

Data Trustees – They Do Work! The Example of Research Data Centers

Daniel Fuß¹ and Marie-Christine Laible²

¹ Leibniz Institute for Educational Trajectories, Germany (Research Data Center)

² Federal Office for Migration and Refugees (Research Data Center)

Abstract. This contribution presents the long established system of accredited Research Data Centers (RDCs). Created in the data-landscape of the social, behavioral, educational, and economic sciences, they enable access to restricted data and bridge interests of data providers and researchers. A distinctive feature of most research data in the above disciplines is the coverage of real persons. Such sensitive data require specific safeguards. The focus of this contribution is on the institutionalized processes of connecting and securing appropriate research data management strategies for this sort of data. It includes quality assurance measures through the accreditation of RDCs, a monitoring system and the regular cooperation of accredited RDCs.

Keywords: Research Data Infrastructure, Quality Assurance, Sensitive Data, Forschungsdatenzentrum, RatSWD, KonsortSWD

1. Overview

An increasingly prominent term in the discussion about the provision of sensitive data for research purposes is that of the ‘Data Trustee’ as data intermediation services (cf. EU Data Governance Act [1]). Data trustees mediate the interests of data providers or data producers and data users. They should ensure a balance between the high level of protection for the often personal data on the one hand and the information depth necessary for high-quality research on the other hand by means of pseudonomization or anonymization processes, contract management etc. These demands largely describe the range of tasks of accredited RDCs. Thus, they are a model of data trustees. In our presentation we take an exemplary look at the rather young RDC of the Federal Office for Migration and Refugees (BAMF-FDZ), which prepares and makes available data from the Central Register of Foreigners (Ausländerzentralregister, AZR) for research purposes.

The network of decentralized RDCs is unique in Germany. RDCs make up an infrastructure that successfully accommodates the different demands of researchers and that has continuously adjusted to new requirements. They are considered to be “best practice” by the Council for Scientific Information Infrastructures (RfII [2, pp. 30–33], cf. also Wissenschaftsrat [3, pp. 81–82]) and have played a growing role in the German scientific system during the last two decades. The first RDC was founded in 2001 by the Federal Statistical Office (FDZ-Bund), followed by five more RDCs until 2008 – the year in which the German Data Forum (RatSWD) introduced the accreditation process. Currently, there are 41 accredited RDCs with a thematic range from economic and insurance data to education and labor market data to health and social data as well as a broad spectrum of data types from large-scale panel studies such as the Socio-Economic Panel Study (SOEP) and the National Educational Panel Study (NEPS) to register and census data to rich qualitative data collections in text, audio, and video formats.

In 2021, these RDCs had 5,517 datasets made available to nearly 53,000 national and international data users, counted more than 88,000 downloads of open access datasets, and recorded 3,101 scientific publications based on RDC data (cf. RatSWD [4]).

All accredited RDCs are members of the Committee for Data Access (CDA). It is made up of the heads of the RDCs and convenes twice a year to share best practices in research data management, but also to exchange experiences with innovations. All efforts of the CDA aim at safeguarding the continuous improvement of the research data infrastructure, which includes advancing the quality and quantity of available data as well as facilitating data access for researchers. Specific working groups are set up to deal with cross-RDC issues and to develop concerted answers to common (user) demands. A coordination office acts as single point of consultation and contact. The RDCs in the CDA are also essential in the context of the Consortium for the Social, Behavioral, Educational and Economic Sciences (KonsortSWD) within the National Research Data Infrastructure (NFDI), particularly due to their central role in various measures around research data management issues.

The CDA cooperates closely with the RatSWD. First, the RatSWD accredits new RDCs based on recommendations from the CDA. Accreditation is bound to some commonly defined mandatory criteria and a couple of standardized information requests. Thus, RDC accreditation ascertains a minimum level of data management and data sharing based on common guidelines (Bug et al. [5]). RatSWD and CDA have also established a reporting and monitoring system. All accredited RDCs submit an annual questionnaire reviewing the use of their data, their services and activities in the previous year. Finally, a general complaints office has been installed for data users to report alleged infringements of accreditation criteria. All three measures are important quality assurance instruments. By accrediting and monitoring RDCs, the CDA and the RatSWD have created a highly connected and sustainable research data infrastructure that guarantees user-friendly research data management and secure data access.

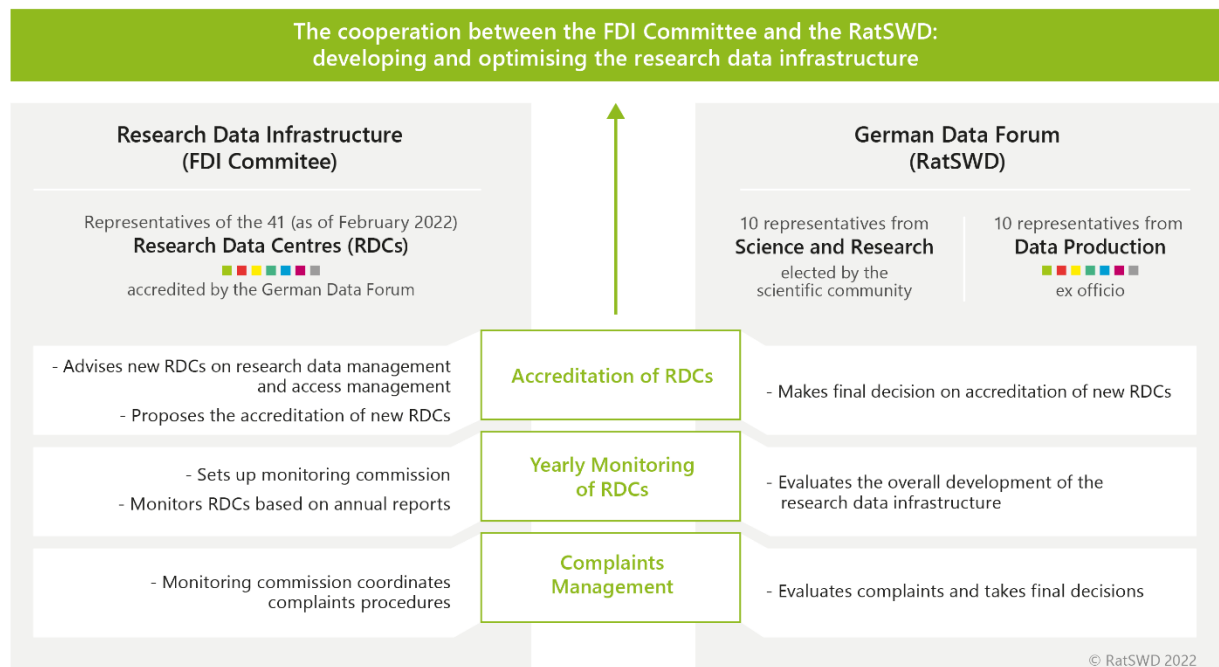


Figure 1. Connections of CDA and RatSWD for quality assurance of research data management in the social, behavioral, educational, and economic sciences.

Data availability statement

-

Underlying and related material

-

Author contributions

-

Competing interests

The authors declare that they have no competing interests.

Funding

-

Acknowledgement

-

References

1. European Union. "Regulation (EU) 2022/868 of 30 May 2022 on European data governance, and amending Regulation (EU) 2018/1724 (Data Governance Act)," Official Journal of the European Union. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868> (16 April 2023)
2. RfII, 2016. „Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland.“ Göttingen, Rat für Informationsinfrastrukturen. <http://www.rfii.de/?p=1998> (16 April 2023)
3. Wissenschaftsrat, 2011. „Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften.“ Berlin. https://www.wissenschaftsrat.de/download/archiv/10465-11.pdf;jsessionid=947594311989891E2366B5CB396E9580.delivery2-master?_blob=publicationFile&v=3 (16 April 2023)
4. RatSWD, 2022. „Research Data Infrastructure accredited by the RatSWD.“ Berlin, German Data Forum. <https://doi.org/10.17620/02671.71> (16 April 2023)
5. M. Bug, S. Liebig, C. Oellers and R. T. Riphahn, „Operative und strategische Elemente einer leistungsfähigen Forschungsdateninfrastruktur in den Sozial- und Wirtschaftswissenschaften,“ Jahrbücher für Nationalökonomie und Statistik, 238, 6, pp. 571–590, 2018, doi: <https://doi.org/10.1515/jbnst-2018-0029>