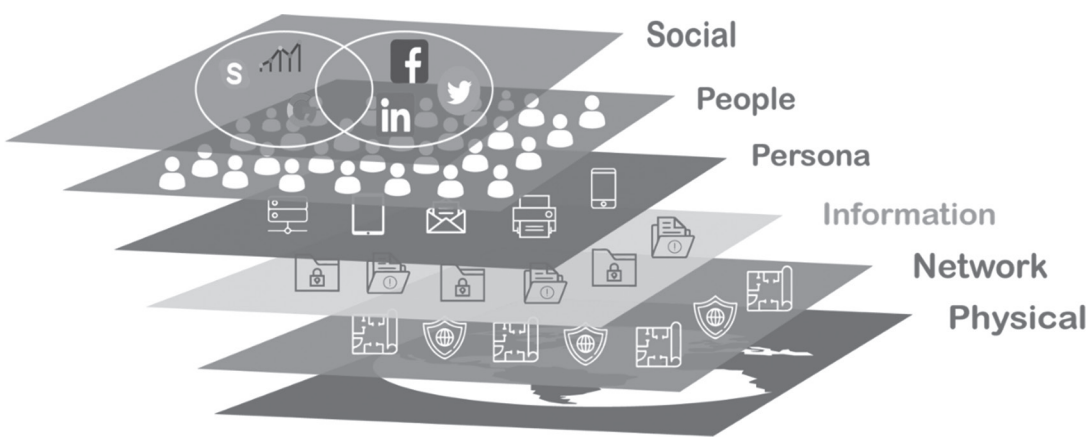


Figure 1: The cyber landscape

- Capabilities that enable collaboration and innovation, but create further opportunities for security breaches.
- Concentration of data and manipulative power to vastly improve efficiency and scale of operations, but means exponentially increasing the risk associated with a privacy and security breach.
- Great capability, but the complexity creates new vulnerabilities and lowers the visibility of intrusions.
- Responsiveness and flexibility, but also permits small changes in a component's design or direction to degrade or subvert system behaviour.
- Democratises capabilities to collaborate and innovate, but also removes safeguards present in systems that require hierarchies of human approvals.

In essence, this means there are inherent vulnerabilities in the software and technologies found in the cyber landscape. This was acknowledged in 2013 by the Internet Engineering Task Force (IETF) in the wake of the Snowden revelations concerning mass surveillance, 'the scale of recently reported monitoring is surprising. Such scale was not envisaged during the design of many Internet protocols'.[12]

There is a parallel to this problem situation from engineering safety; often referenced as the *swiss cheese* model of accident causation. The swiss cheese model arose from analysis of the 1988 Piper Alpha disaster and recognised the existence of inherent flaws or holes, which may align on their own or in combination with more active measures.[13]

Transposing the theory from engineering safety into the cyber landscape, the combination of inherent flaws creates a number of potential *non-malicious* hazards, based upon people, process and technology. Nevertheless, threats may arise from a collection of *malicious* threat actors, who have the motive and means to exploit inherent vulnerabilities or to take active measures to create new ones. The message is simple: there is no ability to deliver absolute security of privacy engineering; hence, any European Cybersecurity Certificates will not provide an absolute guarantee of security. Figure 2 provides a simple summary of potential sources of threats and hazards, the vulnerabilities and the effects these create, but not the range of possible impacts on individuals' and organisations.

Finally, in recent years, the cyber landscape has been further complicated by the increasing use of artificial intelligence (AI) and in particular machine learning (ML) techniques in core operational processes. This includes AI being increasingly being adopted in the fields of consumer products and services.

Sources of Threats & Hazards	Vulnerabilities	Effects

Potential Malicious Actors:
• State/state-sponsored
• Serious Organised Criminals
• Terrorists
• Private Enterprises
• Activists
• Low Level Criminals
• Current/Former Employees
• Lone Actors
 	
Non-Malicious Failures:
• IT systems/software failures
• Human error
• Culture:
  • Trying to get the job done and by-passing the controls! or
  • Ignorant of the policy and process
	Within the organisation:
	• People
	• Processes
	• Software Applications
	• Paper Information
	• Devices
	• Networks
	• Infrastructure
	
	External to the organisation (data processors and supply chain):
	• IT/infrastructure providers
	• Non-IT supply chain/partners, including all of the above
	
	Physical Actions:
	• The physical removal, manipulation or denial of information or related item
		Leads to a breach of confidentiality, integrity and availability of personal data and other data by being:
		• Inaccurate, insufficient or out of date;
		• Excessive or irrelevant;
		• Kept for too long;
		• Disclosed to those who the person it is about does not want to have it;
		• Used in ways that are unacceptable to or unexpected by the person it is about; or
		• Not kept securely.

Figure 2: Summary of the cyber landscape's threats and hazards, vulnerabilities and effects

> Machine learning is the process that powers many of the services we use today — recommendation systems like those on Netflix, YouTube, and Spotify; ... The list goes on…. In all of these instances, each platform is collecting as much data about you as possible — what genres you like watching, what links you are clicking, which statuses you are reacting to — and using machine learning to make a highly educated guess about what you might want next. Or, in the case of a voice assistant, about which words match best with the funny sounds coming out of your mouth.[14]

Thus, AI is creating new and still to be fully understood threats and hazards, including inherent bias, such as the use of criminal risk assessments in the USA.[15] Ultimately, the success of widespread AI will depend upon creating an environment of trust. This will in part be supported by developing more transparent AI to ensure that people have a better understanding of how algorithms make automated decisions, based partly upon the increasing amounts of data collected.

This will become increasingly difficult as more (human) unsupervised learning techniques are applied; however, algorithms are designed and created by people in the first instance, even those evolve on their own through unsupervised learning. There is an increasing body of work[16] examining this area reflecting its potential for both good and harm upon society and individuals. This includes some initial attempts to develop goals and set of principles for holding algorithms accountable.[17] These will not be considered further in this paper, but they provide a useful reference point which can be integrated into the overall by design, by default approach articulated in this paper.

The implication of the cyber landscape's hyper-connectivity and inherent vulnerabilities is the creation of a complex, dynamic and ambiguous environment. This environment is perhaps best viewed as a *system of systems*, that stretches beyond individual organisations boundaries and control from the overlapping privacy, data protection and security perspectives. This complexity, dynamism and ambiguity make decision-making at both the governance level or *top-down*, and the operational and project level or *bottom-up* level, extremely problematic to deliver privacy and security capability.

THE REQUIREMENTS TO SUPPORT BY DESIGN AND BY DEFAULT

In the EDPS's opinion, a number of requirements for a by design and by default approach are identified. There are four by design requirements and a further two for by default. All requirements are of equal importance and form an integral part of governance (GDPR, Article 5(2), the *Accountability Principle*). The requirements will be summarised below, with an aim to show how they link together and enable a systematic framework that can be applied to a by design and by default approach within the cyber landscape outlined above.

By design: Whole project lifecycle approach

The EDPS, in their opinion, reiterates that the GDPR Article 25 requires consideration of safeguards both at the design and operational phase. This means taking a whole project life cycle and clearly identifying the protection of individuals and their personal data within the project requirements. Practically, this means an aligned approach between privacy and security requirements and developing a shared understanding of the cyber landscape with mutually supporting objectives.

The PRIPARE *Privacy and Security by Design Methodology Handbook* provides a foundation for individual organisations to approach this; however, it is potentially intimidating to all but the most well-resourced, technically able and mature organisations. Figure 3 provides a simplified and slightly modified overview of the PRIPARE phases and processes.

Capability						
• Organisational governance and risk management • Organisational privacy and security policies, processes and technologies (Identify, Protect, Detect, Respond & Recover) • Privacy and security culture (values, knowledge, skills, attitudes and behaviours)						
Analysis	Design	Implementation	Verification	Release	Maintenance	Decommission
• Functional description and high-level privacy and security analysis • Legal assessment • Privacy and security plan preparation • Detailed privacy and security analyses • Operationalisation of privacy and security principles	• Design enhanced privacy and security policies, processes and technologies • Design enhanced privacy and security culture	• Privacy and security implementation	• Privacy and security dynamic analysis • Privacy and security static analysis	• Incident management plan created or enhanced to reflect new project • Create system de-commissioning plan • Final privacy and security review • Publish public version of the DPIA	• Privacy and security verification and validation • Execute incident management plan as required	• Execute de-commissioning plan

Figure 3: PRIPARE phases and processes

The PRIPARE methodology, or the amended approach outlined in this paper, should be applied iteratively. This reflects that there is an element of discovery of new issues and risks as any project develops, while the organisation's privacy and security capability, referred to as environment and infrastructure in PRIPARE methodology, is critical to success.

The methodology framework has the potential to be applied to the use of AI techniques, applying additional principles for holding algorithms accountable; however, the author has no known examples or personal experience. Finally, the methodology may be applied in whatever order necessary and it is possible to jump from one phase to other (eg from maintenance to implementation or from design to verification), reflecting that most organisations are working on *brown field sites* with existing processes and technologies. Nevertheless, even new projects rarely start without any preconditions that have an impact upon design and operation.

By design: Risk-based approach

In essence, any risk-based approach seeks to understand the context, scope and criteria of the what is being assessed, accepting that there may be positive, negative or both effects. Having identified the assets to be protected, and or the obligations to be complied with, analysis and evaluation seeks to identify potential scenarios, including their likelihood and impact. Then, it is necessary to determine the appropriate measures or controls to reduce the likelihood or impact of potential negative risks, otherwise known as risk treatment.

The GDPR Article 35 and Recital 90, plus EDPB guidance requires that a contextual based risk-based approach is taken. This reflects also the PRIPARE analysis phase and processes identified in Figure 2. From a GDPR perspective, *risk* is a scenario describing an event and its consequences, estimated in terms of likelihood and severity of impact on the data subject, while from a security perspective, this commonly focuses upon the organisation. In both cases, the emphasis is normally on the negative impact, but positives should always be also identified to provide a holistic perspective to enable judgement.

Although privacy risk is the primary focus in the GDPR, it considers the impact upon the rights and freedoms of data subjects as well. This requires consideration of the

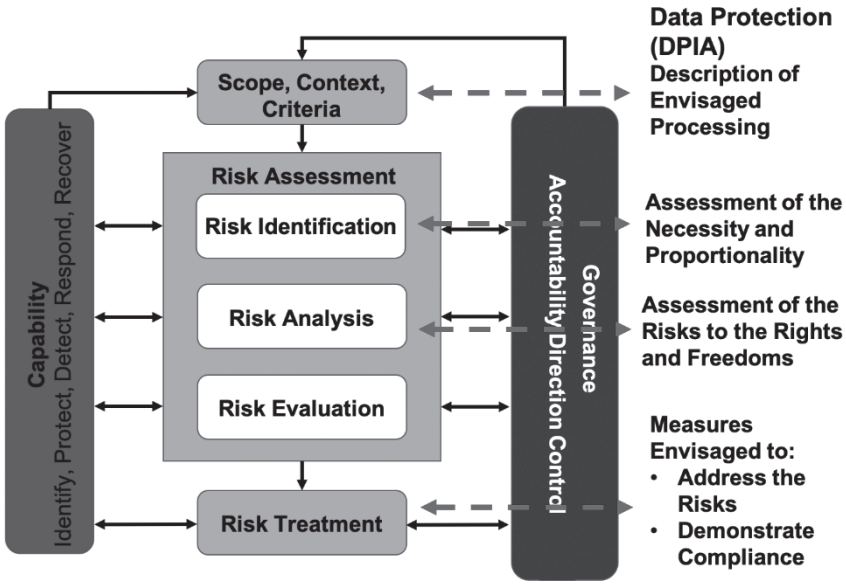


Figure 4: Integrated DPIA and I-SRA risk management processes

Charter of Fundamental Rights of the European Union,[18] from which the right to privacy derives and for which some effects may go beyond just privacy. This can be best achieved through a data protection impact assessment (DPIA), which must focus upon protecting individuals' privacy; however, owing to the dependency upon the overlapping security objectives, a separate information-security risk assessment (I-SRA) should also be conducted.

Both the DPIA and I-SRA should ideally be aligned in the terminology used, to ensure there is clarity and shared understanding between the separate processes. A common root cause of error is using different terminology for the same issue in different risk reporting processes within the same organisation.

ISO 31000:2018 *Risk Management — Principles*[19] provides a standard international good practice framework for risk management. Consequently, when possible, the DPIA section titles should reflect standard risk management process titles and be mapped to the minimum criteria for a DPIA reflected within the GDPR. Figure 4 provides an adapted ISO 31000:2018 process diagram, which helps map DPIA minimum criteria to general risk management process, including those used during I-SRAs. It also shows the linkages to governance and with capability. Communications and consultation with stakeholders are implied in Figure 4, while it is explicit in the original ISO 31000:2018 diagram.

By design: The use of goals and principles approach

The third requirement is the need for risk treatment measures to be appropriate and effective. This is important, due to the obligation under the GDPR Article 25 for the organisation to choose safeguards within the *state of the art* of the art of available technology and the cost of implementation of the measures. As the EDPS makes clear, 'this must not be interpreted in such a way that the measures chosen do not sufficiently mitigate existing risks and the resulting protection is not adequate.'[20] The challenge is how to identify appropriate and effective measures in the complex, dynamic and uncertain cyber landscape, however, when it is impossible to cater for all eventualities.

The common approach to general management, risk and compliance decision-making is to create a set of narrow rules to be followed. When this is practical, it can lead to efficiency and the achievement of desired objectives. But to work well, a set of prescriptive rules needs to cater for all eventualities.

The alternative approach is to define a set of principles that, if consistently used to guide decision-making, will collectively result in the desirable objectives; however, this is challenging for many individuals and organisations who are not familiar with the subject matter, that is, privacy, data protection and security individually and collectively. Equally, most are used to more output focused approaches, that is, rules and standards-based (*tick boxes*) to demonstrate compliance.

While it is not possible to devise an effective set of prescriptive rules for good privacy and security, it is possible to state a set of principles as a guide to decision-making. The GDPR has been drafted so that a set of data protection principles, if used consistently to guide decision-making, will collectively result in the desirable objective of data subjects' privacy.

The same conceptual approach has been adopted and developed from an organisational security perspective by the UK's NCSC. Initially, this was developed to provide security guidance for implementation of the GDPR and has been further evolved for the UK's NIS Regulation obligations for subject organisations. These are both pragmatic and forward-looking approaches to legislation and regulation that tacitly recognise the cyber landscape's complexity, dynamism and uncertainty.

The UK's NCSC CAF takes a set of objectives across the functions of identify, protect, detect, respond and recover. This aligns with, but develops security further, to a more objectives and principles-based approach than the US National Institute for Science and Technology (NIST) Cybersecurity Framework.[21] The US NIST Cybersecurity Framework often has been adopted as the most comprehensive and holistic approach to cyber security and benchmark in many regulated sectors globally.

With some minor adaption, it is therefore possible to align privacy, data protection and security through an adapted CAF, allowing the application of data protection and security principles, to support objectives under the holistic functions of identify, protect, detect, respond and recover. Figure 5 is an adaption of the UK NCSC's NIS Regulation CAF aligning data protection and security and principles.

<table>
<tr><th colspan="8">Objectives</th></tr>
<tr><th colspan="2">Identify</th><th colspan="2">Protect</th><th colspan="2">Detect</th><th colspan="2">Respond & Recover</th></tr>
<tr><td colspan="2">Managing Data Protection and Security Risk</td><td colspan="2">Protecting Against All Threats, & All Hazards</td><td colspan="2">Detecting Data Protection and Security Events</td><td colspan="2">Minimising the Impact of Incidents</td></tr>
<tr><th colspan="8">Principles</th></tr>
<tr><td>Governance</td><td>Risk Management</td><td>Protection Polices and Processes</td><td>Identity and Access Control</td><td>Security Monitoring</td><td>Proactive Security Event Discovery</td><td>Response and Recovery Planning</td><td>Lessons Learned</td></tr>
<tr><td>Asset Management</td><td>Supply Chain</td><td>Data Security</td><td>E2E Security</td><td colspan="4"></td></tr>
<tr><td colspan="2"></td><td>Resilient Networks & Systems</td><td>Staff Awareness & Training</td><td colspan="4"></td></tr>
</table>

Figure 5: Integrated objectives and principles CAF
Note: E2E, end to end.

Inevitably, every organisation that adopts this approach will need to develop their own version of this CAF from the generic version provided here. Nevertheless, it provides a practical and defensible point of reference, especially when used with a whole project life cycle and a PRIPARE methodology, which in this paper has been adapted to reflect the CAF; plus, the integrated DPIA and I-SRA risk-based processes, which form a key part of the CAF and PRIPARE methodology.

By design: Integration of safeguards into processing approach

The fourth requirement follows from the first three and is the obligation to integrate the identified safeguards, measures and controls into the processing. The EDPS identifies that the GDPR includes some safeguards to protect individuals whose data is processed through means that are *external* to the processing itself, such as privacy notices. This requirement instead focuses on the need to protect individuals by directly protecting their data and the way it is managed. This also aligns with the need to integrate organisational security safeguards, measures and controls into processing and overall operations and governance.

The PRIPARE methodology includes a set of guidelines and criteria to operationalise privacy and data protection. These are based upon the work completed by the French data protection supervisory authority, CNIL, while the GDPR was being drafted; however, they provide a useful reference point for developing privacy safeguards, measures and controls. Also, various information (cyber) security standards and frameworks can be applied according to context, taking into account, people, process and technology. The key is to follow a systematic and defensible framework which can document the analysis and subsequent phases of the PRIPARE methodology or similar.

Although much of the PbD literature focuses upon the use of privacy enhancing technologies (PETs), various security technologies, including some that are AI based should be considered. Reflecting the paradoxical nature of the cyber landscape outlined earlier, however, many controls, especially security technologies may create new privacy, data protection and security risks. Again, this requires judgement and documentation of the associated trade-offs. In author's experience, this includes the use of additional DPIAs for some of the more intrusive monitoring technologies of biometrics and individual behaviours both in the physical and virtual worlds.

By default: Only processing that is necessary approach

The first by default requirement is to limit processing to only what is necessary for specified purposes. The EDPS notes that it can be argued,

> that this obligation is already implicit in the 'purpose limitation' and 'data minimisation' principles in both the design and operation phases, the explicit rule stresses the importance of taking technical measures to meet the expectations of the individuals whose data are processed, not to have their data processed for other purposes than what the product and service is basically and strictly meant to do, leaving by default any further use turned off, for instance through configuration settings.[22]

This obligation is entirely aligned with good security practice also; however, as identified earlier, Metcalf's Law and the general *race to market*, has largely evolved with a priority on functionality, scalability and openness, rather than privacy and security. This by default obligation is something that will create tensions with many projects that seek to maximise data collection, including personal data, especially as AI technologies and techniques become ever more pervasive.

Nevertheless, by following the by design approach outlined above, including the investment not just in PETs but in people who design systems and make decisions, it should be possible to increasingly make this by default approach reality. The challenge for many organisation's will always be to understand what data flows occur at levels of the cyber landscape.

By default: Not releasing data to unauthorised people approach

The second and final by default requirement is an extension of the *data minimisation* and *storage limitation* principles. The EDPS identifies that GDPR, Article 25(2) 'establishes a precise obligation by instantiating the general principle in one particular use case: organisations shall set up measures to prevent personal data from being made public by default.'[23]

Again, this obligation is entirely aligned with good security practice also. In practice, this needs clear retention and disposal policies. Also, beyond the policy, an understanding of the nature of non-malicious hazards, especially culture and behaviours, becomes very important. The previously identified openness encouraged on the world wide web leads to many items of information, including personal data being published as a default.

Different cultural perspectives (national, sector, generational and individual) and sometimes legal obligations can lead to non-malicious publishing of personal data. By following the by design approach outlined above, however, especially in understanding culture and legal obligations, decisions can be made in a systematic and defensible way to demonstrate compliance this by default approach.

PRACTICAL BY DESIGN AND BY DEFAULT ASSUMPTIONS TO APPLY

The by design and by default assumptions outlined below are based upon the analysis above, guidance from the UK's NCSC and from the author's academic and professional (including military and working across multiple sectors) experience. Throughout there should be a desire for simplicity of operation, clarity of purpose and integration of capability (people, process and technology), to counter the prevailing complexity, dynamism and uncertainty of the cyber landscape.

The following assumptions provide a basis to apply the PRIPARE methodology and various by design and by default approaches outlined above.

(1) People are the key to privacy, data protection and security capability; consequently, change and all design and defaults need to be *people centric*.
(2) People, processes and technology all have inherent vulnerabilities, which can be exploited by malicious attack, or lead to a non-malicious incident.
(3) The likelihood of arising from the combination of malicious attack and non-malicious incident is assessed to be very high, therefore it should be considered a case *of when* and *not if*.
(4) The potential consequences of malicious attacks and non-malicious incidents may be amplified and display both predicable (linear) and unpredictable (non-linear) effects.
(5) The combination of a by design, and by default approach, supply-chain assurance, operating practice, maintenance and testing should minimise the likelihood and impact of successful attacks or failures, by:
    (a) Good design of systems includes appropriate limits and conditions for operation of each system (people, process and technology) and alternative systems for *critical* functions. This should include the ability to verify and validate, to maintain assurance and trust.
    (b) Good operating practice and maintenance is based in part

on a culture that reflects peer checking, self-assessment, training, accreditation, and internal and external assurance.

(c) Until capability has been tested, it is only a set of *arrangements* which may or may not work. Testing should be resourced and programmed, based upon challenging intelligence-led threats and hazards-driven scenarios and with clear objectives and follow-up learning actions.

SUMMARY

The concept of PbD has evolved along with the complexity, dynamism and uncertainty of the cyber landscape. The recent preliminary opinion on PbD issued by the EDPS provides a useful point of reference for considering the practical implications of a by design and by default approach. This is now an obligation under the GDPR and associated national data protection, plus additional security legislation and regulation for some organisations.

The EDPS has identified a number of aspects to PbD which have been outlined and extended in this paper, to demonstrate the need to consider privacy, data protection and security capability holistically. In doing so, although the purposes are different for each subject, they are aligned and can be mutually supporting in developing a positive strategic change in organisations.

Although reference to PETs and cyber technologies are often referenced for PbD and security by design respectively, ultimately a by design and by default approach requires an investment in people first and continuously thereafter. It is investment in culture: the values, knowledge, skills, attitudes and behaviours that ultimately enables privacy, data protection and security by design and by default to be integrated to deliver effective capability. This enables effective governance and bottom-up projects and operations.

References and Notes

1. European Data Protection Supervisor (EDPS) (2018) 'Preliminary opinion on privacy by design', Opinion 5/2018, 31st May, available at: https://edps.europa.eu/sites/edp/files/publication/18-05-31_preliminary_opinion_on_privacy_by_design_en_0.pdf (accessed 2nd January, 2019).
2. British Standard (2013) 'Code of practice for delivering effective governance', BS 13500.
3. Crespo García, A., Notario McDonnell, N., Troncoso, C., Le Métayer, D., Kroener, I., Wright, D., del Álamo, J. M. and Martín, Y. S. (2015) 'PRIPARE privacy and security by design methodology handbook', 31st December, available at: http://pripareproject.eu/wp-content/uploads/2013/11/PRIPARE-Methodology-Handbook-Final-Feb-24-2016.pdf (accessed 2nd January, 2019).
4. EDPB website: https://edpb.europa.eu/edpb_en (accessed 2nd January, 2019).
5. UK NCSC website: https://www.ncsc.gov.uk (accessed 2nd January, 2019).
6. ENISA website: https://www.enisa.europa.eu (accessed 2nd January, 2019).
7. EC News website: https://ec.europa.eu/commission/news/cybersecurity-act-2018-dec-11_en (accessed 2nd January, 2019).
8. British Standard (2018) 'Cyber risk and resilience – Guidance for boards and executive management', BS 31111.
9. UK MOD Joint Doctrine Note, 'Cyber primer version 2', available at: https://www.gov.uk/government/publications/cyber-primer (accessed 28th December, 2018).
10. 'Metcalfe's law', available at: https://en.wikipedia.org/wiki/Metcalfe%27s_law (accessed 28th December, 2018).
11. Danzig, R. J. (2014) 'Surviving on a diet of poisoned fruit: Reducing the national security risks of America's cyber dependencies', Centre for a New American Security, July, available at: https://www.cnas.org/publications/reports/surviving-on-a-diet-of-poisoned-fruit-reducing-the-national-security-risks-of-americas-cyber-dependencies (accessed 28th December, 2018).
12. IETF (2013) 'Security and pervasive monitoring', 7th September, available at: https://www.ietf.org/blog/security-and-pervasive-monitoring/ (quoted by EDPS paper) (accessed 28th December, 2018).
13. 'Swiss cheese model', available at: https://en.wikipedia.org/wiki/Swiss_cheese_model (accessed 28th December, 2018).
14. Hao, K. (2018) 'What is machine learning? We drew you another flowchart', *MIT Technology Review*, 17th November, available at: https://www.technologyreview.com/s/612437/what-is-machine-learning-we-drew-you-another-flowchart/ (accessed 28th December, 2018).
15. Angwin, J., Larson, J., Mattu, S. and Kirchner, L., (2016) 'Machine bias', ProPublica, 23rd May, available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed 2nd January, 2019).

16. Finley, K. (2016) 'Tech giants team up to keep AI from getting out of hand', *Wired Magazine*, 28th September, available at: https://www.wired.com/2016/09/google-facebook-microsoft-tackle-ethics-ai/ (accessed 2nd January, 2019).
17. Diakopoulous, N. and Friedler, S. (2016) 'How to hold algorithms to account', *MIT Review*, 17th November, available at: https://www.technologyreview.com/s/602933/how-to-hold-algorithms-accountable/ (accessed 2nd January, 2019).
18. Charter of Fundamental Rights of the European Union, available at: https://ec.europa.eu/info/aid-development-cooperation-fundamental-rights/your-rights-eu/eu-charter-fundamental-rights_en (accessed 2nd January, 2019).
19. ISO/FDIS 31000:2018 Risk Management — Guidelines.
20. European Data Protection Supervisor (EDPS) (2018) 'Opinion 5/2018 Preliminary opinion on privacy by design', 31st May, page 11, para 29, available at: https://edps.europa.eu/sites/edp/files/publication/18-05-31_preliminary_opinion_on_privacy_by_design_en_0.pdf (accessed 2nd January, 2019).
21. US NIST Cybersecurity Framework, available at: https://www.nist.gov (accessed 2nd January, 2019).
22. *Ibid*., page 7, para 33.
23. *Ibid*., page 7, para 34.

Automotive viewpoint: How dealerships can streamline GDPR compliance, while minimising data breach and supply chain risks

Received: 29th October, 2018

Jim Steven
is Head of Data Breach Services for Experian Consumers Services in the UK, building on the knowledge, experience and success of Experian's global data breach resolution offering. Jim's passion lies in helping organisations understand their customers, and to build loyalty and retention through customer satisfaction. Jim's focus is to help businesses to take proactive steps in preparing for the ever-growing threat of data breaches. To fully prepare for a data breach, organisations must invest in and test plans regularly, updating as new breaches occur, to ensure there is a comprehensive response plan in place that addresses everything from customer notification, to crisis management, forensics and legal considerations — most importantly to protect those affected, and to protect the organisation's reputation in the longer term. Jim and his team support organisations through this challenge — helping to provide reassurance to organisations and those affected in uncertain times. Prior to joining Experian, Jim worked in the security and risk management industry providing expertise in security risk management solutions, travel risk management, aviation security and corporate security for some of the world's largest security companies.

Experian, 6th Floor, Cardinal Place, 80 Victoria Street, London, SW1E 5LJ, UK
E-mail: Jim.Steven@experian.com

Abstract In the new world of GDPR how can the Automotive industry take positive steps to manage, store and protect an individuals personally identifiable information (PII)? The reality is all organisations are facing the increased threat of a data breach incident and so need to take proactive steps to protect both the business and their customers. Protecting an individuals identity is of paramount importance and a responsibility all organisations should take to ensure they align to regulation, and more importantly put the individual at the heart of the business priorities. What are the crucial considerations and who should drive this cultural change through the dealership and their third party supply chain? Ultimately, understanding consumer sentiments on the topic of managing personal data and identifying the key components of a data breach readiness plan can support the organisation to maintain trust, but respond with confidence when a data breach incident happens.

KEYWORDS: GDPR, data breach response, data breach readiness, personally identifiable

INTRODUCTION

Automotive dealerships are among many UK businesses that have been taking a long hard look at how they collect, store and manage data. In particular, the EU General Data Protection Regulation (GDPR) now requires dealerships to ensure that customers' personally identifiable data is collected and managed securely to reduce the risk of it being stolen or modified. What is more, dealerships must now clearly define processes to ensure that personally identifiable information is only used for specific purposes in line with customer consent, such as financing applications or onboarding programmes.

These requirements are even more challenging as dealerships often share customers' personal data with an extended network of supply chain partners, ranging from vehicle manufacturers, insurance partners and servicing partners, to name a few. This means that dealerships need to protect customer data and ensure it is used in line with customers' consent across the entire partner ecosystem — or risk non-compliance with the GDPR.

With a rising tide of data breaches across the UK and worldwide, effective data protection is vital — both across dealerships and their extended network of supply chain partners. Recent research by Experian and ComRes, however, shows that only 33 per cent of medium and large businesses are 'very confident' about what to do in the event of a data breach in third-party's systems.[1]

RISING FRAUD AND DATA SECURITY RISKS, RESULTING IN DATA BREACHES

While the requirements of the GDPR are clear, many dealerships are working with legacy systems and processes that have the potential to put personal data and its security at risk. In addition, many organisations have not yet fully identified or embedded a culture of data protection, which can create security gaps that hackers, fraudsters and other criminals can take advantage of.

When looking at the purchase itself there are a number of steps that require forms to be completed by the individual that contain personally identifiable information, all forming part of the essential car buying process. The reality is, even these initial steps have the potential to put the dealership at risk of a data breach incident and the resulting regulatory fines.

There are a number of measures dealerships can consider taking to reduce the risks of regulatory non-compliance and reputational damage. These include the following steps, as outlined below.

Understand people, processes and policies

In the first instance, a simple review of current people, processes and policies can help a dealership to identify areas of weakness and risk. Simply understanding if there are any incomplete or inappropriate data collection practices, storage or management of data can help dealerships to significantly reduce the risk of a data breach incident.

Map where personally identifiable data is held

With customer data (both old and new), and employee data, it can be very difficult to keep track of what information is held and where it resides in the organisation or within supply chain partners' organisations. In fact, research from Experian and ComRes shows that 32 per cent of businesses do not know where all their third party suppliers store their customers' personal data.[2]

By conducting an in-depth analysis of customer data (Box 1) and the systems that are used to store and manage it, as well as mapping data, dealerships can see where data flows across the dealership and its partners. This makes it easier to detect potential security vulnerabilities in systems and processes, helping dealerships take measures to minimise security risks. In addition, dealerships can begin to centralise their data and delete aged or duplicated data, helping to reduce their attack surface and minimise the risk of a data breach incident.

Manage data for the lifetime of the customer

Any organisation managing customer data has a responsibility to each and every individual. Therefore, the management and security of data differs little from the way a bank or financial institution would manage or maintain it. Nevertheless,

BOX 1

37 per cent of medium to small businesses, 49 per cent of medium to large businesses and 56 per cent of large businesses plan to carry out data quality checks with third party suppliers.[3]

there are cultural and practical factors that make the level of data management more challenging to achieve in the automotive industry.

One of these is the need to share data securely with supply chain partners. To do this, dealerships need to notify partners about consent parameters at the outset, and constantly review these parameters for compliance with evolving data protection regulations. It should also be clear in any data sharing relationship who the owner of the data is at each stage of the customer interaction, and for the entire lifetime of the customer.

Again, mapping data and data flows makes it possible to identify when and how data is shared with supply chain partners. This allows dealerships to identify and address security gaps, such as inconsistent data sharing processes, or scenarios where personal data is sent by e-mail over unencrypted connections. Dealerships can also look to extend this work and agree with supply chain partners that they apply the same high standards of data protection, once customer data has been shared with them.

Get technology security right

It goes without saying that taking a customer's information and leaving paper forms on a desk in full view is not something that we see very often these days. A large proportion of data is now being captured online and stored — so the IT team have a big part to play in reducing the risk of a data loss or data breach. Nevertheless, only 29 per cent of organisations currently have a formal, written cyber security policy in place according to Experian research.[4]

Regularly performing IT security audits of all systems involved in data collection, management, storage and sharing is key. By integrating standard IT security technologies, including firewalls for perimeter security and internet security and anti-virus software is an important step. There is also the opportunity, however, to consider limiting access to customer databases based on password security, biometric security or other technologies that ensure that only authorised dealership staff can access customers' sensitive personally identifiable information (PII) and financial information.

Define a data protection culture

Experian data shows that consumers place huge trust in brands, with 56 per cent of the general public comfortable to share general data with a company and 29 per cent happy to share their personal, identifiable information.[5] To ensure that this trust is reciprocated, dealerships need to develop a culture and environment that embraces data protection.

Executive level sponsorship should be established at the outset to help dealerships communicate the importance of safeguarding customers and their data across the workforce and embed data protection best practices as 'business as usual' activities. Dealerships can also introduce training programmes to support employees, improving their awareness and

understanding of what the GDPR is and how to work compliantly with data.

Additionally, dealerships can use policies to define how long customer data is retained within the organisation. Typically, it is necessary to retain data for long enough to answer customers' warrantee and repair requests, which normally means holding onto it for the entire customer lifecycle. Once the data has no useful purpose, however, it should be deleted to reduce the dealership's attack surface and minimise the risk of a data breach in the future.

Have trusted advisers managing the customer relationship

In the past, it was normal to fill vacancies with new staff based on personal recommendations and references; however, a number of industry trends are casting doubt on this method of recruiting.

First, insider fraud is increasing across the industry, with many documented cases of customers' funds being diverted into employees' accounts, for example. This shows the need for more rigorous reviews of any potential new starter at a dealership.

Second, large sums of money often change hands on the forecourt, especially when high value vehicles are being sold. Dealerships need to trust the integrity of their salespeople 100 per cent to prevent the risk of theft or fraud.

Third, the new GDPR requirements require a higher level of trust in employees and their integrity with regard to managing and sharing customer data securely and responsibly.

Owing to all these risk factors, dealerships may consider carrying out background checks on new starters and existing employees. These checks are easy to implement and offer both the organisation and its customers peace of mind about who is managing their payments and their personal information.

Technology advancements, regulation and future strategy

The European GDPR is a catalyst for significant changes that will not only make customers data more secure, but also help to improve dealerships' operating practices. For example, by investing in open architectures for data protection, dealerships can simply evolve their data protection platforms as security risks change, and connect into new data protection technology that is likely to come to market in the future. This kind of technology approach also gives dealerships the opportunity to consider utilising low cost platforms that use open application programming interfaces (APIs).

The flexibility provided by open technologies is particularly important given the rapidly changing regulatory landscape. For example, the Internet of Things (IoT) and Connected Car data are not covered by the GDPR today, but it may well be in the future. The fact is that Connected Car data in particular can reveal where a customer is at any point in time and, in some cases, even has credit card data attached to it. This makes it potentially sensitive and likely to be of concern to the regulator in the short- to mid-term.

If dealerships have invested in open infrastructure, they can easily connect into new systems and tools that bring IoT or Connected Car data into the data protection environment. This is one more example of how open technologies can future proof dealerships' technology investments and compliance strategies.

Build out a data breach readiness response plan

It is every business' worst nightmare to have customers' personally identifiable information compromised or stolen. In such cases, the risk to the rights and freedoms of individuals is deemed to

BOX 2

- 25 per cent of businesses do not know if their third party suppliers could notify them within 72 hours of a data breach.
- 33 per cent of businesses have experts at the ready to respond to a data breach[8]

BOX 3

Organisations we work with are still surprised about just how complex it can be to notify customers. This is further compounded when the supply chain is complex and it is a third party who has suffered a breach. Getting ready in advance and agreeing who is in control is the only sure way of ensuring a response reaches those affected, and in good time.

Jim Steven, Head of Data Breach Response, Experian

be high by the regulator, which means dealerships need to have an effective data breach response plan in place (Box 2) to reduce the risk of regulatory fines and reputational damage.

This is a message that has already been heard by the 78 per cent of UK businesses who already have data breach plans,[6] a figure that rises to 92 per cent for the largest businesses.[7] To further improve security, 43 per cent of large organisations also check supplier policies regularly.[6] When it comes to data protection, small, medium enterprises (SMEs) fare worse overall, with only 65 per cent planning for a data breach[6] and just 20 per cent checking their suppliers' policies regularly.[6]

For dealerships of all sizes, however, putting the customer front and centre of a plan will ensure that the right resources and expertise are in place to respond and notify the regulator and the individuals affected. This forward planning will not only help limit damage to customers and the dealership's reputation, it is also an opportunity to demonstrate commitment to customers' rights and the best way to safeguard them from becoming a potential victim of fraud in the future.

Some of the key aspects of data breach readiness are very practical steps, such as where the customer data is stored, checking if it is up to date and if the contact data is accurate. These crucial questions can help dealerships to understand how quickly those affected could be notified if required. If a dealership has large numbers of customers, it may also be necessary to consider whether the resources exist to handle a wave of incoming calls in the event of a data breach. If personally identifiable information has been lost or stolen, providing some form of remediation to those individuals affected (Box 3) — such as a credit or identity monitoring service — will provide customers with reassurance and potentially prevent them becoming a victim of fraud in the future. Planning for a data breach in advance is a step every organisation can take and is the right thing to do by the customer. Crucially, this approach means that a dealership can respond, reassure and recover with confidence.

References

1. Experian (2017) 'The growing trend of plans that don't safeguard businesses or their customers', White paper, January, p. 2, available at: https://www.experian.co.uk/assets/data-breach/white-papers/

experian-white-paper-readiness-vs-reality.pdf (accessed 29th March, 2019).
2. Experian (2017) 'Data breach: Supply chain risk', White paper, November, p. 3, available at: https://www.experian.co.uk/assets/data-breach/white-papers/data-breach-supply-chain-risk.pdf (accessed 29th March, 2019).
3. *Ibid*., p. 7,
4. Experian (2017) 'Data breach response: Readiness vs the reality', White paper, January, p. 3, available at: https://www.experian.co.uk/assets/data-breach/white-papers/experian-white-paper-readiness-vs-reality.pdf (accessed 29th March, 2019).
5. Experian, ref. 2 above, p. 8. Original data sourced from Censuswide data, commissioned by Experian in November 2017 (sample of 2000 UK consumers).
6. Experian, ref. 4 above, p. 5.
7. Experian, ref. 2 above, p. 4.
8. Experian, ref. 4 above, p. 6.

Unified surveillance systems: Data mining with PeekYou, GPS and facial recognition

Received: 30th January, 2019

Jessica Berger

MLIS, CIPM is an information security and privacy consultant whose research on the cyber security of drones informed drone data privacy policy design for the cities of Boston and Toronto. As a former paediatric registered nurse and health columnist concerned with computer-based patient records and confidentiality, she has published articles addressing online safety and the privacy rights of children and patients. She holds a master of library and information science and is an International Association of Privacy Professionals (IAPP) certified privacy programme manager. Through San Jose State University, Jessica served as an information security and records manager intern, originating the Wikipedia Library's privacy programme including security measures around integrated apps, encryption protocols, social media and virtual private network (VPN). She currently offers cyber security analysis and privacy programme consulting in Massachusetts and Connecticut.

Tel.: (413) 527 1869; E-mail: jsberger@tutanota.com; Website: https://privacyprotectordotblog.wordpress.com/

Abstract Unified surveillance systems threaten to unlock a portal to mass surveillance, swift round ups, incarceration and deportation. By combining data from an array of people tracking technologies, governments and corporations can now instantly locate and monitor entire populations in real time. This paper unveils the mechanics of how these technologies are bound together within the fabric of our daily lives, silently invading our privacy. Research methods include examination of more than 29 articles, an exploration of PeekYou and an interview with an accomplished transportation professional from Seattle, Washington. Privacy-by-design offers the promise of freedom from constant corporate and government scrutiny. Ongoing coerced assent to mass surveillance need not remain our global fate. The United States Constitution, in particular the Third, Fourth and Fifth Amendments, offers citizens protection from the systematised misapplication of these invasive and largely covert programmes. As such, it is never too late to alter the course of history so that we amplify these technology's attributes while protecting all people from the abuse inherent in their utilisation.

KEYWORDS: artificial intelligence, bulk data collection, connected technologies, corporate surveillance, data capture, facial recognition software, geolocation data, government surveillance and privacy

INTRODUCTION

Since 2010, the Transportation Security Administration (TSA) has conducted warrantless surveillance of innocent air travelers targeted by the Department of Homeland Security (DHS). Merely sweating too much can land you on the DHS watch list.[1] Mass surveillance in the name of security crowds both our skies and highways. Linking Massachusetts Department of Transportation (Mass DOT) tracking policies to global positioning systems (GPS), Hausman,[2] explains that geo-location tracking involves multiple

modalities, of which GPS is but one. Dunn found that 'Google is tracking your location, even when the setting is turned off'.[3] Connected technologies provide ample opportunity for corporate and government spying. In fact, 100,000 people's movements were traced through cell phone data triangulation.[4] Amazon has profited from its facial recognition system, Rekognition, by selling facial images to the police. The American Civil Liberties Union (ACLU) explains, 'With Rekognition, a government can now build a system to automate the identification and tracking of anyone.'[5]

PURPOSE

This paper introduces the fundamentals of facial recognition software, GPS and PeekYou, with a focus upon the interwoven impacts these inventions have upon our privacy and freedom. Do these combined technologies really enable the tracking and monitoring of entire populations in real time? If so, this triumvirate could readily coordinate the distribution of goods and services to needy populations. In contradistinction, the pandemics of mass incarceration and slavery[6,7] position these technologies as harbingers of doom for those concerned with privacy and freedom. We will explore how these modalities function as a unified surveillance system. In response, we will highlight how the US Constitution offers citizens protection from the systematised misapplication of these invasive and largely covert systems.

DEFINITIONS AND HISTORY OF PEEKYOU, GPS AND FACIAL RECOGNITION SOFTWARE

General definitions related to data mining

'Chelsea Manning has compared life in the US to her time in prison because of surveillance systems, cameras and the presence of police'.[8]

Basic data mining terminology is essential to understanding how disparate information retrieval systems work together synergistically to track people. When people's movements, images, keystrokes or conversations are simply monitored and recorded, this is referred to as *data capture*.[9] Once data is captured by one system, it can be added to data from other systems. Combining data from different sources is referred to as *aggregation*.[10] Matching algorithms applied to this aggregated information pinpoint a specific individual. This process is called *data matching*.[11] Finally, we stumble into 'Data mining, (which) does not discover already existing information; (but) generates new information about existing information'.[12] Privacy, liberty, and legal concerns arise even in the face of claims that this information is anonymised. In separate reports, Weisinger,[13] Barocas[14] and Leonard[15] quoting Cavoukian and Jonas[16] concur that re-identification of individuals is inherent in the data mining process as numerous pieces of personal information are clustered into a set. This data is so specific in terms of activities, addresses and habits that it will re-identify individuals.

History and definition of PeekYou

Bell[17] interviews PeekYou advisory board member Marshall Sponder, who reveals what can be done by combining text analytics with geo-location data picked up from mobile devices. Sponder aspires to a career in politics that leverages his awareness of data mining.[18] Meanwhile, PeekYou went from Beta in 2006,[19] to live in 2007, claiming '50 million users'. This people search engine crawls the web, collocating online materials about a single individual into one profile. The patented algorithm allows 'PeekYou's search engine (to) calculate(s) the likelihood of any URL being associated with an individual'.[20]

Background and definition of Global Positioning System (GPS)

People connected to smartphones associated with Verizon can be found in real time because the phone's whereabouts can be triangulated 'every 7 second(s)'[21] in relation to the nearest mobile phone tower.[22–24] The inclusion of 'black boxes' throughout a car's operating systems allow for monitoring of its route, miles and speed.[25] Related to, but different from GPS, is the Massachusetts Department of Transportation's (Mass DOT) collection of data about drivers on toll roads. Here, automobiles pass under gantries that scan the EZPass to bill the driver for the toll. If drivers do not have an EZPass, a camera takes a photo of the car licence plate and a bill is sent to the driver.[26]

Linking Mass DOT tracking policies to GPS, Hausman explains that geolocation tracking involves multiple modalities, of which GPS is but one; Hausman relates that GPS 'consists of orbiting satellites…detected by a GPS receiver in a cell phone or other mobile device…(with) a clear line of sight to four or more satellites.'[27] Whereas in 2014, only under certain circumstances could a person be immediately found with near precision, now there exists an entire industry called People Tracking Technology.[28,29]

Radio frequency identification devices (RFIDs), vision analytics, raspberry pi and 3D spatial learning are just four of the 15 people tracking technologies available to retailers, government and employers.[30]

History and definitions of facial recognition

Innocent football fans attending the 2001 Tampa, Florida Super Bowl were the unwitting subjects of mass surveillance through facial recognition technology. These sports enthusiasts' faces became images in a *data capture*. Their facial images were *aggregated*, to then become subjects of *data matching* against a criminal database by law enforcement.[31,32] This covert manoeuver was purportedly conducted to assay the efficacy of facial recognition technology as a counter-terrorism tool. Nineteen sports fans faces matched up with those of 19 'petty criminals'.[33] The fact that an entire stadium of people were experimented upon without their consent and treated like 'mug shots' is itself a controversial topic. Agre explains that facial recognition software not only matches faces against a database of other faces, it also 'can extract facial expressions.'[34]

Now, new facial recognition technologies can even identify emotions.[35] Friedland reports that in the future, our government could spy on our faces from afar.[36] Given the ubiquity of surveillance cameras, it becomes impossible to avoid being tracked and monitored. Agre points out that unlike other biometrics, facial recognition technology renders individuals powerless with regards to choice.[37]

AGGREGATION OF DATA FROM PEEKYOU, GPS, AND FACIAL RECOGNITION APPS

By combining the information from PeekYou, GPS and facial recognition software, businesses can covertly uncover a new customer's preferences and assets. Hausman suggests that vendors might provide preferential service to those with rich profiles.[38] Barocas contends that this phenomenon predisposes to prejudicial treatment.[39] While shopkeepers may appreciate FaceFirst's identification of individuals with a history of shoplifting,[40] shoppers attempting to turn over a new leaf will find themselves pegged as a criminal in perpetuity. In fact, Event Technology's sales pitch lauds its ability to immediately identify people on watch lists through facial recognition.[41] Imagine having a medical issue that causes you to sweat, being put on a DHS watch list because of this 'unusual behaviour' and then subsequently being barred from events.

AN INTERVIEW: SEATTLE TRANSPORTATION PROFESSIONAL DISCUSSES PRIVACY BY DESIGN TRAFFIC TECHNOLOGY INNOVATIONS

Fortunately, municipal leaders are working to safeguard the privacy of citizens through adoption of new transportation technologies offering privacy by design. The following interview with an innovative transportation professional from Seattle, Washington, showcases this crucial development.

Q: Is there any way to remain anonymous while driving on toll roads?
A: There are two ways to look at this: Primarily, the answer is 'NO', at least not legally. The vehicle will be identified either through reading the toll tag or through a license plate reader (most systems incorporate these). The only way around this would be to not have a tag or a license plate. However, from another perspective, the answer is 'YES' in that inherently the system cannot positively identify the driver, only the registered owner of the vehicle or account.
Q: Is your license plate scanned if you don't have an EZPass?
A: This is the case for most systems. If there is a lane or option on the signage for 'pay by mail', they're reading the license plate and billing the registered owner.
Q: Is a record of when your EZPass travels below the gantries recorded and held?
A: Yes, at least until payment is received.
Q: Could you describe any privacy-by-design transportation innovations in your area?
A: There is a related situation regarding license plate readers. Wherever you see a system that says, 'X minutes to <someplace>', it's doing that using the technology. In this case it's a two-agent system: The first agent reads license plates optically. It doesn't store any image; it just directly creates a record of that license plate which is assigned a temporary record ID. That system holds the association between the license plate number and the ID, but not any geo-location or timestamp. The second agent receives only the record ID and adds the geo-location (based on the reader that did the detection) and timestamp. That way, one agent only knows the license plate number, but not where or when it was seen, and the other only knows where and when but not who. I would expect that the information would have a TTL (time to live) and be automatically deleted after that expires. This comes up regarding 'Connected Vehicles', too. The standard (for when/if it ever goes live) uses security certificates that are issued to vehicles that are only valid for 5 minutes after it is first used. In the same way, the certificate issuer and the system don't communicate the association between the certificate and vehicle to preserve anonymity.
(Transportation Professional from Seattle, WA, 9th May, 2019, personal communication)

SOCIAL AND PRIVACY ISSUES

Thompson and Thompson define privacy as 'The right to be let alone and the right of individuals to determine when, how, and how much information about themselves is released to others.'[42] Importantly, they emphasise that privacy includes freedom from intimidation to reveal personal secrets in order to attend to one's daily affairs.[43] Friedland notes that we must comply with numerous intrusive online privacy policies to merely conduct our basic banking and healthcare needs.[44] The same is true for social media. Citizens must either comply with these policies or forgo banking, medical care, social media and more; however, there has been some regulatory pushback.

While PeekYou profited from aggregating personal data, they earned the dubious distinction of being subpoenaed by the Federal Trade Commission (FTC). Timberg explains how the FTC highlighted PeekYou's unscrupulous practice of 'onboarding (which) allows markets to load offline information — from magazine subscriptions, store loyalty cards or government records — into cookies that digital advertisers use to target consumers for pitches.'[45]

Connected technologies provide ample opportunity for government and corporate spying. On the corporate front, Alim et al. (2017), of The Electronic Frontier Foundation, found that Google collects and stores vast swaths of personal information about K-12 students who are forced to use their services as a condition of matriculation.[46] Tunick explains how US police favour GPS and video cameras over traditional stakeouts.[47] Friedland introduces the loophole concept of 'The Silent Subpoena'.[48] This legalises warrantless searches in which the government utilises aggregated data.[49] Warrantless searches are also possible through the use of *zero days* — intentional corporate cyber security flaws through which politicians breach encryption.[50] The legality of these surveillance tactics does not render them ethical or constitutional. Tunick corroborates this view, noting parallels between these techniques and those deployed by the Gestapo.[51]

The combined powers of PeekYou, GPS, and facial recognition do indeed allow the government to track mass groups of people in real time. In fact, 100,000 people's movements were traced through mobile phone data triangulation. Bayir reports that in addition to location tracking, the data from these same people's mobile phones was used to extrapolate information about their habits, friends and activities.[52] Data mining could potentially endanger innocent people. For instance, if one is mistaken for a terrorist, it could be possible to be placed on a no-fly list or treated to legalised brutalities such as extraordinary rendition.

DEMOCRACY AND CONSTITUTIONALITY OF THE TECHNOLOGIES

Friedland cites the Third Amendment's protection of citizens from US military home intrusion.[53] Today, this amendment should protect US citizens from government's bulk collection of our online data — much of which is created in citizens' own homes. According to findlaw.com, the Fourth Amendment guarantees the Constitutional right to personal privacy at home, at work and about our persons.[54] Henceforth, the police must obtain a search warrant before investigating a citizen's private affairs.[55] Yet, there are no search warrants attached to licence plate scanners, GPS signal data, facial recognition software, surveillance cameras and PeekYou. Have you ever heard this rap: 'You have the right to remain silent. You have the right to speak to an attorney. Anything you say can and will be used against you'? These rights embody the Fifth Amendment Miranda Rights, which provides those in police custody with protections from self-incrimination.[56] Fifth Amendment rights are subverted by the ubiquity of surveillance cameras when surreptitious information from facial recognition programmes establishes culpability. Friedland relays the import of the Miranda doctrine with regards to testimony obtained under duress.[57] Since we are forced to have our faces scanned, are tracked by GPS and monitored online, one could say that the entirety of our interconnected lives should be protected by the Third, Fourth and Fifth Amendments.

CONCLUSION

The combined surveillance capabilities of GPS, facial recognition apps, people tracking technologies and PeekYou are already covertly embedded within the fibre of our society. Nevertheless, by upholding preexisting Amendments and introducing new privacy legislation specific to these technologies, we can curtail the insidious spread of coerced assent to mass surveillance.[58,59] These technologies hold the potential for an unprecedented coordination of resources for the greater good. It is never too late to alter the course of history such that we amplify these technology's attributes while protecting all people from the abuse inherent in their utilisation.

References

1. NPR (2018) 'Former TSA administrator discusses "Quiet Skies" surveillance program', All Things Considered, available at: https://www.npr.org/2018/07/30/634087400/former-tsa-administrator-discusses-quiet-skies-surveillance-program (accessed 13th May, 2019).
2. Hausman, S. (2014) 'Proliferation of tracking technology reaps security benefits, sparks privacy concerns', SecurityInfoWatch.Com, available at: https://www.securityinfowatch.com/alarms-monitoring/asset-and-gps-tracking/article/11350683/dr-steven-hausman-examines-the-benefits-and-pitfalls-of-the-proliferation-of-tracking-technology (accessed 9th May, 2019). See p. 1, para 1.
3. Dunn, J. E. (2018) 'Google is tracking your location, even when the setting is turned off', *Naked Security by Sophos*, available at: https://nakedsecurity.sophos.com/2018/08/15/google-is-tracking-your-location-even-when-the-setting-is-turned-off/ (accessed 9th May, 2019).
4. Bayir, M. A. (2010) 'Enabling location aware smartphone applications via mobility profiling', Doctoral dissertation, State University of New York at Buffalo. See p. 7 & p. 9.
5. Kan, M. (2018) 'Will Amazon's facial-recognition tech enable mass surveillance?', *PC Mag*, available at: https://www.pcmag.com/news/361346/will-amazons-facial-recognition-tech-enable-mass-surveillance (accessed 9th May, 2019).
6. Gibbons, A. and Marsh, S. (2018) 'Chelsea Manning says life in the US is like being in prison', *The Guardian*, available at: https://www.theguardian.com/us-news/2018/oct/01/chelsea-manning-life-us-like-prison-uk-visit (accessed 8th May, 2019).
7. Ochab, E. U. (2018) 'Human trafficking is a pandemic of the 21st century', *Forbes*, available at: https://www.forbes.com/sites/ewelinaochab/2018/07/26/human-trafficking-is-a-pandemic-of-the-21st-century/#4b0102906195 (accessed 8th May, 2019).
8. Gibbons & Marsh, ref. 6 above, p. 1.
9. Barocas, S. (2014) 'Panic inducing: Data mining, fairness, and privacy', PhD dissertation, New York University, USA.
10. *Ibid*.
11. *Ibid*.
12. *Ibid*., p. 19.
13. Weisinger, D. (2017) 'Big data and data anonymization: is anonymization an illusion?', Formtek, para 4, available at: https://formtek.com/blog/big-data-and-data-anonymization-is-anonymization-an-illusion/ (accessed 10th May, 2019).
14. Barocas, ref. 9 above, p. 8.
15. Leonard, P. (2014) 'Customer data analytics: Privacy settings for "big data" business. *International Data Privacy Law*, Vol. 4, No. 1, pp. 53–68, doi: http://dx.doi.org/10.1093/idpl/ipt032. See p. 63.
16. Cavoukian, A. and Jonas, J. (2012) 'Privacy by design in the age of big data', 8th June, available at: https://jeffjonas.typepad.com/Privacy-by-Design-in-the-Era-of-Big-Data.pdf (accessed 9th May, 2019).
17. Bell, G. (2012) 'Interview with Marshall Sponder, author of social media analytics', *Strategic Direction*, Vol. 28, No. 6, pp. 32–35, available at: https://www.emeraldinsight.com/doi/abs/10.1108/02580541211224102 para 1 (accessed 10th May, 2019).
18. *Ibid*., para 15.
19. Hussey, M. (2007) 'PeekYou emerges from stealth mode with 50 million profiles in beta, providing easy and efficient people search capabilities', *PRNewswire*, 17th July, available at: http://michaelhussey.com/2007/07/17/peekyoucom-beta-launch/ (accessed 7th May, 2019).
20. *Ibid*.
21. Hardin, N. (2017) 'Cell phone surveillance: Tactics, litigation, and next steps', p. 9, para 1, available at: https://vae.fd.org/sites/vae.fd.org/files/training/April_2018/03%20Cell%20Phone%20Surveillance.pdf (accessed 10th May, 2019). Note: this is an unpublished PDF uploaded by Nicole Hardin for public use; she is an assistant federal public defender for the Middle District of Tampa, FL, USA.
22. McMurrer, S. and Newburn, M. (2018) 'Next Generation 9-1-1 update', Board of Supervisors IT Committee Meeting, available at: https://www.fairfaxcounty.gov/boardofsupervisors/sites/boardofsupervisors/files/assets/meeting-materials/2018/oct09-it-next-generation-911-presentation.pdf (accessed 9th May, 2019).
23. Friedland, S. I. (2015) 'I spy: The new self-cybersurveillance', *Washington and Lee Law Review*, Vol. 72, No. 3, pp. 1459–1501, available at: https://heinonline.org/HOL/LandingPage?handle=hein.journals/waslee72&div=35&id=&page=&t=1557365135 (accessed 9th May, 2019), see p.1471.
24. Hausman, ref. 2 above.
25. Friedland, ref. 23 above, p. 1463 & pp. 1473–1474.
26. MassDOT (2016) 'All electronic tolling now activated along I-90', 29th October, archives, available at: http://blog.mass.gov/transportation/massdot-highway/all-electronic-tolling-now-activated-along-i-90/ (accessed 9th May, 2019).
27. Hausman, ref. 2 above, p. 1, para 1.
28. Max, R. (2018) 'People tracking: 15 technologies in 2018 — What's people tracking technology? *Behavior Analytics Retail*, available at: https://behavioranalyticsretail.com/technologies-tracking-people/ (accessed 10th May, 2019).
29. Hausman, ref. 2 above.
30. Max, ref. 28 above.
31. Celentino, J. C. (2016) 'Face-to-face with facial recognition evidence: admissibility under the post-Crawford Confrontation Clause', *Michigan Law Review*, Vol. 114, No. 7, pp. 1317–1353. See p. 1322.
32. Agre, P. (2001) 'Your face is not a barcode — Arguments against automatic face recognition in public places', available at: http://polaris.gseis.ucla.edu/pagre/bar-code.html (accessed 10th May, 2019).
33. Celentino, ref. 33 above, p. 1322.
34. Agre, ref. 32 above, para 11.
35. Gemalto (2018) 'The top 7 trends for facial recognition in 2018', available at: https://www.gemalto.com/govt/biometrics/facial-recognition (accessed 10th May, 2019).
36. Friedland, ref. 23 above, p. 1489.

37. Agre, ref. 32 above, para 7.
38. Hausman, ref. 2 above, p. 1, para 4.
39. Barocas, ref. 9 above, p. 8.
40. Hausman, ref. 2 above.
41. Zenus (2018) 'Facial recognition and events — a comprehensive guide', Event — The Intelligence Platform to Run Better Events, available at: https://www.eventmanagerblog.com/facial-recognition-guide-2018 (accessed 9th May, 2019).
42. Thompson, C.W. and Thompson, D. R. (2007) 'Identity management', *IEEE Internet Computing*, Vol. 11, No. 3, pp. 82–85, doi: http://dx.doi.org/10.1109/MIC.2007.60.
43. *Ibid*.
44. Friedland, ref. 23 above, p. 1475.
45. Timberg, C. (2014) 'What do firms know about you? FTC would pull back the curtain', *The Washington Post*, 28th May, p. 2.
46. Alim, F., Cardozo, N., Gebhart, G., Gullo, K. and Kalia, A. (2017) 'Spying on students: School-issued devices and student privacy', The Electronic Frontier Foundation, available at: https://www.eff.org/wp/school-issued-devicesand-student-privacy (accessed 10th May, 2019).
47. Tunick, M. (2009) 'Privacy in public places: Do GPS and video surveillance provide plain views?', *Social Theory and Practice*, Vol. 35, No. 4, pp. 597–622. See p. 597.
48. Friedland, ref. 23 above, p. 1466.
49. *Ibid*.
50. *Ibid*., pp.1465–1466.
51. Tunick, ref. 47 above, p. 611.
52. Bayir, ref. 4 above.
53. Friedland, ref. 23 above, p. 1498.
54. FindLaw.com (2016) 'Search and seizure and the Fourth Amendment', Thomson Reuters, available at: http://criminal.findlaw.com/criminal-rights/search-and-seizure-and-the-fourth-amendment.html (accessed 10th May, 2019).
55. Friedland, ref. 23 above, p. 1466.
56. 'What is the definition of the Miranda Doctrine?' (2016) IAC Publishing Labs Company, available at: https://www.reference.com/government-politics/definition-miranda-doctrine-530313a00f9e5aee (accessed 10th May, 2019).
57. Friedland, ref. 23 above, p. 1492.
58. Hausman, ref. 2 above, p. 1, para 1.
59. Agre, ref. 32 above, para 12.

Feeling fine! Harmonisation and inconsistency in EU supervisory authority administrative fines

Received: 15th February, 2019

Arye Schreiber
is a dual-qualified lawyer, a data protection consultant and CEO of MyEDPO. Arye advises a broad range of clients, from early stage startups to public companies, NGOs, universities and government agencies. Arye has degrees in law, including MA (Cantab), LLM (University of London) and an MBA (Stanford) degree. In addition to his professional work in data protection, Arye has worked for over a decade in corporate law advising and representing tech corporations. Arye has published many articles in top tier law journals, and has been cited in the leading publications in privacy law. Arye lectures regularly in professional data protection fora, and holds CIPP/E and CIPM certifications, and is a Fellow of Information Privacy (FIP) of the IAPP.

Tel.: +44-203-870-3376; E-mail: arye@myedpo.com

Abstract GDPR has a stated goal of harmonisation in general, and of penalties in particular. This article demonstrates that under GDPR penalties, and especially fines, are inconsistently applied across EU member states, and that GDPR has left many of the most important topics relating to fines to member state legislation. The article starts by showing that the One-Stop Shop mechanism actually incentivises forum-shopping. Next, it is shown that the method of calculating fines is inconsistent and unsettled. Different language versions of GDPR lead to different conclusions as to how to calculate an undertaking's revenue, and the meaning of an undertaking is neither entirely consistent within GDPR itself, nor across member states. The role of regulators is likewise unclear, and in some member states the regulators do not even have the power to impose an administrative fine under GDPR. The role of non-regulators, such as data subjects and representatives of classes of data subjects similarly lacks consistency across member states. Public bodies are another area of disharmony between member states: the scope of applicability of GDPR to public bodies is a matter for member state legislation, and the outcomes are in fact different across member states. Additional areas discussed include: the responsibility and liability of directors and officers of a company; the enforceability of a contract for insurances against GDPR fines; choice of law clauses as governing data being processed under GDPR; and issuance of warnings prior to imposition of fines. In all these areas, GDPR itself and member state law is inconsistent and is far from harmonised. Finally, the role of the economic model of the infringing party in calculation of the applicable fine is unsettled, and is left to member states, and is therefore similarly at odds with a goal of harmonisation.

KEYWORDS: administrative fines, harmonisation, supervisory authorities, insurance, directors' liability, public bodies

INTRODUCTION

The General Data Protection Regulation (GDPR) has introduced a new regime of administrative fines and other sanctions to EU data protection law and practice. Member state laws, supervisory authority

opinions and guidance, and the former Article 29 Working Party (WP29) guidelines, have all contributed to the development of the new powers vested in the supervisory authorities. This paper identifies some of the key emerging issues in this area: how and why fines are imposed, how they are assessed, how the risks of fines can be managed, who may be fined and more. As emerges from the paper, many of these topics are unsettled and, between various member states, inconsistent.

The GDPR's recitals lay out the legislative purposes of the GDPR. The Data Protection Directive 95/46/EC (the 'DPD') sought to 'harmonise' data protection (Recital 3) among member states, 'but it has not prevented fragmentation in the implementation of data protection across the Union' (Recital 9). The solution is in passing the GDPR; 'Consistent and homogenous application of the rules for the protection …of personal data should be ensured throughout the Union' (Recital 10). This includes not only the applicable law, but also the penalties for its violation: 'In order to strengthen and harmonise administrative penalties for infringements of this Regulation, each supervisory authority should have the power to impose administrative fines' (Recital 150).

Even within the area of administrative fines under the GDPR, there are unresolved discrepancies that challenge the harmonisation goals of the GDPR, and threaten the predictability and effectiveness of administrative fines across member states. Ten such areas are briefly detailed now.

EUROPEANISATION OF DATA PROTECTION AND THE RACE TO BE THE ONE-STOP SHOP

DPD Article 24 empowered member states to provide for sanctions for violations, but did not so much as mention fines. Administrative fines were levied under the member states' acts giving effect to the directive. Under the DPD fines, in so far as they were imposed at all, were localised. The maximum fine was set by the implementing laws in each member state. In Romania, the maximum fine was 500 million Lei, which, after the 2005 conversion, is 50,000 Romanian Lei, currently approximating €10,500. In Belgium, for example, it was €600,000, over 50 times greater than the Romanian maximum. The scope and effectiveness of administrative fines was entirely the prerogative of the member state, and indeed nothing in the DPD required administrative fines as such. The GDPR has made a dramatic departure from that model, grants supervisory authorities the power to issue administrative fines (with exceptions, discussed below), and moreover does so in a way that ostensibly promotes harmonisation across the member states. Under the GDPR, the maximal fines, the criteria for assessing fines, and even the scope of the infringements to which the fines relate, are Europeanised.[1] This is a part of the Europeanisation of data protection law under the GDPR. As noted by Lynskey,[2] the GDPR introduces several novel structures into the data protection regime; one that Lynskey focused on in particular is the administrative fines. Lynskey queried:

> once the consistency mechanism is engaged it is solely the lead authority that addresses a final decision to the data controller. It would also therefore seem logical to assume, although not expressly stipulated by the GDPR that it is solely that lead authority that can impose an administrative fine on the data controller (and therefore that each supervisory authority that is an addressee of the EDPB [European Data Protection Board] decision cannot impose an administrative fine on its own territory). Given the enhanced administrative fines foreseen by the Regulation, which are arguably now criminal in nature as a result of their severity, one could query whether the imposition of sanctions by multiple Member States would comply with the principle of *ne bis in idem*.

The GDPR does in fact address this, simply stating that 'the imposition of criminal penalties for infringements of such national rules and of administrative penalties should not lead to a breach of the principle of *ne bis in idem*, as interpreted by the Court of Justice.'[3] In other words, the GDPR clearly answers Lynksey's question in the negative. How exactly that will be carried out in practice remains to be seen. This is particularly interesting as the supervisory authorities imposing an administrative fine collect the fine to the coffers of that member state, according to member state law. Thus, if the French authority imposes a €50M fine on Google (as is discussed below), that is a €50M boon to the French treasury. Perhaps the relevant supervisory authority extracts or justifies its budget based, inter alia, on its ability to finance itself, and more than finance itself, through the fines it imposes. This in turn will clearly lead to a rush to impose fines, especially on the biggest companies and deepest pockets, such as Google.[4] WP29 has rightly stated that a 'harmonized approach to administrative fines in the field of data protection requires active participation and information exchange among Supervisory Authorities.'[5] As a result, one can expect that some supervisory authorities will become known as more business friendly, others less so, with some jurisdictions thereby becoming preferred locations under the one-stop shop mechanism (see Recitals 127–128).

THE NUMBER AND SIZE OF THE FINES

In its first year, the GDPR has dramatically increased both the number of investigations and also the magnitude of the fines, and this even with respect to infringements that took place under the DPD. Regarding the number of complaints and fines: according to the EDPB, in the 8 months since the GDPR came into force, there were 95,180 complaints filed with data protection authorities.[6] This represents a very significant increase in the number of complaints and investigations at the data supervisory authorities since the GDPR came into effect. For example, the Information Commissioner's Officer (ICO) has recorded a 133 per cent increase in the number of data protection cases it is currently handling,[7] compared with its pre-GDPR caseload. The number of fines issued in total is clearly not yet very high, because of the processing period of fines, but reports indicate that as of end of January 2019, there have been 91 fines imposed under the GDPR.[8]

Interestingly, the size of the fines has increased, and the GDPR seems to have had an effect even on fines issued under the DPD. For example, several files that were under investigation by the ICO under the DPD and Data Protection Act (DPA) 1998 were concluded after 25th May, 2018, when the GDPR was already in effect, and in two of those cases (Equifax[9] and Facebook[10]) the ICO imposed the maximum available fine under DPA 1998 — namely £500,000[11] — which it had never previously done. Regarding one of these fines, the Information Commissioner, Elizabeth Denham, said: 'We considered these contraventions to be so serious we imposed the maximum penalty under the previous legislation. The fine would inevitably have been significantly higher under the GDPR.'[12] This indicates that since under the GDPR, which was in force at the time this fine was imposed, the fine could have been potentially very much higher, the largest fine possible under the DPA 1998 no longer seems large, and was therefore imposed.

There are several open questions as to how fines are calculated under the GDPR. One ongoing argument between violators and authorities is the identity of the controller. Under the DPD, Facebook claimed that Facebook Ireland is the controller of data by Facebook in Europe; under the DPD, this view was promptly rejected by regulators and courts.[13]

Under the GDPR, Facebook's position is still less tenable, and the consequences for viewing Facebook, Inc., the US parent company, as the controller, or the undertaking in question, has very significant ramifications. GDPR Article 83(5) sets a maximum for an 'undertaking, up to 4% of the total worldwide annual turnover of the preceding financial year.'[14] Recital 150 explains that an 'undertaking' could mean an entire corporate group, a position substantiated by Court of Justice of the European Union (CJEU) case law. Here, it is noteworthy that the GDPR itself refers and defers to EU competition law in the definition of an undertaking.[15] Yet this is a comparison that is far from obvious:

> Competition law seeks to avoid economic harm, namely a negative impact on the parameters of price, quality, choice and innovation which affect efficiency or consumer welfare. While data protection law can also prevent such economic harm (for instance, by tackling information and power asymmetries), this is not the sole objective of the data protection rules. These rules also seek to prevent harm to fundamental rights, such as privacy, non-discrimination and freedom of association. There are therefore many circumstances in which data protection and competition law will have no mutual influence. For instance, even if an undertaking's data processing policy complies with competition law, it may entail a violation of the right to privacy. Equally, not all competition law concerns are data protection concerns: for instance, personal data processing plays no role in many markets. It is also important to acknowledge that the methods employed in each field are distinct and, in this regard, data protection law appears more akin to consumer protection law.[16]

The relationship between competition law and data protection law is in its infancy, and there have been several major mergers in recent years, motivated in large part by the personal data sharing post-merger, giving rise to new opportunities to explore the relationship between these previously almost unrelated areas of law. For companies driven largely by personal data, the use or alleged misuse of personal data may provide an opportunity to test the relationship between competition law and data protection law. This was the case in Facebook's investigation by the German Competition Authority (GCA) for anti-competitive products, which essentially required users to agree to extensive data sharing — an alleged abuse of both data protection rules and of competition rules. The GCA ultimately found that various data protection violations could stand in their own right as anti-trust violations, since they were exclusionary and constituted anti-competitive abuse.[17] In this way, fines might be levied for anti-competitive behaviour, based entirely upon violations of the GDPR. At the very least, in such cases, the definition of 'undertaking' and other GDPR provisions drawing on competition law, will make sense.

Returning to the case of Facebook, its topline revenue globally in its previous financial year was US$40.653bn. Four percent of that sum amounts to US$1.623bn. That is approximately 3250 times the £500,000 that the ICO recently imposed on Facebook. As noted by Voigt and von dem Bussche,[18] the term 'undertaking' is used elsewhere in the GDPR with a narrower meaning; in Article 4(19), the GDPR offers the following definition: ''group of undertakings' means a controlling undertaking and its controlled undertakings.' In this definition, an 'undertaking' is clearly not a corporate group. In a conflict between the recital (Recital 150) and an article of the GDPR (Article 4(19)), the latter ought to be definitive. Yet both WP29 and supervisory authorities have already assumed the broader, indeed broadest, interpretation of 'undertaking' in the context of imposition of administrative fines.[19]

Likewise the definition of 'of the preceding financial year' is not settled. The French law, for example, provides 'chiffre d'affaires annuel mondial total de l'exercice précédent', which is a year, not necessarily a 'financial year' (Article 83(5)). This, in the present example, is very much to Facebook's advantage, as in the last calendar quarter of 2018, Facebook announced dramatically increased earnings, at an annualised rate of about US$67bn.

Thus both 'undertaking' and 'financial year' may be applied in a variety of ways, with no requirement for these to be harmonised. Moreover, member states may specifically reserve the right to determine the definition of 'undertaking', 'turnover' and 'financial year', which the UK has done, for example.[20]

Evidently, some of the most important definitions regarding administrative fines are not settled and need not be harmonised. Even the purpose of the fines is still largely discretionary, as discussed presently. The relationship between competition law and data protection law, enshrined in the GDPR, is in its earliest stages, and the way these affect each other beyond the definition of 'undertaking' may have very far-reaching effects on administrative fines and beyond.

REGULATORS AND NON-REGULATORS AND THE PURPOSE OF ADMINISTRATIVE FINES

Data protection laws have existed for some time now, but there has long been a norm of corporations seeking 'to structure compliance by adapting to external mandates in ways that most easily achieve the appearance of legitimacy... focusing on easily visible indicators of compliance, rather than meaningful incorporation into firm decision making.'[21] In recent years, 'greater transparency around privacy failures has enabled nonregulators… to become credible enforcers.'[22] The GDPR has expanded the enforcement role of non-regulators in many ways. Most noticeably, data subjects have a host of access rights (GDPR Articles 12–23), with a resulting right to lodge complaints about data processors.[23]

Recital 129 sets out the powers that the GDPR gives regulators, such as 'investigation, corrective powers and sanctions, and authorisation and advisory powers, in particular in cases of complaints from natural persons …to bring infringements of this Regulation to the attention of the judicial authorities and engage in legal proceedings. Such powers should also include the power to impose a temporary or definitive limitation, including a ban, on processing.' The GDPR thus puts special emphasis on the role of the natural person,[24] in the enforcement process.[25] Article 80 goes further and empowers data subjects to mandate a 'not-for-profit body… to lodge the complaint on his or her behalf', which is a roundabout way for groups to sue for their collective privacy rights,[26] or for interest groups to pursue GDPR violations that go against their group values. The non-profit NOYB, an acronym for None Of Your Business, founded by Max Schrems, quickly became perhaps the most prominent of such groups. Schrems was famously instrumental in pre-GDPR legislation,[27] but NOYB filed multiple complaints on 25th May, 2018, including one that led to the largest ever data protection fine. The influence that non-profit data protection advocacy groups will have on the data protection landscape and on fines is without precedent, since they had no standing under the DPD, but from the experiences of the first months of the GDPR, it appears that the non-regulators' influence will be considerable.

The roles of regulators are also not entirely settled. Bennet and Raab detail[28] the varied roles that data protection authorities fill, including ombudsmen, auditors, consultants, educators, policy advisors, negotiators and finally enforcers. But the

role of enforcers is far from obvious. In the Republic of Kosovo (not currently an EU member), for example, the data protection authority does not have the power to impose fines for violations of data protection law.[29] Some EU member states likewise do not. GDPR Article 83 states that administrative fines are to be 'effective, proportionate and dissuasive'. Yet after detailing the powers of supervisory authorities to impose administrative fines, the GDPR envisages a reality in which supervisory authorities do not have the power to impose a fine. Article 83(9) states:

> Where the legal system of the Member State does not provide for administrative fines, this Article may be applied in such a manner that the fine is initiated by the competent supervisory authority and imposed by competent national courts, while ensuring that those legal remedies are effective and have an equivalent effect to the administrative fines imposed by supervisory authorities.

Recital 151 explains that certain member states have not granted supervisory authorities the power to impose a fine. Notably, supervisory authorities in Denmark and Estonia do not have the power to impose an administrative fine under their respective national law. The Recital explains how the administrative fines may, nonetheless, be imposed in a consistent manner:

> The rules on administrative fines may be applied in such a manner that in Denmark the fine is imposed by competent national courts as a criminal penalty and in Estonia the fine is imposed by the supervisory authority in the framework of a misdemeanour procedure, provided that such an application of the rules in those Member States has an equivalent effect to administrative fines imposed by supervisory authorities. Therefore the competent national courts should take into account the recommendation by the supervisory authority initiating the fine. In any event, the fines imposed should be effective, proportionate and dissuasive.

The outcome here is that the GDPR instructs the independent courts of a member state how to issue misdemeanor fines and administrative fines. It remains to be seen to what extent a national court considers itself bound by the recitals of the GDPR. The roles of different member state supervisory authorities are thus clearly not harmonised and not settled.

PUBLIC BODIES

Another important aspect of administrative fines yet to be clarified is how they will be applied to public bodies. This author is unaware of warnings and fines issued to public authorities, so far; however, aside from the practice of imposing fines on public authorities, there are some aspects of the GDPR left open on this matter, some matters of interpretation that are untested, and areas left to member state legislative discretion. Notably, there is member state discretion with respect to Article 83(7). The article states:

> Without prejudice to the corrective powers of supervisory authorities pursuant to Article 58(2), each Member State may lay down the rules on whether and to what extent administrative fines may be imposed on public authorities and bodies established in that Member State.

Different member states have reached very different conclusions in this regard. French data protection law applies the same administrative fine rules to public authorities as to non-public ones.[30] Others place limits: The UK DPA 2018 has reserved for the secretary of state the power to determine whether and to what extent administrative fines may be imposed on public authorities.[31] The Irish DPA 2018 specifically empowers the supervisory authority to impose

administrative fines on public authorities, but limits the fines to €1m.[32] In some member states this was a hotly debated topic; in the House of Commons, it was determined that certain public authorities, such as parishes, would be excluded from the definition of 'the term "Public Authority" under GDPR'.[33] In the Danish law's legislative history, this matter was likewise a subject of considerable wrangling:

> One of the main topics discussed with regards to the adaption of the GDPR to the Danish legal system was whether or not public authorities should be subject to fines. The Ministry of Justice had not decided on this in the first draft of the Data Protection Act that was published for public consultation. However, just before the first parliamentary reading the Ministry of Justice added a section in § 41 of the Data Protection Act that provides that public authorities can be sanctioned with fines as well as private actors. Under the first reading in Parliament, the Minister of Justice, Søren Pape Poulsen, stated that the government found it reasonable and fair to sanction public authorities as well as for private actors for infringements of the Data Protection Act and the GDPR.[34]

As is apparent from this small sample of approaches, different member states often do not have a settled jurisprudence on this, and there is no harmonised approach. There are also some interpretive matters that remain open.

Returning to Article 83(7): 'each Member State may lay down the rules on whether and to what extent administrative fines may be imposed on public authorities and bodies established in that Member State'. How is this sentence to be read? The word 'public' clearly qualifies 'authorities', but does it qualify 'bodies'? In other words, does this section apply to both public authorities, and to bodies, established in a member states, or does it apply to public authorities and public bodies established in that member state? Recital 154, for example, states: 'The reference to public authorities and bodies should in that context include all authorities or other bodies covered by Member State law on public access to documents.' There, it is made clear that 'public' does not qualify 'bodies'.[35] Conversely, Article 41(6) states that 'This Article shall not apply to processing carried out by public authorities and bodies'. In that case, it is clear that public qualifies 'bodies'.[36] At any rate, one wonders what the point of fining a public body could be. The administrative fines are collected by the state, which has the power to promptly return the fine to the public body, and the matter would be more complicated where quasi-national authorities, privatised national services, state-managed companies, local authorities and so on would be concerned. Member state jurisprudence and legislation differs in this area, and so there is little hope or aspiration for harmonisation.

INDIVIDUALS AND OFFICERS

Under the GDPR, any act of an employee, presumably acting in their capacity as such, can be attributed to the employer.[37] More complex is the imposition of fines on individuals for corporate violations. Some member states specifically authorise imposing sanctions on directors and officers of violating legal entities. For example, the Irish Data Protection Act 2018 provides that where a corporate entity has committed an offence, and it is proven to have been with the 'consent or connivance of, or to be attributable to any neglect on the part of' a director, manager, company secretary, officer or a person purporting to be one of those, then that person may be found guilty of and may be punished for that offence as if it were they who committed it.[38] UK law similarly provides that 'The director, manager, secretary, officer or person, as well as the body corporate, is guilty of the offence and liable to be proceeded against and punished accordingly.'[39]

The UK Supreme Court ruling in *Vestergaard* is interesting in this regard.[40] The case involved former employees of Vestergaard who started a competing business. One of the employees took along with him some trade secrets of Vestergaard and used them at the new business — a breach of his duty of confidence to this former employer. Another of the employees demonstrated that she had no knowledge, nor constructive knowledge, of the misappropriation of the trade secrets, and it was found that she was therefore not in violation of her duty of confidence. The case is pertinent to the GDPR in light of the requirements of Article 28(3)(b) that a processor must ensure 'that persons authorised to process the personal data have committed themselves to confidentiality.' Constructive knowledge, meaning that the person ought to have known, may be sufficient for a finding of breach of a duty of confidence. In *Vestergaard*, the courts were not in agreement, and in the future one may expect robust deliberation as to the role of constructive knowledge, vicarious responsibility and the boundaries of executive responsibility under the GDPR. These cardinal questions are within the realm of member state law, introducing further lack of harmonisation in the law and additional motive for forum shopping of sorts.

INSURANCE

Insurance is an additional area that has potential to influence greatly the world of administrative fines.[41] Put succinctly, administrative fines generally have a criminal law character, and they are intended to dissuade deviant behaviour. Yet, where a party has insured against such fines, that takes the sting out of the supervisory authority's tail.[42] Where there is insurance against fines, the fines will generally fail their essential purpose, calling into the question the point, if any, of imposing them. For this reason, in many states there is a public policy that restricts the validity of insurance against regulatory fines. Where insurance is valid, the premia paid to insurers essentially means that one perpetrator's fine is spread across many parties — the insurer's or underwriter's clients. The preparatory work of the insurers means that they ought to best understand the risks of each party, and can set the premium for each insured party to match their chances of being fined. Thus, in some respects the fines may be viewed as being amortised. But the public policy remains widespread across member states, that regulatory fines ought not to be insurable. Insurance of fines under the GDPR have not yet been the subject of case law, to the author's knowledge, and in the meantime there remains a public policy challenge to such insurance, meaning that this contract may be in violation of public policy, and may be found to be unenforceable.

One review of the area suggests that only Finland and Norway (the latter is not an EU member, but an European Economic Area (EEA) member) enable insurance against GDPR fines.[43] In Finland, this is qualified by the *mens rea* such that deliberate or gross negligence violations are not insurable; for some EU member states, the insurability of GDPR fines is unclear; and for most EU member states, GDPR fines are uninsurable. This entire area of law remains completely within the remit of member states, and is as yet untested.

CHOICE OF LAW

The factors listed above, such as director liability and the insurability of GDPR administrative fines, may subsequently influence the one-stop shop doctrine and the almost inevitable forum shopping. Under DPD, there was a significant forum shopping problem, and this ought to have been largely ameliorated by the broad and direct applicability of the GDPR's provisions. Nevertheless, insurability and director liability are matters for local law, and may therefore play into both choice of

law provisions and lead supervisory authority election. Brkan wrote, of the state of choice of law provisions under the DPD:

> The current doctrine and practice is divided regarding the question whether the parties to a contract can freely choose data protection law that is applicable for processing of data and for data protection breaches in a framework of this contract.[44]

She concludes that, in contrast with DPD's Article 4(1), the GDPR 'unifies EU data protection rules and hence no longer contains an overarching conflict-of-law provision.'[45] It is true that there is no overarching provision, but there is certainly a strong interest for parties to choose their applicable law.

WP29 has stated emphatically: 'The GDPR does not permit "forum shopping".'[46] Indeed, with respect to identifying the lead supervisory authority, there is a mechanism in place to ensure that the identity of the lead supervisory authority follows the jurisdiction of the main establishment. It appears, however, that this will not generally affect contractual terms. In other words, where a data protection agreement states that the laws of, say, Finland, will govern, then even if the lead supervisory authority of the processor and controller is the Commission nationale de l'informatique et des libertés (CNIL) in France, the contract, its terms, interpretation and so on ought to be governed by Finnish law, or at least by the Finnish data protection law.

Choice of law issues may be expected to raise some interesting challenges of this kind for supervisory authorities and courts, and it remains to be seen how they are to be contended with.

WARNINGS

Several supervisory authorities have issued various forms of warnings or notice to alleged violators of the GDPR. Indeed, the GDPR generally encourages or envisages a warning being issued prior to a fine being imposed,[47] and the published notices give some useful insight into the supervisory authorities' aims in the fines that may follow the warnings. The warnings offer some insight into the factors that may be considered by the supervisory authorities. The general factors are listed in the GDPR, but each supervisory authority may place the emphasis where they see fit.

There has been at least one case of a supervisory authority issuing a warning and notice to a non-EU entity. The UK's ICO issued a notice to AggregateIQ (AIQ), a company providing data and data analytics in connection with political campaigns. According to the ICO, AIQ had violated the lawfulness, transparency and fairness principles, and the purpose limitation and data minimisation principles. AIQ was issued with a warning,[48] and the ICO specifically considered whether 'the failure has caused or is likely to cause any person damage or distress', as required by section 150(2) of the DPA 2018. This introduces an additional element, in this case based on local law, of 'distress', as a factor in possible imposition of sanctions.[49]

In July 2018, CNIL issued warnings[50] to Teemo, Inc. and Fidzup SAS, two companies allegedly collecting and retaining geolocation data, and this is in contravention of the GDPR. The companies were warned to obtain consent and correct other data practices within a period of 3 months. In both cases, the companies had developed SDK – software development kits. This is a module of code that other app builders could include in their apps, and which collects various data — returning it to the app builders and owners, but also to the authors of the SDK, in this case, Teemo and Fidzup. Teemo's SDK collected geolocation data; Fidzup's enabled sending a targeted advertisement to any user who was near a Fidzup point of sale installation. CNIL found the alleged consent of the users

inadequate, the data retention excessive and information that ought to have been provided to the data subjects had not been provided as required. In this case, the violators were given 3 months to correct the situation; that appears to be more than enough time, and it seems from the CNIL notice that if indeed they apply the necessary fixes, they will be saved from a fine.

In addition to formal data protection authority warnings, there may be a variety of notices and warnings prior to a formal complaint and investigation. Microsoft was the subject of a fairly damning review, commissioned by the Dutch Ministry of Justice, of the data protection practices among offices of a major Microsoft customers — government institutions in Holland.[51] The review found several high-risk clusters of activity, noting that Microsoft's services as used by the Dutch government, reflect a lack of transparency, unlawful storage of special categories of personal data, lack of purpose limitation and more. This has acted as a warning to Microsoft, but has the potential for massive fines. The goal of making Microsoft services compliant may be better served in this case by the warning than by the fines.

It therefore appears, thus far in the evolution of administrative fines, that warnings ought to be taken very seriously, and that a full and effective response to a warning may entirely avoid a fine. The warnings issued may further elucidate likely considerations in the imposition of a fine, and generally offer a chance to rectify a violation, with exceptions, as discussed presently.

FINES IN PRACTICE

To date, supervisory authorities have imposed only a handful of fines under the GDPR. There are important indications of the various elements considered by the supervisory authorities, and these are elaborated on presently.[52] To that end, several instances are briefly discussed below.

The first fine issued under the GDPR and the new Bundesdatenschutzgesetz (BDSG) was issued by Landesamt für Datenschutzaufsicht (LfDI), the data protection authority of Baden-Wuertemberg.[53] The case involved a social dating site that had stored user passwords in clear text, inadequately protected the data and then suffered a breach. The LfDI emphasised that the company's cooperation and transparency in the investigation was exemplary, as well as its responsiveness to the LfDI's demands. These expressly motivated the LfDI to impose a relatively modest fine of €20,000. The commissioner, Dr Stephan Brink, said that the LfDI was not in a competition to impose the highest possible fine, but was tasked with protecting the rights of data subjects.

Another early GDPR fine was issued by the Austrian data protection authority. The Austrian Datenschutzgesetz, the data protection act, in section 11 specifically states that first-time infringements will generally be met with a warning.[54] Notwithstanding that, in the case of a betting establishment, the owner had installed CCTV which was filming public spaces outside the establishment. It was found that the business did not keep records of processing, did not delete data and had no justification for the same, and it did not give notice that there was video surveillance in place.[55] The fines for these infringements were €2400 for the first and €800 for the latter three, totalling €4800.[56]

More significant fines were levied in the case of a Portuguese hospital. Centro Hospitalar Barreiro Montijo[57] was fined €400,000, a very significant sum. Of this, €150,000 was for not adhering to the data minimisation principle; another €150,000 was for not putting in place appropriate technical and organisation measures to protect the data from unlawful access; and €100,000 for lack of data security measures commensurate with the risks of the data. Of particular interest is the first

of these infractions: the hospital had 985 users defined as 'doctors' in its central data system, but only had 296 actual doctors on the staff. Several considerations played into the relatively harsh fines imposed by the Portuguese data protection authority, the Commission nationale pour la protection des données (CNPD). One was the sensitivity of the data, namely medical data. Another was that the hospital did not report the breaches, but an investigation was begun after media reports of data mismanagement. The apparent willful neglect of the hospital with respect to data security and the knowing and egregious lack of data minimisation, were central in the imposition of this very significant fine. This fine underscores the *mens rea*'s role in determining the size of the fine imposed. Shortly after this case, another GDPR fine was imposed that was two orders of magnitude greater, and which depended less on the *mens rea* of the perpetrator and more on its economic model, and is discussed next.

THE VIOLATOR'S ECONOMIC MODEL

On 21st January, 2019 the French supervisory authority, CNIL, imposed the largest data protection fine yet, that of €50m, on Google.[58] The main violations by Google were that the consents obtained for their Android operating system were invalid, principally on account of the lack of specificity, with one act of consent for Android ultimately leading to personal data being used in Google Search, YouTube, Google Home, Google Maps, Playstore, Google Pictures and more. As a result, Google was actually collecting a vast amount of personal data with no lawful basis whatsoever, in violation of the GDPR. Google was thus essentially flaunting several of the central tenets of the GDPR, and did so with respect to a very large number of data subjects. In explaining the magnitude of the fine, CNIL noted (the following is an unofficial English language translation):

> The amount and the publicity of the fine, are justified by the severity of the infringements of the principles of transparency, information and consent; the violations are continuous not limited in time; and the economic model of the company is partly based on the ads' personalization.

In other words, several aspects played into the severity of the fine. Most notably, the severity of the violations of two pillars of data protection, that is, lawful processing and transparency. Likewise, the massive number of data subjects affected was an important factor. More interestingly here, is the last sentence quoted above: 'the economic model of the company is partly based on the ads' personalization.' This is in line with the expectation of Recital 149 that member states will legislate for penalties that include 'deprivation of the profits obtained through infringements' of the GDPR. French law did in fact previously include such a provision:

> The amount of the financial penalty provided under Article 45 Section I shall be proportional to the severity of the breaches committed and to the profits derived from said breach.'[59]

The amended French data protection law does not, however.[60] Rather, the French law simply references GDPR Article 83 for criteria that may be considered in imposing a fine.[61] Article 83(k) states: 'any other aggravating or mitigating factor applicable to the circumstances of the case, such as financial benefits gained, or losses avoided, directly or indirectly, from the infringement.' In this case, it was enough for CNIL to state that Google's business model involves making money from personalised advertising, to establish that there was an aggravating factor. In other words, CNIL

was not trying to impose an account of profits, but viewed the business model as an aggravating factor. Even if French law had in fact provided for an accounting of profits, presumably CNIL would still try to avoid demanding an account, which would make the case inordinately complicated and may ultimately show that the profit from the infringing act was very much less than the fines.

Recital 149 allows for member states to grant supervisory authorities the power to impose an account of profits for breaches of the GDPR. In the EU, Article 13 of the 2004 Enforcement Directive provides for an account of profits as a remedy in intellectual property cases. This has not typically included breach of confidence.[62] The remedy of account of profits is generally available in intellectual property violations, but the Court of Appeal in *Vestergaard* indicated that Article 13 of the Enforcement Directive applied.[63] This is connected with a broader issue — that of the 'propertisation' of data.[64] As data are increasingly viewed as property, data protection rights will increasingly be viewed as intellectual property rights. The propertisation of data, the availability of an account of profits as a remedy and the use of the perpetrator's economic model as an aggravating factor in assessing a violation, are all factors largely dependent on member state law, and have yet to be clarified in the context of the GDPR.

CONCLUSIONS

It has been shown above that though the GDPR sought to harmonise data protection laws generally and administrative fines in particular, there remain many considerations and factors that are untested, unsettled and generally open to member state law. These include warnings, the role of regulators and non-regulators such as public interest groups, choice of law, insurability, directors' and officers' liability and the use of the perpetrators economic model in consideration of fines. These and other factors lead to a conclusion that although the situation may be improved as compared with the DPD, the GDPR most certainly has not yet harmonised EU data protection law, and especially the fines imposed under the GDPR.

References and Notes

1. Giurgiu, A. and Larsen, T. A. (2016) 'Roles and powers of national data protection authorities', *European Data Protection Law Review*, Vol. 2, No. 3, pp. 342–352.
2. Lynskey, O. (2016) 'The Europeanisation of data protection law', Cambridge Yearbook of European Legal Studies, Vol. 19, pp. 252–286; see p. 274 for quote.
3. Recital 149.
4. One may thus contemplate the following: if the fines are imposed based on pan-European criteria, and on the basis of infringement across the EU, ought the fines therefore not be shared across the relevant member states, say in proportion to the number of relevant data subjects in each, or perhaps by some other mechanism between member states?
5. Working Party 29 Opinion 253, adopted 3rd October, 2017: Guidelines on the application and setting of administrative fines for the purposes of the Regulation 2016/679.
6. EU Commission, https://ec.europa.eu/commission/sites/beta-political/files/190125_gdpr_infographics_v4.pdf (accessed 10th May, 2019).
7. Khanna, M. (2018) '2018 GDPR Study', Prosults Ltd, available at: https://iapp.org/media/pdf/resource_center/2018_GDPR_Study.pdf (accessed 10th May, 2019).
8. DLA Piper (2019) 'GDPR data breach survey', February, DLA Piper LLP, available at: https://www.dlapiper.com/en/uk/insights/publications/2019/01/gdpr-data-breach-survey/ (accessed 10th May, 2019).
9. Denham, E. (2018) ICO, available at: https://ico.org.uk/media/2259808/equifax-ltd-mpn-20180919.pdf (accessed 10th May, 2019).
10. Denham, E. (2018) ICO, available at: https://ico.org.uk/media/action-weve-taken/mpns/2260051/r-facebook-mpn-20181024.pdf (accessed 10th May, 2019).
11. ICO (2018) 'ICO issues maximum £500,000 fine to Facebook for failing to protect users' personal information', 25th October, available at: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2018/10/facebook-issued-with-maximum-500-000-fine/ (accessed 10th May, 2019).
12. Denham, E. (2018) See her interview here: https://vimeo.com/296670132/d04544e679 (accessed 10th May, 2019).

13. Van Canneyt, T. (2015) 'The Belgian Facebook recommendation: How the nomination of a single EU Data controller is under fire', 20th May, available at: https://privacylawblog.fieldfisher.com/2015/the-belgian-facebook-recommendation-how-the-nomination-of-a-single-eu-data-controller-is-under-fire (accessed 10th May, 2019). See discussion in Bu-Pasha, S. (2017) 'Cross-border issues under EU data protection law with regards to personal data protection', *Information & Communications Technology Law*, Vol. 26, No. 3, pp. 213–228.
14. The French law, for example, provides 'chiffre d'affaires annuel mondial total de l'exercice précédent', which is a year, not necessarily a 'financial year' (Article 83(5)). This, in the present example, is very much to Facebook's advantage, as Facebook announced dramatically increased earnings for the last calendar quarter of 2018, at an annualised rate of about US$67bn. Note that member states may reserve the right to determine the definition of 'undertaking', 'turnover' and 'financial year'; see, for example, UK DPA 2018 s.159.
15. As noted by WP29 Opinion 253, footnote 4. Recital 150 refers to Articles 101 and 102 of the Treaty on the Functioning of the European Union; those provisions provide a wide interpretation of 'undertaking', based on competition law precedent. See Voigt P. and von dem Bussche, A. (2017) 'The EU General Data Protection Regulation (GDPR)', Springer, Cham, Switzerland, pp. 212–213; Golla, S. (2017) 'Is data protection law growing teeth? The current lack of sanctions in data protection law and administrative fines under the GDPR', *Journal of Intellectual Property, Information Technology and E-Commerce Law*, Vol. 8, No. 1, pp. 70–78, see p.76 for discussion and sources on the meaning of 'undertaking'. That article pre-dates WP29, and subsequent fines, all of which support the broadest interpretation of undertaking as, essentially, a corporate group.
16. Costa-Cabral, F. and Lynskey, O. (2017) 'Family ties: The intersection between data protection and competition in EU law', *Common Market Law Review*, Vol. 54, No. 1, pp. 11–50, para 2.2. See extensive discussion in Santos Silva, A. R. (2017) 'Towards the incorporation of privacy in EU competition law: How data protection harms can reduce the quality of goods and services', Master's thesis, Tilburg University.
17. Colangeo, G. and Maggionlio, M. (2018) 'Data accumulation and the privacy–antitrust interface: Insights from the Facebook case', *International Data Privacy Law*, Vol. 8, No. 3, pp. 224–239.
18. Voigt and von dem Bussche, ref. 15 above.
19. Golla, ref. 15 above.
20. UK DPA 2018 s.159.
21. Bamberger, K. A. A. and Mulligan, D. K. (2015) 'Privacy on the ground', MIT, Cambridge, MA, p. 28.
22. *Ibid*., p. 230.
23. A right that data controllers must specifically bring to the attention of the data subject; see Articles 12(4), 13(2)(d), 14(2)(e), 15(1)(f), and with respect to BCRs 47(2)(e).
24. In this recital the text refers to a 'natural person', but in Article 57(1)(f) the regulation refers to a 'data subject'. An important difference, but one which is not the focus of this paper.
25. See further Recital 141.
26. See at length Taylor, L., Floridi, L., and van der Sloot, B. (eds) (2017) 'Group privacy', Springer, Cham, Switzerland.
27. *Schrems v Data Protection Commissioner* [2014] IEHC 310 (18 June 2014).
28. Bennett, C. J. and Raab, C. D. (2006) 'The governance of privacy', MIT, Cambridge, MA, pp.133–143. More recently, and a comparison of the roles of DPAs in the DPD and GDPR compared: Giurgiu, A. and Larsen, T. A. (2016) 'Roles and powers of national data protection authorities' *European Data Protection Law Review*, Vol. 3, pp. 342–352.
29. See discussion in Zenjnullahu, N. (2016) 'Imposition of monetary sanctions as a mechanism for protection of personal data', *European Data Protection Law Review*, Vol. 1, pp. 80–90.
30. Loi Informatique et Libertes Act No. 78-17, 6 January 1978. Article 3.
31. DPA 2018 s.56(b)(1).
32. Irish DPA 2018, s.141.
33. UK DPA 2018 s.7, see Hansard, 9th May, 2018, column 790.
34. Overby, T. (2018) 'The Danish adaptation of GDPR', June, available at: https://blogdroiteuropeen.com. First reading of the Data Protection Act in Parliament, available at: http://www.ft.dk/samling/20171/lovforslag/L68/BEH1-20/forhandling.htm (Danish) (accessed 10th May, 2019).
35. Likewise in Recital 108, and Recital 158.
36. Likewise Recital 92.
37. See 'Short Opinion 2 – Supervisory Sanctions' of the German Data Protection Conference – DSK, available at: https://www.datenschutz-berlin.de/fileadmin/user_upload/pdf/publikationen/kurzpapiere/DSK_KPNr_2_Sanktionen.pdf (accessed 10th May, 2019).
38. Irish DPA 2018, s.146.
39. UK DPA 2018, s.198.
40. *Vestergaard Frandsen v Bestnet Europe* [2013] UKSC 31.
41. Baker, J. (2019) 'Data breach insurance: A three-part problem', IAPP, The Privacy Advisor blog, 29th January, available at: https://iapp.org/news/a/data-breach-insurance-a-three-part-problem/ (accessed 10th May, 2019).
42. There seems little punitive element in administrative fines under the GDPR, but member state laws add criminal stigma, prison sentences and other traditional elements of criminal sanction.
43. 'The price of data security: A guide to the insurability of GDPR fines across Europe', DLAPiper and Aon report, May 2018, available at: https://www.aon.com/attachments/risk-services/Aon_DLA-Piper-GDPR-Fines-Guide_Final_May2018.pdf (accessed 10th May, 2019).

44. Brkan, M. (2016) 'Data protection and conflict-of-laws', *European Data Protection Law Review*, Vol 2, No. 3, pp. 324–341.
45. *Ibid*., p. 340.
46. WP29 244 (2016) 'Guidelines for identifying a controller or processor's lead supervisory authority', 13th December, p. 7.
47. Recital 151, Article 58(2)(a).
48. Denham, E. (2018) Information Commissioner, Enforcement Notice, available at: https://ico.org.uk/media/action-weve-taken/enforcement-notices/2260123/aggregate-iq-en-20181024.pdf (accessed 10th May, 2019).
49. The First Tier Tribunal (Information Rights) overturned ICO's fine in *Scottish Borders Council v Information Commissioner* [2013] EA/2012/0212, since it was found that there was not a likelihood that harm would materialise. See discussion in Ceross, A. (2018) 'Examining data protection enforcement actions through qualitative interviews and data exploration', *International Review of Law, Computers & Technology*, Vol. 32, No. 1, pp. 99–117.
50. CNIL, available at: https://www.cnil.fr/fr/applications-mobiles-mises-en-demeure-absence-de-consentement-geolocalisation-ciblage-publicitaire (accessed 10th May, 2019).
51. Nas, S. and Roosendaal, A. (2018) 'DPIA Diagnostic Data In Microsoft Office Proplus', Ministry of Justice and Security for the benefit of SLM Rijk (Strategic Vendor Management Microsoft Dutch Government), available at: https://regmedia.co.uk/2018/11/16/microsoft-office-gdpr-fail.pdf (accessed 10th May, 2019).
52. There are fairly comprehensive surveys of fines and actions under DPD – see Ceross, ref. 49 above; less so under the GDPR, owing to the short time that the GDPR has been in force.
53. 'Kooperation mit Aufsicht macht es glimpflich' (unofficial translation: 'Cooperation with the Supervisory Authority makes it easy'), 2018, available at: https://www.baden-wuerttemberg.datenschutz.de/wp-content/uploads/2018/11/LfDI-Baden-W%C3%BCrttemberg-verh%C3%A4ngt-sein-erstes-Bu%C3%9Fgeld-in-Deutschland-nach-der-DS-GVO.pdf (accessed 10th May, 2019).
54. As does the UK Data Protection Act 2018, s.115(9). See also Austrian Federal Act for the Protection of Individuals with regard to the Processing of Personal Data (Data Protection Act — DSG), available at: https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10001597 (accessed 10th May, 2019).
55. Austrian data protection authority, Datenshutz behörde ('DSB'), available at: https://www.dsb.gv.at/documents/22758/116802/Straferkenntnis+DSB-D550.038+0003-DSB+2018.pdf/fb0bb313-8651-44ac-a713-c286d83e3f19 (accessed 10th May, 2019).
56. Compared with a merely symbolic fine of 1500 CZK (less than €100) issued in 2007 by the Czech data protection authority against Rynes. See at length the appeal to the Czech court: František Ryneš v Úřad pro ochranuosobníchúdajů (Office of the Protection of Personal Data), 113/2012, and of course the CJEU appeal of the same name, which overturned the court's annulling the fine; František Ryneš v Úřad pro ochranuosobníchúdajů C212/13.
57. Menezes Monteiro, A. (2019) 'First GDPR fine in Portugal issued against hospital for three violations', IAPP, available at: https://iapp.org/news/a/first-gdpr-fine-in-portugal-issued-against-hospital-for-three-violations/ (accessed 10th May, 2019).
58. CNIL, https://www.cnil.fr/en/cnils-restricted-committee-imposes-financial-penalty-50-million-euros-against-google-llc (accessed 10th, May, 2019).
59. Loi no. 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés et diverses dispositions concernant la protection des données à caractère personnel, Article 47.
60. Ordonnance no. 2018-1125 du 12 décembre 2018 prise en application de l'article 32 de la loi no. 2018-493 du 20 juin 2018 relative à la protection des données personnelles et portant modification de la loi no. 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés et diverses dispositions concernant la protection des données à caractère personnel.
61. Article 20(3)(7).
62. See discussion in Johnson, P. (2013) '"Damages" in European law and the traditional account of profits', *Queen Mary Journal of Intellectual Property*, Vol. 3, pp. 296–306. In the EU, this is mandated by Article 13 of Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights.
63. *Vestergaard Frandsen v Bestnet Europe* [2009] EWHC 1623 para 56, per Jacob LJ.
64. See at length: Evans, B. J. (2011) 'Much ado about data ownership', *Harvard Journal of Law and Technology*, Vol. 25, No. 1, pp. 69–130. More generally, on the propertisation of data protection, see Pearce, H. (2018) 'Personality, property and other provocations', *European Data Protection Law Review*, Vol. 4, No. 2, pp. 190–208; Schreiber, A. (2009) 'Privacy: Proprietary or human right?', *Intellectual Property Quarterly*, No. 1, pp. 99–138.

Country profile — Chile

Received: 28th August, 2018

Oscar Molina

Oscar's practice focuses on issues related to the protection and defence of intellectual property assets, technology contracts, privacy, domain names, software licensing and e-commerce. He also has significant experience in advertising issues, regulation in the entertainment industry and consumer protection. He holds an LLM degree from New York University and from the National University of Singapore. Oscar has been recognised by the prestigious legal ranking Chambers & Partners in the area of intellectual property and recommended in the Leaders League in the area of Technology and Digital Services. He currently serves as Chapter Co-Chair for Chile at the International Association of Privacy Professionals (IAPP) and has been certified by the same organisation in privacy management.

International Trademark Association (INTA) – Amicus Brief Subcommittee Chair;
International Association of Privacy Professionals (IAPP), Co-Chair for Chile;
El Golf 150, 4th Floor, Santiago, Chile
E-mail: omolinad@gmail.com

Andrea Céspedes

is an Associate at Albagli Zaliasnik and a member of the firm's Intellectual Property, IT & Data Protection Group. Andrea advises national and international clients on matters of intellectual property. Her practice is focused on copyright and technology contracts, distribution and franchise agreements, e-commerce agreements, personal data contracts and policies, as well as general legal advice related to transactional intellectual property and new technology matters. She also participates in registration and litigation procedures brought before the Chilean Trademarks and Patents Office. Andrea is a member of the International Association of Privacy Professionals (IAPP).

Albagli Zaliasnik, El Golf 150, 4th Floor, Santiago, Chile
Email: acespedes@az.cl

Abstract This paper provides a general description of the Chilean personal data protection regime, especially in relation to the main piece of legislation that governs this matter in the country; that is, Law 19.628 'On the protection of private life', commonly referred as the 'Personal Data Protection Law'. It describes the main elements of the law, in addition to other relevant norms and compares it to the General Data Protection Regulation (GDPR) in Europe, where relevant. Chile was the first Latin American country to adopt a personal data protection law in Latin America; however, as the description of the current legal scenario shows, it is currently lagging behind in comparison to modern legislation such as the GDPR, which includes relevant distinctions, rights and principles not present in current Chilean law.

KEYWORDS: data protection, Chile, privacy, GDPR

LAW AND REGULATORY FRAMEWORK

In Chile, personal data protection is addressed in several specific laws, as well as scattered provisions in related or complementary norms. Nevertheless, the following laws give the primary framework on the subject:

(1) Constitution of the Republic of Chile, Article 19 No. 4: This is the most elementary legal source for all matters

relating to the right to privacy and to data protection. In 2018, the constitution was amended to explicitly include the right to data protection, which now establishes 'The respect and protection of private life and the honor of the person and his family, and furthermore, the protection of personal data. The treatment and protection of this data will be put into effect in the form and conditions determined by law.'

(2) Law 19.628 'On the protection of private life', commonly referred as 'Personal Data Protection Law' (henceforth DPL): This is Chile's most comprehensive regulation on data protection and applies to all data collection and processing on Chilean territory. It contains provisions addressing: (i) the regulation of data subjects' rights; (ii) additional requirements on the processing of personal data relating to economic, financial, banking or commercial debts; (iii) the processing of personal data held by public institutions; and (iv) the liability of the person/entity in charge of the database for violations of the Law. The last update made to the DLP dates from 2012.

Chile joined the Organization for Economic Co-operation and Development (OECD) in 2010, pledging to reform and amend relevant laws to the organisation's standards, including the DPL. Since then, there have been various attempts to reform current law to include the OECD's recommendations.

In early 2017, a new Data Protection Bill (henceforth 'the Bill') was introduced for discussion in the Chilean Congress. Although it is still at an early discussion stage, there seems to be political will to have the Bill approved by the end of 2019. It is inspired in the European General Data Protection Regulation (GDPR) and includes new heavy fines for infringements, additional rights for data subjects such as the right to data portability, new categories of data (biometric and geolocalisation), regulations for international data transfers and the creation of a Data Protection Agency to monitor compliance of the new law, among other relevant changes.

DATA PROTECTION AUTHORITY

In Chile there is no public agency that monitors compliance with personal data regulation. Enforcement of Chilean data protection law is done by the local courts, which have no *ex officio* investigative powers and will proceed only when requested by a data subject, as follows:

- The *Jueces de Letras*, or territorial civil jurisdiction, judges exercise jurisdiction in the first instance over violations of the DPL.
- The Appeals Court exercises jurisdiction in the first instance in connection with constitutional actions, including those involving alleged breaches of the constitutional right to personal data protection. It is also the Appeals Court (with second instance jurisdiction) that has jurisdiction over matters involving alleged violations of the DPL.
- The Supreme Court hears appeals involving constitutional violations. Also, when a citizen's petition for removal, information, modification or blocking of their personal data from a public or private database is denied on 'national security' grounds under the DPL, it also has jurisdiction in the first instance over such claims.

The Bill seeks to amend this lack of a specialised supervisory authority, granting administrative powers to the 'Counsel for Transparency', a public institution currently in charge of ensuring transparency of information in the public sector. If the Bill is finally approved, the Counsel for Transparency would become, in effect, the Chilean Data Protection Authority.

Consequently, unlike Europe where each country has a national data protection agency that ensures compliance with the

GDPR, in Chile the local courts have had exclusive powers to enforce the DPL, which has only happened on rare occasions.

BREACHES OF DATA PROTECTION

To this date, there are no administrative sanctions or criminal penalties considered in the DPL in relation to data breaches. Breaches of personal data that result in damages are handled by local courts, through legal actions presented by affected parties. In addition, there are no mandatory data breach notifications established in the law. An exception is the banking sector in which notification of cybersecurity incidents, which may involve personal data breaches, have recently been made mandatory.

In this sense, our current DPL differs substantially from the GDPR, which establishes a clear obligation of notification of data breaches to the supervisory authority and details the information that needs to be included in relation to the informed breaches.

SCOPE

The DPL applies to all industries and types of organisations, both public and private. In Chilean law, personal data is defined as 'any data related to information of any type concerning identified or identifiable natural persons', hence any information that falls within that definition will be subject to the DPL.

Therefore, it shares with the GDPR the declared scope of covering all industries and organisations, including public institutions, bodies, offices and agencies. Nevertheless, given the low level of enforcement of the DPL, in practice the scope of the law is very limited.

COMMUNICATIONS, MARKETING AND SURVEILLANCE LAWS

The DPL does not explicitly mention these issues, with the exception of electronic marketing which is covered in Article 4 of the law. It is one of the cases in which the general requirement to obtain consent from data subjects may not apply, if the personal data is obtained from a publicly available source and the data is needed to provide direct commercial communications.

This issue is also convered in the Consumer Rights Protection Act, which allows direct marketing communications, provided that the content of the message includes contact information through which the recipient of the messages may opt out.

Regarding the interception of communications, that issue is covered in several laws, such the Computer Crime Act, Criminal Procedure Code, Terrorism Act, State Intelligence Law, Drugs and Narcotics Law and Free Competition Act, all of which provide the authority to intercept communications in the exercise of their respective authorities' investigative powers.

OTHER LAWS

Other data related rules are scattered in different legal bodies, including the following:

(1) Labor Code: Article 2, paragraph 7 (prohibition on consideration of financial data in the recruitment process); Article 5, paragraph 1 (duty of the employer to respect the constitutional rights of his or her employees, especially the right to privacy); and Article 154 (obligation of the employer to keep employees' personal data confidential).
(2) Consumer Rights Protection Act: Article 37, paragraph 4 (obligation on the supplier of goods or services to provide notice about possible personal data processing, where such processing is not specifically authorised by law).
(3) Tax Code: Article 30, paragraph 4 (obligation of data collector processing data related to tax declarations to maintain that information as confidential).

(4) General Banking Law: Article 14 and Article 154 (both on banking secrecy). Additionally, the Superintendence of Banks and Financial Institutions ('SBIF' in Spanish) has issued an update on the banking regulation to mandatory the notification of cybersecurity incidents in the banking sector.
(5) Patient Rights and Duties Act (Law No. 20.584), regulates the privacy of patient health records and requires that these are to be kept by the healthcare provider for 15 years.

PERSONAL DATA FORMATS

Under Chile's current DPL, 'personal data' is any data related to information of any form concerning identified or identifiable natural persons. Data concerning legal persons or non-identified or non-identifiable natural persons is not 'personal data' under this law.

The only distinction made by the current DPL regards 'sensitive data', which is defined broadly as personal data that refers to any physical or moral characteristics of any person, or to facts or circumstances of his or her intimate sphere, such as personal habits, racial origin, political ideologies and opinions, religious beliefs, physical and mental health, and sexual life.

The Bill under current discussion creates new categories of data a such as geolocalisation data, biometric data and the data of teenagers and children.

EXTRA-TERRITORIALITY

The application of the law is limited to the data controller established or operating within the jurisdiction of Chilean. Furthermore, international transfers of personal data are not explicitly mentioned in the current DPL. In practice, data subjects need to consent to the international transfer of their data for it to be lawful.

The Bill currently being discussed does regulate in detail the requirements to proceed with international transfers of personal data. If approved, the Data Protection Authority should create a list of countries that are authorised to receive personal data from Chile, without pre-approval. Currently, the European Commission does not consider Chile as offering an adequate level of data protection. It remains to be seen after the approval of the Bill, if the country's status will be reconsidered.

Unlike the GDPR, there are no provisions regarding the extraterritorial scope of the law and neither does the current Bill being discussed in Chilean Congress.

COVERED USES OF PERSONAL DATA

The DPL defines processing of personal data in a very broad manner as any operation or set of operations or technical procedures, automated or not, that permits the collection, storing, recording, organisation, working on, selection, extraction, comparison, interconnection, dissociation, communication, providing, transfer, transmission or elimination of personal data, or the use of said data in any way. Consequently, under this definition, a data controller who has a legal basis to use personal data may proceed with practically any type of use, provided that they have secured express consent or a legal exception applies.

Chilean law does not use the terms data controllers or data processor, preferring the concept of 'responsible person'. The person in control of the database is the natural or legal person that is responsible for the decisions related to the processing of personal data.[1] They must fulfill the following obligations:

- Use personal data only for purposes that are not contrary to the public interest, respect data subjects' human rights (especially privacy) and the rights that are given to them by the DPL.[2]
- Obtain informed consent from data subjects before processing their personal data (unless exemptions apply).[3]
- Use personal data only for the purposes for which they were collected, except

for personal data obtained from publicly accessible sources.[4]

- Keep personal data accurate and up-to-date.[5] The person in charge of the database shall: eliminate from it any personal data which has not been legally obtained, as well as any data that is out of date or for which the authorisation to process such data has expired; rectify any inaccurate or incomplete data; and block access to any personal data if its accuracy or expiration cannot be verified.[6]
- Take reasonable measures to secure personal data.[7]

LEGITIMATE PROCESSING OF PERSONAL DATA

Any personal data processing, including collection and transmission, must be specifically authorised by law or consented to by the data subject. Consent must be explicit and provided in written or electronic form. The data subject must be informed of the purpose of the processing of his or her data, and, if relevant, the possibility of the data being publicly communicated. Therefore, entities are required to obtain data subjects' prior, explicit, specific and informed consent in Chile. The consent given by the data subject is revocable; however, revocation will not have retroactive effect and must be stated in writing.

As noted, consent is not required for processing of personal data that is expressly authorised by law. The most relevant types of data that may be processed without consent are:

(1) Personal data issued or collected from *publicly available sources* if such data is:
   i. 'financial data'(economic, banking or commercial data collected from publicly available sources);
   ii. related to a specific group of individuals and which disclose information about, for example, their occupation, educational, titles, address or date of birth; or
   iii. information used to direct commercial or marketing communications (although the data holder must be able to opt out).

(2) Personal data processed by private legal persons (ie business organisations) for their exclusive use, or for the use of their associates and the entities to which they are affiliated, if the purpose of said processing is:
   i. statistical; or
   ii. to establish prices.

Even where a data subject's consent is not required, a data subject may still exercise their rights (eg of access and correction) over personal data.

Sensitive data may not be processed, unless there is (1) legal authorisation; (2) the data subject has consented; or (3) the sensitive data is necessary to grant healthcare benefits to its holder.

Current law does not consider 'legitimate interest' as a base for processing personal data, which is a substantial difference with the GDPR. Nevertheless, 'legitimate interest' is being discussed in the Bill as an additional legal basis to consider.

DATA HANDLING RESPONSIBILITIES OF THE DATA CONTROLLER

There are no notification obligations in the law, either to inform of current processing or of data breaches. Nevertheless, the data controller is obliged to handle the database containing personal data '*with due diligence, being held accountable for the damages*.''. In practice, it is the local courts that determine if the data controller has fulfilled the standard of care set by the law.

CONTROL OF USE OF PERSONAL DATA

As a general rule, once the data subject has given consent to the data controller ('responsible person') to use their personal data,

they have conceded control over it. The data controller is limited, however, by the finality principle contained in our law, by which the use of data can only be limited to the purposes for which it was granted. Consequently, a controller would not be legally authorised to use the individual's data for a different purpose to that which they were originally authorised.

In addition, Chilean law explicitly includes a series of rights, which grants some control to the data subject, such as:

- *The right to be informed*.[8] Prior to giving consent, the data subject must be informed of the purpose of the data processing and whether the data will be made publicly available.
- *The right to data access*.[9] The data subject can request, free of charge, access to their personal data, as well as information about the sources and recipients of such data, the purpose of the processing and the identity of third parties to whom that data is being transferred to regularly.
- *The right to rectify data*.[10] If data is wrong, inaccurate or incomplete, the data subject may request the modification of such data.
- *The right to eliminate or block data*.[11] If the personal data is not stored legally (eg no consent was obtained) or if the data is no longer up-to-date or the authorisation to process the data has expired, then the data subject will be able to request that the person in charge of the database eliminate their data from it. Data subjects also have the right to request the elimination or blocking of personal data stored in a database, if such data was given voluntarily by the data subject, or if the data is being used to send marketing communications.

These rights are collectively referred as to 'ARCO' rights after their initials in Spanish (*acceso, rectificacion, cancelacion* and *oposicion*).

Data subjects are entitled to exercise these rights free of charge, every six months.[12] When data subjects use their right to rectify, eliminate or block data, they can request a free copy of the altered entry of the database. If new rectifications or eliminations of data are made, the data subject will be able to request free copies of the updated entry. Each new request may be made not less than six months after the last.

In comparison, the GDPR established additional rights that are not explicitly recognised by current law, such as the right to be informed, to restrict processing and to data portability. Nevertheless, the inclusion of these rights is currently being discussed in the Bill.

DATA ACCURACY

The person in charge of the database (the natural or legal person that is responsible for the decisions related to the processing of personal data),[13] must keep personal data accurate and up-to-date.[14] The person in charge of the database shall: eliminate from it any personal data that has not been legally obtained, as well as any data that is out of date or for which the authorisation to process such data has expired; rectify any inaccurate or incomplete data; and block access to any personal data if its accuracy or expiration cannot be verified.[15]

It is required by law that the responsible person must take these measures to ensure that the data is accurate, notwithstanding the petitions made by any given data subject (Article 6).

AMOUNT AND DURATION OF PROCESSING PERSONAL DATA

There is no explicit data minimisation principle in Chilean law, nor does the DPL contain any explicit maximum or minimum data retention periods. Nevertheless, through the application of the purpose limitation principle, which establishes that personal data may only be used and processed to accomplish the purposes for which they were originally collected, the data controller should dispose of the data after it has

fulfilled its declared purpose. Reuse of said data, without securing additional consent, represents a breach of the purpose limitation principle.

Other sectorial laws impose minimum data retention periods, such as law No. 20.584, on patient's rights and duties, which requires that patient health records are kept by the respective healthcare provider for at least 15 years.

USE OF PERSONAL DATA

The DPL follows the purpose limitation principle, by which data may only be used and processed to accomplish the purposes for which it was originally collected. As an exception, however, if the data has been obtained from publicly available sources and it falls under any of the categories included in Article 4 of the DPL, the data may be used for various purposes, as express and informed authorisation is not required.[16]

SECURITY OF PROCESSING PERSONAL DATA

All personnel involved in the treatment of personal data have a legal obligation of confidentiality related to data that is not publicly available, even after they end their contractual relation.

The security of personal data contained in databases is an obligation of the 'responsible person', as defined above. This person must maintain the database and keep it '*with due diligence, being held accountable for the damages*.'

This article does not distinguish the nature of the damages (*moral* or common, losses, etc), resulting from a breach of security measures.

NOTIFICATION OF A PERSONAL DATA BREACH

Currently, there are no general data breach notifications obligations in the law. As of September 2018, the Superintendence of Banks and Financial Institutions ('SBIF' in Spanish) has issued an update on the banking regulation to oblige notification of cybersecurity incidents in the banking sector. In this case, the norm makes it mandatory to notify about security incidents, not only to the authorities (SBIF) and the affected clients (under certain conditions), but also the industry itself.

Other than this sectorial requirement, however, there are no general provisions that require data controllers or processors to notify authorities or individuals of breaches, which is a clear difference with the GDPR.

Nevertheless, it is likely that the current reform of the DPL will include mandatory data breach notifications. The extent and scope of those notifications remains to be determined.

INTERNAL CONTROLS

As previously noted, Chilean law does not explicitly distinguish the figure of a data controller or data processor as such, but rather prefers the concept of 'responsible person', who is accountable for mitigating harm or damage to an individual as a result of processing their personal data.

The 'responsible person', that is the natural person, legal entity or public body that makes decisions related to the treatment of personal data, is responsible for ensuring that personal data is protected in accordance with applicable law. The general duty of care that the law imposes is that of 'due diligence'.

There are no general requirements to appoint a data protection officer (DPO), as whoever falls under this definition becomes the 'responsible person'.

The 'responsible person' must also respond to the inquiries of any individual regarding their personal data and its modification, deletion or blocking. If the responsible person provides no answer within two business days, the affected person can initiate a civil procedure before the corresponding authorities.

As a related control measure, when the processing of private databases is delegated to a third party — in effect a data processor, following the GDPR's nomenclature — the contract between both parties must be in writing and include the conditions stipulated in the processing.

RECORD KEEPING

The DPL does not establish a general obligation on record keeping or documentation, except for the case in which the responsible person (data controller) has implemented an 'automated transmission procedure', where it is mandatory to track:

(1) the inquirer's identity;
(2) the motive and purpose of the data requests; and
(3) the specific data that is being transferred.

REGISTRATION WITH SUPERVISORY AUTHORITY

There is no register of private data controllers. Regarding data banks that are under a public authority, the Civil Registry and Identification Service will keep the Register of Personal Data Banks in charge of Public Organisations.

This is certainly different to the GDPR, in which the data controller must carry registration of their process and inform the authority as requested.

ENFORCEMENT AND PENALTIES

Responsible entities are liable for breaches of the DPL in relation to denials of data subject rights (access, rectification etc), and could face a fine between US$80–4000, to be determined by the court. If the person in charge of the data bank (the data controller) is a public service the court may sanction the head of the service with the suspension of their position for a period of five to 15 days.

As previously noted, there are no specific penalties for data breaches, as these are handled directly by local courts, if an affected individual request the courts' intervention. If the legal action is successful, the court will impose monetary compensation based on the damage that has resulted from the breach.

There is consensus within the Chilean jurisdiction that the current DPL has a low enforcement ratio and the few penalties that it includes are rarely imposed. For these reasons, the reform Bill seeks to increase the range of sanctions, in the same manner as the GDPR, distinguishing between mild, serious and very serious offences. The fines imposed could range from US$80 to 350,000 or 4 per cent of annual income, depending on the existence of malice.

References

1. Law 19.628, Article 12, n).
2. Law 19.628, Article 1, second paragraph.
3. Law 19.628, Articles 4 and 10.
4. Law 19.628, Article 9, first paragraph.
5. Law 19.628, Article 9, second paragraph.
6. Law 19.628, Article 6.
7. Law 19.628, Article 11.
8. Law 19.628, Article 4, second paragraph.
9. Law 19.628, Article 12, first paragraph.
10. Law 19.628, Article 12, second paragraph.
11. Law 19.628, Article 12, third and fourth paragraphs.
12. Law 19.628, Article 12, fifth paragraph.
13. Law 19.628, Article 12, n).
14. Law 19.628, Article 9, second paragraph.
15. Law 19.628, Article 6.
16. Law 19.628, Article 9, first paragraph.

Book review

'The Handbook of Privacy Studies. An Interdisciplinary Introduction'

By Bart Van Der Sloot and Aviva De Groot (Eds)
Amsterdam University Press, Amsterdam; 2018; 456 pp.; ISBN 9789462988095; €29.95

This new handbook represents an impressive attempt to summarise encyclopaedically the state of the art of relevant legal and social science knowledge on privacy in general. For a topic that is subject to constant mutations from intermittent technological change, and where cultural differences lead to dissimilar rules, perceptions and practices in different jurisdictions and different cultural communities, this is a tall order. The editors deserve credit and praise for having undertaken this unparalleled task — in a hardcopy publication on top of that. The book consists of substantial academic chapters (numbered), interspersed by short essays (un-numbered), often aimed at providing a particular perspective or experience (which may be academic, but one that requires a different, shorter and more personal format). The book is structured as follows:

The impressive breadth of this project makes it stand out; indeed, there seems to be no direct competition on the book market. Certainly, reviewing edited books with numerous chapters by authors with different backgrounds can be a daunting task, as it can be impossible to do justice to all texts without summarising and commenting on them one by one, which would require a much overlong review. Suffice to say that there is much to learn from the 11 chapters and the nine shorter 'in-between' essays. Despite some shortcomings, which will be addressed next, it is a rich resource which I certainly expect to draw upon shortly in my own work. Yet in geographical terms, the authors represent a rather narrow corner of the world — narrow if the book is intended for a European or even a worldwide readership — whereas cultural and legal traditions in countries other than those represented via the authors' affiliations remain quite unaccounted for. Judging from the author biographies at the end of the book, 15 authors are based in the Netherlands, eight in the USA, two in the UK and one in Belgium, while in one case the exact geographical affiliation cannot be determined with certainty. Anyone familiar with the balancing acts needed in European data protection meetings will be aware that, even within one continent, traditions and perceptions differ markedly. The exception to this rule is of course Chapter 11, with its look at extra-European cultures. We learn from the Introduction that the book stems from a conference and, as such, it is understandable that some countries should be over-represented, because expertise on a particular subject matter is often concentrated in a few places. Nevertheless, there would have been a case for inviting some additional perspectives from other traditions and jurisdictions and, as regards expertise, countries such as Estonia or Sweden could also have furnished some interesting ideas. Two authors are German-speaking but based at Dutch or British universities. If this is a potential source of bias, it has to be acknowledged that Continental European traditions are not overtly visible, though Chapter 2 mentions some German traditions, such as the BVerfG *Volkszählung* judgment.[1]

Some chapters (eg the Introduction), while listing impressive amounts of sources, have too few footnotes. When Chapter 2 mentions 'other schools' in connection with *Volkszählung*, it would have seemed natural, precisely in a reference work, to include a footnote pointing readers to the relevant publications. The same applies to other references made to issues over which the literature is divided: a reference work ought to be more explicit. Later, the Roessler essay does mention this seminal decision, making up for some of the Anglo-Dutch bias, yet the link between the BVerfG doctrine and its alternatives is still not made visible. The textbook character is highlighted by detailed explanations, as in Chapter 2, as to what makes a law, what is the EU, what distinguishes it from the Council of Europe, etc. It is precisely in a text of this genre that references ought to be pinpointed and signposted, if nothing can be taken for granted.

Despite its inherent interest and novelty, the book could have profited from some additional editing. Although multi-author volumes should allow authors to write in their own style (having edited several collective volumes, I always resisted the temptation to impose an impersonal style on contributors), references should follow a unified system. In this book they do not and, this being a handbook (thus a practical tool), the practical value of the book depends on the reference system, *inter alia*. Chapter 2 is followed by a copious bibliography, much of which has not actually been quoted in the text; it feels rather like a guide to further reading. The discussion on the balancing of conflicting norms and rights (pp. 123–124) is well-written and inspirational (like the whole of that chapter), yet does not include

any references to who said what, despite making it clear that authors are divided over the issues identified. The users of a handbook cannot be expected to know beforehand: it is precisely to find out that they are using the handbook. Chapter 3, by contrast, includes numerous and very specific references, using the APA format, as do Chapters 4 and 10, for instance. Chapters 5 and 10 rely on footnotes, much in the way familiar to the legal community. Chapter 10 is a medical chapter, but the footnote system is not the one familiar to medical journals (continuous numbering). In Chapter 2, the page numbers of articles and book chapters are missing though the word 'pages' is always included. Chapter 6 is by far the most bewildering, relying on 121 footnotes, with a bibliography entitled 'Further reading' following the chapter. Titles are grouped thematically, but appear in the most haphazard order imaginable: neither alphabetical, nor chronological. Authors' initials may appear before or after their surnames, or not at all, while a report from the British House of Lords (written 'Lord' using the singular) is referred to using the French word, 'Rapport' rather than the English word 'Report' (apart from the fact that 'Report' is not an author); it does not seem to appear in any of the footnotes. While I hope that the quality of the text is good (unlike the authors, I am not an expert on 'Privacy from an Intelligence Perspective'), the state in which the reference list is found raises the question whether the text *itself* has been prepared with the necessary diligence as can be expected in an academic publication. Applying the usual standards for academic work, in particular as regards references, submission to a self-discipline is beneficial, not only in that it helps readers find the works quoted, but also so that we can double-check the veracity of all claims made in the text.

On balance, and with the caveats set out earlier in this review, this remains a novel and useful reference work. It stands out by integrating both legal and social science perspectives, thereby filling a gap in the extant literature. The editors are to be congratulated for their efforts in bringing together contributions from so many fields and specialisms. Whether it is an interdisciplinary collection, as suggested in the subtitle, is debatable, for as we learn from the Introduction (p. 12), authors were asked to 'keep to their disciplines'. This seems to have been a wise decision, and the book may alternatively be characterised as pluridisciplinary, in that it juxtaposes different disciplinary perspectives.

Dr Jacob Kornbeck
Brussels, Belgium
Journal of Data Protection & Privacy

Reference

1. BVerfG, Urteil v. 15 Dezember 1983, Az. 1 BvR 209, 269, 362, 420, 440, 484/83.

Data Protection & Privacy – Free Online Access

Subscribers to *Journal of Data Protection & Privacy* can also access the full content of the journal online at no further charge. Simply register with our host company ingentaconnect then contact HSP to validate your access. Only fully paid subscribers will be able to receive free online access via INGENTACONNECT.

INSTRUCTIONS FOR REGISTERING WITH INGENTACONNECT

1. Go to http://www.ingentaconnect.com/content/hsp and click the 'Register' button that appears in the box on the right hand side of the screen.
2. Click the button marked 'Register' that appears to the right of 'Individual Registrations'.
3. Enter the information in the boxes shown on the page. *We cannot process your registration without this information.*
4. Please choose a username and password that match the following criteria. Your username must be at least four (4) characters in length and may have either alpha or numeric characters. Your password must be at least six (6) characters in length. Password may not contain username and must contain at least one (1) alphabetic and one (1) numeric character.
5. Click 'OK'.

If you have any problems registering, please e-mail either: uksales@ingentaconnect.com
or: GwenY@henrystewart.co.uk

Once you have registered, please e-mail

— your INGENTA ID number;
— your username;
— your e-mail address.

to: gweny@henrystewart.co.uk so that we can validate your online access.

If you have recently paid, please allow up to 14 days for your access to be validated.

INSTRUCTIONS FOR VIEWING ARTICLES ONLINE

1. Go to: http://www.ingentaconnect.com/content/hsp.
2. Enter your username and password in the box to the right of the screen.
3. In the section 'Quick Search', enter *Journal of Data Protection & Privacy* in the box 'Search For'. Then click the button 'Publications' and click 'Search'.
4. You should then see a screen that will take you into the content of the journal.

For assistance, e-mail: support@ingenta.com

HENRY STEWART PUBLICATIONS