Bálint Ferencz* – Bettina Büki**

Three Ways of Secure Data Reusability in Europe: German Research Data Centres, Finnish Findata and the French Secure Access Data Centre

Abstract

The importance of the data economy has been recognised by the European Commission (hereinafter: Commission); hence, since the release of the data strategy, a set of legislative initiatives has been launched. One of these is the Data Governance Act (hereinafter: DGA) which intends to persuade Member States to set up or strengthen their already existing data intermediaries. In order to understand the motivations of the Commission, this article presents its digital agenda alongside the measures implemented to enhance the reuse of data. The article then provides an assessment of three data sharing services which are highlighted in the DGA's impact assessment. These intermediaries are the German Research Data Centres *(Forschungsdatenzentrum)*, the Finnish Findata and the French Secure Access Data Centre *(Centre d'accès sécurisé aux données)*. The article introduces the main characteristics and *modus operandi* of these bodies and finally provides a comparison of them. The comparison identifies the focal points these bodies face, such as non-profit objectives, the legal background and accessibility and security issues. As a conclusion, it seems that the main structure and principles underlying their functioning of these bodies are rather similar.

Keywords: data centre, data governance, data sharing, data intermediaries, Findata, CASD, *Forschungsdatenzentrum*

* Bálint Ferencz is a PhD candidate researching the theoretical relationship between law and informatics and examining current developments in the field of data policy, Artificial Intelligence, blockchain and smart contracts (balintov@caesar.elte.hu, balintov.ferencz@gmail.com).

** Bettina Büki is an LL.M candidate focusing on digital and data economy, having experience in the European Commission Directorate General of Communication Networks, Content and Technology – Artificial Intelligence, Technologies and Systems for Digitising Industry (buki.bettina@gmail.com).

DOI: 10.54148/ELTELJ.2022.1.81

I Introduction

As one of the most common analogies holds (despite its inadequacy[1]), data are considered the new oil, that is needed for Europe in order to help strengthen its economy amidst cruel economic competition with China and the USA. One of the hardest enterprises of the current Commission is to create the Digital Single Market through several legal initiatives.

While the General Data Protection Regulation (hereinafter: GDPR)[2] already put the European data protection policy on the global centre stage, it is apparent that the Commission would like to take a step further and aims to arrange a structure where both economic goals and data protection measures could co-exist in order to raise the EU to become a global challenger to China and the USA. In the last two years, the 'legislative pentagon'[3] (i.e. Digital Services Act, Digital Market Act, Artificial Intelligence Act, Data Governance Act and the forthcoming Data Act) has shown the main aims of the Commission and key actions which are about to be taken.

The article briefly presents the above-mentioned legislative initiatives without providing any remarks. The main focus of the article is three institutions which are highlighted by the impact assessment of the Data Governance Act (hereinafter: DGA) as good examples of data intermediaries.[4] Although the DGA gives a special role to these types of bodies (see for instance its recital 22), these have not yet been examined deeply from the perspective of legal scholarship. As such, the article's primary aim is to provide an overview of the RDC (Germany), Findata (Finland) and Secure Access Data Centre (France). It will describe the key features, such as the legal status and tasks of these institutions, their modus operandi and other peculiarities which are worth mentioning. Furthermore, the three bodies will be compared and assessed based on common patterns, such as their primary aims, their legal background and the mode of data accessibility. Although it is clear that there are other similar hubs, one-stop-shops and resembling institutions in Europe, due to their exemplary status in the DGA's impact assessment, only these are considered here.

---

1 Lauren Scholz, 'Big Data Is Not Big Oil: The Role of Analogy in the Law of New Technologies' (2018) SSRN Electronic Journal <https://www.ssrn.com/abstract=3252543> accessed 12 April 2021, DOI: https://dx.doi.org/10.2139/ssrn.3252543.

2 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation) (2016) OJ L 119/1.

3 The expression is used by Floridi, although he holds the GDPR as one of the pentagon's elements. See: Floridi L, 'The European Legislation on AI: A Brief Analysis of Its Philosophical Approach' (2021) SSRN Electronic Journal <https://www.ssrn.com/abstract=3873273> accessed 15 December 2021, DOI: https://dx.doi.org 10.2139/ssrn.3873273.

4 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) COM(2020) 767 final 13.

As regards the methodology applied and sources used, it must be noted that, in addition to the relevant laws, the website of the given body served most often as the primary source. Since these bodies are rather new, the online sources could serve as the most up-to-date information.

II European Digital Agenda

In the European Union (hereinafter: EU), there is currently a high level of willingness to implement the EU data strategy that was unveiled on 19th February 2020[5] in order to create a single market for data, where data can flow easily across sectors and countries while respecting EU values.[6] One of the six main priorities of the European Commission (hereinafter: Commission) for 2019–2024 is to create a Europe that is fit for the digital age and to empower people with a new generation of technologies.[7] Since digitalisation has a huge impact on people's lives, the Digital Decade aims to strengthen Europe's digital sovereignty while setting standards and focusing on data, technology and infrastructure.[8] The Commission's target is to make the EU a role model for a society empowered by data, where data flows freely in order to help businesses, researchers, public administrations and people to make better decisions based on non-personal data available to all.[9] While it seemed earlier that the Commission based its strategies not only on economic considerations but also on common European values, this strategy has already been criticised, as it brings back the standard approach of the Commission by intending to tackle the digital challenges by economic means mainly.[10]

Since 2013, the Commission has taken several steps to facilitate the development of the data-agile economy, such as the Public Sector Information (hereinafter: PSI) Directive, the Regulation on the free flow of non-personal data, the Open Data Directive and the General Data Protection Regulation.[11] At the same time, supporting the data-centric

[5] 'Shaping Europe's Digital Future' <https://ec.europa.eu/commission/presscorner/detail/en/ip_20_273> accessed 15 December 2021.

[6] '23 November 2021 – EU Open Data Days – Publications Office of the EU' <https://op.europa.eu/en/web/euopendatadays/23-november-2021/#The-EU-data-strategy-towards-a-single-European-market> accessed 15 December 2021.

[7] 'The European Commission's Priorities' <https://ec.europa.eu/info/strategy/priorities-2019-2024_en> accessed 13 December 2021.

[8] 'A Europe Fit for the Digital Age' <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age_en> accessed 13 December 2021.

[9] 'European Data Strategy' <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en> accessed 13 December 2021.

[10] Paul Keller, Alek Tarkowski, 'Digital Public Space – A Missing Policy Frame for Shaping Europe's Digital Future' (2021) Open Future <https://openfuture.pubpub.org/pub/digital-public-space-policy-frame/release/2> accessed 21 November 2021.

[11] 'A Europe Fit for the Digital Age' (n 8).

economy serves a greater aim as 'one of the EU's main strategic questions in the 21st century is how control over its data asset may be taken back and how technological vulnerability leading to data loss may be decreased'.[12]

The Commission focuses on generating value through the reuse of public sector information that has a significant potential in new services, increases the transparency of governments or simply helps to address societal challenges. The 2003/98/EC Directive on the reuse of public information was created in order to stimulate the further development of a European market for services based on information flowing from the public sector, strengthen competition and enhance the use and application of PSI in business processes.[13] The Commission first adopted a proposal for a revision of the PSI Directive in 2011 and then in 2018. The new Directive (2019/1024) supersedes the previous rules. Under the Open Data Directive, minimum rules are established – with regard to the exceptions –, for the re-use of existing documents held by public sector bodies of the Member States in order to stimulate innovation.[14]

III Reuse of Public Data in Europe

The European Strategy for data wants to ensure Europe's competitiveness and data sovereignty based on the belief that, with the right policies and investments, Europe can seize the opportunities associated with a paradigm shift and become a leader in data. From 2018 to 2025, the global data volume will grow from 33 to 175 zettabytes, the value of the data economy in the EU27 will growth from EUR 301 billion to EUR 829 billion, and the ratio between centralised computing facilities and smart connected objects will reverse (from 80% to 20% and from 20% to 80%).[15] Making data available for companies, individuals and public stakeholders helps economic growth, competitiveness, job creation, sustainability improvement and societal progress.[16] The strategy intends to create fair and clear rules for access and use of data while data can flow within the EU and across sector for everybody's benefit and respect privacy and data protection and competition law.[17]

---

[12] Tóth András, 'A Tisztességes Adatkereskedelmet Biztosító Szabályozás Szükségességéről' (2021) 62 Állam- és Jogtudomány 100–121, 112, DOI: https://doi.org/10.51783/ajt.2021.3.05.

[13] European Commission, 'Open Government Data & the PSI Directive' (2014) <https://data.europa.eu/sites/default/files/training_1-1_open_government-and-the-psi_en.pdf> accessed 13 December 2021.

[14] 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 (2019) OJ L 172/56.

[15] European Commission Directorate General for Communication, *The European Data Strategy: Shaping Europe's Digital Future* (Publications Office 2020) <https://data.europa.eu/doi/10.2775/645928> accessed 15 December 2021, DOI: https://doi.org/10.2775/645928.

[16] 'A European Strategy for Data' <https://digital-strategy.ec.europa.eu/en/policies/strategy-data> accessed 15 December 2021.

[17] Ibid.

As part of the data strategy, on 25 November 2020, the Commission proposed a regulation on data governance in order to boost data sharing across sectors and Member States and overcome technical obstacles to the reuse of data. The DGA focuses mainly on public sector data subject to the rights of others, such as personal data (e.g. health data) – but without affecting the application of the GDPR –, data protected by intellectual property rights, trade secret or statistically confidential data.[18] It sets conditions for the re-use of protected public sector data while increasing trust in data intermediaries. The DGA aims also to support European data spaces in the fields of health, environment, energy, agriculture, mobility, finance, manufacturing, public administration and skills involving not just public players but private ones, too. Where a sector-specific Union legal act requires public sector bodies, providers of data sharing services or registered entities providing data altruism services to comply with specific requirements, the sector-specific Union legal act should also apply.[19] The DGA complements not only the Open Data Directive but (since it addresses data held by public sector bodies that are subject to rights of others and therefore fall outside the scope of the Open Data Directive, which focuses on public sector data as well) also the Data Act, which is about to be issued at the time of writing. The DGA does not aim to grant, amend or remove the substantive rights on access and use of data, because such measures are envisaged for the Data Act.[20] The European Council and the European Parliament have already reached a provisional agreement on the DGA on 21 November 2021 under the Slovenian Presidency; therefore, the aim of the French presidency is the promulgation of the act.[21] After the final approval, the provisions will apply 15 months afterwards.

Unfortunately, data sharing in the EU is hampered by an absence of appropriate structures and processes; therefore there is limited data-handling capacity and data reuse in the public sector.[22] Even though, thanks to the GDPR, there is an increased awareness of personal data protection, this is not always matched in the public sector. Public sector bodies find it difficult to reuse public data since there is huge lack of technical capacity and legal competence to process requests to reuse public data.

In spite of this, there are some Member States which have already established various institutions in order to facilitate secure conditions for the reuse of public data. In Germany, Research Data Centres facilitate access to sensitive data for researchers, in France the Secure

---

[18] A. van de Meulebroucke, L. Deschuyteneer, 'Data Governance Act Tackles Re-Use of Public Sector Data | Eubelius' (28 January 2022) <https://www.eubelius.com/en/news/data-governance-act-tackles-re-use-of-public-sector-data> accessed 22 February 2022

[19] 'European Data Governance Act' <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act> accessed 15 December 2021.

[20] European Commission, Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) COM(2020) 767 final.

[21] 'European Council and Parliament reach agreement on Data Governance Act' <https://eudatasharing.eu/news/european-council-and-parliament-reach-agreement-data-governance-act> accessed 13 February 2022.

[22] European Commission (n 4) 12.

Access Data Centre allows the secure processing of statistical micro-data, and the Finnish data permit authority Findata aims to provide researchers with a one-stop-shop service for receiving a permit to process data from a range of public registers for health and social protection.

IV Research Data Centres (Germany)

Research data are highly important for the scientific community and policy consulting, since these data help find answers to various research questions.[23] In the early 2000s, an intensive discussion went on in Germany on how to grant access to microdata from official statistics.[24] As an answer, the German Federal Ministry of Education and Research published the report 'Commission to Improve the Information Infrastructure between Research and Statistics (KVI)' after three professors, Richard Hauser, Gert G. Wagner, and Klaus F. Zimmermann, published a memorandum with the title 'Conditions for the success of empirical economic research and research-based policy advice in economic and social policy'.[25] The KVI aimed to improve the interrelation between researchers and statistics, and one of the recommendations of the report was to establish Research Data Centres (hereinafter: RDC) on the premises of the public data producers.[26]

In Germany, there are currently two RDCs of official statistics: the RDC of the Federal Statistical Office, which was established in autumn 2001 and funded by the Federal Ministry of Education and Research, and the RDC of the Statistical Offices of the Federal States, established in April 2002. Both RDCs have the same objective, the coordination of data and services for scientific use of official statistics microdata.[27] The RDCs' main aim is to support scientists who are working only on empirical scientific projects described within the data application process, such as master's or doctoral theses, and also research projects that are funded either by third parties with their own resources or on behalf of ministries. Furthermore, RDCs strive to improve microdata and to adapt to the changing needs of

[23] Daniela Hochfellner and others, 'Datenschutz Am Forschungsdatenzentrum' 4–5 <https://www.researchgate.net/profile/Daniela-Hochfellner/publication/254421148_Datenschutz_am_Forschungsdatenzentrum/links/02e7e535704e4a6468000000/Datenschutz-am-Forschungsdatenzentrum.pdf> accessed 14 December 2021.

[24] Sylvia Zühlke and others, 'The Research Data Centres of the Federal Statistical Office and the Statistical Offices of the Länder' (2004) 124 Schmollers Jahrbuch: Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften 567, 567.

[25] 'Development' *(KonsortSWD)* <https://www.konsortswd.de/en/ratswd/german-data-forum-ratswd/development/> accessed 14 December 2021.

[26] 'About RDC | Research Data Centre' <https://www.forschungsdatenzentrum.de/en/about-rdc> accessed 14 December 2021.

[27] Ralf K. Himmelreicher, Hans-Martin Gaudecker, Rembrandt D Scholz, 'Nutzungsmöglichkeiten von Daten Der Gesetzlichen Rentenversicherung Über Das Forschungsdatenzentrum Der Rentenversicherung (FDZ-RV)' (Max-Planck-Institut für demografische Forschung 2006) MPIDR WORKING PAPER WP 2006-018 <https://www.demogr.mpg.de/papers/working/wp-2006-018.pdf> accessed 15 December 2021, DOI: https://doi.org/10.4054/MPIDR-WP-2006-018.

scientists. The data infrastructure, which needs continuous improvement, is based mostly on functionally centralised data storage and the regionalised infrastructure.[28]

Centralised data storage is required, since the scientific analyses mostly relate to more than one federal state; therefore, the RDCs of the Statistical Offices of the Federation and the federal states make it possible for centralised data storage to allow official microdata from every federal state to be provided and used in all regional locations of both RDCs. This is a very important improvement, since the majority of official statistics in Germany are compiled in a decentralised manner by the Statistical Offices of each federal state.[29] The regionalised infrastructure enables RDCs to be close to science, since data users have various opportunities to visit safe centres (see them in details in part 2 of this chapter) that are distributed over all of Germany. In order to achieve these objectives, RDCs offer different ways of data access via which differently anonymised data products are provided.[30]

1 Legal Framework

The legal background of the use of RDCs are laid down in the Federal Statistics Law (*Bundesstatistikgesetz* – BstatG.) of Germany.[31] Section 16 (1) item 4 and 16 (6) are the most relevant parts – with regard to of absolutely, formally and factually anonymised data – of the law that regulates the use of data for scientific projects. In order to ensure the provisions in Section 16 (6), RDCs check all statistical results based on the data provided to ensure statistical confidentiality. With the amendment of the BstatG. in 2005, it is possible now to merge data from, for example, different environmental and economic statistical sources.[32] According to Section 16 (1), linking data from different data producers is possible only with the prior written consent of the data subject.[33] Moreover, Section 16 (6) guarantees the legal requirements for a broader access to individual data from official statistics: researchers from independent scientific institutions who are bound by secrecy are allowed to access factually and formally anonymised data.[34]

---

[28] Statistische Ämter Des Bundes Und Länder Forschungdatenzentren, 'General terms of use' (19 February 2020) <https://www.forschungsdatenzentrum.de/sites/default/files/rdc_general_terms_of_use.pdf> accessed 15 December 2021.

[29] 'Statistics' *(Federal Statistical Office)* <https://www.destatis.de/EN/About-Us/Our-Mission/bundesstatistik.html> accessed 14 December 2021.

[30] 'Über Die FDZ | Forschungsdatenzentrum' <https://www.forschungsdatenzentrum.de/de/ueber-die-fdz> accessed 14 December 2021.

[31] Bundesstatistikgesetz <https://www.gesetze-im-internet.de/bstatg_1987/BJNR004620987.html> accessed 15 December 2021.

[32] Anja Malchin, Ramona Pohl, 'Firmendaten der amtlichen Statistik: Datenzugang und neue Entwicklungen im Forschungsdatenzentrum' (2007) 76 Vierteljahrshefte zur Wirtschaftsforschung 8, 13, DOI: https://doi.org/10.3790/vjh.76.3.8.

[33] Florian Köhler, '10 Jahre Forschungsdatenzentren Der Statistischen Ämter – Angebot Und Nachfrage Nach Amtlichen Mikrodaten –' (2012) (June) Statistische Monatshefte Niedersachsen 333.

[34] Ibid.

In order for a scientific institution to have access to microdata, an RDC legally checks the eligibility of the applicant, since the data may be only used by persons who are enrolled in the institution, if their thesis or dissertation is supervised by the institution, they are employees of it or have a guest researcher's status. One further criterion is that users are committed to statistical confidentiality in accordance with section 16 (7) of the BStatG when using a Scientific Use File or visiting a safe centre.[35]

RDCs are bound to specific users, meaning that RDCs may only give access to official microdata to higher education or other institutions entrusted with tasks of independent scientific research. Those who are fall outside that scope have the possibility to keep in contact with the enquiry services of the respective Statistical Offices of the Federation and the Federal States.[36]

2 Use of RDC

Entitled users have two ways of accessing official microdata such as the on-site and off-site use that also can be combined with each other's. They are different from each other mainly in the anonymity of the usable data as well as how they are provided.

The on-site use means that guest researchers can analyse microdata inside the RDC PC workplace (i.e. safe centre). The data in the centres are already protected, not just through the regulation of data access but also through the equipment that researchers can use in the PC workplace; therefore, the microdata – depending on the data sensitivity – can be provided factually or formally anonymised. The PC workplaces are equipped with the common statistical programs, and a separate PC workplace is also available for e-mail communication and internet searches. Throughout their on-site use, researchers have the opportunity to execute remotely when there is no direct access to the data. During this procedure, data users receive data structure files instead that help program codes – that are applied by staff at the statistical offices to analyse the original data – to be prepared using the statistical programs SPSS, SAS, Stata or in some cases R.[37]

For a researcher who wants to use datasets off-site, RDCs offer different Use Files, such as Scientific Use Files (hereinafter: SUF), Public Use Files (hereinafter: PUF) and Campus Files. SUF are standardised datasets that contain factual anonymised microdata.[38] In contrast with the on-site use, SUF offer a lower potential for analyses, but they are suitable

[35] Statistische Ämter Des Bundes Und Länder Forschungdatenzentren (n 28).

[36] 'Terms of Use | Research Data Centre' <https://www.forschungsdatenzentrum.de/en/terms-use> accessed 14 December 2021.

[37] 'Access | Research Data Centre' <https://www.forschungsdatenzentrum.de/en/access> accessed 14 December 2021.

[38] Maurice Brandt, Anja Crössmann and Christopher Gürke, 'Harmonisation of Statistical Confidentiality in the Federal Republic of Germany' (Joint UNECE/Eurostat work session on statistical data confidentiality, 2009) 3–8 <https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2009/wp.5.e.pdf> accessed 14 December 2021.

for large scientific projects, since the factual anonymisation of the microdata can be used outside the statistical offices. Only researchers who are working for registered research institutions located in Germany have permission to use SUF. Due to legal restrictions, those researchers who do not fit these instructions are obliged to access microdata on-site, except if they are using SUF that are offered for the SAS, SPSS, and Stata analysis programs or are provided with the according input routines. For legal reasons, SUF cannot be sent to foreign countries.[39]

PUF's microdata are absolutely anonymised; therefore, only selected variables are available, and variables with high degree of subject-related detail are aggregated. Deeper special delimitations can usually not be made on the basis of PUF. Registered users can have access to agriculture, household, and social welfare statistics.[40] In order to encourage the use of microdata in university teaching, RDCs also offer Campus Files that contain anonymised microdata that can be used by students to acquire methodological knowledge of analysing official microdata.[41] Campus Files are provided for free and for scientific teaching purposes.[42] Currently there are only two categories on the website: Health and Household. Under the category of Health, users can find diagnosis-related group statistics that can be requested starting from the survey year 2005, while microdata concerning Microcensus fall under the Household category.[43]

Working with German microdata in RDC is not free, since the fee depends on the number of used statistics (i.e. number of different data sets provided), survey years and ways of data access. It can easily happen that, for the given project, the data have to be processed in a particular form and only for the project, in which case additional costs can arise. The use of data sets is not unlimited, because the project is tied to a specific purpose; they can be used normally for a period of three years, with the possibility of extension for another three years.[44]

---

[39] Köhler (n 33) 335–336.

[40] 'Public Use Files | Research Data Centre' <https://www.forschungsdatenzentrum.de/en/node/6065#> accessed 14 December 2021.

[41] Markus Zwick, 'CAMPUS-Files – Kostenfreie Public Use Files für die Lehre' (2008) 2 AStA Wirtschafts- und Sozialstatistisches Archiv 175, DOI: https://doi.org/10.1007/s11943-008-0035-x.

[42] Heike Wirth, 'Microdata Access and Confidentiality Issues in Germany' (2008) <https://www.gesis.org/fileadmin/upload/forschung/programme_projekte/sozialwissenschaften/Amtliche_Mikrodaten/wirth_manchester_census_final.pdf> accessed 14 December 2021.

[43] 'Campus Files | Research Data Centre' <https://www.forschungsdatenzentrum.de/en/campus-files> accessed 14 December 2021.

[44] 'User Charge | Research Data Centre' <https://www.forschungsdatenzentrum.de/en/user-charge> accessed 14 December 2021.

3 Anonymity of Microdata

Anonymisation is the way of rendering personal data anonymous.[45] According to Preamble (26) of the GDPR, anonymous data is 'information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable'. Microdata of official statistics are subject to strict confidentiality, therefore RDC makes only anonymised data available that can be absolute, factual or formal.

The above-mentioned PUF and Campus Files are absolutely anonymised, meaning that data are modified by coarsening or by removing individual variables to a degree that the identification of the respondents is no longer possible. Absolutely anonymised microdata are available to all interested persons or institutions and for methodological teaching. If de-anonymisation cannot be ruled out completely and if only unreasonable time, cost and manpower effort make the allocation of data to the respective statistical unit possible, we talk about de-facto anonymised microdata. Different anonymisation procedures can be applied in order to achieve this; for example, the reduction of information or the modification of information.[46]

Based on German law, only de-facto anonymised data can be made available to scientific institutions for the exclusive purpose of scientific projects and may only be used by foreign scientists on the secure premises of the statistical offices. At the RDC, factual anonymity is a matter of the remaining informational value of the data, the parameters of a use of data, the concomitant possibilities for-deanonymisation and the access conditions.[47] Formal anonymity means that the direct identifiers and auxiliary characteristics are deleted from the data set, but, at the same time, the functional and regional structures and all other characteristics remain unchanged. In safe centres and remotely, through remote performance, data users have the opportunity to analyse formally anonymised microdata.[48]

4 Evaluation

The right to data privacy and data confidentiality are important issues in Germany; the more data are collected and merged with other data sources, the higher attention to data security is needed by statistical agencies and researchers.[49] Since the existence of RDCs, there is a successfully implemented data infrastructure in Germany that makes access to

[45] Agencia Española de Protección de Datos and European Data Protection Supervisor, 'AEPD-EDPS Joint Paper on 10 Misunderstandings Related to Anonymisation' (27 April 2021) <https://edps.europa.eu/data-protection/our-work/publications/papers/aepd-edps-joint-paper-10-misunderstandings-related_en> accessed 15 December 2021.

[46] Brandt, Crössmann and Gürke (n 38) 5–8.

[47] Zühlke and others (n 24) 572–573.

[48] Köhler (n 33) 334.

[49] Wirth (n 42) 24–26.

numerous scientific analyses of microdata in all fields of official statistics possible. There is a huge potential in German microdata provided by RDC, because they reflect valid information about different German enterprises, a wide variety of sectors or fields (see the data list provided[50]) – such as health care – that are reliable sources for scientific research or policy decisions.[51]

V Findata

Not only do statistical data have great potential, but also the role of artificial intelligence (hereinafter: AI) is highly important, since it may be applied effectively in the field of health care. Since medical data can be produced every second, they may be aggregated and analysed in order to tackle diseases or only to provide some assistance to people for making their lives healthier. As a recent study by the World Health Organization (hereinafter: WHO) observed, '[i]n recent years, artificial intelligence has made great progress in the detection, diagnosis, and management of diseases'.[52] Although this method may seem straightforward, the reality always reminds us that real world is more complex than that: despite expectations, during the Covid-19 pandemic there was no AI system which could have helped to spot the virus[53] and, in Singapore, using AI in order to replace missing medical doctors turned out to be less successful than it promised before.[54] These findings prove that making these systems work in real life needs high quality data from a trustworthy environment.

Health care, as a sector where data-economy may flourish, is also in the focus of the Commission: in their 2018 study, a chapter was devoted to this sector.[55] Possible

[50] 'Alle Daten | Forschungsdatenzentrum' <https://www.forschungsdatenzentrum.de/de/alle-daten> accessed 16 March 2022.

[51] Malchin and Pohl (n 32).

[52] Digital Health and Innovation, Medical Devices and Diagnostics, *Generating Evidence for Artificial Intelligence Based Medical Devices: A Framework for Training Validation and Evaluation* (World Health Organization 2021) 7 <https://www.who.int/publications/i/item/9789240038462> accessed 16 March 2022.

[53] Will Douglas Heaven, 'Hundreds of AI Tools Have Been Built to Catch Covid. None of Them Helped.' MIT Technology Review (30 July 2021) <https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/> accessed 15 November 2021.

[54] Will Douglas Heaven, 'Google's Medical AI Was Super Accurate in a Lab. Real Life Was a Different Story.' MIT Technology Review (27 April 2020) <https://www.technologyreview.com/2020/04/27/1000658/google-medical-ai-accurate-lab-real-life-clinic-covid-diabetes-retina-disease/> accessed 15 November 2021.

[55] European Commission. Directorate General for Communications Networks, Content and Technology and Deloitte, *Study on Emerging Issues of Data Ownership, Interoperability, (Re-)Usability and Access to Data, and Liability: Final Report.* (Publications Office 2018) <https://data.europa.eu/doi/10.2759/781960> accessed 21 November 2021 DOI: https://doi.org/10.2759/781960.

applications are also named, such as predicting infection, analyse patient population etc.[56] The Commission aims to create data space in the medical field as well.[57]

Even though the medical AI field stands in front of a steep development boom, this is one of the most sensitive and problematic fields in terms of data sharing/usage. As a recent experiment showed, on the one hand, there are major privacy concerns when it comes to sharing medical data (even in anonymised form) and, on the other hand, there are fears as regards data quality.[58] The data quality worries are echoed in the recent WHO study as well, which states that 'unforeseen errors at data entry level can lead to catastrophic effects when deployed at scale if performance errors go unchecked'.[59] In the study, a set of recommendations are provided for researchers in order to conduct data management properly.[60] Proper data management is needed in order to render it possible for medical AI services to provide adequate explanations expected by the GDPR and the forthcoming Artificial Intelligence Act.[61]

The sensitive nature of medical data is why Findata may be seen as one of the most innovative initiatives among data intermediaries, since its goal is to share medical data held by the public sector. Due to this enterprise, Finland is one of the leading countries in secondary use of health data in Europe, according to the Open Data Institute.[62] As they put it in their specific country report, 'Finland should be rightfully proud of the global leadership shown in creating a legislative framework for the secondary use of health and welfare data'.[63]

1 Overview of Procedure and the Legal Framework

Findata's procedure is based on the Act on secondary use of health and social data, enacted in 2019.[64] Besides that, useful guidelines are provided on the official Findata site, from which

[56] Ibid.

[57] 'European Health Data Space' (Public Health – European Commission, 18 September 2020) <https://ec.europa.eu/health/ehealth/dataspace_en> accessed 28 November 2021.

[58] Annie Sorbie and others, 'Examining the Power of the Social Imaginary through Competing Narratives of Data Ownership in Health Research' (2021) Journal of Law and the Biosciences, DOI: https://doi.org/10.1093/jlb/lsaa068.

[59] Digital Health and Innovation, Medical Devices and Diagnostics (n 52) 3.

[60] Ibid 41.

[61] Miranda Mourby, Katharina Ó Cathaoir and Catherine Bjerre Collin, 'Transparency of Machine-Learning in Healthcare: The GDPR & European Health Law' (2021) 43 Computer Law & Security Review, DOI: https://doi.org/10.1016/j.clsr.2021.105611.

[62] Mark Boyd and others, 'Secondary Use of Health Data in Europe' Open Data Institute 38, 6.

[63] The Open Data Institute, 'Finland Profile FINAL' *(Google Docs)* 2 <https://docs.google.com/document/d/1qnK7wlK3gPLBRoPaRw1yGNKWsNW6VMuRIuzwQ-Vtztw/edit?usp=drive_web&ouid=117576810781307564193&usp=embed_facebook> accessed 30 November 2021.

[64] Act of secondary use of the health and social data (552/2019).

the pre-screening criteria for data permit applications document[65] was its main use in this report.

Overall, the Act is comprehensive and clear on nearly all aspects of the procedure. It sets out the main definitions, the parties involved on behalf of the Finnish public sector, the requirements to submit an application, the combination and process for the requested dataset, and the most crucial deadlines for conducting the procedure. Although the most important detail omitted here is the exact method for calculating the fees, this information may be found on the website as well.

2 Key Players

In this legal relationship, three key players may be identified: the applicant, Findata and the data holder controllers. The Act only describes in detail the tasks and rights of Findata and the controllers. The applicant is mentioned only in relation to certain obligations. Thus, the Findata and the controllers shall be presented below according to their roles based on the Act.

Findata can be found in the Act as the 'Data Permit Authority'. Although it seems to be some independent legal person (with a public law background), in reality it is only a separate unit within the National Institute for Health and Welfare. The operational guidance belongs to the Ministry of Social Affairs and Health. At first sight, this arrangement may seem rather inflexible since Findata is not an autonomous legal person that may run its own business. Nevertheless, this solution serves both flexibility and the expected guarantees related to medical data: the National Institute for Health and Welfare is an independent institution under the Ministry of Social Affairs and Health and this relationship is based on a four-year performance agreement.[66] Moreover, inside the Institute, Findata enjoys nearly complete independence;[67] its leadership is intertwined with the Ministry, given that the Ministry appoints the director. As a critical assessment, the political influence may be brought up as a negative factor, since the Ministry could interfere in the functioning of Findata.[68] On the other hand, there is a broad list of organs which supervise the activities of Findata, most importantly the Parliamentary Ombudsman and the Data Protection Ombudsman, to which an annual report shall be submitted 'regarding the processing of

[65] Finnish Social and Health Data Permit Authority, 'Pre-Screening Criteria for Data Permit Applications' (25 October 2021) <https://findata.fi/findata-pre-screening-criteria-for-data-permit-applications/> 15 December 2021.

[66] 'Administrative Branch' (Ministry of Social Affairs and Health) <https://stm.fi/en/administrative-branch> accessed 28 November 2021.

[67] 'Data Permit Authority – THL' [Finnish Institute for Health and Welfare (THL), Finland] <https://thl.fi/en/web/thlfi-en/about-us/organisation/departments-and-units/data-permit-authority> accessed 28 November 2021.

[68] 'About Us' (Findata) <https://findata.fi/en/about-us/> accessed 28 November 2021.

health and social data and the related log data'.[69] In short, the political influence is balanced by the professional supervision and the body's independence, so this structure should not lead to the conclusion that political considerations could jeopardise the data protection measures. As regards the entire system, it could be observed that a proper balance has been formed by giving it rather broad economic and professional freedom but still keeping Findata's activities within the state's administrative structure with all of its guarantees and transparency obligations.

Additionally, by virtue of the Act, a steering group assists and guides the functioning of Findata. The steering committee makes proposals to the institution and the Ministry on various matters which concern the operations of Findata in an annual action plan, with an associated budget, report on operations, financial statements etc. (Section 8). Besides that, a high-level expert group has been set up in order to 'provide guidelines on anonymisation, data protection and data security for Findata's operations'.[70]

Regarding the operational competences of Findata, it is basically the heart of the whole secondary usage procedure. Findata may be deemed a 'one-stop-shop' body, which is responsible for providing security environments for applicant identification (Section 21-22), remote access to the disclosed data (Section 17), examining and deciding on data permits and data requests and combining data from controllers' registers, and anonymising or pseudonymising the data sets (Section 14). Whereas Findata is the central player, its effectiveness relies heavily on the controllers' willingness to cooperate since they provide the data from which Findata serve the final products.

For the controllers, Section 6 gives an exhaustive list which covers virtually all actors which stores medical data (for instance the Social Insurance Institution of Finland, Finnish Medicines Agency Fimea and public services organisers of social and health care). The Act burdens these bodies in particular with four main obligations: 1) providing descriptions for its datasets 2) maintaining advisory services in order to satisfy applicants' enquiries (Section 10) 3) providing Findata with the necessary information during the decision-making phase in order to make a well-founded decision on the feasibility of the data permit/data request (Section 36) and 4) in the event of a granted data permit, provide the desired data for Findata (Section 36). According to Section 36, Findata may also request data from those private providers set forth in the Client Act.

3 Summary of the Act and Findata's Procedure

In line with the Act, the main service of the Findata is to combine datasets from several state databases and provide them to the applicant. Since this type of service could be very

[69] Ibid.

[70] Johanna Seppänen, 'Social and Health Data Permit Authority – Johanna Seppänen PhD, Director' <https://www.ehalsomyndigheten.se/globalassets/dokument/seminarier/finland_findata.pdf> accessed 15 December 2021.

attractive to so many businesses, the combination and processing may be served only for a 'greater good,' such as statistics, scientific research or education (for other examples see Section 2). Although the list of potential applicants is not given (i.e. it could be a for-profit business or a natural person as well), in order to demonstrate the purpose of the application, a data utilisation plan must be submitted to Findata. For instance, according to the pre-screening criteria, in case of a scientific research the following are essential to get approval: 1) an appropriate research plan, 2) the name of the principal investigator, 3) that the results shall be published in scientific publications and 4) the research produces new information.

The Act distinguishes two main types of service among the definitions provided: data permit and data request. The difference lies in the outcome: while someone who was eligible for a data permit shall be awarded secret personal data (i.e., in most cases, a combined database), the data request holder obtains aggregated statistics. In order to put some flexibility in the procedure, Findata is entitled to reclassify a data permit as a data request on the bases of the consent of the applicant (Section 43). The reason for this one-way channel may be is that most applicants seek a data permit since greater value lies in a given database than in receiving aggregated statistics.

The permit application shall be detailed and it must contain all essential information, from billing details through a thorough description of the data requested up to data processing specifics. In order to provide effective assistance in setting the data description (such as giving the register-specific lists of variables and extraction-related delimitations), by virtue of the law, the relevant authorities are obliged to provide a data description, which is available on a dedicated website[71] although only in Finnish (Section 13). Besides that, both the relevant authorities and Findata maintain an advisory service, through which the aspiring applicant may gain information from the controllers as regards the data content of the available registers and the suitability of the data in the registers for their needs and additional information on various matters related to their application (Section 13).

The application shall be submitted via the data request management system (Section 16). Findata has three months to take the decision which may be extended by an additional three months. During this phase, Findata gathers information about the feasibility of the application from the controllers (Section 36) and the applicant may be asked for further clarification. The controllers' contribution is vital also for calculating a fee (they must deliver an estimated cost), which is communicated, together with the final extraction description, to the applicant. The applicant may accept or reject them both; in other words, it is not possible to request some sort of decrease in the fee or reduce the service already offered.

Although the fees are not specified in the Act, they are also based on the law, namely Decree 1168/2020 of the Ministry of Social Affairs and Health on charges for work carried out by the health and social data permit authority Findata. The fees are fully transparent since they are published in a clear manner on the website (for example a data request and

[71] 'Etusivu – Aineistokatalogi' <https://aineistokatalogi.fi/catalog> accessed 27 November 2021.

a data permit for EU business is 1,000 EUR, while the processing fee is 115 EUR/hour). According to the information set forth there, the first invoice and payment are due after the decision on the data request or the data permit. The price for processing and delivering the data (which was given along with the decision) shall be invoiced after the delivery of the results to the applicant.

Having granted the data permit (and received the first payment), Findata takes the necessary steps to deliver the desired outcome. It collects the necessary data from the controllers, combines them, makes them secret via anonymisation or pseudonymisation[72] and then make the final result available via a secure hosting service (Section 51). By virtue of the law, this procedure shall be conducted within 60 days from which the controllers have 30 working days for the data handover. As the main rule, the secure hosting service is a remote access environment (Kapseli) provided by Findata, for which there is a monthly charge as well and exception may be made only 'if the data utilisation plan and the data permit state a separate reason that necessitates it'. After the disclosure, the applicant has 30 working days to review the delivered data and indicate any problems which it has experienced. According to Section 43, a data permit can only be granted for a fixed period. This is elaborated more in the pre-screening criteria, which explicitly state that the maximum period is five years unless there is a proven justification for a longer period.[73]

4 Findata and the GDPR

As mentioned before, due to the sensitive nature of medical data recognised by virtue of Article 9 Paragraph 1, compliance with the GDPR is essential. Nonetheless, the Act mentions the GDPR when it states that 'the provision of this Act is supplementary to those laid down in' the GDPR (Section 2). The harmonisation between the Act and the GDPR has been criticised by the assessment drafted by the Commission in which it observes that the Act 'does not stipulate the legal basis that should be used for further processing in public sector research'.[74]

---

[72] While personal data loses its personal nature due to anonymisation, pseudonymised data still qualifies as personal data according to the GDPR at first glance (Recital 26). This opinion has been debated by Mourby and others, who claim that pseudonymised data could qualify as non-personal data if the relationship could be restored only with substantial difficulty. See in detail: Mourby M and others, 'Are "Pseudonymised" Data Always Personal Data? Implications of the GDPR for Administrative Data Research in the UK' (2018) 34 Computer Law & Security Review, DOI: https://doi.org/10.1016/j.clsr.2018.01.002.

[73] Finnish Social and Health Data Permit Authority, 'Pre-Screening Criteria for Data Permit Applications' (25 October 2021) <https://findata.fi/findata-pre-screening-criteria-for-data-permit-applications/> accessed 28 November 2021.

[74] Consumers, Health, Agriculture and Food Executive Agency, *Assessment of the EU Member States' Rules on Health Data in the Light of GDPR* (Publications Office 2021) 70 <https://data.europa.eu/doi/10.2818/546193> accessed 28 November 2021, DOI: https://doi.org/10.2818/546193.

Given the fact that one of the duties of Findata is to anonymise or pseudonymise the data received from the controllers, it is rather clear that the Findata works with personal data. In that case however, on the one hand the legal role of Findata should be stipulated (is it a controller or a processor?); on the other hand, the legal basis of the data processing must be based on the GDPR.

As regards the first question, the difficulty of the situation derives from several factors: Findata receives data for its own purposes therefore 1) it has no data unless it is provided 2) Findata acts as dictated by the applicant and not by the controllers. By virtue of the GDPR, Findata could be controller, joint controller or processor.

Findata could be classified as a processor if there would be some sort of hierarchical relationship between Findata and the controller in which 'the processor obeys the dictates of the controller'.[75] In this particular case, it cannot be true since Findata acts in the name of the applicant and the controller cannot prescribe any order to Findata. One key factor which is often highlighted by the relevant commentaries is that data controllers must determine rather precisely the task(s) of the data processor.[76] If that were the case, Findata may not fulfil its obligations towards the applicant since Findata should strictly follow the controllers' instructions.

In line with the GDPR, the controllers and Findata may be joint controllers together. Although this seems a viable options at first glance, in order to apply this provision Findata and the controllers must jointly determine the purposes and means of processing. It is rather clear, however, that Findata and the controllers have different aims and different activities and so they cannot perform processing together.

This type of cooperation resonates to the example of where several companies use the same camera system (hence the same records) but the decisions regarding the data processing are taken independently, in which case the companies do not qualify as joint controllers.[77] It can be applied in this case: Findata and the controllers use the same data, although for different aims. Nevertheless, this analogy is not perfect either: it is debatable in a narrow sense whether they use the same database (since probably a modified one is provided for Findata which can hardly be used for the daily operation of the controller) and their independence is not full in the sense that Findata is the one that makes requests to the controller. Even if it may be concluded that the legal status could be clarified as more in line with the GDPR regime in the Act, the website of Findata explicitly says that Findata is a data controller[78] which is quite evident according to the reasoning above.

---

[75] Christopher Kuner, Lee A. Bygrave and Christopher Docksey (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (Oxford University Press 2019) 160.

[76] Péterfalvi Attila, Révész Balázs and Sziklay Júlia (eds), *Magyarázat a GDPR-ról [Commentary of the GDPR]* (Wolters Kluwer Hungary 2018) 85.

[77] Ibid, 83.

[78] 'Data Protection and the Processing of Personal Data' (Findata, 10 September 2021) <https://findata.fi/en/data-protection-and-the-processing-of-personal-data/> accessed 12 December 2021.

Another essential topic is the legal basis under which the Findata processes personal data since there is a general prohibition on processing personal data concerning health. This general prohibition can only be overwritten in certain cases given in the GDPR itself.

Three exceptions could be considered applicable to the activities of Findata: Article 9 (2) h), which defines exceptions for medical cases and the management of health or social care systems and services and Article 9 (2) i), which prescribes exceptions for public health matters, and lastly Article 9 (2) j), which creates exceptions for scientific reasons.

The first exception seems the most suitable, since it provides a wide discretion in terms of medical matters. Although a medical research project may not fall under this provision,[79] other bodies could claim for the services of Findata if it is needed in order to develop their functioning pursuant to Section 41. The reference for the GDPR's exception is included in this section of the Act.

Article 9 (2) i) may be understood in a rather narrow way 'that is intended for use by public health authorities, NGOs and other entities working in areas such as disaster relief and humanitarian aid, and similar bodies'. Since there is no indication in the Act that Findata has some sort of obligation to provide services in such grave situations, this exception does not seem applicable.[80]

The third exception, namely scientific research, aligns perfectly with Findata's goals. This provision requires appropriate safeguards in order to guarantee high standards regarding data protection, although it is a rather vague obligation and 'it is not specified what is meant by "suitable and specific measures to safeguard the fundamental rights and the interests of the data subject"'.[81] According to the cited authors,

> In light of the lack of specificity in the text, and absent more detailed guidance from the EDPB, controllers and processors will have to design safeguards based on principles underlying the GDPR, such as proportionality, data minimisation and data security. This can include a variety of measures based on the purposes of processing and the sensitivity of the data, such as encryption, minimising the amount of sensitive data processed, training personnel who handle personal data and placing personnel under a duty of confidentiality.

As we have seen from Findata's procedure, these principles prevail during the whole procedure: for example a detailed research plan must be given, the claimed set of data must be as limited as possible and the (anonymised or pseudonymised) data is delivered only via the secured environment provided by Findata for certain people and for fix period to avoid any chance to reidentification.

Given this reasoning, it is rather surprising that this provision is not directly invoked by the Act. Although Section 38 covers the data permit for scientific research and statistics,

[79] Kuner, Bygrave and Docksey (n 75) 379.
[80] Ibid 380.
[81] Ibid.

it points only to the Data Protection Act (1050/2018) effective in Finland. According to Section 6, Article 9 (1) of the GDPR is not applicable to data processing for scientific or historical research purposes or for statistical purposes, and in there is a wide range of measures which must be implemented in order to safeguard the rights of the data subject. Although, from Findata's side, these safeguards have been adopted (for instance one of the measures is the pseudonymisation of personal data, which is already performed by Findata), the from controllers' side this raises some questions as regards the meaning of the provision (i.e. whether all the measures must be taken or is it only a set of recommendations instead).

There are some other use-cases where the Act invokes another GDPR provision. In order to produce educational materials (Section 39) and client data necessary for the planning and reporting duty of authorities covered in Section 6 (Section 40) the legal basis for processing personal data is Article 9 (2)(g). This is rather curious in light of the following: 'To process sensitive data, the public interest must be 'substantial', in contrast to the conditions for processing personal data based on a task carried out in the public interest under Article 6(1)(e), where there is no such requirement'.[82] It is added that some examples may be found in Recital 46, none of which seems to be applicable to the case of Findata (for instance 'humanitarian purposes, including for monitoring epidemics and their spread or in situations of humanitarian emergencies, in particular in situations of natural and man-made disasters'). In both clauses the requirement of having a data permit is included so most probably – in line with the information found on the website[83] – the legal basis for processing data to serve the data permit is Article 9(2)(g).

Another GDPR-related issue is the transfer of the data. Whereas the primary option for data delivery is Findata's own Kapseli secure environment, hypothetically, an applicant may request transfer to other platform. According to the information found on the Findata website, in this case the GDPR rules prevail in other words the transfer may be carried out only in certain cases outside the EU/EEA in line with Chapter V of the GDPR.[84]

5 Current State of Play

Findata started to build up its structure and organisation in 2019 and it took nearly a year to begin to operate. From 1st January 2020, it started to receive data requests and, from 1st April 2020 data permit applications began to arrive.[85] According to their statistics published on their website, in the course of its history (up until 23rd November 2021) the number of submitted applications was rather high (598), nevertheless the number of pending applications and application under process are quite high as well (198).[86]

[82] Ibid 379.

[83] 'Data Protection and the Processing of Personal Data' (n 78).

[84] 'Data Permits' (Findata) <https://findata.fi/en/data-permits/> accessed 12 December 2021.

[85] 'What Is Findata?' (Findata) <https://findata.fi/en/what-is-findata/> accessed 28 November 2021.

[86] 'Findata' (Findata) <https://findata.fi/en/> accessed 12 December 2021.

According to the assessment cited above, 'Findata has an indicative budget of 1 million EUR' (which was higher in the starting years), which seems quite high but still rather modest compared with other similar bodies; for instance, Health Data Hub France was granted initial funding of 36 million euros for four years.[87]

6 Evaluation

By establishing Findata, Finland created a popular, transparent and secure process in order to provide medical data for scientific research. Although the full potential of this organisation may be exploited later, it seems that the whole arrangement from the applicants' submission up to the delivery of the results has been considered deeply and wisely. The only criticisms that could be brought up are the relative slowness (as regards the above-mentioned timelines and waited applications) and the possible political influence through Findata's leadership.

VI The Secure Access Data Centre

The role of AI is indispensable, since the development of machine learning algorithms depends on large volumes of data, which overall boost the data economy; therefore it is not surprising that AI is not only used by Findata, but also plays a significant function in the French data policy. France represented its AI strategy in 2018 for a 5-year period based on the French AI policy report[88] that aims to establish an open data policy in order to implement AI applications and pool assets together. The French strategy has a huge focus on infrastructure, highlighting data policy initiatives such as the CASD secure Data Hub aiming to help exchange sensitive protected data for research and development projects securely [89] by hosting private data from the bank, service, transport and private health industry and by making them available to researchers or private operators on a voluntary database to support the development of value-added services.[90]

The Secure Access Data Centre (hereinafter: CASD) is a public interest group in France aiming to organise and implement secure access services for confidential data[91] – that

[87] Consumers, Health, Agriculture and Food Executive Agency (n 74) 111.

[88] Cédric Villani, 'For a Meaningful Artificial Intelligence. Towards a French and European Strategy' (2018) <https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf> accessed 15 December 2021.

[89] 'France AI Strategy Report | Knowledge for Policy' <https://knowledge4policy.ec.europa.eu/ai-watch/france-ai-strategy-report_en> accessed 14 December 2021.

[90] CNL, 'Topics for Consideration 2019' 8, <https://www.cnil.fr/sites/default/files/atoms/files/topics-for-consideration-2019.pdf> accessed 15 December 2021.

[91] Jean-Pierre Le Gléau and Jean-François Royer, 'Le centre d'accès sécurisé aux données de la statistique publique française: un nouvel outil pour les chercheurs' (2011) 130 Courrier des statistiques 1, 2.

cannot be published as open data[92] – for non-profit research and study and to promote the technology developed to secure access to data in the private sector. In 2010, the Group of National Schools of Economics and Statistics (hereinafter: GENES) and the National Institute of Statistics and Economic Studies (hereinafter: INSEE) jointly carried out the CASD project in the framework of the 'Investment of the Future Programme' with the recognition of the Equipment of Excellence. CASD was granted funding managed by the National Research Agency through to the end of 2019.[93] As part of a consortium agreement in 2012, other institutions endorsed the objectives pursued by the project in order to allow the development of CASD as a service for access to confidential data for research, study, evaluation and innovation.[94]

The ECOO1832598A interministerial decree by the Ministry of Economy and Finance of 29th December 2018 created CASD as a public interest group – having a legal change in its status – in order to bring together the State, represented by the Director General of the National Institute of Statistics and Economic Studies (INSEE), the Group of National Schools of Economics and Statistics (GENES), the National Centre for Scientific Research (CNRS), the Polytechnic School *(L'école politechnique)* and HEC Paris.[95] These institutions represent ministries, public establishments of a scientific, cultural and professional nature, science and technology and higher education; as such, the General Assembly of the Group is able to make informed decisions for the orientation of the group since the members have experience from several spheres and several disciplines.

In the scope of CASD, there are three advisory committees, namely the Scientific Council, the Data Producers Committee and the Information System Security Policy Monitoring Committee. In the Scientific Council – as the governance body – there are 14 experts with an international background in the field of data analysis, data processing and dissemination. Their main tasks are to give scientific orientation, ensure technological, methodological, legal and ethical oversight of access to confidential data related to international developments and suggest partnerships with similar centres while ensuring that CASD is represented well in France and on abroad. The Council has an important role in assisting the General Assembly and the director of the organisation in innovation,

---

[92] Emile Marzolf, 'Comment l'État veut s'emparer des données pour améliorer la gestion de ses RH | À la une | Acteurs Publics' (Comment l'État veut s'emparer des données pour améliorer la gestion de ses RH | À la une | Acteurs Publics, 4 October 2021), <https://www.acteurspublics.fr/articles/comment-letat-veut-semparer-des-donnees-pour-ameliorer-la-gestion-de-ses-rh> accessed 14 December 2021.

[93] 'Equipex – Le CASD – Centre d'accès sécurisé aux données' <https://www.casd.eu/en/le-centre-dacces-securise-aux-donnees-casd/partenaires/> accessed 14 December 2021.

[94] 'Convention Constitutive Groupement d'Intérêt Public CASD' (8 October 2018) <https://www.casd.eu/wp/wp-content/uploads/CASD-conv-const-20181008_V3.00_signee.pdf> accessed 14 December 2021.

[95] Arrêté du 20 décembre 2018 portant approbation de la convention constitutive du groupement d'intérêt public « Centre d'accès sécurisé aux données » Journal officiel de la République française, texte 53 sur 202, 29 decembre 2018; <https://www.casd.eu/wp/wp-content/uploads/joe_20181229_0301_0053.pdf> accessed 15 December 2021.

ethics and scientific strategy. The Data Producers Committee helps the main bodies related to data access conditions, documentation, archiving and dissemination of information, while the Information System Security Policy Monitoring Committee assists in matters of information system security governance.[96]

The group – as part of its research service missions – is responsible for implementing secure services access to confidential data. Its participation in the operations of matching and anonymising data, documenting and archiving confidential data and in developing access to confidential data at national, European and international level in connection with other data provision mechanisms are also key elements.[97]

As part of its valuation missions, in particular with the competitive sector, the group initiates the provision of advice and expertise in its areas of expertise to the State and other French entities, to provide tools or security services in its areas of competence and to provide the technology for securing access to data for private interest purposes. CASD aims to ensure that data depositors store, make available and use their data and protect the confidentiality of such data, maintain a high level of infrastructure and quality of service that allows users to access data under good conditions; and provide secure and equitable access to accredited data users, allowing for advanced processing and analysis under the best working conditions.

1 Technology

CASD is a benchmark in the very sensitive and complex world of data security, since the Group anticipates needs, innovates and helps build a regulatory framework compatible with a digital society that is open and protective. If the individual data produced by the public sphere are increasingly voluminous and of high added value from a scientific point of view, their use has so far remained difficult for reasons of confidentiality. These personal data are generated and produced by CASD partners, but mostly by INSEE, official statistics, administration and health data producers.[98]

CASD offers a secure infrastructure, called 'secure bubbles,' to data producers, guaranteeing a very high level of security. The so-called SD-Box – which can only be installed in the premises of a legal entity[99] – is an autonomous terminal designed by CASD, which very simply consists of a single unit with all elements necessary for the

[96] 'Governance and Missions – Le CASD – Centre d'accès sécurisé aux données' <https://www.casd.eu/en/le-centre-dacces-securise-aux-donnees-casd/gouvernance-et-missions/> accessed 14 December 2021.

[97] 'Convention Constitutive Groupement d'Intérêt Public CASD' (n 94).

[98] Kamel Gadouche, 'The Secure Data Access Centre (CASD), a Service for Datascience and Scientific Research – Courrier Des Statistiques N3 - 2019 | Insee' (22 June 2021), <https://www.insee.fr/en/information/5014754?sommaire=5014796> accessed 14 December 2021.

[99] Jean-Pierre Le Gléau, 'L'accès aux données confidentielles de la statistique publique – De la sensibilité des données économiques à la sensibilité des données de santé' (2014) 2 Statistique et société 27, 30.

services it must provide enclosed.[100] It allows remote access to a secure infrastructure where confidential data are safeguarded. Due to the SD-Box, users return to the familiar interface of a workstation, but they only have access to data for which authorisation has been granted.[101] The SD-Box meets key IT security requirements while it is easy to install; it has automated maintenance, low dependence on local IT infrastructure and has a low impact on the user IT environment.[102]

The SD-Box is a key tool for accessing the whole CASD environment from outside, since it establishes a secure web link with the CASD central infrastructure that is designed to enable the processing of detailed confidential data. The main principles are guaranteeing the highest level of security to preserve the confidentiality and integrity of data, enabling users to benefit from workspace and minimising the SD-BOX's impact on their IT systems.[103]

2 Compliance with GDPR and Security

Since CASD grants access to confidential data and hosts personal data from several major institutions, it is a key requirement that the Group meets the requirements for data protection. CASD is ISO 27001 compliant in the field of Information Security Management, which is a reference to taking into account the best practices in the field of personal data protection. By being certified, CASD assures users that any information relating to an identified or identifiable person, directly or indirectly within this infrastructure, is based not just on legislation but on certifications issued by competent and authorised agents.[104] In November 2021, CASD participated on the Club 27001 annual conference to share experiences of the implementation of the ISO 27001 standard and discuss best practices.[105]

Health Data Hosting certification has also been obtained by CASD, which is highly valuable for the organisation since it is very active in granting access to health data collected during healthcare activities such as prevention or diagnosis. The management system of the IT infrastructure protects sensitive data and secures all information while preventing the

[100] Nathalie Picard and Kamel Gadouche, 'L'accès aux données très détaillées pour la recherche scientifique', (Université de Cergy-Pontoise 2017, THEMA Working Papers 2017/06) 9–10.

[101] 'Secure Data Access Center | ENSAE Paris' <https://www.ensae.fr/centre-acces-securise-aux-donnees/> accessed 14 December 2021.

[102] 'SD-Box – Le CASD – Centre d'accès sécurisé aux données' <https://www.casd.eu/en/technologie/sd-box/> accessed 14 December 2021.

[103] 'Infrastructure – Le CASD – Centre d'accès sécurisé aux données' <https://www.casd.eu/en/technologie/infrastructure/> accessed 14 December 2021.

[104] 'L'ISO 27701, une norme internationale pour la protection des données personnelles | CNIL' (2 April 2020) <https://www.cnil.fr/fr/liso-27701-une-norme-internationale-pour-la-protection-des-donnees-personnelles> accessed 14 December 2021.

[105] Célia Seramour, 'La conférence annuelle du Club 27001 se tiendra le 4 novembre' (Le Monde Informatique) <https://www.lemondeinformatique.fr/actualites/lire-la-conference-annuelle-du-club-27001-se-tiendra-le-4-novembre-84699.html> accessed 14 December 2021.

risks of cyber-attacks; therefore a trusted e-health and patient follow-up environment can by promoted by CASD. It also has Health Data Security Standard certification, which was developed on the basis of a rigorous risk analysis in order to put the appropriate security measures in place. The Standard is applicable to the National Health Data System. Based on a decree (22 March 2017), the data that is available in the System classified sensitive data.[106] These certifications guarantee secure hosting and data processing infrastructure services via the SD-Box – ensuring biometric access control and encrypted connection – installed in the establishment under a contract with CASD.[107]

3 Access to Data

In order to access confidential data as a user, it is necessary to be authorised either by the Statistical Secrecy Committee (hereinafter: CSS) in France or directly by the data depositor. Through the Confidential Data Access Portal, it is necessary to create an account while signing a confidentiality agreement. The complete file must be submitted to the Committee, which conducts its deliberation and sends its result to the project leader. After the project is greenlighted, it is compulsory to sign an agreement with data producers and the French Archives, which sends it back not just to the project leader but also to the CASD, which concludes the legal procedure. When requesting the right to access some data sources (justice, higher education, housing, baking, rural development, marine, etc.), it is necessary to contact the data depositor, who will send an authorisation document to CASD, that can start the process of creating access.[108]

At the end of both procedures, project members have to sign a contract – since the CASD is a paid service that must be contracted in advance – and participate in an enrolment session; this is a mandatory step. During the awareness training, users will receive essential information on legal, statistical and IT issues while getting an Access Card to the SD-Box with the applicant's encrypted own fingerprint, allowing CASD to grant users access to confidential data[109] and that cannot be lent to anyone under any circumstances.[110] It is very important that data provided by CASD are accessible from abroad too, since all countries in the European Union are subject to the ongoing accreditation.[111]

---

[106] 'Certifications & Security – Le CASD – Centre d'accès sécurisé aux données' <https://www.casd.eu/en/technologie/securite-certifications/> accessed 14 December 2021.

[107] Marcel Goldberg and Marie Zins, 'Le *Health Data Hub* (fin). De multiples problèmes et des solutions alternatives?' (2021) 37, 277, DOI: https://doi.org/10.1051/medsci/2021017.

[108] 'Procédures d'habilitation – Le CASD – Centre d'accès sécurisé aux données' <https://www.casd.eu/gerer-son-projet/procedures-dhabilitation/> accessed 15 December 2021.

[109] 'Contractualisation – Le CASD – Centre d'accès sécurisé aux données' <https://www.casd.eu/gerer-son-projet/contractualisation-2/> accessed 14 December 2021.

[110] 'Séance d'enrôlement – Le CASD – Centre d'accès sécurisé aux données' <https://www.casd.eu/gerer-son-projet/seance-denrolement-2/> accessed 14 December 2021.

[111] 'FAQ' <https://www.casd.eu/gerer-son-projet/faq/> accessed 15 December 2021.

4 Projects

With its technology, CASD is active both in the public and in the private sector, since it makes data available for tax data from INSEE and the Ministries of Justice, National Education, Agriculture and Food, Economy and Finance and access must be provided by them. In the private sector, there is a long list of companies that are in cooperation with CASD.[112] Since CASD is a division of GENES, secure dissemination of data is allowed through it, because GENES is a trusted third party. The data producers provide access to their data through CASD while keeping complete ownership of the data. These factors result in a rising demand on the user's side: since the beginning of the project there has been a 30% increase in the number of projects: 80% of the requests come from public institutions and 20% from private organisations in the energy, transport, banking and insurance sector which can analyse, process and cross-reference data with several sources and collaborate with other users from different countries involved in the same project.[113] Furthermore, CASD is highly active in projects based on health data; for example, it has already provided SNDS data to identify drugs that protect against Parkinson's disease,[114] or evaluate the impact of comedications on chemotheraphy efficacy for breast cancer,[115] medicine, surgery and odontology data in order to evaluate hospital activities[116] or administrative data in order to develop an algorithm for tracking fragility and dependence in health insurance; such ant indicator will lead to relevant conclusions in the field of health surveillance, research and disease prevention among the elderly.[117] Proving that CASD strives to open up new opportunities for scientific research and statistical studies, it has signed a partnership agreement with the Banque de France in order to make nearly 75 sources of banking data available on the CASD.[118]

---

[112] 'CASD – Le CASD – Centre d'accès sécurisé aux données' <https://www.casd.eu/en/le-centre-dacces-securise-aux-donnees-casd/le-casd> accessed 14 December 2021.

[113] 'Secure Data Access Center | ENSAE Paris' (n 101).

[114] 'Use of the SNDS for the Identification of Drugs Protective of Parkinson's Disease – Le CASD – Centre d'accès sécurisé aux données' <https://www.casd.eu/en/project/use-of-the-snds-for-the-identification-of-drugs-protective-of-parkinsons-disease/> accessed 14 December 2021.

[115] 'Analyse des relations entre comédications et réponse à la chimiothérapie pour un cancer du sein – Le CASD – Centre d'accès sécurisé aux données' <https://www.casd.eu/project/analyse-des-relations-entre-comedications-et-reponse-a-la-chimiotherapie-pour-un-cancer-du-sein/> accessed 14 December 2021.

[116] 'Traitement des données du PMSI par la société IRIS CONSEIL SANTE – Le CASD – Centre d'accès sécurisé aux données' <https://www.casd.eu/project/traitement-des-donnees-du-pmsi-par-la-societe-iris-conseil-sante/> accessed 14 December 2021.

[117] 'Développement d'un algorithme de repérage de la fragilité et de la dépendance dans les bases médico-administratives de l'Assurance Maladie – Le CASD – Centre d'accès sécurisé aux données' <https://www.casd.eu/project/developpement-dun-algorithme-de-reperage-de-la-fragilite-et-de-la-dependance-dans-les-bases-medico-administratives-de-lassurance-maladie/> accessed 14 December 2021.

[118] 'Les données de la Banque de France bientôt disponibles sur le CASD – Le CASD – Centre d'accès sécurisé aux données' <https://www.casd.eu/les-donnees-de-la-banque-de-france-bientot-disponibles-sur-le-casd/> accessed 14 December 2021.

5 Evaluation

The main objectives of the organisation are to maintain a high level of infrastructure, allowing users to access data safely and analyse them under the best working condition while guaranteeing the data protection rules. CASD, as a trusted third party, underpins its activity by these strong imperatives that remain at the heart of the approaches it undertakes.[119]

VII Common Patterns

In this section some of the identified common patterns for all three institutions will be presented in order to demonstrate the main challenges and issues with which they face. This is not intended to be a comprehensive enumeration, although these are the most essential concerns which it is advised to consider while establishing bodies like these.

1 Primary Aims

All of these organisations have been founded to provide appropriate data and services for scientific, statistical research and non-profit goals. This may be seen most clearly in the case of RCD, since the main condition for eligibility is enrolment in an institution with an ongoing thesis or dissertation under professional control. Findata provides data primarily for scientific research and, besides that, there is a lower price for data needed for a thesis. The consortium agreement of CASD back in 2012 highlighted the importance of research as well.

Paradoxically, this approach underlines the importance of the Commission's main purpose, which is opening data to the for-profit private sector in a wider manner. While the priority character of academic scientific research is indisputable, in order to strengthen the European economy the private sector should benefit from these services and data sets.

2 Legal Background

The most basic similarity may be spotted in terms of the 'big picture': there is some sort of legal background behind all these bodies. Nevertheless, the solutions vary from country to country: Finland dedicated a sectoral, separated act in order to regulate the process and cooperation of the relevant bodies, Germany inserted these types of activities to laws in effect and France arranged it by a decree. Given the fact that all legal systems have their own standards and peculiarities, the differences are not surprising. As a matter of fact, from an outer 'perspective' the Finnish model seems the most straightforward in terms of its clarity

[119] 'Governance and Missions – Le CASD – Centre d'accès sécurisé aux données' (n 96).

and comprehensibility (i.e. the relevant rules may be found in one source, and it may be found quite easily since acts generally have the highest legislative rank).

This legal arrangement entails some kind of centralisation as well. Whereas the desire to set up a one-stop shop is obvious in the case of Findata, it is fairly evident in the case of RCD and CASD too: RCD takes care of the centralisation of the data and CASD provides an Access Card to the SD-Box with confidential data from various sources.

Although this centralisation makes the process and the access to data easier, it may raise some privacy concerns as well. This is where conceptual problems occur: data access is highly important (and it is effective if the individual elements may be reached, too) but effective safeguards are needed in order to make sure that one cannot misuse these opportunities (given the fact that it is getting harder to find the boundary between personal and non-personal i.e. anonymised data as Purtova, for instance, demonstrated[120]).

3 Accessibility

Due to the above-mentioned privacy issues, all of the bodies need to take serious security measures. One of the common solutions is that the institutions rarely give the data to the client without constraints: Findata makes it available only for certain person/s via its secure environment, registering all the logs related to operations there, RCD provides full potential to the data in on-site premises and CASD devised its own tool to keep the data safe.

Additionally, there is a selection process for the aspiring applicant, during which the aims of the applicant and the applicant himself/herself/itself are examined. This raises the question whether these services should be available to persons/business outside the EU. While fair competition would demand equal terms, this could raise security and digital sovereignty issues as well. In this regard practices also diverge: whereas the German core services are available only for domestic students, Findata provides services with higher prices for outside the EU.

On the one hand, these safeguards seem quite appropriate, on the other hand there is a hazard in joint data generated by data intermediaries possessed by them (and in this way indirectly by the state). In order to prevent any misuse from the state's side, there must be established a proper institutional check as implemented in the case of Findata.

VIII Conclusion

The aim of the article was to present how the three bodies indicated in the impact assessment of the DGA have been set up by the three Member States. In order to get an

[120] Nadezhda Purtova, 'The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law' (2018) 10 Law, Innovation and Technology 40, DOI: https://doi.org/10.1080/17579961.2018.1452176.

overall point of view, the European Digital Agenda has been introduced with the most important milestones.

The first example was the German RCD with the main purpose of providing data for academic research. As was shown, the background of the RCD is more traditional, in the sense that it has been incorporated into the existing institutional structure rather than creating new networks. This example demonstrates how crucial is the anonymisation and the confidentiality of data, as it is a major issue tackled by the relevant German laws.

The second example was the Finnish Findata, which is a specific sectorial intermediary with the sole purpose of providing services in the medical field. The functioning and the structure of Findata is rather clear due to the dedicated Act and there are also available materials on its website which makes its process rather transparent. The establishment of Findata put emphasis on compliance with GDPR and created effective safeguards as well.

The third example was the French CASD. While the Finnish and the German examples were more conservative in terms of their obligations and services, CASD plays a role in legislative work by providing assistance on various matters and it provides data via a special tool which has been produced in order to prevent the misuse of confidential data.

Finally, the main points were analysed, aiming to identify the primary hurdles which may occur related to these bodies. Three common patterns have been demonstrated: the primary aims of the organisations, their legal background and the accessibility of data. It has been shown that the primary aim is to contribute to scientific and statistical research (thus private, profit-orientated activities are not backed up by these services), their legal background is settled but using various modes and the accessibility (security) of the data provided are extensively safeguarded.