Master's Degree in Transnational Law

Compatibility between the DGA's framework on the re-use of publicly held personal data and the GDPR

Mariana Correia Martins Barreira Fernandes

No.143821006

Supervisor: Professor Nuno Sousa e Silva

Lisbon, September 10th 2023

Pelo fulgor do estio, pelo azul do claro dia,
Pelas flores que esmaltam os campos, pelo sossego dos pastos,
Pela exactidão das rosas, pela Sabedoria,
Pelas pérolas que gotejam dos olhos dos amantes,
Pelos prodígios que são verdadeiros nos sonhos,
Pelo amor, pela liberdade, pelas coisas radiantes.

Natália Correia in Ode à Paz

À minha Mãe

Table of Contents

Glossary

AI Artificial Intelligence

Art(s). Article(s)

A29WP Article 29 Working Party

APIs Application Programming Interfaces

CJEU Court of Justice of the European Union

DSM Digital Single Market

DGA Data Governance Act

DPA(s) Data Protection Authority(ies)

DPIA Data Protection Impact Assessment(s)

DS Data Subject(s)

EC European Commission

EU European Union

EDPB European Data Protection Board

EDPS European Data Protection Supervisor

EDS European Data Strategy

GC General Court

GDPR General Data Protection Regulation

ICO International Organization for Standardization

IEC International Electrotechnical Commission

IoT Internet of Things

MS Member State(s)

NODD New Open Data Directive

NPDR Free Flow of Non-Personal Data Regulation

Personal Data PD

PSB Public Sector Bodies

PSI Public Sector Information

PSID Public Sector Information Directive

Rec(s). Recital(s)

RP Research Paper

SMEs Small and Medium Enterprises

SPEs Secure Processing Environments

TFEU Treaty on the Functioning of the European Union

Keywords

access to data, anonymisation, data protection law, data governance, DGA, GDPR, non-personal data, open data, personal data, PD, re-use of data, risk of re-identification

Chapter I. Introduction

1.1. Background and Motivation

Emerging technologies such as AI, serverless computing, augmented reality or IoT are impacting the future of every industry and human being. From helping logistics companies to predict average arrival times, assisting researchers in solving cancer or health analysis provided by a smartwatch on our wrist. These are some of the main drivers of innovative digital economy developments in which data is the primary cell.

Data is a vital resource for developing technologies in public and private sectors, constantly being generated, processed and applied in significant amounts. There has been an exponential growth in data volume, expected to reach 175 zettabytes[1] by 2025. The speed-up development inherent to the tech industry features urgent need for companies to access more relevant data to produce state-of-the-art products and provide new services effectively.[2]

The potential and value of data do not expire after a single use, additional value can be found in secondary uses.[3] Therefore, digital economy's business actors have a significant interest in accessing and re-using data held by other market players. Data re-use in new models can boost competition and support business innovation by facilitating the design of sustainable and high-end solutions. This will foster effective competitiveness in the digital era and benefit consumers[4]. Yet, most of the data is controlled by incumbent companies and retained for market reasons or public administrations[5], both collecting vast amounts of data in a plethora of domains and activities, as healthcare, manufacturing and financial services industries.

[1] David Reinsel, John Gantz and John Rydning, 'The Digitization of the World from Edge to Core', International Data Corporation, 2018.

[2] World Development Report 2021: Data for Better Lives, *The World Bank*, 2021, 93.

[3] Stefaan G. Verhulst, 'Unlock the Hidden Value of Your Data', Harvard Business Review, 2020.

[4] OECD, 'Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by 'Big Data'', OECD Digital Economy Papers No. 222, OECD Publishing, 2013.

[5] Thorhildur Jetzek and others, "Data-Driven Innovation through Open Government Data", Journal of Theorical and Applied Electronic Commerce Research, 9(2), 2014, 102.

Data held by public bodies is essential to foster public policies and services[6]. There is a potential value in the availability and secondary uses of public sector data – it may promote democratic citizen participation and prevent abuse of power[7], push for efficient decision-making by re-users/end-users[8], promote accountability[9] and furthermost stimulate disruptive products/services by competitive companies of all sizes[10].

One may state access to public data is a critical factor for a wealthier data-driven economy with functional data-based business models throughout all economic sectors within the EU[11]. Particularly for start-ups and SMEs, access to relevant data may be interpreted as an "essential resource"[12] for scientific R&D[13] and making informed decisions. As such, data retention and "power advantage"[14] of more prominent market players constitute a barrier to entry for SMEs and start-ups in the highly competitive digital market, hindering the application of technical expertise and potential benefits for the EU economy and social progress – job creation[15], innovation and productivity growth[16]. In contrast, data remains an under-utilised asset rather than having impact on the public good.[17]

[6] OECD, The Path to Becoming a Data-Driven Public Sector, "*The application of data in the public sector to generate public value"*, OECD.

[7] Filipa Urbano Calvão, '*Direito da Proteção de Dados Pessoais: Relatório sobre o programa, os conteúdos e os métodos de ensino da disciplina'*, (Universidade Católica Portuguesa. Porto, 2018), 19.

[8] Empowering people with open data, The Official Portal for European Data, 2022.

[9] Agustí Cerrillo-i-Martínez, 'The Reuse of Public Sector Information in Europe and Its Impact on Transparency', European Law Journal, 770-7922012, 18(6).

[10] Open data and Entrepreneurship, Data.europa.eu, 2018.

[11] Josef Drexl and others, 'Data Ownership and Access to Data', Max Planck Institute for Innovation and Competition RP No. 16-10, 2.

[12] COM(2020) 66 final, 2.

[13] COM(2011) 882 final, 6.

[14] *Ibid* 8.

[15] The DynEmp: Measuring job creation by startups and young firms, OECD, 2021.

[16] OECD (6).

[17] Stefaan G. Verhulst (3)

Aligned with Government 2.0's rise[18] (Web 2.0. transformation within governmental contexts to better solve collective services and processes), the European Commission (EC) has long stepped up efforts to increase availability and facilitate re-use of public data focusing on generating value for society and businesses through re-use of public sector information (PSI). With the PSID[19], a common legal framework for a European market for government-held data has been established at the EU level. Simultaneously, efforts have been taken to guarantee a free flow of data altogether scenario in which there are no legal barriers to cross-border data flows with the adoption of the General Data Protection Regulation (GDPR)[20] and the Free Flow of Non-Personal Data Regulation (NPDR)[21].

Later in the wake of the GDPR, lawmakers have agreed on a revised Directive on open data and the re-use of public sector information[22] (NODD) in force since 2019, renaming PSID rules. NODD states any individual/entity may request a public sector body to release data for re-use. It aimed to increase the availability and quality of data held by national PSB and boost reusability. Yet, its application's scope is somewhat limited as it enables the Member States (MS) to legislate on exceptions and discharges a broad range of data. This leads to the circumstance the access and re-use of tremendous amounts of data are not allowed. Art. 1(2) NODD states it excludes specific categories of data protected by different legislations – documents, as sensitive data protected by national security/defence/public security reasons, statistical and commercial confidentiality (business, professional and company secrets); data protected by third parties' intellectual property rights; and (parts of) documents containing persona data (PD) whose full access is excluded or re-use would be incompatible with data subjects (DS)' protection of privacy and integrity under the GDPR. Hence, personal data publicly held continues to be "underutilised"[23] as opposed to being exploited to the fullest in

[18] Teresa M. Harrison, Theresa A. Pardo and Meghan Cook, 'Creating Open Government Ecosystems: A Research and Development Agenda', 2012, Future Internet, 900.

[19] Directive 2003/98/EC on the re-use of public sector information (PSID).

[20] Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

[21] Regulation (EU) 2018/1808 on a framework for the free flow of non-personal data in the EU, OJ L 303/59.

[22] Directive (EU) 2019/1024 on open data and the re-use of public sector information (recast).

[23] Rec 5, Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and amending Regulation(EU) 2018/1724 (Data Governance Act)

secondary use for European research and innovation[24] and by SMEs and start-ups who "typically find it more difficult to collect relevant data themselves"[25], being acknowledged as a "source of market power"[26]. The interplay between NODD's rules and the GDPR proves the protection of PD overrides any potential for the re-use of PD, making one question the existence of a truly free flow of PD.

It was the EU's intention to move a step forward with a European Strategy for Data[27](ESD) which intends to make the Union a real role model for a society empowered by data based on common, interoperable European data spaces (ecosystems with data infrastructures and governance frameworks to ease data pooling, access and sharing in different sectors[28]).

The EU based on internal market competence (Art. 114 TFEU) aspires to a European way of data governance to ease data sharing across MS and sectors through the Regulation on European Data Governance - Data Governance Act (DGA), and the complementary Proposal for a Regulation on harmonised rules on fair access to and use of data (Proposed Data Act) which establishes the rules on who can use and access what data for which purposes across sectors.

The horizontal framework established in the DGA is over-arching and built up of 4 pillars: the re-use of certain data held by public sector bodies (PSB) which is subject to the rights of third parties, being out of the NODD's scope, namely personal data (Chapter II); the regulation of data sharing intermediaries by setting up a number of operating requirements (Chapter III); the encouragement of data-altruism to create more tools and opportunities for individuals to donate data for the common good (Chapter IV); the functioning of the monitoring competent authorities (Chapter V) and the creation of a European Data Innovation Board (Chapter VI).

Materialising the ESD[29], the European legislators reached a political agreement on the DGA on November 30, 2021 and on the April 6, 2022 the Parliament approved a draft. On the 16th of

[24] Rec 6 DGA.

[25] Rec 25 DGA.

[26] Josef Drexl and others (11) 9.

[27] A European Strategy For Data, COM(2020) 66 final.

[28] SWD(2022) 45 final, 2.

[29] COM(2020) 66 final.

May 2022, the Council approved the DGA's final text, which entered into force 20 days after the publication in the EU Official Journal on June 23, 2022. The DGA becomes applicable on September 24, 2023.[30]

DGA aims to create incentives for the re-use of relevant data for scientific research purposes and by start-ups and SMEs and mechanisms to facilitate the re-use of specific categories of protected public sector data. Due to its potentiality to broader the secondary use of data held by PSB, including of PD, the question of compatibility between the DGA's universal framework on the re-use of PD and the GDPR arises. This is the matter we aim to examine.

1.2. Research Question

The research question we will tackle in this paper lies in determining "the *extent in which the DGA's conditions for re-use of personal data held by public sector bodies are compatible with the GDPR and mitigate the risk of re-identification".* DGA intends to act as a horizontal legal foundation based on which the initial nine individual sectors' legislation, eg. on health, mobility and financial data spaces[31], can be built up, promoting the re-use of PD, as defined in GDPR, held by PSBs. The trust in re-use and the development of a European common data space depends on protecting PD. In particular, the possibility of re-identification[32] when de-identified anonymised data, erstwhile PD, is re-used must be considered. The analysis of the interplay between the DGA and the GDPR regarding such re-use assumes paramount importance to figure out to what extent the DGA intends to safeguard the protected nature of the data and mitigate the risk of re-identification in technical and legal terms. It is our objective to analyse the continuous effort from EU lawmakers to unlock PSI potential and the process of building

[30] Art. 38 DGA.

[31] Oscar Corcho and Elena Simperl, 'The European Data Spaces: A report on challenges and opportunities', Data.europe.eu, 2022, 6.

[32] Nadezhda Purtova, 'The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law', Law, Innovation and Technology, 2018, 41.

up the free flow of data within the EU. Another essential question we additionally ought to seek is how the DGA's regime deals with mixed datasets, combining personal and non-personal data, introduced by the NPDR. Namely, we should find out whether the newer Regulation contributes to developing a feasible solution to processing (and reusing) those datasets. After all, this assessment should prompt conclusions about whether the EU has designed an adequate framework for re-use of publicly held datasets, stimulating new business models and social innovation.

All in all, the main aim of this paper is to examine the possible impact of the DGA's framework on the re-use of PD held by PSB and its consistency with existing EU law, especially the GDPR. We aim to properly frame the foundations of the European Data Governance framework established in the DGA and reach conclusions about whether an efficient use of PD across the EU can be achieved.

Chapter II. Data at the centre of the European Strategy for the Digital Economy

2.1. Data as the driver of the European Digital Economy

The EC has gradually been proposing laws to facilitate the creation of a data-driven economy, which depends on access to large amounts of quality datasets as a means for social development by setting up better-fit products and services and improving European social welfare[33]. The capacity to fetch public and private data from public and private market players, guaranteeing the availability of extensive and high-quality training data, is paramount for businesses to train models and generate economic growth and innovation in the age of AI.

The DGA presents a data definition as *"*any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audiovisual recording*"*[34]. Priorly, a similar definition could be found in the Council Framework on attacks against information systems as "computer data"[35].

One may realise in EU Law's wording the terminology data and information tends to be used indiscriminately[36]. From a technical standpoint, there is a distinction between data and information: data on its own has no meaning[37] (raw facts or mere figures), then it only acquires a meaning when it is processed and provides significant information. Focusing on the information's syntactic level, ICO and IEC define data as a reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing[38].

[33] COM(2020) 66 final, 1.

[34] Art. 2(1) DGA. Equally reproduced in Art. 2(1) Proposed Data Act

[35] Art. 1(b) of the Council Framework 2005/2022/JHA OJ L 69

[36] Thomas Streinz, '*The Evolution of European Data Law'*, 2021, Paul Craig and Cráinne de Búrca, The Evolution of EU Law, OUP, 904.

[37] See Topic 1.1 Data, Information and Knowledge, Cambridge International AS & A Level Informational Technology, 9626.

[38] ISO/IEC 2382-1, Section 2, 01.01.02.

EU Law distinguishes different types of data - public, open and private data, as well as personal data and non-personal data.

The rationale behind open data (OP) that advocates making more public sector data available at higher quality to facilitate re-usability is no big news in the EU since the trend of open access and re-use of public sector data for community good has been long established. It has encouraged the possibility for individuals and companies to access and re-use public data by adopting the PSID, amended in 2013 and led to the NODD.

Public data or PSI means information public bodies or public undertakings, the latter covered in the NODD only, generate, collect or pay for, including a wide range of data, as statistical, publicly funded research data or even digitized books[39]. Open data stands for public data made available in an open and machine-readable format and can be widely accessed and freely re-used by anybody at all for any purpose[40] with zero or minimum legal restrictions.

These practices seek to improve accessibility towards government data re-use and boost private sector engagement. The EU has played a key role in setting the OD agenda with governments pioneering OP policies and initiatives in Europe and abroad. A reference must be made to the Estonian OD Green Book on the disclosure of PSI in a machine-readable format which led to the re-launch of the OD portal[41](including datasets from distinctive sectors/domains), or the Swedish dataportal.se,[42] to promote easier dissemination of data for re-use.

With the emergence of advanced technology and augmented ability of private and public sector actors to collect and process data, privacy and data protection are increasingly placed as a central theme in contemporary societies. Fundamental rights to privacy[43] and data protection[44] have become a top priority for European legislator, culminating in GDPR's implementation, dealing exclusively with PD. GDPR formalises a definition of PD in Art 4 (1) as any

[39] COM(2011) 882 final, (10), 1.

[40] Open Knowledge Foundation, 'The Open Definition - Defining Open in Open Data, Open Content and Open Knowledge' (*Opendefinition.org*), 2022.

[41] See https://avaandmed.eesti.ee/

[42] See https://www.dataportal.se/en

[43] Art. 8(1) ECHR.

[44] Art.16 TFEU.

information which is related to an identified or identifiable natural person. This definition limits GDPR's scope of application, helping to identify whether a specific data processing is subject to data protection rules.

One may argue EU data protection reform has negatively influenced sustainability and progression of OD and data re-use, since the NODD, agreed at the EU level in 2019, excludes PD from its scope of application. This circumstance may hint at how difficult it became to reconcile other legislation with the GDPR complex data protection rules.

2.2. GDPR: Fit for the digital age?

The GDPR's adoption, replacing the national regimes under Directive 95/46/EC, shows an attempt to unify data protection law at the EU level. It corresponds to a far-reaching move - from a Directive to a Regulation - aiming to achieve full harmonisation in implementing modern data protection rules in the internal market. GDPR is considered a "remarkable legislative achievement of supranational lawmaking"[45] and it claims to give DS "control back over their PD"[46] while it improves businesses' data protection compliance and promotes the free flow of PD.

GDPR central principles are influenced by the 1980 non-binding OECD Guidelines Governing the Protection of Privacy and Transborder Data Flows of Personal Data[47], renamed OECD Privacy Framework[48]. By then, the framework included, among others[49]: principles of collection limitation; of lawful and fair means of collection, and of data quality coupled with purpose specification.

[45] Streinz, (36), 913.

[46] Rec 68 GDPR.

[47] OECD, 'OECD Guidelines Governing the Protection of Privacy and Transborder Data Flows of Personal Data', 1980; Ruth Boardman and others, '*European Data Protection Law and Practice'* (Eduardo Ustaran, 3rd edn, The International Association of Privacy Professionals, 2023) 7.

[48] OECD, 'The OECD Privacy Framework', 2013

[49] *Ibid* 14.

One may see a parallel with the lawful, fairness and transparency requirement to the processing of PD along with purpose limitation and data minimisation principles in Art. 5 GDPR. The GPDR, directly applicable since May 2018, has inserted data protection on every company board's agenda, yet it has not accomplished a fully harmonised and efficient implementation nor accomplished a truly free flow of data[50]. As it installs several restrictions on PD processing, it wedges innovation boost through potential data access and re-use throughout the EU[51].

First, as degree of use and subsequent implementation of specification clauses by each MS varies, a "degree of fragmentation"[52] of regimes within the EU MS stands out. Considering the re-use of PD in research, different implementations of Art. 89 GDPR safeguards and derogations[53] from the general prohibition hinders cross-border data processing and precludes cutting-edge knowledge developments.

Belgian Data Protection Act states anonymous data shall be used for research purposes, yet, if the research purpose cannot be met, use of pseudonymous data is allowed and, if the purpose cannot be met with such data, use of non-pseudonymised data[54] is allowed. It includes safeguards as motivation why pseudonymised data is (or is not) used, the reasons why DS' rights may endanger or render the purposes impossible, and data protection impact assessments (DPIAs) when sensitive data is used[55]. In case of re-use, the initial controller shall anonymise or pseudonymise the data before communicating it to other controller for secondary processing for research, and a controller-controller agreement shall be concluded - except if the data were made public, law provides a mandate for research or re-use for other purposes is forbidden[56]. Portuguese law merely establishes anonymisation or pseudonymisation as a safeguard and adds consent regarding processing for scientific research purposes may cover several areas of

[50] EC, Staff Working Document, Impact Assessment Report accompanying the DGA Proposal, SWD(2020) 295 final, 18.

[51] *Ibid* 14.

[52] EC, Data protection as a pillar of citizens' empowerment and the EU's approach to the digital transition - two years of application of the GDPR, COM(2020) 264 final, 2020.

[53] Els Kindt and others, 'Study on the appropriate safeguards under Art. 89(1)GDPR for the processing of personal data for scientific research: Final Report', EDPB, 2021.

[54] Belgian Act 30.07.2018, Art. 197

[55] *Ibid* Art. 191

[56] *Ibid* Art. 194.

research or specific fields or research projects only, respecting ethical standards[57]. Finnish law establishes a data controller to benefit from the exception for scientific purposes shall develop a research plan, allowing for sensitive data processing for scientific research purposes on the condition research fulfils generally approved ethical principles for science[58]. French law limits the exemption from the prohibition on collecting sensitive data to public research, excluding private research.[59] One may conclude GPDR's diverse implementation *de facto* does not facilitate an efficient and economical approach to research purposes, especially if projects have a cross-border nature.

Due to GDPR's high compliance costs, companies in general and SMEs in specific[60], find the GDPR's compliance particularly challenging. Chen and others concluded companies subject to the GDPR experienced an average reduction of 8% in profits - SMEs were the most affected with 8.5%. Gal and Aviv argue small parties find data collection unprofitable and larger players may benefit from the legal uncertainty[61]. This situation reinforces barriers to entry for businesses and precludes growth opportunities. Also, GDPR may not be adequate for Big Data's era described by value, variety, velocity, veracity and volume (5V's)[62]. For a valuable Big Data analysis, it is crucial to have access to big quantities of unstructured data from diverse sources whose processing outcome could be used for different purposes[63]. One may say data minimisation and purpose limitation conflict with that exigency and hinder large-scale data re-use. Nicholas Martin and others have identified product abandonment, entrepreneurial

[57] Art. 31 Law no. 58/2019, of 8th August.

[58] Section 31 Finnish Data Protection Act.

[59] Art. 44.6 *Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés*. It is Francisco Paes Marques' opinion that profit-oriented scientific research does not leave out the benefits to society in '*Comentário ao Regulamento Geral de proteção de Dados e à Lei nº. 58/2019*', (A. Barreto Menezes Cordeiro Almedina 2022), 530.

[60] Chinchih Chen, Carl Benedikt Frey, and Giorgio Presidente, 'Privacy Regulation and Firm Performance: Estimating the GDPR Effect Globally', Working Paper No. 2022-1, The Oxford Martin Working Paper Series on Technological and Economic Change, 2022, 11-18.

[61] Michal Gal and Oshrit Aviv, 'The Competitive Effects of the GDPR, Journal of Competition Law and Economics', 2020, 4.

[62] Anil Jain, 'The 5 V's of Big Data', IBM Watson Health Perspectives, 2016.

[63] Tal Zarsky, 'Incompatible: The GDPR in the Age of Big Data, Seton Hall Law Review', Vol. 47, 4(2), 2017.

discouragement and data minimisation as "negative, innovation-constraining responses"[64] to the GDPR rules which reduce start-up innovators' access to input data and discourage the development of products and technologies highly dependent on Big Data, as a result of lack of data.[65]

One may state GDPR fails to achieve the desired free flow of data within the EU as data's protection *per se* is not the concern of the EU legislator here. Instead, GDPR aims to protect the individual person against any danger resulting from the use of their PD[66]. Although Rec 4 GDPR states data protection rights must be balanced against other fundamental rights and freedoms in the internal market as freedom to conduct business, some argue it is an unbalanced Regulation since it privileges the protection of the individual person over market integration.[67] One can speak with the GDPR one's interests of control and access to data have prevailed over any economic interest linked to the re-use of PD. Contrarily, EU initiatives as the NPND and the DGA show an opposite move, that we look in turn.

2.3. Non-Personal Data Law as a sidestepping from the GDPR

It is crucial to mention at the EU level there is no data covered by a classical property right, namely by the most wide-ranging one, the ownership right. Notwithstanding, in the words of Nuno Sousa e Silva: "the future may bring new forms of "ownership" over ideas, information or data"[68], especially considering much information is treated already "at least at the contractual

[64] Nicholas Martin and others, 'How Data Protection Affects Startup Innovation, Information Systems Frontiers' (2019) 21:1307–1324, 1318.

[65] *Ibid 1319.*

[66] Herbert Zech, 'A legal framework for a data economy in the European Single Market, Journal of Intellectual Property', 2016, 463.

[67] Streinz (36) 915.

[68] Nuno Sousa e Silva, 'Quando O Segredo É a 'Alma Do Negócio''– Definição De Um Conceito (When is There a Trade Secret – About the Concept), 2013, 5.

level as property, despite its intangible nature"[69]. That is the case when a provider (trader) supplies digital content or service to DS (consumers), and DS provide their PD to the former[70].

DS do not own their PD [71], yet they hold some rights under GDPR, as the right to access and get information PD, to rectification and erasure, to restriction of processing and to data portability (Arts. 13, 14, 16, 17, 18, 20 GDPR). These rights are grounded in informational self-determination[72], not property theory, enshrined in Rec 7 GDPR. In the words of Filipa Urbano Calvão what is at stake is "the protection and control over personal information by the DS on its own"[73].

The topic of whether should be established data ownership rights has been discussed by scholars, resulting in divergences. Data ownership or right to use data can be understood as "the allocation of data by means or at least along the lines of exclusive rights"[74], which implies data owners would be able to trade property rights.

Zech argues the establishment of a data ownership right could ensure a fair profit allocation in consequence of the data processing and analysis[75] between companies and right holders who would negotiate position and benefit. Contrarily, Purtova[76] and Drexl and others[77] are quite reluctant to the creation of a new exclusive right to use data. Purtova challenges the creation

[69] Nuno Sousa e Silva, 'Direito E Robótica - Uma Primeira Aproximação (Robots and the Law – a First Take)', Revista da Ordem dos Advogados, 2017, 532.

[70] Art. 3 Directive (EU) 2019/770 on certain aspects concerning contracts for the supply of digital content and digital services.

[71] Bjorn Lundqvist, *'*Big Data, Open Data, Privacy Regulations, Intellectual Property and Competition Law in an Internet of Things World, Faculty of Law*'*, University of Stockholm RP No. 1, (2016), 10.

[72] Florent Thouvenin, 'Informational Self-Determination: A Convincing Rationale for Data Protection Law?', 2021, 5.

[73] Filipa Urbano Calvão (7) 14.

2018, 14.

[74] Zech (66) 463.

[75] Zech, *'*Information as Property', JIPITEC, 6 (3), (2015), 197.

[76] Nadezhd Purtova, 'Do Property Rights in Personal Data Make Sense after the Big Data Turn?: Individual Control and Transparency', 10(2) Journal of Law and Economic Regulation November 2017, Tilburg Law School RP No. 2017/21, 2017, 1.

[77] Josef Drexl and others (11) 3.

and management of property rights in PD since, due to its dynamic nature, enforceability "might prove problematic"[78]. Drexl and others state data protection law does not motivate the control of data over the use of data[79]. This reasoning is based on Art. 1(3) GDPR, which states the free movement of PD throughout the EU shall not be restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of their PD. In parallel, from an economic viewpoint, the authors argue as long data is a public good and the principle of the public domain of information must prevail, the recognition of an exclusive right in data would interfere with other EU freedoms as freedom to conduct a business and fair competition[80]. In addition, Lynskey concludes a property rights regime would require a framework establishing statutory public domains, such as the exclusion of anonymous data, to limit individual control over data[81]. Finally, problems arise regarding the legal object of such right, since data is a changeable asset besides figuring out who is the owner of the data, since the creator of the data may not correspond to the person entitled to the data[82].

The EU's rationale seems to align with the majority of the literature on data ownership when in late 2018, as part of the Digital Single Market (DSM) strategy and Building a European Data Economy initiative in specific, the NPDR was published. This legislation aims to remove unjustified barriers to the movement of non-personal data in the EU and thus unlock the potential of a European data-driven economy. NPDR may be considered the European lift to the free movement of different types of data within the EU as GDPR proves to be insufficient to unleash the potential of data-driven innovation.

While addressing the obstacles to the free flow of other than PD within the EU and between informational systems, it aims to facilitate cross-border data exchanges by empowering companies to store non-personal data anywhere in the EU. When PD is at stake, as long as it is truly anonymised, meaning no direct or indirect personal identifiers and becoming non-personal data, the processing will not fall within the GDPR's scope but straightforward under NPDR rules. This creates incentives for companies to process data in an anonymised format to any

[78] Purtova (76) 15.

[79] Josef Drexl and others (11) *3.*

[80] *Ibid* 2.

[81] Orla Lynskey, *The Foundations of EU Data Protection Law* (Oxford University Press, 2015), 244-246.

[82] Andreas Rahmatian, 'Debts, Money, Intellectual Property, Data and the Concept of Dematerialised Property', Journal of Intellectual Property, Information Technology and Electronic Commerce Law, Vol. 11(2), 2020, 195.

data processing thereafter is not covered by GDPR's scope[83] and boosts *de facto* the free flow of data. At the same time, anonymised data may not be particularly useful for some purposes and the innovation potential of PD is hindered[84].

An issue related to the interplay of GDPR and NDPR arises when mixed datasets are at stake[85]. One can define a mixed dataset as a collection of data in which personal and non-personal data are present, picturing most datasets[86]. A parallel application is due as it is stated in Art. 2(1), 1st part NPDR: the PD portion falls under the scope of GDPR and the remaining non-personal data part falls under the NPDR regime existing whenever possible in separated data blocks[87].

The notion of PD (data directly or indirectly relates to an identified or identifiable natural person) is wide in the sense non-personal data might indirectly identify a natural person when mixed with other information and thus qualifies as PD[88]. Additionally, it is considered a dynamic and context-dependent[89] qualification as identifiability may vary throughout time. Barreto Menezes Cordeiro argues due to the technological developments PD's concept needs "punctual surgical amendments"[90].

[83] Emily M. Weitzenboeck and others, 'The GDPR and Unstructured Data: Is Anonymization Possible?', International Data Privacy Law, 2022, 2.

[84] Inge Graef, Raphael Gellert and Martin Husovec, 'Towards a Holistic Regulatory Approach for the European Data Economy: Why the Illusive Notion of Non-Personal Data is Counterproductive to Data Protection', TILEC Discussion Paper, DP 2018-028, 2018, 5.

[85] Raphael Gellert and Inge Graef, 'The EC's proposed DGA: some initial reflections on the increasingly complex EU regulatory puzzle of stimulating data sharing', TILEC Discussion Paper, DP 2021-006, 2021 5.

[86] EC, Guidance on the Regulation on a framework for the free flow of non-personal data in the EU COM(2019) 250 final, 2019.

[87] Nuno Sousa e Silva and Benedita Cunha Pinto, 'Internet das Coisas(IoT): alguns desafios jurídicos', 27.

[88] Lundqvist (71) 11.

[89] Graef, Gellert and Husovec (84) 5.

[90] António Barreto Menezes Cordeiro, *'Dados Pessoais: Conceito, Extensão e Limites'*, Centro de Investigação de Direito Privado, 2018, 3.

The fact is it is questionable whether nowadays a personal and non-personal data distinction can be drawn in practice, especially considering both growing availability of data points and algorithmic capacity to interlink DS and infer personal information from non-personal data.[91]

The application of Art.2 NODD becomes excessive burdensome and unusable in practice[92], leading to the GDPR's exclusive application. Besides, Art. 2(1) second part introduces ambiguously the term "inextricably linked" related to mixed datasets, stating GDPR's application to this type of dataset shall not be undermined. The EC comes to clarify it can refer to a situation whereby a dataset contains PD as well as non-personal data and separating the two would either be impossible or considered by the controller to be economically inefficient or technically unfeasible.[93] GDPR fully applies to the whole dataset, also when the PD's component is minimal as the applicable provisions of the GDPR must be fully complied with in respect to the PD part. One may assume this interpretation establishes a prevalence of GDPR over NPDR rules and shows how challenging it is to strike a balance between the added value around sharing non-personal data and protecting individuals' rights. This takes one question its proportionality since it seems to prove Purtova's opinion data protection risks to become the "law of everything"[94]. The focus is now on the DGA, namely to what extent it may bring a solution to this problem.

[91] Michèle Finck and Frank Pallas, 'They who must not be identified – distinguishing personal from non-personal data under the GDPR', International Data Privacy Law, 2020, 10 (1), 1.

[92] EDPS, Comments of the EDPS on a Proposal for a Regulation of the European Parliament and of the Council on a Framework for the Free-Flow of Non-Personal Data in the EU, 2018, 3.

[93] COM(2019) 250 final, 10.

[94] Purtova (32) 41.

Chapter III. EU-wide Governance Framework

3.1 The European way of Data Governance

Promoting a EU data-driven economy characterized by a dynamic ecosystem in which different players interact in the DSM is a longtime EU leading priority. As a response to the 2013's European Council focused on the digital economy, innovation and services, the EC has established an extensive digital transformation policy[95], focusing on boosting the EU data economy through data-driven innovation (business' and governments' capacity to make use of information from improved data analytics to develop better services and goods and improve the daily life of individuals and organisations, namely SMEs[96]).

Today's digitalization means the volume of generated data is increasing rapidly, however, it is the availability of data holds the potential to unlock the EU data-driven innovation in the public and private sectors, enabling economic growth, job creation and development of better-fit and cheaper products and services, based on cutting-edge technologies as AI and IoT[97].

A study[98] shows Europe as a whole is falling behind during a second wave of innovation areas as Quantum Computing and AI, i.e. transformational technologies intrinsically driven by and dependent on processing large amounts of data, being outpaced by the United States of America and China. Aware of that, the EC with the EDS intends to make further progress on EU data-driven innovation from which European citizens can benefit. The EC tries to catch up with those wealthier economies and compete for the next decades while unleashing data value for economy and social welfare.

[95] COM(2014) 442 final.

[96] *Ibid.*

[97] SWD(2020) 295 final (50) 1.

[98] Jacques Bughin and others, 'Innovation in Europe – Changing the game to regain a competitive edge*'*, McKinsey Global Institute, 2019.

Authors as König and Lundqvist argue the superior concern of EU digital policy starting in 2010 is of economic growth and competitiveness, being centered on data as a key resource[99] and the data protection's overriding goal are commercial reasons[100]. Despite this, we tend to second Thomas Streinz's way of seeing that, in what concerns PD, an "economistic framing of data as a resource"[101] which supports data availability was left behind until the EDS and the DGA, which will be explored in this Chapter.

The EDS calls for the creation of an internal market for data in which data could be used regardless of its physical storage location in the EU in a manner compliant with applicable law, a single European Data Space. It stimulates the creation of nine sectoral common European data spaces for data sharing and pooling in areas as health, the financial sector, manufacturing, public administration and tourism, and the European Open Science Cloud. These data spaces should make data findable, accessible, interoperable and re-usable (FAIR principles), guaranteeing cybersecurity. A "European data space" refers to a data infrastructure with tailored governance mechanisms will enable secure and cross-border access to datasets in the targeted thematic areas[102].

For the EU to take a leading role, the EC states the essentiality of creating a clear and enabling legal framework for secure data sharing and increased data availability based on improved governance structures, which ensure the quality of data available for cross-sector use and re-use in the common sectoral data spaces. Besides, the EC presented its intention to enable the use of data held in public datasets for scientific research purposes and facilitate the use of individuals' data for the public good – data altruism, both in a manner compliant with the high level of protection of data protection law.

[99] EC, Digital Europe Programme, Call for Proposals - Preparatory Acts for Data Spaces, DIGITAL-2021-PREPACTS-DS-01, 6.

[100] Lundqvist (71) 13.

[101] Streinz (27) 948.

[102] EC, Annex to the Commission Implementing Decision on the financing of the Digital Europe Programme and the adoption of the multiannual work programme for 2021 – 2022, C(2021) 7914 final, 19.

Given that, those are the issues tackled in the DGA which creates the basis for a European way of data governance, applicable to public bodies and independent actors[103]. The EU tries to strike a balance between a wide use, free flow of data and the maintenance of the EU fundamental values as high privacy, security, safety and ethical standards. Relevance is given to greater data infrastructures and governance mechanisms at the EU level however, no provisions defining a clarification of data governance or governance can be found neither in the DGA *per se* nor in the EDS.

A unified definition of data governance cannot be found since it differs depending on the research field. From an information system's view in which an intra-organisational governance is embodied, Abraham and others resume data governance as the authority and control exercise by an enterprise over the management of strategic data assets based on a cross-functional framework in which decision rights, accountability, data policies and procedures are formalized[104]. This notion of data governance reflects the "micro economic corporate management concept" around data governance Hoffmann and Otero[105] describe. In the governance and regulation research field, this understanding of data governance corresponds to Regulation or regulatory governance as a governance's sub-class Black defines as an "organised attempt to manage risks or behaviour to achieve a publicly stated or set of objectives"[106].

In addition, Grafenstein draws a more society wide-level or sectorial approach to data governance denotes data governance's objective is, from a multi-stakeholder way of seeing, to manage and accommodate divergent and conflicting interests in data use, in which regards the

[103] Kristina Schreiber, Patrick Pommerening, Philipp Schoel, '*The New Data Governance Act A Practioner's Guide'*, Nomos Beck Hart

[104] Rene Abraham, Johannes Schneider and Janvom Brocke, 'Data Governance: A conceptual Framework, structured review, and research agenda', International Journal of Informational Management, 49, 2019, 425.

[105] Jörg Hoffmann and Begoña González Otero, 'Demystifying the Role of Data Interoperability in the Access and Sharing Debate', Max Planck Institute for Innovation & Competition RP No. 20-16, 2020, 21.

[106] Julia Black, 'Learning from regulatory disasters', LSE Law, Society and Economy Working Paper, 4/2014, 2014, 3.

value and risk involved –"value-for-risk dilemma"[107]. The author decomposes a data governance framework into three analytical layers: regulatory layer (laws concerning the collection, sharing and re-use of data along with enforceable contractual agreements); organisational layer (structures, processes and practices entangled) and finally, the technology layer (software and hardware infrastructure involved in the data processing).

One may argue the latter better reflects the European approach to data governance. EU law intends to regulate data through the development of a non-static and shapable framework for data governance with common rules and practices in the MS in respect to fundamental rights, complementing previous legislation concerning competition, privacy and access to and re-use of information. We can argue the DGA is mainly made up of regulatory and organisational layers to foster trust and data availability with few references to computer software and hardware infrastructures in specific.

3.2. Summary description of the DGA

The non-rival and non-exclusive nature of data and of the access to and use of data[108] means different market players can use data unlimited times without diminishing the other's use[109], adding an argument in favour of data sharing and OD availability[110].

Yet, data sharing continues to be limited in the EU due to the following obstacles reported[111]. First, there is a lack of trust in data sharing since companies fear a competitive advantage loss when engaging in data sharing and re-users do not comply with the contractual agreements. Data intermediaries cannot scale up due to the low trust in services. Second, in general, PSB lack the technical and legal capacity to deal and process requests to re-use specific categories

[107] Maximiliam Grafenstein, 'Reconciling Conflicting Interests in Data through Data Governance. An Analytical Framework (and a Brief Discussion of the DGA Draft, the Data Act Draft, the AI Regulation Draft, as well as the GDPR)' (May 9, 2022), HIIG Discussion Paper Series No. 2022-02, 2022, 5.

[108] Zech (66) 195.

[109] Hal Varian, 'AI, Economics and Industrial Organisation', NBER Working Paper 24839, 2018, JEL N. L0, 9.

[110] Bertin Martens, 'The Impact of Data Access Regimes on AI and Machine Learning', JRC Digital Economy Working Paper 2018-2019, 2018.

[111] SWD(2020) 295 final (50).

of publicly held DS to the rights of third parties. Namely, concerning requests to use PD, PSB are not sufficiently equipped to make the data available for use in a manner GDPR-compliant due to the lack of technical privacy-enhancing mechanisms, hindering potential use for machine learning and research purposes. Although DS are more and more willing to share PD for the common good and scientific research purposes[112], until now, there was a lack of established data altruism rules and processes, hence PD sharing remains not adequately developed in the EU. Also, technical obstacles related to interoperability, as shortage of common standards, findability and uncertainties regarding data quality have hindered cross-border access and re-use by different stakeholders.

The DGA translates an action to tackle those issues at the EU level, rather than by the MS separately, and reinforce the single market for data by creating a harmonised governance framework for data exchanges. The DGA's objective is to set conditions for boosting the European data spaces' development, leaving room for applying rules targeting specific sectors.

Contrarily to GDPR, the DGA does not defining its territorial scope, albeit it expressly introduces GDPR-like safeguards to govern international transfers of non-personal data held by the public sector in Art. 5.

The DGA aims to create a legal regime for the re-use, within the EU, of specific categories of protect data publicly held by PSB (Chapter II). Re-use must be interpreted as the use by natural or legal persons of data held by PSB, for commercial or non-commercial purposes other than the initial purpose within the public task for which the data were produced, except for the exchange of data between PSB purely in pursuit of public tasks – Art. 2(2) DGA. It intends to establish a notification and supervisory framework for data intermediation services (Chapter III) and a framework for voluntary registration of entities collect and process data made for altruistic purposes (Chapter IV). Lastly, to prevent fragmentation between the MS, it creates a framework for the establishment of a supervisory European Data Innovation Board (Chapter VI).

[112] *Ibid.*

The DGA's complementary Proposed Data Act[113], addresses issues about who can use and access data generated in the EU across all economic sectors, with a main focus on business to government (B2G) and business to business (B2B). As PSB are not legally obliged to make protected data available under the DGA, they fall out of the scope of Chapter III of the Proposed Data Act. Subsequently, the re-use of publicly held protected data – government to business (G2B) data sharing – is not targeted by the Proposed Data Act. Hence, no further developments regarding the Proposed Data Act will be addressed in this work.

3.3. DGA and NODD: Complement or Overlap?

PSB as National Bureaus of Statistics, Weather Institutes, Patent Offices or Hospitals collect, produce, reproduce, and disseminate a vast amount of data at the cost of public budgets. PSI data is used to fulfil public tasks or provide services of general interest by PSB and public sector undertakings (PSU). A data-intensive cycle happens in political, economic, legal, social or geographical areas.

The availability and re-use of the data can contribute to improve competition in the internal market through the development of new data-driven applications, promoting economic growth and social participation. That was the PSID's intention to set minimum standard rules for the development of practices and policies - to make existing documents held by PSB and PSU of the MS available for dissemination and re-use.

MS have started to establish re-use policies and making documents available as OD, defined in Chapter 2.1, which led to a divergent set of rules and practices across MS linked to PSI's exploitation. Since it risks to create a barrier to the cross-border offer of products and services and hindering pan-Union application based on comparable datasets, minimum harmonisation was needed and settled with the NODD.

The NODD came to boost the EU's data economy by increasing the availability of valuable publicly held and publicly funded data for re-use, a minimum harmonisation framework for national rules and practices regarding what public data are available for re-use, considering consistency with the relevant general and sectoral specific access regimes. It aims to tackle the

[113] COM(2022) 68 final.

provision of real-time access to dynamic data through adequate technical means, the emergence of new forms of exclusive arrangements and exemptions to the principle of charging the marginal cost.

The NODD's regime governing re-use and practical arrangements for facilitating the re-use applies to existing documents (any content or part of such content whatever the medium (paper/electronic form), sound, visual/audiovisual recording) held by PSB – Art. 2(6) NODD. It applies to documents held by selected PSU (Art. 1(b) NODD), along with to publicly funded research data, which show an augmentation of the scope of application when compared to PSID.

The NODD lays down a general obligation for PSB to make all existing documents reusable for commercial or non-commercial purposes[114], unless access is restricted or excluded under national law or subject to other exceptions. On the contrary, PSU are not obliged to do so and retain the right to authorise or not the re-use of existing documents.

Existing documents shall be made available for re-use by PSB, following a request lodged by an applicant/potential re-user (a person/legal entity able to re-use documents held by PSB or PSU for commercial or non-commercial purposes, other than the initial purposes within the public task or to provide services in the general interest for which documents were produced)[115].

The PSB shall process requests for re-use, whenever it is possible and appropriate through electronic means. If so, the PSB shall deliver the documents for re-use or finalise the needed licence offer the applicant within a reasonable time consistent with time limit established for the processing for access to documents in national regimes. In the absence of such provisions, PSB shall process the request and finalise the decision within 20 working days of receipt, which may be extended by other 20.[116].

The re-use shall in principle be free of charge, nonetheless PSBs may charge for marginal costs of data reproducing, providing or disseminating information, as well as for the process of anonymisation of PD or actions may take to protect commercial confidential information.

[114] Arts. 2 (11)(1) &3 NODD.

[115] *Ibid.*

[116] Art. 4 NODD.

Under Art. 12 NODD, re-use shall be open to every market actor, even if other market actors already exploit added-value products, avoiding lock-in effects. This denotes the non-rivalry data nature and Varian's way of seeing[117] - the focus when data is at play should be on data access rather than on data ownership.

The PSB and PSU are encouraged to produce and make the existing documents fall within the scope of the NODD in accordance with open by design and by default principle. Besides, conditions for re-use are established in Chapter III NODD. When delineating technical solutions for the documents' re-use, recommendations to guide and strengthen interoperable public services must be considered by MS. Data interoperability and access are "inherently intertwined"[118] concepts, dependent on each other. Data interoperability, which ensures systems work together, includes not only syntactic interoperability (interoperation of data format and structure used whenever information/services are shared between distinct parties), but also semantic interoperability (involved agents, services and applicants exchange information in a meaningful manner and formats are preserved, on and off the web)[119]. Application Programming Interfaces (APIs) and data standards are considered the primary technical enablers of data interoperability, hence only the provision of information via these means may guarantee efficient access and re-use of data, boosting data-driven innovation and economy.

Art. 5(1) NODD states PSB and PSU shall make documents and linked metadata available in any pre-existing format/language and, whenever possible and appropriate, in electronic means in open, machine-readable, accessible, findable and reusable formats and in a manner compliant with open data standards. The notion of possible and appropriate is vague, yet EU legislator in Art. 5(3) and (7) tries to clarify in general there is no obligation for PSB or PSU to comply with those conditions for re-use, including the creation, adaptation or provision of extractions, when these actions go beyond a simple operation and involve disproportionate effort. Only when dynamic data and high-datasets are at play there is an obligation for the PSB and PSU to provide information via suitable APIs, enabling machine-to-machine data exchanges[120], and as a bulk

[117] Varian (109) 9.

[118] Hoffmann and Otero (105) 9.

[119] *Ibid* 9.

[120] *Ibid* 13.

download, whenever relevant. This raises questions regarding what constitutes a simple operation and what makes an effort disproportionate or proportionate.

One can easily argue the general provisions of the conditions for re-use require and expect a less active role from PSB and PSU than the desirable and necessary for data be considered OD material and allow the development of better products and services. According to Ubaldi, to achieve openness[121], data needs to be available and accessible to anyone - meaning available in convenient, modifiable and disaggregated form. Besides to ensure re-use, redistribution and the opportunity to inter-mix datasets, data needs to be provided in a timely manner and an open, machine-readable format (includes XML, XSLT formats and excludes PDF files

According to the OD Maturity Report 2022[122], different OD policies are established in the 27 MS. While in Denmark and Portugal, the NODD implementation law continues to be core OD frameworks, Bulgaria has gone a step forward[123]. Bulgarian E-Government strategy 2019-2025 embeds principles, as reduction in the use of unstructured data and establishes data generated by the public sector shall be in an open-machine-readable format to allow for reuse. In addition, in Sweden, the European Interoperability Framework[124] is implemented by the Agency for Digital Government to support Sweden's public sector in publishing data for re-use and a new law has been proposed to transpose the NODD into national law[125]. However, across the EU, PSB continue to lack legal and technical abilities and legal expertise to make data available and deal with re-use requests, namely in which regards data protected by third parties' rights and PD in specific.

In this respect, the DGA's Chapter II creates a legal regime for the re-use of certain categories of protected data held by PSB not subject to obligations for PSB to make it available for re-use for commercial and non-commercial purposes under the NODD, excluding data held by PSU.

---

[121] Barbara Ubaldi, 'Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives", OECD Working Papers on Public Governance, 22, 2023, 24.

[122] Giulia Carsaniga and others, Open Data Maturity Report 2022, Data.europa, eu, 2022, 17.

[123] See Danish Act Amending the Law on the Reuse of Public Sector Information, of 10 May 2021; Portuguese Law no. 68/2021 and Digital Bulgaria Program 2025.

[124] European Interoperability Framework – Implementation Strategy, COM(2017) 134 final

[125] Available at: https://www.regeringen.se/rattsliga-dokument/proposition/2022/04/prop.-202122225

Besides, contrarily to the NODD, it seems the DGA does not create a legal obligation but rather an option[126] on PSB to allow the re-use of such data.

The categories subject to re-use under the DGA fall outside the scope of the NODD[127], which excludes, among others, documents covered by rights of third parties or containing PD. One can argue the DGA complements and extends the NODD's scope of application and function in this respect[128] however, the fact those documents be adapted by PSB to be accessed and re-used[129] raises concerns regarding a possible overlap and legal uncertainty on the applicable regime.

De-personalized data (data once was PD but has been rendered and manipulated to qualify as anonymous data[130], not being subject to the GDPR rules) included in a document do not fall outside the scope of the NODD and can be made available for re-use. That is the case established in France, ranked second in 2019 OECD Open Usable and Re-usable (OUR) Data Index[131], where national law[132] lays down an obligation for the PSB to anonymise PD before allowing re-use. Yet, according to the EDPB, former A29WP, the re-use of PD under assessment must be made on a case-by-case analysis[133].

[126] Julie Baloup and others, 'White Paper on the DGA, CiTiP Working Paper', KU Leuven Centre for IT & IP Law, 2021, 16.

[127] Art. 1(2)(d), (h) NODD

[128] Irma Klünker and Heiko Richter, 'Digital Sequence Information between Benefit-Sharing and Open Data – How to Advance the Legal Framework', Max Planck Institute for Innovation and Competition RP No. 22-1, 2022, 23.

[129] Baloup and others (126) 17.

[130] Michèle Finck and Frank Pallas (91) 1.

[131] OECD, OECD Open, Useful and Reusable data (OURdata) Index: 2019, 2020, 20.

[132] L 312-1-2, *Code des Relations entre Le Public et L'Administration (CRPA).*

[133] A29WP, Opinion 6/2013 on Open Data and Public Sector Reuse, 2013, 6.

Chapter IV. Surpassing barriers to the re-use of personal data

4.1. Re-use conditions and institutional requirements for PSB under the DGA

Aware of the importance and value potential of allowing for more data to be available for re-use, namely data held by PSB, some MS have established structures, processes or legislation at the national level to facilitate the re-use of protected data, namely PD. In Germany, research data centres prepare and make available research data for scientific purposes[134]. Besides, in Finland, the Health and Social data permit authority Findata, whose activities are based on the Act on the secondary use of health and social data[135], grants access to health and social data for re-use for specific purposes as: scientific research, statistics, development and innovation operations or education. Being considered "highly advanced in its approach to facilitate access to data for research"[136], Findata's legal bases for data processing are Arts. 6(1)(e) and 9(2)(g) GDPR. Yet, other MS have not started to legislate in this respect. An uncoordinated way of tacking the "raw material of the DSM"[137] exacerbates legislative and administrative fragmentation in the EU. This triggers an efficient EU-level intervention to unlock the potential of re-using data held by PSB.

Given that, with the aim to unlock the socio-economic potential underlies data held by the major data suppliers in society[138], while preserving European rights and values, in the DGA's Chapter

[134] *Rat für Sozial- und Wirtschaftsdaten* (German Data Forum), 'The German Data Forum (RatSWD) and Research Data Infrastructure: Status Quo and Quality Management', 2018, 7.

[135] Chapter 1, Sector 1, Act on the Secondary Use of Health and Social Data, Finland Ministry of Social Affairs and Health 552/2019.

[136] Johan Hansen and others, 'Assessment of the EU MS' rules on health data in the light of GDPR Specific Contract No SC 2019 70 02 in the context of the Single Framework Contract Chafea/2018/Health/03´, DG Health and Food Safety, EC, 2021.

[137] COM(2014) 442 final, 3.

[138] Gustavo Magalhães and Catarina Roseira, Open Government and the private sector: An empirical view on business models and value creation, Government Information Quartely, 37, 2020, 101248, 2.

II the legislator aims to set a harmonised legal regime built on basic conditions under which PSB may allow the re-use of certain types of data protected by law and cannot be made freely available as OD under the NODD's regime. As such, the provisions outlined encourage the PSB to make available a broader range of data such as commercial business information, namely trade secrets, PD and data covered by intellectual property rights[139]. Regarding PD in specific, it is DGA's objective to encourage the re-use of PD in public datasets, which are frequently not made available[140] for research purposes and innovation activities in the public interest, despite that possibility under the GDPR, E-Privacy and Law Enforcement Directives.

The DGA does not create a right to re-use data but lays down conditions under which re-use would be permitted. PSB are not obliged to allow, being just encouraged to facilitate it[141]. G2B data sharing turns out to be only voluntary, rather than mandatory. This may lead one to question what the actual impact of DGA in the re-use of these categories of data will be, since it seems more of a modest than an effective measure. In any case, it is important to state what are the underlying institutional requirements for PSB and conditions for re-use as follows.

The DGA opts for a low-intensity regulatory option and calls for the instalment of a data re-use single information point/one stop shop[142] where potential data users can find a publicly available searchable asset list[143] which contains an outline of available data resources at a sectoral/regional/local level, with significant information - the data format, size and conditions for re-use.

MS should designate, establish or contribute to the competent bodies' establishment to support the PSB in charge of granting access or refusing re-use. Based on Rec. 26 DGA, competent bodies should be entrusted with the role to provide assistance to PBS on how to structure and store data through APIs' implementation and to make data interoperable, transferable and

[139] Art. 5 DGA.

[140] Rec. 6 DGA.

[141] Art. 1(2) DGA. Additionally, the fact that it is voluntary might raise the question of whether allowing for the re-use of data and guaranteeing the conditions for the allowance should be considered a public task in the public interest laid down by law.

[142] SWD(2020) 295 final (50) 24.

[143] Art. 8(2) DGA.

searchable, while considering the processing's best practices, regulatory and technical standards and secure processing environments (SPEs).

DGA recognises the importance of protecting the sensitive nature of data protected by third parties' rights, to guarantee respect for such rights or to diminish the negative impact on fundamental rights, the principle of non-discrimination and data protection. To allow a broader range of protected data to be available for re-use, specific technical and legal measures must be implemented. Art 5(3) DGA requires PSBs or competent bodies, after the access request and before granting access for the re-use of protected data, to ensure PD has been anonymised, so it does not allow the identification of DS and commercial confidential information rendered, namely modified, aggregated, treated or by other means so no confidential information is disclosed. These obligations are considered "challenging and costly"[144] to fulfil by the PSB since they are not entities specialised in data processing and not sufficiently equipped to provide those environments. Yet, competent bodies must "have adequate legal and technical capacities and expertise to be able to comply"[145]. The voluntary entrustment of competent bodies with those roles may be considered essential to guarantee the mechanisms for access and re-use of data are simple[146] and data is efficiently accessible and reused. Otherwise, PSB risk not having the capacity to fulfil legal obligations, which may lead to a chilling effect and preclude data re-use and innovation.

Like the NODD, to foster a competition environment and avoid "risk that powerful players in the market get exclusive access to the data"[147], conclusion of exclusive access agreements between PSB and re-users for the re-use of publicly held protected data which might have the objective or effect the creation of exclusive rights are restricted to circumstances where an administrative act or contractual agreement are justified and necessary for the supply of a product/provision of a service, both in general interest, with the maximum duration of 12 months[148]. The exclusive right to re-use data shall be granted in accordance with EU or national law based on necessity requirements which shall be transparent and made publicly available

[144] Baloup and others (126) 25.

[145] Art. 7(4) DGA.

[146] American Chamber of Commerce to the EU, 'Our Position DGA', 2021, 5.

[147] SWD(2020) 295 final (50) 25.

[148] Art. 4(4) DGA.

online. A regular review should be conducted to verify whether the grant of the exclusive right remains necessary, based on a market analysis. As such, in principle, exclusive agreements are to be avoided.

Based on Art. 5(1) DGA, PSB, making data available under national law, should set and publish the conditions upon which re-use of protected data would be allowed. The conditions should be restricted to those necessary to guarantee the rights and interests of third parties in the data and the integrity of the PSB information systems. As such, these conditions must be non-discriminatory, transparent, proportionate and objectively justified, regarding the data category and nature and re-use purpose.

One may see PSB's transparency requirement is line with the transparency duty under Art. 12 GDPR. DGA considers the purpose for data re-use, contrarily to the NODD which makes data available for undefined purposes and may expose parties to "severe interference with their rights and protected interests"[149]. This shows what Baloup denotes "a shift from the open data approach to purpose-based re-use of data"[150] to enhance the amount of publicly held data which can be re-used. The final free decision of the PSB or the competent body on the request for re-use should be provided within 2 months from the receipt of the request and may be contested before MS' Courts, according to Art. 9(1) and (2) DGA.

Non-discrimination access to data is a bottom-line condition to prevent the creation of "new forms of economic exclusions"[151], hence re-use conditions may not restrict competition. However, the DGA incite MS to promote access data by SMEs and start-ups to foster innovation, which is the main goal of the EDS. This approach which privileges smaller entrants significantly differs from the GDPR's one which does not provide derogations based on data operators' size since it is not considered an "indication of the risks the processing of PD it undertakes can create the individuals"[152].

[149] Baloup and others (126) 18.

[150] *Ibid* 18.

[151] Ubaldi (121) 25.

[152] COM(2020) 264 final (36) 23.

Additionally, DGA outlines the conditions for re-use should promote and privilege re-use of protected data for scientific research, which should not be considered as discriminatory as well[153]. Here DGA's approach goes in line with the GDPR, which privileges further processing for scientific research purposes[154] and does not distinguish commercial and non-commercial scientific purposes. It is true most scientific research has gradually been moved to private sector[155], which often conducts applied research ("original investigation undertaken to acquire new knowledge, directly primarily towards a specific practical aim or objective"[156]) - private sector's share of global expenditure on science represents around 70%.[157] Current wording suggests a larger player requesting data access for commercial scientific purposes may benefit from the privileged access regime.

However, DGA does not oblige PSB to allow the re-use of data held by them to a potential data user. A natural or legal person qualifies as data user when has lawful access to personal or non-personal data and the right to use data for commercial and non-commercial purposes – Art. 2(9) DGA. Then, PSB as data holder (a legal person who has the right to grant access to certain personal or non-personal data according – Art. 2(8) DGA) may grant or reject an access request made by the applicant, through the single information point.

Regarding fees, in contrast to the NODD which establishes a free of charge re-use of data as a principle, enabling the charge of the marginal cost for the re-use[158] and above that in few specific events only, the DGA seems to leave more possibility for PSB to charge fees for allowing data re-use based on the costs of processing it requests[159]. Such fees should be transparent, non-discriminatory, proportionate, objectively justified and not restrict competition. It should be limited to the necessary costs connected with data reproduction, provision and dissemination, the clearance of rights, acts of preparation of PD and

[153] Rec. 15 DGA.

[154] Art. 6(4) GDPR.

[155] Manuela Fernández Pinto, 'Open Science for Private Interests? How the Logic of Open Science Contributes to the Commercialization of Research', 5 Frontiers in Research Metrics and Analytics, 2020, 2.

[156] OECD, 'Research and Development (R&D)', OECD iLivrary, 2015.

[157] International Science Council, 'Science in the Private Sector', 2022.

[158] Art. 6(1) NODD

[159] Mirelle van Eechoud, 'Study on the Open Data Directive, Data Governance and Data Act and their possible impact on research', EC, 2022, 27.

commercially confidential data, as anonymisation, the maintenance of the SPE, the right's acquisition to allow re-use by third parties outside the public sector and the provided assistance in seeking consent from DS and permission from data holders whose rights and interests may be affected by re-use. Findata services consider the fee for the data request or data permit, working hours for combining, pre-processing, pseudonymising and anonymising the data and remote access environment.[160]

In the event anonymisation or modification of data do not respond the re-user's needs, remote access within a SPE controlled or provided by the PSB may be allowed if it does not weaken the rights and interests of third parties and subject to fulfil any requirements to carry out a DPIA (Arts. 35-36 GDPR) and consult DPAs. That is the case of the remote processing environment for researchers provided by the Finish Data Permit Authority which has proved to meet the Secondary Act's high requirements in a 2021 data security audit[161]. In a private cloud solution based on the ePouta platform researchers have a separated workspace, which can be accessed remotely[162]. It is crucial to mention legislator's option to enable the processing of data within the boundaries of a SPE provided by the data re-user, such as data processing in the cloud, assigning the supervision to the PSB only. This may be interpreted as a bold attempt of the legislator to overcome the lack of resources of public administrations and make (more) data available in a GDPR-compliant way[163].

In case the risk to the rights and interest of protected parties are not considered to be minimal, access and re-use of the data within physical premises can be allowed.[164] When on-premises or remote re-use of data within a SPE is allowed, PSB should impose conditions that guarantee the integrity of the functioning of the SPE technical systems. An on-premises or virtual SPE should allow innovative processing of data to which access can be granted under conditions controlled by PSB, in a compliant way – safe reading rooms for data.[165] PSB are able to set and supervise data processing actions (data display, storage, download and exportation) along with the calculation of derivative data through algorithms. PSB not only should retain the right to

[160] Johan Hansen and others (136) 103.

[161] ICT for Brilliant Minds, 'Audit gives green light to the data security of the Findata's operating environment'.

[162] *Ibid.*

[163] SWD(2020) 295 final (50) 13.

[164] Art. 5(3)(c) DGA.

[165] SWD(2020) 295 final (50) 25.

verify the process, the means and results of data processing pushed forward by the re-user to protect the data, but also the right to prohibit further use of the results whenever it compromises the rights and interests of others – Art. 5(4) DGA. This possibility for PSB to block the use of the results may discourage innovation and research by potential data re-users, creating a chilling effect[166].

The re-use of commercial confidential information by a re-user should be subject to the adherence to a confidentiality obligation, prohibiting the disclosure, as a non-disclosure agreement that re-user may have gathered that compromises third parties' rights. Regarding de-personalized data, re-users should be prohibited from re-identifying the DS to whom data relates and take technical and operational measures to prevent DS's re-identification. If that happens, the re-users shall inform the legal persons whose rights and interests may be affected. Besides, re-users shall notify in the event of a data breach results in the re-identification of the DS.

The fact Art. 5 DGA's current phrasing on the conditions for re-use suggests that each MS could have its own conditions' criteria to allow the re-use of data may lead to lengthy and fragmentated processes between MS[167], rather than an efficient response from the PSB and the EU, which in the end hinder the free flow of data within the EU.

4.2. Parallel regime for the re-use of protected data for scientific research and by SMEs and start-ups

The DGA leaves each PSB with the responsibility to decide the conditions under which data would be permitted for re-use, therefore actual risk of fragmentation and burdensome complexity for potential data re-users may remain.

It is clear the privileged position the re-use for scientific research purposes in the public interest and by SMEs and start-ups occupies within the DGA. PSB are encouraged to facilitate the

[166] American Chamber of Commerce to the EU (146) 5.

[167] *Ibid 5.*

access to data by SMEs and start-ups "which find more difficult to collect relevant data"[168] and to design conditions for re-use in a way that favours such purposes as a rule, not being considered a discriminatory condition, in a manner compliant with State aid rules. Conducting scientific research contributes to stimulate both scientific knowledge and economic development[169] as well to address real world problems such as the case of scientific research in the context of the COVID-19 pandemic.

The DGA distinguishes between commercial and non-commercial purposes in fees. PSB are encouraged to charge a zero or discounted fee for non-commercial research purposes (*any type of research-related purpose regardless of the organisation or financial structure of the research institution in question, with the exception of research that is being conducted by an undertaking with the aim of developing, enhancing or optimizing products or services* – Rec. 25 DGA). PBS are encouraged to give the same benefits to SME and start-ups as well as civil society and educational establishment, providing a list of privileged categories of re-users is made public. That said, one may assume the legislator do not incentivise PSB to provide data at a discounted fee to R&D departments of commercial companies (which do not qualify as SMEs or startups).

Regarding re-use of PD in particular, European legislator states, despite the possibility of using PD for research or innovation in the public interest, there is an "insufficient use of such data"[170] due to sensitive nature. PSB find it challenging to process re-use requests of PD as they lack appropriate legal and technical capacity[171] to safeguard the DS's rights and interests.

With DGA, EU legislator tries to reconcile privacy and innovation by facilitating the re-use of PD, whose value is high for research and AI applications. Besides approaching consent as the "last resource"[172] legal basis for the lawful processing of personal under the GDPR for research

[168] Rec. 25 DGA.

[169] Paul Quinn, 'Research under the GDPR – a level playing field for private and public sector research?', Life Sciences, Society and Policy 2021', 17:4, 2021, 1.

[170] Rec. 6 DGA.

[171] EC, 'Synopsis report of the public consultation on the revision of the Directive on the reuse of public sector information', 2018.

[172] Access Now, Position on the DGA, 2021, 5.

and innovation, DGA sets PSB should have technical means to guarantee the protection of DS and make SPE available.

The DGA does not create a new legal basis under the GDPR for data holders to share PD without proper authorization, as stated in its Rec 5. If PD is at stake, data processing should be based on the legal bases set out in Arts. 6 and 9 GDPR[173].

As it is often not possible to fully identify a specific purpose of PD processing for scientific research purposes at the time of data collection, GDPR "introduces some flexibility" [174] on specification and embodies a broader notion of DS's consent to certain areas of scientific research in Rec. 33. However, the request for DS's consent may undermine research's objectives, becoming "ill suited"[175] and extremely burdensome to contact each DS when big pools of data are at play[176]. This is in conformity with DGA's rational in Art. 5(6) which states by default data holders will not be asked to provide consent[177]. Only when PSB cannot ensure PD has been anonymised nor provide access within a SPE nor GDPR legal basis for transmitting data can be found, it is PSB's obligation to make best efforts to help potential re-users in seeking DS' consent. DGA's current wording does not define best efforts albeit it states the assistance must be "feasible without a disproportionate burden on the PSB*"*.

The impossibility and unfeasibility to contact subjects is already reflected in Arts. 14(5)(b) and (5)(c) GDPR. In those cases in which PD have not been obtained from the DS, controllers (re-users) are excepted from providing information to DS[178] (i) whenever it proves impossible/ would involve disproportionate efforts, namely for processing for scientific research purposes under Art. 89(1) GDPR's conditions and safeguards or (ii) is likely to render impossible/ seriously impair the processing's objectives or (iii) where the obtaining or disclosure of PD is

[173] Rec. 7 DGA.

[174] Graça Canto Moniz, '*Manual de Introdução à Proteção de Dados Pessoais'*, (1st edn Almedina 2023), 77.

[175] Marcello Ienca, Effy Vayena and Alessandro Blasimme, 'Big Data and Dementia: Charting the Route Ahead for Research, Ethics, and Policy', 5 Frontiers in Medicine, 2018, 5.

[176] *Ibid* 8.

[177] Access Now, (172), 5.

[178] Art. 14(1)(2) GDPR. The information to be provided includes, namely, the controller's identity and contacts, where applicable, of the controller's representative and data protection officer, the purposes of the processing as the legal basis for the processing as well as the information it came from publicly accessible sources.

established in EU or MS law to which the controller is subject and provides appropriate measures to protect the DS's legitimate interest.

The DGA notably favours anonymisation when it states PSB or the competent body have the obligation to ensure PD has been pre-processed and "*anonymised*"[179], to grant access for re-use and not to allow the DS' identification.

One may question whether in today's world with technological developments absolute anonymous data is still possible, matter which will be analysed in the following chapter.

Anonymous data no longer qualifies as PD being out of GDPR's scope[180]. Hence, the processing of this data "does not trigger the application of the data protection law"[181](eg. data minimisation and purpose limitation). On the contrary, pseudonymous data is still "explicitly and importantly, personal data"[182] since it can be attributed to a specific DS with the use of additional information.[183] It is crucial to refer the legislator removed the wording "pseudonymise personal data" which was present in the Art. 5 DGA initial draft. This goes in line with A29WP's opinion which states "pseudonymisation is not a method of anonymisation"[184]. Pseudonymous data diminishes risk of direct identification to DS[185], playing an important role in the GDPR as an "useful security measure"[186] under Art. 32 GDPR. PSB and competent bodies may enable the re-use of pseudonymous data within SPE, having the re-users a limited access to this data. Pseudonymous data may be more valuable than anonymous data for researchers due to its individual-level granularity and possibility to easily match data

[179] Art. 5(3)(a)(i) DGA.

[180] Rec. 26 GDPR.

[181] Purtova (32) 4.

[182] Lilian Edwards, *'Data Protection: Enter the GDPR',* Law, Policy and the Internet (2018), 15.

[183] Art. 4(5) GDPR.

[184] A29WP, Opinion 05/2014 on Anonymisation Techniques, WP216, 2017, 3.

[185] Mourby M and others, 'Are "Pseudonymised" Data Always Personal Data? Implications of the GDPR for Administrative Data Research in the UK', 34 Computer Law & Security Review, 2018, 223.

[186] A29WP (184) 3.

records[187]. This rational is in conformity with Ohm's statement that "data can be either useful or perfectly anonymous but never both"[188].

When PSB cannot ensure PD is pre-processed and fully anonymised, nor able to guarantee a SPE for the re-use of PD, it is essential to analyse whether there is a legal basis for transmitting the data under the GDPR, other than consent (Art. 5(6) GDPR). Otherwise, the PSB assisted by the competent bodies shall support re-users in seeking consent, whenever it is feasible and does not lead to a disproportionate burden.

In Rec. 159 GDPR, the European legislator "interprets scientific research broadly"[189] and does not distinguish between scientific research pursuing public interest from private or purely commercial ones. Distinctive potential legal bases may be available to researchers[190] - use of consent, processing is necessary for compliance with a contract or legal obligation, reasons of public interest or legitimate interest pursued by the controller or by a third party (Art. 6 (1)(a)(c)(e)(f) GDPR).

Under Art. 6 (1)(e) GDPR, both public and private entities conducting research could furnish a legal basis for processing PD, in the absence of the DS's consent, "for the performance of a task carried out in the public interest or the exercise of official authority vested in the controller".

In practice, the different legislation of the MS varies in terms of the extent the general interest legal basis can be used by commercial research form, as SMEs or start-ups. That said, the probability specific national law facilitates the use of the public interest legal bases for lawful processing of PD by commercial organisations is low.[191] Poland implemented the exemptions

[187] Information Commissioner's Office, 'Anonymisation: Managing Data Protection Risk Code of Practice', 2012, 36.

[188] Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization', UCLA Law Review, U of Colorado Law Legal Studies RP No. 9-12, 2019, 4.

[189] Tiago Rodrigues de Oliveira, '*Direito da Proteção de Dados Perspetivas Públicas e Privadas'* (Domingos Soares Farinho, Francisco Paes Marques and o de Freitas, 1st edn, Almedina 2023), 414.

[190] Quinn (169) 7.

[191] *Ibid.*

outlined in Art. 89(1) GDPR in Law on Higher Education[192] which is applicable to research institutes and universities, conducting research as an activity for public interest and limits the qualification to institutions which are not purely commercial entities.[193] Therefore, private entities conducting research and trying to use Art. 89(1) GDPR's legal basis for the data processing and a waiver for other obligations cannot benefit from Art. 89 GDPR's exemptions[194].

Often, these organisations rely on Art. 6(1)(f) GDPR legal basis for process PD when the processing is necessary for purposes of the legitimate interests pursued by controller or by a third party, except where such interests are overridden by the DS's interests or fundamental rights and freedoms. Rec. 47 explains data controllers should consider reasonable expectations of DS based on the relationship with the controller. Besides, the EDPB clarifies processing for research purposes (including marketing research) could constitute a legitimate interest, provided the controller implemented sufficient safeguards[195].

PD can be re-used or "further processed" for a purpose other than that for which they were initially collected, providing purpose limitation requirements – Arts. 5(1), 6 (4) and 89 (1) GDPR - are met. It has two components: purpose specification (PD should only be collected by specified, explicit and legitimate purposes); compatible use (implies further processing must not be incompatible with the purposes for which PD were first collected). Data controller to allow for the re-use of data, unless the DS consents to a new specific purpose, should to ascertain if processing for another purpose is compatible with the initial purpose, consider any link between the purposes for which the PD have been collected and the purposes of the intended further processing, the context in which the PD have been collected, in particular regarding the relationship between DS and the controller, whether special categories of PD are processed – Art. 9, or whether criminal convictions and offences are processed, pursuant to Art.

[192] Act of 20 July 2018, The Law on Higher Education and Science, Constitution for Science - Ministry of Science and Higher Education, Ministry of Science and Higher Education, 2018, 3.

[193] Els Kindt and others, 'Study on the Appropriate Safeguards under Art. 89(1) GDPR for the Processing of Personal Data for Scientific Research Final Report', Milieu Consulting SRL, 2019, 56.

[194] *Ibid.*

[195] A29WP, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679.

10, the possible consequences of the further processing for DS and the existence of appropriate safeguards (including encryption or pseudonymisation).

Art. 4(2) GDPR defines processing as any (or set of) operation(s) performed on PD and includes adaption or alteration, whichever de-identification technique is to be considered as further data processing, requiring a legal basis to the processing. Therefore, anonymisation constitutes an act of data processing and must be legitimated under GDPR. Data collection must be legitimated based on Arts. 6(1) and 9(2) GDPR and the processing to render data anonymous by the data controller, the PSB or the competent body acting as a substitute[196] must be justifiable and compatible with the initial purpose.

In 2019 the Danish Data Protection Authority (DPA) found a taxi company failed anonymisation and violated purpose limitation principles since it was not able to prove phone numbers were collected for explicitly stated and legitimate purposes (Art. 5 (1)(b), (2) GDPR). The company claimed it anonymised the customers' data after the lawful two-years retention policy. Yet it deleted the customers' names only and retained phone numbers for five years for business purposes. The DPA concluded phone numbers were not de-identified since Taxa 4x35 could use it to identify/link the customers indirectly[197].

It is important to stress data protection law presumes some purposes are not considered incompatible with the data processing's initial purposes, namely further processing for the four purposes: archiving purposes in the public interest, scientific or historical research purposes and statistical purposes[198]. Art. 89 (1) GDPR provides when processing operations take place for the purposes abovementioned, appropriate safeguards for the rights and freedoms of the DS must be implemented. Technical and organizational measures, as pseudonymisation and anonymisation techniques and ethical policies and reviews' proceedings of data processing[199] must be implemented to ensure respect for data minimisation – Art. 5(1)(c) GDPR. One may argue no additional legal basis is required for the processing of PD for scientific research

[196] Baloup and others (126) 21.

[197] European Data Protection Board, 'The Danish Data Protection Agency proposes a DKK 1,2 million fine for Danish taxi company', 2019. Datatilsynet, 'Tilsyn med Taxa 4x35's behandling af personoplysninger', 2019.

[198] Art. 5 (1) GDPR.

[199] Quinn (169) 15.

purposes granted there was a valid primary legal basis for the data collection and Art. 89(1) GDPR structures are in place. After all, this legal provision tends to benefit the accustomed larger companies and controllers over SMEs and start-ups, as the formers have access to a greater pool of data capable to conduct ampler scientific research based on that and are able to warrant the inherent economic and organizational costs to comply.

Chapter V. DGA and GDPR: Compatibility or Conflict on the re-use of personal data?

5.1. Principle of Purpose Limitation vs. Anonymisation

As analysed before, discerning whether specific data is of personal nature or not under the GDPR is crucial to determine its scope of application[200]. At the same time, the application of proper de-identification techniques, consisting of a broad spectrum of tools for GDPR compliance, plays a crucial role in facilitating data secondary uses while helping to protect individuals' privacy[201].

De-identification means "a process of removing the association between a set of notifying attributes and the DS"[202]. Techniques may vary between methods can reduce privacy risk to a moderate level as pseudonymization or stronger methods which leave out (most or) all privacy risks as anonymisation. On the contrary, there is an increasing risk of re-identification, consisting of the risk of de-identified data to be (re-)identified, known as DS's identity disclosure or attribute disclosure[203].

Rec. 26 GDPR introduces the concept of pseudonymous data, distinctive from anonymous data, as an intermediate level of de-identification. This Rec. presents a risk-based legal test to

[200] Michèle Finck and Frank Pallas (91) 11.

[201] Heung Youl Youm, 'An overview of De-identification Techniques and Their Standardization Directions', IEICE Transactions on Information and Systems E103.D(7), 1448.

[202] *Ibid* 1449.

[203] Mahsa Shabani and Luca Marelli, 'Re-Identifiability of Genomic Data and the GDPR: Assessing the re-identifiability of genomic data in light of the EU GDPR', 20 EMBO reports, 2019, 1.

distinguish between personal and non-personal data[204], stating that "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". Hence, data that cannot be attributed to a specific DS without the use of additional information must be considered pseudonymous data and falls under the data protection law's regime. Thus, as pseudonymous data qualifies as personal, GDPR's principles such purpose limitation and data minimisation apply to them. The additional information must be kept separately and secured through organizational or technical measures, namely through encryption, to prevent re-identification of the DS[205]. The Rec. concludes data protection principles do not apply to anonymous information, namely "information which does not relate to an identified or identifiable natural person or to PD rendered anonymous in such a manner the DS is not or no longer identifiable". This regime's exclusion shows EU legislator considers anonymous data does not infringe any "legally enforceable interest"[206], hence it is out of GDPR's scope.

In 2022 the Greek DPA fined telecommunications companies due to PD breach and illegal data processing under GDPR[207]. It was claimed PD was anonymised for analytics purposes, however, the DPA considered datasets in question were pseudonymized rather than anonymised. Then, it considered statistics' extraction from the pseudonymized databases was a further purpose of processing, which may be compatible with the original purposes but subject to Art. 89 GDPR's conditions. In this case, DS were not adequately informed about the relevant processing.

Regarding the re-use of PD held by PSB, in DGA the legislator leaves no doubt[208] and states both in Recs. 4 and Art. 1(3) the EU and national legislation on data protection must be followed and prevail in the event of conflict with the DGA. Besides, it conveys the DGA does not create

[204] Michèle Finck and Frank Pallas (91) 14.

[205] Rec. 29 GDPR

[206] Barreto Menezes Cordeiro (90) 2.

[207] Hellenic DPA, 'Fines imposed due to personal data breach and illegal data processing by telecommunications companies', 2022.

[208] Tiago Sérgio Cabral, 'Breve incursão sobre a figura dos *Data Intermediation Services* e a sua compatibilidade com o RGPD', O Contencioso da União Europeia e a cobrança transfronteiriça de créditos - Volume III, Coleção UNIO E-book, 2022, 130.

an additional legal basis for PD's processing nor recast any obligations and rights outlined in GDPR or e-Privacy Directive.

Data controllers can only use PD for a new purpose if either the secondary data use is compatible with the original purpose for which data has been collected, acquires new consent or has a clear obligation or function set out in legislation, according to purpose limitation. GDPR considers processing for research purposes as a compatible lawful operation (Rec 50 GDPR) and enables data controllers to re-use PD previously collected under a lawful legal basis for secondary scientific research purposes. Otherwise for all other purposes, to ascertain whether a new purpose is compatible with the original one, the requirements embodied in Art. 6(4) GDPR must be taken into account any link between the original purpose for which the PD have been collected and the secondary purpose; the context in which PD have been collected; the nature of PD, namely whether it qualifies as sensitive data; any possible consequence resulting from further data processing for the DS and whether appropriate safeguards such as encryption or pseudonymisation are in place.

In that regard, taking a close look at Art. 5 DGA, EU legislator seems to incentivize PSB to seek and ensure anonymisation of data at first instance. Then, any data processing therefrom fall outside GDPR's scope. As data will not fall within GDPR's scope of application, purpose limitation and data minimisation are not applicable, hence data becomes easier to (re-)use[209]. It must be noted the reference to "pseudonymisation" in this Art. has been removed which makes one to assume *de novo* the incentive for anonymisation de-identification technique as a primary option and fadeout of GDPR provisions, namely in which regards data secondary purposes of further processing or "re-use".

Only in the event anonymised data does not fulfil potential re-user's needs, on-premises or remote re-use of the data within a SPE may be allowed by the PSB. Here, principle of purpose limitation needs to be respected, therefore new purposes other than archiving purposes in the public interest, scientific or historical research purposes and statistical purposes need to be compatible with the initial purposes of the processing of data.

[209] Weitzenboeck and others (83) 2.

5.2. A look towards mixed datasets

Progressively datasets incorporate different types of data. Currently it is common a dataset consists of "a mix of personal and non-personal data"[210], the so-called mixed datasets. As outlined earlier in Chapter 2.3., NNPD's provisions on mixed datasets results in legal uncertainty regarding which rules (NNPD or GDPR) apply in respect of each part of the dataset as it becomes more and more demanding to distinguish between the different categories of data in a dataset due to the dynamic, broad notion and context-dependent nature of PD[211]. In parallel, EC's guidance on mixed datasets and "inextricably linked" parts of datasets leads back to the solely application of GDPR to these datasets.

Trying to avoid (additional) complex interplays, the final text[212] mentions DGA should be without prejudice to GDPR and the e-Privacy Directive, even in circumstances "where personal and non-personal data in a data set are inextricably linked". However, no further explanation regarding datasets in which there is inextricably linked data is given. As such, joining GDPR and NPDR, DGA does not define the concept "inextricably linked", nor does it add progress in relation to which regime is applicable to mixed datasets. Therefore, mixed datasets held by PSB potentially will end up subject to GDPR rules[213] and PSB put under Art. 5 DGA's obligations to guarantee the conditions for re-use, namely the anonymisation of the data and privacy-enhancing environments under Art. 32 GDPR.

[210] Graef, Gellert and Husovec (83) 6.

[211] Paul M Schwartz and Daniel J. Solove, 'Reconciling Personal Information in the United States and EU', 102 California Law Review 877 (2014), UC Berkeley Public Law RP No. 2271442, GWU Legal Studies RP No. 2013-77, GWU Law School Public Law RP No. 2013-77, 2013, 886.

[212] It should be noted no reference to "inextricably linked" datasets were outlined in drafts. This shows a clear intention of the legislator to acknowledge, albeit ambiguously, those possibilities in DGA's final text.

[213] COM(2019) 250 final (54) 9.

5.3. Assessing the Risk of Re-identification

EU lawmakers have relied (and continue to rely) upon anonymisation to surpass and balance complex questions[214]. The DGA is not an exemption and seem to resort to anonymisation to strike a balance between the added value around re-use of (non-)personal data and the protection of individual' rights under GDPR, albeit it recognises the "risk of re-identification of non-personal, anonymized data".

Accordingly, safe and robust de-identification-anonymisation techniques as anonymisation, differential privacy, generalization, suppression and randomization[215] applied before the "semi-public"[216] release of publicly held PD may assist PSB to guarantee the anonymisation of data and individuals are no longer identifiable.

However, it has to be said whereas data may be treated as anonymous, as state-of-the-art technologies and the growing amount of online data about individuals can be collected[217], "anonymisation is increasingly difficult to achieve"[218]. As such, some authors argue there is a growing and everlastingly underlying residual risk of re-identification[219].

As outlined earlier, GDPR entails a broad scope of the concept "personal data" in a crystal-clear move to achieve "flexibility, allowing it to be applied to various situations and developments affecting fundamental rights, including those not foreseeable"[220]. Covering all information relating to an identified or identifiable person, either directly or indirectly, the essential criteria to determine whether data is personal, is the test embodied in Rec 26 GDPR, based on the risk of identification. To determine whether a natural person is identifiable, "*all the means reasonably likely to be used*, namely singling out, either by the data controller or by other parties to identify the natural person directly or indirectly" (emphasis added) shall be

[214] Ohm (187) 1738.

[215] Rec. 7 DGA.

[216] Youl Youm (201) 1451.

[217] Weitzenboeck and others (83) 2.

[218] A29WP, Opinion 03/2013 on Purpose Limitation, WP203, 2013, 31.

[219] Finck and Pallas (91) 35.

[220] COM(2010) 609 final, 5.

considered. To determine whether means of identification are reasonable likely to be used to identify the natural person, objective factors as costs and time required for identification, state of the art technologies and technological developments must be considered. All in all, this risk-based test[221] constitutes "one of reasonable likelihood of identification"[222] by the data controller or by another person, considering only objective factors as state-of-the-art technologies at processing's time and technological developments.

The circumstance full and absolute anonymisation is no longer possible has been pointed out by A29WP[223] and researchers such as Finck and Pallas[224], Ohm[225], Purtova[226] and Barreto Menezes Cordeiro[227]. According to Finck and Pallas' interpretation, where there is a reasonable risk of identification of the DS, data ought to qualify as PD; otherwise data ought to be treated as non-personal, nonetheless identification cannot be exempted with total certainty[228]. This interpretation is in line with Menezes Barreto Cordeiro's argument adoption of a reasonable likelihood criteria by the EU legislator "reflects the factual impossibility to guarantee the absolute anonymity of the data collected"[229]. Both points of view are in conformity with the EU data protection law's risk-based approach, adopted by the CJEU on the 2016 judgement on dynamic IP addresses - C-582/14 *Breyer*[230]. The case brought by Patrick Breyer against the Federal German Government was referred to the CJEU for preliminary ruling to determine whether the data in the hands of the Federal German Institutions operating the websites in question was data relating to an identifiable person. The Court, following Advocate General's opinion and based a middle ground between an objective and relative criterion, states "it is not required all the information enabling the identification of the DS must be in the hands of one person" to information be treated as PD. Additionally, the Court held information would not be considered PD in cases where the DS's identification was prohibited by law or when

[221] Weitzenboeck and others (83) 8.

[222] *Ibid* 8.

[223] A29WP (184) 3.

[224] *Ibid.*

[225] Ohm (188) 1742.

[226] Purtova (32) 48.

[227] Barreto Menezes Cordeiro (90) 17.

[228] Finck and Pallas (91) 14.

[229] Barreto Menezes Cordeiro (90) 17.

[230] Case C-582/14, ECLI: EU: C: 2016:779, Patrick Breyer v *Bundesrepublik* Deutschland, 2016.

identification was '*practically impossible* on account of the fact it requires a disproportionate effort in terms of time, cost and man-power, so *the risk of identification appears in reality to be insignificant* (emphasis added)." Following CJEU approach, the French DPA describes anonymisation as processing made up of a set of techniques which render identification of the DS "practically impossible"[231]. From another angle, one may state robust anonymisation may only lessen re-identification risk to a certain "acceptable threshold"[232] which depends on criteria as mitigation controls, impact on individual's privacy in the event of re-identification and any motive and capacity of an attacker behind a data breach. As such, in the event data can be considered to have a small risk of re-identification, it is to be considered fully anonymized and thus out of GDPR's [233].

To the contrary, A29WP, in the not formally binding Opinion 05/2014 on Anonymisation Techniques, defines anonymisation as "technique applied to personal data to achieve *irreversible de-identificatio*n (emphasis added)"[234] and considers it an instance of further processing compatible with the original purposes of data collection under the Art. 5(1)(b) GDPR. A29WP seems to embrace a stricter approach, when compared with the GDPR's rational, and include a "zero-risk test"[235]. Resultantly, the outcome of the application of anonymisation techniques such as randomization - noise addition, permutation or differential privacy - and generalization - aggregation, k-anonymity, l-diversity and t-closeness - applied to PD should be "as permanent as erasure"[236]. In order to determine whether de-identification has occurred, A29WP specifies three specific criteria which can be interpreted as "three different ways of deriving new attributes about individuals"[237] which ought to be taken into account: the possibility to isolate some or all records identify an individual within the dataset (singling out as stated in Rec 26); the possibility to link records regarding the same individual or group of individual within the same or two different databases (interlinking) and the possibility to infer

[231] Commission National de l'Informatique et des Libertés, 'Anonymisation de Données Personnelles', 2020.

[232] European Medicines Agency, External Guidance on the Implementation of the European Medicines Agency Policy on the Publication of Clinical Data for Medicinal Products for Human Use, 2017, 47.

[233] *Ibid.*

[234] A29WP (184) 7.

[235] Finck and Pallas (91) 15.

[236] A29WP (184) 6.

[237] Gergely Acs, Claude Castelluccia and Daniel Le Métayer, 'Testing the Robustness of Anonymization Techniques: Acceptable versus Unacceptable Inferences - Draft Version', 2016, 1.

an DS' information with significant probability (inference). A29WP considers pseudonymous data continues to permit DS to be singled out and inter-linkable, being less privacy-friendly than anonymous data, since it allows for identifiability (meaning there are chances of a data subject to be identified when additional information is added).

An highlight must be given to EU GC's ruling of April 26, 2023 Case T-557/20, ECLI:EU:T:2023:219 – *Single Resolution Board (SRB) v European Data Protection Supervisor (EDPS*) which provided clarity by concluding pseudonymized data transmitted to a data recipient will not be considered PD in the event data recipient does not have the means to re-identify the DS. The case in question involved the processing of personal comments produced by the shareholders and creditors of Banco Popular, then shared with a third-party consulting firm Deloitte by decision of SRB. The comments shared were filtered, categorized, and aggregated, if identical, and associated with an alphanumeric code (33-digit globally unique identifier randomly generated) at the time of reception. Deloitte did not have access to identifying data used for registration. While EDPS asserted SRB had breached Art. 15 GDPR obligation to inform DS PD would be shared with a third party, GC followed *Breyer'*s reasoning, adopting a risk-based approach and stressed to determine whether pseudonymized data transmitted to a data recipient constitutes PD, it is necessary to consider the latter's perspective. Yet, please note GC did not address concrete conditions for data to be considered anonymous and the ruling is currently under appeal.

It becomes clear there is a divergence as to the approach to be taken to determine whether anonymisation has occurred. Based on a risk-based approach in line with GDPR, which seems to be "gaining steam"[238], consideration ought to be taken to possible means to be used by the data controller or third party to identify the natural person based on reasonable likelihood of identification, considering the long-established evidence of the probability DS being re-identified from the original variables and datasets with "astonishing ease"[239]. In opposition,

[238] Andrew Burt and Sophie Stalla-Bourdillon, '*The definition of 'anonymization' is changing in the EU: Here's what that means'*, The International Privacy Professionals Association, 2023.

[239] The classic example of Governor William Weld's re-identification where Dr. Sweeney was able to re-identify the Governor's allegedly anonymised set of medical data by combining and linking two different databases. It demonstrates computer science's ability/ease to re-identify DS. Daniel Barth-Jones, '*The 'Re-Identification' of Governor William Weld's Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now'*, 2012.

according to A29WP's restrictive standpoint, to data controllers guarantee anonymisation complies with the GDPR, the erasure of original datasets would be "the only remaining solution"[240]. Regarding PSB in specific, option would be barely impossible in practice due to the common need to keep a record of the original to fulfil further legal obligations[241]. Besides, one can argue would result in a potential chilling effect on the re-use of data and subsequentially in a negative impact on EU innovation and economy[242]. This potential impossibility of the PSB's to comply with data protection law in practice exacerbates what Purtova describes as risk of the GDPR "becoming the law of everything"[243]. Furthermore, the EC seems to seek to find possible alternatives to a future data protection law's structure by financing a project whose objective refers "a law regulating everything is meaningless".[244]

The anonymisation process and implementation highly influence the likelihood of the re-identification risks. Given the GDPR's risk-based approach, one may state whereas before transmission de-identification techniques have been sufficiently strong so that identification of the DS is no longer reasonably likely, PSB may ensure data has been anonymised according to 5 (3)(a)(i) DGA and thus falls outside the GDPR's scope of application. In addition, PSB as data controllers keep a monitoring obligation and must adopt technical and organizational measures to keep up with the dynamic nature of PD and technological advancements, and thus avoid re-identification of any DS. The EU legislator through the DGA's framework on the re-use of data supports a "semi-public disclosure"[245] of data limited to an authorized group of re-users (individuals or organizations), hence stricter than a public release under the NODD, and subject to a formal request and approval process. DGA incentives PSB and re-users enter into non-disclosure agreements in which re-users may be subject to data confidentiality obligations to prevent disclosure of any information threaten rights and interests of third parties, being prohibited from re-identifying the data subjects to whom data relates and obliged to take technical and organizational measures, ensuring GDPR's compliance. DGA outlines a

[240] Weitzenboeck and others (83) 3.

[241] According to Portuguese Law, clinics and medical offices must keep a record of patients' clinical processes. Ordinance no. 136-B/2014, of 3th July.

[242] Barreto Menezes Cordeiro (90) 25.

[243] Purtova (32) 41.

[244] EC, 'Understanding information for legal protection of people against information-induced harms', 2016.

[245] Youl Youm (201) 1451.

notification obligation of the re-users to notify PSB[246] in the event of any data breach (any unauthorized re-use of the non-personal data, *whose outcome is the re-identification of the data subjects).* This obligation does not exclude GDPR's notification obligation to the DS and DPAs[247]. One may see a similar protective approach in the event access and re-use of data is permitted within a SPE, where PSB may impose conditions for re-use so that integrity of the technical structure of the SPE is secured. Besides, PSB reserve the right to verify the process, any means and outcome results of the data processing so that integrity of the PD is secured, what is more, the right to block the use of any result hindering the protection of PD. One may argue DGA, acknowledging an inherent risk of re-identification, aims to guarantee a realistic and "acceptable threshold"[248] of risk of re-identification, while it gives great consideration to the protection of PD, through the addition of obligations on the re-users and the assignment of rights to the PSB to guarantee effective safeguards.

5.4. DGA's offbeat approach to consent

In the current EU legal framework, consent given by the DS occupies a central placement to empower DS to have control over any collection and processing of their PD [249]. Consent (statement or clear affirmative action *freely given, specific, informed and unambiguous*[250]) may be considered "the most well-known lawful basis" [251] to process PD under the GDPR.

Yet it is crucial to mention consent constitutes only one of the six hierarchically equal legal bases[252] to process PD under the GDPR (Arts. 6 and 9). Consent may not be "the most appropriate in many cases"[253]. Considering scientific research purposes, it is often not possible to fully identify a specific purpose of PD processing at the time of data collection (Rec. 33 GDPR). Therefore, although GDPR embodies a broader notion of DS's consent to certain areas

[246] Art. 5 (5) DGA.

[247] Rec. 15 DGA.

[248] European Medicines Agency (232) 47.

[249] Eleni Kosta, '*Consent in European Data Protection Law, Nijhoff Studies in EU Law'*, (Martinus Nijhoff Publishers, Volume 3, 2013), 1.

[250] Art. 4 (11) GDPR.

[251] Data Protection Commission, 'Guidance Note: Legal Basis for Processing Personal Data', 2019, 2.

[252] *Ibid.*

[253] *Ibid.*

of scientific research, the request for DS's consent may not be feasible, as explored in Chapter 4.2., undermining the research's objectives. Simultaneously, the use of consent as a legitimate basis for the processing of PD has faced some criticism concerning its legitimacy. As such, it is questionable whether nowadays it is possible for individuals to make a truly informed decision about each data processing of PD and foreseen possible consequences and outcomes.[254]

The DGA seems to take these circumstances into account and in Art. 5(6) sets out only when PSB cannot ensure data has been anonymised nor are able to provide access to data within a SPE nor GDPR legal basis, other than consent, for transmitting data can be found, it is obligation of the PSB to make best efforts to help potential re-users in seeking consent of the DS whenever neither anonymisation or modification techniques fulfil re-user's needs. Notwithstanding, DGA states PSB's assistance must be *"feasible without a disproportionate burden on the PSB".* Contrarily to history in data protection law, DGA suggests by default DS will not be asked to provide consent[255], showing an approach to consent as the "last resource" legal basis[256] or last option while anonymisation of the data is treated as the favoured one.

[254] Purtova (73) 11.

[255] Access Now (172) 5.

[256] *Ibid 5.*

Chapter VI. Conclusion

This paper aimed to explore the recent approved EU transversal data governance's framework, with a particular focus on its conditions for the re-use of PD held by public authorities' provisions and analyse its compatibility with the established European data protection law.

The EU through the DGA works towards the goal to ensure clear and harmonised conditions for access to and re-use of PD, as of underestimated and underused, and achieve the desired EU common data space. The analysis of the DGA's regime shows provisions of Chapter II aim to complement NODD's scope of application, yet they may create some overlap with national open access regimes. Moreover, albeit the DGA does not make progress on the matter of mixed datasets, it states unambiguously the prevalence of GDPR's regime in case of any conflict of the DGA with the GDPR. Nevertheless, it further shows the EU legislator continues to rely on anonymisation as a sidestep to exclude data from the GDPR's scope of application and balance complex questions. In the present case, the DGA relies on anonymisation to strike a balance between the added value around re-use of anonymised data by re-users and the protection of individual's rights under the GDPR. Additionally, the EU legislators on purpose recognises a possibility of re-identification and takes that into account when drafting the conditions for the re-use of data to be implemented by MS and further established by the PSB in specific. Whereas the legislator treats robust anonymisation as the first option to be taken and consent as the last resource legal basis under the GDPR, we argue great consideration is given to the interests and fundamental rights of the individuals represented in the data, and thus can assume compatibility of regimes.

The DGA's parallel and privileged re-use regime for scientific purposes and by SMEs and startups entails the EU in the right direction to facilitate the re-use of data for European research and innovation by both public and private entities. Notwithstanding the great effort deployed to make more data available, it is doubtful whether this framework will clearly have an impact on data availability and re-use, since no access right is established and PSB will not be legally bound to provide access and re-use of data. Besides, we suggest a diverse implementation of different condition's criteria in the 27 MS may lead to complex and fragmented processes, what hinders initial DGA's goal to address fragmentation of the internal market and data economy.

Bibliography

Abraham R, Schneider J and vom Brocke J, 'Data Governance: A Conceptual Framework, Structured Review, and Research Agenda' (2019) 49 International Journal of Information Management 424 < https://www.sciencedirect.com/science/article/abs/pii/S0268401219300787 >

Access Now, Position on the DGA, (2021) < https://www.accessnow.org/wp-content/uploads/2021/03/Access-Nows-position-on-the-Data-Governance-Act.pdf >

Acs G, Castelluccia C and Daniel Le Métayer, 'Testing the Robustness of Anonymization Techniques: Acceptable versus Unacceptable Inferences', [2016] < https://hal.inria.fr/hal-01399858 >

American Chamber of Commerce to the EU, 'Our Position DGA', (2021) < https://www.amchameu.eu/system/files/position_papers/data_governance_act_final.pdf >

Baloup J and others, 'White Paper on the Data Governance Act', CiTiP Working Paper, KU Leuven Centre for IT & IP Law, (2021) < https://www.researchgate.net/publication/352690055_White_Paper_on_the_Data_Governance_Act >

Barreto Menezes Cordeiro A, *'Dados Pessoais: Conceito, Extensão e Limites'*, Centro de Investigação de Direito Privado, (2018) < https://alvarovelho.net/index.php/recursos/documentos?task=download.send&id=519&catid=0&m=0>

Barth-Jones D, 'The 'Re-Identification' of Governor William Weld's Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now', (2012) < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2076397 >

Black J, ‘Learning from regulatory disasters’, LSE Law, Society and Economy Working Paper, 4/2014, (2014) < https://eprints.lse.ac.uk/60569/1/WPS2014-24_Black.pdf >

Boardman R and others, *‘European Data Protection Law and Practice’* (Eduardo Ustaran, 3rd edn, The International Association of Privacy Professionals, 2023)

Bughin J and others, ‘Innovation in Europe: Changing the Game to regain competitive edge’, McKinsey Global Institute Discussion Paper, 2013 < https://www.mckinsey.com/~/media/mckinsey/featured%20insights/innovation/reviving%20innovation%20in%20europe/mgi-innovation-in-europe-discussion-paper-oct2019-vf.pdf >

Burt T and Stalla-Bourdillon S, ‘The definition of 'anonymization' is changing in the EU: Here’s what that means’, The International Privacy Professionals Association, 2023, < https://iapp.org/news/a/the-definition-of-anonymization-is-changing-in-the-eu-heres-what-that-means/ >

Cambridge International A Level, ‘Topic 1.1 Data, Information and Knowledge’ < https://www.cambridgeinternational.org/Images/285017-data-information-and-knowledge.pdf >

Canto Moniz G, ‘*Manual de Introdução à Proteção de Dados Pessoais’*, (Almedina, 1st edn, 2023)

Cerrillo-i-Martínez A, ‘The Reuse of Public Sector Information in Europe and Its Impact on Transparency’, 18 European Law Journal 770, (2012) < https://onlinelibrary.wiley.com/doi/abs/10.1111/eulj.12003 >

Chen C and others, ‘Privacy Regulation and Firm Performance: Estimating the GDPR Effect Globally’, Working Paper No. 2022-1, The Oxford Martin Working Paper Series on Technological and Economic Change, (2022) < https://www.oxfordmartin.ox.ac.uk/downloads/Privacy-Regulation-and-Firm-Performance-Giorgio-WP-Upload-2022-1.pdf >

Drexl J and others, 'Data Ownership and Access to Data - Position Statement of the Max Planck Institute for Innovation and Competition of 16 August 2016 on the Current European Debate' [2016] SSRN Electronic Journal < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2833165 >

Edwards L, 'Data Protection: Enter the GDPR', Law, Policy and the Internet, (2018), < https://ssrn.com/abstract=3182454 >

Fernández Pinto M, 'Open Science for Private Interests? How the Logic of Open Science Contributes to the Commercialization of Research' , 5 Frontiers in Research Metrics and Analytics, (2020) < https://www.frontiersin.org/articles/10.3389/frma.2020.588331/full#B13 >

Finck M and Pallas F, 'They Who Must Not Be Identified—Distinguishing Personal from Non-Personal Data under the GDPR', 10 International Data Privacy Law 11, (2020) < https://academic.oup.com/idpl/article/10/1/11/5802594?login=false >

Graef I, Gellert R and Husovec M, 'Towards a Holistic Regulatory Approach for the European Data Economy: Why the Illusive Notion Of Non-Personal Data is Counterproductive to Data Innovation' (2018) < https://www.researchgate.net/publication/328060880_Towards_a_Holistic_Regulatory_Approach_for_the_European_Data_Economy_Why_the_Illusive_Notion_of_Non-Personal_Data_is_Counterproductive_to_Data_Innovation >

Graef I and Gellert R, 'The European Commission's Proposed Data Governance Act: Some Initial Reflections on the Increasingly Complex EU Regulatory Puzzle of Stimulating Data Sharing', TILEC Discussion Paper No. DP2021-006, (2021) < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3814721 >

Grafenstein, Maximilian, 'Reconciling Conflicting Interests in Data through Data Governance. An Analytical Framework (and a Brief Discussion of the Data Governance Act Draft, the Data Act Draft, the AI Regulation Draft, as well as the GDPR)' (May 9, 2022), HIIG Discussion Paper Series No. 2022-02, (2022) < https://ssrn.com/abstract=4104502 >

Harrison TM, Pardo TA and Cook M, 'Creating Open Government Ecosystems: A Research and Development Agenda', 4 Future Internet, (2012) < https://www.mdpi.com/1999-5903/4/4/900/html >

Ienca M, Vayena E and Blasimme A, 'Big Data and Dementia: Charting the Route Ahead for Research, Ethics, and Policy', 5 Frontiers in Medicine, (2018) < https://www.frontiersin.org/articles/10.3389/fmed.2018.00013/full >

International Organization for Standardization, 'ISO/IEC 2382-1' (1993) < https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-1:ed-3:v1:en >

International Science Council, 'Science in the Private Sector - International Science Council', (2022) < https://council.science/actionplan/science-private-sector/ >

Lundqvist B, 'Big Data, Open Data, Privacy Regulations, Intellectual Property and Competition Law in an Internet of Things World', (2016) Faculty of Law, University of Stockholm Research Paper No. 1, (2016), < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2891484 >

Jain A, 'The 5 V's of Big Data - Watson Health Perspectives' (Watson Health Perspectives, (2016) <https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data >

Jetzek T, Avital M and Bjorn-Andersen N, 'Data-Driven Innovation through Open Government Data' , (2014) < https://www.researchgate.net/publication/260929913_Data-Driven_Innovation_through_Open_Government_Data >

Kindt E and others, 'Study on the appropriate safeguards under Article 89(1)GDPR for the processing of personal data for scientific research: Final Report', EDPB, (2021) < https://edpb.europa.eu/system/files/2022-01/legalstudy_on_the_appropriate_safeguards_89.1.pdf >

Kosta E, '*Consent in European Data Protection Law'*, Nijhoff Studies in EU Law, (Martinus Nijhoff Publishers, Volume 3, 2013)

Lynskey O, '*The Foundations of EU Data Protection Law'*, (Oxford University Press, 2015)

Lubarsky B, 'Re-identification of anonymized data', 1 Geo. L. Tech. Rev. 202, (2017) < https://perma.cc/86RR-JUFT >

Magalhaes G and Roseira C, 'Open Government Data and the Private Sector: An Empirical View on Business Models and Value Creation', 37 Government Information Quarterly 101248, (2020) <https://www.sciencedirect.com/science/article/abs/pii/S0740624X17302629 >

Martin N and others, 'How Data Protection Regulation Affects Startup Innovation' Information Systems Frontiers (2019), < https://link.springer.com/content/pdf/10.1007/s10796-019-09974-2.pdf>

Mourby M and others, 'Are "Pseudonymised" Data Always Personal Data? Implications of the GDPR for Administrative Data Research in the UK', 34 Computer Law & Security Review, (2018) < https://reader.elsevier.com/reader/sd/pii/S0267364918300153?token=D5528605176C6C9EAEABF5E7051EDB003B13FF38882398842DB7AB6708B194F7395B45402CF0CBE552DD9966A620897B&originRegion=eu-west-1&originCreation=20220827144220 >

OECD, 'Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by "Big Data"', OECD Digital Economy Papers, No. 222, OECD

Publishing, (2013) < http://dx.doi.org/10.1787/5k47zw3fcp43-en>

OECD Library, 'DynEmp: Measuring Job Creation by Start-Ups and Young Firms - OECD', (2020) < https://www.oecd.org/sti/dynemp.html>

OECD Library, 'The Path to Becoming a Data-Driven Public Sector', (2019) < https://www.oecd.org/gov/the-path-to-becoming-a-data-driven-public-sector-059814a7-en.htm>

OECD, 'OECD Open, Useful and Reusable data (OURdata) Index: 2019', (2020) < https://www.oecd.org/governance/digital-government/ourdata-index-policy-paper-2020.pdf>

OECD, 'The OECD Privacy Framework', (2013) < https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf>

Ohm P, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization', UCLA Law Review, Vol. 57, p. 1701, U of Colorado Law Legal Studies Research Paper No. 9-12, (2010) < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006 >

Paes Marques F, *'Comentário ao Regulamento Geral de proteção de Dados e à Lei nº. 58/2019'*, (A. Barreto Menezes Cordeiro Almedina 2022)

Purtova N, 'Do Property Rights in Personal Data Make Sense after the Big Data Turn?: Individual Control and Transparency', 10(2) Journal of Law and Economic Regulation November 2017, Tilburg Law School Research Paper No. 2017/21, (2017) < https://ssrn.com/abstract=3070228 >

Purtova N, 'The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law', 2018 Law, Innovation and Technology 10(1), (2018) < https://ssrn.com/abstract=3036355 >

Rahmatian A, 'Debts, Money, Intellectual Property, Data and the Concept of Dematerialised Property', [2020] < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3684502 >

Quinn P, 'Research under the GDPR – a Level Playing Field for Public and Private Sector Research?', 17 Life Sciences, Society and Policy, (2021) <https://lsspjournal.biomedcentral.com/articles/10.1186/s40504-021-00111-z>

Reinsel D, Gantz J and Rydning J, 'The Digitization of the World from Edge to Core', (2018) < https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

Sérgio Cabral T, 'Breve incursão sobre a figura dos Data Intermediation Services e a sua compatibilidade com o RGPD', O Contencioso da União Europeia e a cobrança transfronteiriça de créditos - Volume III, Coleção UNIO E-book, (2022) <https://repositorium.sdum.uminho.pt/bitstream/1822/77932/3/O%20Contencioso%20da%20Uni%c3%a3o%20Europeia%20e%20a%20cobran%c3%a7a%20transfronteiri%c3%a7a%20de%20cr%c3%a9ditos_Vol%20III.pdf >

Sousa e Silva N, Direito E Robótica - Uma Primeira Aproximação (Robots and the Law - a First Take), Revista da Ordem dos Advogados, (2017) < https://ssrn.com/abstract=2990713 >

Sousa e Silva N, 'Quando O Segredo É a 'Alma Do Negócio''– Definição De Um Conceito (When is There a Trade Secret – About the Concept), (2013) < https://ssrn.com/abstract=2378066 >

Sousa e Silva N and Cunha Pinto B, '*Internet das Coisas(IoT): alguns desafios jurídicos'* (2022), Estudos de Direito do Consumo, AAFDL

Steinz T, 'The Evolution of European Data Law', Paul Craig and Gráinne de Burca, The Evolution of EU Law (3rd edn, OUP, 2021), 902-936, SSRN Electronic Journal < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3762971 >

Rat für Sozial- und Wirtschaftsdaten (German Data Forum), 'The German Data Forum (RatSWD) and Research Data Infrastructure: Status Quo and Quality Management', 2018 < https://www.konsortswd.de/wp-content/uploads/RatSWD_Output1.6_QualityMgmt.pdf >

Reinsel D, Gantz J and Rydning J, 'The Digitization of the World from Edge to Core', International Data Corporation, 2018, < https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf >

Rodrigues de Oliveira R, '*Direito da Proteção de Dados Perspetivas Públicas e Privadas'* (Domingos Soares Farinho, Francisco Paes Marques and Tiago Fidalgo de Freitas, 1st edn, Almedina, 2023)

Thouvenin F, 'Informational Self-Determination: A Convincing Rationale for Data Protection Law?', 12, (2021) < https://www.jipitec.eu/issues/jipitec-12-4-2021/5409/thouvenin_pdf.pdf >

Ubaldi B, 'Open Government Data' OECD Working Papers on Public Governance', 22, [2013] <https://www.oecd-ilibrary.org/governance/open-government-data_5k46bj4f03s7-en >

Urbano Calvão F., '*Direito da Proteção de Dados Pessoais: Relatório sobre o programa, os conteúdos e os métodos de ensino da disciplina'*, (Universidade Católica Portuguesa. Porto, 2018)

Varian H., 'Artificial Intelligence, Economics and Industrial Organisation', NBER Working Paper 24839, JEL N. L0, (2018) < https://www.nber.org/system/files/working_papers/w24839/w24839.pdf >

Verhulst S G, 'Unlock the Hidden Value of Your Data', Harvard Business Review, (2020), < https://hbr.org/2020/05/unlock-the-hidden-value-of-your-data>

World Bank, 'World Development Report 2021: Data for Better Lives' (2021) < https://www.worldbank.org/en/publication/wdr2021 >

Youl Youm H, 'An overview of De-identification Techniques and Their Standardization Directions', IEICE Transactions on Information and Systems E103.D(7), (2020) < https://www.jstage.jst.go.jp/article/transinf/E103.D/7/E103.D_2019ICI0002/_article >

Zech H, 'A Legal Framework for a Data Economy in the European Digital Single Market: Rights to Use Data' Journal of Intellectual Property Law & Practice, 2016, Vol. 11, 460-470, (2016) < https://ssrn.com/abstract=2873135 >

Zech H, 'Information as Property', JIPITEC 6 (3) 2015, (2015) < https://www.jipitec.eu/issues/jipitec-6-3-2015/4315/zech%206%20%283%29.pdf >

EU Legislation

Council Framework Decision 2005/222/JHA of 24 February 2005 on attacks against information systems [2005] OJ L 69

Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information [2013] OJ L 345

Directive (EU) 2019/770 of the European Parliament and of the Council of 17 November 2003 on certain aspects concerning contracts for the supply of digital content and digital services [2019] OJ L 136

Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information(recast) [2019] OJ L 172

Regulation (EU) 2016/679 of the European Parliament and of the Council 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC ("*General Data Protection Regulation*") [2016] OJ 119

Regulation (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union [2018] OJ L 303/59

Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and amending Regulation(EU) 2018/1724 (Data Governance Act) [2022] OJ L 15

Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of Data (Data Act) – COM(2022) 68 final - 2022/0047(COD)

Grey Literature

Article 29 Working Party, Opinion 03/2013 on Purpose Limitation, WP203, (2013) < https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf >

Article 29 Working Party, Opinion 05/2014 on Anonymisation, (2014) < https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf >

Article 29 Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, (2016) < https://ec.europa.eu/newsroom/article29/items/612053 >

Article 29 Working Party, Opinion 06/2013 on Open Data and Public Sector Information ("PSI") Reuse, (2013) < https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp207_en.pdf >

Bulgarian Government, 'Digital Bulgaria Program 2025', (2022) < https://www.mtitc.government.bg/en/category/85/national-program-digital-bulgaria-2025-and-road-map-its-implementation-are-adopted-cm-decision-no73005-12-2019 >

Carsaniga and G and others, 'Open Data Maturity Report 2022', Data.europa, eu, (2022) < https://data.europa.eu/sites/default/files/landscaping_insight_report_n8_2022.pdf >

Commission National de l'Informatique et des Libertés, 'Anonymisation de Données Personnelles', 2020 < https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles >

Corcho O and Simperl E, 'Data.europa.eu and the European Common Data Spaces: : A report on challenges and opportunities', Data.europe.eu, (2022) < https://data.europa.eu/sites/default/files/report/EN_data_europa_eu_and_the_European_common_data_spaces_0.pdf>

Data.europa.eu, 'Empowering People with Open Data | Data.europa.eu' (2019) < https://data.europa.eu/en/datastories/empowering-people-open-data >

Data.europa.eu, 'Open Data and Entrepreneurship | Data.europa.eu', (2018) < https://data.europa.eu/en/datastories/open-data-and-entrepreneurship >

Data Protection Commission, 'Guidance Note: Legal Basis for Processing Personal Data', (2019) < https://www.dataprotection.ie/sites/default/files/uploads/2020-04/Guidance%20on%20Legal%20Bases.pdf>

Datatilsynet, 'Tilsyn med Taxa 4x35's behandling af personoplysninger' (2019), < https://www.datatilsynet.dk/afgoerelser/afgoerelser/2019/mar/tilsyn-med-taxa-4x35s-behandling-af-personoplysninger >

European Commission, Communication From The Commission To The European Parliament, The Council, The Economic And Social Committee And The Committee Of The Regions: A comprehensive approach on personal data protection in the European Union, COM(2010) 609 final, (2010) < https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0609:FIN:EN:PDF>

European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of The Regions: A European Strategy for Data, COM(2020) 66 final, (2020) <https://ec.europa.eu/info/sites/default/files/communication-european-strategy-data-19feb2020_en.pdf>

European Commission, Communication from the Commission to the European Parliament and the Council: Data protection as a pillar of citizens' empowerment and the EU's approach to the digital transition - two years of application of the General Data Protection Regulation, COM(2020) 264 final, (2020) < https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0264>

European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of The Regions:

European Interoperability Framework – Implementation Strategy - COM(2017) 134 final, (2017) < https://eur-lex.europa.eu/resource.html?uri=cellar:2c2f2554-0faf-11e7-8a35-01aa75ed71a1.0017.02/DOC_1&format=PDF >

European Commission, Communication From The Commission To The European Parliament, The Council, The Economic And Social Committee And The Committee Of The Regions: Open Data: An Engine for innovation, growth and transparent governance – COM(2011) 882 final, (2011) < https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF>

European Commission, 'Synopsis report of the public consultation on the revision of the Directive on the reuse of public sector information' (2018), < http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51544 >

European Data Protection Board, 'European Data Protection Board, 'The Danish Data Protection Agency proposes a DKK 1,2 million fine for Danish taxi company' (2019), < https://edpb.europa.eu/news/national-news/2019/danish-data-protection-agency-proposes-dkk-12-million-fine-danish-taxi_en >

European Medicines Agency, 'External Guidance on the Implementation of the European Medicines Agency Policy on the Publication of Clinical Data for Medicinal Products for Human Use' (2017) < https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data_en-1.pdf>

European Commission, 'Commission Staff Working Document Impact Assessment Report Accompanying the document Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act)', SWD(2020) 295 final, (2020) < https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020SC0295 >

European Commission, 'Commission Staff Working Document on Common Data Spaces', SWD(2022) 45 final, (2022) < https://ec.europa.eu/newsroom/dae/redirection/document/83562 >

European Commission, 'Communication from the Commission to the European Parliament and the Council - Two Years of Application of the General Data Protection Regulation, (2022) < https://ec.europa.eu/info/law/law-topic/data-protection/communication-two-years-application-general-data-protection-regulation_en >

European Commission, 'Digital Europe Programme (DIGITAL) Call for Proposals Preparatory Actions for Data Spaces' (2021) < https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/digital/wp-call/2021/call-fiche_digital-2021-prepacts-ds-01_en.pdf >

European Commission, 'Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union, Communication', COM(2019) 250 final, (2019) < https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=COM:2019:250:FIN>

European Commission, 'Digital Europe Programme (DIGITAL) Call for Proposals Preparatory Actions for Data Spaces' (2021) < https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/digital/wp-call/2021/call-fiche_digital-2021-prepacts-ds-01_en.pdf >

Hellenic Data Protection Authority, 'Fines imposed due to personal data breach and illegal data processing by telecommunications companies', (2022) < https://www.dpa.gr/en/enimerwtiko/news/fines-imposed-due-personal-data-breach-and-illegal-data-processing >

Johan Hansen and others, 'Assessment of the EU Member States rules on health data in the light of GDPR: Specific Contract No SC 2019 70 02 in the context of the Single Framework Contract Chafea/2018/Health/03', DG Health and Food Safety, European Commission (2021) < https://health.ec.europa.eu/system/files/2021-02/ms_rules_health-data_en_0.pdf >

Van Eechoud M, 'Study on the Open Data Directive, Data Governance and Data Act and their possible impact on research', Commission, 2022 < https://op.europa.eu/en/publication-detail/-/publication/a313139b-1147-11ed-8fa0-01aa75ed71a1>

National Legislation

Act on the Secondary Use of Health and Social Data, Finland Ministry of Social Affairs and Health 552/2019, < https://stm.fi/documents/1271139/1365571/The+Act+on+the+Secondary+Use+of+Health+and+Social+Data/a2bca08c-d067-3e54-45d1-18096de0ed76/The+Act+on+the+Secondary+Use+of+Health+and+Social+Data.pdf?t=1559641328000>

Belgian Act 30.07.2018, <https://www.dataprotectionauthority.be/publications/act-of-30-july-2018.pdf >

Danish Act Amending the Law on the Reuse of Public Sector Information, of 10 May 2021, < https://www.retsinformation.dk/eli/accn/A20210176429 >

Finnish Data Protection Act, <https://www.finlex.fi/en/laki/kaannokset/2018/en20181050.pdf >

Code des Relations entre Le public et L'administration, < https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000033205514/>

Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, <https://www.legifrance.gouv.fr/loda/id/LEGIARTI000037822800/2019-06-01/#LEGIARTI000037822800 >

Ordinance no. 136-B/2014, of 3th July , < https://dre.pt/dre/detalhe/portaria/136-b-2014-25345134 >

Portuguese Law no. 58/2019, of 8th August, < https://files.diariodarepublica.pt/1s/2019/08/15100/0000300040.pdf?lang=EN >

Portuguese Law no. 68/2021, of 22nd August, < https://diariodarepublica.pt/dr/detalhe/lei/68-2021-170221042 >

Case Law

Case C-582/14, ECLI: EU: C: 2016:779, *Patrick Breyer* v *Bundesrepublik* Deutschland, [2016] < https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62014CJ0582>

Case T-557/20, ECLI:EU:T:2023:219, *Single Resolution Board (SRB) v European Data Protection Supervisor (EDPS)*, [2023] < https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62020TJ0557#t-ECR_62020TJ0557_EN_01-E0001>