

German Data Forum (RatSWD) (Ed.)

Research Report

The German Data Forum (RatSWD) and Research Data Infrastructure: Status Quo and Quality Management

RatSWD Output, No. 1 (6)

Provided in Cooperation with:

German Data Forum (RatSWD)

Suggested Citation: German Data Forum (RatSWD) (Ed.) (2018) : The German Data Forum (RatSWD) and Research Data Infrastructure: Status Quo and Quality Management, RatSWD Output, No. 1 (6), Rat für Sozial- und Wirtschaftsdaten (RatSWD), Berlin, <https://doi.org/10.17620/02671.30>

This Version is available at:

<https://hdl.handle.net/10419/224191>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

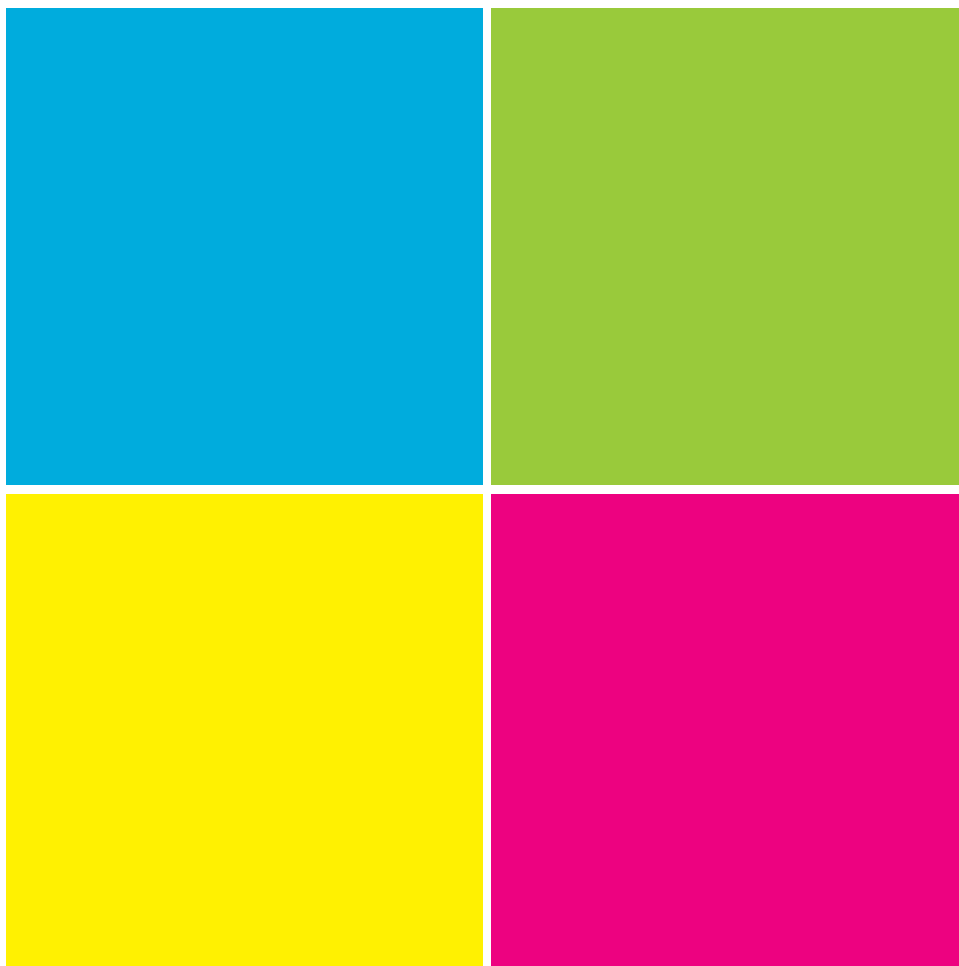
Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The German Data Forum (RatSWD) and Research Data Infrastructure: Status Quo and Quality Management



SPONSORED BY THE



Federal Ministry
of Education
and Research

German Data Forum (RatSWD)

The German Data Forum (RatSWD) and Research Data Infrastructure: Status Quo and Quality Management

Contents

| | | |
|--|---|-----------|
| 1 | Overview of the research data infrastructure of the German Data Forum (RatSWD) . . . | 6 |
| 1.1 | About the German Data Forum (RatSWD) | 6 |
| 1.2 | The German Data Forum (RatSWD) research data infrastructure | 6 |
| 1.3 | Responsibilities of the research data centres (RDCs) | 7 |
| 2 | Quality management | 9 |
| 2.1 | Self-organisation of the research data infrastructure as a model of success | 9 |
| 2.2 | Accreditation of research data centres (RDCs) | 9 |
| 2.3 | Monitoring and evaluation | 11 |
| 2.4 | Complaints office and procedures | 12 |
| 3 | Facts and figures about the current activities of the accredited RDCs | 13 |
| 3.1 | Structure of the research data centres (RDCs) | 13 |
| 3.2 | Storage and quality assurance | 14 |
| 3.3 | Data provided | 15 |
| 3.4 | Services provided | 16 |
| 4 | International orientation of the research data infrastructure | 19 |
| Appendix A: At a glance 1: The network of the German data infrastructure. | | 23 |
| Appendix B: At a glance 2: Bottom-up governance to secure the quality of the research data infrastructure. | | 24 |
| Appendix C: Mandatory and information criteria for research data centres (RDCs) to gain accreditation from the German Data Forum (RatSWD) | | 25 |
| Appendix D: Development and description of the German Data Forum (RatSWD) research data infrastructure | | 28 |

1 Overview of the research data infrastructure of the German Data Forum (RatSWD)

1.1 About the German Data Forum (RatSWD)

Established in 2004, the German Data Forum (Rat für Sozial- und Wirtschaftsdaten, RatSWD) is an independent council advising the German federal and state governments with regard to expanding and improving the research data infrastructure for the empirical social, behavioural, and economic sciences. It provides an institutionalised forum for researchers and data producers to discuss access to high-quality and scientifically relevant data.

For the research fields represented by the German Data Forum (RatSWD), it is important that public and private actors make data about their operations and activities available for scientific inquiry. These data are a unique reflection of the societal and economic status quo and its evolution over time. However, major efforts are sometimes necessary to convince the institutions behind the data collections to view these data as scientifically interesting research data that can help society move forward. This kind of advocacy work is done at the German Data Forum (RatSWD), and it continues to be a major challenge even now, more than ten years since the forum was established. Another core responsibility of the German Data Forum (RatSWD) is to help ensure that research data from research projects and long-term studies (such as panel studies), which are often publicly funded, are made available to a large number of researchers.

That is why the German Data Forum (RatSWD) consists of eight elected representatives of the social, behavioural, and economic sciences and eight representatives of the most important data producers. The equal representation of data users and producers and the strong will to work together to face the present and future challenges make the German Data Forum (RatSWD) a body offering both depth and breadth of expertise. It plays a key role in further developing research infrastructures for the empirical social, behavioural, and economic sciences and in making the general legal and political environment more research-friendly.

To achieve these goals, the German Data Forum (RatSWD) advises the establishment of research data centres (RDC) as a model solution for flexible and comprehensive access to sensitive data for science and research. Since their inception, the RDCs have played a pivotal role in the social, behavioural, and economic sciences and strengthened the international competitiveness of the German research landscape. The German Data Forum (RatSWD), as of 2017, has accredited 31 RDCs. The RDCs coordinate their activities to optimise their services (see also the following chapter). The German Data Forum (RatSWD) has played a major role in supporting the development of these model solutions. They have led to a surge in accessible research data in the past decade.

1.2 Research data infrastructure

Partly due to data sensitivity and respective legal regulations, the data offered by the RDCs must remain at their respective data producing institutions. Therefore, the decentralised structure of the research data infrastructure is a tried and tested way to satisfy the demands of data producers, data users in science and research, and data protection (see also Appendix D for an overview over the RDCs accredited by the German Data Forum and about the available data).

At each RDC, researchers have access to data, including to official statistics data and data that are relevant to specific disciplines and research areas. The RDCs established so far provide a central infrastructure for social science, behavioural science, and economic research and help increase the competitiveness of the German research landscape. The reasons for setting up an RDC can

be diverse: Some RDCs were established based on external recommendations; others were created on data producers' own initiative to promote research and to help build a network of public data producers. Whatever the motivation, all RDCs share the desire to expand and strengthen the research data infrastructure in Germany by offering access to high quality data. As RDCs guarantee at least de facto anonymity of their sensitive personal data, they can continue to allow researchers access to their data under German data protection legislation. For all RDCs, being accredited by the German Data Forum (RatSWD) not only means a seal of standards and quality but also yields additional benefits, including the opportunity to network with other accredited RDCs, to receive best practice solutions for improving their own infrastructures, and to participate in the exchange of knowledge and experience among RDCs.

To foster a productive dialogue among the RDCs, the "Standing Committee Research Data Infrastructure" (FDI Committee) was established in 2009. The committee's primary task is to safeguard and improve the research data infrastructure on an ongoing basis, that is, the quantity and quality of the data, their availability, and their accessibility (for example, metadata, access ways). As a central and open committee, it represents all accredited RDCs, for instance, by jointly formulating the interests of the RDCs and communicating them to funding agencies, policy makers, and research representatives, as well as data protection authorities and ethics commissions at the national and international level. In its capacity as an advisory board, the FDI Committee supports the German Data Forum (RatSWD) by complementing the latter's strategic focus with a focus on the day-to-day challenges of the research data infrastructure. The FDI Committee's decisions are forwarded to the German Data Forum (RatSWD) as recommendations. As a result, a parallel structure of two bottom-up initiatives has emerged: The German Data Forum (RatSWD) as a political advisory board made up in equal parts of data users and data producers and the FDI Committee, made up of representatives of the individual RDCs, at the infrastructure's operative level. The joined efforts of these two components have significantly expanded the research data infrastructure, guided by the concrete needs of researchers and public data producers.

1.3 Responsibilities of the research data centres (RDCs)

RDCs are accredited by the German Data Forum (RatSWD) according to standardised and transparent criteria. This procedure safeguards high-quality data access and strict adherence to data protection regulation. Moreover, the criteria enable the objective evaluation of the operations of the RDCs. RDCs fulfil a range of important tasks for the scientific community. Their main responsibilities are:

- **Providing researchers with user-friendly, transparent, and high-quality access to data.**
So far, this data has concerned mostly microdata that can be analysed statistically. The data are collected as part of official statistics, administrative operations, research projects, or scientific survey programmes. In their capacity as mediators, the RDCs help improve cooperation between data users and data producers.
- **Ensuring that data users comply with federal data protection policies and, if applicable, with policies specific to individual research areas, by taking appropriate technical and organisational measures.**
Depending on the level of anonymisation (see info box 1), datasets are offered for off-site use (via download or mail order) in the form of Scientific Use Files (SUF), Public Use Files (PUF), or Campus Files (CF). Moreover, the generation of synthetic data can be an option to support research needs. To facilitate access to highly sensitive microdata, the RDCs offer the option of on-site use. In this case, users can access the data at a guest researcher workstation on the premises of the RDC.
- **Ensuring equal treatment of all data users by means of transparent and standardised application and access policies.**
Incoming applications are not assessed with regard to the content of the proposed research; they are only reviewed in terms of their compliance with contractual or data protection policies.

- Creating easy-to-analyse data products featuring quality-assured, standard-compliant metadata and comprehensive documentation.

The RDCs present information on their respective data services via their websites, in data and method reports, at scientific events, or in individual advising sessions.

- Conducting independent research using the data they offer.

This helps ensure that each RDC has strong expertise regarding the data and their quality. At the same time, ongoing scientific discussions about methods and contents can inform the advising services provided to data users. Research activities by RDC staff do not involve any exclusive access to data products.

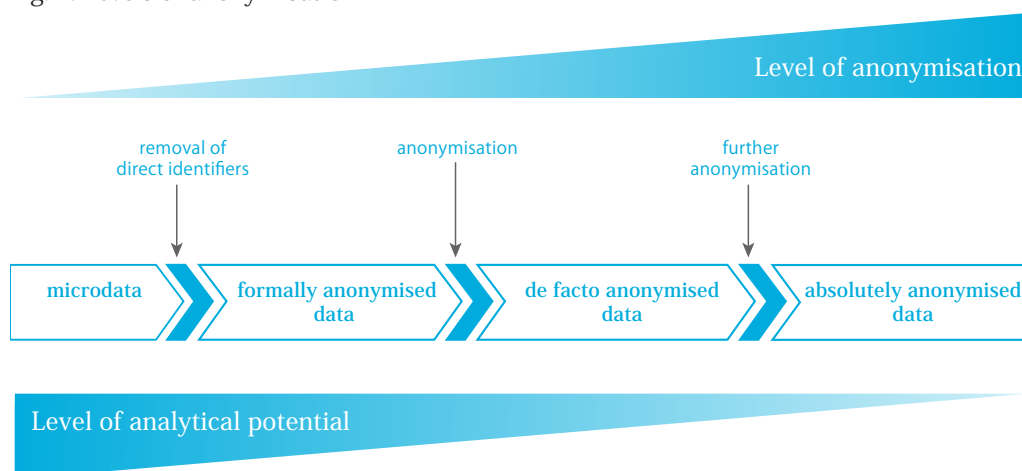
Info box 1: Anonymisation concepts

To comply with the legal requirements regarding anonymisation while at the same time preserving the analytical potential of the data for researchers in the best way possible, the research data centres pursue a variety of jointly agreed concepts regarding data anonymisation and data access. The basic rule is: the more restrictive the policies for accessing a given dataset, the less rigid the anonymisation measures typically applied to that dataset. In such cases, the RDC in question may provide guest researcher workstations featuring special security measures. Anonymisation measures range from formal anonymisation (i.e. deletion of direct identifiers such as names and addresses) to de facto anonymisation (i.e. individual information can be re-attributed only with disproportionate effort) and absolute anonymisation (i.e. re-identification is practically impossible).

To create formally anonymised data, direct identifiers such as names and addresses are either deleted or stored separately from other data and replaced with pseudonyms.

Anonymisation measures leading to de facto or absolute anonymisation as defined in the relevant German laws and regulations, include the deletion, the imputation, and the aggregation of variables with a high re-identification potential such as regional indicators, respondents' age, or information on business sectors. Subsampling is another method to ensure this kind of anonymisation.

Fig. 1: Levels of anonymisation



2 Quality management at the RDCs

2.1 Self-organisation of the research data infrastructure as a model of success

A decentralised network of now 31 RDCs (as of January 2018) accredited by the German Data Forum (RatSWD) provides researchers with low-cost and simple access to a wide range of research data. See Appendix D for an overview of the development and description of the German Data Forum (RatSWD) research data infrastructure. What all of these RDCs have in common is that they also offer access to sensitive individual data on natural persons, households, or businesses.

Depending on the content or the unit of observation, making such data available is subject to various legal requirements, most notably by the EU General Data Protection Regulation and the German Federal Data Protection Act but also state data protection legislation, the German Social Code, and the Federal Statistics Act.

The RDCs make an important contribution to ensuring compliance with the legal and regulatory requirements. At the same time, they take care to preserve the data's analytical potential for researchers and their interests in the best way possible.

2.2 Accreditation of research data centres (RDCs)

To ensure that the RDC are able to fulfil the tasks outlined in paragraph 1.3 and to safeguard the quality of the research data infrastructure, the German Data Forum (RatSWD) has developed standards and criteria for accrediting RDCs. Since 2015, the RDCs have been monitored and evaluated on a yearly basis.

Basic prerequisite of accreditation

One basic prerequisite for an RDC to be accredited by the German Data Forum (RatSWD) is that the RDC can provide proof of being fully operational.

An RDC is considered to be fully operational if it

- has been in operation for at least six months and
- can provide proof of a minimum of three external data users. Acceptable proof includes data user contracts or research requests based on the analysis of the RDC's data.

As accreditation may be relevant for RDCs to become eligible for external funding, there is also the option of applying for accreditation even before or when becoming fully operational. Accreditation prior to becoming fully operational may be granted on a conditional basis – in the form of provisional accreditation.

Mandatory criteria for accreditation

The three mandatory criteria that all RDCs accredited by the German Data Forum (RatSWD) have to meet and report on as part of the annual monitoring process¹ are:

- They must provide at least one data access path.
- They must provide sufficient data documentation.
- They must have a strategy in place to ensure the long-term availability of the data.

Access barriers requiring additional administrative efforts and resources at the RDCs are undesirable because otherwise scientific data collections may not even be made available in the first place. That is because as long as there is no external funding to pay for making data available, doing so means extra work at many RDCs – work that is still not considered a must but is generally worthy of recognition.²

Information criteria for accreditation

For first-time accreditation, each RDC has to answer questions about additional information criteria³ that all accredited RDCs must answer in the annual monitoring process as well.

These criteria are designed to give the German Data Forum (RatSWD) an overview of the activities already completed, the activities planned for the future, and the quality of services. The focus is on the following aspects:

- Scope and development of the social, behavioural, and economic data provided
- Method for timely data provision
- Provision of tools
- Quality assurance of datasets
- Data protection safeguards in due consideration of the interests of researchers
- Service concept
- Single entity comprising institution and research data centre
- Provision of all datasets relevant to research
- Overlap and distinct features compared to existing RDCs
- Research activities
- Multiple provision of the same data (multiple hosting, not hosting at multiple sites)
- Time to process applications
- Staff

Procedure for first-time accreditation

If an RDC applies for the German Data Forum (RatSWD) accreditation for the first time, it begins by presenting its data and its current (or planned) activities to the FDI Committee. After reviewing the application, the FDI Committee submits a recommendation to the German Data Forum (RatSWD).

- If the applicant RDC fails to meet the three mandatory criteria, the application for accreditation is turned down (for the time being).
- If the RDC is fully operational, meets the three mandatory criteria, and does not show any major deficiencies regarding the information criteria, accreditation is recommended with immediate effect.
- If the RDC is fully operational, meets the three mandatory criteria but shows major deficiencies regarding the information criteria, the German Data Forum (RatSWD), after checking with the RDC, looks into whether these deficiencies can be addressed in the short or medium term. Other RDCs may offer support and consultations to help improve the situation. Upon improvement, the

1 See Appendix C for information on the concrete operationalisation of these criteria for the purpose of accreditation and annual monitoring.

2 An alternative to making research data available through an RDC is to pass them on to a data archive or other facility.

3 The mandatory and information criteria, as well as their respective operationalisation, are listed in Appendix C.

German Data Forum (RatSWD) can recommend provisional accreditation, which is subject to certain obligations.

- Provisional accreditation may likewise be recommended if the applicant RDC is not yet fully operational but will become so in the near future and if the RDC can provide plausible evidence that it will meet the mandatory criteria.
- When granting provisional accreditation, the German Data Forum (RatSWD) will contact the RDC at the end of the provisional period asking for a statement showing that the RDC has complied with its obligations.
 - If this statement receives a positive evaluation, the German Data Forum (RatSWD) will discuss the case and may go forward with the accreditation.
 - If the RDC's development is considered insufficient, provisional accreditation is not changed to permanent accreditation.

2.3 Monitoring and evaluation

Monitoring commission

The FDI Committee elects a monitoring commission for a three-year term concurrent with the German Data Forum's (RatSWD) appointment period. For a schematic overview, see Appendix B.

The main task of the monitoring commission is to collect and assess the research data centres' annual reports. Moreover, the commission handles complaints regarding RDC accreditation criteria and provisional accreditations (see also chap. 2.2).

Provisional accreditation

If an RDC's application for first-time accreditation has resulted in provisional accreditation involving certain obligations, the German Data Forum (RatSWD) asks the FDI Committee and its monitoring commission to monitor the RDC's compliance with the obligations in the following ways:

- After the agreed provisional period, the German Data Forum (RatSWD) requests that the RDC submits documentation showing its compliance with the obligations. These documents are forwarded to the monitoring commission.
- After reviewing the documents, the monitoring commission issues a statement to the German Data Forum (RatSWD) evaluating the documents submitted by the RDC:
 - If the statement is considered unproblematic (i.e. positive), the obligations are fulfilled, and the accreditation procedure ends. The German Data Forum (RatSWD) discusses the case and may grant accreditation.
 - If the statement is considered problematic (i.e. negative), the German Data Forum (RatSWD) discusses the case and may decide to deny accreditation.

Annual reporting

All accredited RDCs contribute to annual reporting by completing a questionnaire. As with accreditation, the questionnaire is based on the mandatory and information criteria.

Annual reporting is an instrument to assure the quality of the entire research data infrastructure of the German Data Forum (RatSWD). It is meant to assess whether RDCs comply with the accreditation criteria. Furthermore, it serves as the basis for a joint annual activity report of all accredited research data centres giving a transparent account of the quality and range of the data services offered by the RDCs.

The FDI Committee's monitoring commission reads and evaluates the questionnaires. If the monitoring process reveals deficiencies at an RDC, the RDC is informed and requested to submit a written statement. Depending on the assessment of the deficiencies and the explanations provided by the RDC, the monitoring commission and the evaluation commission may initiate further steps or end the procedure.

Evaluations requested by RDCs

If an RDC itself requests an evaluation – for instance because it has accomplished a significant quality improvement and wants that improvement confirmed by the German Data Forum (RatSWD) – the German Data Forum (RatSWD) will establish an evaluation commission within a three-month period. Normally, that commission is composed of members of the German Data Forum (RatSWD) and the FDI Committee, but the German Data Forum (RatSWD) may also choose to bring in outside expertise. The evaluation commission delivers a report to the German Data Forum (RatSWD) assessing the RDC's level of development. In such cases, the applicant RDC bears the costs.

2.4 Complaints office and procedures

To professionalise complaint management and to make it more transparent, a complaints office was established at the German Data Forum (RatSWD) office. If data users identify deficiencies in the data of an accredited research data centre, and if they cannot find a solution directly with the specific RDC, they are able to submit the issue to the German Data Forum's (RatSWD) complaints office. The complaints office's mandate is limited to matters of compliance with the accreditation criteria of the German Data Forum.

If data users note major shortcomings in the data services of an accredited RDC, they should first approach the RDC directly and try to find a solution. If the problem cannot be resolved, users may direct their concern to the German Data Forum (RatSWD) complaints office. For that case, the German Data Forum (RatSWD) has created a detailed complaints procedure.

3 Facts and figures about the current activities of the accredited RDCs

The annual reporting of RDCs (see chap. 2.3) provides information on the research data infrastructure's size, performance, and the general progress of the research data infrastructure. The following section presents the results from the reporting year 2016, when there were 30 RDC accredited.

3.1 Structure of the research data centres (RDCs)

By 2016, the German Data Forum (RatSWD) had accredited 30 research data centres. These employed a total of 225.7 staff (see Tab. 1) in full-time equivalents in 2016.

Tab. 1: Staff in full-time equivalents (FTEs)
Responses from 30 research data centres (n=30)

| | Number in FTEs | Range in FTEs |
|--------------------|----------------|---------------|
| Total staff | 225.7 | 0.375 – 36 |
| Academic staff | 143.2 | 0.25 – 24 |
| Non-academic staff | 58.1 | 0 – 12 |
| Student assistants | 24.4 | 0 – 3.6 |

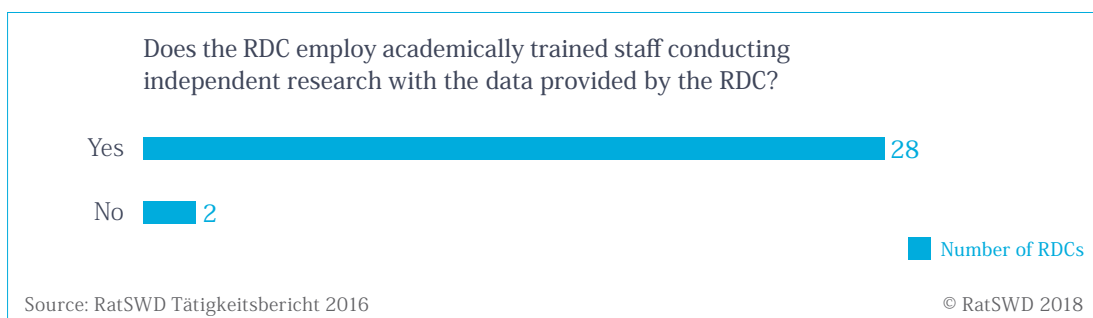
Source: RatSWD Tätigkeitsbericht 2016

© RatSWD 2018

The academic staff constitutes the largest proportion of the total staff of RDCs.

Twenty-eight RDCs employed academically trained staff performing independent research with the data held at the RDC (see Fig. 2). That research refers to content-related issues, explores questions of methodology, or seeks to achieve technical advancements. As a general rule, it is desirable for RDCs to engage in research of their own because this is the best way for staff to become familiar with their own data, which helps to improve data, services, and user consulting.

Fig. 2: Independent research by academic staff
Responses from 30 research data centres (n=30)

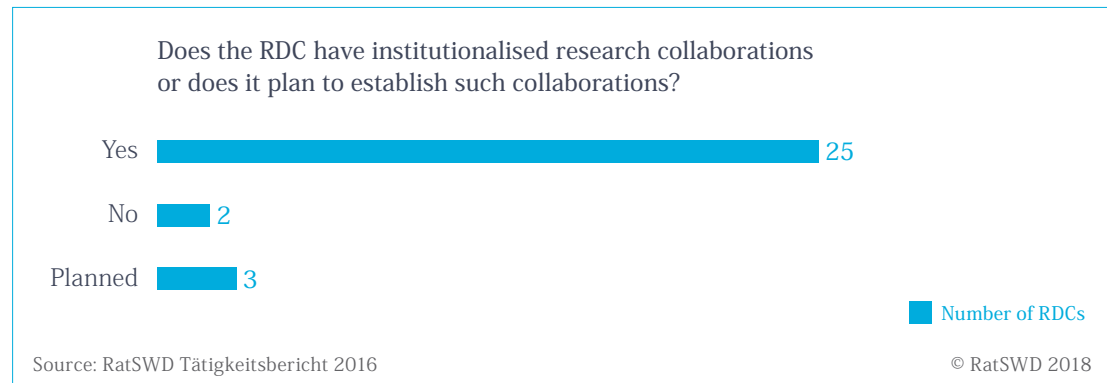


Most RDCs use their data to conduct own independent research.

In 2016, 25 of the 30 accredited RDCs were part of research collaborations (see Fig. 3). Of the five RDCs not involved in research collaborations, three are currently planning to enter collaborations.

Fig. 3: Research collaborations

Responses from 30 research data centres (n=30)



Almost all RDCs engage in or are currently planning research collaborations.

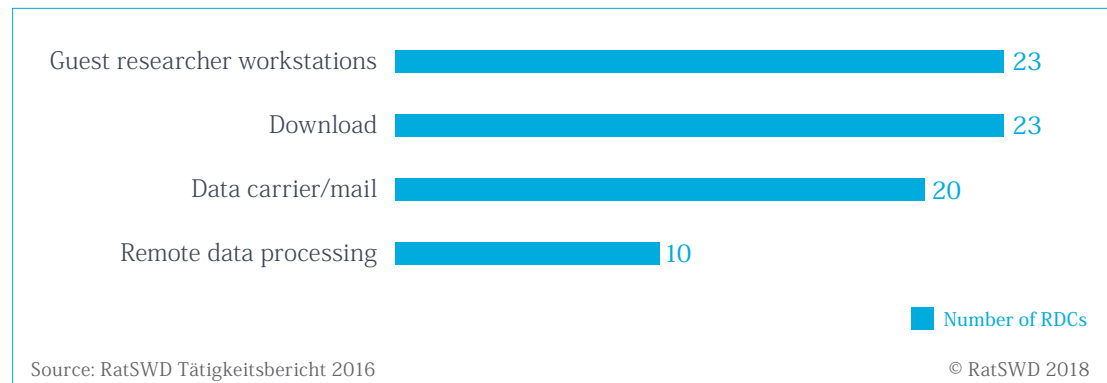
3.2 Data storage and access

The RDCs provide a wide range of access paths to their data. Generally, there are two distinct basic ways of accessing the data: on-site (i.e. on the physical premises of the RDC) and off-site (i.e. outside the RDC). Several RDCs offer multiple paths of access.

Guest researcher workstations with special security provisions for on-site data access are available at 23 RDCs (see Fig. 4 and Appendix A). Researchers can use these workstations to access confidential data. Ten RDCs offer the possibility of using data via (controlled) remote data processing, for instance, via remote access or by sending programme syntax to the RDC staff. Compared to the previous year, off-site services have seen particular growth at many research data centres. A total of 23 RDCs now allow researchers to download datasets for off-site use. Receiving data by e-mail or on data carriers is possible at 20 research data centres.

Fig. 4: Data access paths

Responses from 30 research data centres (n=30). Multiple responses were possible.



The RDCs provide a wide range of access paths for their data.

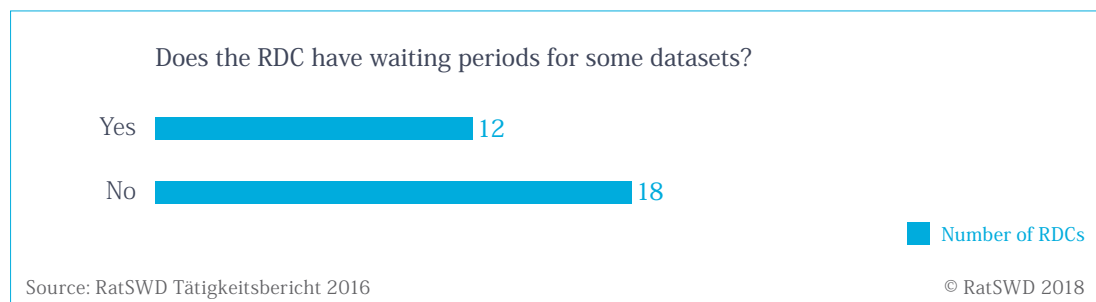
At each RDC, various procedures are used to store the data in databases, on servers, or in specialised data archives, where backups are performed on a regular basis. Data are stored in different formats (e.g., csv, SPSS, or Stata). Specialised data archives use multi-site redundant storage on different media – that is, the archived data are duplicated and stored at multiple physical sites simultaneously using different storage media (hard drives, CDs, etc.).

3.3 Data provided

In 2016, the RDCs offered in sum 3,214 datasets. A total of 18 RDCs do not have waiting periods, that is, the data are made available immediately after their preparation (see Fig. 5). The other twelve RDCs report waiting periods of a few months up to a maximum of three years. For the most part, waiting periods are only used for specific datasets; the standard case at all RDCs is making data available as soon as the preparation and documentation process is completed.

Fig. 5: Waiting periods

Responses from 30 research data centres (n=30)

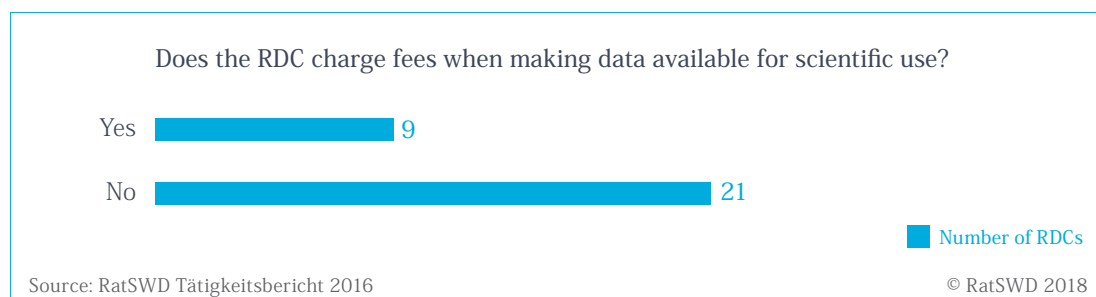


The datasets of RDCs are subject to different waiting policies.

Of the 30 RDCs accredited up until 2016, 21 report not charging any fees for providing access to data (see Fig. 6). The fees charged by the nine other RDCs are low (in the two-digit euro realm) and mostly used to cover the costs for media and contracting.

Fig. 6: Data access fees

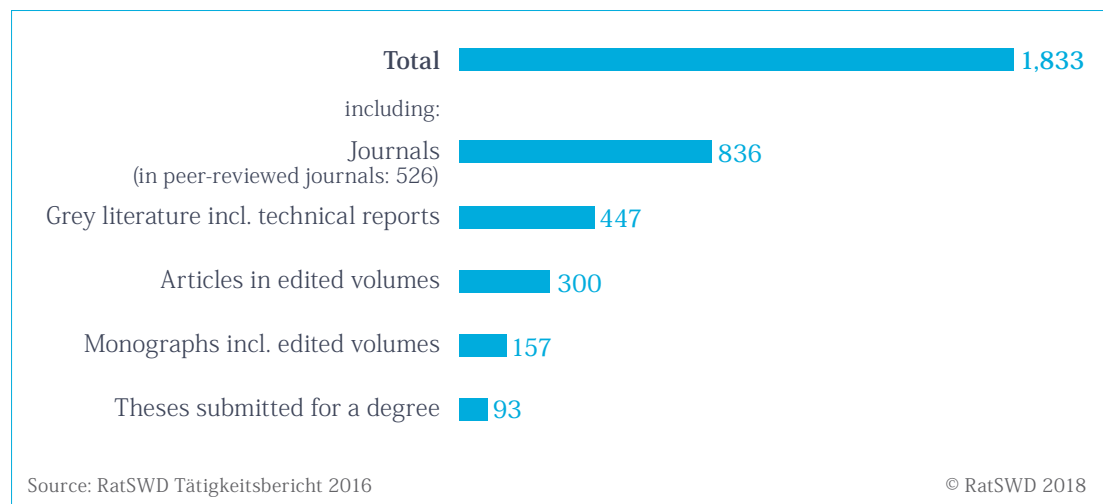
Responses from 30 research data centres (n=30)



Most RDCs make their data available for free.

For the 2016 report year, 1,833 publications were registered that use data provided by the RDCs (see Fig. 7). There are five RDCs for which publication figures are not collected. Thus, the reported numbers underreport the actual publication outlets and present only minimum figures. Generally, articles in peer-reviewed journals are the most frequent type of final publication.

Fig. 7: Number of final publications in 2016 based on the research data provided
Responses from 25 research data centres (n=25)



The most frequent type of final publication using the RDC's research data are peer-reviewed journal articles.

3.4 Services provided

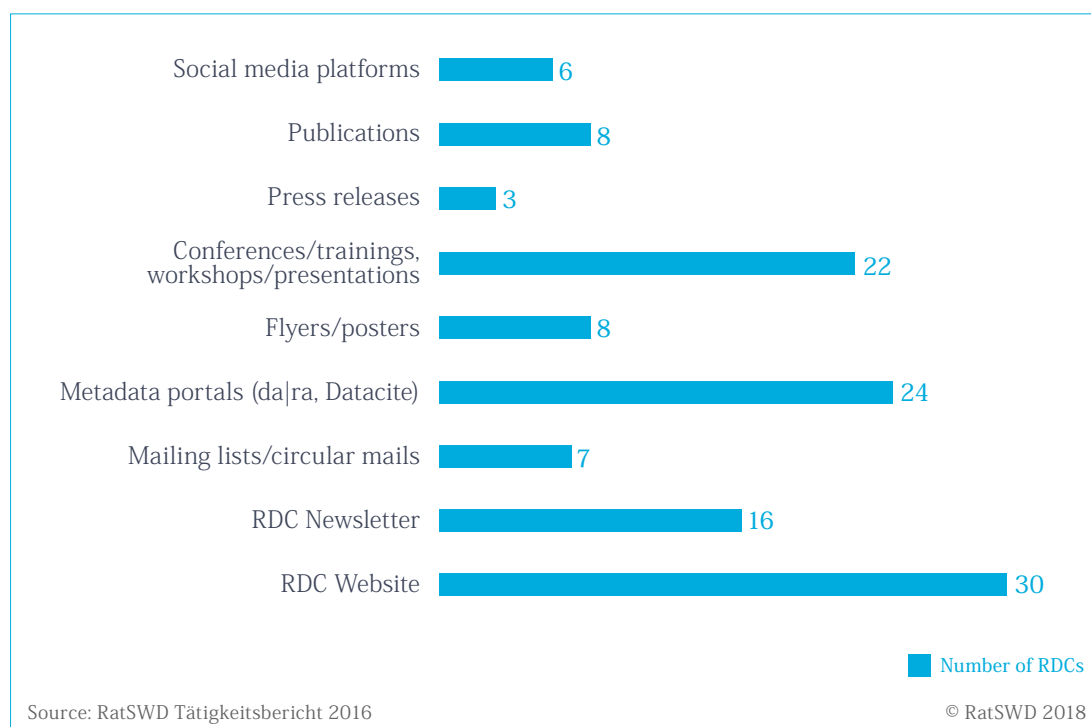
Provision of data documentation and tools

With the exception of one RDC, all current research data centres provide information and documentation materials; these vary, however, by RDC and dataset. Metadata for surveys usually include at least questionnaires and codebooks, sometimes supplemented by method reports, dataset descriptions, quality reports, classifications, technical papers, and variable lists. Moreover, users frequently have access to dataset-specific tools such as syntaxes (do files, XLM files, or web applications) or theory-based documents providing information about the theoretical background of individual questions or variables. Generally these materials are provided online via the RDC's website and can be accessed as open access materials without prior registration. If data are delivered as a package, this package also includes the available tools in most cases. Usually, data documentation is downloadable as open access PDF files. Furthermore, there is a remarkable amount of English-language documentation to make data services accessible to international audiences.

The RDCs use a variety of methods to introduce their data to the scientific community. Figure 8 indicates the various channels used by the RDCs to present their data. The predominant outlets are the research data centres' own websites. Other important channels for RDCs to present their data and to attract additional users include using metadata portals and assigning persistent identifiers, most notably by registering the data in da|ra.⁴ Presenting the data at (international) conferences, trainings, and workshops, is another frequently used method by RDCs to spread information about their data and their services. This also includes giving courses at universities to inform students.

Fig. 8: Information channels used to present data services

Free-text responses from 30 research data centres (n=30). Multiple responses were possible.



Websites, metadata portals, and direct communication are the three main information channel of RDCs to present their data to the scientific community.

Processing time from application to data transfer

Researchers usually have to apply to use data of an RDC that are not openly available for download. After submitting an application, many RDCs conclude formal contracts with the researchers to regulate the data usage. Researchers may have to wait between one hour and up to eight weeks until they receive the requested data, depending on the RDC in question. If online access is possible, the requested data are usually made available immediately or within a few hours. If data have to be made accessible, users should allow for at least one extra day. Longer processing times are to be expected if users request access to data containing sensitive information, if they want the data to be prepared in a special way, or if special permissions have to be obtained before the data may be released.

⁴ Registration agency for social and economic data (da|ra offers the DOI registration service in Germany for social science and economic data). Further information: <https://www.da-ra.de/home> (12/21/2017)

Service concepts

All RDCs employ contact persons whom external researchers may contact by phone, by e-mail, or in person at the respective RDC. If specialised expertise is necessary, questions are usually forwarded to in-house specialists. In addition, the RDCs offer advice on the right choice of dataset, on a dataset's analytical potential, or on data management, including the required measures for data protection and data security. Aside from offering professional development opportunities, RDC staff gives workshops and trainings, or presentations at user conferences, seminars, and universities. Other ways in which some RDCs support data users include providing FAQs and information on their websites.

Evaluation and quality assurance

User feedback and evaluations are institutionalised to different degrees at the various RDCs and the intervals at which they are performed vary as well.

Aside from formal evaluations, user surveys such as feedback questionnaires also play a major role. The efforts to continuously improve services are as much guided by the formal evaluations as by the feedback provided by users. Standardised user surveys at regular intervals are a quality assurance tool for nearly half of all RDCs. Other centres plan to introduce such surveys in the future.

4 International orientation of the research data infrastructure

Research institutions and research infrastructures have increasingly adopted an international orientation. The degree of internationalisation in the existing research data infrastructure can be described in four dimensions:

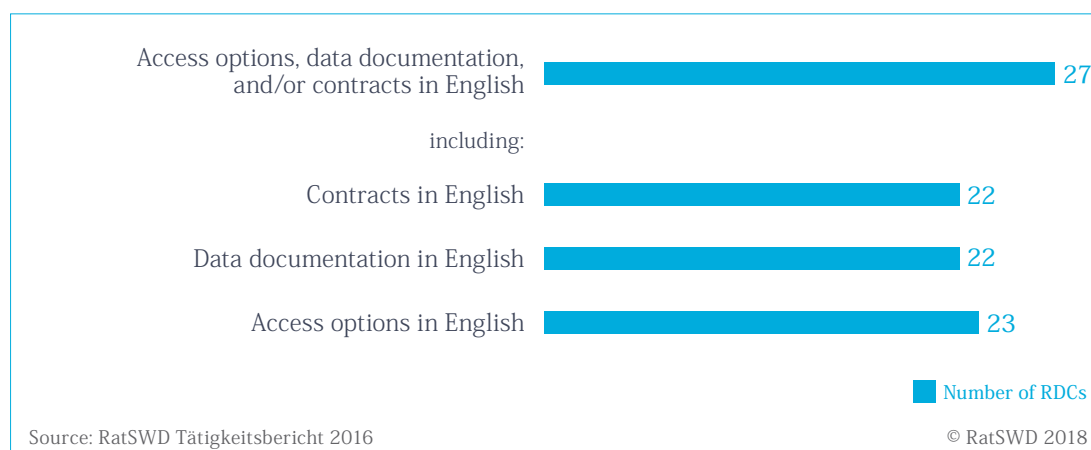
- 1) use of internationally accepted standards,
- 2) data documentation in English language,
- 3) participation of research data centres in international collaborations,
- 4) the utilisation of the infrastructure by researchers abroad.

1) Being guided by and using internationally accepted standards of data description and data provision serves to ensure interoperability and harmonisation across national borders. In particular, the use of persistent identifiers and the related standardisation regarding the description of essential characteristics and contents of each set of research data is now common practice at all RDCs. This helps to ensure that all data can be transferred directly to international research contexts.

2) Nearly all RDCs provide information in English about their data and the services they offer. Twenty-seven RDCs offer English-language contracts, access options, and/or data documentation (see Fig. 9). Standard services also include English-language data documentation containing not only technical but also methodological descriptions and information.

Fig. 9: English-language support for researchers

Responses from 30 research data centres (n=30). Multiple responses were possible.

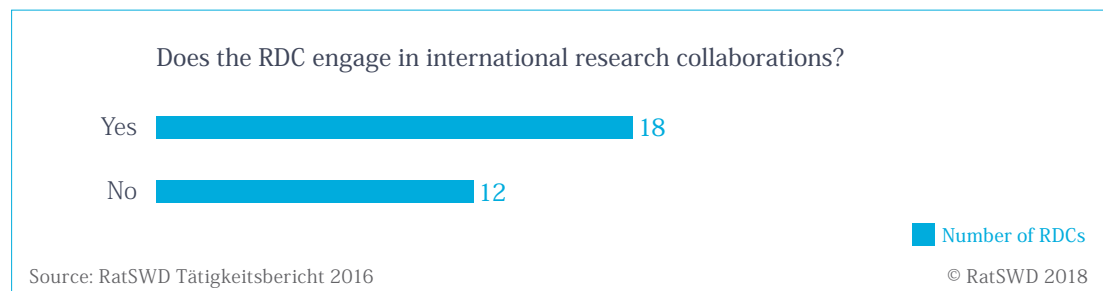


Almost all RDCs offer information in English-language about their data and services.

3) In addition, 18 RDCs are involved in international research collaborations (see Fig. 10). At the same time, there has been an overall intensification of contacts and collaboration with international research data infrastructure institutions (UK Data Archive, Data Archiving and Networked Services [DANS], etc.). The goal of the collaborative projects in this context is to expand the international research data infrastructure. One example is the concluded project “Data without Boundaries”⁵, which involved multiple RDCs, national statistical offices, universities, and data archives. Another example is the ongoing project “Synergies for Europe’s Research Infrastructures in the Social Sciences” (SERISS)⁶, which is about harmonising cross-national social sciences surveys.

Fig. 10: International research collaborations

Responses from 30 research data centres (n=30)



Many RDCs engage in international collaborations.

Another way of embedding the RDCs in international networks is to have non-German researchers or representatives on centres’ advisory boards. These boards consist of representatives of the various academic disciplines as well as experts from the fields of IT and archive management. Of the 30 accredited RDCs in 2016, 20 had an advisory board and 18 of them included members with an international background (see Fig. 11).

Fig. 11: International advisory board members

Responses from 20 research data centres (n=20)



The advisory boards of most RDCs include members with an international background.

⁵ <http://www.dwbproject.org> (12/21/2017)

⁶ <https://seriss.eu> (12/21/2017)

4) Finally, internationalisation also refers to researchers abroad using the data offered by an RDC. The percentage of international users varies widely across the individual research data centres. A total of 23 RDCs maintain contacts to research institutions abroad or provide assistance to users from foreign institutions (see Fig. 12). The internationalisation of data access is particularly beneficial for German researchers who work abroad but need access to German data. Performing comparative and cross-country analyses in the social and economic sciences thus becomes possible.

Fig. 12: Contacts to research institutions abroad

Responses from 30 research data centres (n=30)



The majority of RDCs has contacts to research institutes abroad.

Appendix

Appendix B: At a glance 2: Bottom-up governance to secure the quality of the research data infrastructure

The German Data Forum (RatSWD) is an advisory board made up of elected representatives from the data-using scientific community in the social sciences and appointed representatives of major German data producers. The FDI Committee consists of representatives from all accredited research data centres (RDCs). It complements the German Data Forum's (RatSWD) strategic focus with a focus on day-to-day challenges and quality assurance.

Accreditation

The German Data Forum (RatSWD) has developed standards and criteria for the accreditation of RDCs. Fully operational RDCs need to meet three mandatory criteria (in addition to further information criteria):

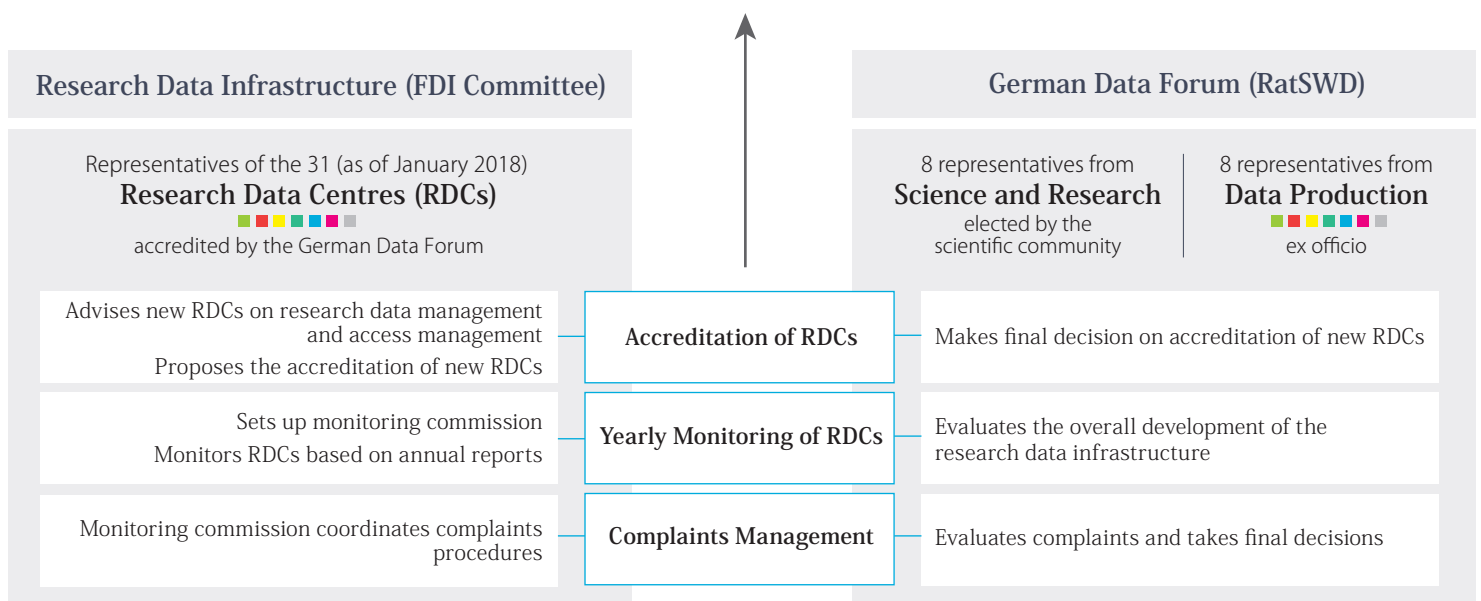
- provide at least one data access path
- provide sufficient data documentation
- ensure the long-term availability of the data

Monitoring and Complaints Management

The FDI Committee elects a monitoring commission which manages:

- the annual monitoring of all accredited RDCs
- complaints from the research community about data access policies in RDCs

The cooperation between the FDI Committee and the RatSWD:
developing and optimising the research data infrastructure



Appendix C:

Mandatory and information criteria for research data centres (RDCs) to gain accreditation from the German Data Forum (RatSWD)

To be accredited by the German Data Forum (RatSWD), a research data centre must meet three mandatory criteria (M). Furthermore, it must answer questions about additional information criteria (I) to enable the German Data Forum (RatSWD) to assess the scope and quality of the RDC's operative business.









The individual criteria and their operationalisation are described below:

| Criterion | Operationalisation | Response options | Mandatory criterion (M); Information criterion (I) |
|---|---|--|--|
| Minimum of one data access path | What is the primary access path (minimum one) through which data are made available? What were the reasons for choosing this/these access path(s)? | Yes, No. If there is no access path, accreditation cannot be granted. | (M) |
| Provision of sufficient data documentation | Does the RDC provide data documentation? | Yes, No. If yes, which? If no, then accreditation cannot be granted. | (M) |
| Concept for long-term data availability | Does the RDC guarantee data access for a minimum of 10 years (legally mandated storage period)? | Yes, No. If no, then accreditation cannot be granted. | (M) |
| Provision of social, behavioural, or economic science data | What are the thematic fields and research contents covered by the data provided by the RDC? From which institutions does the RDC obtain its research data? | Self-placement within the social, behavioural, and economic sciences; assignment according to the professional associations eligible to nominate candidates for the German Data Forum (RatSWD) elections | (I) |
| Provision of tools | Does the RDC offer tools that go beyond mere documentation (e.g., codebooks, variable descriptions, syntax)? | Yes, No. If yes, which tools and where? | (I) |
| Quality assurance of datasets | Is reviewing data (for quality) part of the responsibilities of the RDC? | If yes, please list the procedures used. | (I) |

| Criterion | Operationalisation | Response options | Mandatory criterion (M); Information criterion (I) |
|---|---|--|--|
| Data protection in due consideration of the interests of researchers | List of the procedures used to safeguard both data protection and the interests of researchers as well as the legal foundations used for the anonymisation of personal data and their legal equivalents. | | (I) |
| Service concept | Which services are available to users (e.g. contact persons, advisory services, training offered by RDC staff, workshops, etc.)? | | (I) |
| Single entity comprising institution and RDC | Does the institution have multiple RDCs? If yes, why is it impossible to integrate the data into an existing RDC? | Yes, No. If yes, please explain. | (I) |
| Provision of all datasets relevant to research | Can users access all datasets relevant to research? | Yes, No. If no, please explain. | (I) |
| Overlap and distinct features compared to existing RDCs | Are there already RDCs which offer the potential for overlap? | Yes, No. If yes, how is the work divided between the RDCs concerned? If yes, which policies are in place regarding copyrights and usage rights, if applicable? | (I) |
| Research activities | Does the RDC employ academically trained staff performing independent research with the data offered by the RDC? Are there institutionalised (e.g. contractual regulated) research collaborations or are they planned? | Yes, No. If no, please explain. | (I) |

| Criterion | Operationalisation | Response options | Mandatory criterion (M); Information criterion (I) |
|--|---|--|--|
| Multiple provision of the same data (multiple hosting, not hosting at multiple sites) | Are the research data already available at another institution? | Yes, No. If yes, please list the data concerned. If yes, what is the difference between the data and why is it reasonable to offer them at multiple sites? | (I) |
| Time to process applications | What is the average processing time from the moment when all user information and documents required for creating a contract are available and the moment when the data are transmitted? Does the RDC charge fees for providing data access? | | (I) |
| Staff | How many staff does the RDC (plan to) employ? | | (I) |
| Contribution to developing the infrastructure | What is unique about the RDC and otherwise not available in the research data infrastructure? Where does the RDC see a need for advice or for sharing knowledge? | | (I) |

Appendix D: Development and description of the German Data Forum (RatSWD) research data infrastructure

| | |
|------|---|
| 1999 | The "Commission to Improve the Informational Infrastructure between Research and Official Statistics" (KVI) recommends the establishment of research data centres (RDCs). |
| 2001 | <p>The German Data Forum (RatSWD) Founding Committee is set up.</p> <p>The following RDC is founded:</p> <p></p> <p>Research Data Centre of the Federal Statistical Office (FDZ-Bund)*</p> <p>Germany-wide access to official statistics microdata from the following fields: population, education, health, business, agriculture, environment, administration of justice, finance, and taxes.</p> <p>www.forschungsdatenzentrum.de/en</p>  |
| 2002 | <p>The following RDC is founded:</p> <p></p> <p>Research Data Centre of the Statistical Offices of the Länder (FDZ-Länder)*</p> <p>Germany-wide access to official statistics microdata from the following fields: population, education, health, business, agriculture, environment, administration of justice, finance, and taxes.</p> <p>www.forschungsdatenzentrum.de/en</p>  |
| 2003 | <p>The following RDCs are founded:</p> <p></p> <p>Research Data Centre German Microdata Lab at GESIS (RDC GML)*</p> <p>Research based services for researchers working with microdata from European and German official statistics: Tools for data management and data analysis. Metadata (MISSY): Comprehensive data documentation for official microdata on a detailed level. Knowledge transfer: Consulting, Training, User Conferences. On research using official microdata on methodological and substantive topics. Established 1987.</p> <p>www.gesis.org/en/institute/research-data-centers/rdc-german-microdata-lab</p>  <p></p> <p>International Data Service Centre at the Institute for the Study of Labour (FDZ IZA, IDSC)*</p> <p>National and international labour market datasets with standardised information (eddi-conferences.eu). 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).</p> <p>http://idsc.iza.org</p>  |

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 im 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.

<http://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.

http://forschung.deutsche-rentenversicherung.de/FdzPortalWeb/discontent.do?id=main_fdz_english



2008

The following RDCs are accredited:



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 (FDI 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.

www.diw.de/en/diw_02.c.222518.en/research_data_center_of_the_soep.html



* The research data centres Federal Statistical Office, Statistical Offices of the Länder, GML, IZA, BA im IAB, and RV were established prior to the foundation of the German Data Forum (RatSWD) and became part of the German Data Forum (RatSWD) 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 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.

www.share-project.org



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).

www.gesis.org/en/institute/research-data-centers/rdc-international-survey-programs



Research Data Centre Elections at GESIS (RDC Elections)

Access to German national election surveys (federal elections and state elections), Politbarometer, Forsa-Bus, ARD Deutschlandtrend. The RDC's largest project at this point is the German Longitudinal Election Study (GLES).

www.gesis.org/en/institute/research-data-centers/rdc-elections



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.

www.gesis.org/en/institute/research-data-centers/rdc-allbus



2010

The following RDCs are accredited:



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.

www.uni-bielefeld.de/en/soz/fdzbo



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://fdz.rwi-essen.de>



2011

The following RDCs are accredited:



LMU-ifo Economics & Business Data Center (EBDC)

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.

www.cesifo-group.de/de/ifoHome/facts/EBDC.html



Research Data Centre of the Robert Koch Institute (RDC RKI)

Data on the state of health and health-related behaviour of Germany's resident population, collected on the basis of nationally representative studies.

www.rki.de/puf

ROBERT KOCH INSTITUT



Research Data Centre of the Federal Centre for Health Education (FDZ BZgA)

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.

www.bzga.de/home



2012

The following RDCs are accredited:



**Research Data Center Wissenschaftsstatistik of the Stifterverband
(RDC Wissenschaftsstatistik)**

Data on the research and development activities of German companies, on the financial volume, structure, and regional distribution of research and development activities (R&D), and on R&D staff in the business sector.

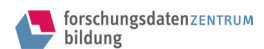
www.stifterverband.org/research_data_center



**Research Data Centre Education at the German Institute
for International Educational Research (DIPF) (FDZ Bildung)**

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.

www.fdz-bildung.de



**Research Data Center of the Leibniz Institute for Educational Trajectories
at the University of Bamberg (RDC-LifBi)**

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.

www.lifbi.de/en-us/home.aspx



ZEW Research Data Centre for European Economic Research (ZEW-FDZ)

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.

<http://kooperationen.zew.de/en/zew-fdz>



2013

The following RDCs are accredited:



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

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.

www.dji.de/en



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

The following RDCs are accredited:



Research Data Center Archive for Spoken German at the Institute for the German Language (**FDZ-AGD**)

Data on spoken German in interactions (conversation corpora) and data on domestic and non-domestic varieties of German (variation corpora).

<http://agd.ids-mannheim.de>



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/institute/research-data-centers/rdc-allbus/



2015

The following RDCs are accredited:



Deutsche Bundesbank Research Data and Service Centre (**RDSC Bundesbank**)

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.

www.bundesbank.de/Navigation/EN/Bundesbank/Research/RDSC/rdsc.html



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.

www.iwh-halle.de/en/research/data-and-analysis/research-data-centre



2017

The following RDC is accredited:



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>



Imprint

Editor:

German Data Forum (RatSWD)
Chausseestraße 111
10115 Berlin
Germany
office@ratswd.de
www.ratswd.de/en

Editorial:

Mathias Bug, Nora Dörrenbächer

Translation:

Carsten Bösel

Layout:

Claudia Kreutz

Icons:

Made by Freepik from www.flaticon.com

Berlin, March 2018

This report is the result of a project, which 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).

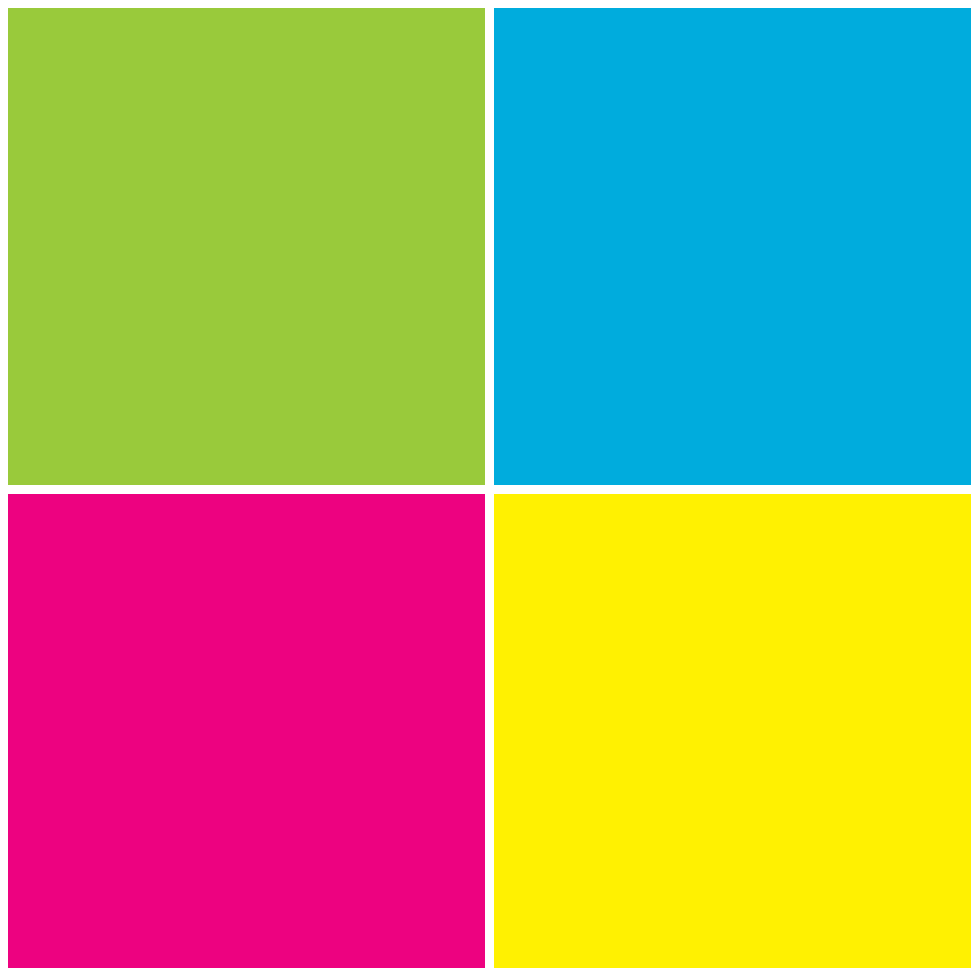
doi: 10.17620/02671.30

Citation:

RatSWD [German Data Forum] (2018): The German Data Forum (RatSWD) and Research Data Infrastructure: Status Quo and Quality Management. RatSWD Output 1 (6). Berlin: German Data Forum (RatSWD). <https://doi.org/10.17620/02671.30>

■ Established in 2004, the **German Data Forum** (Rat für Sozial- und Wirtschaftsdaten, RatSWD) is an independent council. It advises the German federal government and the federal states (Länder) in matters concerning the research data infrastructure for the empirical social, behavioural, and economic sciences. The German Data Forum (RatSWD) has 16 members. Membership consists of eight elected representatives of the social, behavioural, and economic sciences and eight appointed representatives of Germany's most important data producers.

The German Data Forum (RatSWD) offers a forum for dialogue between researchers and data producers, who jointly issue recommendations and position papers. The council furthers the development of a research infrastructure that provides researchers with flexible and secure access to a broad range of data. The German Data Forum (RatSWD) has accredited 31 research data centres and fosters their interaction and collaboration.



www.ratswd.de/en