April 2020

Statement of the German Council for Scientific Information Infrastructures (RfII)

Designing Trusted Intermediaries for Data Exchange – On Experiences from Science

The Council for Scientific Information Infrastructures (RfII) was established in 2014 to provide advice to the German Federal Government, the *Länder* (state governments) and the academic communities on the development of modern and sustainable infrastructures for scientific information. RfII recommendations serve both to inform the debate and to support the development and communication of German positions in the wider international arena. In line with its mandate, the council has set up a working group to examine issues related to the establishment of trusted intermediaries to build confidence in the exchange of digital data in today's society: *Datentreuhandstellen* (data trustee agencies). In the following statement the council offers initial recommendations on how experiences and examples from within the scientific system could aid the current debate and policy formulation.

****

The German Federal Government is addressing the issue of *Datentreuhänderschaft* (data trusteeship) as part of the development of a comprehensive data strategy. A recent discussion paper by the government underlines that efforts are being made to "significantly increase the responsible provision and use of data by individuals and institutions".[1] New data monopolies should be prevented and "fair participation" should be ensured. The European Commission's data strategy also advocates the establishment of data pools or data spaces "in strategic sectors and domains of public interest". Such data spaces should facilitate both increased access to data for a wide range of actors and benefit society.[2]

The debate on *Datentreuhänderschaft* has centered on how data can be made available by neutral parties in a form that is to some extent withdrawn from competition – hence the reference to the concept of a trustee or fiduciary (explained in more detail below). This is primarily a non-scientific debate, but it is analogous to related negotiations in science. For example, the concept is based on the exchange and fair re-use of data, which can mean equitable access to data is guaranteed amongst actors. Such intermediaries are likely to encourage wider data pooling. This supports innovation policy, but also holds considerable potential for helping to address major societal challenges in areas such as health, climate, mobility or poverty prevention. Data trustees could be deployed where data are exchanged in competitive environments or in situations where a strong asymmetry of power or resources exists –

[1] Die Bundesregierung (2019): Eckpunkte einer Datenstrategie, p. 1.
[2] European Commission (2020): A European strategy for data. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Brussel, p. 21.

for example between global internet companies, start-ups, researchers or individual consumers.

Whilst the concept of *Datentreuhänderschaft* is drawing increased attention in Germany it is often discussed in relative vague terms. There is currently no public or political consensus on how data trustees or data trustee agencies should be created.[3]

The RfII has taken the issues involved in building broad trust in data exchange into consideration. The basic principle of neutral parties providing sustainable data services offering users fair terms and conditions is central for facilitating access to or re-use of newly created data sets. However, in broader terms, data trustees could also become part of established data infrastructures for society at large.

A range of issues must be taken into account in designing a trustee role, which would both guarantee trust in the handling of personal or other sensitive data and ensure their secure, long-term storage. In examining this issue, the RfII has built on an exchange of experience within scientific fields to draw lessons from existing and established models. Many scientific communities have a long tradition of sharing quality-assured data and have developed regulations and norms to safeguard the rights of data providers and users. Outside of its own frameworks, organisations, and experience, science also has specific needs and expectations of external agencies tasked with acting as data trustees.[4]

The RfII is an advisory body that deals with the establishment and future development of information infrastructures in science. Against this backdrop, it has defined the concept of data trusteeship and formulated recommendations regarding their wider role in civil society and the economy as well as their function in research and development.

EXPLANATION OF TERMS

General considerations on the concept of trusteeship

The idea of a fiduciary assumption of responsibility, or duties, has a long standing tradition in Germany. It is briefly outlined here with the understanding that there is some leeway for adaptation. In legal terms, a trusteeship is a legal relationship between a fiduciary/trustee and a settlor. The settlor transfers or grants the trustee certain rights with regard to the trust property, which the trustee may exercise externally to the relationship, in their own name, or in the name of a third party. In their relationship with the settlor, however, the trustee remains bound by agreed terms, conditions and limitations. A trust agreement is not regulated in de-

[3] In its 2019 report, the Commission "Competition Law 4.0", appointed by the Federal Ministry of Economics and Energy, also initially advocated to simply examine the establishing of data trustees and to consider various models. Cf. BMWi - Bundesministerium für Wirtschaft und Energie (2019): Ein neuer Wettbewerbsrahmen für die Digitalwirtschaft. Bericht der Kommission Wettbewerbsrecht 4.0.

[4] Cf. RatSWD - Rat für Sozial- und Wirtschaftsdaten (2019): Big Data in den Sozial-, Verhaltens- und Wirtschaftswissenschaften: Datenzugang und Forschungsdatenmanagement, Berlin, p. 21ff. The RatSWD is already formulating a catalogue of tasks and criteria for the design of a data trustee agency in the economic and social sciences.

tail by law: Depending on the agreement underpinning the trust relationship, different legal regulations are applicable. A trust relationship can be set up either to safeguard the interests of the trustee (*eigennützige Treuhand* /"equitable trust"), or the interests of the settlor (*fremdnützige Treuhand* /"extraneous trust"). However, a trustee can also act in cases of a potential conflict of interests involving different persons ("double-sided trust"). In the latter, the fiduciary relationship serves to protect the interests of each party by involving a third party acting as an unselfish trustworthy agent.

In economics, the term is applied more broadly. For example, the safeguarding of foreign interests can be understood as a trusteeship. In the English-speaking context, a "trustee" generally means a reliable broker. In the German-speaking sense, the term *Datentreuhandstellen* is rooted in an open, rather pragmatic notion of the role of a trustee.[5]

Datentreuhänder (data trustee)

The term *Datentreuhänder* has emerged in recent German-speaking reflections on digitisation processes as a metaphor or analogy describing agencies which act as intermediaries between one or more data providers and a user demanding access to data. A number of examples of the use of this term in different applications already exist. In medical applications for example, a *Datentreuhänder* is defined as a "legally, spatially and personally autonomous and independent agency which is ideally subject to a special duty of confidentiality".[6] The term is also sometimes used in a broader sense. For example, the *Bundesdruckerei* defines a *Datentreuhänder* as "an independent trustworthy authority that conveys data between data providers and data users in a secure and legally compliant manner".[7] In the case of personal data, this may in particular mean that the trusted intermediary undertakes the pseudonymisation or anonymisation of data, and that data is only made available in pseudonymised, anonymised or aggregated form.

Within the scientific community, there are no fiduciary relationships in the strict legal sense. However, scientific institutions often take on the role of providing neutral, i.e. non-commercial and content-independent data archives. Examples are bio-banks or research data centres in the social and economic sciences.

Legal and data policy requirements for the creation and, where appropriate, codification of the role of one or more *Datentreuhänder* are a comparatively new concept in Germany. They are articulated, for example, in connection with large (and often in terms of their integrity vulnerable) data sets created through modern information technologies, through data har-

[5] In particular, the idea of data ownership is thus waived. Cf. the recommendation of the Data Ethics Commission of the German Federal Government; Datenethikkommission der Bundesregierung (2019): Gutachten der Datenethikkommission, Berlin, p. 18.

[6] Pommering, Klaus et al. (2014): Leitfaden zum Datenschutz in medizinischen Forschungsprojekten. Generische Lösungen der TMF 2.0 (Schriftenreihe der TMF, 11), Berlin, p. 209.

[7] https://www.bundesdruckerei.de/system/files/dokumente/pdf/BDR.de_Datentreuhaender.pdf (last accessed on: 03.04.2020). For the definition of the term „Datentreuhänder", see also RatSWD - Rat für Sozial- und Wirtschaftsdaten (2017): Handreichung Datenschutz (RatSWD Output, 5 (5)), Berlin, p. 16.

vesting and processing. Today, there is considerable interest in making medical, mobility or commercial usage data generated in everyday applications of the Internet of Things more accessible for further use, i.e. for third parties, social, or research purposes. In civil society, data trustee frameworks are also being discussed as an option to enable fairer and more individually empowered handling of consumer data (Personal Data Trusts, Personal Information Management Systems). Here, the focus is more on individual data sovereignty. In this specific context, the Data Ethics Commission of the German Federal Government has explicitly stated that the individual must be protected from "alleged advocates" primarily representing economic self-interests.[8]

RECOMMENDATIONS

From the RfII perspective, established scientific data access, analysis or re-use practices can in some circumstances serve as a model for the development of data infrastructures in other societal sub-sectors. The remote sensing data from satellite missions and large telescope arrays show how data can be shared in an efficient and equitable manner within a large community of heterogeneous stakeholders. This example also demonstrates that the long-term availability of data can be adequately ensured. Both aspects are essential for enabling the analysis of patterns within complex systems. Large-scale research, for example at CERN's Large Hadron Collider, with its highly complex and collaborative data archiving activities, is another example of how large volumes of data can be shared and analysed by researchers around the globe – based on guaranteed standards for the quality of these data. For centuries, the German library system, which is decentralised in terms of local administration but centrally coordinated with regard to its core activities, has evolved and adapted forms and requirements for metadata content. Research data centres in the social and economic sciences provide models for legally compliant access to and re-use of data that can be sensitive for reasons of corporate ownership or data privacy. Each of these examples from the scientific sphere show how the exchange of data, which is quality-assured by using uniform standards, can be fostered and fairly organised amongst the relevant actors.

Based on such experiences, the RfII envisages data trustee agencies in an institutional sense, namely as a particular type of data infrastructure.

Designing data infrastructures for science, economy and society in a balanced manner

The current discourse on intensifying the re-use of data and establishing models for data trusteeship provide potential markers for policy makers on which directions the numerous projects and initiatives in the field of data infrastructures should be developed strategically. The International Data Spaces Association is developing a reference model for platforms to facilitate the access, use and secure exchange of data between or with companies. The GAIA-X project aims to create an open "ecosystem" of trusted tools, services, storage and

[8] Datenethikkommission (2019): Gutachten der Datenethikkommission, p. 21.

computing capacities for companies and organisations throughout Europe. With a view to public welfare and innovation policy goals, the European Commission has proposed creating data pools or data spaces in nine specific sectors, including industry, environment and health. Public authorities around the world are compiling sources of public information and data, and making them available via portals such as the Transparency Register in Germany, GovData, or the European Data Portal. Taking into account the particularities of research activity, similar efforts have been on-going internationally to establish scientific data centres in recent years.

The character of business and consumer data may differ from that of scientific data sets, but what both have in common is the necessity to meet certain quality criteria. Criteria such as trustworthiness, legal conformity and transparency of access are of similar importance in both worlds. Here, the interests of commercial actors as well as those of scientists, politicians and social groups or individual citizens can be brought together.

The idea of a recognised system of data trustee agencies, which guarantees a basic principle of efficient, equitable and legally compliant use of quality-assured data collections, being consistent in terms of data sovereignty, is from the RfII perspective exceedingly useful for the development of a wider data strategy. The development of such data infrastructures should be promoted by a social framework or mandate, and if necessary supported by public funding. However, the areas in which trusteeship models can and should be applied need to be carefully examined. Specifically, it should be made explicitly clear whether such intermediary frameworks have a remit to merge data collections (with a focus on quality assurance if this is part of their mission), or rather to simply facilitate and coordinate the exchange of data between distributed resources and users.

Ensuring access for research and development

The RfII sees great potential in the concept of data trustee agencies. Such an approach could be useful in organising the exchange of confidential or sensitive data not only within sectors but also to help boost cooperation among competitors in industrial innovation processes, making available and merging health-related data, public administrative data, etc. In this sense, such infrastructures could act as an interface between science, industry and society that would augment the economic and public welfare-oriented innovation potential associated with data sharing.

In cases where these agencies belong to the public sector, the RfII considers it important to ensure the widest possible access to data for research and development purposes. With respect to current considerations in competition law, the RfII recommends that any regulatory measures to open up the data of large platforms to commercial actors also enables access for publicly funded research and development to an appropriate extent.

The RfII would like to support and clearly articulate the needs of science in this process: Sectoral data monopolies or other inappropriate restrictions on access must be avoided. Accordingly, policies to regulate data access should include a research clause, taking into account the wider interests of scientific research.

In addition, there are numerous examples of data platforms providing data access for scientific purposes through research contracts, cooperation agreements, or the provision of research data sets. Depending on the level of data protection required, trustee agencies can perform a variety of tasks, including the development of data products for specific cases of re-use and for different groups of users. Such scenarios have already been tested in the research sector and are being used successfully in many applications.

Establishing quality assurance for trustee agencies

Data trustees should act primarily as reliable neutral parties, in a transparent and fair manner to organise the use of data amongst potentially competing actors. However, a fuller definition of the functional role that such infrastructures would fulfil is still pending. To avoid ambiguity, or future plasticity in the application of the data trusteeship concept in the German-speaking debate, the RfII proposes that the term data trustee agencies should be established and protected by a quality seal or specific certification. In the opinion of the RfII, the term should be associated with a quality charter that is underpinned by appropriate quality assurance measures, ensuring the desired neutrality. The quality of the data should also be taken into account when defining the criteria required for this purpose. Although data trustee agencies are not, or should not be, explicitly responsible for data quality assurance, they should at least enable an assessment of the data provided and pass on information on data quality to users.

Since data can vary immensely in the degree of sensitivity, as well as in the intended use and conditions for re-use, it is appropriate to provide for different levels of quality assurance. For example, in the case of high value or high-risk potential data sets, checks by third parties, such as independent certification bodies or state monitoring bodies, will be necessary. If the level of importance is lower, self-regulatory mechanisms such as codes of conduct could be considered. The scientific sphere is familiar with implementing grading classifications of this kind, for example in the certification of "trustworthy long-term archives". Acquiring a quality seal or an accreditation often also entails alignment with agreed community norms or best practices, wherein data centres organise the exchange of experience and know-how in developing common procedures.

Promoting cross sectoral co-design and regulation of data trustee agencies

The RfII suggests initiating discussions between actors involved in policy making, business, administration and science promptly, in order to develop concrete proposals for a system of certified data trustee agencies, ultimately, also at the European level. In this process, science should not only contribute its long standing experience in developing regulations for access and re-use of data but should also clearly articulate its own specific needs. The design characteristics and regulations for data trustee agencies will have to be determined with particular regard to the nature and criticality of the data provided and the purposes of their use. Funding initiatives to elaborate such requirements and regulatory needs in more detail would help build momentum to advance the development of this kind of data infrastructure in a robust and manageable manner.

Imprint

Courtesy translation

German Council for Scientific Information Infrastructures (RfII)
Head Office
Papendiek 16, 37073 Göttingen, Germany
Phone +49 551 392 70 50
Email info@rfii.de
Web www.rfii.de