

Science and the humanities need legislation on research data!

Positioning of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)

Securing the future viability of our country is inconceivable without the fundamental contribution of science and the humanities. This was demonstrated impressively in the handling of the coronavirus pandemic. The latter also showed clearly that in addition to interdisciplinary approaches in science and the humanities, it is crucial to provide cross-sectoral access to data and to enable the interlinking of data. Currently, there are considerable hurdles in terms of the ability to access and interlink data generated outside research. In times of multiple crises, all sectors need to be even better equipped to deal with current and future challenges. Secure and regulated access to data is essential here. A need for action in this area has already been recognised at European level, as reflected in the European strategy for data. Research data legislation that provides research with legally secure access to data – and therefore going well beyond the narrowly defined requirements of the Data Act still under discussion at European level – makes a fundamental national contribution here, thereby enriching the European discourse.

Despite the organisational challenges and considerable effort involved, the data generated through publicly funded research projects is usually made available following the FAIR principles within the research community, but also increasingly to other sectors as well. By contrast, science and the humanities lack systematic access to relevant data generated in other sectors.¹ As a result, it is not possible to fully exploit the potential for innovative research and the transfer of findings to practical application. This shortcoming can only be remedied if research data legislation is consistently conceived as a “data access law for the research community”. In order to improve this situation, access to data generated by official bodies and authorities in particular requires urgent improvement, as does the use of this data for the

¹ Deutsche Forschungsgemeinschaft (2022): *Open Science as Part of Research Culture. Positioning of the German Research Foundation*, Bonn; DFG Interdisciplinary Commission for Pandemic Research (2021): *Daten für die gesundheitsbezogene Forschung müssen besser zugänglich und leichter verknüpfbar sein*, Bonn; Wissenschaftsrat (2022): *Digitalisierung und Datennutzung für Gesundheitsforschung und Versorgung. Positionen und Empfehlungen*, Cologne.

purpose of research. Otherwise, severe limitations will be imposed on the capacity to answer certain research questions and put forward evidence-based recommendations and forecasts that contribute significantly to the common good and the provision of public services – even though the methodological approaches required for this are essentially available in science and the humanities.

Global challenges can only be tackled based on international partnerships and cooperation. For this reason, German research must remain internationally competitive. Without improved access to data and the capacity to interlink it, this can no longer be guaranteed. While other European countries with comparable legal frameworks have significantly improved data access and the use of data by the research community², science and the humanities in Germany are increasingly at risk of being left behind. As a result, researchers are increasingly turning to data from other countries.

For the above reasons, the DFG welcomes the federal government's objective as set down in the coalition agreement concluded by the SPD, Bündnis 90/Die Grünen and FDP for the 20th legislative period and since reaffirmed in the Future Research and Innovation Strategy: "Our aim is to comprehensively improve and simplify access to research data for public and private research by means of a research data act and introduce provisions which give research the right to access data"³. Several European regulations have already established a legal framework and this should be taken up and further developed in national legislation. The goal of national legislation must firstly be to create the conditions for the untapped potential of data collected outside the research community to be used more effectively than before in contributing to the gain of knowledge. In order to achieve this, the existing barriers to access must be dismantled. Secondly, the aim must be to establish legal certainty by clearly defining the expectations of all parties involved. This is because clear-cut legal and technical standards facilitate the handling of data and reduce the risks involved for all those who provide and work with data.

Statutory regulation through research data legislation must consistently allow for further developments in organisation, process and infrastructure. Here it is necessary to draw on existing structures (e.g. the National Research Data Infrastructure, NFDI). In the social, behavioural and economic sciences in particular, accredited research data centres are already

² Rat für Informationsinfrastrukturen (2022): *Datenpolitik, Open Science und Dateninfrastrukturen. Aktuelle Entwicklungen im europäischen Raum*, Göttingen.

³ www.bmbf.de/bmbf/de/forschung/zukunftsstrategie/zukunftsstrategie

being used today based on a data trust model so as to concretely implement legally defined access to data produced by official bodies and authorities. In other areas, structural requirements of this kind are still lacking.

Even though improving data exchange with the private sector is just as relevant to research as exchanging data with official bodies and authorities, the requirements and framework conditions differ significantly between the public and private sector, which means that specific regulations, differentiated approaches and different transition periods are required. Public sector data is usually collected using taxpayers' money and should in principle be accessible to the research community. In the case of the private sector, suitable structures and processes must first be identified in order to be able to establish the appropriate conditions for statutory regulation.

Nevertheless, the handling of data must be viewed and conceived holistically according to the notion of a data ecosystem. A data institute needs to be established for this purpose as this will be able to make a vital contribution, providing the participation of all sectors is guaranteed – including research. Together with other actors in the field of research, the DFG has extensive experience in dealing with data from different fields of research, thereby enabling it to contribute to ongoing debate.

Fundamental aspects that must be considered in the formulation and implementation of a research data act include the following:

- a) Legal aspects
 - ▶ Since publicly funded research and teaching fundamentally serve the common good, privileged access to data generated in other sectors should be guaranteed.
 - ▶ Access to data generated in the private sector requires separate consideration, since data sets are usually essential to business and innovation processes. The preservation of data sovereignty is a particularly important and complex factor in this area. In terms of using data generated in the private sector therefore, it will first be necessary to intensify the sharing of data between research and industry in practice and develop framework conditions for this to happen. For this reason, we recommend that research data legislation should initially focus on data held by the public sector. The only other area that should be equally covered by such legislation is access to data that is generated collaboratively between research and industry.
 - ▶ Legislation on research data should in principle be conceived of as an overarching form of statutory regulation, even though specifications will be necessary. The Health

Data Use Act (GDNG) is a good example of this. Data from numerous different fields can be relevant to answering most questions in research, so this fact should not be counteracted by legal differentiation.

- ▶ In order to ensure the validity of subsequent research results, the focus should be on data that is collected by means of structured and quality-assured processes or that is of particular importance to research. It is neither realistic nor necessary to require full access to data held by third parties or extend comprehensive archiving requirements to areas outside research.
- ▶ A right to refuse to give evidence and safeguard clauses for the data used in research must be established so as to ensure the data is used in a trustworthy manner. Liability issues should be regulated in such a way that the risk to researchers using the data is minimised and research processes are enabled.
- ▶ In order to enable unbureaucratic and non-discriminatory access to data, it is necessary to reduce legal fragmentation in data protection, in particular based on a uniform interpretation of the GDPR and also a uniform conception of the granting of permission in federal and state law. It must be possible to guarantee aspects relevant to research such as the verifiability of research results, and these must be sensibly balanced with the requirement to delete data. The concept of data economy also deserves critical reflection.

b) Organisational and infrastructural aspects

- ▶ Data-sharing infrastructures minimise the risk to all parties, promote harmonisation and standardisation, and reduce costs.
- ▶ When setting up new structures, existing structures must be used or integrated; in particular, the NFDI and other central subject-specific research data infrastructures must be taken into account.
- ▶ It is imperative to ensure clear identification and the dismantling of structural obstacles to data interoperability, in particular by establishing persistent identifiers and using binding metadata standards and open formats in public bodies and authorities.
- ▶ The use of data for publicly funded research must not become a financial resource issue. Expenses incurred by data holders for providing data to third parties should at most have to be compensated in the form of an expense allowance.
- ▶ The use of data trust structures is a key starting point in terms of maintaining the confidentiality of information when dealing with particularly sensitive data. Separation of data storage and data use via trustees makes it easier to work with trusted or

personal data. The highest data security standards can be met by means of IT and organisational measures.

- ▶ The quality and comparability of datasets cannot be guaranteed without access to relevant software, so access to the latter is equally important. This can be seen in connection with research instrumentation, for example, and information management at hospitals and surgeries.

c) Overarching aspects

- ▶ Systematic cooperation with regard to the exchange of data between research, official bodies and authorities and the private sector should be stimulated in a targeted manner by means of funding measures in order to illustrate the mutual benefits in practice while at the same time broadening the basis of trust.
- ▶ Research should collaborate with the private sector to develop models that enable secure and trustworthy access to data held by the private sector – not least in order to be able to draw on such models to develop data trust structures and further elaborate the legal framework conditions.
- ▶ High levels of administrative complexity in the accessing and interlinking of data particularly affect researchers in the early stages of their scientific career: regulatory clarity contributes to the development of creativity and innovative capacity in all phases of a scientific career.

Deutsche Forschungsgemeinschaft

Kennedyallee 40 · 53175 Bonn · Postal address: 53170 Bonn

Tel.: + 49 228 885-1 · Fax: + 49 228 885-2777 · postmaster@dfg.de · www.dfg.de

DOI: 10.5281/zenodo.8223970