Activities Report 2018
of the Research Data Centres (RDCs)
accredited by the
German Data Forum (RatSWD)
Activities Report 2018
of the Research Data Centres (RDCs)
accredited by the
German Data Forum (RatSWD)
Fig. 1: Locations and guest researcher workstations of accredited research data centres in 2018. Most RDCs keep their German acronym FDZ.

International Locations (FDZ BA at IAB)

- USA
  - Cornell University, Ithaca**
  - Harvard University**
  - Princeton University**
  - University of California, Berkeley**
  - University of California, Los Angeles**
  - University of Michigan**

- Canada
  - Vancouver School of Economics**

- UK
  - UK Data Archive, University of Essex**
  - Centre d’Accès Sécurisé aux Données (CASD), Palaiseau (Paris)**

© RatSWD 2019
## Contents

Preface by the chairs of the FDI Committee ................................................................. 7

1 Overview of the research data infrastructure of the German Data Forum (RatSWD) . . 8
   New accreditations

2 The structure of the research data centres (RDCs) .................................................. 11
   Staff
   Cooperation and research activities
   Number of scientific publications

3 Archiving and quality assurance .............................................................................. 15
   Concepts for making data available on a long-term basis
   Archiving
   Quality assurance

4 Availability and use of data ...................................................................................... 19
   The range of available data
   Waiting periods and fees
   Research output based on RDC research data
   Data use

5 Services for Users .................................................................................................... 23
   Service quality assurance
   Data access paths and data formats
   Provision of tools
   Time period between the application and the transmission of the data
   Publicising the research data
   Skills development

6 Internationalisation and the current development of the research data infrastructure .................................................. 30
   Internationalisation
   Innovation and improvement of the research data infrastructure
   Further developing the research data infrastructure

7 Special topic: data protection .................................................................................. 33

8 Complaints management .......................................................................................... 36

Appendix

Appendix A: Development of the RatSWD research data infrastructure and RDCs ........ 38
Appendix B: The monitoring commission .................................................................... 44
Appendix C: Contributors to the Activities Report 2018 .............................................. 45
Preface by the chairs of the FDI Committee

Both policymakers and the public expect researchers to provide responses to key questions on the security of social cohesion and economic prosperity in an era of global challenges, such as climate change, migration, and digital transformation. Researchers in the social, behavioural, and economic sciences are making a significant contribution by identifying opportunities for shaping a liveable future and developing solutions for the challenges ahead. For that purpose, researchers depend on access to data that is as current as it is comprehensive. This is even more important for innovative research approaches. The research data centres (RDCs) accredited by the German Data Forum (Rat für Sozial- und Wirtschaftsdaten - RatSWD), which are organised in the FDI Committee (‘Standing Committee Research Data Infrastructure’), facilitate access to research data in the most transparent way possible.

The Activities Report 2018, of the now 34 accredited RDCs, highlights the dynamic development of the research data infrastructure in the social, behavioural, and economic sciences. In 2018, three new RDCs were accredited, and many other research institutions have expressed an interest in establishing RDCs and accreditation.

In addition to becoming a successful model for organising access to research data, the establishment of RDCs with the German Data Forum’s (RatSWD) accreditation has resulted in a significant expansion of the data available to researchers in the social, behavioural, and economic sciences. The accreditation safeguards adherence to minimum standards and increases the visibility of the research data held by RDCs. Their academic staff makes sure that those data are systematically aligned with the needs of the scientific community. The close ties with the research community are reflected by the high number of research papers coming out of RDCs. The connections between the RDCs (see Fig. 3), the distribution of RDC locations all over the country (see Fig. 1), and the numerous guest researcher workstations are proof of a vibrant and growing research data infrastructure that serves the research community.

The key figures of 2018 highlight the successful development of the accredited RDCs: by the cut-off date for this publication, 31 December 2018, the RDCs employed a total of 285 staff in full-time equivalents (FTEs), of which academic staff accounts for 175 FTEs. The RDC staff generated 528 scientific publications in 2018. This is a success in many ways: (re-)using in-house research data creates and secures close ties to the research community, strengthens the competencies of RDC staff to support users, thus contributing to quality assurance, and making the datasets better known. At least 2,074 scientific publications are fully or partly based on the 3,940 datasets made available by the RDCs. A more systematic collection of publications would probably yield an even higher number. Data citations are an important tool to document the relevance of data availability for empirical research. RDCs have laid the foundation for this development with their widespread use of persistent identifiers for research data; this cultural shift is now also beginning to take root on the user side. The RDCs should seek to highlight the benefits of data citation even more in the future.

In 2018, 369 additional datasets were made available to data users; these datasets have considerably expanded the opportunities for re-using research data. The demand for these data was high: in 2018, 46,464 data users were registered at RDCs – 9,081 new users signed 5,112 data use contracts.

Safeguarding data protection is a key task of the RDCs. Almost every RDC processes sensitive microdata and makes them accessible to researchers – either by de-facto anonymising the data or via secure data access paths. It was therefore a logical step for RDCs to prepare for the introduction of the GDPR with considerable internal and external support. Several adjustments were necessary at RDCs, particularly in contract management and in declarations of consent. This year’s report contains a special chapter dedicated to the ins and outs of GDPR implementation at the RDC-level. Data anonymisation procedures, on the other hand, have hardly changed.
In preparation for new challenges and opportunities in research data management, the FDI Committee focused strongly on developing new areas of activity last year: advancing the archiving structure and data access, preparing common guidelines, and expanding training opportunities at RDCs. These areas of activity were aimed at helping RDCs to understand the current dynamic within the research data infrastructure and to channel it in productive ways. It is the goal of all RDCs to play an active part in shaping the future of data access and to continuously improve it for the benefit of their users. All this feeds back into the RDCs’ overarching goal: improving the services they provide to the research community.

To foster cooperation beyond the national level, the German Data Forum (RatSWD) founded the GO FAIR Implementation Network Eco-Soc (EcoSoc-IN) in 2018. Seventeen RDCs joined the network with an individual mandate. GO FAIR is committed to improving the findability, accessibility, interoperability, and re-usability of research data. The FAIR criteria are closely intertwined with the values of the German Data Forum’s (RatSWD). They will serve as the principles guiding the future development of German and European research infrastructures. However, the comprehensive implementation of the FAIR principles requires substantial investment into existing infrastructures. The role of research funding organisations in financing research infrastructures is changing accordingly, as is demonstrated by the GWK’s commitment to funding and setting up the National Research Data Infrastructure (NFDI).

The German Data Forum (RatSWD) and the FDI Committee wish to thank the members of the monitoring commission for dedication in responsibly developing this report. By advancing the development of monitoring processes, the monitoring commission is continuously helping us to develop a better understanding of the infrastructure and what it is capable of.

Lastly, we wish to express our sincere appreciation for the assistance provided by the German Data Forum’s (RatSWD) business office in supporting the monitoring activities and creating this publication.

Dr. Pascal Siegers  
(Chair of the FDI Committee)

Dr. Jan Goebel  
(Deputy Chair of the FDI Committee)
1 Overview of the research data infrastructure of the German Data Forum (RatSWD)

The FDI Committee (‘Standing Committee Research Data Infrastructure’) oversees a dynamic and decentral network of 34 research data centres (RDCs) that are accredited according to the German Data Forum’s (RatSWD) criteria (as of July 2019). The steady increase of RDCs shows that the RDC model is fit for the future and that the German Data Forum’s (RatSWD) accreditation has become established as a seal of quality among research funders and data users.

The RDC network has seen substantial expansion with the accreditation of three new RDCs (see p. 10): the accreditation of the RDCs Qualiservice and eLabour has added two very strong collaborative research projects that expanded the availability of qualitative data. Moreover, both RDCs have a long-standing expertise in dealing with the processing and archiving of qualitative interview data in accordance with data protection.

The RDC of Germany’s Federal Motor Transport Authority (Kraftfahrt-Bundesamt), which has now obtained provisional accreditation, is adding more official register data to the infrastructure’s data portfolio. It has kicked off by making the anonymised microdata from the Register of Driver Fitness (Fahreignungsregister FAER) available to researchers.

How it all began

The Commission on Improving the Informational Infrastructure (KVI) was established in 1999 as a response to initiatives from within the scientific community. It presented a comprehensive report in March 2001. One of its key recommendations was to set up RDCs at major public data producers. The Founding Committee leading to the German Data Forum (RatSWD) was set up that same year.

Evolution and consolidation

Since 2001, the founding of additional RDCs has led to a consistent development of the research data infrastructure. There were various reasons for institutions to set up RDCs: some sought to implement the recommendations issued by commissions like the KVI, the German Council of Science and Humanities, or scientific advisory groups; other RDCs were commissioned by their own institutions with the aim of fostering research. What they all had in common was the aim to expand and strengthen the research data infrastructure in Germany by improving access to research data for the scientific community.

The German Data Forum (RatSWD) was founded in 2004. To promote a productive dialogue between the RDCs, the German Data Forum (RatSWD) set up the FDI Committee. Its main task was to continuously secure and improve the research data infrastructure, i.e., expanding the quality and the quantity of data and data access.

<table>
<thead>
<tr>
<th>The KVI is founded</th>
<th>Founding Committee German Data Forum (RatSWD)</th>
<th>The German Data Forum (RatSWD) is founded</th>
<th>The FDI Committee is set up</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>2001</td>
<td>2004</td>
<td>2009</td>
</tr>
</tbody>
</table>

1 A full list of RDCs is included in the appendix on page 38.

To assure the quality of the research data infrastructure, the German Data Forum (RatSWD) defined minimum standards and accreditation criteria in 2010. These standards and criteria were modified several times and regular monitoring and evaluation processes were implemented between 2015 and 2018. Elected from the members of the FDI Committee, this is when the monitoring commission took up its key operative task.

German Data Forum (RatSWD) accreditation is a seal of quality for the RDCs. Before accreditation is awarded, RDCs are reviewed for compliance with the German Data Forum’s (RatSWD) mandatory criteria. Accredited RDCs also receive extensive support: they can connect with other accredited RDCs, receive information about best practice solutions to help guide the ongoing development of their own infrastructures, and participate in the exchange of knowledge and experiences with other RDCs. Fig. 3 shows the numerous connections between the accredited RDCs in 2018, beyond cooperating within the FDI Committee. Considerable innovative potential for research is derived from this close cooperation as well as from the resulting opportunities to merge datasets.

By recommendation of the German Council for Scientific Information Infrastructures (RfII), the Federal Government initiated the establishment of a National Research Data Infrastructure (NFDI) in 2018. Not primarily a technical infrastructure, the aim of the NFDI is to precipitate a cultural change in the scientific community that values sharing of research data as an integral component of good scientific practice. The prerequisites of this, however, are professional standards for research data management, which need to be implemented across the communities. The RatSWD, too, aims at fostering this change. The FDI Committee began addressing these issues as early as 2017. In 2018, following a needs analysis among the German Data Forum’s (RatSWD) accredited RDCs, the FDI Committee specified these broader areas of activity towards the needs of the RDCs and their respective research communities.

These long-term efforts aim to ensure that the future of research data infrastructure stays globally competitive and that important developments concerning the increasing digital transformation and internationalisation are addressed early on.

---


4 For more background information and a list of members of the monitoring commission, see Appendix B on page 44.
New accreditations

Research Data Centre eLabour (FDZ eLabour)
eLabour is a collaborative project of several research institutions from industrial and occupational sociology working together with partners from IT, information science, and data infrastructures. Their aim is to archive comprehensive research material from qualitative studies in labour sociology (since the 1970s) and to make them available for scientific re-use in a customised research environment. Moreover, eLabour is developing the competencies and support services required by this task and establishing standards for research data management of qualitative data from industrial and occupational sociology. At the same time, it fosters and conducts secondary analysis research with these data. eLabour data are available as scientific use files (SUFs) and campus files (CFs) via download or remote processing.5

Research Data Centre Qualiservice (FDZ Qualiservice)
Qualiservice is the follow-up project of the Special Collaborative Centre 186 ‘Status Passages and Risks in the Life Course,’ which was founded 20 years ago. Members of Qualiservice include SOCIUM, PANGEA, the State and University Library Bremen (SuUB), the university library of Humboldt-Universität zu Berlin as well as its Special Subject Collection Folklore Studies and Ethnology, and GESIS. Qualiservice is also in charge of archiving the data from the Special Subject Collection Folklore Studies and Ethnology. Moreover, it is responsible for archiving a wide range of qualitative data collected through a variety of survey methods (context data, microdata, media data, transcripts, field notes) and qualitative data from neighbouring disciplines such as sports studies, religious studies, and spatial research. It also archives additional received or inherited qualitative data.

The RDC currently makes data available as pseudonymised SUFs, PUFs, and CFs for downloading or on data storage mediums sent by regular mail. Additionally, users may access the data via a guest researcher workstation.

Research Data Centre at Kraftfahrt-Bundesamt (FDZ im KBA)
Provisional accreditation6

The RDC at the Kraftfahrt-Bundesamt (KBA), Germany’s Federal Motor Transport Authority, is based in Flensburg. It is the result of a research project funded by the Federal Ministry of Transport and Digital Infrastructure (BMVI), which was launched in October 2017 for a period of two years. The project was later extended until late 2020. It is planned to make the RDC permanent by 2021. Its aim is to facilitate access to the entry data of the Register of Driver Fitness (Fahreignungsregister, FAER), which records traffic safety data from road users and corresponding demographic information. The data are particularly interesting for researching traffic safety issues and are available for downloading as SUFs. A guest researcher workstation is planned.

---

5 See page 27 for more details on the range of research data types.
6 Provisional accreditation may be granted to RDCs that are not yet fully operational, see also https://www.ratswd.de/en/info/monitoring-and-complaints-management.
2 The structure of the research data centres (RDCs)

The information presented in the following chapters was taken from the German Data Forum’s (RatSWD) annual quality survey, which was developed by all accredited research data centres (RDCs). In the 2018 reporting year, 34 RDCs, including the three newly accredited RDCs, took part in the survey.

### Staff

<table>
<thead>
<tr>
<th></th>
<th>2017 (n=31)</th>
<th>2018 (n=34)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total staff</td>
<td>264.03</td>
<td>284.84</td>
</tr>
<tr>
<td>Academic staff (Acad)</td>
<td>154.5</td>
<td>175.51</td>
</tr>
<tr>
<td>Non-academic staff (Non-acad)</td>
<td>67.14</td>
<td>42.39</td>
</tr>
<tr>
<td>Student assistants (Stud)</td>
<td>42.39</td>
<td>40.15</td>
</tr>
</tbody>
</table>

Range of FTEs in RDCs

- 0.4 – 34.6
- 0.25 – 23
- 0 – 11.6
- 0 – 6

Range of FTEs in RDCs

- 0.2 – 32.5
- 0.25 – 18.9
- 0 – 10.1
- 10 – 6

On the 31 December 2018 the 34 RDCs employed a total of about 285 staff in full-time equivalents (FTEs) (see Fig. 2). The increase in staff at RDCs continued in 2018. This is not only due to the newly accredited RDCs but also to an increase in staff, particularly academic staff. The number of student assistants saw a slight decrease. About 20 percent of RDCs do not employ students at all. As shown by the FTE range, staff numbers vary strongly across RDCs. The data here do not show a specific pattern, for example, that availability of student assistants depends on overall staff make-up, or on the RDC’s key working areas. Instead, the numbers underscore the diversity of accredited RDCs united by the common standards of the research data infrastructure.
Cooperation and research activities

The cooperative relationships between the German Data Forum’s (RatSWD) accredited RDCs have shown to be stable over time. Beyond the activities within the FDI Committee, the RDCs cooperate closely and continuously (see Fig. 3). Moreover, working relationships to institutions outside of the FDI Committee play an important role.

Cooperation within and beyond the FDI Committee has created a multi-faceted, interdisciplinary network of RDCs.
Two thirds of the RDCs maintain institutionalised working and research relationships with national partners (see Chapter 6 for international research cooperation). Two RDCs are planning to establish such relationships in the future, while 11 RDCs are not currently involved in such institutionalised relationships. The partners for cooperation cited in the survey mainly come from other scientific and semi-scientific institutions. First, this type of cooperation requires high-quality work from RDC staff and, secondly, it makes sure that the infrastructure demands of scientific data users are channelled directly into the RDCs.

Cooperating with external institutions fosters professional expertise at RDCs as does the in-house independent research conducted by their academic staff.

In 32 of the surveyed RDCs there is academic staff who conduct their own independent research (see Fig. 4). These research activities can encompass research-driven issues, the development of methodological innovation, or advancing technological developments. In-house independent research enables RDC staff to perform data-related, support-related, and consulting services on an adequate professional level and to help RDC users work with the data. At 18 RDCs, the academic staff dedicates a fixed share of their working time to research activities. At most RDCs, this amounts to a maximum of half of their work hours; only two RDCs dedicate more. On average, two thirds of staff work hours are dedicated to research activities.
The experiences gained by RDC staff, by conducting independent research with in-house data as well as from external data sources, ultimately foster the quality of the RDC’s user services. It helps to identify potential enhancements and weak spots when accessing the data, to sharpen awareness for quality assurance, and to explore (research/statistical) opportunities and limitations of the datasets. The academic staff’s research activities are documented by a range of publications.

Fig. 5: Number of scientific publications of RDC staff
Please indicate the number of scientific publications produced by your RDC staff, regardless of the type of data and whether the publication was prepared during RDC working hours.

The RDCs have significantly increased their publication output compared to the previous year. The staff at RDCs produced a total of 528 publications in 2018 (see Fig. 5). Regardless of the types of data used and whether the publication was prepared during RDC work hours, this underscores the high level of skills at RDCs. A total of 192 articles were published in journals, including 152 in peer-reviewed journals, which is also an indication of the research output’s high quality. An added 69 articles were published in edited volumes. The largest share of publications was featured in grey literature, a total of 239 publications, which includes working papers as well as technical reports. The latter are an essential component of research data management and re-using data. If we apply this figure to the number of academic staff, this amounts to more than three publications per FTE. If we include non-academic staff, assuming that all staff is equally involved in creating publications, the RDCs still produced two publications per FTE in 2018.
3 Archiving and quality assurance

Concepts for making data available on a long-term basis

Making sure that data used for research is available in the long-term is good scientific practice and firmly established in most research institutions. More and more scientific journals expect that published empirical findings can be replicated, which entails that the data are properly stored for further use and re-use purposes. Long-term data preservation is a central task of research data centres (RDCs). This also includes the obligation to delete the data held by RDC users after a fixed time period if this is required by the respective contracts with respondents and data providers.

In addition to physically storing the data (in different versions, if necessary), the RDCs support processes related to data re-use by ensuring that the data can be read, interpreted, and used in the future. Data environments can change over time. A key challenge is to store data in formats that ensure readability with future technologies. Aside from data archiving, RDCs must also safeguard data against loss as well as unauthorised and unwanted manipulation.

All RDCs hosting prepared datasets and/or data documentation also ensure their long-term availability and re-use (see Fig. 6). Having access to the prepared data also puts them in charge of data preservation. RDCs that host raw data but do not ensure their long-term storage are the exception. Centres hosting questionnaires and other metadata are usually in charge of their preservation. Overall, the results show that RDCs employ diverse archiving and re-use concepts to provide data users with comprehensive material for their research.

Fig. 6: Data preservation at RDCs

Does your RDC ensure the availability of data on a long-term basis (according to the rules of good scientific practice and for at least ten years after they were last used) and in standard formats? (Multiple answers possible)

Preservation of data documentation 31
Preservation of prepared data 33
Preservation of other metadata 30
(bibliographical, technical, administrative, and legal information)
Preservation of questionnaires 25
Preservation of project-related data 26
Preservation of raw data 23

Long-term data preservation is a core task of the RDCs.
Archiving

Thirty-three RDCs use locally redundant storage, that is store data on multiple storage mediums, and 20 of these RDCs report that they store data in multiple locations. Only one RDC restricts itself to a simple backup on in-house servers. The RDCs use a variety of procedures to backup data in databases, on servers, or in specialised data archives at regular intervals. Data are also preserved in a variety of data formats (such as CSV, SPSS, and Stata). Specialised data archives employ distributed and redundant storage on multiple storage mediums, i.e., the archived data are duplicated and secured in different physical locations on different storage mediums (hard drives, CDs, etc.). The RDCs have different ways of handling updates of data and metadata, where necessary. Most often, users receive the most current version of a dataset (20 RDCs). Nine RDCs archive and make available all versions; five RDCs host datasets that do not require updating.

In 2018, the number of RDCs with certification for their archiving solutions increased (from seven in 2017 to ten in 2018). Most are certified with the ‘Core Trust Seal’.7

Quality assurance

In most RDCs, one of the staff’s key tasks is to assist in data checks. At RDCs of institutions that gather data themselves, quality assurance usually starts during the testing of survey instruments and throughout every step of the data collection process. Usually, the data go through various stages of preparation, during which they are checked for completeness, consistency, and plausibility and any data errors or gaps are fixed, where possible. Four out of 34 RDCs performed all quality assurance measures themselves. The majority of RDCs perform some these checks themselves. Only four RDCs do not perform data checks at all.

RDCs who partially perform data checking tasks usually share the responsibility of data quality assurance with the primary researchers or the data producers (who can be third-party institutes). Fifteen RDCs have created clear guidelines for data checks.

---

7 Respondents in the 2018 reporting year also cited the Data Seal of Approval and the certification of ICSU World Data Systems. Both have now been merged with the Core Trust Seal. The aim of such certifications is to help researchers identify trustworthy repositories for archiving and curating research data. They provide information on a repository’s adherence to transparent standards laid out in the certification criteria.
Several RDCs provide regular documentation of their quality assurance measures in quality reports or data reports on individual data products. This is also reflected by the high number of working papers and technical reports (see Fig. 5) which also include data reports.

Fig. 8: Types of data checks
Which types of data checks are performed at your RDC?
(Multiple answers possible)

The RDCs perform appropriate data checks for each dataset. The types of checks strongly depend on the survey method.

The most common data checking measures at RDCs include checking consistency with data documentation, label checks, and completeness checks. About half of the RDCs also perform annual comparisons (see Fig. 8).
Another component of data quality assurance is correcting data errors, also during data generation. The distribution of responsibilities is comparable to that for data checks: three RDCs correct the data themselves; 24 do so only partially. Seven RDCs do not perform data correction measures at all.

The most common task in this department is the generation of additional variables. Coding missing values is also widespread. Correcting implausible values is a key task at 19 RDCs; harmonisation of longitudinal data at 18 RDCs. Nine RDCs stated that they are responsible for imputing missing values, which is a significant increase compared to the previous year (see Fig. 9). There are clear guidelines for data correction at 12 out of 27 RDCs; at 18 RDCs, data corrections are made transparent to data users at the individual level.

Overall, it can be said that the RDCs extended the scope of their quality assurance tasks compared to 2018.
4 Availability and use of data

The range of available data

As of the 31 December 2018, 3,940 datasets were available in the RDCs. In 2018, the RDCs added 369 datasets with a digital object identifier (DOI) or were suitable for DOI registration (see Fig. 10). The number of newly added datasets varies strongly across RDCs. One RDC added 68 new datasets, some RDC added a single digit number. Since a dataset can contain several individual studies, the number of available studies is significantly higher.

Thirty-one RDCs assign persistent identifiers (PIDs) to their datasets to ensure their long-term findability and citation. All RDCs that do not yet use PIDs, like DOIs, are currently in the process of implementing this practice.

Embargo periods and fees

Seventeen RDCs do not have embargo periods, meaning that the data are made available immediately after they are received and prepared (see Fig. 11). Three RDCs report fixed waiting periods ranging from six months to no more than two years. Embargo periods at all other RDCs hinge on certain requirements, for example, a research project’s end date, or other specifications by funding organisations or data givers.
In 2018, 34 RDCs reported that they do not charge any fees (see Fig. 12). The fees reported by seven RDCs were mostly in the two-digit or low three-digit euros range (charged per dataset, survey year, and data access path, or depending on the costs of specific anonymisation measures).

Research output based on RDC research data

Fewer RDCs are charging fees for data provision.

There is a continuous increase in the number of publications.
In the 2018 reporting year, 30 RDCs reported a total of 2,074 publications (see Fig. 13) based on the research data they provide (four RDCs did not provide any information in this category). However, many researchers neglect to inform the RDCs about a publication or to send a copy. Since many RDCs do not have the resources to continuously gather data on citations (which authors do not yet handle consistently either), it is safe to assume substantial underreporting here. The fact that researchers use datasets from several RDCs also contributes to this fact.

Despite these caveats with gathering data on citations, there was a slight increase in publications based on research data provided by the RDCs. Articles in journals continue to be the most reported type of publication. The number of articles in peer-reviewed journals is particularly high and is also the fastest growing category.

**Data use**

*Fig. 14: Data use at RDCs in 2018*

<table>
<thead>
<tr>
<th>Downloads</th>
<th>External data users</th>
<th>New external data users</th>
<th>Data use contracts on 31/12/2018</th>
<th>New data use contracts</th>
</tr>
</thead>
<tbody>
<tr>
<td>71,488</td>
<td>46,464</td>
<td>9,081</td>
<td>28,640</td>
<td>5,112</td>
</tr>
<tr>
<td>n=11</td>
<td>n=32</td>
<td>n=26</td>
<td>n=29</td>
<td>n=32</td>
</tr>
</tbody>
</table>

The research data infrastructure was used by more than 46,000 people in 2018.

One of the core tasks of the RDCs is to provide researchers with comprehensive and flexible data access and to continually expand and improve it. Due to the specific data protection regulation and other legal provisions, researchers are offered a range of data access paths. Due to this flexibility in access paths and differences in how data are gathered by RDCs on contracts, projects, and data users, it is difficult to determine a precise number of data users. Since double counting of contracts and projects or between access paths could not be ruled out in previous years, the surveying of downloads, data users, and data use contracts was modified in 2018. The new data are therefore not comparable to those from last year’s report. However, by surveying existing and new data users separately in 2018, this report can now present comparative values that make it possible to trace developments over time.

**Downloading data**

Overall, 14 RDCs make datasets available as free downloads. At the majority of RDCs, downloading requires prior registration. However, there are numerous open datasets that are available for download without prior registration, which creates obstacles to quantifying data access. Furthermore, not all data access paths facilitate the gleaning of usage statistics. In some cases, the technical infrastructure also interferes with quantifying downloads. All this contributes to undercounting of data users. The eleven RDCs, which are able to provide information on user numbers reported 71,488 downloads of open datasets in 2018.
The standard case, however, is that RDC-held research data are made available only after users have registered or signed a contract. For data protection reasons, contracts on data access and usage contain explicit references to research projects (purpose limitation), that is a separate contract must be drawn up for every research project using the data. However, there are no formal templates for such contracts. How RDCs design their contracts is governed by the freedom of contract and the contractual depth is determined by legal provisions and requirements. Access to official statistics data, for example, is legally restricted by a string of laws and regulations. Access to survey data, too, is subject to data protection regulation. This applies particularly to sensitive personal data. Other data, however including regionalisation and land use data, are openly accessible for some purposes and subject to licensing for others. This diversity is also reflected by the contract design used at RDCs. This is true for the data themselves as well as the signatory parties: contracts can be with individuals, projects, or entire research facilities. They can cover entire data collections, collections of studies, or individual datasets. It has proven a complex task to quantify such diversity. For this reason, this year’s Activities Report differentiates three subsections: external data users, data use agreements, and the data users, or datasets, covered by the agreements (see Fig. 14).

External data users
In the 2018 reporting year, the RDCs had a user base of 46,464 (data from 32 RDCs) external data users. External users are data users that are not affiliated with an RDC, or RDC’s home institution. Thirty-two RDCs were able to provide data on the number of external data users. One RDC is not yet fully in service and could thus not provide any information on data users; one RDC did not provide any data.

In 2018, the number of external data users increased by 9,081 persons. Twenty-six RDCs were able to provide data on additional data users in 2018. Overall, this means that the number of new data users in 2018 might be higher.

Data use agreements
On 31 December 2018, 29 RDCs reported 28,640 existing data use agreements in the German Data Forum's (RatSWD) research data infrastructure. Thirty-two RDCs signed 5,112 new data use agreements in 2018.

Thirteen RDCs reported that each dataset required an individual data use agreement. Twenty RDCs allow for contracts covering several users. Fifteen RDCs restrict use to specifically stated individuals. On average, contracts provided access to 1.9 persons. Contracts of 10 RDCs are signed at a project-level; six RDCs draw up institution-level contracts.

At 21 RDCs, data use agreements govern access to several datasets. In eleven cases, only one person may use them; nine RDCs extend access to include several people. On average, contracts covered 1.7 persons. Project level contracts were used by 12 RDCs; six RDCs offered institution-level contracts. This facilitated a much more differentiated picture of the number of data users in 2018. Going forward, it will be possible to give an even more differentiated account of these developments in the future.
5 Services for Users

All RDCs offer customer support tailored to the needs of external researchers, which is provided by designated RDC staff via telephone, email, or face to face (see Fig. 15). RDC staff are thus entrusted with the complex task of offering customised support to researchers. This service is made possible by the fact that RDC staff use their in-house datasets to conduct independent research and are thus uniquely familiar with the data’s analytic potential. Help with specific in-depth issues is provided by other in-house specialists.

Additionally, RDCs offer workshops, training, conferences, and FAQs. Some RDCs also offer skills development opportunities (training, workshops, etc.) for data users to use the whole analytic potential of specific datasets (see Chapter 6). Lastly, some RDCs provide online support such as online tutorials or discussion forums.

Fig. 15: User services at RDCs
Which services for user does your RDC provide?
(Multiple answers possible)

- User support (phone, e-mail, face-to-face)
- Access to tools (e.g., codebooks, variable descriptions, syntax files)
- Workshops/seminars
- Conferences
- FAQs
- Trainings (e.g., summer/winter schools)
- Online courses (e.g., MOOCs*, webinars)

All RDCs focus on tailored user support.

© RatSWD 2019

* Massive Open Online Courses
Service quality assurance

Most RDCs use process-integrated measures to assure the quality of their services. The most important measure is the feedback from users, which is either given without prompting and channelled directly back into RDC processes or deliberately collected, for example, through user surveys or feedback rounds.

Only six RDCs do not evaluate their services at all. Depending on RDC policy, user surveys are conducted either after data use or at fixed regular intervals. Insights gleaned from user survey results are used by the RDCs to further develop their offered services. Other mechanisms include integrated user feedback through a variety of user boards or committees, feedback from external experts, and an informal exchange among RDC staff. RDCs place great importance on implementing user feedback and continually aim at improving the organisation of events and their service offering. The results of user surveys are discussed internally on a regular basis; some RDCs also document this process.

Data access paths and data formats

Access to sensitive data for researchers is made possible by data use agreements, which spell out the scope of data use and the specific requirements users must comply with when handling the data. The RDCs monitor whether data use is limited to scientific purposes.

Fig. 16: Data access paths offered by RDCs

Which data access paths does your RDC offer?
(Multiple answers possible)

Data downloads are the most commonly used access path in 2018.
To date, 29 RDCs facilitate off-site data access by downloading datasets, most often in the SUF format. The amount of RDCs offering guest researcher workstations with specific safety measures for on-site data use saw an increase from 24 to 28 on the previous year (see Fig. 16). Guest researcher workstations enable researchers to access sensitive data (see Fig. 1 for a map of guest researcher workstations in Germany). Data access via (controlled) remote access is offered by 14 RDCs, for example via a remote desktop or by submitting analysis scripts (remote execution). The RDCs expanded their off-site services in 2018. Data transmission via email is available at six RDCs and, at 12 RDCs, via data storage mediums sent by regular mail. Some RDCs also make data available in the cloud or via other online platforms.

In addition to providing data for on-site use via guest research workstations, a core task of RDCs is to provide data for off-site use. Where data access via download is permitted, SUFs are the preferred format. PUFs and CFs are made available for downloading by half of the RDCs (see Fig. 17). Overall, 2018 saw an increase in off-site data use at RDCs. For more information on data access paths and data access formats, see Info Box 1 on page 27.
### Provision of tools

**Fig. 18: Provision of tools to work with the data**

Which tools do you offer for which data access path?  
(*Multiple answers possible*).

<table>
<thead>
<tr>
<th>Tool Description</th>
<th>Open Access</th>
<th>After registration</th>
<th>After signing a data use contract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset descriptions</td>
<td>28</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Codebooks</td>
<td>25</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Study descriptions</td>
<td>24</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Variable descriptions/documentation</td>
<td>24</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Methodological reports</td>
<td>23</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Survey instruments (e.g., questionnaires, interview guides)</td>
<td>21</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Analysis/user instructions</td>
<td>17</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Value description</td>
<td>15</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Field reports</td>
<td>14</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Quality reports</td>
<td>12</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Variable correspondence list</td>
<td>10</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Code (SPSS syntax files, STATA do-files)</td>
<td>10</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Scales handbooks</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Transcripts</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

| Number of RDCs in 2018 (n=34) |

Most tools for working with the data are open access.

Almost all RDCs provide users with additional information and documentation on their datasets. They vary according to the RDCs and the datasets. Tools for working with the data are open access, enabling researchers to gather information even before accessing the dataset. The most common tools include dataset descriptions, codebooks, variable descriptions, variable documentation, study descriptions, and methodological reports. More than two thirds of these tools are open access.

Common dataset-specific tools include code (e.g., do-files, syntax files), technical metadata descriptions (e.g., XML files), web applications (e.g., metadata search systems), and theory-focused documents, which inform users about the background of certain questions and variables. Access to documentations such as code and transcripts is often restricted until after signing a data use contract.

---

8 Open access means that scientific literature and other material are available openly online.
The use of some tools is restricted until after registration or signing a contract because they can contain sensitive information, which are restricted for data protection reasons or a syntax that is useless without the appropriate dataset. In some cases, data are provided only after formal review to guarantee non-commercial use.

Normally, tools for working with the data are made available online as open access via the RDC websites and can be accessed without prior registration. If the research data are provided to users in a package, this package tends to contain all the available tools.

Moreover, some RDCs offer other support tools, for example, help with citing research data or by providing information on board notes, seating plans, or interview protocols in qualitative research.

**Time period between the application and the transmission of the data**

Data accessible via online access are usually available immediately or within a few hours (for example, after a simple registration process, or via direct download from the data catalogue). Sensitive, less anonymised data may not be downloaded online. In these cases, the time period between signing the data use contract and the transmission of the data by the RDC can range between one hour and several weeks. Users may encounter longer processing times when requesting specially prepared data, or where data release requires special permits, for example, separate review for data use abroad.

---

**Info box 1: Data access paths and data access formats**

**Guest researcher workstations**

These are specially secured workstations at research data centres (RDCs) enabling researchers to access data that are at least formally anonymised. It is common for guest researcher workstations not to have uncontrolled internet access and to disable local saving of files.

**Remote data processing**

This data access path enables researchers to perform data analysis at RDCs without being on location. Researchers submit analysis scripts to a RDC’s staff, which they write at their respective workplaces (sometimes using a structured dataset to test their code). Depending on data sensibility and/or legal regulation, the resulting files are checked at the RDC (output control) before being sent back to the researchers. This method is called remote data processing or remote access.

**Scientific use Files (SUFs)**

SUFs are research datasets that are de-facto anonymous datasets but still have considerable analytic potential.

**Campus Files (CFs)**

Highly anonymised research datasets created for academic teaching purposes.

**Public Use Files (PUFs)**

Anonymised research datasets without use restrictions that can be shared for non-academic purposes.
Publishing the research data

RDCs have an interest in the extensive use of their data. They use a multitude of channels to publicise their data offering (see Fig. 19). Self-promotion is the most important channel, particularly via RDC websites. This is followed by presentations at (international) conferences, training, and workshops, and publications on the available research data. Metadata portals like da|ra and DataCite also play a central role in publishing research data and attracting more users. Widespread use of these services is a promising base for improving the accessibility (and findability) of RDC data for other additional disciplines at the national and international level. Publishing press releases and content on social media is still less widespread.
Skills development

In addition to user services, 14 RDCs also offer at least one skills development measure to train researchers to work with their datasets. While there was an increase in RDCs offering skills development measures such as training events, conferences, and online courses on the previous year, the number of RDCs offering workshops/seminars saw a slight decline. This has affected the absolute number of skills development measures. The overall decrease is a result of the smaller number of workshops and seminars.

Fig. 20: Number of skills development measures at RDCs

(Multiple answers possible)

Most RDCs organise workshops and seminars.
6 Internationalisation and the current development of the research data infrastructure

Internationalisation

International collaborations and cooperation are common in science and research. Research results are increasingly published in international journals, thus giving way to new forms of scientific discourse.

This internationalisation of the research landscape is also everyday practice at research data centres (RDCs). Responding to the demands from within the scientific community, research facilities and research infrastructures are increasingly international by default. This development is paralleled by the international orientation of the RDCs, which, on the one hand, have created data access paths and data documentation to foster research by the global scientific community in Germany and, on the other hand, set up guest researcher workstations abroad, for example, in France, the United Kingdom, and at several universities in the US and Canada.

The RDCs are continually expanding their services for international researchers.

Overall, 25 RDCs facilitate international data access by providing access paths in English. At 24 RDCs, this includes data documentation; 23 also offer use agreements in English. Furthermore, international users are provided with tailored English user support and communication via email and phone. Lastly, the RDCs also offer translated contracts, newsletters, user workshops, conferences, and training events in English.
In addition to supporting the global scientific community with using their data, the RDCs work together with international researchers on specific issues in international research partnerships. In 2018, 13 RDCs reported that they maintained international research partnerships (see Fig. 22), which predominantly included projects, working groups, and other forms of collaboration with European facilities such as universities, RDCs, data archives, and research institutes.

The benchmark for the internationalisation of RDCs is the use of their research data by researchers abroad. The share of international data users varies very strongly across RDCs. In total, 24 RDCs have contacts to international research facilities (see Fig. 23). This underscores the fact that research data from most German RDCs are used abroad and that investing in English metadata is a sound measure.
Innovation and improvement of the research data infrastructure

2018 saw a broad range of innovations and improvements of the research data infrastructure. In summary, most RDCs reported that they continuously expanded their data range and the availability of new datasets.

A striking innovation of 2018 was the increase in data access points at a national and international level – often combined with deploying remote access solutions. Many RDCs invested heavily into expanding the infrastructure. Other areas of improvement included optimising work flows and improving documentation, particularly metadata, data search, and online message boards. Moreover, some RDCs introduced new software for user contract management and for visualisation of spatial and temporal data.

Further developing the research data infrastructure

The following segment cites a range of issues where RDCs indicated a need or an interest in knowledge exchange. Data protection continued to be an important issue throughout the reporting year 2018, with many RDCs indicating an interest in knowledge transfer and support, particularly regarding the implementation of the GDPR. Other key points included exchanging information on use agreements and data transfer agreements, authoritative legal advice on data protection and copyright issues in connection with novel, unconventional data, for example, data generated through web scraping, secondary research data use, and anonymisation measures in light of the new European data protection regulation.

In research data management, five key issues were identified as particularly relevant for the future of RDCs:

1. RDCs wish to increase knowledge sharing in data processing, including issues like methods and tools for data harmonisation, generating synthetic datasets, data documentation, opportunities and challenges of providing and linking survey data with regional indicators, and linking traditional data with sensitive data.

2. Issues related to metadata, including adapting metadata capture for data archiving to the DDI standards, standardised procedures for exchanging previously structured metadata, and metadata capture in general.

3. Issues related to data access, including alternative and innovative data access paths, expanding the use of the RDC-in-RDC approach, remote data processing, (automated) output control, and implementing remote access solutions.

4. RDCs raised the following issues related to qualitative data: processing, archiving, and making available qualitative data and the legal aspects pertaining to these issues.

5. Expanding the research data infrastructure and improving networking among the RDCs continues to play an important role, particularly in issues related to the RDC-in-RDC approach.

The RDCs also wish to foster knowledge exchange regarding the following issues: further development and cooperation in remote data processing, harmonising processes among the RDCs, certification procedures (costs/benefits), opportunities of the use of geocodes, linking information based on georeferenced data, linking datasets from different RDCs, application and contract management, data rating and selection criteria, and interoperability of data on the variable level.
7 Special topic: data protection

The introduction of the European General Data Protection Regulation (GDPR) changed the legal framework for handling personal data across Europe. This has also affected most research data centres (RDCs) in Germany. Many documents, including declarations of consent, contracts, and application forms, had to be modified to meet the GDPR’s new requirements.

This has resulted in an increased need for consultation on behalf of the RDCs and made data protection a key topic of the German Data Forum (RatSWD) in 2018. Thirty-two RDCs reported that they had sought expert advice on GDPR issues. Eleven of them consulted with their in-house data protection commissioner. The in-house legal departments were also able to give support. Home institutions of RDCs increasingly offered training opportunities on the new data protection regulation, which were attended by RDC staff. Some sought advice from external data protection commissioners as well as lawyers. Additionally, the FDI Committee hosted a presentation on the ‘Role of the GDPR and the German implementation laws for RDCs’ in April 2018, which resulted in an intense discussion among RDCs about the GDPR’s relevance and the practical consequences for data anonymisation, data archiving, and storage of user data. Upon looking back on the GDPR’s introduction at its July 2018 meeting, the German Data Forum (RatSWD) summarised that the interpretation of the GDPR tends to vary strongly in its practical application. There were no legally binding precedents at the time.

Several German Data Forum (RatSWD) press releases9 addressed the relevance of the GDPR for RDCs. The FDI Committee provided its members with a handout on post-GDPR ‘de-facto anonymisation’ and ‘personal data’. This measure contributed to making sure that the RDCs were supported and supporting each other on this year’s central issue. The majority of RDCs reported that they benefitted from the FDI Committee network while transitioning to the new regulation. The RDCs cited presentations, discussions within the FDI Committee, and informal cooperation among each other as the most important tools on the way.

---

Twenty-three of the surveyed RDCs reported that legal adjustments regarding data protection were necessary in 2018 (particularly due to changes in contracts and use agreements). Nine RDCs reported that this was not necessary. Many RDCs also changed the way they documented data protection measures and handled user management (see Fig. 24). Only very few RDCs were affected in areas such as data collection, data archiving, and anonymisation measures.

Nine RDCs also function as data repositories. Here, the new data protection regulation changed many things. One RDC had to fully re-examine and partially adjust its technical procedures. Moreover, several RDCs had to examine, update, and adjust declarations of consent, data protection measures, and transfer agreements to the GDPR. At some RDCs, the day-to-day processes hardly changed at all and no adjustments had to be made.

It has also been shown that data protection issues were mostly addressed by data providers. External data users rarely approached the RDCs with questions regarding the changes in data protection regulation. There were a few exceptions pertaining to specific issues of research data use, for example, issues like joint data use by co-authors, sustainable data use, anonymisation, and linking data.

---

10 RDCs that act as a data repository make a data server available as a repository, which external researchers and research projects can use to store and archive their research data for secondary use.
Info box 2: the new European General Data Protection Regulation (GDPR) at the RDCs

The GDPR has been effective since 24 May 2016 and became enforceable in all European Union Member States on 25 May 2018. Due to the regulation’s stark similarities to German data protection law, the effects on scientific research in Germany have been manageable. The GDPR’s research article (Article 89) enshrines the particular social relevance of science and research into data protection law. New documentation and proof obligations were added.

The legal definition of ‘de-facto anonymised data’ in the Federal Statistical Act (old version) has been omitted. This aspect is particularly relevant to RDCs accredited by the German Data Forum (RatSWD). Generally speaking, the access paths provided by RDCs are based on data that are anonymised to an extent that safeguards data protection while maintaining the highest degree of analytic potential for research (see Fig. 25).

The German Data Forum (RatSWD) sought to clarify the new legal situation. The result was that, according to experts, the GDPR’s definition of anonymity is not to be understood in absolute terms. Anonymity encompasses a continuous assessment of the risk of re-identification against the backdrop of technological developments and generally accessible data sources. Consequently, RDCs can continue to apply the concept of ‘de-facto anonymisation.’

Although the GDPR is now enforceable across the EU, the legal practice in science and research will only become more concrete in the following years. This will be based on practical examples and case studies as well as the statements put out by the European Data Protection Board.

---

Fig. 25: Degrees of anonymisation based on risk assessment

© RatSWD 2018
8 Complaints management

One of the main tasks of the German Data Forum (RatSWD) is to assure and improve the quality of research data centres (RDC) services. Since its inception, the RatSWD has acted as a dedicated point-of-contact for complaints relating to RDC data and services.

In addition to overseeing the annual monitoring process, of which the present activities report is one outcome, the German Data Forum (RatSWD) monitoring commission also handles complaints put forward by research data users.

The German Data Forum (RatSWD) set up a complaints office at its business office to professionalise complaints management and make it more transparent. The complaints office ensures a swift and professional response to complaints and feeds the results back into RDC processes to further improve the data infrastructure.

If data users become aware of major shortcomings in the data services of an accredited RDC, it is recommended to first approach the RDC directly to try to find a solution. If the problem cannot be solved, users may direct their concern to the complaints office. The complaints office's mandate is limited to issues concerning compliance with the German Data Forum's (RatSWD) accreditation criteria. The German Data Forum (RatSWD) is not responsible for delays during everyday procedures or for staff conduct at RDCs. Complaints of this nature should be directed to the RDC in question.

For more detailed information about the procedures, see the German Data Forum's (RatSWD) Output 8 (5)\(^{11}\) or the updated version of that output on the German Data Forum (RatSWD) website\(^{12}\).

---


\(^{12}\) https://www.ratswd.de/en/info/complaints-office
Appendix

Appendix A: Development of the RatSWD’s research data infrastructure and RDCs ............... 38
Appendix B: The monitoring commission ................................................................. 44
Appendix C: Contributors to the 2018 Activities Report ............................................. 45
Appendix A

Development of the RatSWD’s research data infrastructure and RDCs

Last update: 06/08/2019

1999

The Commission to Improve the Informational Infrastructure between Research and Official Statistics (KVI) recommends the establishment of research data centres (RDCs).

2001

The German Data Forum (RatSWD) Founding Committee is set up.

The following RDC is founded:

| Research Data Centre of the Federal Statistical Office (FDZ-Bund)* |
| Germany-wide access to official statistics microdata from the following fields: population, education, health, business, agriculture, environment, administration of justice, finance, and taxes. |
| www.forschungsdatenzentrum.de/en |

2002

The following RDC is founded:

| Research Data Centre of the Statistical Offices of the Länder (FDZ-Länder)* |
| Germany-wide access to official statistics microdata from the following fields: population, education, health, business, agriculture, environment, administration of justice, finance, and taxes. |
| www.forschungsdatenzentrum.de/en |

2003

The following RDCs are founded:

| Research Data Centre German Microdata Lab at GESIS (RDC GML)* |
| https://www.gesis.org/en/gml/gml-home |

| International Data Service Centre at the Institute for the Study of Labour (FDZ IZA, IDSC)* |
| National and international labour market datasets with standardised information. Research with, methods and resources for using online data for labor economics and social science. Development of tools and methods for remote access (statsdirect.org) and remote processing (JoSuA). |
| https://www.iza.org/organization/idsc |

2004

The German Data Forum (RatSWD) is founded.

The following RDCs are founded:

| Research Data Centre of the German Federal Employment Agency at the Institute for Employment Research (FDZ BA at IAB)* |
| Data on persons, households, and employers, as well as combined datasets consisting of survey data and administrative research data in the fields of social security and labour market, and employment research. |
| https://fdz.iab.de/en.aspx |
**Research Data Centre of the German Pension Insurance (FDZ-RV)**

Data on the insurance accounts of individuals insured in the Federal Pension Insurance. The accounts contain data on the insured persons' insurance history and the pension and rehabilitation benefits they received.

https://www.eservice-drv.de/FdzPortalWeb/dispcontent.do?id=main_fdz_english

---

**The following RDCs are accredited:**

2008

- **Research Data Centre of the Federal Institute for Vocational Education and Training (BIBB-FDZ)**
  
  Firm-level and individual-level datasets of vocational education research dealing primarily with the attainment and use of vocational knowledge and skills.
  
  www.bibb.de/en/53.php

- **Research Data Centre of the Institute for Educational Quality Improvement (FDZ IQB)**
  
  German datasets from the major national and international school performance studies and national studies measuring educational standards.
  
  www.iqb.hu-berlin.de/fdz

2009

- **Establishment of the Standing Committee Research Data Infrastructure (RDC Committee) of the RatSWD**
  
  The following RDCs are accredited:

- **Research Data Center of the Socio-Economic Panel Study at DIW Berlin (RDC SOEP)**
  
  Data from representative annual surveys of private households. The SOEP-CORE sample features topics such as income, employment, education, and health. In addition, there is the longitudinal innovative sample (SOEP-IS), which enables external researchers to contribute research projects of their own.
  
  https://www.diw.de/en/diw_02.c.222518.en

- **Research Data Centre of the Survey of Health, Ageing and Retirement in Europe (FDZ SHARE)**
  
  Data from the multidisciplinary and cross-national panel study "Survey of Health, Ageing and Retirement in Europe" (SHARE), which produces microdata on health, socio-economic conditions, and social and family networks of approximately 123,000 individuals aged 50 or older in more than 20 European countries and Israel. The seventh wave of SHARE was collected in 2017.
  
  http://www.share-project.org/data-access.html

- **Research Data Centre International Survey Programmes at GESIS (RDC International Survey Programmes)**
  
  Internationally comparative survey data from more than 70 countries on nearly all social science topics: Comparative Study of Electoral Systems (CSES), European Values Study (EVS), Eurobarometer, European Election Studies (EES), International Social Survey Programme (ISSP).
  

---

* The research data centres Federal Statistical Office, Statistical Offices of the Länder, GML, IZA, BA at IAB, and RV were established prior to the foundation of the German Data Forum (RatSWD) and became part of the research data infrastructure in 2004. In these cases, the year of the RDCs' foundation is listed. All other RDCs were accredited after 2004 by the German Data Forum (RatSWD). With these RDCs, the year provided is the year of their accreditation.
Research Data Centre Elections at GESIS (RDC Elections)
Access to German national election surveys (federal elections and state elections), Politbarometer, Forsa-Bus, ARD Deutschlandtrend and Surveys for the Federal Government. The RDC's largest project at this point is the German Longitudinal Election Study (GLES).
https://www.gesis.org/en/elections-home/elections-home

Research Data Centre ALLBUS at GESIS (RDC ALLBUS)
Data from the Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) and German General Social Survey (GGSS) in English, on the attitudes, behaviours, and social structure of the German population.
https://www.gesis.org/en/allbus/allbus-home

Akkreditierung folgender FDZ:

Research Data Centre for Business and Organizational Data (FDZ-BO)
Quantitative and qualitative business, organizational data, linked employer and employee data, and data from employee and member surveys.
http://www.fdz-bo.diw.de

Research Data Centre of the German Centre of Gerontology (FDZ-DZA)
Data from the long-term German Ageing Survey (DEAS) on the changing life situations and ageing processes of people in mid- and older adulthood, and from the German Survey on Volunteering (FWS), a representative survey programme with a focus on voluntary activities and civic participation in Germany.
www.dza.de/en/fdz.html

Research Data Centre PsychData of the Leibniz Institute for Psychology Information (FDZ PsychData at ZPID)
Pooled quantitative datasets from both basic research and applied psychology; data archiving with a focus on longitudinal studies, large-scale survey studies, and development testing.
www.psychdata.de/index.php?main=none&sub=none&lang=eng

Research Data Centre of the German Family Panel (FDZ pairfam)
Datasets from the ’Panel Analysis of Intimate Relationships and Family Dynamics’ (pairfam), a representative, interdisciplinary longitudinal study for the analysis of private living arrangements in Germany.
www.pairfam.de/en

Research Data Centre Ruhr at the RWI – Leibniz Institut for Economic Research (FDZ Ruhr at RWI)
Specialisation on regional data: socioeconomic data measured by 1 square km grids. Aside from geo-referencing data on a scientific basis, the RDC provides various individual-level and employer-level data collected in RWI research projects.
http://en.rwi-essen.de/forschung-und-beratung/fdz-ruhr
### 2011

The following RDCs are accredited:

<table>
<thead>
<tr>
<th>RDC Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMU-ifo Economics &amp; Business Data Center (EBDC)</td>
<td>Datasets of German companies, including survey data collected by the ifo Institute on firms’ business status, innovativeness, and investment behaviour, as well as external data on corporate financing and governance structure. Merged panels of the aforementioned two data sources are also available. <a href="https://www.ifo.de/en/EBDC">https://www.ifo.de/en/EBDC</a></td>
</tr>
<tr>
<td>Research Data Centre of the Robert Koch Institute (RDC RKI)</td>
<td>Data on the state of health and health-related behaviour of Germany’s resident population, collected on the basis of nationally representative studies. <a href="https://www.rki.de/puf">https://www.rki.de/puf</a></td>
</tr>
<tr>
<td>Research Data Centre of the Federal Centre for Health Education (FDZ BZgA)</td>
<td>Data from nationally representative surveys, repeated at regular intervals, measuring the population's susceptibility to health education and prevention campaigns, as well as the knowledge, attitudes, and behaviour in the general population concerning the health issues addressed by BZgA. <a href="https://www.bzga.de/home/bzga">https://www.bzga.de/home/bzga</a></td>
</tr>
</tbody>
</table>

### 2012

The following RDCs are accredited:

<table>
<thead>
<tr>
<th>RDC Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Data Center Wissenschaftsstatistik of the Stifterverband (RDC Wissenschaftsstatistik)</td>
<td>Data on the research and development activities of German companies, on the financial volume, structure, and regional distribution of research and development activities (R&amp;D), and on R&amp;D staff in the business sector. <a href="https://www.fdz-wissenschaftsstatistik.de">https://www.fdz-wissenschaftsstatistik.de</a></td>
</tr>
<tr>
<td>Research Data Centre Education (FDZ Bildung) at the German Institute for International Educational Research (DIPF)</td>
<td>The hosted datasets include approaches of qualitative educational research such as video data, transcriptions, contextual materials and survey tools of quantitative educational research such as questionnaires and assessment tests. The collected datasets refer to the quality of instruction and to the quality of schools. <a href="https://www.fdz-bildung.de">https://www.fdz-bildung.de</a></td>
</tr>
<tr>
<td>Research Data Center of the Leibniz Institute for Educational Trajectories at the University of Bamberg (RDC-LiBi)</td>
<td>Longitudinal data from the National Educational Panel Study (NEPS), which was launched in 2010 with more than 60,000 panel participants in six starting cohorts to study skills formation, educational processes, educational decisions, and educational returns in formal, non-formal, and informal contexts across the lifespan. <a href="http://www.lifbi.de/en-us/home.aspx">www.lifbi.de/en-us/home.aspx</a></td>
</tr>
<tr>
<td>ZEW Research Data Centre for European Economic Research (ZEW-FDZ)</td>
<td>The ZEW-FDZ provides microdata from ZEW firm surveys on innovation activities, the development of young firms, the use of information and communication technologies, and further topics. Data from individual and expert surveys are also accessible – for example, the ZEW Financial Market Survey. <a href="https://kooperationen.zew.de/en/zew-fdz/home.html">https://kooperationen.zew.de/en/zew-fdz/home.html</a></td>
</tr>
</tbody>
</table>
### 2013

<table>
<thead>
<tr>
<th>The following RDCs are accredited:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Data Centre of the German Youth Institute (FDZ-DJI)</td>
</tr>
<tr>
<td>Data from the surveys on children and young people growing up and the life situations of adults and families, conducted in regular intervals since 1988.</td>
</tr>
<tr>
<td><a href="http://www.dji.de/en/the-dji.html">www.dji.de/en/the-dji.html</a></td>
</tr>
</tbody>
</table>

### FDZ SFB 882

Qualitative and quantitative datasets from inequality research

The RDC was discontinued in 2016. Depending on the data type and basis, the data of RDC SFB 882 were transferred to different organisations: IAB data were handed over to FDZ BA im IAB; qualitative data with an organisational connection were handed over to FDZ BO; the remaining data were handed over to the SOBI archive at the University of Bielefeld (currently under development). (Last update on 09/14/2017)

https://sfb882.uni-bielefeld.de/de/fdz-sfb882.html

### 2014

<table>
<thead>
<tr>
<th>The following RDCs are accredited:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Data Center Archive for Spoken German at the Institute for the German Language (FDZ AGD)</td>
</tr>
<tr>
<td>Data on spoken German in interactions (conversation corpora) and data on domestic and non-domestic varieties of German (variation corpora).</td>
</tr>
<tr>
<td><a href="http://agd.ids-mannheim.de">http://agd.ids-mannheim.de</a></td>
</tr>
</tbody>
</table>

| Research Data Center Programme for the International Assessment of Adult Competencies (PIAAC) at GESIS (RDC PIAAC) |
| German and international data of the Programme for the Assessment of Adult Competencies (PIAAC). For Germany, additional regional data and longitudinal data are available. |
| www.gesis.org/en/piaac/rdc |

### 2015

<table>
<thead>
<tr>
<th>The following RDCs are accredited:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deutsche Bundesbank Research Data and Service Centre (RDSC Bundesbank)</td>
</tr>
<tr>
<td>Various datasets on banks, securities, investment funds, and enterprises, as well as combinations of those; data from the Panel on Household Finances, a representative study on the structure and composition of households’ wealth.</td>
</tr>
</tbody>
</table>

<p>| Research Data Centre of the Halle Institute for Economic Research (RDC-IWH) |
| Company data from panel studies and longitudinal studies on development trends in East Germany's manufacturing and construction sectors, as well as on the choice of locations of multinational companies in East Germany and in Central and Eastern Europe. |</p>
<table>
<thead>
<tr>
<th>Year</th>
<th>Accredited RDCs</th>
</tr>
</thead>
</table>
| 2017 | Research Data Centre for Higher Education Research and Science Studies (fdz.DZHW)  
Quantitative and qualitative research data from the field of higher education and science studies, especially the DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification (Studienberechtigtenpanel), the DZHW Graduate Panel (Absolventenpanel), the DZHW Social Survey, and the DZHW Science Survey.  
[https://fdz.dzhw.eu/en](https://fdz.dzhw.eu/en) |
| 2018 | Research Data Centre eLabour (FDZ eLabour)  
Qualitative data from studies in industrial and occupational sociology with a focus on the changing nature of work, including open and semi-standardised interviews, observations, and expert interviews.  
[http://elabour.de](http://elabour.de) |
| 2019 | Research Data Centre Qualiservice (FDZ Qualiservice)  
Qualiservice focuses on archiving, curating and providing qualitative research data from a range of disciplines. Its secure, flexible, and research-oriented services include processing primary qualitative studies for secondary use, comprehensive user support, long-term preservation, and the provision of archived research data as well as relevant context information.  
[https://www.qualiservice.org](https://www.qualiservice.org)  
Research Data Centre at Kraftfahrt-Bundesamt (FDZ at KBA)  
(Provisional accreditation)  
The Forschungsdatenzentrum im Kraftfahrt-Bundesamt (FDZ im KBA) makes the anonymised microdata of the Register of Driver Fitness (Fahrzeigmzinsregister, FAER) available to researchers as scientific use files.  
[https://www.kba.de/DE/Statistik/Forschungsdatenzentrum/forschungsdatenzentrum_node.html](https://www.kba.de/DE/Statistik/Forschungsdatenzentrum/forschungsdatenzentrum_node.html) |
| Guest status is granted to the following RDCs (Last update: August 2019): |
| Research Data Centre of the Hans-Böckler-Foundation at the WSI (RDC WSI)  
The RDC WSI provides access to the data of the WSI Works Councils Survey until 2011.  
[https://www.boeckler.de/wsi_51469.htm](https://www.boeckler.de/wsi_51469.htm) |
| Research Data Centre of the German Centre for Integration and Migration Research (DeZIM.fdz)  
DeZIM.fdz organises access to research data collected by the German Centre for Integration and Migration Research. In addition, DeZIM.fdz provides comprehensive support for working with these data and a range of other key methodological issues.  
[https://www.dezim-institut.de](https://www.dezim-institut.de) |
Appendix B

The monitoring commission

For quality assurance purposes, the German Data Forum (RatSWD) agreed to establish a monitoring commission in July 2016. Its main task is to collect and assess the regular reports handed in by the RDCs. Moreover, the commission monitors compliance with the obligations arising from provisional accreditation. The FDI Committee elects the commission from its own membership for a three-year term concurrent with the German Data Forum (RatSWD) appointment period. The commission thus enjoys a special level of trust and legitimacy. It consists of four members of the FDI Committee and two deputy members (to replace elected members, if required) and the German Data Forum (RatSWD) chairpersons sit in as guests.

Members of the monitoring commission

**Maurice Brandt** (Chair since May 2019)
Research Data Centre of the Federal Statistical Office (FDZ-Bund)

**Dr. Lea Eilers** (Chair from July 2018 – September 2018)
Research Data Centre Ruhr at the RWI – Leibniz Institute for Economic Research (FDZ Ruhr at RWI)

**Dr. Cornelia Lang** (Mitglied seit April 2019)
Forschungsdatenzentrum des Leibniz-Instituts für Wirtschaftsforschung Halle (FDZ-IWH)

**Holger Quellenberg** (Chair from October 2018 – April 2019)
Research Data Centre of the German Youth Institute (FDZ-DJI)

**Dr. Karsten Stephan** (Member until December 2018)
Research Data Centre for Higher Education Research and Science Studies (fdz.DZHW)

Deputy members of the monitoring commission

**Dr. Sandra Gottschalk**
ZEW Research Data Centre for European Economic Research (ZEW-FDZ)

**Tatjana Mika**
Research Data Centre of the German Pension Insurance (FDZ-RV)

Standing guests of the monitoring commission

**Prof. Regina T. Riphahn, Ph.D.**
Chair of the German Data Forum (RatSWD)

**Prof. Stefan Bender**
Vice chair of the German Data Forum (RatSWD)
Appendix C

Contributors to the 2018 Activities Report

**Maurice Brandt** (Vorsitz)
Research Data Centre of the Federal Statistical Office (FDZ-Bund)

**Dr. Lea Eilers**
Research Data Centre Ruhr at the RWI – Leibniz Institute for Economic Research (FDZ Ruhr at RWI)

**Dr. Cornelia Lang**
Research Data Centre of the Halle Institute for Economic Research (RDC-IWH)

**Holger Quellenberg**
Research Data Centre of the German Youth Institute (FDZ-DJI)

**RatSWD Business Office**
Marie Bormann
Dr. Mathias Bug
Aktuelle Entwicklungen der Forschungsdateninfrastruktur in den SVW
This report is the result of a project that is funded by the Federal Ministry for Education and Research (reference number: 01UW1402). Unless otherwise stated, the responsibility for this publication lies with the German Data Forum (RatSWD).

doi: 10.17620/02671.44

Citation suggestion:
Der Rat für Sozial- und Wirtschaftsdaten (RatSWD) berät seit 2004 die Bundesregierung und die Regierungen der Länder in Fragen der Forschungsdateninfrastruktur für die empirischen Sozial-, Verhaltens- und Wirtschaftswissenschaften. Im RatSWD arbeiten acht durch Wahl legitimierte Vertreterinnen und Vertreter der sozial-, verhaltens- und wirtschaftswissenschaftlichen Fachdisziplinen mit acht Vertreterinnen und Vertretern der wichtigsten Datenproduzenten zusammen.

Er versteht sich als institutionalisiertes Forum des Dialoges zwischen Wissenschaft und Datenproduzenten und erarbeitet Empfehlungen und Stellungnahmen. Der RatSWD engagiert sich für eine Infrastruktur, die der Wissenschaft einen breiten, flexiblen und sicheren Datenzugang ermöglicht. Solche Daten werden von staatlichen, wissenschaftsgetragenen und privatwirtschaftlichen Akteuren bereitgestellt. Der RatSWD hat 34 Forschungsdatenzentren akkreditiert, deren Kooperationen er fördert.