

Berlin, September 2010

Criteria of the German Data Forum (RatSWD) for the Establishment of Research Data Infrastructure

This paper presents the aims, functions, and characteristics of the members of the research data infrastructure of the German Data Forum, which consists of the research data centers (FDZ, *Forschungsdatenzentren*) and data service centers (DSZ, *Datenservicezentren*).

The research data centers (RDCs) and data service centers (DSCs) are accredited and supported by the German Data Forum (RatSWD) with the aim of improving the research data infrastructure for the social, economic, and behavioral sciences, both at a German as well as international level. Whilst pursuing this goal, the German Data Forum also bears in mind that infrastructure also has to be installed in areas that go beyond the scope of traditional infrastructure as given by governmental statistics (for example, departmental research, evaluation studies, research based surveys, and research focal points using public funding).

Aim of the German Data Forum's Support for Research Data Infrastructure

The complexity of economic and social transformation, in addition to the progress made in science and information technology, have fundamentally changed the data requirements of scientific analyses of modern societies. The use of statistical aggregates (tables) is no longer sufficient. Currently, there are much greater methodological and content-related demands to provide science with microdata for deeper empirical analysis. Closely related to this is the provision of access to anonymized microdata, with which differentiated empirical analyses can be conducted. In supporting research data infrastructure, the German Data Forum's goal is to create a data environment in which available information can be utilized to its maximum potential.

History

Since the request of the Federal Ministry of Education and Research (BMBF, *Bundesministerium für Bildung und Forschung*) in 1999, the Commission on Improving the Statistical Infrastructure between Research and Statistics (KVI, *Die Kommission zur Verbesserung der Informationellen Infrastruktur zwischen Wissenschaft und Statistik*), and, after its founding, the German Data Forum, have been responsible for developing suggestions on how Germany's data infrastructure may be improved. One of the central recommendations of the Commission (2001) was to establish RDCs and DSCs. Shortly thereafter, four RDCs were founded – the Institute for Employment Research (IAB, *Institut für Arbeitsmarkt- und Berufsforschung*) at the Federal Employment Agency (BA, *Bundesagentur für Arbeit*), the Research Data Center at the German Pension Insurance (FDZ-RV, *Forschungsdatenzentrum der Rentenversicherung*), the Research Data Center at the German Federal Statistical Office and at those of the *Bundesländer* (FDZ-SB, *Forschungsdatenzentrum des Statistischen Bundesamtes und der Statistischen Ämter der Länder*), and at the Leibniz Institute for the Social Sciences (GESIS, *Leibniz-Institut für Sozialwissenschaften*) – all of which are funded by the BMBF and have been accredited by the German Data Forum. Additionally, alongside

the research data centers and data service centers, the recommendations of the KVI further endorse the expansion of existing research infrastructure and the construction of further research data centers at important data producers and establishments involved in fields of scientific research. It is envisaged that these additional RDCs specialize according to their areas of competence. In doing so, not only can the access to data be expanded, but also the quality of access and the quality of the data can be improved and organized in an understandable manner. The objective of this exercise is to improve the access to all forms of data, especially those relevant for the social, economic, and behavioral sciences.

To advise and coordinate this expanding research infrastructure, so as to create synergy effects, is one of the central tasks of the German Data Forum.

After having evaluated the German Data Forum in 2009, the German Council of Science and Humanities emphasized in its recommendations that:

“The dynamics of the creation of new research data centers requires particular attention from the German Data Forum. In the future, it will become necessary for the German Data Forum to perform considerable integration and coordination tasks in order to ensure an ongoing quality in the provision of data. The German Data Forum should continue to offer prompt, user-friendly, and open data access and avoid prohibitive fee-based models of data provision. In this context, the development of sustainable models of financing has high priority” (own translation).

In order to fulfill these demands and to ensure their implementation in an administratively comprehensible manner, the German Data Forum has developed criteria for accrediting establishments as German Data Forum RDCs or German Data Forum DSCs.

Criteria for Accreditation as a German Data Forum Research Data Center or Data Service Center

- 1) In the RDCs and DSCs accredited by the German Data Forum, research data are prepared and made available for scientific purposes. Additionally, service is offered for data users at these centers.
- 2) Research data form the cornerstone of scientific understanding. Accordingly, they can be used for scientific purposes independent of their original collection purposes. This includes publicly funded data that were originally collected for statistical purposes within the framework of administrative decisions related to evaluating research based infrastructure or other associated objects of assessment.
- 3) User access is of course subject to field-specific ordinances of data protection and data security. The role of the German Data Forum is to facilitate access to data.
- 4) Access to microdata is governed by legal regulations that ensure equal treatment of data users. Accordingly, the data centers provide for transparent and standardized access arrangements. This additionally implies that no application for access is treated preferentially or rejected as a direct result of the contents of that application. The data centers are not responsible for judging the intentions of specific research projects; rather, they only assess applications based on legal measures pertaining to data protection and contractual legitimacy. If contractual or legal restrictions apply to the analysis of data, these must be published at the same time as the data is released. Assessments that do not violate data protection regulations and are contractually permitted are essentially allowed to be independently published by the user.
- 5) In addition to the provision of access, the data centers produce detailed documentations of data. Furthermore, information about available data and data centers are made available in standardized form via the internet, data and methodology reports, and the provision of individualized advice. The tasks of

the data centers additionally include organizing and participating in scientific events to present available data, as well as presenting data and how they can be accessed to potentially interested parties (in particular non-university affiliated research institutes, polytechnical colleges, and universities). The data centers actively participate in both discussions of the analytical potential of available data and possibilities for the use and development of data infrastructure for scientific purposes.

- 6) The data centers accredited by the German Data Forum should, to a specified extent, conduct research. Only he who researches holds knowledge about the data, current research methodologies, and content based discussions. Accordingly, data centers will be better able to advise and supervise data users adequately. The assigned tasks of the data centers must not be limited to simple service oriented tasks, for this would be equivalent to withdrawing from the scientific community. Conducting scientific work in the data centers facilitates career development and further qualification as well as participation at scientific events and the publication of papers in journals.
- 7) Conducting research in the data centers is not connected with exclusive access to the data. The data present in RDCs and SDCs are available to all researchers, at least via controlled remote data processing or guest researcher workstations.
- 8) The accreditation of RDCs and DSCs is performed by the German Data Forum. In order to be listed as German Data Forum research data infrastructure, RDCs and DSCs must declare that they will act according to these German Data Forum criteria. Furthermore, at the certification stage a report concerning the work of the relevant RDC or DSC is to be produced according to the specifications of the German Data Forum. Once accredited, the relevant RDC or DSC becomes a member of the standing committee of the RDCs and DSCs. A presentation of the work/data and organizational form of the newly accredited RDC/DSC is desirable.
- 9) In order to ensure documentation of the long-term compliance of the RDCs and SDCs with the German Data Forum criteria, and to demonstrate the common

progress and relevance of the research data infrastructure, all members of the German Data Forum accredited research data infrastructure are required to produce standardized annual reports according to the specifications of the German Data Forum. Further details are to be specified by the standing committee of the research infrastructure and its indicator catalogue.

- 10) For those RDCs and DSCs that are still in their founding stages and do not yet possess one of the databanks commonly available to scientific researchers, guest status in the RDC/SDC committee may be awarded. The German Data Forum and the standing committee deem it of great use to communicate and engage in mutually beneficial dialogue with potential institutions of Germany's research data infrastructure at as earlier stage as possible. The German Data Forum is responsible for granting the aforementioned guest status. Further details are determined by the standing orders of the Standing Committee on research data infrastructure.
- 11) Accreditation as part of official German Data Forum research data infrastructure does not entail start-up financing or a recommendation for funding. Should the Federal Ministry of Education and Research or another ministry or funding body ask the German Data Forum to assess concrete applications for funding, the German Data Forum is required to principally obtain this assessment from established external experts.