I. Introduction

"Society Needs Research, Research Needs Data".[1] Access to data for scientific (and academic) research is prerequisite to making knowledge gains, particularly in relation to issues of societal importance. In the healthcare sector, for example, data are crucial for studying side effects of drugs and vaccines and coming up with more effective treatment possibilities. The research community needs access to the databases of platforms of the internet business giants in order to study the algorithms being employed, for these are of relevance to issues of discrimination, overblocking (censorship), disinformation tactics and user manipulation. Spotify, for example, has developed speech recognition software that makes song and ad recommendations based on the user's mood.[2] One risk this entails is creating a projection bias that leads to bad consumer purchases.[3] In the mobility sector, greater access to research data could make possible the development and testing of intelligent transport systems. In the energy sector, similarly, power consumption data and the driving factors could be analyzed to reap energy efficiency gains.

Complaints over insufficient access to research data are becoming increasingly louder.[4] As far back as 2015 however, *Havel*, in a dissertation entitled "Information Access Rights for Academic-Scientific Research" (title translated), called for broad information access rights for research purposes,[5] and *Wielsch* outlined media-specific principles for accessing intellectual property in his *habilitation* work entitled "Principles of Access: Knowledge Sharing and the Law"[6]. The study at hand builds upon the ideas developed in those publications. The Competition Law 4.0 Commission, too, has now come out in favor of research data access for scientific purposes (albeit by a narrow majority).[7]

De lege lata, access to social, behavioral and economic research data in Germany takes place essentially within a system comprising 39 research data centers which are accredited by the Council for Social and Economic Data (Council SWD). These national research data centers are to be strengthened under the National Research Data Infrastructure Initiative, but without abandoning the system as a federal infrastructure. Access channels to research data centers, for example, are to be rendered more transparent and uniform, and a state-wide network of access points (e.g. guest

---

[1] Title of the 8th Social and Economic Data Conference held by the Council on Social and Economic Data (RatSWD) 2020.
[2] *Savage*, Spotify wants to suggest songs based on your emotions, URL: https://www.bbc.com/news/entertainment-arts-55839655, last accessed 7/16/2021.
[3] *Heidhues/Köster/Köszegi*, Steering Fallible Consumers, 2021, pp. 10-11, URL: http://www.personal.ceu.hu/staff/Botond_Koszegi/steering.pdf, last accessed 7/16/2021.
[4] *Peichl/Bachmann/Riphahn*, FAZ 8/6/2021, URL: https://www.faz.net/aktuell/wirtschaft/forschern-haben-in-deutschland-zu-wenige-daten-zur-verfuegung-17471899.html, last accessed 08/20/2021; in contrast to the status quo seen by *Hevers* in 2015 in Informationszugangsansprüche des forschenden Wissenschaftlers, p. 470 f., with the observation that this is likely due to the fact that at that time lots of data was made available voluntarily.
[5] *Havel*, Informationszugangsansprüche des forschenden Wissenschaftlers, p. 453.
[6] *Wielsch*, Zugangsregeln – Die Rechtverfassung der Wissensteilung.
[7] Final Report, p. 46.

researcher workplaces) and a federal archiving infrastructure are to be set up.[8] While a crucial foundation for empirical research, the enabling of access via research data centers is not the primary focus of this study. Rather, this study primarily concerns the ensuring of access to data collected by public and private-sector entities which are not already retained in research data centers *de lege lata*, such data being highly relevant for the sectors being studied. Research data centers could, however, be integrated into the data access ecosystems outlined in this paper.

There are various ways to guarantee data access under substantive law, as contractual information claims can ensure data access just as well as data protection law.[9] Antitrust grounds for access claims are another possibility.[10] However, none of these access claims are principally relevant to the interests of the research community, enabling access as well—significantly—by parties who have no scientific interests. This paper is explicitly concerned with research clauses which ensconce privileged access for research purposes. Accordingly, access rights of a general nature are not the concern herein.

To date, data access on grounds other than contractual clauses, data protection law and antitrust principles has only been provided in a rudimentary manner. Only a few clauses are found that allow or require data access for research purposes (research clauses), such as § 19 (3) of the Copyright Service Provider Act (UrhDaG), § 5a of the Network Enforcement Act (NetzDG), § 8 of the draft Federal Cancer Register Act and § 303e of Social Code, Book 5 (SGB V). Data falling under the data access rights are furthermore limited in scope, and access requirements are structured in a highly heterogeneous fashion. Pursuant to Art. 31 of the draft Digital Services Act (DSA-E), for example, researchers may only have data access via a third party structured as a public authority. Finland employs a similar model for healthcare data, having set up a national authority (Findata) to coordinate research data access. Similar approaches are being taken in Australia as well. Both Australia and the UK are looking at improving data access, for research and other purposes, via 'data hubs' for data aggregation and enrichment. The landscape of data access possibilities for science and research purposes and the infrastructure in and through which data access is enabled is extremely varied. This heterogeneity is further increasing as the overarching perspective across sectors continues to be ignored.

This paper is aimed at studying possibilities for data access and the corresponding infrastructure under both German law and select foreign legal regimes in order to identify and further develop best practices for ultimately structuring functioning data access ecosystems. Data access infrastructure and data access rights have to be considered simultaneously, for the most watertight data access right is ultimately pointless without sufficient infrastructure enabling the data access. And optimal data access

[8] https://www.konsortswd.de/konsortswd/tasks/datenzugang/, last accessed 7/16/2021.
[9] *Specht-Riemenschneider*, Data access rights - A comparative perspective, in: Bundesministerium der Justiz und für Verbraucherschutz/Max-Planck-Institut für Innovation und Wettbewerb, Data Access, Consumer Interests and Public Welfare, 2021, p. 402 ff.
[10] For a comprehensive discussion see *Podszun*, Handwerk in der digitalen Ökonomie, p. 76 ff.

infrastructure is conversely of little use without legal rights to data access. In addition, arising conflicts over the data access infrastructure between the respective rights and interests of parties entitled to data access (as a matter of freedom of research), of parties obligated to allow access to data (whose protected trade secrets and database protection rights in particular may be affected) and third parties (whose rights to informational self-determination may be affected by data access) could be resolved, for example, by making data exclusively accessible within secure environments maintained by a data trustee, or by anonymizing/pseudonymizing data before allowing access. Specifics regarding the means of providing data access and the handling of the relevant accessed data (subsequent usage rules) can also be structured so as to reduce risks.

These data access ecosystems will be developed in sector-specific fashion, allowing them to be optimally tailored to the rights and interests of the parties affected by data access within a given sector. The sectors under consideration are healthcare, internet (online platforms principally), mobility and energy, as specified by the client commissioning the study. The relevant laws of Finland, Canada, Australia, France, the UK and India are looked at here in view of the noteworthy advances these countries have made in ensuring data access for research purposes.

Australia's My Health Records Act (2012), the Canadian province of British Columbia's General Directive – Access to Health Data for Research (2018), Finland's Act on the Secondary Use of Health and Social Data (2019) and France's Code de la santé publique (2018) have established best practices in the healthcare sector. The UK guarantees data access for research purposes via 'Research Data Hubs'.

The Ministry of Transport of New South Wales, Australia has set up a platform for the mobility sector (Transport for NSW – Open Data Hub), making government and third-party data accessible on a contractual basis. Scientific organizations are eligible as contracting parties for explicit purposes of "policy research".[11] Regarding the mobility sector, the main focus is thus on Australia as a source of best practices.

The UK has a similar platform in planning for the energy sector. Siemens is developing "YODA" as a national platform for accessing energy data (Your Online Digital Architecture) based on recommendations outlined by an "Energy Data Task Force"[12]. Such platforms may grant privileged data access for research purposes. The UK is thus the leader regarding best practices for the energy sector.

Existing data access rights are more limited regarding the mobility, energy and internet sectors, thus the approach is to focus on healthcare first before addressing the other three sectors. Best practices

[11] https://opendata.transport.nsw.gov.au/open-data, last accessed 7/16/2021.
[12] https://www.gov.uk/government/groups/energy-data-taskforce, last accessed 7/16/2021.

should be adopted into national and/or European law based on non-sector-specific insights, which is to say that research data access structures existing in other countries for one sector, such as healthcare, may be borrowed in appropriately adapted form for enacting national/European laws for the energy sector, as an example.

The need for research data access regulation is being debated for B2B, G2B and C2B business models, to be addressed in the draft Data Act announced for the fourth quarter of 2021,[13] but the mandatory data access rules under discussion only cover part of the data access regime that is called for. The data access ecosystems outlined in this study can still be usefully referenced when developing infrastructure to accommodate data access rights of other players, including competitors and state actors, for example. However, data access rights will of course have to be worked out specifically for all respective actors in view of the applicable particulars.

[13] Inception Impact Assessment dtd 5/28/2021 on the planned Data Act of the EU Commission, available at: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13045-Data-Act-&-amended-rules-on-the-legal-protection-of-databases_en, last accessed 7/16/2021.

III. Study findings

1. Data exists in four states with regard to accessibility: open, public, shared and closed. Public administration data have to undergo a status change from closed to open or public (a process already underway); private data have to undergo a status change from closed to shared.

2. Data access standards should be defined based on the 'Five Safes' model, a data access classification framework of five interrelated risk dimensions which need to be factored in when drafting data access legislation.

3. The 'FAIR' principles furthermore should be respected with regard to data access. This means data should be Findable, Accessible, Interoperable and Reusable.

4. Regarding constitutionality, data access rights and the legal rights of the data-collecting parties must be structured so as to take account of the constitutional rights of third parties and the constitutionally guaranteed interests of the parties seeking access. These differing interests must be appropriately weighed and balanced. For parties who are obliged to grant access, the protection of trade secrets, as anchored in Art. 12 or Art. 14 of the German Constitution (GG) depending on one's view, or in Art. 17 of the Charter of Fundamental Rights of the European Union (CFR) or Art. 6, 15, 16 CFR, again depending on one's view,[14] is a primary issue to be considered, as is the protection of intellectual property legally ensconced in Art. 14 GG and Art. 17 CFR, including for example database copyright per § 2 of the Copyright Act (UrhG) and the sui generis rights per § 87a UrhG as well as freedom of profession in general. The right to informational self-determination per Art. 2 para. 1, Art. 1 para. 1 GG and/or Art. 7 and 8 CFR of the third parties affected by the accessing of data must be respected on one hand, and the freedoms of research and inquiry accruing to parties seeking access under Art. 5 para. 3 and/or Art. 13 CFR must be respected on the other hand.

5. Enabling research and scientific inquiry represents a legitimate purpose for impinging the legal rights accruing to parties who are obliged to grant access to data. Legislators have traditionally enjoyed very broad leeway in assessing suitability, the limits of which are perhaps not exceeded if data access for research purposes is granted.

---

[14] *Wollenschläger*, in: von der Groeben/Schwarze/Hatje, Europäisches Unionsrecht, Art. 17 GRC mgn. no. 16; *Wollenschläger*, in: von der Groeben/Schwarze/Hatje, Europäisches Unionsrecht, Art. 16 GRC mgn no. 8; *Aplin*, Right to Property and Trade Secrets, in: Geiger, Research Handbook on Human Rights and Intellectual Property, 2015, pp. 421-437; *Breuer*, Staatliche Berufsregelung und Wirtschaftslenkung, in Isensee/Kirchhof, Handbuch des Staatsrechts: Band VIII, Grundrechte: Wirtschaft, Verfahren, Gleichheit, 3. Auflage, 2010, § 171 mgn. no. 38.

6. The EU's legislative competence regarding research data access follows from its internal market competence per Art. 114 TFEU; the federal government's competence follows from Art. 74 para. 1 no. 13 in conjunction with Art. 72 para. 2 GG.

7. Under data protection law, data access for research purposes is allowed *de lege lata*, though in some cases clarification would be desirable to afford greater legal certainty.

8. Data access for research purposes irrespective of existing contractual relationships, theoretically possible antitrust claims and accessibility for individual researchers under data protection law can be achieved via these five structurally different regulatory instruments, which evidence differing degrees of effectiveness:

- A direct constitutional access right to data for research purposes
- Proper 'research clauses'
- Open data laws
- Transparency regulations with reporting requirements
- Permissions for granting data access

9. A direct constitutional data access right for research purposes is rejected *de lege lata* by majority view. A right to demand access to new sources of information does not follow from either freedom of information or freedom of research.

10. Open data laws are developing favorably; the widening of the scope of § 12a of the E-Government Act (EGovG) is applauded. Rights to access specific data of government agencies arise from the Freedom of Information Act (IFG), the Environmental Information Act (UIG) and the Consumer Information Act (VIG) as well as the corresponding state laws. Rights hereunder are awarded to everybody, including to research entities. § 12a EGovG, however, does not contain a right of its own.

11. Transparency and reporting obligations are insufficient for purposes of science and research, as data access permits only serve to provide a legal basis under data protection law for the granting of access; they do not establish a right to access data for research purposes.

12. Data access for conducting research in the sectors concerned in this study could be most effectively guaranteed by anchoring bona fide 'research clauses' in the law, which is to say recognizing subjective rights to data access. A distinction must be made between research clauses under private and public law respectively, i.e. clauses pertinent to a private-sector organization on the one hand and a public-sector entity on the other as the party whose data is accessed.

13. If access to data held by private organizations is afforded by way of a subjective right to data access, this research clause can be either derived or original. The clause is derived if it is based on existing data

access rights, i.e. if researchers are to enjoy the same privileges as those who currently have a right to data access, as for example under § 19 (3) UrhG. A subjective right to data access is original if it is introduced for research purposes specifically, not following the example of data access rights of other parties, as for example under § 5a NetzDG. An original 'research clause' is also found in Art. 31 DSA-E.

14. In bona fide research clauses granting access to data held by public-sector entities, a distinction must furthermore be made between bound decisions and rights to a decision free of discretionary error, and freedom of research must be taken into account in such a discretionary decision.

15. The following *de lege lata* situation can be seen looking at existing research clauses in national law:

a) Research clauses exist *de lege lata* in the internet, healthcare and mobility sectors, but not in the energy sector. There are also research clauses regarding archiving/administration, but these are not a focus in this study which concerns research data access in the healthcare, internet, mobility and energy sectors.

b) Existing data access rights have a very narrow scope of application all in all, which is subject to two specifications, regarding both the data recorded and the data holders required to grant access. There are no general rights to data access for research and scientific purposes across sectors; there are only a few, highly limited sector-specific data access rights.

c) Where data access rights concern data held by public-sector entities, parties required to grant access are more likely to have discretion, in the healthcare sector at any rate, if the data access poses a greater risk to the data subjects' right to informational self-determination. The more stringent the protective measures implemented, via anonymization solutions, for example, the more probable it is that a bound decision ought to be made. From a research point of view, such a bound decision is preferable. The precautionary protections should thus be stringent enough to serve as a justification for a bound decision. In the mobility sector, however, there is discretion with regard to the accessing of non-personal data as well. This is due to potential conflicts with other legal rights, such as the protection of trade secrets. Where data access rights concern data held by private-sector organizations, the protection of the rights and interests of third parties is principally ensured through the limitations for data access found in the legal provisions which grant the respective right.

d) The commercial character of scientific research does not exclude data access rights under all provisions analyzed.

e) Many research clauses (though not all) require research which is in the public interest, or in some cases, research which benefits the common good. This is because the term 'common good' has a dogmatic value relating to the protection of fundamental rights: It is by definition a matter of especially serious public interest, which makes it highly suitable grounds for allowing encroachment upon the

fundamental rights of parties obligated to provide research data access. The more the common good can be advanced through commercial research as well, the more easily the latter can be construed as falling within the scope of a data access right in research clauses.

f) Additional application requirements are only specified in a few research clauses. Where there are such additional requirements, however, such as in the form of a protection concept to be submitted with an application filing, these are aimed at protecting the fundamental rights and freedoms of third parties or the parties obliged to grant access.

g) The requirements which must be met by a protection concept to be submitted with the application filing are more stringent if data access poses a greater risk to the fundamental rights of the parties who must grant access or of third parties. Protection concepts are therefore absolutely recommendable for protecting the interests of the research community. For the lower the risk to the rights and interests of parties affected by the data access, the wider the scope of a data access right in accordance with the Five Safes model may be.

h) Unlimited data access for science and research purposes is not guaranteed *de lege lata* in a majority of research clauses, as restrictive provisions apply. This, too, represents a balancing of the fundamental rights accruing variously to entitled parties, obligated parties and third parties.

i) Limiting provisions of research clauses are less stringent the narrower the scope of the research clause is and the more extensive the requirements are for the protection concept to be submitted with the application filing. This is because the fundamental rights of the parties obliged to grant access and of third parties are already largely taken into account by way of the defined scope, application structuring and protection concept, enabling less stringent limiting provisions.

j) Several research clauses tie data access to the purpose of (scientific) research. Tying data access to purpose means the data can initially only be used for the specified purposes, e.g. to conduct scientific research projects. Alternatively or in parallel there may be a follow-up usage requirements pertaining to both the original data (e.g. anonymization after research project completion; follow-up use for differing purposes) and to the research findings.

k) Further data access requirements are stated in some research clauses in relation to the criterion of necessity, with ‘special justification’ being called for in some cases as limitation, while other research clauses provide for a review by a scientific committee to decide regarding data access. All these additional requirements have in common that they restrict data access for research purposes in the interest of balancing conflicting legal rights by providing for further protecting ‘safeguards’. Such safeguards may be enacted in substantive law, by way of the criterion of necessity, or formally, by way of a board review and decision.

l) In the healthcare sector there are comprehensive provisions governing compensation for data access in the Data Transparency Fee Regulation, which applies for data access granted in line with §§ 303a ff. SGB V.

m) In the internet sector, fee regulations exist, in national law at any rate, providing for a right to appropriate reimbursement of costs incurred in connection with providing access to data. These are to ensure appropriate compensation for accrued expenses in relation to data access. Reimbursement claims can effectively render accessing data unattractive for research purposes, creating a *de facto* barrier for data access. This is why § 5a NetzDG provides that such costs may not pose a major obstacle to exercising a data access right. § 287 (1) ZPO applies regarding cost calculation. A cost cap of 5,000 Euros furthermore applies.

n) A deadline for enabling data access is generally not specified in the research clauses studied. Only the DSA-E features a requirement that data access must be granted within a 'reasonable period'. General principles may also be referred to for access provisions for the public sector which provide for a discretionary decision, according to which the proper exercise of such discretion implies that access must be granted within a reasonable period of time. Data access requests can vary greatly regarding their scope and urgency, thus it is not likely that standardized deadlines will become ensconced in law. On the other hand, researchers need to have some idea of the point in time at which they can file suit without the risk of having to bear costs if a right is recognized in the meanwhile. If researchers are not to be required to set a deadline in every individual case, it will be unavoidable to have deadlines that can vary but are specific. Using phrasing like 'without delay' is a possibility with the simultaneous definition of a standardized maximum time limit. It must be ensured, however, that parties obligated to provide data access can extend the deadline depending on the nature and scope of the requested data access so as to avoid an undue burden.

o) It must be noted that the placement of the burden of proof is subject to the general principles governing the burden of proof.

p) Enforcement depends on the relevant jurisdiction. Whether a civil court or an administrative court is competent ought to be determined by application of the general principles. There are only special considerations if the party obliged to provide access is a private individual but data access can only be requested via a third party such as a 'digital services coordinator'. If this third party fails to respond, the requestor may take legal action against them via administrative procedure. The administrative court is always competent if either this is specifically required by law or the requirements of the general clause in § 40 (1) sentence 1 half sentence 1 of the Administrative Courts Ordinance (VwGO) are met and the dispute is not specifically required to be adjudicated by another court.

16. The following must be noted regarding the granting of research data access in the healthcare sector within the legal systems of India, France, Canada, Finland, Australia and the UK:

a) The proposed regulatory framework for non-personal data in India is highly innovative, but as the examples cited in the independent report demonstrate, the framework functions primarily to ensure data access by the government for purposes aligned with the public welfare. The data accessible for scientific research should not and indeed cannot be limited to data which has been declared of relevance to the public interest as a 'high-value dataset' (HVD), typically by a government authority. Rather, the research community requires access to data free of predefined limits. Whether the research itself is then in the public interest is a separate question to be examined. A central intermediate instance serving as a contact for researchers to arrange access to data collected by third parties can facilitate the data access process for research purposes. It is not required, however, that the data be centrally held by that intermediary. Rather, the intermediary could arrange access to data which firms have stored locally, saving the latter time and resources connected with data access requests.

Regarding research data access in India, the effective interposing of a data transmission intermediary is one important element, but others include the fact that parties at whom data access rights are directed can be private-sector entities, and that data access rights are non-sector-specific, being designed horizontally.

b) In France, a commission of experts formed in 2018 recommended that data rooms be set up and made usable by and accessible to as many parties as possible.

In December 2019, an initial iteration of this 'health data hub' was unveiled, which commenced real-time operation in April 2020. The hub is fed data from the national healthcare data system, including health insurance, preventive check-up, maternity and child protection-related data, among other types, thus representing an aggregate of health data collected by the state which is held by a central administrative instance. An ethics committee which was formed under an order entitled "Arrêté du 26 mai 2020 portant nomination des membres du Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé" decides on data access. The committee reviews whether a given research project makes a relevant contribution toward public interests. The use of data for commercial marketing purposes is prohibited, and the data may not be sold.[15] The accessing of personal data within the national health data system can only be approved pursuant to L 1461-3 Code de la Santé Publique, for purposes in line with L-1461-1. Under L 1461-1 para. 2 no. 2, these purposes include conducting research, studies and evaluations as well as pursuing innovation in the

[15] *Stuwe*, Presentation: Health Data Hub - Overview, strategy and lessons learned, 2020, available at: http://ehaction.eu/wp-content/uploads/2020/10/D3S2_HDH-Louisa_Stuwe-new_version.pdf, last accessed 7/16/2021.

areas of healthcare and medical social caregiving. The language employed in L 1461-1 and L1461-3 indicate that this represents a bona fide research clause in the sense of a data access right. Access to the data specified under 1461-2 Code de la Santé Publique has been free of charge. Access to health data other than as specified under Article L. 1461-2 is likewise free of charge for research conducted exclusively for public administration purposes. A further reflection process to be concluded in 2022 is focusing on the development of business models and financing possibilities around access to data from the Health Data Hub. A ministerial decree effective since April 21 has allowed the hub to obtain[16] and subsequently provide a broad spectrum of health data for research purposes in the public interest.

c) Canada has a large number of regulations in place enabling government entities to guarantee access to data for scientific purposes. All these data access provisions appear to only concern the permission to data access, however, and therefore fail to guarantee any rights to such data access.[17] This is evident by looking at the wording of healthcare regulations as compared to regulations in the administrative sector where there is an unquestionable right to data access, for example under Art. Chapter IV Division 2 125 of the Act Respecting Access to Documents Held by Public Bodies and the Protection of PERSONAL INFORMATION [18].[19]

An overview of these data access permits is found in the paper: "Expert Panels on Timely Access to Health and Social Data for Health Research and Health System Innovation". It is striking that all data may be used "for approved data purposes", the decision on use generally being made by a Research Ethics Board (REB) or other specially appointed body. Data usage agreements furthermore have to be concluded between the data custodian and the researcher/research entity. The ministry has a "model research agreement" for this. Requirements for data access vary greatly, furthermore. Eligibility to file an application is generally limited, however, to specific researchers within Canada.

d) Bona fide research clauses in the sense of a data access right for research purposes are recognized under Finland's laws for the healthcare sector, including the Biobank Act and the Act on Secondary Use of Health and Social Data particular in (hereinafter the 'Secondary Use Act'). 'Secondary use', in contrast to primary use, means using data for specific purposes other than the purposes for which they

[16] Arrêté du 21 avril 2020 complétant l'arrêté du 23 mars 2020 prescrivant les mesures d'organisation et de fonctionnement du système de santé nécessaires pour faire face à l'épidémie de covid-19 dans le cadre de l'état d'urgence sanitaire, available at:
https://www.legifrance.gouv.fr/download/file/JQCAhy2BjS_uSuRmKba4o_yPpUVXDsxSS7PEreByYJg=/JOE_TEXTE, last accessed 7/16/2021.
[17] See Office of the Information & Privacy Commissioner for British Columbia, Access to Data for Health Research, 2018, p. 10, available at https://www.oipc.bc.ca/guidance-documents/2115, last accessed 7/16/2021.
[18] Act Respecting Access to Documents held by Public Bodies and the Protection of Personal Information, available athttp://legisquebec.gouv.qc.ca/en/pdf/cs/A-2.1.pdf, last accessed 7/16/2021.
[19] Further data access rights for the administrative sector are found in the Access to Information Act and in Section 5 of the Statistics Act, in conjunction with the Statistics Canada Policy on the Use of Administrative Data Obtained under the Statistics Act.

were collected; see Art. 3 Secondary Use Act[20]. Research is one such privileged purpose constituting an allowable secondary use. Article 38 Secondary Use Act provides solely that *"a data permit may be granted"*, although para. 2 requires that freedom of research be given due consideration regarding the granting of permits. This could be understood as grounds for a data access right in cases where a data access right does not conflict with the rights and interests of the party obligated to provide access or of third parties (e.g. for non-disclosure reasons) and data protection is adequately safeguarded through precautionary measures. Data are gathered from public entities, such as national data repositories and healthcare and social welfare care data archives. Registered data are also gathered from private providers of social and healthcare services. Data access is granted either by the data-holding authority itself (e.g. the data repository) or by a new authority, namely Findata (the Data Permit Authority), which is operated by the Finnish Institute for Health and Welfare but otherwise independent from the institute's activities. Findata is overseen by the Ministry of Social Affairs. When Findata grants a data access permit, Findata collects the data from the data-holding entities and compiles and pseudonymizes or anonymizes the data as necessary (referred to as data 'preparation') before making it available to the applicant via a secure hosting service specifically set up for this purpose. Access may only be granted to data made available on the basis of consent under data protection law within the scope of such consent.

e) In Australia, access to research data is guaranteed under the My Health Records Act. The My Health Record System is a state-run system for providing patient health data for 'primary use' purposes (providing them healthcare services) and 'secondary use' purposes, such as science and research. Recipients of healthcare services have a personal healthcare record on file in this system which is created either when the patient registers accordingly or, where the opt-out model is in place by ministry order, where they have not opted-out. The system operator runs the National Repositories Service where the most important data in the patient's healthcare record are stored. Other data records are kept on electronic file by registered repository operators. In aggregate these records constitute the patient's personal health file. For this data to be accessed, the Data Governance Board, which is comprised of various experts and obtains advice from various bodies, must approve a request to use the data for research purposes. The requestor has to agree to the terms of use[21] and attach a risk management plan to the filing based on which the Board assesses the risk of data loss or improper use, among other factors.[22] The data subject's consent is always required if personal data are to be

[20] *Lilja*, Secondary use of health data – the new Finnish Act, 2019, available at: https://www.roschier.com/newsroom/secondary-use-of-health-data-the-new-finnish-act/, last accessed 7/16/2021.
[21] *Australian Government - Department of Health*, Framework to guide the secondary use of My Health Record system data, 2018, p. 31, available at: https://www1.health.gov.au/internet/main/publishing.nsf/content/F98C37D22E65A79BCA2582820006F1CF/$File/MHR_2nd_Use_Framework_2018_ACC_AW3.pdf, last accessed 7/16/2021.
[22] *Australian Government - Department of Health*, Framework to guide the secondary use of My Health Record system data, 2018, p. 47, available at:

accessed. Data are not forwarded to insurers. In deciding on data access requests, the Board applies the 'safe people principle', evaluating the knowledge, competencies and incentives of the requestor regarding the appropriate storage and use of data. The Board is not an ethics board.

e) The UK has a system consisting of seven Research Data Hubs[23] set up in October 2019 by an independent, registered non-profit organization called Health Data Research UK. There is no right to data access, as the decision to grant data access is made by a committee applying the criteria adopted by that committee. If personal data are to be made available, the consent of the data subject is required.[24] The exemplary data hub 'Insight' was set up by the Data Trust Advisory Board (Data TAB) based on various practical access criteria that afford the possibility for appropriate review and are at the same time practical, efficient and scalable. The Data TAB decides who is to receive access to data for what purposes.

17. In the mobility sector there are no comprehensive regulations of model character. The Australian state of New South Wales utilizes an Open Data Transport Hub designed to enable data-driven innovation in the mobility sector and 'policy research', which is accessible by anyone on contractual terms.

18. Research data access is also underdeveloped in the energy sector. In the UK, only one intermediary platform is being developed for accessing energy data, which is done on contractual terms. It has not been decided what specific data may be involved.

19. For the internet sector, there are no regulations governing research data access. National laws can provide orientation for drafting bona fide research clauses which appropriately balance rights and interests.

20. Data access rights have to be designed in parallel with data access infrastructures to ensure that the rights are sufficiently effective. Research data centers and data trustees are important elements in these data access infrastructures, which include personal information management systems (PIMS). Data access infrastructures and rights together comprise a data access ecosystem.

21. National and international research clauses can be analyzed to devise sector-specific guidelines for research data access in order to build a larger research data access ecosystem in the healthcare sector and ensconce health research data access in law. For the healthcare sector, a heterogeneous system

https://www1.health.gov.au/internet/main/publishing.nsf/content/F98C37D22E65A79BCA2582820006F1CF/$File/MHR_2nd_Use_Framework_2018_ACC_AW3.pdf, last accessed 7/16/2021.

[23] Improving UK Health Data - Impacts from Health Data Research Hubs, 2021, available at: https://www.hdruk.ac.uk/wp-content/uploads/2021/04/Improving-UK-Health-Data-Impacts-from-Health-Data-Research-Hubs-v2.pdf, last accessed 7/16/2021.

[24] Improving UK Health Data - Impacts from Health Data Research Hubs, 2021, available at: https://www.hdruk.ac.uk/wp-content/uploads/2021/04/Improving-UK-Health-Data-Impacts-from-Health-Data-Research-Hubs-v2.pdf, last accessed 7/16/2021.

of original research clauses is recommended with central data stores (e.g. pre-existing central registers like the Federal Cancer Register), decentral/central data stores (e.g. regionally distributed registers like the state cancer registers—organization on the state level not being necessary) and fully decentral data storage, by healthcare service providers, for example. Data from private healthcare service providers could likewise be fed into this system of central, decentral/central and entirely decentral data storage. Data donation should also be possible, via PIMS in particular. Electronic patient records could be managed as a PIMS. In designing regulations governing data access rights, the general norm should be to have a narrowly defined purpose tied to research for the common good. For if so, the scope of parties authorized to access data does not need to be limited to entities conducting non-commercial research. Insurance companies should be excluded from such rights. Data access is not to be contingent upon 'necessity'. Presentation of a protection concept should be required, however, irrespective of whether data is anonymized or transmitted in the form of personal data in order to appropriately balance conflicting rights and interests. A standardized data access request procedure should be in place (ideally uniform internationally) to effectively guarantee data access. It is also advisable to borrow the structure seen in some foreign legal systems (France, Canada, Finland, Australia) of having an authority involved which decides questions of ethics in research. An additional requirement for request filing should be receiving approval by such a Research Ethics Board (REB). In Canada, this authority decides whether under a given request the data concerned will be used 'for approved data purposes'. In Australia, the Data Governance Board performs this function. Separation between the two instances seems to be recommendable, as ethical issues connected with research projects should primarily be decided by individuals versed in ethical philosophy, while questions of whether research serves what is seen as the public interest is another issue—fulfillment of the conditions for data access rights to accrue under substantive law being of course a legal issue. Subsequent data usages must be clearly defined. The data concerned must be pseudonymized, anonymized or erased at the earliest possible time. Sale of the data and use for commercial advertising purposes should furthermore be prohibited. A complete ban on further forwarding of the data (sharing/disclosure/transfer etc.) could be considered. This would reduce the risk of improper use, thus data access rights for research purposes under the Five Safes model would likely be broader than if forwarding of research data were allowed. The de-anonymizing of anonymized data should be barred by law as a criminal offense.[25] Anonymization standards should be defined at the same time to afford legal clarity and fulfill the principle of specificity. Data access may not unduly impinge the rights or interests of third parties. This should be implemented as a limitation provision following the example

[25] *Specht*, Das Verhältnis möglicher Datenrechte zum Datenschutzrecht, GRUR Int. 2017, 1040, 1046-1047; Gutachten der Datenethikkommission, 2019, p. 132, available at: https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf?__blob=publicationFile&v=6, last accessed 7/16/2021.

set in Art. 15 GDPR. The scope of compensation provisions should be limited to cover administrative costs only. The Data Transparency Fee Regulation should be referenced for orientation in this regard. Deadline provisions for the administrative area should always adequately take available human and other resources into account. It must also be ensured that the provision of data is properly reviewed. Flexible deadline provisions involving a final deadline of maximum length are thus preferable to rigid provisions which would not or not adequately accommodate the amount of work required in a given case. It is recommendable to require data permit decisions to be made without delay, with reference to Finland's Secondary Use Act, and in no case later than three (3) months from the date of receipt of the complete request application by the authority. Data access decisions formally represent an administrative act (*Verwaltungsakt*). The appropriate path of legal recourse in the event of the request's rejection is thus to file for a judgment of performance (*Verpflichtungsklage*) for an approval by administrative act. The principles governing the burden of proof apply in principle, but research should be deemed to be in the public interest if it is conducted by public research institutions and the research findings (anonymized) are made publicly available. In parallel to a system comprised of central, decentral-central and fully decentralized research data storage, flexible data trust structures should be provided for and a secure legal basis created for these.

22. A model research data access clause for the internet sector can be arrived at by studying national and international research clauses, to be ensconced in law in a number of identified laws as both derived and original research clauses. In the internet sector, access to data held by private and public-sector entities should be guaranteed through indirect data access structures existing within state organizations. Specifically, digital service coordinators per Art. 31 DSE-E or similar instances with decision-making authority regarding data access should make the decision, relieving private organizations of such responsibility. There does not have to be any restriction of data access rights to specific research projects. If no such restrictions are implemented, asymmetrical regulation is recommended, i.e. regulation which is directed only at private-sector companies of a certain size so as to avoid an undue economic burden on other companies with data access obligations. Data access requirements can be mostly similar to those existing for the healthcare sector, though in contrast, private organizations in the internet sector are obligated to provide data in response to a legitimate data access request. This may involve considerable effort, requiring compensation. Yet requiring excessive compensation can significantly impair the data access right in practice. Thus pursuant to § 5a NetzDG compensation is to be capped at 5,000 Euros, the appropriateness of the amount otherwise being up to the courts in line with § 287 ZPO. This appears to be the only feasible approach. The specific amount of the cap should be set on a case-by-case basis, ideally based on evidence. Both researchers and parties required to provide access must have options for objecting to the digital service coordinator's decision. If a party obligated to provide data access refuses such access, the entitled

party should be able to take direct action against the former without having to await further action by the digital service coordinator. If the digital service coordinator rejects a request by an entitled party, the latter can file for a judgment of performance with an administrative court.

23. National and international best-practice regulations are lacking in the mobility and energy sectors, thus a commission of experts should first be formed like in the UK to develop recommendations, principally regarding the data concerned in data access rights and the corresponding data access infrastructure. The various approaches being taken in the mobility sector of using mobility data rooms and platforms should be considered as a whole.

XIII. Policy recommendations

This study yields 15 summary recommendations for policy action:

1. Data access standards should be defined based on the 'Five Safes' model, a data access classification framework of five risk dimensions which need to be factored in when drafting data access legislation.

2. The 'FAIR' principles furthermore should be respected with regard to data access. Data should be Findable, Accessible, Interoperable and Reusable.

3. It is recommended to draft bona fide research clauses providing for subjective rights to research data access in all of the sectors concerned here, while at the same time the recommendations adopted vary in scope depending on the sector.

4. In the healthcare sector, an overarching research data access ecosystem should be established and anchored in a data access law for healthcare research. The anchoring of original research clauses is recommended for the healthcare sector and the creation of a system comprised of central data stores (e.g. pre-existing central registers like the Federal Cancer Register), decentral/central data stores (e.g. multiple regionally distributed registers holding similar data, like the state cancer registers—organization on the state level not being necessary) and fully decentral data storage, by healthcare service providers, for example. This system of centralized, hybrid and fully decentralized data storage, modeled after structures in Finland, Australia and the UK, can provide for the usage of both publicly held data and privately held data of specific types. Data donation should also be possible, via PIMS in particular.

5. In designing regulations governing data access rights, the general norm should be to have a narrowly defined purpose tied to research for the common good. For if so, the scope of parties authorized to access data does need not to be limited to entities conducting non-commercial research. Insurers should be excluded from access rights, as is done in Australia. Data access is not to be contingent upon 'necessity'. To balance conflicting rights and interests, it should be required, however, to submit a protection concept for data access like that provided for in the Network Enforcement Act, irrespective of whether the access right concerns anonymized or personal data.

6. A standardized data access request procedure should be in place (ideally uniform internationally) to effectively guarantee data access. It is furthermore recommended to borrow the idea seen in various foreign legal systems (France, Canada, Finland, Australia) of having one instance (Research Ethics Board) responsible for deciding questions of research ethics and another instance (Data Governance Board) responsible for deciding whether the conditions for data access in substantive law have been met and how far such access should go.

7. Follow-up uses of data must be clearly defined. The data concerned must be pseudonymized, anonymized or erased at the earliest possible time. The sale of the data and their use for commercial advertising purposes should furthermore be prohibited. A complete ban on further data sharing/transfer could be considered. This would reduce the risk of improper use, thus data access rights for research purposes under the 'Five Safes' model would likely be broader than if forwarding of research data were allowed. The de-anonymizing of anonymized data should be barred by law as a criminal offense. Anonymization standards should be defined at the same time to afford legal clarity and fulfill the principle of specificity.

8. Data access may not unduly impinge the rights or interests of third parties. This is to be implemented via limitation provisions. The scope of compensation provisions should be limited to cover administrative costs only. The Data Transparency Fee Regulation should be referenced for orientation in this regard. Deadline provisions for administration should always adequately take available human and other resources into account. It must also be ensured that the provision of data is properly reviewed. Flexible deadline provisions involving a final deadline of maximum length are thus preferable to rigid provisions which would not or not adequately accommodate for the amount of work required in a given case. It is recommendable to require data permit decisions to be made without delay, with reference to Finland's Secondary Use Act, and in no case later than three (3) months from the date of receipt of the complete request application by the authority.

9. Data access decisions formally represent an administrative act (*Verwaltungsakt*). The appropriate path of legal recourse in the event of a request's rejection is thus to file for a judgment of performance (*Verpflichtungsklage*) for an approval by administrative act. The principles governing the burden of proof apply in principle, but research should be deemed to be in the public interest if it is conducted by public research institutions and the research findings (anonymized) are made publicly available.

10. In parallel to a system comprised of central, decentral-central and fully decentralized research data storage, flexible data trust structures should be provided for and a secure legal basis created for these.

11. For the internet sector, data access rights for research purposes should be ensconced in a set of identified laws via derived and original research clauses, drawing upon the model research data access clauses proposed herein. Derived research clauses should, at a minimum, be anchored in the following:

- Art. 33a DSA
- § 20 para. 1a GWB
- § 19a GWB
- § 32e GWB
- Art. 17 DMA

Original research clauses should be anchored in the following, in particular:

- the AI Regulation of the EU Commission
- the Data Act of the EU Commission

Data access rights for sectors not discussed in this study similar to the model clause developed herein could be implemented via integration into many clauses which previously only provided for data access permission. These include in particular:

- § 476 StPO
- § 98 SGB XI
- § 119 SGB XII
- § 42a BZRG
- § 75 SGB X
- § 66 PStG
- § 14 para. 2 in conjunction with § 15 TPG
- § 24a para. 1 AZRG
- Landeskrankenhausgesetze (Hospital Acts of the German states)
- § 35 para. 7 HMG
- § 88a para. 4 AufenthaltsG

Provisions the reasoning of which is formulated to indicate that they are intended to enshrine data access rights but whose wording does not explicitly state this can already be interpreted to grant subjective data access rights. This applies for example to § 1g and § 63a para. 5 of the Road Traffic Act (StVG). Amending the wording would be desirable here, too, for clarification.

12. In the internet sector, access to data held by private and public-sector entities should be guaranteed through indirect data access structures existing within state organizations. Specifically, digital service coordinators per Art. 31 DSE-E or similar instances with decision-making authority regarding data access should make the decision, relieving private organizations of such responsibility.

13. There does not have to be any restriction of data access rights to specific research projects. If no such restrictions are implemented, asymmetrical regulation is recommended, i.e. regulation which is directed only at private-sector companies of a certain size so as to avoid an undue economic burden on other companies with data access obligations. Data access requirements can otherwise be mostly similar to those existing for the healthcare sector. In the internet sector, in contrast to healthcare, private organizations are obligated to provide data in response to a legitimate data access request. This may involve considerable effort, requiring compensation. Yet requiring excessive compensation can significantly impair the data access right in practice. The decision on the specific compensation

figure for data access should thus be evidence-based, avoiding an undue burden impairing the research data access right.

14. Both researchers and parties required to provide access must have options for objecting to the digital service coordinator's decision. If a party obligated to provide data access refuses access, the entitled party should be able to take direct action against the former without having to await further action by the digital service coordinator. If the digital service coordinator rejects a request by an entitled party, the latter can file suit for a judgment of performance with an administrative court.

15. National and international best-practice regulations are lacking in the mobility and energy sectors, thus a commission of experts should first be formed to develop recommendations, principally regarding the data concerned in data access rights and the corresponding data access infrastructure. The various approaches being taken in the mobility sector of using mobility data rooms and platforms should be considered as a whole.