Rat für
Informations
Infrastrukturen

ENHANCING RESEARCH DATA MANAGEMENT: PERFORMANCE THROUGH DIVERSITY

Recommendations regarding structures, processes, and financing for research data management in Germany

ABOUT THIS PUBLICATION

Adopted on 03 May 2016

Rat für Informationsinfrastrukturen (RfII)
Head Office
Papendiek 16
37073 Göttingen
Germany

Phone +49 551 392 09 59
E-mail info@rfii.de
Web www.rfii.de

DESIGN, TYPESETTING, AND PRINTING
Basta Werbeagentur GmbH, Göttingen

SUGGESTED CITATION
RfII – German Council for Scientific Information Infrastructures: Enhancing Research Data Management: Performance through Diversity. Recommendations regarding structures, processes, and financing for research data management in Germany, Göttingen 2016, 90 p.

The RfII prefers to use gender-sensitive language. The male form used in individual cases applies equally to females. We have forgone the consistent use of two designations to improve readability.

The German National Library lists this publication in the German National Bibliography; detailed bibliographic data can be found on the website at http://dnb.dnb.de.

CONTENTS

List of abbreviations

BMBF	Federal Ministry of Education and Research
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DFG	German Research Foundation
DINI	German Initiative for Network Information
DMP	Data Management Plan
ELIXIR	European Life-Sciences Infrastructure for Biological Information
ERIC	European Research Infrastructure Consortium
ERA	European Research Area
ESFRI	European Strategy Forum on Research Infrastructures
EUDAT	European Data Infrastructure
FAIR	Findable, Accessible, Interoperable, Reusable
GBIF	Global Biodiversity Information Facility
GESIS	Leibniz-Institute for the Social Sciences
GFBio	German Federation for Biological Data
GWDG	Joint computing centre by the University of Göttingen and the Max Planck Society
GWK	Joint Science Conference
HPC	High Performance Computing
HRK	German Rectors' Conference
KII	German Commission on the Future of Information Infrastructures
NEPS	German National Education Panel Study
nestor	German competence network for digital preservation
NFDI	National Research Data Infrastructure
NHR	National High Performance Computing
NoMaD	Novel Materials Discovery Project
OECD	Organisation for Economic Co-Operation and Development
OpenAIRE	Open Access Infrastructure for Research in Europe
RADAR	Research Data Repositorium
RatSWD	German Data Forum
RDA	Research Data Alliance
RDM	Research Data Management
RfII	German Council for Scientific Information Infrastructures
SOEP	German Socio-Economic Panel
WR	German Council of Science and Humanities

EXECUTIVE SUMMARY

The digital revolution is fundamentally transforming research data and methods; indeed, science on the whole is in a state of flux. Mastering this transformation poses major challenges for stakeholders in the domains of science and policy, in part because the precise nature of the transformation is not yet clear. The process of digitalisation creates immense opportunities, but it must be structured proactively. To this end, the establishment of effective governance mechanisms for research data management (RDM) is of fundamental importance and will be one key driver for successful research and innovation in the future.

In its position paper, the German Council for Scientific Information Infrastructures (RfII) makes a series of recommendations concerning how research data should be managed in the future. The RfII was tasked by Germany's Joint Science Conference (GWK) with formulating broad-based recommendations for the science system in Germany as a whole. Consequently, the recommendations presented here have ramifications for a variety of stakeholders in the domains of politics and science. The RfII is convinced that the adoption of new ways in dealing with research data as well as long-term preservation and accessibility will be a significant, common challenge for all stakeholders in the years ahead.

The position paper describes current policies and practices for managing research data and discusses a number of conflicting priorities in science policy. While there are several good examples of research data management[1] in Germany, there is an overall absence of coordination, and current efforts often take the form of parallel, project-based initiatives. Universal access to services for data management is lacking, as the key players at present are individual institutions and organisations, and their efforts often suffer from limited financing and/or excessive niche focus. High staff turnover means that valuable know-how is frequently lost. Furthermore, the range of services being provided is impaired by the absence of governance mechanisms which could impart greater strategic direction. In addition, there is a risk of international competitive disadvantage for Germany due to unresolved issues in the domains of quality assurance, legal compliance, data privacy, and data security. In this way, there is a clear need for action in a variety of areas.

Based on the foregoing findings, the RfII has developed a series of recommendations:

[1] The RfII has a broad conception of research data that encompasses both analogue and digital collections.

Adjusting funding mechanisms

With regard to funding policy, the RfII recommends implementing long-term funding mechanisms in line with the long-term nature of research data management. Current project funding schemes do not foster long-term sustainability, thus jeopardising the development of services needed in the long run. Policy-makers need to establish clear pathways for institutions and organisations to obtain the resources they require over the long term. Without tying up subsidies in a fixed manner, funding phases for RDM infrastructures should be designed so that applicants have a clear understanding of their chances for securing long-term financing and the steps that are necessary to obtain it. This will boost the value that is derived from public investment. The goal of funding policy should be to develop an ecosystem of sustainable infrastructures that provides researchers in Germany with universal and reliable access to data management services. Such an ecosystem should also designed to grow dynamically over time.

Efficiency and coordination based on a (distributed) national infrastructure

The RfII recommends the establishment of a *Nationale Forschungsdateninfrastruktur* (National Research Data Infrastructure, or NFDI), which will serve as the backbone for research data management in Germany. The NFDI should be implemented as a national collaborative network that grows over time and is composed of various specialised nodes. The establishment of such a network is recommended on a step-by-step basis, as this will ensure the overall management system remains flexible while also facilitating the productive integration of diverse resources. The NFDI will provide for coordination, cooperation, and common standards. As a network-based, dynamic organisational structure, the NFDI will be composed of nodes of various sizes. Some of these nodes will take the form of broad-based "service centres", while others will be specialised "centres of excellence" for specific subareas. These centres can be established within various existing organisations and institutions. The difficult issue of developing intelligent solutions for the long-term archiving/provisioning of research data will be among the areas addressed by the NFDI.

Responsible data culture

Monitoring and quality assurance

The current transition to digital processes means that nearly all of the "soft" factors for scientific activity are in a state of flux. The responsibilities that fall to researchers have to be recalibrated. The RfII thus makes various recommendations regarding data quality assurance, the adoption of a legal framework for data reuse (based on the Open Science model), and data privacy and protection. These recommendations aim to define the responsibilities borne by researchers and their organisations during all phases of the "data life cycle". Accordingly, policy-makers and scientists need to understand and set forth good scientific practice for research in the digital age. Scientific organisations also have a clear role to play in this area: monitoring and evaluation systems should be designed to create incentives for good research data management while also engendering trust among researchers and the lay public. Good data management practices go hand in hand with research that is cutting edge and has a strong practical value for society as a whole.

Comprehensive human resources development

The RfII additionally recommends that due attention be devoted at all levels to human resources development. Adequately qualified individuals are required in large numbers for data-intensive research and teaching. The German labour market is dangerously lagging behind global developments in this area. Accordingly, the RfII sees a pressing need to educate a new generation of highly capable researchers and specialist employees for new occupations in the area of data management. At the same time, new skills and greater awareness need to be promoted at management levels, as communication and process management are frequently decisive for solving infrastructure problems. Consequently, there is an urgent need to develop new occupational profiles and fields of study. Ultimately, it is the individuals behind the system, with their skills and aptitudes, who generate new scientific knowledge and added value by integrating data, information, and knowledge from heterogeneous sources.

Stronger international ties

The RfII attaches particular importance to international collaboration, for the flow of digital information (like research) knows no borders. The RDM ecosystem in Germany will have to develop within a broader European and global context. This does not only mean competition, but more importantly active collaboration and mutual learning. The RfII believes that Germany has a particular role to play in this area, and calls upon policy-makers and scientists to develop closer networking ties to their counterparts in other nations.

Actively steering the transition

Changing course requires time, cost, and effort. The RfII believes that considerable investments are required at various levels of the German scientific system in order to ensure its future efficacy. The tasks in need of attention will require a long transition process. The scientific system will need to change in nearly all areas if research data and methods are to become truly digital in coming years. The need for change also applies to existing infrastructures, repositories, and archives, which need to reorient towards new tasks. At this current tipping point, the RfII calls upon decision-makers in the domains of science and policy to make vigorous efforts to facilitate the necessary transition phase. This call for action extends to Germany's *Länder* and the national government, which are urged to undertake effective and rapid measures to overcome the current fragmentation of efforts in the area of research data management. This fragmentation cannot be overcome without mobilising substantial resources within the science system, including its funders.

The RfII emphasises that the challenges are complex and that collaboration will be decisive for success. Scientific policy-makers and the scientific community are equally responsible for ensuring that joint action takes place. For all those involved, the overarching concern should be to support the international competitiveness of German scientific research and make efficient use of public funding. With effective collaboration Germany can actively push the digital transformation to ensure maximum benefit for science, industry, and society.

This English translation of the position paper is an abridged version of the original German version. The original German version contains extensive appendices with in-depth information on the following topics: clarification of additional relevant terms, history of information infrastructure policy in the Federal Republic of Germany, sample “scenarios” of the creation of research data, as well as other facts and data related to information infrastructures.[2]

2 Cf. http://www.rfii.de/download/rfii-empfehlungen-2016/ (last checked on 17/08/2016).

1 CHALLENGES AND TASKS

The digitality of data collection, workflows, publication channels, and usage methods in the sciences is progressing at dramatic pace. Understandably, scientific policy and research financing initially focused on large-scale research facilities in this context. Challenges like the digitisation of analogue knowledge bases and agenda items relating to access to digital knowledge such as *open access* or *information literacy* arose as well.

In the meantime, significantly more fundamental questions regarding research data and the future management of digital research data have come into the forefront in the fields of research, education, and knowledge transfer. Due to the enormous amounts and heterogeneity of such data – which is highly specific and expires relatively quickly – there is a need for action. Optimal exploitation as well as storage and utilisation of this data require a political framework that does not exist yet. Nonetheless, the sustainable use of quality-assured research data for scientific purposes adds significant value to public investments in research and has a high potential for creating value in science and business. The digital transformation is characterised by system-relevant tipping points at which it is necessary to decide which path to take to enable new, high performance structures. A failure to act at this point will have a negative impact. There is not only a risk of being left behind, but also of pursuing undesirable developments.

However, the call for *research data management* involves a complex web of requirements and design questions. Research data is not only comprised of the (final) results of research. Instead, research data comprises all data generated in the course of scientific activity, including large amounts of data used for documentation purposes in scientific projects generated through measurements and through selecting, preparing, collecting, and storing information. The management of *digital* research data entails a wide range of tasks, including organisational measures extending beyond the normal activities of researchers in the narrower sense. These measures are necessary to make the results reproducible in the digital world and to make data available for reuse throughout the entire *data life cycle*, from the collection, processing, and analysis to the archiving of the data.[3]

Since the turn of the millennium, numerous international initiatives dealing with data collections, hosting, standardisation, juridification, and use have arisen. There is also activity in this area in Germany, and numerous recommendations, agendas, and appeals have been submitted by the major players in the German scientific system.

[3] The RfII has created an explanation of terms for this and other frequently used terms and phrases. It is included in this position paper as a glossary.

The present situation is characterised by *diversity* in the positive sense, and by *fragmentation* in the negative sense. The transformation has seen numerous separate developments and different stages of development. Researchers and service partners, universities and non-university facilities, disciplines and project consortia, research funders and ministries are experimenting with options usually obtained through a *bottom-up* approach. Decision-makers in science and politics have favoured the creation of new structures from the *bottom-up* over consolidation and comprehensive systematisation. Many of the self-organised infrastructures are still in the experimental stage and are technically and organisationally incompatible for the most part. Maintaining all these systems is not financially feasible over the long term; there are technical, organisational, and financial limits restricting the ability to subsequently merge them into one system; sensible interventions are needed to reduce expenses in the upcoming years. For these reasons, a consolidation phase should be planned with the future in mind to secure Germany's top position in the sciences.

The rapid advance of international competition is more than just a marginal issue in the handling of research data. Research in Germany is highly networked throughout Europe and across the globe, and it needs broad access to global resources in order to remain competitive. However, software, communication, storage, and publication services from private vendors are also used on a massive scale by researchers in Germany. Research data is being released on the Internet at an unprecedented scale and is thus being transferred to spheres not governed by German law. In terms of scientific policy, restrictive publication/subscription models from commercial providers are an additional cause for concern. The open science paradigm of the global availability of data also has a downside. Until now, there has been little discussion of the effects of maximum *openness* of research data on the scientific and economic competitiveness of data producers in different countries, regions, and institutions.

Indeed, there is no lack of information on the need for regulations, resources, and infrastructures or of recommendations for future developments in research data management. In spite of this, decision-makers in science and politics are having a hard time finding approaches for implementing measures, not to mention investing in services and research data infrastructures from a strategic perspective. This is comprehensible since the requirements for the management of digital research data actually affect almost all core processes in research, education, and knowledge transfer. As a result, it becomes necessary to restructure the existing organisation. Fully reacting to the challenges will thus lead to more or less significant changes in science, which is not the only reason why such regulatory decisions involve taking responsibility.

A data infrastructure in line with demand cannot be planned from the *top down* for the German scientific system, which is decentralised and networked for good reason, due to the wide range of disciplines, institutions, and forms of research it encompasses. However, decisive top-down stimuli are needed. The principle of giving priority to "stimulated self-organisation"[4] favoured by both the German Council of Science and Humanities (WR) and the Joint Science Conference (GWK) must also apply to the complex field of research data management. Nevertheless, a framework must be created for this purpose. The recommendations presented by the RfII here therefore pursue the goal of inducing more *oriented* and more *coordinated* behaviour of the various stakeholders involved as well as obtaining sustainable *support* from them according to the principles of a learning system. The objective is to obtain a balance between the following conflicting priority areas:

- Project funding vs. institutionalisation
- *Top-down* vs. *bottom-up* coordination of processes
- Open data vs. data sovereignty when conducting research (cooperation vs. competition)
- Discipline- and institution-specific solutions vs. comprehensive services
- Data protection vs. freedom of research
- Government funding vs. commercial funding
- Infrastructure services provided by government offices vs. services offered by large research facilities or higher education institutions
- Investments in infrastructures vs. investments in minds
- National activities vs. a European and global environment

The practical hurdles encountered in the form of funding eligibility conditions, institutional barriers, and the lack of recognition for infrastructure services urgently need to be overcome in the short term. Furthermore, there are legal questions that need to be answered in a manner that meets the needs of science. Studies also show that many researchers take little or no advantage of the new opportunities. In light of these practical and cultural hurdles, it is evident that clear rules, a target system, and suitable support functions are needed. The new structure should allow and promote indirect coordination and establish control points and mechanisms that subsequently bring together what belongs together.

4 GWK (2013) – Drucksache 13.48; WR (2012) – Empfehlungen zu Informationsinfrastrukturen, p. 81.

2 CURRENT SITUATION

2.1 RESEARCH DATA, INFORMATION INFRASTRUCTURES, AND SCIENTIFIC POLICY

The key role of infrastructures

Data management requires an infrastructure – and this is where scientific policy for shaping the transformation is required. Research infrastructures have always been of central importance to the development of science. Ever since the production of digital research data started increasing in the course of scientific activity and began to be used for documentation purposes in particular (e. g. for measurements, simulations, selection, preparation, collection, and storage), the special importance of information infrastructures has become evident. The term 'information infrastructure' refers to technically and organisationally networked services and facilities for working with data, information, and inventories of knowledge in the sciences.[5] In turn, these are generally closely linked to digital methods, processes, discipline-specific services, and specific forms of research. The discussion is not only about devices and databases, but also about the research process itself because almost all research tools today have corresponding IT components and researchers communicate with each other using digital networks. The methods used by libraries, information centres, and archives to supply scientific information have also become digital.

Furthermore, information infrastructures – in the broader sense of services and processes as already mentioned – are a key to handling the sheer amount of data produced as well as to adapting to digital methods of managing (and accessing) digital research data. This includes legal aspects and aspects relating to scientific culture.

For this reason, stakeholders in the scientific system have reacted to the digital transformation in the last few years by releasing a number of statements, studies, and position papers on fundamental topics. Such papers discussed the *open data* paradigm, the systematic design of data infrastructures, the incentives needed to promote a change of culture in the sciences, and new job profiles, among other topics.

In 2010, a high level expert group of the European Commission suggested a collaborative data infrastructure that supports seamless access to trustwor-

5 Information infrastructures or (in a somewhat stricter sense) e-infrastructures should be understood as specific forms of research infrastructures. Cf. the explanation of terms in the glossary.

thy data as well as its sustainable use across different systems.[6] As a result, EU-funded initiatives with German involvement were launched that work towards the goal of developing such an infrastructure. Examples include the Research Data Alliance (RDA), European Data Infrastructure (EUDAT), and the planned European Open Science Cloud (EOSC).[7]

Position papers in Germany and the EU

In Germany, the discussion has become more intense since the publication of the overall concept of the German *Commission on the Future of Information Infrastructures* (KII) and a series of related recommendations from the Council of Science and Humanities in 2011 and 2012.[8] The German Initiative for Network Information (DINI), an association of libraries, computer centres, and media centres; the German Research Foundation (DFG); and the German Rectors' Conference (HRK) published statements that also reflect their particular institutional roles. From the perspective of research, the focus is on questions regarding the accessibility and protection of the growing amount of data expected. Data management and data-based science – including digital data analysis methods – must be supported and facilitated. Central challenges repeated over and over include the sustainable financing of suitable infrastructures, further education and training of the specialists working in science and in information facilities, and the development of the corresponding rules, policies, and commitments, but also the formulation of legal provisions.

The position paper of the research data working group in the *Priority Initiative Digital Information* formulates a vision of science in the spirit of the *Berlin Declaration*: by 2025, "Researchers in all disciplines [should be able to] access all research data quickly and easily in a straightforward process in order to carry out research at the highest level and produce excellent results".[9]

The digital agenda adopted in 2014 by the German federal government also includes better access to research data as a goal. The agenda as well as other

---

6 European Commission (2010) – Riding the wave, p. 4: "Our vision is a scientific e-infrastructure that supports seamless access, use, re-use, and trust of data. In a sense, the physical and technical infrastructure becomes invisible and the data themselves become the infrastructure […]".

7 Research Data Alliance – https://rd-alliance.org; EUDAT – https://www.eudat.eu/what-eudat; EOSC – https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud (links last checked on 25/04/2016). For information on the European Cloud initiative, see also the European Commission (2016) – Data and knowledge economy and BMBF (2016) – Eckpunkte European Cloud Initiative (unpublished).

8 Cf. KII – Kommission Zukunft der Informationsinfrastruktur (2011) – Gesamtkonzept, and WR (2012) – Empfehlungen zu Informationsinfrastrukturen (as well as the recommendations from the previous years referenced therein).

9 Allianz-Initiative Digitale Information – AG Forschungsdaten (2015) – Research Data at Your Fingertips, p. 1.

measures are understood in this regard as being part of an internationalisation strategy for the sciences.[10]

It thus acknowledges the need for action in a field that, in the early digital phase of the Federal Republic of Germany in the 1970s, was initially organised centrally for the sciences by the government using so-called "specialised information systems". Since the 1980s, though, responsibility has been increasingly transferred to the sciences. The players involved in the digital transformation of science were able to take advantage of the opportunities arising in conjunction with liberalisation and direct responsibility. Distributed activities and competitive elements have ensured dynamic development.[11]

Diversity of project-based approaches to solutions in Germany

In the meantime, a primarily decentralised *"landscape"* of solutions and infrastructures has developed – stimulated in part by funding programmes (cf. 2.2). Some of these are oriented towards data-based services (storage, calculation, transfers), some are more geared towards information services (indexing, informing, researching), while others are based on specific scientific methodologies and forms of research (observation, experimentation, measurement, simulation, hermeneutic interpretation, theoretical analysis, design/development).

Nevertheless, the status quo of self-organisation in the sciences stimulated in this manner creates some problems: developments often progress incrementally at different rates, heterogeneous stakeholders often take uncoordinated action, the development of comprehensive quality assurance standards[12] is painstaking, investments are almost impossible to coordinate, and resource problems lead to mistakes and stagnation. Diversity thus remains unguided, and synergies cannot be created through the universal (re)use of data. For these reasons, science and politics both have an equally high interest in consolidation and systematisation projects. In addition, many services still do not appear to be reliable to researchers due to the prevailing method of financing them on a project basis. To some extent, this explains why the use of services and resources and widespread acceptance of research data management have fallen behind expectations. This then harms the competitive position of Germany as a research location as a result.

Problem descriptions vs. implementation deficit

The RfII finds that there are already numerous informative statements and problem diagnoses available in and around the policy area of future research data management.[13] It is therefore necessary to begin by tackling all items

10 Some scientific policy goals of the digital agenda are also found in the federal government's strategy for realising the European Research Area. Cf. BMBF (2014) – Strategie EFR; Bundesregierung (2014) – Digitale Agenda 2014-2017.

11 Cf. WR (2012) – Empfehlungen zu Informationsinfrastrukturen, p. 35.

12 To gain an understanding of standards, cf. also the explanation of terms in the glossary.

13 Cf. RfII (2015) – Auftakterklärung, p. 3.

that, in light of the complexity of the situation, eliminate the existing implementation deficit in spite of the availability of problem descriptions.

In addition to the complexity of the requirements, it is also necessary to consider the current institutional situation in this regard. The German federal/state government structure as well as the balance between universities/higher education institutions and non-university research facilities, including departmental research, should be taken into consideration in future decisions on scientific policy. At the structural level, the traditional division of labour between computer centres, libraries, and the more “thematic”, discipline-based, or method-based data centres and archives should be taken into account. The service profiles of information infrastructures should be coordinated as well as possible.

2.2 OVERVIEW OF THE “LANDSCAPE”

Where is research data created?

Digital research data is generated throughout the entire spectrum of the scientific system. Certain forms of data-intensive, large-scale research facilities (e. g. for nuclear physics, Earth observation, astrophysics, or climate modelling) are typically encountered in *non-university research*. However, research at *universities* produces enormous amounts of digital data as well (e. g. in genomics/proteomics, pharmaceutical research using high throughput screening, archaeology/classical studies, research using imaging, audio and video data, and language corpuses). Universities also get involved in establishing large data networks (e. g. for geodata or healthcare data). In addition, the diversity of fields and projects produces numerous small and highly heterogeneous databases. Furthermore, *scientific libraries, collections, archives,* and *museums* also produce large amounts of data (e. g. core samples, biomaterials, or artwork), especially through the digitisation of analogue inventories (e. g. text, images, cultural assets, and natural history archives). Finally, data from *departmental research* and from measurements and surveys taken by *public authorities* also need to be mentioned; large amounts of such data are produced by the various levels of administration and by many non-governmental facilities (e. g. environmental data, agricultural data, social data, economic data, health insurance data, corporate data, data from telecommunication applications, etc.). All these types of data and combinations thereof are essential to research.

Working with research data

The situation is similarly complex when working with research data. It must be noted in particular here that digitality facilitates distributed collaborative research (EU consortia, special research areas, excellence clusters, postgraduate schools, etc.). In interdisciplinary constellations in particular, *heterogeneous methods and forms of research*, with their different types of data and requirements on data quality and data processing, meet head on. Over the last few years, an increasing number of *virtual research environments* have been established. The goal of such environments is to seamlessly integrate (individual) digital tools and to network resources. At the same time, they are also intended to support *communication between researchers* parallel to the research process. Low threshold services in this context include free Web 2.0 tools (LabArchives, GitHub, or even cloud services like Dropbox) that are used to exchange data between different disciplines. Highly integrated digital research environments typically require more or less homogeneous methods and communities.[14] At universities, research data is also used in *science education*; it should thus also be possible to integrate this data into digital learning environments.

Digital publication of research data

Contributions submitted for the digital publication of research findings are increasingly being linked with research data – at least in disciplines that work empirically or experimentally. The publication of research data is also becoming more and more common, for example in data journals in the earth and environmental sciences, nanotechnology, and the social sciences.[15] Institutionally, classic *publishers* play a role, although just a few global players dominate the scientific publishing industry from an economic perspective. However, international journals are also published by *professional associations* (e. g. in physics), and the importance of digital *self-publishing* is increasing. In addition, service providers like the Dryad data repository, which is financed by member organisations, or the Figshare company have emerged to serve those willing to publish: they publish the data corresponding to authors' papers – usually for a fee. *Online networks* dedicated to science (Academia.edu, Mendeley, ResearchGate), which in some cases offer variable services to registered members at conditions similar to those of publishers, are also used for self-publishing.[16]

Providers of services for the exchange, information-based and usage-oriented collection, storage, indexing, preparation, and possibly even management of scientific data and information are typically *libraries*, but they can also be archives, *research data centres, research information centres*, or *repositories* for specific subjects or fields. The German landscape includes an abundance of facilities with different sponsors, consortium structures, commercial contract partners, service portfolios, concepts for use, and user groups from business and science. Some German data centres and data collections function as components of distributed transnational infrastructures that are supplied by

---

14 For information on the term "community", see the glossary.

15 Cf. examples of data journals at http://proj.badc.rl.ac.uk/preparde/blog/DataJournalsList (last checked on 21/04/2016) or Pangaea for the Earth and environmental sciences: https://www.pangaea.de (last checked on 25/04/2016).

16 Academia.edu – https://www.academia.edu; Mendeley Data – https://data.mendeley.com; ResearchGate – https://www.researchgate.net (last checked on 25/04/2016).

various national "nodes" (such as GBIF, NoMaD, ELIXIR, and the Europeana repository of cultural heritage). Wherever such "nodes" have formed, it has sometimes led to the consolidation of the corresponding individual activities at the national level (see 2.4). At the European level, the publicly funded EUDAT platform, which was designed as a distributed system dedicated to science and offers various data services, went online in 2014.[17]

Data archiving, long-term availability

Many of the tasks of a digital archive are similar to those of a large data centre, the difference being that they are designed from the outset for different storage periods and the – ideally permanent – storage of the digitally archived records, which will then be available in the future for currently unknown forms of use. Funding projects over the last few years have shown that there are still difficult questions regarding long-term archiving – for example regarding the tracking of data usage. Currently, some basic problems in the long-term archiving of digital artefacts (objects that were born digital or turned digital) can technically, logistically, and organisationally only be solved in part, since research data as well as metadata[18] needs to be archived on changing hardware together with the (quickly outdated) programs needed to use the data. Furthermore, organisational and legal standards for digital archiving are frequently inconsistent or unclear (for example where public authorities produce material for digital archives). A list of examples illustrates the range of German institutions dealing with digital archiving: the German National Library, the Bavarian State Library, the three German National Libraries (of Medicine, Economics, and Science and Technology), the German Literature Archive Marbach (DLA), the German Federal Archives, and the state archives involved in various cooperation projects.[19] Special infrastructural challenges also arise wherever turned-digital objects are archived together with conventional physical data ("original" items in collections, fabric samples, archaeological finds, core samples, etc.).

The organisations involved in research data management

In addition to the institutional stakeholders, the picture is characterised by a number of private scientific, commercial, local, transregional, and even international *temporary initiatives*, some of which transfer, distribute, and archive research data on an *ad hoc* basis while others do so with a medium-term perspective. Under these conditions, the fate of research data is currently left to chance in some cases. Nevertheless, projects funded for a limited period of time still predominate, not only in science-driven activities, but also where research data management is pursued in an institutional framework, such as re-

---

17 GBIF – http://www.gbif.org; NoMaD – https://nomad-coe.eu; ELIXIR – https://www.elixir-europe.org; Europeana – http://www.europeana.eu; EUDAT – https://www.eudat.eu/what-eudat (last checked on 25/04/2016).

18 Cf. the explanation of terms in the glossary.

19 In addition, around 20 facilities dealing with archiving have joined together in nestor, the German competence network for digital preservation: http://www.langzeitarchivierung.de (last checked on 25/04/2016).

search data and information centres, repositories, computer centres, libraries, archives, or collections.

In terms of their tasks, the profiles of digital information infrastructure services are becoming increasingly similar, even when coming from different sources. A high level of interoperability must be the goal here even though this is not always possible in every case. In the fields and institutions mentioned, new forms of scientific and IT data expertise are therefore needed. Questions on the quality, productivity, and integrity of the research processes as well as on the reuse (or possibly the reproduction) of data need to be evaluated collaboratively and cooperatively. The same applies to the generation of scientific and technical metadata. For this reason, there has been intense debate for years regarding the scope and type of personnel requirements for the area described – from *digital* library expertise and the *curation* of research data to informational and methodical interface expertise.[20]

2.3 SPONSORSHIP – FINANCING – FUNDING PROGRAMMES

Sponsorship and financing in the federal scientific landscape are as diverse as the types of data infrastructures themselves. When attempting to get an overview, it must be kept in mind that separate facilities for research data management are the exception and not the rule. Instead, there is a wide 'base' of distributed responsibility for the data, which usually lies with the research facilities or even the researchers themselves. The result is that research data – often in large amounts – is mainly stored wherever it is created: in local IT solutions, at the sites of large-scale facilities, in computer centres, in libraries and library networks, and in archives, but also at sites belonging to the contract partners of public researchers (publishers, IT companies, media partners, foundations, etc.) or in the cloud.

The established German research data repositories are often oriented towards specific disciplines or fields. They are often located in non-university facilities, are generally collaborative, and are often operated together with international partners. In addition to sponsorship funding, they receive financing from the DFG, the Federal Ministry of Education and Research (BMBF), and the research framework programmes of the EU. This can be seen by analysing the international registry of research data repositories at re3data.org.

The approximately 220 services with German participation registered at re3data.org are up against the wide 'base' mentioned at the beginning, which encompasses the entire institutional landscape. The following figures are avail-

---

20 To understand the terms interoperability, metadata, and data curation see also the explanation of terms in the glossary.

able: 110 German universities as well as 233 technical colleges, some with relatively large central facilities for storing data and others with smaller facilities; around 280 facilities for non-university research; several thousand scientific and scientifically relevant libraries, archives, and museums – each with repositories operated by external partners or independently by the facilities themselves. In addition, there are the departmental research and specialised information facilities of the German federal government as well as a heterogeneous group of facilities of the German federal and state authorities that possess data. Here are only a few by way of example: nine regional public broadcasters; almost one hundred computer centres belonging to the German federal government; and numerous public authorities with millions of data records, very few of which can be accessed openly to date.[21]

Overall system: many levels, division of labour, and collaboration

Research data infrastructures are located in facilities at all levels of the scientific system and are thus funded either by the German states (e. g. universities, museums), or else by the scientific organisations jointly funded by the German federal and state governments. Federal and state administrations as well as municipalities and companies are also involved as sponsors (archives, museums).

The tasks in the data life cycle – collection, indexing, storage, analysis, archiving, and access – are occasionally shared between the members of a collaborative network. According to the preliminary analysis of the German Council for Scientific Information Infrastructures, the universities and the non-university research facilities in particular are active in the framework of partnerships; therefore, in the area of the operation and funding of data repositories, the stereotype of a separation between *infrastructure facilities* as providers and *universities* (or *research facilities*) as users is not necessarily evident. Conversely, the information facilities operated by the Leibniz Association or the federal authorities who manage data, for example, have become more and more oriented towards research over the last few years.[22]

Financing – observed throughout the entire life cycle

It is not unusual for new information infrastructures in particular – and in view of the increasingly interlaced cooperation between computer centres as 'computation' sites and libraries as media supply sites – to require years of development until they reach maturity. In contrast to physical research infrastructures, digital services are not implemented through individual measures such as their

[21] The figures provided in the following come from a survey conducted by the RfII head office in February 2016. A study conducted by the German Federal Ministry of the Interior (BMI) as well as the German Commission of Experts for Research and Innovation (EFI) determined that the potential of open government data in Germany could be expanded significantly. Cf. EFI – Expertenkommission Forschung und Innovation (2016) – Jahresgutachten; Klessmann et al. (2012) – Open Government Data.

[22] It would appear that the determined evaluation activities of the German Council of Science and Humanities calling for greater orientation towards research has taken effect.

foundation or *construction*, but pass through beta and test phases that – supported by the response and dedication of a more or less fast growing user community – then lead step by step to further developments. Many databases, software tools, analysis platforms, and similar services emerged from research projects or were initiated as projects, e. g. through infrastructure funding programmes run by the German Research Foundation or the BMBF. It is therefore possible to obtain third-party funding for ten or more years for various development stages.

The start-up phase of a service being established in the area of research data management is guided by the concerns – as well as the financing conditions – of innovative data production, regardless of whether it is innovative in terms of form or content. This phase is characterised by a high percentage of research and development activity that follows the needs of the overall project. Research and development work in this phase is typically performed by PhD students and young postdocs, among others. The findings are publishable and help to sharpen the profile of the young scientists.

If the project infrastructures are opened later and their "products" offered to a wider public, then a consolidation phase follows, although no financing is available for this phase: an increasing number of researchers outside the project use whatever is available there. Opening up research data infrastructures for such forms of use usually has a positive effect on their quality. Problems and errors are detected quickly by users, and in the ideal case, a brief message is sent to the administrators and developers, who then make the necessary changes or improvements. In addition, new ideas from the expanded group of users flow into the development process.

Difficult financing of services

If the services gain a reputation in this manner in the research landscape, then users will not only expect high availability[23], but also advice and support in case of problems. Operations are then divided into a research area that focuses on promoting the further development of the infrastructure and a service area (help desk and feedback). Young scientists are not suited to performing these kinds of ongoing tasks in the area of scientific services, which is why the need for additional highly qualified personnel for this area is growing. Theoretically speaking, the continued operation of the services through institutional resources must be achieved by this time, or earlier if possible. Experience has shown, though, that universities and research facilities are seldom ready to offer any resources beyond those that can be provided by the existing computer centre with just a little extra effort. Third-party funding for the operation of mere services is rare due to the funding structures. It is possible, though, to apply for third-party funding for the development of the digital infrastructures.

[23] For a definition of the term 'availability' see also the detailed explanation of terms in the glossary.

There is a danger in this case that existing services will be relabelled as research projects or that the wheel is reinvented again and again. The misallocation of research funds then masks an infrastructure problem.

The infrastructures created during projects and considered to be of significant national importance were included in the joint research funding programmes of the German federal and state governments after evaluation on a case-by-case basis. Examples include the national longitudinal studies like the German National Educational Panel Study (NEPS) and the German Socio-Economic Panel (SOEP).[24] Some large data infrastructures in non-university research facilities are established directly by the German federal government and/or state governments – one example is the German Climate Computing Center (DKRZ) – while in other cases, their establishment is financed indirectly through the budgets of the sponsoring organisations, for example like the data centres of the Max Planck Society.

Few options for long-term financing

The expansion of research data services and the continuation of projects are left up to the scientific partners and their sponsoring organisations for the most part. Actual success depends to a great extent on the willingness and capabilities of the sponsoring organisation. The ability to offer universal scientific services sustainably under these conditions requires affiliation with a traditional organisation with a nationwide presence, e. g. an organisation in the library and archive system, the Leibniz Association, the Helmholtz Association, or other non-university scientific research facilities. Another option is to create new forms of cooperation, e. g. between universities at the state level or between universities and non-university research facilities. Other examples of stable self-organisation include the DFN-Verein, the association promoting the German National Research and Education Network, or the Gauss Alliance (whereby the German Council of Science and Humanities sees a potential for continued structural development here).[25]

In the federal system of the Federal Republic of Germany, there is extremely little room for manoeuvre in terms of expanding the research data infra-

---

24 The NEPS longitudinal studies were funded from 2009 to 2013 by the BMBF. Continuation was achieved after evaluation by the German Council of Science and Humanities with the foundation of the Leibniz Institute for Educational Trajectories (LIfBi). Cf. GWK (2013) – Ergebnisse der Sitzung (Pressemitteilung), p. 2; Evaluation Leibniz Institute for Educational Trajectories: http://www.wissenschaftsrat.de/nc/arbeitsbereiche-arbeitsprogramm/evaluation.html#c20161 (last checked on 25/04/2016). The German Socio-Economic Panel (SOEP) received project funding from 1983 to 2002 from the German Research Foundation. Since 2003, two thirds has been funded as a "service facility" by the German federal government, and the remaining third was funded by the Federal State of Berlin. It is located in the German Institute for Economic Research (DWI), which is an institute of the Leibniz Association. Cf. http://www.diw.de/en/diw_02.c.299771.en/about_soep.html (last checked on 25/04/2016).

25 WR (2015) – Empfehlungen zur Finanzierung des NHR.

structures used and funded together across state borders and across sectors. Bottom-up processes – for which there is no lack of incentives – often lead to short-lived solutions or even dead ends at the present time in spite of strong commitment: the picture seems to be blurry and unsustainable. The limited resources available are used inefficiently. Even when continuation is achieved through an institution, the forms, players involved, and criteria of evaluations are inconsistent.

Heterogeneous funding landscape for information infrastructures

Many research funders have recognised the need for research and development in the areas of digital, data-intensive methods and special technical, tailor-made tools. For years, innovative approaches have been supported in the form of project funding not only in computer science research, but also – in the sense of infrastructure-like support solutions – at interdisciplinary interfaces between discipline-based and IT-based research. Increasingly, statements on *sustainability* or corresponding commitments, for example to ensure continuation, are needed.

This is namely the case now when applying for dedicated infrastructure funding. However, continuation of a collaborative (or joint) service usually becomes complicated if resources belonging to the home state of a university or research facility are required for long-term financing. Similarly, the short-term projects of several – but not all – German states are hard to 'roll out' in a subsequent phase in the sense that they become a nationwide solution. Incentives to fund projects at the national level (for example through the German Research Foundation or BMBF) seldom lead to long-term solutions in spite of the commitment of the parties involved.

The funding landscape created in this manner remains unstable and heterogeneous. For this reason, the first coordinated activities to preserve scientific resources developed in projects and to establish national scientific information infrastructures have been initiated in recent years.[26] A new instrument to regulate the development of large-scale scientific information infrastructures is the National Roadmap for Research Infrastructures. It was first published by the BMBF in 2013 (see also 2.4).

Lack of authority for the development of the overall system

What remains is a large gap in authority for the targeted design and development of the overall system that particularly affects the area in the *middle* consisting of specialised, thematic, or even network-like infrastructures. Funding programmes that primarily want and should fund content-driven projects are unprepared for this, not only in financial terms. What is more, evaluations and developments that were ensured during a build-up phase are difficult to imple-

[26] Particularly noteworthy in this context is the system of specialised information systems coordinated by the German Research Foundation and the medical informatics funding concept of the BMBF as well as the German Data Forum (www.ratswd.de) or projects like GFBio (www.gfbio.org).

ment throughout the spectrum of the scientific system beyond the usual forms of funding.

Substantial need for financing and investment

Determining the total financing and investment requirements for scientific information infrastructures is considered a difficult task by all scientists and scientific policymakers involved in Germany and other countries due to the wide range of aspects that need to be taken into consideration. In any case, there is a consensus that there is a substantial need for financing and investment at various levels of the overall system.

The pure cost of maintaining the technical operation of a database can be estimated accurately based on experience. It is clear that the existing diversity binds resources and that there is a potential for synergies if the appropriate initial investments are made. In the area of data retention, large infrastructures can be more efficiently operated than small infrastructures due to economies of scale. Of course, a strong increase in the amount and use of data can be expected, which means a well-integrated research data landscape in Germany will not save money overall, but will instead increase the capability of the scientific system as a whole. This is urgently needed in order to maintain international competitiveness.

Substantial personnel requirements

There is also a frequently underestimated and simultaneously high demand[27] for resources for qualified personnel in areas that do not even exist yet – not even in terms of education or training programmes. It is the 'minds' that, through the integration of data, information, and knowledge from heterogeneous sources and across domain boundaries, create new scientific knowledge and thus ensure value is created from the data.[28] The first lines of funding have already stated the use of digital information infrastructures as a funding goal. This is a welcome development in terms of the sustainable use of digital resources and of the urgent need to develop expertise.

On top of that, managing the archiving and the long-term availability of data and data services has not been adequately clarified, and these tasks are technically and organisationally complex. Storing digital scientific collections stably over time, and possibly permanently, is a huge challenge in terms of science funding. This will be impossible to accomplish under the conditions of primarily project-based financing schemes.[29]

---

27 Estimates for the EU assume several hundred thousand positions need to be filled. Cf. the interview on 09/03/2016 with Barend Mons, Chair of the High Level Expert Group for the European Open Science Cloud, http://primeurmagazine.com/weekly/AE-PR-05-16-58.html (last checked on 25/04/2016).

28 Cf. also the explanation of the phrase 'dynamic knowledge integration' in the glossary.

29 In its 2015 position paper, the nestor competence network pointed out that long-term archiving is primarily financed on a project basis at the present time. Cf. Nestor (2015) – Positionspapier Digitale Langzeitarchivierung.

2.4 GERMANY BY INTERNATIONAL COMPARISON

The inequality and non-simultaneity of international development

Global access to scientific information has changed fundamentally in the last 15 years due to the Internet. In this context, the need for information and education, but also business and innovation policy trends, form the background for the discussion of scientific policy. Citizen movements are demanding open access to information, and open access to data in particular is suggested in studies and in the recommendations of the OECD and the G7/G8 countries.[30] Parallel to this, a UN report warned of the growing inequality internationally between "data-rich" and "data-poor" countries and of the "invisibility" of certain groups of people in terms of their data.[31]

The international landscape, though, is not only characterised by inequality and non-simultaneities in terms of how science and infrastructure providers handle research data. It is also heterogeneous with regard to strategies, paths of development, and even *questions of culture*.[32] This has a significant impact on the formation of scientific collaborations and also, from a German perspective, leads to the necessity to make adjustments and state a position.

The establishment of distributed international information infrastructures at the operational level is promoted and augmented at the level of scientific policy through the roadmap programmes in which Germany participates. At the level of scientific processes, the development of standards and best practices in international networks is also important.

International policy initiatives for research data management

Policy initiatives, among other things, are of interest to develop national information infrastructures. Internationally, a distinctly heterogeneous picture of RDM, an area that is also experiencing major changes, is evident.[33] This picture reflects the differences between the current situations in each country, including differences in terms of their size, economic power, or cultural identity, or if the state organisation is structured centrally or as a federation. Accordingly, the interplay between bottom-up and top-down activities is also more or less pronounced depending on the country.

Australia and Canada: bottom-up driven strategies

The expansive Commonwealth of Australia, for example, practices the principles of "equality". Bottom-up activities for the harmonisation of individual RDM topics stand in contrast to national top-down policy initiatives that are a central

30 Allianz der Wissenschaftsorganisationen (2003) – Berliner Erklärung; Chan et al. (2002) – Budapest Open Access Initiative (website); G8 (2013) – Open Data Charter; OECD (2008) – Recommendation on Public Sector Information.

31 IEAG (2014) – A world that counts, pp. 3 and 7.

32 This perspective is represented in a study conducted by the OECD (2015) – Making Open Science a Reality.

33 Direct insights were gained in an expert consultation of the International Orientation working group in April 2016. The descriptions for the sample countries in the following are by no means complete.

component of the changing economic structure. Through the National Collaborative Research Infrastructure Strategy (NCRIS), a project was initiated in 2009 with the Australian National Data Service (ANDS) to act as a bridge-builder during the merger of bottom-up and top-down activities. An essential element for the effectiveness of a project is a long project duration (e. g. 2 x 10 years in the framework of NCRIS funding). Successful initiatives receive special support from ANDS to provide an incentive to other initiatives.

Canada also exhibits a highly decentralised state organisational structure. The development of RDM to date has been primarily based on grass roots initiatives that then triggered selective top-down coordination due to high demand. For example, the Canadian Association of Research Libraries founded the Portage Initiative, which provides support for RDM. Discussion forums like Research Data Canada and the Leadership Council for Digital Infrastructure are attempting to effect stronger top-down structures in RDM without any financial support.

Netherlands: collaborative structures

The Netherlands is a completely different example. Top-down structures have been established in the RDM and open access strategy. The goals include the verification of data, promoting transparency, and quality assurance in the sense of good scientific practice, in addition to building up data collections. In the Netherlands, the national solutions designed to require the cooperation of several organisations stand out in particular, e. g. Research Data Netherlands in the area of science and the cross-sector Netherlands Coalition for Digital Preservation (NCDD) in the area of archiving. In 2005, DANS was founded as a shared service facility of the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO) to enable permanent access to digital research data. Researchers are integrated directly into the implementation process of the national RDM strategy. For example, in a pilot project of the NWO, approval for research funding since January 2015 has required submission of a data management plan and a successful scientific project assessment. The experience gained from this pilot project will be integrated into a comprehensive policy on open access to data. The government is also pursuing an open government vision that includes open data, e-government, and citizen participation.[34]

USA and EU: data management increasingly becoming a duty

The National Science Foundation (NSF) in the USA requires, in contrast to Netherlands, the submission of data management plans together with the project application. Since 2007, the National Institutes of Health (NIH) in the USA have required recipients of their funding to publish their findings in the form of a publication and allow access to the corresponding data. In the cur-

[34] Research Data Netherlands – http://www.researchdata.nl; NCDD – http://www.ncdd.nl/en/about-the-ncdd; DANS – http://www.dans.knaw.nl; Open Government Partnership Netherlands – http://www.opengovpartnership.org/country/netherlands (last checked on 25/04/2016).

rent research framework programmes of the EU, the requirement to publish research data is being tested in the framework of an "open data pilot". Based on the intermediate results, expansion to other participants in the research framework programme can be expected. Data management plans are a requirement for the projects taking part in the programme.[35]

Convergence, key issues

Overall, a remarkable convergence towards the following key issues can be seen internationally:

- There is a wide range of activities aimed at developing RDM strategies and RDM plans (roadmaps). One essential element would appear to be supplementing the bottom-up activities with national strategies and other instruments. These still appear to be more or less pronounced depending on the country.
- Data management plans are considered a useful tool for creating incentives and motivating the researchers.
- Sustainability has been recognised everywhere as a central problem and is being solved in part using long-term funding instruments.
- Everyone sees the necessity for data quality management (data quality controlling), but there has been a lack of measurable indicators to date, and evaluations are only available to some extent.
- The necessity to develop education and training programmes (degree programmes) has been recognised.

Germany: numerous individual initiatives

Similar developments can be observed in Germany. Research data management has gained speed through the activities of the Alliance of Science Organisations (since 2010), then through the recommendations of the German Rectors' Conference) and of the German Council of Science and Humanities (2012) as well as through its addition to the digital agenda of the German federal government. Various initiatives of German states (Länder), the Helmholtz Data Federation, the guidelines of the German Research Foundation, and some lines of funding have been initiated as a result of specific developments in order to more firmly establish research data management in institutions. However, information infrastructures for research data management with nationwide coverage do not exist, and no national strategy has been formulated either.

ESFRI: internationally distributed information infrastructures for the EU

Since 2002, strategic development of transeuropean research infrastructures has taken place in the EU through the roadmap of the European Strategy Forum on Research Infrastructures (ESFRI).[36] The ESFRI roadmap, which was first published in 2006, is an important tool for consolidating infrastructures and the processes associated with them. It receives financial support through measures in the EU research framework programmes, but the infrastructures on

35 European Commission – DG for Research and Innovation (2016) – Guidelines on Data Management.

36 Cf. http://www.esfri.eu/about (last checked on 25/04/2016).

the roadmap are essentially financed by the member states and associated countries according to the principle of variable geometry. The history of the ESFRI projects is a good demonstration of the patience required in the area of infrastructure development. In 2013, ten of the 48 projects in the 2001-2010 roadmaps were considered "success stories", and 17 others have since reached the implementation phase and are referred to in the Roadmap 2016 as "landmarks". The ESFRI evaluation in 2013 showed that most research infrastructures still needed "substantial support and guidance" even after several years of development. This applied to the management of the infrastructure, the stability of its financing, and the commitment of its stakeholders. In addition, project management, the user strategy, and risk evaluation were also assessed to be areas requiring expansion. Ethical aspects have not been adequately handled either.[37] The typical sustainability problems of infrastructures receiving funding on a project basis should be solved by making a series of adjustments. For example, starting in 2016, inclusion to the new ESFRI roadmap requires sustainable funding commitments from the participating countries and a higher level of maturity of the infrastructure.

Linking roadmap processes in the EU

The ESFRI roadmap process has been integrated little by little into roadmap processes at the national level.[38] As a result of a pilot process, Germany published its first "Roadmap for Research Infrastructures" in 2013. It contains 27 projects, 16 of which can also be found on the ESFRI roadmap.[39]

In August 2015, the national roadmap process for research infrastructures was started in Germany based on the experience gained in the pilot phase.[40] It consists of two complementary evaluation processes: a science-driven evaluation that will be conducted by a mandated committee of the German Council of Science and Humanities, and an economic evaluation of the expected costs, the maturity of the project, and the feasibility of implementation, which are closely based on the controlling criteria of the BMBF for executing large projects and which are organised by the corresponding project management organisation. Both evaluations are conducted with the help of external international experts and are expected to take a vote on the research policy priorities of long-term oriented research infrastructure projects in the summer of 2017. The national roadmap process for research infrastructures is not a funding programme, but a process for supporting and providing a basis for strategic research policy decisions. It is basically intended to promote the selected projects, although

37 European Commission (2013) – Assessment of ESFRI Projects.

38 Subsequently, 22 member states have implemented national roadmaps. Cf. European Commission (2015) – European Research Area. Facts and Figures 2014, p. 21. For an overview of the status in each member state, see http://www.esfri.eu/national-roadmaps (last checked on 25/04/2016).

39 BMBF (2013) – Pilotprojekt nationale Roadmap; WR (2013) – Bewertung Forschungsinfrastrukturvorhaben.

40 BMBF (2016) – Nationaler Roadmap-Prozess.

an application for support can only be submitted after successful completion of the roadmap process. The development of the evaluations by the German Research Foundation (Commission on IT Infrastructures and LIS – Scientific Library Services and Information Systems) has demonstrated over the last 15 years how slow the evolution of evaluation processes progresses when there are no national roadmaps to support them.

In the course of the European roadmap processes and the international cooperation, networked information infrastructures will be formed that will be fed by several institutional or national "nodes". Distributed locations not only form networks, but also integrate themselves into an extensive international infrastructure. The European Commission is planning the formation of a European Open Science Cloud as well as a European data infrastructure for high-performance computing in the coming years. The initiative is intended to promote the development of open science and the creation of a European Digital Single Market.[41]

Special organisational structures

Distributed information infrastructures are now considered separate organisational structures requiring special governance as well as a special legal form.[42] At the European level, a separate legal form was created in 2012, the *European Research Infrastructure Consortium* (ERIC). ERICs are legally independent constructs with defined governance structures that hire personnel, may be receivers of third-party funding, and are not required to pay value-added tax. Setting up an ERIC requires the approval of the European Commission.[43] In Germany, it has been possible to found ERICs since 2013, although the ERIC framework primarily regulates collaborations between international partners; consequently, an ERIC cannot be founded as a purely German enterprise. There is no national equivalent. Representatives from the humanities and cultural sciences, for example, have pointed out that digital research infrastructures need specific organisational forms, but these have not yet been developed.[44]

German participation in internationally distributed information infrastructures usually requires consolidation at the national level. One example of this is the ELIXIR life sciences repository, where the lengthy formation of national nodes delayed Germany's participation for a long time.[45] A successful example, on the other hand, is the DARIAH project supported by 18 European countries with its

41 European Commission (2016) – Data and Knowledge Economy.

42 GSF-Global Science Forum (2014) – IDRIS. Issues and Options.

43 European Commission (2010) – ERIC Legal Framework.

44 DARIAH-DE (2016) – Memorandum digitale Forschungsinfrastrukturen (website).

45 Consolidation now appears – after a preliminary phase of about ten years – to have been achieved in the form of the de.NBI network (German Network for Bioinformatics Infrastructure), cf. https://www.elixir-europe.org/news/elixir-and-denbi-agree-collaboration-strategy (last checked on 25/04/2016).

German node DARIAH-DE.[46] There are already 21 project partners from Germany involved in DARIAH; DARIAH has reached the legal status of an ERIC.

Overall, Germany is one of the very active countries when compared to other countries in the EU, both in terms of the development of internationally networked research infrastructures as well as in terms of its participation in these infrastructures. For example, Germany participates financially in 18 of the 48 ESFRI projects, and several research infrastructures are located in Germany.[47] However, the problem of long development times, the unsolved problem of long-term financing, and the problem of acceptance by researchers still exist. There are only a few comparable initiatives in global scientific cooperation networks.[48] Another exemplary project worth mentioning is in the area of geosystem and Earth system research, namely the Global Earth Observation System of Systems (GEOSS)[49], of which Germany is a member.

Standardisation and best practice – global initiatives and their admission in Germany

The trend towards networking information infrastructures across borders creates a need for standardisation and harmonisation. Independent of the roadmap activities organised as *top-down* activities, a variety of best practices and standardisation initiatives have developed internationally from the *bottom up*. These initiatives have produced recommendations and technical standards, but the large number of recommendations and technical standards produced sometimes encourages fragmentation and the formation of silos within the international landscape instead of reducing it.[50] In this situation, the global RDA founded in 2013 – with German participation as well – is endeavouring to achieve consolidation.

Initiatives in individual EU countries like the *Data Seal of Approval* developed in the Netherlands or the German nestor Seal for Trustworthy Digital Archives have also attracted interest internationally and are correspondingly networked (cf. 2.5). The German Data Forum (RatSWD), which very actively contributed nationally to the development of "good practices" for research data centres, adopted an internationalisation strategy in 2014. Our impression to date, though, is that the offer to exchange experience and contribute to the development of the RDA and other organisations in Germany is still relatively un-

---

46 DARIAH-DE – https://de.dariah.eu; DARIAH-EU – https://dariah.eu (last checked on 25/04/2016).

47 BMBF (2014) – Internationale Kooperation.

48 However, both the OECD Global Science Forum as well as the global associations of research promoters are discussing common initiatives to form globally networked research infrastructures.

49 GEOSS – http://www.earthobservations.org/geoss.php (last checked on 25/04/2016).

50 See also Dally/Fless/Förtsch (2012) – Altertumswissenschaften (Classical Studies) for more information.

known in general. However, the RDA-DE[51] is currently becoming a multiplier at the national level.

The effects of these initiatives on the promotion of the international *research data market* and its infrastructures will most likely depend on a reliable and adequate supply of resources. Currently, the initiatives are supported by the dedication of voluntary experts, although this also comes at the expense of research or duties in the corresponding home organisation and offers few incentives. Sporadic participation and a lack of strategic feedback in the home organisations pose a risk in this regard.

2.5 ENABLING STRUCTURES AND SPECIAL DEVELOPMENT TASKS

One aspect of the consolidation of data management initiatives is to create orientation. Researchers and decision-makers in the scientific system are just as dependent on this as the wider public, businesses, and politicians. The need for improved findability and identifiability of data is now being served by the first registries and databases tailored to meet the needs of research based in Germany: the RIsources database of the German Research Foundation, for example, maintains information on around 300 research infrastructures (about two thirds of which are information infrastructures and data repositories). re3data.org, a registry that specialises in data repositories, lists almost 250 internationally oriented research data repositories with German participation. These services are for the most part run collaboratively by two or more partners. The GESIS information centre has established a registry called da|ra especially for social science research data. Databases of this type are currently one of the few sources where it is possible to find quantitative data on the development of the future German infrastructure landscape, which means they are also of interest to users from the area of infrastructure policy. The potential of these registries, however, has only been recognised within smaller circles.[52]

Findability through identifiers

Persistent identifiers are a prerequisite for the functionality of such registries, which could also be referred to as *enabling structures*. Identifiers guarantee the findability and networking of individual published data sets, people, and organisations. In recent years, globally operating services have been developed

51 Experts from Germany who are active in the RDA organise the exchange of experience through workshops – cf. http://www.forschungsdaten.org/index.php/RDA-DE (last checked on 25/04/2016).

52 German Research Foundation RIsources – http://risources.dfg.de/home_en.html; da|ra – http://www.da-ra.de; re3data – http://www.re3data.org (last checked on 25/04/2016). The information on the number of services registered was obtained in April 2016.

for this purpose.[53] DataCite[54] is a successful bottom-up initiative with considerable German participation.

The currently prevailing boom promotes the creation of numerous different kinds of services for different user groups. The first dedicated scientific standardisation and quality management initiatives for information infrastructures offer the potential for orientation and professionalisation in this regard. The path chosen offers a voluntary certification process organised within science itself. In Germany, this includes accreditation through the German Data Forum (RatSWD) in the area of social and economic data, the nestor Seal for Trustworthy Digital Archives (DIN 31644 standard), and the DINI certificate for open access repositories, which was established several years ago. Internationally, for example, the EUDAT and RDA are striving to create a uniform framework for certificates and accreditation.[55]

Certification and accreditation as a step towards professionalisation

The number of certified or accredited data infrastructures thus far is relatively small. Nevertheless, it is becoming apparent that, in conjunction with certification initiatives such as those in Germany associated with the German Data Forum (RatSWD) or in the form of the international *Data Seal of Approval* community, communities of interest are forming that are developing their own particular standards and processes and offering to exchange experiences.[56] Commitment initiatives designed for a large number of users like the recent *Joint Declaration of Data Citation Principles*, which will be augmented on the operational side by the *Data FAIRport* initiative for the development of tools, are an easy way to achieve this.[57] Whether or not both approaches will be successful is still unknown at the present time. There is also fragmentation internationally when it comes to best practice initiatives (cf. 2.4).

Shared services, potential synergies

Undoubtedly, there is a potential for synergies between services, especially when the application layer containing the services accessible to the user and

---

53 For example the Digital Object Identifier System (DOI) – https://www.doi.org and the ORCID Initiative (Open Researcher and Contributor ID) – http://orcid.org (last checked on 25/04/2016).

54 DataCite – https://www.datacite.org (last checked on 25/04/2016), the central business office is located at the German National Library of Science and Technology (TIB) in Hannover.

55 German Data Forum – http://www.ratswd.de/en/data-infrastructure/info; nestor seal – http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.html; DINI certificate – https://dini.de/dini-zertifikat/english (last checked on 25/04/2016).

56 31 research data and data service centres at the German Data Forum, and 57 services (one third of which are German) in the Data Seal of Approval initiative. Cf. http://www.ratswd.de/en/data-infrastructure/rdc and http://datasealofapproval.org/en/community (last checked on 25/04/2016). All figures as of April, 2016.

57 Data Citation Synthesis Group (2014) – Joint Declaration of Data Citation (website). Signatories commit to follow certain guidelines for the fair re-use of research data based on the model of the various open access declarations; Data FAIRport – http://www.datafairport.org (last checked on 25/04/2016).

data retention, i. e. the physical storage systems and their associated services, are considered separable tasks. This is reflected in the 'separate' inter-university cloud initiatives for science in Germany, the merger process in the area of joint catalogues, and regional/technical cooperations between different universities or between universities and non-university research facilities. Examples include the operation of a joint computing centre (GWDG) by the University of Göttingen and the Max Planck Society or the establishment of the interdisciplinary RADAR repository by a consortium of universities, Helmholtz and Leibniz institutes, and information facilities.[58] Cluster projects receiving EU infrastructure support and the working groups of the Research Data Alliance have explicitly stated the identification of potentially universal digital architectures and services as their goal.[59]

Potentials for synergies also exist between the area of research and other sectors in similar transformation processes – such as memory institutions and archives or in the (frequently poorly networked) public administrations. In addition to technical and procedural standardisation to enable networked data exchange and quality assurance, the question of long-term archiving in particular must be mentioned as a common task with obvious potential for synergies, which need to be further examined in each case.

58 GWDG – https://www.gwdg.de; RADAR – https://www.radar-projekt.org (last checked on 25/04/2016).

59 For a list of comprehensive services see Field et al. (2013) – Realising the full potential of research data.

3 STANCE ON FUNDAMENTAL ISSUES

When the RfII provides recommendations for the complex area of research data management at the system level, it states its positions on obvious, unresolvable, and conflicting issues. These positions will be described briefly – as answers to fundamental questions arising in the corresponding issues. These answers are simultaneously maxims and form the background for the subsequent recommendations of the RfII (Chapter 4).

Top-down vs. bottom-up

The Joint Science Conference (GWK) has already stated that processes of self-organisation in science should be given priority when considering *bottom-up and top-down approaches to political control* in their expectations regarding the work of the RfII. The members of the RfII agree with this in principle. In light of the challenges of handling research data, though, top-down incentives are also needed. However, it is actually a matter of triggering processes without subjecting them to a centralised approach to a solution. The goal is the *dynamic integration of distributed knowledge*[60] in order to enable good science. Accordingly, one of the RfII's suggestions involves the establishment of coordinating organisations and incentive systems that promote cooperation. They should take effect in a dedicated national framework, but should also link different, even distributed stakeholders and institutional components with each other.

Project funding vs. institutionalisation

The classic funding policy conflict between *project funding and the need for institutionalisation* of what is indispensable to the system is particularly evident in the area of research data management. The value of digital data and the benefits of using this data depend directly on the strongly organised, permanent, and technical availability of the data. The RfII also endorses the testing of new digital processes and methods in the form of projects for the area of infrastructures. Nevertheless, no permanent, sufficiently transparent, and resource-efficient network of services for the storage and use of research data in Germany can be created through project funding. Against this background, the RfII sees all stakeholders responsible to achieve long-term solutions, when services have proven themselves and relevant expertise has been built up.

Standardisation vs. diversity

Possibly the greatest challenge is rooted in the internal heterogeneity of discipline-specific requirements, forms of research, and methods – i. e. in science itself. It relates to the conflict between the necessity for technical and organisational *standardisation* and the necessity for technical and organisational as well as content *diversity*, on which digitised research thrives today in spite of all the internal differentiation. The RfII considers the creation of universal conditions for interoperability a decisive factor in this regard. Transitions and (from the perspective of research content) "translatability", for which coordinated

60 Cf. the explanation of terms in the glossary.

standards and interfaces are prerequisites, represent a qualitative feature of a system that on the one hand ensures diversity, but on the other hand opens up variable – as well as fast-growing and enabling – interconnections.

Individual solutions vs. comprehensive services

Similar concerns about interoperability apply to technical requirements. Nonetheless, in choosing between *individual solutions* on the one hand and *generic services* on the other hand, it is indisputably possible to organise basic services sensibly across different disciplines and different forms of research. The RfII considers the principle of a cross-location and cross-institutional approach to be reasonable, possibly based on collaborative services, a significantly greater willingness to cooperate, as well as new forms of cooperation between scientific organisations. Cooperation is required between universities as well as between universities and non-university research facilities, between science, culture, and public administration, or even across state and regional borders.

Openness vs. data sovereignty and data protection

The goal of *open use of data* also leads to conflicting priorities. In some cases, *sovereignty over research data* will still be desired for good reason, for example for quality assurance purposes or to prevent or regulate the use of the data for commercial purposes. Furthermore, regulatory issues relating to *data protection must be taken into account* to the full extent of German and European standards. The RfII subscribes to the open science paradigm in principle, as is currently the case for numerous scientific organisations. At the same time, the RfII does not deny that dealing with the boundary between science and business raises a series of questions. To establish the most *open* culture for sharing research data possible, especially among the people involved in science, it is necessary to erect legal barriers to prevent global players from exploiting easy access to data in order to claim ownership of something that was intended to circulate *freely* (in the sense of a scientific common) among researchers.

Government funding vs. commercial funding

Accordingly, it is necessary to strike a balance between *government funding and commercial funding* with the appropriate foresight. Cooperation partners such as publishers or telecommunications companies classically belong to the scientific system; the same has applied for a long time to providers of other fee-based media services and even to software providers. Of course, dependencies can arise wherever the government backs too far away from infrastructure issues. This is not only critical financially, but can also endanger the robustness of the processes and the diversity of the methods used in research itself – for example due to monopolisation, bankruptcy, or a lack of redundancy of external solutions. The RfII therefore also endorses utilising public-private partnerships, but stresses the need to design them to be reversible. An excess of privatisation must be prevented, especially within the realm of what will remain relevant to research data and information infrastructures over the long term.

Responsibilities of the federal government and the state governments

In the multi-tiered system existing in the Federal Republic of Germany, conflicting priorities also arise due to the *division of authority between the federal government and the state governments (Länder)*. To handle the large amount of heterogeneous research data being produced everywhere, which is rarely only locally relevant and is impossible to handle using *distributed* solutions, high performance, quality assured, and efficient solutions can only be found when policymakers also emphasise collaboration. In many cases, the requirements of dynamic knowledge integration would be fulfilled for the most part if national solutions, or even processes for determining priorities, could be created through intelligent, cross-border cooperations between the German states (Länder). Consequently, problems in research data management do not *per se* require either the German federal government or non-university research facilities to be the dominant player. The RfII considers the German federal government and the German states to be mutually responsible for taking on the challenges arising in the context of research data and information infrastructures in a coordinated manner. Knowledge integration does not stop at state borders and thus requires national solutions that counteract the fragmentation of activities currently prevailing in the area of research data management.

Investment in infrastructures vs. investment in 'minds'

The final important area of conflicting priorities to be mentioned is the funding policy *alternative between investing in infrastructures and investing in 'minds'*. In the area of digital infrastructures, the focus for years was on investing in equipment and technology – namely in mainframe computers, cable networks, software, licenses, and energy. For several years now, a paradigm shift has been taking place, in which infrastructures need to be oriented towards increasingly complex services, presented to users, and finally financed: hardware, software, etc., only become tools for research when used in conjunction with qualified expertise that is available at all times. In the increasingly data-intensive domain of science, it is necessary to create new fields of activity. The RfII therefore calls for the establishment of infrastructures for managing research data – at all levels of qualification existing in science – to always be considered a matter of investing in minds as well.

The previously mentioned objective of achieving the dynamic integration of data, information, and knowledge can only be realised step by step. This means structures, processes, and financing routes need to be developed towards a form of convergence that does not conflict with diversity or *distributedness*. However, determination and the willingness to realise these developments is required because the alternative is stagnation and the increasingly inefficient use of public resources. Consequently, it is necessary to take these steps as soon as possible.

4 RECOMMENDATIONS

4.1 LONG-TERM FINANCING OF INFRASTRUCTURE PROJECTS

In view of the rapid development of digital technologies and the changes to how research is conducted due to this development pressure, it appears appropriate to test infrastructure solutions in research settings as well as in the form of projects. In addition, it is necessary to gain a clear impression of what long-term financing processes could look like. The objective should be to create a system of sustainable information infrastructures that offer reliable work structures, enable dynamic development, and ensure high quality provision of basic services in Germany (cf. also 4.2).

From the RfII's perspective, services dependent on project funding over a long period of time pose a risk to the development of a sustainable infrastructure.[61] On the one hand, the goal should be to design the project funding scheme so that there is a clear understanding before the project is completed of whether or not long-term financing will be available for the infrastructure and which steps are necessary to obtain this financing. The small-scale approaches associated with project-based financing as well as the loss of know-how associated with high staff turnover stand in contrast to the need for consolidation, standardisation, and the availability of expertise over the long term. When it comes to the permanent archiving/publication of their own research data, researchers will lack trust in cases where the prospects for secure long-term financing are unclear. On the other hand, approving large amounts of funding over long periods can be a disadvantage in areas of science that are subject to rapid changes. For this reason, mechanisms must be available for making changes in cases where funding measures are found to be ineffective.

Even successful infrastructure projects can only continue development for a period of about ten years at most and stay 'on the market' when receiving funding on a project basis (cf. also 2.3). This is why important, successfully 'fostered' infrastructure projects such as DARIAH-DE have now reached a structural limit. Sponsoring the sustainable operation of such research infrastructures having only little or no connection to institutionally financed facilities unfortunately appears to be infeasible in the current framework in Germany because a suitable sponsorship structure is lacking.

Phase model for the development of information infrastructures

4.1.1 To sensibly augment the use of public funds in infrastructure projects, the RfII recommends a phase model that enables consolidation and allows decisions on which route to take to transfer the project into a suitable sponsorship

[61] This assessment is based on the results of the German Research Foundation workshop "Concepts for successful scientifically relevant information infrastructures" (2014).

structure to be made. The transition from the project phase to long-term financing/consolidation also needs to be based on an evaluation. The initial goal of the evaluation is to analyse the scientific quality, acceptance, and relevance of the infrastructure in question. In general, the RfII considers the FAIR data principles to be reasonable guidelines.[62] In a second step, the options for long-term financing are to be examined on a case-by-case basis. As part of the examination, suitable experts should provide recommendations for continued operation and permanent consolidation of the infrastructure, including its personnel resources (cf. 4.1.2 and 4.1.3). Scientific policy aspects should also be examined. On the one hand, this includes technical and geographic aspects as well as the institutional focus, while on the other hand should ensure adequate plurality and competition within research as well as analyse the development potential of the projects in the new structure. This phase model should also be used as a guide when the project phase is financed through institutional funding from a non-university research facility.

Although research projects should always be designed and executed so that the corresponding research data collection can be transferred or integrated into an overall research data infrastructure, the following options for long-term funding are still available at a minimum:

- Continued operation as a singular infrastructure
- Incorporation into an existing infrastructure
- Integration into an existing central service already available in a national infrastructure for long-term reuse (cf. 4.2)
- Merger of critically under-equipped infrastructures into overall models of operation
- Archiving the status reached
- Improvement of the cost-benefit ratio through reduction to the core elements with the highest priority from a long-term perspective.

Planning capability and minimum standards

To improve the ability to plan these alternative models, comprehensive minimum standards for data retention must apply to all projects and throughout the entire data life cycle. Any additional requirements arising in a project must be integrated into the comprehensive standard over the medium term. This facilitates active support for the processes of standardisation during the establishment and operation of the particular data infrastructures.

To avoid critical funding gaps and to ensure that the wrong incentives are not created due to variable project funding conditions, the transfer of successful projects to a suitable sponsoring organisation or to institutional funding, name-

[62] FAIR = Findable, Accessible, Interoperable, Reusable: https://www.force11.org/group/fairgroup/fairprinciples (last checked on 25/04/2016). Cf. also diagram the Data FAIRport initiative in section 2.5.

ly to basic funding from the sponsoring organisation, must be organised at an early stage.

Independent evaluation

4.1.2 In the development scheme, the establishment of an infrastructure can be funded by institutions or through consecutive third-party funding. In general, the corresponding funder is then responsible for conducting evaluations in the project phase. Breaking points are created due to gaps in authority, e. g. when no suitable funding programme is available and several funders or sponsors will be providing partial funding, instead of one sponsor providing all funding. If institutionalisation is being considered, then an independent body must provide guidance in the form of an evaluation for the decision on which path to take in the next operating phase. It is also recommended to perform such evaluations on smaller infrastructure projects. Suitable players in the scientific system can be appointed to conduct these evaluations. They need to take into account the needs of the scientific system as a whole as well as other relevant aspects in addition to scientific excellence (cf. 4.1.3).

Concept for regular evaluations

4.1.3 The scientific evaluation system in Germany is traditionally oriented towards topical and methodical *discipline-specific* evaluation procedures and procedures oriented towards research *content*. Initial concepts for the evaluation of information infrastructures are available, but only a few have been thoroughly tested or are generally accepted. At the same time, it has been impossible until now to obtain easy access to pools of experts with the necessary cross-disciplinary expertise.

The RfII therefore recommends the development of a concept for regular evaluations of research data infrastructures that takes the particular methodical and cultural aspects of disciplines into account and states the basic goals and criteria for such infrastructure-based evaluations. In addition to evaluating the scientific quality and the user orientation, the data’s relevance to society and its potential for commercial use should also be evaluated.

4.2 FOCUS AND DIVISION OF LABOUR

The current situation in the management of research data in Germany is characterised by numerous activities and investments that, owing to circumstances, are poorly coordinated. Although the demand for coordinated services is increasing, not enough services have been created to establish a reliable, nationwide research data management system. If an efficient, nationwide system is to be created, then restructuring and aggregation according to function is necessary (cf. 4.1.1).

Such an aggregation cannot be designed according to a master plan and then implemented from the *top down* due to the high level of complexity involved.

In order to create new structures organised as collaborations, a combination of strong networking incentives and at least a minimum of indirect coordination is required. The most important tasks are to develop support structures and bundle existing expertise in data curation and analysis methods using an overall process.

The goal is to expand the capacity available for research data management and, in the long term, to build an integrated system of research data infrastructures. The RfII has provided an outline for this without describing any of the implementation decisions in detail.

Provision of basic storage infrastructures and services

National research data infrastructure

4.2.1 Many aspects of research data management are of a generic, and thus transferable, nature even when the data generated in the scientific system is highly heterogeneous as a whole. With a view to costs and efficiency, generic services can and should be designed and offered collaboratively. The RfII therefore suggests founding an association that bundles the existing expertise and ensures the provision of basic storage infrastructures and services as well as the fast transfer of know-how in the scientific system. This *National Research Data Infrastructure* (NFDI) should be established in the form of a network, be designed for use by all disciplines and communities[63], and integrate the existing large-scale information facilities as well as the ESFRI projects and the repositories of user groups at the national level, provided the needs of these user groups are sufficiently homogeneous. As a network, the NFDI must have a system of governance designed to allow action to be taken and to control development (initially in this case: a transition process). Nevertheless, there should be participation options available that take the diversity of the stakeholders involved appropriately into account (cf. 4.4).

Overcoming fragmentation

4.2.2 The *National Research Data Infrastructure* (NFDI) is a networked, distributed infrastructure for the creators and users of data, specialised services, and data services with common user access and a strategically positioned, decision-making coordination unit. An open, *evolving* structure created for the purpose of integration allows data from different communities to be combined for further scientific analysis, including big data applications, and is designed to overcome fragmentation. The basic principle of a collaborative process permits relevant specialisation, a focus on key topics, and the foundation of centres and "nodes" to perform higher-level tasks.

As a functional network, the NFDI must meet the following national challenges affecting the overall system:

[63] Cf. the explanation of terms in the glossary. The more narrowly and more or less socially defined term "(specialised) community" stands in contrast to the more taxonomic term "discipline".

- The definition of overall minimum standards and quality management for describing and retaining data;
- The development of generic data analysis methods and their application to specific data and questions;
- The development, establishment, and provision of generic data services and data repositories as well as interfaces to connect distributed local repositories;
- Basic training, advanced training, and education.[64]

The ability to link to European developments, specifically the European Open Science Cloud, must be taken into account as well as the links to existing repositories and structures and to individual disciplines and specific services. Interfaces are needed for this purpose that connect the various layers of data and enable connections to generic research data management services. In addition to the collaborative development of the technical infrastructure for storing data and providing professional user support, the methods and software for handling data and for data analysis (research, aggregation, visualisation, etc.) could also be developed collaboratively.

Planning the IT, development of methods, and service in the same context

When establishing the NFDI, it is recommended that the technical features, the support, and the development of methods/data analysis are viewed and planned in context. The concepts for the establishment of the National High Performance Computing (NHR) programme could be used as a reference in this regard.[65]

Due to the highly dynamic developments and the lack of clarity regarding future requirements for research data management, the processes for further development of the NFDI should be designed with the necessary level of openness. At the same time, it is necessary to coordinate this infrastructure with a wide circle of scientists and scientific organisations to ensure that the existing structures are included and their needs are met to the greatest extent possible.

Establishing overall infrastructure centres

4.2.3 As sponsors of scientific (or scientifically relevant public) facilities, the German federal and state governments, and possibly municipalities and private foundations, possess different capabilities and conditions for the establishment of research data infrastructures. In light of the great institutional need (cf. 2.3), it is recommended to develop infrastructure services together with data producers, operators, and data users that operate nationwide and ensure basic provision of research data repositories together with the corresponding user support services. These centres should have the capability to link to interna-

---

64 The list should not be considered complete. Tasks associated with data protection and data security – which are important in all areas of research data management – must also be considered, for example.

65 WR (2015) – Empfehlungen zur Finanzierung des NHR.

tional initiatives. Especially in the currently rapidly developing area of research data repositories, which should store research data from individual projects *professionally*, local structures will only have scientific value over the long term when connected to comprehensive repositories. Tasks such as long-term archiving (see 4.3) and the operation of large collections of research data require continuity as well as a critical amount of expertise and resources wherever necessary. In turn, it is recommended that decision-makers at the management level in universities, research facilities, and infrastructure facilities work wherever possible in network structures (which may need to be created) that can be integrated into the NFDI and that they consider establishing connections to larger infrastructure centres. The latter in particular also applies to the development of cross-institutional data infrastructures in scientific communities and research networks.

Step-by-step networking

4.2.4 The high specificity of these tasks will entail the step-by-step establishment of the NFDI. First of all, universal standards[66] for describing data are needed that are comprehensible across communities or even across different forms of research and that provide researchers with the information needed to determine if the data is of interest to their research (cf. also 4.6.1).

Once the minimum requirements are fulfilled, additional steps can deal with the technical standardisation of the infrastructures to be connected to the NFDI. This will enable a wide circle of researchers access to various resources and lay the foundation for combining, and therefore assessing, data from different sources. The minimum requirements include common standards for the quality of the data as well as their description using metadata, documented rules for access rights (possibly in the sense of open data), and rules for reuse.

Jointly operated services

It is recommended to enable universal access to the connected services for registered users, to establish central services for access rights and read/write permissions, and to create general rules to control how different types of users access the data. Other enabling structures are to be defined in good time in a detailed concept. It is also recommended that the experience gained from the establishment of similarly situated initiatives is reviewed.[67]

Services for good scientific practice

4.2.5 Larger infrastructure centres can provide services for the establishment of repositories in the form of hardware and in the form of methodology know-how and consulting. This would create economies of scale which, in combination with the professionalisation of the services, would lead to greater cost efficiency and ensure data quality. The RfII recommends that infrastructure

66 To gain an understanding of standards, cf. also the explanation of terms in the glossary.

67 For example the approaches pursued in projects like OpenAIRE (https://www.openaire.eu), Europeana (http://www.europeana.eu), and the Deutsche Digitale Bibliothek (German Digital Library, http://www.deutsche-digitale-bibliothek.de, all links last checked on 25/04/2016).

providers create incentives that will convince researchers of the benefits of the services offered through the provider's expertise in methods and impressive service. This naturally includes good management of the research data as a standard component of good scientific practice. Furthermore, the researchers creating the data must always have sovereignty over the data they create. When researchers build new repositories from the *bottom up*, it is recommended that they remain conscious of costs and use existing services already offered by infrastructure facilities.

Federated structure and connection to local services

4.2.6 Within the research data infrastructures, there will not only be a division of labour, but a functional differentiation as well. In view of the complexity of the task of data retention, the level of specialisation, and the scope of user support/dynamic knowledge integration, it would appear advantageous to make several distinctions. The RfII considers the following, non-exhaustive list of distinctions reasonable:

- The broad-based infrastructure centres providing basic services in the NFDI were already discussed in section 4.2.4. It is recommended to implement competitive control elements so that these infrastructure centres compete for users and funds. This could enable centres of excellence with different areas of focus to emerge that consider the special demands of large communities while simultaneously covering the needs of smaller communities by offering generic services. Such centres of excellence could be created in facilities already managing large community repositories, for example.[68]
- Some areas of the scientific system generate high-volume data collections (high throughput screenings, simulations). It is recommended to link data centres with high analysis and processing needs to correspondingly equipped high performance computing (HPC) centres of excellence over the long term.
- In a national federated infrastructure, it will still be necessary to provide financing to smaller players who take care of the provision of basic services. Examples of such players include libraries, archives, small computer centres, or the IT infrastructures of scientific institutions. They form a separate layer in the NFDI that is also based on cooperation, and their functionality is to be integrated into the system through suitable organisational structures.
- Scientific data is collected at different locations and in different types of facilities. It is recommended that these resources are connected to appropriate centres of excellence for the corresponding discipline or even to larger infrastructure centres in the future (cf. 4.2.4). There must be a case-

[68] The boundary between an infrastructure centre (which is normally a broad-based service provider) and a centre of excellence (which primarily reacts to specific needs) is not well defined, and in individual cases may even be non-existent.

by-case analysis to determine whether data should be physically integrated into a larger data centre for the purpose of efficient access or long-term storage. In principle, smaller research data centres operated at the level of working groups, for example, should be allowed to integrate themselves into suitable host facilities.

- Data storage services (scientific clouds) must be expanded slowly over the long term and in line with demand in order to offer users the greatest benefit possible. It is recommended that smaller local storage infrastructures are linked together virtually at first by implementing standardised data access protocols. Over the long term, such locally stored data should also be migrated to ensure adequate long-term availability of the data and to lower costs.

The large infrastructure centres should be able to perform backup functions for each other. It is recommended to link the establishment, operation, and expansion of a given infrastructure centre to regular evaluations.

Long-term process of transition

4.2.7 Financing the NFDI is a challenge that transcends the normal framework of ongoing development processes in research. In addition to requiring an indispensable increase in the commitment of institutionally funded scientific stakeholders – which is only possible within limits, though – it is also a joint task of the German federal and state governments. It is necessary to coordinate this infrastructure across the broad scope of the scientific system to ensure that the existing structures are included and their needs are met to the greatest extent possible. The RfII assumes such a transition phase will take 15-20 years.

For this reason, adaptive and long-term functional coordination structures as well as suitable pilot projects must be developed for the NFDI network. Due to the rapid pace of development, the processes for further development of the NFDI must be designed with the required degree of openness. Once the NFDI's specific needs are clear, the RfII recommends examining its inclusion in the national roadmap for research infrastructures (cf. 4.13).

4.3 LONG-TERM ARCHIVING AND LONG-TERM AVAILABILITY OF RESEARCH DATA

The challenge of ensuring the long-term availability, verifiability, and usability of selected research data[69] does not only apply to the domain of cultural heritage and, in general, publishing, but also to the digital data created and used during the research process. This entails complex tasks extending far beyond the simple job of data storage that also bind additional IT resources. On

[69] "Selected" means there are also questions regarding relevance criteria (see also 4.3.3).

the one hand, architectures, processes, and services that ensure universal and discipline-independent functionality (quality assurance, standardisation according to internationally accepted criteria, data persistence, access security in the sense of authentication/authorisation, and search functions) need to be programmed, operated, and developed (further). On the other hand, discipline-dependent aspects such as documentation of the discipline-specific context of the data must also be taken into account.[70] In its 2015 position paper[71], the nestor competence network stated the primary tasks at the general level in the area of digital preservation as they relate to digital research and education. However, ensuring the long-term availability of data, especially research data, is associated with requirements extending beyond those stated by nestor.

Scientific discourse on archiving periods

4.3.1 The RfII encourages scientific discourse on the distinction between storage for the duration of a project and storage for significantly longer archiving periods. The long archiving periods necessary for digital research data have not been adequately taken into account at the present time. The German Research Foundation, in its recommendations for "Safeguarding Good Scientific Practice", suggests retaining the primary data for a period of ten years (but not necessarily in cases where open access is provided).[72] This retention period ensures the reproducibility of research processes and the verifiability of results – but only for a relatively short period. For many disciplines as well as basic research problems, the period for which research data should remain available for reassessment or verification is much longer.

Guidelines for long-term digital archiving and availability

4.3.2 In general, a variety of aspects need to be considered and guidelines are required for long-term digital archiving. First, it is necessary to protect the data (in the same state as it was stored) against risks. Methods for protecting data include redundant – and possibly even geographically distributed – data storage as well as procedures to check the integrity of the data regularly and restore it when necessary. Such measures are sufficient for relatively short periods of several years, although sooner or later it will be necessary to migrate the data once the physical data media need replacing. The longer the data needs to be retained, the more likely it will be necessary to take measures that not only preserve the data technically, but also preserve its content (referred to as *content preservation*). Necessary measures could include migrating the data to new formats or making changes to adapt to new hardware and software environments. If the data is stored in formats that cannot be read any more in new software environments, then long-term use of data can only be ensured by performing complex conversions (format migrations). To minimise the frequency of such format migrations, the data objects should be archived in formats that will be processable for a long time (e. g. csv, xml). In light of the

70 Neuroth et al. (Ed.) (2012) – Langzeitarchivierung.

71 Nestor (2015) – Positionspapier Digitale Langzeitarchivierung.

72 DFG (2013) – Sicherung guter wissenschaftlicher Praxis, p. 21.

wide variety of metadata formats used in the various disciplines, the RfII sees an urgent need for stronger agreements on standards in this field. The RDA is active internationally in this area, for example.

It is recommended that researchers and their facilities create an adequate RDM plan and develop the tools, services, and workflows needed throughout the data life cycle in good time and in cooperation with the NFDI service providers (cf. also 4.6.2 for information on data management plans). In the future data producers (researchers) should have tools available on-site that help them configure their data so that it meets the requirements of long-term availability (including its so-called *curation*). Developing and providing such tools is the task of discipline-specific and discipline-independent data centres in the framework of the NFDI.

Relevance assessment by the discipline itself

4.3.3 To make decisions regarding the scope and volume of data to be archived over the long term (possibly including preservation and preparation) and choose the right paths for ensuring long-term availability of the data, it is necessary to link scientific considerations, in particular those related to the culture of the corresponding discipline, with the unavoidable economic considerations. In general, it can be assumed that not all data will need to be stored forever. Generic cost-benefit analyses in the area of long-term archiving are nevertheless difficult because the potential scientific relevance (as well as the potential economic value) of research data in the future cannot be estimated with certainty. In the opinion of the RfII, an initial assessment of the relevance of a given research data collection can be conducted by the corresponding discipline itself. Otherwise it will be necessary – especially for *orphaned* data, similar to the handling of analogue stores – for data archivists to decide whether data stores and data collections should be preserved or destroyed based on previously defined criteria.

Long-term archiving organised as an ongoing task

4.3.4 Long-term archiving and ensuring long-term availability are two of the most important tasks of a national research data infrastructure. For the area of long-term availability of research data, the RfII recommends tackling the problem of (solely) project-based financing of potentially permanent data collections very swiftly. The resources needed for long-term availability should be provided through a financing plan coordinated over the long term that allows organisations to proceed step-by-step but that also takes the national importance of the task into account. The tasks associated with long-term archiving must be performed and coordinated in the form of ongoing information infrastructure tasks using the (possibly networked) facilities available in the framework of the NFDI (cf. 4.4).

Except for a relatively small number of – generally discipline-specific – data centres, there are no established roles and responsibilities in Germany, which is why the specific authorities and responsibilities need to be clarified. In addition, the problem of deciding which tasks need to be performed for the specific disciplines and which need to be performed across all disciplines needs to be solved. Facilities offering long-term archiving services for research data also need to describe the scope of these services (*service levels*) and the quality requirements placed on the data to be archived with sufficient transparency. A closely coordinated process in the NFDI that performs these tasks should identify and utilise synergies between the sciences and the information infrastructure facilities when establishing services. Thus far, it has been impossible to overcome barriers of authority in the area of ensuring long-term availability, which is why the creation of clearly defined and communicated roles and responsibilities based on proven expertise is particularly urgent.

Facilities using the services should share in the cost

4.3.5 If data is archived for the long term outside of the facility that produced the data, i. e. long-term archiving is achieved using a service, then suitable cost and business models are required. In order to sustainably establish long-term availability, the RfII considers institutional efforts as well as support from the German federal and state governments essential, although it also endorses requiring the facilities using the services to pay an appropriate share of the costs.

Payment models could be designed with separate plans for different storage periods and user groups. The EU state aid laws must be followed provided that the services are not only limited to science, but are also meant to be used by commercial enterprises.

4.4 NEW PLAYERS/RESPONSIBILITIES TO BE ESTABLISHED

To establish a network-like National Research Data Infrastructure (NFDI), basic processes of understanding and collaborative structures must be organised. Even though the NFDI is designed to bundle existing players in an integrated manner, new roles and responsibilities must be established.

Step-by-step formation of the network

Joint action of the German federal and state governments

4.4.1 The NFDI should be built up as a network in a needs-based and step-by-step process. An initial construct (*association of founders*) in the sense of a core group is required. Furthermore, commitments must be made right from the start, and it must be possible to answer resource questions so that common understandings can lead to operative implementations. The German federal and state governments, in their authority as funders of research facilities and universities, should mandate cooperation and good research data management practices in a suitable manner at the facilities they fund. The RfII recommends that the German federal and state governments use a coordinated procedure based on a suitable instrument, e. g. a roadmap for the NFDI. Suitable stakeholders should create a strategy paper for this purpose in good time. The field of long-term archiving requires special effort and a high degree of institutional continuity in order to finally realise economies of scale and synergy effects (cf. 4.3.4). The RfII recommends bundling resources early in a network of centres of excellence and interdisciplinary infrastructure centres within the NFDI that meet the specific requirements for long-term availability. The German federal and state governments should cooperate during the foundation of centres of excellence for long-term archiving and coordinate the available capacities.

Creation of network structures to contribute to the NFDI

4.4.2 To contribute to the NFDI, the sponsors of scientific (or of scientifically relevant government) facilities have various options and prerequisites they can capitalise on for developing a basic supply of research data repositories together with data producers, facility operators, and data users (cf. 4.2.3). It is recommended that decision-makers at the level of university rectors and research/infrastructure facility directors join network structures that can be integrated into the NFDI. This recommendation also applies to the development of cross-institutional data infrastructures in scientific communities and research networks.

Reaching an understanding on evaluation criteria

4.4.3 The evaluation and certification processes for information infrastructures have not yet been established in the required breadth in the scientific system (see sections 2.3 and 2.5 as well as 4.1.1). The RfII recommends scientific organisations, possibly including the association of NFDI founders, to first agree on which universal criteria and certifications could contribute to networking at the national level and could be integrated into their corresponding processes. Similarly, the RfII suggests the establishment of a cross-discipline pool of experts composed of national and international consultants (cf. 4.1).

Wide participation

4.4.4 Broad discussion is necessary to coordinate the existing structures and needs for a national research data infrastructure. The private sector should participate in the development of a national research data infrastructure in a suitable manner since commercially organised information infrastructures play an important role in the scientific system. The same applies in particular to the cultural sector, to the area of the participation of non-professionals in research processes (*citizen science*), and to the data producers and infrastructure providers in government agencies and public authorities. Researchers, in their dual role as producers and reusers of data, should coordinate their interests with the infrastructures. Conversely, infrastructures as service facilities should try to coordinate with researchers. The RfII recommends creating a forum for both purposes. Furthermore, it also recommends establishing suitable forms of participation during the realisation of the NFDI, e. g. advisory boards or *communities of practice*.

Networking international activities

4.4.5 The need for *internationalisation* (cf. 4.8) should serve as impetus for the creation of networks of representatives from the German scientific system who are active members of international committees. Research funders, or even the RfII, could establish a suitable forum for this purpose.

4.5 NEW OCCUPATIONS, DEGREE PROGRAMMES, TRAINING PROGRAMMES

The stronger processes at universities or research facilities in general are affected by digitisation, the more urgent the need to systematically bundle expertise becomes. Data services, data management, and the design of information infrastructure services are scientifically complex challenges of a new kind. At the same time, education and research are increasingly calling for professionalism and international connectivity in such new fields of expertise, which sometimes require skills more oriented towards scientific methods and other times more towards infrastructures or service and management, as prerequisites. The present situation does not yet reflect this systematic approach to the new challenges, which affect the areas of *education*, *research*, and *infrastructure services* equally. Instead, there is a trend towards "flowery" professional titles[73] without any connection to the training or degree programmes responsible for these areas.

The RfII welcomes the current processes of communication at higher education institutions regarding suitable new degree programmes and underscores the necessity not only for a rapid, but also differentiated transformation. This transformation must take the diversity of subjects and research methods into account as well as the wide range of requirements arising between the more technical and organisational education and training profiles and the profiles with more scientific content. The numerous recommendations regarding general tasks as well as the challenges posed by specific services that have been produced up to now in the scientific system for the development of an information infrastructure rightly demand that education providers (higher education institutions, professional associations, etc.) offer specific programmes and demonstrate their willingness to implement them.

Teaching information literacy in all subjects

4.5.1 The development of information infrastructures must be augmented by increasing information literacy at all levels. Information infrastructures do not only consist of IT structures and software solutions. It is just as important, if not more so, to provide services and expertise in scientific methods. The term "information literacy" must be defined broadly because it merges various technical, communicative, social, and organisational subskills with discipline-specific skills.[74] The RfII recommends integrating modules for teaching information literacy and data management skills throughout the entire spectrum of degree programmes currently offered (i. e. from engineering, medicine, and the legal sciences to the humanities and social sciences) to train future generations of highly capable researchers. Additionally, further training and education paths should be opened up – ideally with international coordination – in science and

73 Hanraths (2015) – Hacker und Missionare.

74 Cf. HRK (2012) – Hochschule im digitalen Zeitalter, p. 6.

administration. The training and education programmes must be designed with the conflict between the methodologically-based disciplines and the topic-based disciplines in mind.[75]

Setting up specialised degree programmes

4.5.2 In addition, full degree programmes should be introduced that lead to new occupations such as that of a (digital) *documentalist*, *data librarian*, *data archivist*, or *data scientist*, for example, with specialisations in the corresponding subjects. In terms of the qualification profiles of these occupations, the German Rectors' Conference (HRK) has stated that they all "imply a high degree of ability to integrate and operate at the interface between research and the infrastructure facilities, as well as between the various infrastructure facilities themselves. This means that the conventional separation, for example between the work of library and computer centre staff in managing data, now appears obsolete".[76] Due to the foreseeably high need (cf. 2.3), new degree programmes should be developed and offered at a large number of higher education institutions by the existing library, archive, and information science centres in close cooperation with the computer sciences and other relevant scientific disciplines on site. In addition, the RfII recommends actively recruiting new students for the new occupations and education programmes.

Obtaining qualified personnel through dual vocational training programmes

4.5.3 The tasks and services of research data management also include standardised and recurring activities that do not necessarily need to be performed by academically trained personnel. In the framework of the dual vocational training system in Germany, new training programmes closely related to research data (e. g. to become a specialist in market and social research) have been created in recent years. Such vocational education paths must be developed further and promoted. Utilising vocationally trained personnel opens up new opportunities for further and simultaneously efficient professionalisation of the fields closely related to research data in infrastructure facilities.

Competition for minds

4.5.4 When designing new processes of teaching and learning in the context of an increasingly digital science, particular attention should be paid to ensuring more women work in these promising new occupations and follow these attractive new career paths. Furthermore, qualification options for older professionals as well as immigrants should be offered in order to cover the high demand for specialists. These new academic and non-academic job profiles will be in high demand inside and outside higher education institutions: academically educated data specialists are needed wherever large or complex data volumes need to be managed and integrated into complex organisational environments. For this reason, challenging data-related tasks in fields where research is not conducted (in the health sector, in public administrations, in memory in-

75 The German Council of Science and Humanities (WR) differentiates between "simulation sciences" and "simulation-based sciences" in a position paper. Cf. WR (2014) – Positionspapier Simulation, p. 11. A similar distinction may also make sense for the data sciences.

76 HRK (2012) – Hochschule im digitalen Zeitalter, p. 16.

stitutions, in the cultural sector, etc.) should also be taken into account when shaping the contours of these fields of expertise. On the other hand, it can be assumed that there will be a 'competition for minds', which then means it is important to keep qualified personnel in research. In such a competitive environment, this can only be achieved through an attractive design of new career paths.

4.6 NEW "DATA CULTURE"

The digitality of research data poses a wide variety of new challenges in quality assurance. In addition to the 'inherent' quality and integrity of the data, the quality of the data management system and metadata, adherence to internationally accepted standards, and the quality of the infrastructure itself also play an important role.[77]

Ensuring sufficient data quality over the entire data life cycle requires a joint effort. Each of the producers, processors, and various reusers of research data must assume responsibility for ensuring the quality of the data. This requires a new culture of open data and data sharing[78], but also agreements on standards. Establishing a common discourse on quality with all players is a goal that will take time and a large number of intermediate steps to achieve. In some areas of the scientific system, collaborative structures are already being built for this purpose that channel the negotiation processes with the intention of implementing common (quality) standards and tools in the medium term.[79]

Discipline-specific standards and reviews for data

4.6.1 The RfII recommends that professional associations, scientific societies, the review boards of the German Research Foundation, and recognised infrastructure centres pay particularly close attention to adequate, but also sufficiently dynamic, scientific standards and quality criteria. Opportunities for cross-discipline cooperation as well as collaboration with other disciplines and forms of research must be considered in the framework of data platforms and virtual research environments.

Describing data quality using standardised metadata designed for interoperability wherever possible plays a decisive role in this regard. Researchers should first ensure the quality of their research data by producing good documentation and reports. In particular, the level of uncertainty in the data must be doc-

---

77 Cf. the detailed explanation of the term "data quality" in the glossary.

78 Cf. the explanation of terms in the glossary.

79 In some cases, interest groups form in the context of building distributed infrastructures, which was the case in the ELIXIR initiative for the life sciences cluster, for example (cf. 2.4, FN 43), while in other cases infrastructure providers set standards, as was the case in the 30 or so accredited research data centres of the German Data Forum (RatSWD) or the international Data Seal of Approval community (cf. 2.5, FN 54).

umented. Corresponding commitments in science form the basis for offering quality-assured research data in the repositories.

In a further step, suitable *review* procedures should be conceived and promoted in sufficiently networked communities. Ideally, combined solutions that take generic as well as discipline-specific quality criteria into account will be designed and developed by the communities. Contact with (potential) data users and the specific professional networks of each discipline should be established early in this process (cf. 4.2). Interoperability is also an important goal of quality standards, which normally arise from the *bottom up*. In the medium term, binding mechanisms resembling the *peer review* process used to assure quality in scientific journals should be established for research data publications.

Data management plans as planning tools

4.6.2 The RfII recommends requiring the submission of data management plans (DMPs) as instruments for planning and vigorously promoting their widespread use in the research process – not only when research is funded externally, but as an integral part of good scientific practice. Data management plans are already commonly used by data producers in science, business, and administrations as a tool for quality management. The German federal and state governments as well as the funding organisations can set standards for DMPs for research projects funded by them. Furthermore, universities and research facilities should firmly establish data management institutionally through internal policies.[80] While doing so, the implementation of these policies and their effect on various subject areas and different forms of research should be tracked in accompanying research projects to counteract bureaucratisation, which can have a negative impact on science.

External quality assurance

4.6.3 The RfII recommends adding mechanisms for regular external quality assurance (e. g. monitoring and evaluation) to the NFDI, thus promoting transparency for the scientific users and funders alike. The process quality of the repositories and services is of particular importance since they must accompany the entire value chain and be tightly integrated into the development and application of regulations, standards, and policies. For producers and scientific re-users as well as for parties from the realms of business and politics interested in data, information infrastructures as "quality service providers" could become much sought-after points of contact.

The quality of repositories and other services is also measured based on their ability to process data according to its relevance. It is now possible to classify data itself, its potential forms of reuse, and its relevance. Parameters for measuring the relevance include, for example, the scientific benefit, prospects for commercial use, the socio-political need for the data, or its long-term cultural significance. These parameters are defined primarily based on estimates

[80] Cf. also the practical recommendations of the HRK (2015) – Handlungsoptionen für Forschungsdatenmanagement; and LERU (2013) – Roadmap for Research Data.

provided by the corresponding user groups (cf. also 4.3.3). A rough guide to high-quality data handling, e. g. the transparent specification of data curation levels and data curation expenses, would be a great help to numerous information infrastructures and should be available soon.

Existing initiatives that develop quality requirements for repositories as well as catalogues of criteria for archiving digital objects for science[81] should be integrated into the development of a framework for quality management. Furthermore, *Quality of Service* (QoS) criteria regarding aspects such as performance, availability, conformity, etc., must be defined for the quality of individual services. These criteria should be defined based on the current standards of comparable information infrastructures, e. g. in e-government. In all quality measures, it must be ensured that suitable monitoring procedures are established to support the evaluations in the system.

4.7 INCENTIVES FOR USE AND ACCEPTANCE WITHIN SCIENCE

The behaviour of researchers is a decisive factor in the success of research data management initiatives. Data producers often foreground the direct use and sharing of their own or other researcher's data within the narrow confines of their own community. By contrast, there is often less awareness of the significance of professional management, accessibility, and long-term availability. The rules of good scientific practice and guidelines for the creation of data management plans attempt to bring about a cultural transformation in this context. Nevertheless, there remains a series of obstacles to overcome before individual researchers will accept offers to manage their research data. In addition to the points already discussed in 4.6, it is also necessary to establish a "research data culture" based on the acceptability and attractiveness of a digital framework.

On-site support for research data management

4.7.1 Universities and research facilities should support their researchers, offer services for research data management, and work towards creating user-friendly structures. It is precisely in this phase of transformation that researchers as well as the organisations funding research are forced into the role of consultants and "brokers".

User-oriented design of services

4.7.2 Services established using large amounts of resources are not always used as intensely as they should be. A common reason for this is a lack of sufficient adaptability between the tools offered and the often very specific or simply imprecisely determined requirements of the researchers with respect to the usage costs and user-friendliness. For this reason, the requirements of us-

81 Cf. the examples mentioned in 2.5: accreditation process of the German Data Forum (RatSWD), nestor Seal, and the Data Seal of Approval.

ers should be considered at an early stage in the development and evaluation of services and tools for research data management. Whether or not the infrastructure fulfils these requirements should be checked *a posteriori*.

Legal advice for researchers

4.7.3 There are typically significantly more legal uncertainties associated with digital data traffic than with conventional data exchange methods. The rules of good scientific practice and data management policies put more pressure on individual researchers – and especially on the leaders of large research projects – through the use of key words and phrases like "data responsibility" or "data quality". The negative image of a legal "grey area" here can override trust in the otherwise generally conceded and protected freedom in science to manage the data material essential to research. The RfII recommends ensuring that information and clarification as well as legal advice are available in all organisations, preferably on site. This aspect of the digital transformation, which is highly relevant to the behaviour of all parties concerned, also needs to be tackled at the level of the scientific system as a whole. Suitable communication platforms to promote exchange between research facilities, research funders, science-oriented foundations, and the German Rectors' Conference should be created soon.

Further development of incentive systems

4.7.4 Based on similar suggestions, the RfII recommends ensuring the citability of data[82] and the establishment of data usage and data citation indexes provided that the corresponding fields have suitable publishing cultures available for this purpose. These indexes should be designed so that they act as an incentive to provide and use research data, and therefore to manage it. Linking a set of data or a piece of software to its producer can increase this researcher's reputation and act as a positive incentive to motivate the researcher's performance. On the other hand, the intensity with which data sets are used can serve as a relevance criterion for the development of the infrastructure offering the data. It can also give research communities and funders ideas for development. The RfII recommends discussing informative indicators and the possible effects of their application at the level of scientific communities, e. g. through the review boards of the German Research Foundation. Impetus for the discussion of how to further develop incentive systems and the already critically reflected practice in the area of bibliometrics[83] is desirable.

Recognising digitality in research as a field of expertise

4.7.5 The RfII recommends that researchers view the organisation of research processes under the conditions imposed by digitality as a field of expertise that will become very important in the medium and long term. Beyond training and education (see 4.5), both young scientists and established researchers should

[82] For example the Data Citation Synthesis Group (2014) – Joint Declaration of Data Citation (website) (FORCE 11) or the DataCite Metadata Working Group (2015) – Metadata Schema. Cf. also the discussion of enabling structures in section 2.5.

[83] The effects of citation indexes are becoming increasingly controversial; cf. the critical document recently created by Doove (2016) – Amsterdam Call for Action (website).

not view information literacy just as an additional qualification but should instead consider it a key aspect of the methodologies and discourses on methods in their own area of expertise. Suitable measures for promoting understanding within science in this matter could include the establishment of virtual discussion forums or journals oriented towards methods.

4.8 INTERNATIONAL RESEARCH AND INFORMATION INFRASTRUCTURE POLICIES

Science thrives on the exchange of data at the international level and can only remain competitive internationally and contribute to solving global social challenges through this exchange. National concepts for managing research data must therefore ensure international interaction and also proactively help to shape this interaction. Germany, as a strong member state of the EU, is already actively participating in the formation of the European Research Area (ERA). Even this process is characterised by a lack of synchronisation between the various developments in the individual member states, which also applies to the management of research data.[84] This applies to an even greater extent to coordination at the global level, and Germany is also active at this level through a wide variety of forums and organisations.

The development of an information infrastructure system for research data in Germany is subject to systemic conditions that are different from those of other countries. Such conditions include the wide distribution of responsibilities for scientific policy and the variety of financing routes. Top-down requirements, for example for the management of data, that need to be fulfilled to receive public financing have not yet been realised for the most part. Nevertheless, a national strategy should also be strengthened using elements that have proven themselves in other leading nations in science. There must be an examination and possible testing of which of these elements can be implemented under the conditions in Germany. In this context, the traditional strengths of the German science system – i. e. decentralisation and productive diversity while simultaneously being far removed from politics and enjoying a high level of self-organisation – should not only be maintained and cultivated, but should by all means be incorporated into international concepts. In addition, research data management must be considered an organisational task that is actively pursued and for which the German stakeholders assume responsibility in international strategy discussions.

Learning by comparing systems

4.8.1 To enable the derivation of politically assessable proposals for decisions in light of the very diverse approaches taken internationally to shape the digital

[84] Cf. section 2.4 "Germany in international comparison" as well as ERAC (2016) – Open Research Data.

transformation in science, the RfII suggests following development trends in other European countries as well as worldwide in an intercultural system comparison. Simple comparisons between approaches to solutions that ignore the specific initial situation of each country or nation, though, can be misleading. The RfII therefore deems it important to systematically gather detailed information on the different national research data strategies and compare them to the intended developments in Germany and in the European context. The RfII believes particular attention needs to be paid to the methods used by different countries for the management of so-called "disruptive" developments, which are considered drivers of innovation in the transformation taking place. To keep pace in an environment like this, Germany must promote methodical approaches and introduce them into a national scientific discourse. The solutions based on these approaches must be compatible with international ones and therefore, mindful of international heterogeneity, need to be sufficiently flexible. In light of the dynamic developments in this field, the RfII will continue to monitor international developments, compare them to national objectives, and provide strategic assessments within the scope of its appointed tasks.

4.8.2 The RfII sees a need for new forms of participation management to ensure German participation in science and technology policy-making at the international level. International developments are being shaped more rapidly and more profoundly by new and changing standards. These standards are formulated from the *bottom up* and specifically for each discipline using the networked world of information, but are also occasionally implemented from the *top down* by interest groups through formal standards (e. g. ISO standards). Both approaches are successful strategies in the technology sector for asserting one's own interests and then benefiting from this when successful. Over the last two decades, corresponding momentum has been gaining. The RfII considers it necessary to encourage representatives of German science to increase their participation in the corresponding processes and provide them with support in this regard. Only through more intense, active participation in such international processes of development will it be possible for researchers from Germany to bring in and advance the approaches and proposed solutions developed in Germany in international coordination processes. Furthermore, people holding key positions in international committees should collaborate closely with national committees and prepare their potential successors for the tasks in international committees through communicative integration. Acquiring people to perform such tasks must be a common objective for all stakeholders involved in organising research (cf. also 4.4.5). Scientific organisations could contribute to ensuring that the work done in international committees receives more appreciation inside the community than it has up to now.

Practical international education and training

4.8.3 In Germany, there is much discussion of the disadvantages and risks associated with the digital transformation in science, research, and development.

By contrast, the awareness that there is nevertheless an urgent need to understand new technological capabilities in their international contexts as well in order to proactively shape the digital transformation with all its opportunities and risks is less prevalent – in research, at universities and in the education sector, and even at the management level. In addition to internationally coordinated training and education paths at the national level (cf. 4.5), the RfII therefore encourages, in the sense of enabling transformation, the development of practical learning opportunities in the framework of international initiatives and information infrastructures. Furthermore, it promotes support for practical research on and with existing data infrastructures through fellowships and project funding.

International cooperation on equal footing

4.8.4 Internationality requires organisational interfaces, not least because of the internal complexity inherent in Germany. A prerequisite for successful internationalisation is the ability of national information infrastructures to act as strong cooperation partners. In particular, the RfII sees a duty to support shared structures in the ERA and break down barriers. Active international collaboration is facilitated by clearly formulated, generally accepted, and implemented concepts for governance, technical and organisational connectivity, sustainable financing options, quality criteria, and user orientation. The RfII also sees a need for additional, rigorous incentives to consolidate national self-organisation in the sense of a lobby and as a "voice" advocating information infrastructures – with the goal of achieving a governance system capable of taking action so that it can cooperate as an equal with its international counterparts. Good examples of this include the very active players involved in the European ESFRI projects and in the establishment of HPC resources (e. g. the Gauss Centre for Supercomputing, Gauss Alliance). Research policy should support similar developments through funding measures as well as through expert evaluations.

4.9 USE AND EXPLOITATION

Data is often referred to as the raw material of the information age. The legal status of research data is often unclear, and it is generally not protected against appropriation or even illegal use. This is especially true in cases where the researchers transferred the data themselves, for example to online networks with a commercial interest or to a commercial cloud (cf. also 2.2). The "openness" of research data therefore not only creates new ways to share and use data for scientific purposes (open data, open science[85]), but also the ability to economically exploit this data.

85 Cf. the explanation of terms in the glossary.

Commercial access to research data can be harmless, and even desired, and access can be controlled transparently and take scientific interests (as well as data protection, copyright, and data security issues) into account. However, data can also be accessed in networks under unclear legal conditions or even without being noticed. In extreme cases, subsequent privatisation (meaning the appropriation of "open" data by economic players in the international legal space) can prevent further use of the data in an open science framework or its exploitation on a clearly defined legal basis (e. g. through patents).

In the interest of the German economy, science and politics are called on to meet the challenge of creating a reliable and transparent regulatory framework that protects the value of knowledge and promotes the generation of new knowledge because the data economy is globalised, and value is created over and over again in countries with different data usage regulations.

Data responsibility

4.9.1 Science needs to actively shape its interface with the economic sector by handling research data internally before it reaches said interface. The RfII believes that scientific organisations and research facilities are required to establish binding rules of conduct to try and shape the unregulated flow of data when using the services of private software, communication, and storage providers. All stakeholders should work towards affecting a cultural transformation that links the subject of openness for scientific use with the idea of responsibility for data. Policies, commitments, and licenses that speak out against appropriation (for example Creative Commons, although it is not legally binding) are possible tools. A discourse that raises awareness of the dual nature of the buzzword "openness" in all stakeholders in research – even including students – is also required.

Creating legal certainty

4.9.2 The ability to publish research data, i. e. to provide it in digital form (separately or in essays), is one of the greatest and most welcome opportunities of the digital age. However, suitable minimum standards for designing data publications that ensure the accessibility and reusability of the published data over the short and medium term are required (cf. 4.6.1). The creation of legal standards for the growing area of social networks is equally important. The RfII endorses the clarification of legal issues as well as the establishment of dedicated services in science that are subject to German law (for example in the area of clouds, cf. 2.5 and 4.2.6).

Research-friendly licensing models for data and text

4.9.3 If research data is to be published, then licensing models designed to be research-friendly are needed that do not prevent the reuse and, in the case of text, the application of text mining methods as well. Regarding the design of legal policies for a future-proof German copyright law, the RfII refers to the declaration of the Alliance of Science Organisations published on 1 December 2014.[86]

[86] Cf. Allianz der Wissenschaftsorganisationen (2014) – Zum Urheberrecht.

4.10 MORE KNOWLEDGE ABOUT RESEARCH DATA

To avoid making misguided decisions on the path to a balanced *ecosystem* for research data, knowledge about the changing realities in the scientific system is needed. Current scenarios for research data management are based on examples from pioneering scientific communities that very actively conduct research based on data. Furthermore, studies occasionally provide information on the attitudes of researchers towards paradigms like open science or data sharing[87] and what they would like to see in the supporting IT environment.[88] Overall, though, there is a lack of data, facts, and models for describing the highly fragmented and varying landscape in the area of research data management in Germany. There is hardly any information available on the type and number of infrastructure initiatives, and in particular on their benefits/successes, operating costs, investments made, or any other basic figures.[89] It is even more difficult to get information internationally. Typologies for comparisons, planning, and forecasting are completely lacking according to the knowledge of the RfII, except for a few attempts to sketch a rough outline.[90]

Database for research and infrastructure policy

4.10.1 The RfII recommends establishing a national database concerning research and information infrastructure policy. In connection with the Core Data Set on Research Activities[91] currently in development, other digitisation indicators for science and society should also be considered. Systems created to acquire data on the infrastructure system should be compatible with international solutions. Beyond the national level, the RfII sees a need for the systematic, preferably annual, monitoring of developments in neighbouring European countries and at the level of globalised research and technology policy as well as in the data economy (cf. 4.8.1).

Accompanying research on digitisation in science

4.10.2 Particularly relevant is an understanding of researchers' behaviour across the entire spectrum of methods and forms of research, surveys that evaluate organisational models in terms of legal policies and economics, as well as sound risk assessments in the area of data protection and data security. The RfII urgently suggests initiating targeted, social-scientific accompanying research for this purpose.

87 Cf. also explanation of terms in the glossary.

88 For example the KE – Knowledge Exchange (2014) – Sowing the seed, the Austrian survey conducted by Bauer et al. (2015) – Forschende und ihre Daten, or the paper by Fecher et al. (2015) – A Reputation Economy.

89 Regarding the lack of an institutional "map", cf. Meyer-Doerpinghaus/Neuroth (2015) – Stärkung Informationskompetenz, p. 81. The re3data.org directory nevertheless lists almost 250 data repositories with German participation, the RIsources database of the German Research Foundation lists around 200 information infrastructures (as of April 2016).

90 Cf. e. g. the data pyramid of the EU High Level Expert Group on Scientific Data in: European Commission (2010) – Riding the wave, p. 18.

91 For information on documenting research infrastructures in research reports, cf. also WR (2016) – Kerndatensatz Forschung, p. 43 and other pages.

4.11 DATA PROTECTION

Data protection is a key legal policy topic of the digital age, but there are opposing interests in terms of protecting research data. On the one hand, science needs access to protected data. Without the ability to work with personal, partially individualised knowledge, research and scientific progress in areas of health, education, work, environmental protection, and services of public interest become impossible. On the other hand, science also needs to ensure data protection. In German law, this is achieved on the basis of the informational self-determination and personal rights in Art. 2 in connection with Art. 1 of the German Constitution. Internationally, though, data protection in the sense of *privacy* is formulated in somewhat different legal terms.

German data protection regulations are strict by international comparison and take the conflict between the need for protection and the justifiable interests of researchers according to Art. 5 of the German Constitution into account. Nevertheless, new challenges for data protection arise due to digital technology that enables low-threshold access to protected data, the enormous growth of personal data or data that can be traced back to individuals, the resulting impossibility of anonymisation[92] in some fields, and the global character of data exchanges. Likewise, problems relating to data *security*, meaning technical as well as forensic/criminological questions, must be more closely integrated into the topic of data *protection* (cf. also 4.12).

For the moment, the RfII has only pointed out the importance of this topic in this position paper. At the European level, the EU General Data Protection Regulation[93] has created a new legal framework for the member states that also makes it necessary to adapt current German data protection legislation accordingly within a transition period of two years. The regulation contains general provisions as well as provisions regarding the processing of data for scientific purposes. Implementation at the national level must be utilised as an opportunity to take a look at the growing requirements and possibilities as well as at the opportunities and risks that will arise in the future due to technological and methodological developments in the area of data analysis. The RfII will be presenting corresponding recommendations in the form of a "Data protection compass for research" (working title) in the near future.

4.12 DATA SECURITY

Digital research data is vulnerable. For this reason, data security not only forms the basis for proper scientific work, but is also fundamental to the quality as

[92] In the biomedical sector, for example, it is possible to identify the direct biological relatives of a person whose tissue sample has been analysed.

[93] Council of the European Union (2016) – Standpunkt zur Datenschutz-Grundverordnung.

well as the scientific, economical, and societal value of research data. It is necessary to differentiate between the terms *safety* (operational safety) and *security* (protection against attacks) in this context. If networked systems for retaining data are not designed with adequate technical robustness or can be attacked (i. e. using malware, hacking, etc.), science may encounter new kinds of hazards.

From the perspective of an infrastructure, research data security consists of two decisive aspects: firstly *secure storage* (also over the long term) of research data in sufficiently high-performance infrastructures that are also equipped with sufficient redundancy. The second aspect regards special technical and organisational *cyber security* measures for preventing unauthorised access, espionage, and sabotage.

Prioritising data security in the operation of infrastructures

4.12.1 The RfII recommends that the responsible stakeholders focus much more closely on technical and organisational measures for data security during the development of information infrastructures than has previously been the case. In addition to the NFDI, such stakeholders will include computer centres and information infrastructures in general, those organisations that bear their costs, and research funders. During the evaluation of infrastructure projects, especially during an evaluation or accreditation for long-term financing of services and solutions (cf. 4.1), data security and data protection must be treated with equal importance.

Raising researchers' awareness of data security

4.12.2 Like data protection, data security depends to a great degree on the behaviour of those involved in research and education. In this respect, all scientific facilities are called upon to consider data security aspects as well when discussing data protection.

4.13 FINANCING AND SUSTAINABILITY

Creating framing conditions for the sustainable management of digital research data in Germany entails personnel, communicative, and financial expenditures. The RfII cannot reliably estimate or provide any robust figures for the total expenditures required to accomplish the recommended restructuring. The amount required is very likely substantial, though – particularly considering the complexity of the tasks and the number of stakeholders as well as the inevitably long time for the required reorganisation and quality assurance processes.

However, such efforts are directed towards overcoming a critical threshold in the development of the overall system of science. It is therefore a matter of going through a transition phase in which joint action is required for the special, and in many respects also unique (and therefore crucial), establishment and transition process. For the benefit of new levels of research and operational productivity as well as international competitiveness, science and politics must reshape important core processes of research, which in the meantime has become characterised by digitality. The objective is to leverage potential, increase efficiency, and avoid the risk of undesired developments (i. e. a niche focus, project-based and stand-alone solutions, personnel problems, lacking or heterogeneous standards, the loss of usage rights, data losses, etc.). This generates investment and transition costs that exceed the normal level of resources allocated to research, education, and infrastructure development.

The goal of the recommendations provided by the RfII, and specifically the recommendation to establish an NFDI, is to open up new perspectives in the German scientific system for further quantitative and qualitative development at the highest level. Without additional resources, it will be impossible to realise the urgently needed transformation to achieve this goal. Additionally, and taking into account a sustainable data management that is adequate for science, it is necessary to develop sufficiently low-threshold forms of distributing the costs of data-related services that control the use of the scarce resources available sensibly.

Financial planning using a roadmap

4.13.1. The RfII believes that special efforts are necessary from the scientific community and the providers of grants in order to realistically finance an intelligently guided and thoroughly effective process of transition in German research data management. The RfII emphasises the importance of ensuring this effort is a joint effort. For this reason, a financing plan (national roadmap) is recommended that includes vigorous special measures by the stakeholders as well as strong incentives during a transition phase lasting several years. Otherwise, the RfII foresees the uncontrolled 'proliferation' of the status quo into a costly situation characterised by undesired developments and inefficiency that would weaken the capability of science in Germany.

Costs for the establishment and the transition phase that exceed the normal scope of investment in the scientific system arise particularly in the following five fields from the RfII's perspective:

- Establishment of the NFDI, *top-down* policy development (project/possibly including institutionalisation of the core tasks)
- Communication and coordination processes at the level of scientific communities and disciplines (incentives/possibly including projects)
- Reorganisation processes for existing services (incentives/possibly including long-term financing solutions)
- Training and education measures for handling data in a research setting (incentives/possibly including projects)
- Communication and monitoring as a cross-sector activity (projects/possibly including institutionalisation of the core tasks).

Contribution model for joint financing

4.13.2. Tailored contribution models for sustainable research data management in Germany should be found. The RfII recommends an experimental phase for this purpose in which incentives are provided and forms of allocating costs are developed and tested (whereby it is possible to differentiate accordingly, e. g. into communities, user groups, or similar categories).

In general, contributions help to stress the importance of the services and create incentives for resource-saving behaviour. However, the transfer of costs should not have a discouraging effect. In addition, the sustainable management of research data should not be subject to commercial considerations that focus on break-even prices, let alone market prices. This would destroy scientific competition, and in general destroy the specific value of the social system of science and the research achievements this system brings forth.

Monitoring the transition process

4.13.3. Monitoring measures and forms of moderation should accompany the process of transition (while accounting for effectiveness and efficiency aspects). They should examine questions regarding the effect of financing models for research data services on the overall system as well as accompany the medium-term *learning* process for designing a dynamic environment for the *digital* science of tomorrow. Learning processes can affect the reallocation of resources or even the shifting of tasks between facilities.

In the end, basic services also need be ensured for data-intensive research in the framework of dedicated scientific business models. In addition to acquiring data, the intensive use of existing data, and its reuse in research in particular, are also primary goals when developing sustainable terms of use. It is just as important to ensure that researchers are motivated so that as many as possible actively support high-quality research data management and include their data in an overall, quality-assured *research data ecosystem* subject to German law.

5 OUTLOOK AND PRIORITISATION OF THE RECOMMENDATIONS

After reviewing the current situation, the German Council for Scientific Information Infrastructures (RfII) has provided comprehensive recommendations. The aim of these recommendations is to point out which courses to take in order to guide a scientific system in Germany that is essentially oriented towards the availability of research data in digital form. Numerous aspects of research data management will subsequently need to be developed in more detail.

The RfII combines its suggestions with an urgent message to politicians and society that digitisation will bring about many changes, but that these changes will promote research in Germany and all over the world and can progressively develop for the benefit of every citizen. It must once again be emphasised that scientific policy-makers and the scientific community are equally responsible – each within their own areas – for ensuring joint action takes place. Shaping the transformation prospectively, supporting the international competitiveness of German scientific research in a strategic manner, and making efficient use of public funding must be the overarching concerns. Through effective collaboration, Germany can actively advance the digital transformation to ensure that science, industry, and society achieve the maximum benefit.

The expectations with which the RfII approaches decision-makers in science and politics can be summarised in five points. These bullet points name areas in which urgent action is needed. Action should not be taken in successively, but should and can be taken simultaneously in all areas. The RfII attaches equal importance to all of the following points:

- Adjusting funding mechanisms
- Ensuring efficiency and coordination based on a (distributed) national infrastructure
- Promoting a "research data culture"
- Setting up an overall monitoring and quality assurance system
- Developing human resources at all levels

The recommendations of the RfII are directed at the Joint Science Conference (HRK) of the German federal and state governments, but also at other stakeholders, namely those in science itself, who will be catalysts and shapers of the transformation ahead. The RfII has therefore decided to specify the priorities of the points it has recommended based on target groups.

Table 1 summarises high priority recommendations directed equally at the German federal and state governments as well as to scientific organisations. They point out the "major lines" that need to be developed together. *Table 2* contains a summary of additional recommendations that are primarily directed at scientific organisations, including their representatives and member organisations. They involve problems that need to be solved in the framework of scientific self-organisation. *Table 3* provides an overview of recommendations for researchers, their scientific societies, and management. They include tasks that need to be implemented, supported, and actively lived by the corresponding target group or that primarily require a solution to be found within the scientific community.

Table 1: Recommendations for the German federal and state governments and scientific organisations

No. in text	Recommendations with the highest priority
4.1.1, 4.1.2	Phase model for the development of information infrastructures – ensuring planning capability and minimum standards – organising an orderly transition into a suitable sponsoring organisation based on independent evaluations
4.2.1	Establishment of a National Research Data Infrastructure (NFDI) – bundling expertise and creating universal access to services for research data management
4.2.3	Collaborative organisation of services in network structures – establishment of universal infrastructure centres and centres of excellence
4.2.7, 4.4.1, 4.13.1	Joint special efforts by the German federal government, German states, and scientific organisations – using roadmaps as a tool for coordination and finance planning for the NFDI
4.3.4, 4.3.5, 4.4.1	Sustainability and long-term availability – clarification of long-term financing solutions and responsibilities in the area of long-term archiving
4.6.2	Use of data management plans as planning tools in the research process

No. in text	Recommendations with high priority
4.2.2	Treating technical features, support, and the development of methods as interconnected issues and planning accordingly
4.5.3	Professionalising fields closely related to research data by utilising vocationally trained personnel – promoting training programmes closely related to research data in the dual vocational training system in Germany
4.6.3	Adding an external quality assurance system to the NFDI
4.9.3	Development of research-friendly licensing models for research data (incl. text)
4.8.1, 4.10.1, 4.10.2, 4.13.3	Increasing supervisory knowledge through monitoring and accompanying research
4.13.2	Developing and testing forms of cost allocation to find tailored contribution models for the NFDI

Table 2: Additional recommendations for scientific organisations in Germany

No. in text	Recommendations with the highest priority
4.1.3	Development and establishment of a concept for the regular evaluation of research data infrastructures
4.3.2, 4.3.3	Standardisation of guidelines, data management plans, and data curation in long-term archiving
4.4.3 4.4.4	Criteria and certifications for the networking of information infrastructures; structures and needs in terms of a national research data infrastructure – achieving consensus at the national level
4.5.1, 4.5.2, 4.8.3	Creating training and education programmes as well as specialised major degree programmes; widespread teaching of information and data management skills at all levels; offering practical learning opportunities in the framework of international initiatives and information infrastructures
4.6.1, 4.6.2	Promoting and disseminating standards and quality criteria by representatives from the various disciplines and infrastructure providers – using data management plans as a tool for good scientific practice
4.7.1 4.7.2	Offering support and services for research data management and working towards fulfilling the needs of users
4.7.5	Taking into account the conditions digitality imposes on the organisation of research processes – in terms of competences needed – when selecting and recruiting new staff
4.8.4	Creating shared structures in the European Research Area and achieving cooperation on equal footing
4.9.1	Establishment of binding rules of conduct to shape the flow of data between science and business – the goal: establishing a culture which links openness and responsibility for data together

No. in text	Recommendations with high priority
4.2.4	Development of comprehensive services, standardised data descriptions, and universal accessibility for the NFDI
4.2.5 4.2.6	Differentiation of the facilities in the NFDI according to their function – as an infrastructure provider, create incentives to use the infrastructure by offering expertise in methods and impressive services
4.4.5, 4.8.2	Creation of a network of internationally active representatives from the German scientific system who are active members of international committees – ensuring German participation in science and technology policy-making
4.5.4	Keeping qualified personnel in research data management by creating attractive new career paths in research
4.7.4	Establishment of data usage and data citation indexes to the extent appropriate within a discipline or field
4.9.2	Establishment of dedicated data publication services in science and in social media that are also subject to German law

Table 3: recommendations for researchers, their scientific societies, and management in Germany

No. in text	Recommendations
4.2.3, 4.2.5, 4.4.2	Prioritise working in network structures that can be integrated into the NFDI; using existing services; considering establishing connections to larger infrastructure centres
4.3.1	Leading a scientific discourse on the distinction between storage for the duration of a project and storage designed for significantly longer archiving periods.
4.3.2, 4.3.3	Planning the management of the data and developing the tools, services, and workflows needed throughout the data life cycle in good time – assessments of the relevance by data archivists and communities
4.4.4	Coordinating the interests between data producers, reusers of data, and infrastructure centres; contributing to a common forum
4.6.2	Requiring the submission of data management plans as instruments for planning and vigorously promoting their widespread use in the research process – as an integral part of good scientific practice.
4.7.2	As a facility, ensuring that the needs of users are considered during the development and evaluation of services and tools for research data management and checking to ensure that these needs were subsequently fulfilled
4.7.3	Ensuring that information and counselling as well as legal advice regarding the legally compliant handling of data are available in all organisations
4.7.5	As a researcher: recognising the organisation of research processes under the conditions imposed by digitality as a field of expertise
4.9.1	Establishing binding rules of conduct to shape the flow of data between science and business – the goal: establishing a culture which links openness and responsibility for data
4.12.1	Ensuring a greater focus on technical organisational measures for data security during the development of information infrastructures
4.12.2	Broadening the discussion on data protection to include data security aspects in order to affect a corresponding change in the behaviour of those involved
4.13.1	Making special efforts locally to the extent possible in order to help establish an NFDI

BIBLIOGRAPHY

Allianz der Wissenschaftsorganisationen (2003): Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen, Berlin, online at: http://openaccess.mpg.de/68053/Berliner_Erklaerung_dt_Version_07-2006.pdf, last checked on: 25/04/2016.

Allianz der Wissenschaftsorganisationen (2014): Zum Urheberrecht. Entfristung von Paragraf 52a UrhG nur ein Zwischenschritt – Allgemeine Bildungs- und Wissenschaftsschranke muss kommen (Pressemitteilung), no place, online at: http://www.allianzinitiative.de/fileadmin/user_upload/www.allianzinitiative.de/pm_wissenschaftsschranke_2014.pdf, last checked on: 25/04/2016.

Allianz-Initiative Digitale Information – AG Forschungsdaten (2015): Positionspapier „Research data at your fingertips". Schwerpunktinitiative „Digitale Information" der Allianz der deutschen Wissenschaftsorganisationen (ed.), no place, DOI: 10.2312/allianzfd.001.

Bauer, Bruno et al. (2015): Forschende und ihre Daten. Ergebnisse einer österreichweiten Befragung. Report 2015. Version 1.2, Wien, DOI: 10.5281/zenodo.31935.

BMBF – Bundesministerium für Bildung und Forschung (2013): Roadmap für Forschungsinfrastrukturen. Pilotprojekt, Bonn, online at: http://www.bmbf.de/pub/roadmap_forschungsinfrastrukturen.pdf, last checked on: 25/04/2016.

BMBF – Bundesministerium für Bildung und Forschung (2014): Strategie der Bundesregierung zum Europäischen Forschungsraum (EFR). Leitlinien und nationale Roadmap, Bonn, online at: http://www.bmbf.de/pub/BMBF_Forschungsstrategie.pdf, last checked on: 25/04/2016.

BMBF – Bundesministerium für Bildung und Forschung (2014): Internationale Kooperation. Aktionsplan, Berlin, online at: https://www.bmbf.de/pub/Aktionsplan_Internationale_Kooperation.pdf, last checked on: 25/04/2016.

BMBF – Bundesministerium für Bildung und Forschung (2016): Der Nationale Roadmap-Prozess für Forschungsinfrastrukturen. Investitionen für die Forschung von morgen, Bonn, online at: https://www.bmbf.de/pub/Nationaler_Roadmap_Prozess_fuer_Forschungsinfrastrukturen.pdf, last checked on: 25/04/2016.

BMBF – Bundesministerium für Bildung und Forschung (2016): European Cloud Initiative. Eckpunkte (unpublished), Berlin.

Bundesregierung (2014): Digitale Agenda 2014-2017, online at: https://www.digitale-agenda.de/Content/DE/_Anlagen/2014/08/2014-08-20-digitale-agenda.pdf?__blob=publicationFile&v=6, last checked on: 25/04/2016.

Chan, Leslie et al. (2002): Budapest Open Access Initiative (website), Budapest, online at: http://www.budapestopenaccessinitiative.org/read, last checked on: 25/04/2016.

Council of the European Union (2016): Standpunkt des Rates in erster Lesung im Hinblick auf den Erlass der Verordnung des Europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung). 5419/16. Interinstitutionelles Dossier 2012/0011 (COD), Brussels, online at: http://data.consilium.europa.eu/doc/document/ST-5419-2016-REV-1/de/pdf, last checked on: 25/04/2016.

Dally, Ortwin/Fless, Friederike/Förtsch, Reinhard (2012): Altertumswissenschaften, in: Neuroth, Heike et al. (ed.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme, Boizenburg/Göttingen: Hülsbusch/Universitätsverlag Göttingen, pp. 161–178, online at: http://nestor.sub.uni-goettingen.de/bestandsaufnahme/kapitel/nestor_bestandsaufnahme_008.pdf, last checked on: 25/04/2016.

DARIAH-DE (2016): Memorandum zur nachhaltigen Bereitstellung digitaler Forschungsinfrastrukturen für die Geistes- und Kulturwissenschaften in Deutschland (website), Berlin, online at: http://dhd-blog.org/?p=6559, last checked on: 25/04/2016.

Data Citation Synthesis Group (2014): Joint Declaration of Data Citation Principles (website). Maryann Martone, San Diego (eds.), online at: https://www.force11.org/group/joint-declaration-data-citation-principles-final, last checked on: 25/04/2016.

DataCite Metadata Working Group (2015): DataCite Metadata Schema for the Publication and Citation of Research Data. Version 3.1, DOI: 10.5438/0010.

DFG – Deutsche Forschungsgemeinschaft (2013): Sicherung guter wissenschaftlicher Praxis. Denkschrift. Empfehlungen der Kommission zur Selbstkontrolle in der Wissenschaft, Bonn, online at: http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf, last checked on: 25/04/2016.

Doove, John (2016): Amsterdam Call for Action on Open Science (website), Amsterdam, online at: https://wiki.surfnet.nl/display/OSCFA/Amsterdam+Call+for+Action+on+Open+Science, last checked on: 25/04/2016.

EFI – Expertenkommission Forschung und Innovation (2016): Gutachten 2016. Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands, Berlin, online at: http://www.e-fi.de/fileadmin/Gutachten_2016/EFI_Gutachten_2016.pdf, last checked on: 25/04/2016.

ERAC – European Research Area Committee and Innovation Committee (2016): Opinion on Open Research Data, Brussels, online at: http://www.earto.eu/fileadmin/content/Website/ERAC_Opinion_on_Open_Research_Data.PDF, last checked on: 25/04/2016.

European Commission (2010): Legal framework for a European Research Infrastructure Consortium – ERIC. Practical Guidelines, Brussels, online at: http://ec.europa.eu/research/infrastructures/pdf/eric_en.pdf, last checked on: 25/04/2016.

European Commission (2013): Assessing the projects on the ESFRI roadmap. A high level expert group report, Brussels, online at: https://ec.europa.eu/research/infrastructures/pdf/jd-final-aegreport-23sept13.pdf, last checked on: 25/04/2016.

European Commission (2015): European Research Area. Facts and Figures 2014. EUR 26803 EN, Brussels, DOI: 10.2777/72967.

European Commission – Directorate-General for Research and Innovation (2016): Guidelines on Data Management in Horizon 2020. Version 2.1, Brussels, online at: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, last checked on: 25/04/2016.

European Commission (2010): Riding the wave. How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission, no place, online at: http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=6204, last checked on: 25/04/2016.

European Commission (2016): European Cloud Initiative – Buildung a competitive data and knowledge economy in Europe. COM (2016) 178 final. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Brussels, online at: http://data.consilium.europa.eu/doc/document/ST-8099-2016-INIT/en/pdf, last checked on: 25/04/2016.

Fecher, Benedikt et al. (2015): A Reputation Economy. Results from an Empirical Survey on Academic Data Sharing (Working Paper Series, 246). RatSWD – German Data Forum (ed.), Berlin/Kiel, online at: http://www.ratswd.de/dl/RatSWD_WP_246.pdf, last checked on: 25/04/2016.

Field, Laurence et al. (2013): Realising the full potential of research data. Common challenges in data management, sharing and integration across scientific disciplines. Version 3, no place, online at: http://orca.cf.ac.uk/66034/1/ESFRI_Common_Challenges_v1.pdf, last checked on: 25/04/2016.

G8 (2013): Open Data Charter. Policy Paper, Lough Erne, online at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207772/Open_Data_Charter.pdf, last checked on: 25/04/2016.

GSF – Global Science Forum OECD (2014): International Distributed Research Infrastructures (IDRIS). Issues and Options. OECD – Organisation for economic co-operation and development (ed.), Paris, online at: http://www.oecd.org/sti/sci-tech/international-distributed-research-infrastructures.pdf, last checked on: 25/04/2016.

GWK – Gemeinsame Wissenschaftskonferenz: Ergebnisse der Sitzung der Gemeinsamen Wissenschaftskonferenz (GWK) am 28. Juni 2013. Pressemitteilung 11/2013, Berlin/Bonn, online at: http://www.gwk-bonn.de/fileadmin/Pressemitteilungen/pm2013-11.pdf, last checked on: 25/04/2016.

GWK – Gemeinsame Wissenschaftskonferenz (2013): Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen. Ergänzter Bericht des Ausschusses (GWK Drucksache 13.48, unpublished), Bonn.

Hanraths, Tobias (2015): Hacker und Missionare. Neue Berufsbilder, in: Süddeutsche Zeitung, online at: http://www.sueddeutsche.de/karriere/2.220/neue-berufsbilder-hacker-und-missionare-1.2733698, last checked on: 25/04/2016.

HRK – Hochschulrektorenkonferenz (2012): Hochschule im digitalen Zeitalter. Informationskompetenz neu begreifen – Prozesse anders steuern. Entschließung der 13. Mitgliederversammlung der HRK am 20. November 2012 in Göttingen, Bonn, online at: http://www.hrk.de/uploads/media/Entschliessung_Informationskompetenz_20112012_01.pdf, last checked on: 25/04/2016.

HRK – Hochschulrektorenkonferenz (2015): Wie Hochschulleitungen die Entwicklung des Forschungsdatenmanagements steuern können. Orientierungspfade, Handlungsoptionen, Szenarien. Empfehlungen der 19. Mitgliederversammlung, Kiel, online at: http://www.hrk.de/uploads/tx_szconvention/Empfehlung_Forschungsdatenmanagement__final_Stand_11.11.2015.pdf, last checked on: 25/04/2016.

IEAG – Independent Expert Advisory Group on a Data Revolution for Sustainable Development (2014): A world that counts. Mobilising the data revolution for sustainable development. United Nations Secretary General, no place, online at: http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf, last checked on: 25/04/2016.

KE – Knowledge Exchange (2014): Sowing the seed. Incentives and motivations for sharing research data, a researcher's perspective, Copenhagen, online at: http://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf, last checked on: 25/04/2016.

KII – Kommission Zukunft der Informationsinfrastruktur (2011): Gesamtkonzept für die Informationsinfrastruktur in Deutschland. Empfehlungen im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder. With the collaboration of Sabine Brünger-Weilandt, online at: http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Infrastruktur/KII_Gesamtkonzept.pdf, last checked on: 25/04/2016.

Klessmann, Jens et al. (2012): Open Government Data Deutschland. Eine Studie zu Open Government in Deutschland im Auftrag des Bundesministerium des Innern, Berlin, online at: http://www.bmi.bund.de/SharedDocs/Downloads/DE/Themen/OED_Verwaltung/ModerneVerwaltung/opengovernment.pdf?__blob=publicationFile, last checked on: 25/04/2016.

LERU – League of European Research Universities – Research Data Working Group (2013): LERU Roadmap for Research Data. Advice Paper No. 14, no place, online at: http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf, last checked on: 25/04/2016.

Meyer-Doerpinghaus, Ulrich/Neuroth, Heike (2015): Die Stärkung von Informationskompetenz im Kontext des Forschungsdatenmanagements: Eine Herausforderung für Hochschulen und Politik, in: ZfBB – Zeitschrift für Bibliothekswesen und Bibliographie 62, Nr. 2, pp. 80–84, DOI: 10.3196/186429501562237.

Nestor (2015): Digitale Langzeitarchivierung in einer Informationsinfrastruktur für Forschung und Bildung, no place, online at: http://files.dnb.de/nestor/berichte/nestor-Positionspapier.pdf, last checked on: 25/04/2016.

Neuroth, Heike et al. (ed.) (2012): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme, Boizenburg/Göttingen: Hülsbusch/Universitätsverlag Göttingen, online at: http://nbn-resolving.de/urn:nbn:de:0008-2012031401, last checked on: 25/04/2016.

OECD – Organisation for economic co-operation and development (2008): OECD Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information [C(2008)36]. OECD Ministerial Meeting on the future of Internet Economy in Seoul (Korea) 17-18 June 2008, Paris, online at: http://www.oecd.org/sti/44384673.pdf, last checked on: 25/04/2016.

OECD – Organisation for economic co-operation and development (2015): Making Open Science a Reality. Final Report, Paris, online at: https://www.innovationpolicyplatform.org/sites/default/files/DSTI-STP-TIP%282014%299-REV2_0_0_0_0_0.pdf, last checked on: 25/04/2016.

RfII – Rat für Informationsinfrastrukturen (2015): Auftakterklärung, Göttingen, DOI: 10.5281/zenodo.50662.

WR – Wissenschaftsrat (2012): Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020. Drs. 2359-12, Berlin, online at: http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf, last checked on: 25/04/2016.

WR – Wissenschaftsrat (2013): Bericht zur wissenschaftsgeleiteten Bewertung umfangreicher Forschungsinfrastrukturvorhaben für die Nationale Roadmap (Pilotphase). Drs. 2841-13, Berlin, online at: http://www.wissenschaftsrat.de/download/archiv/2841-13.pdf, last checked on: 25/04/2016.

WR – Wissenschaftsrat (2014): Bedeutung und Weiterentwicklung von Simulation in der Wissenschaft. Positionspapier. Drs. 4032-14, Dresden, online at: http://www.wissenschaftsrat.de/download/archiv/4032-14.pdf, last checked on: 25/04/2016.

WR – Wissenschaftsrat (2015): Empfehlungen zur Finanzierung des Nationalen Hoch- und Höchstleistungsrechnens in Deutschland. Drs. 4488-15, Stuttgart, online at: http://www.wissenschaftsrat.de/download/archiv/4488-15.pdf, last checked on: 25/04/2016.

WR – Wissenschaftsrat (2016): Empfehlungen zur Spezifikation des Kerndatensatz Forschung. Drs. 5066-16, Berlin, online at: http://www.wissenschaftsrat.de/download/archiv/5066-16.pdf, last checked on: 25/04/2016.

ONLINE RESOURCES

Last checked on: 25/04/2016. Resources in alphabetical order.

Academia.edu – Platform for academics
https://www.academia.edu

DANS – Data Archiving and Networked Services
http://www.dans.knaw.nl

da|ra – Registration agency for social and economic data
http://www.da-ra.de

DARIAH-DE – Digital Research Infrastructure for the Arts and Humanities / Germany
https://de.dariah.eu

DARIAH-EU – Digital Research Infrastructure for the Arts and Humanities
https://dariah.eu

DataCite
https://www.datacite.org

Data FAIRport Initiative – Find, Access, Interoperate & Re-use Data
http://www.datafairport.org

DFG RIsources (RI = Research Infrastructure)
http://risources.dfg.de

DINI Certificate – German Initiative for Network Information
https://dini.de/dini-zertifikat

DOI – Digital Object Identifier System
https://www.doi.org

DSA – Data Seal of Approval
http://datasealofapproval.org/en/community

ELIXIR – The European life-sciences Infrastructure for Biological Information
https://www.elixir-europe.org
https://www.elixir-europe.org/news/elixir-and-denbi-agree-collaboration-strategy

EOSC – European Open Science Cloud
https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

ESFRI – European Strategy Forum on Research Infrastructures
http://www.esfri.eu/about-esfri
http://www.esfri.eu/national-roadmaps

EUDAT – European data infrastructure
https://www.eudat.eu/what-eudat

Europe: Demand for qualified personnel – Interview with Barend Mons on 09/03/2016
http://primeurmagazine.com/weekly/AE-PR-05-16-58.html

Europeana – Europeana Foundation
http://www.europeana.eu

FAIR Data Principles – for comment
https://www.force11.org/group/fairgroup/fairprinciples

GBIF – Global Biodiversity Information Facility
http://www.gbif.org

GEOSS – Group on Earth Observations
http://www.earthobservations.org/geoss.php

German Digital Library
https://www.deutsche-digitale-bibliothek.de

GFBio – German Federation for Biological Data
http://www.gfbio.org

GWDG – Society for Scientific Data Processing, Göttingen
https://www.gwdg.de

Mendeley – Reference manager and academic social network
https://www.mendeley.com
https://data.mendeley.com

NCDD – Netherlands Coalition for Digital Preservation
http://www.ncdd.nl/en/about-the-ncdd

nestor – German competence network for digital preservation
http://www.langzeitarchivierung.de

nestor-Data Seal of Approval
http://www.langzeitarchivierung.de/Subsites/nestor/DE/nestor-Siegel/siegel_node.html

NoMaD – The Novel Materials Discovery Laboratory
https://nomad-coe.eu

OpenAIRE – Open Access Infrastructure for Research in Europe
https://www.openaire.eu

Open Government Partnership Netherlands
http://www.opengovpartnership.org/country/netherlands

ORCID-Initiative (Open Researcher and Contributor ID)
http://orcid.org

Pangaea – Data Publisher for Earth & Environmental Science
https://www.pangaea.de

PREPARDE Project –Data Journals List
http://proj.badc.rl.ac.uk/preparde/blog/DataJournalsList

ResearchGate – Academic social network
https://www.researchgate.net

RADAR – Research Data Repository
https://www.radar-projekt.org

RatSWD – German Data Forum
http://www.ratswd.de

http://www.ratswd.de/forschungsdaten/fdz
http://www.ratswd.de/forschungsdaten/info

RDA – Research Data Alliance
https://rd-alliance.org

RDA-DE – Research Data Alliance Germany
http://www.forschungsdaten.org/index.php/RDA-DE

Research Data Netherlands
http://www.researchdata.nl

re3data.org – Registry of Research Data Repositories
http://www.re3data.org

SOEP – German Socio-Economic Panel
http://www.diw.de/de/diw_02.c.221178.de/ueber_uns.html

Wissenschaftsrat – Evaluation Leibniz Institute for Educational Trajectories (LIfBi)
http://www.wissenschaftsrat.de/nc/arbeitsbereiche-arbeitsprogramm/evaluation.html#c20161

GLOSSARY

TABLE OF CONTENTS

INTRODUCTION

The German Council for Scientific Information Infrastructures (RfII) wants to use comprehensible language in its recommendations. To this end, it endeavours to ensure that frequent terms with multiple meanings are used consistently.

The German Council for Scientific Information Infrastructures has compiled a selection of terms initially relevant to the selected topic of focus: research data. Their use in concept and position papers on scientific policy (in particular from the EU, German Federal Government, German Länder, sponsors, scientific communities, etc.) has been taken into account, and the terms themselves are put in a neutral context wherever possible. Buzzwords will be avoided.

The clarifications of the terms below are not intended to provide comprehensive definitions, but are meant as guides to the formulations used by the German Council for Scientific Information Infrastructures.

TERMS

ACCESSIBILITY, AVAILABILITY, AUTHORISED ACCESS

Successful and effective research is based on accessibility to research findings and research resources with as few restrictions as possible, i. e. on 'open', low-threshold usage (open information, open data, open educational resources). German scientific organisations have called for open access to scientific knowledge numerous times – not only as an educational resource, but also as a research resource. Scientists need information infrastructures that provide access to data, information, and knowledge with the lowest possible practical, economic, and legal barriers. One key to achieving this, for example, are open access publishing models – licensing models at the institutional, regional, national, or community level that allow free access for non-commercial (e. g. scientific) purposes – or forms of information and data sharing up to the economic collectivisation of data and information (commons).
The availability of data and information extends beyond the technical, organisational, and legal issues to be resolved for access. Availability comprises all of the following aspects:

a. The indexing of data, from the assignment of resource IDs/labels/identification codes and metadata to the sophisticated automatic and/or intellectual indexing based on disciplinary specific standards;
b. The processing of imports and exports, up to and including their further distribution and reusability in new contexts (open science, open educational resources);
c. Storage, up to and including short-term or long-term archiving, with the goal of enabling unlimited, unrestricted access and full readability through continuous data protection.

- References
  → Metadata → Open access, open data – data sharing, open source

COMMUNITIES, DISCIPLINES

The term 'community' refers to a group (association, union) of researchers that are well interconnected socially and follow similar rules of conduct. Communities can form based on a common area of interest (e. g. the 'climate community', 'Heidegger community'), but also based on specific methods (e. g. the 'HPC community'), or even theories (e. g. the 'neurocritical community'). The term is often used rather loosely. Researchers can belong to numerous communities, and communities themselves can form or dissolve quickly in some cases.
In comparison with 'community', the term 'disciplines' – or more traditionally when emphasising educational aspects, the term 'fields' – is a broader term, whereby the order of disciplines is taxonomical (at universities or in the form of the specialised departments of the German Research Foundation (DFG)). With this in mind, 'interdisciplinarity' – as a separate mode – can in turn explicitly become the problem and topic.
There is an abundance of research papers in science studies on the topic of the difference between the terms 'field' and 'discipline' as well as on the controversial question of the minimum conditions characterising a 'discipline' (such as professional associations and journals, a recognised system of full professorships, self-reproduction through its own degree programmes, etc.).
In the context of the requirements on information infrastructures, it is not easy to state whether it is better for the measures/services/facilities to address the (smaller, temporary, relatively homogeneous) communities at the level of action or to address the (larger, intrinsically heterogeneous) disciplines and fields. For this reason, the term 'scientific community' frequently used by the German Council of Science and Humanities (or less frequently: 'specialist community') includes both definitions in an intentionally ambiguous manner.

DATA

Data forms the basis and is part of a scientific value-creation cycle of data, information, and knowledge that simultaneously generates economic findings. Information can be obtained from structured data, and knowledge from verified information. From a scientific perspective, data is used to gain and validate (empirical) knowledge. From an economic perspective, data is used to create (economic) value. From a societal perspective, data is relevant to education and can be used in many different ways for political purposes.
In general, the term 'data' also includes non-electronic and non-digital data from surveys, calculations, measurements, censuses, text, and the contents of historical documents. Nevertheless, the term 'data' is usually used in the sense of 'digital data'.
Scientifically relevant data is also produced outside of the sciences, for example statistical data from administrative bodies, mobility data, and telecommunication data. The percentage of unused, hidden, or largely untapped data is high.
The value of data depends decisively on verifiable data quality. Through the anonymisation of data, it is possible to balance data protection requirements and the desire for scientific analysis.
The processing and presentation capabilities of digital data have revolutionised and characterised the information and knowledge society, and they therefore require adequate, publicly accessible, and publicly controlled information infrastructures.

- References
  → Research data, research data management → Information infrastructures

DATA CURATION

In the classical sense, curation refers to the selection and documentation of the items in a collection. Data curation encompasses a wider range of tasks that includes the creation of data and the transformation of data into so-called *rich metadata*, for example. Data curation serves to help find, understand, and use the data stored in information infrastructures. In the framework of data management, data curation is a long-term service task that must be performed in addition to the 'basic service' of simply storing the data (long-term archiving). This service task includes the continuous maintenance of the metadata as well as ensuring the reusability of the data in terms of content and topicality. In light of the increasing data intensity of the sciences, curation is rapidly gaining importance as a prerequisite for reuse.
Ideally, infrastructures have curators who ensure that the tools are user-friendly/user-oriented and transform raw data into scientifically relevant information. On the one hand, the range of tasks to be performed to curate data depends on technical standards, but also on the demands of users. Data curation thus requires highly qualified personnel with knowledge of scientific subjects as well as of information technology in order to guarantee the interoperability of the data sets in a repository, for example. This is not achieved when researchers simply store uncurated data, which also reduces the potential for creating value within the data life cycle.

- References
  → Data life cycle → Research data, research data management → Dynamic knowledge integration → Metadata → Accessibility, availability

DATA LIFE CYCLE

The data life cycle is a model that describes the cyclical character of the work done with data of all kinds – including information – in its various stages of processing and use within the process of creating scientific value. The most important stages in this cycle include data generation (e. g. measurements), data preparation, data evaluation/analysis, storage (including long-term archiving), and making the data available through publication (e. g. in databases and repositories, as publications in journals, and on online platforms). The cycle even includes data reuse in additional or new research contexts, which can also arise from academic settings.
The cyclical character of the model emphasises the fact that use and reuse of the data generate new results in the form of further research data. As a consequence, the data management system used throughout this life cycle must ensure that the results can be reproduced across every stage of the cycle. Furthermore, during the data life cycle it is necessary to decide which data should be stored, published as a separate data set, or integrated into a publication, as well as how long it should be kept available and if and when it can be deleted. These decisions are currently being made based on various criteria by research teams or individual researchers, among others, who manage the data they generate all by themselves. Action must be taken to determine the best possible organisational procedures up to long-term archiving, e. g. through the development of relevance criteria and scenarios.

- References
  → Research data, research data management

DATA QUALITY

The term ‘data quality’ refers to the quality and reliability of data objects themselves. A prerequisite for the use of the data objects for scientific purposes is that the data they contain (e. g. of instruments, calculation algorithms) is obtained using documented standards and accepted methods, and that these standards are transparent and ensure sustainability. Whenever possible, the level of uncertainty in a piece of data should be suitably quantified.
The evaluation of the data quality is based on the requirements to be defined, for example on the accuracy of values measured, depending on the research question and therefore on the purpose for which the data will be used. In addition, the quality of the data depends greatly on whether or not the data sets and data collections contain sufficient information (if possible, in the form of standardised technical and topical metadata) on their generation, processing, and methods of analysis. This is an essential requirement for the reuse of the data and the reproducibility of research results.
The required completeness and currency of the data and its supplemental information, which depend on the application, as well as the long-term availability and citability of the data are in turn prerequisites for the quality of the information infrastructures and services that allow it to be stored, located quickly (*retrieval*), accessed, and reused (also in the context of long-term archiving).

- References
  → Research infrastructures → Information infrastructures → Metadata

DIGITAL LITERACY, INFORMATION LITERACY

The term ‘information literacy’ refers to a key competence for various strata of the information society. It includes more specific terms such as media literacy, digital literacy, reading literacy, writing literacy, and numerical literacy. Information literacy is considered a prerequisite for full communication skills and proper conduct. It deals with the handling as well as the use of informational materials with different formats, scopes, and references by user communities having varying quality requirements. Information literacy can roughly be considered to have an instrumental/professional component (school, university, working world) and a reflective/general education dimension. Achieving this literacy should enable users to use information on their own accord and promote the responsible evaluation and use of information. With this definition in mind, information literacy thus expressly includes the creation of content for third parties (i. e. the creation of messages). With the help of digital information infrastructures, it is necessary to expand the teaching of information and media literacy systematically at all levels of the educational system.

- References
  → Data quality → Accessibility, availability

DYNAMIC KNOWLEDGE MANAGEMENT / KNOWLEDGE INTEGRATION

The phrase ‘dynamic knowledge integration’, which has not been encountered frequently in Germany to date, is used by the German Council for Scientific Information Infrastructures as a figurative term for the new world of the complex management of research data. The term originates from phrases such as *knowledge integration* or *knowledge integration dynamics* as they have been used since the 1980s in numerous studies on organisational and knowledge management, especially in the research and development departments of large industrial companies.

The intentionally general term describes the goal of collecting data, information, and knowledge from researchers and research groups in order to solve complex problems across multiple research cycles. This enables higher quality innovation processes and the development of synergies. In this cyclical process, research data (research materials) is made into new results (products), which in turn become research data for the next evaluation phase. The production of knowledge over the course of several cycles is associated with how transparently this dynamic process of attaining knowledge is documented and how verifiable it is. This is especially relevant when it becomes clear in later cycles that measurements or interpretations from previous cycles are wrong or were under the influence of factors that were unknown at the time. Many research and development problems can only be solved efficiently over time when different disciplines, fields, and communities work together through the dynamic integration of distributed knowledge resources.

This not only requires the willingness and ability to communicate and cooperate closely in internal and external networks, but also suitable tools, processes, and infrastructures. These need to be built in cooperation by scientists, data specialists (data scientists), and suitable institutions and then developed further to form powerful and sustainable information infrastructures that meet the needs of researchers.

Interoperability and integrability should ensure that the results obtained through research processes (research data, publications, project documentation) are stored, documented, indexed, and made reusable and available for additional value creation in information infrastructures directed towards sustainability, accessibility, and availability according to transparent international standards.

- References
  → Data life cycle → Research infrastructures → Information infrastructures → Integration, interoperability

INFORMATION INFRASTRUCTURES (RFII), E-INFRASTRUCTURES (EU)

Information infrastructures are technically and organisationally networked services and facilities for accessing and maintaining databases, information bases, and knowledge bases. In the context of the RfII's counselling work, they primarily serve research purposes, are often objects of research, and always function as an enabler.
Information infrastructures must always take into account that knowledge bases in universities, research facilities, archives, libraries, and museums are available in purely analogue or digital form or in a combination of analogue and digital forms. The purpose of the digitisation of analogue knowledge bases is to integrate and merge digitised data and native digital data into uniform, integrated work environments with the goal of achieving dynamic knowledge integration. Like the term 'e-infrastructures', the term 'information infrastructures' commonly encountered in Germany is also increasingly being used to refer to the digital information and communication technologies employed in research.
The performance of digital information infrastructures depends significantly on the amount invested in digitising the content, user-friendly access methods, technical features, international standards, and effective tools. The level of information literacy of the users and personnel and the associated quality of the custom services provided are equally relevant.

- References
  → Data → Research infrastructures

INTEGRATION, INTEROPERABILITY

According to its Latin origin, the word 'integration' refers to the undamaged restoration of an object, but over time it has taken on a variety of meanings when used as a general term as well as when used as technical jargon. In connection with information infrastructures as well as with IT systems, software, and data-related questions, it is particularly necessary to differentiate between the following meanings (which in some cases overlap) of the term:
Integration can be used to refer to a process or an act of connecting/merging items in which the items connected/merged remain intact or even unchanged; in this case, integration is an effect. With this in mind, data can be integrated into software environments; new requirements can be integrated into established standards; distributed knowledge can be merged to form a new whole (integrated knowledge), etc.
However, integration can also be used to refer to those types of connections in which the 'seamlessness' of the transitions between the items connected together is emphasised; in this case, integration ('good' integration or 'integrity') tends to be a property or a state. Complex IT systems can therefore be 'highly integrated', a landscape of research infrastructures can be characterised by a high level of integration, etc.
One aspect of integration in this sense is 'interoperability'. This term is intended to refer to the decisive practical issue that devices or systems should be able to communicate/cooperate with (possibly unknown) other devices or systems – even in flexible scenarios. Transparent and sufficiently standardised interfaces are technical requirements for interoperability.
Finally, integration can be viewed as a process that plays out on the external boundaries (in which case it characterises the scale of openness/integrability of the whole it then bounds). In this sense, the term 'integrability' or the phrase 'structures designed for integration' is used in connection with IT as well as in the context of organisations, cultures, etc.

- References
  → Dynamic knowledge integration → Metadata → Standards, standardisation

METADATA

Metadata is data about data. Metadata is used to describe data and to help users find data, evaluate the suitability of the data found for the intended purpose, and integrate the data found into the user's own system environment. Metadata in this case typically encompasses the description of the syntax (e. g. format information), semantics (e. g. information on data content), quality (e. g. information on the accuracy or uncertainty of data), and legal aspects (e. g. the rights of use) as well as of the source of the data, how it was prepared, and the path used to access the data described (e. g. an Internet address).
Metadata is often found in completely different degrees of structure and harmonisation and in varying degrees of completeness. There are numerous international and discipline-specific standards, which often come from the library sciences, for creating and harmonising metabases. A well-known example is the Dublin Core standard, and the international metadata standard "Resource Description and Access" was recently introduced as a common framework to describe digital and analogue resources in libraries, archives, and museums. The Metadata Standards Directory Working Group of the Research Data Alliance supports the development, application, and transparency of metadata standards for research data and also maintains a directory.
Processes and developments for the automated derivation (annotation) of metadata from existing and emerging digital databases make it possible to exploit metadata for further processing. In such cases, the metadata itself can become research data.

- References
  → Data → Data quality → Integration, interoperability → Standards, standardisation

OPEN ACCESS, OPEN DATA – DATA SHARING, OPEN SOURCE

Open Access (OA) (or the following for extended access requirements: Open Data and Open Content) refers to the low-threshold, preferably free and broad access to the findings coming from the scientific system. Nowadays, numerous research funders and a significant proportion of researchers themselves prefer open access (the industry is also demanding such "open" access to the findings of publicly-funded research projects). Two models for the public co-financing of open access publications have been established: 'green' open access (second publication in digital form) and 'gold' open access (first publication in digital form with the goal of achieving the greatest possible visibility and availability immediately).
Under no circumstances should open access be understood as allowing completely unrestricted exploitation. OA publications are also subject to legal regulations. This also applies to models for the exchange of data (open data and data sharing) or program code (open source) on a non-monetary basis implemented at the technical level or otherwise practised and agreed on. Using one of the various types of Creative Commons licenses available, the owner of the copyright and the related or neighbouring rights can grant additional freedoms (e. g. for further processing or exploitation). The demand for open access not only to publications, but also to research data has become increasingly important in the international discourse. Internationally, and somewhat belatedly also in Germany, the demand for open educational resources has become widespread in educational policies, too.
In contrast to the business world where assets are monetised, an increasingly digitised science appreciates the public domain aspect of research resources, of intermediate and final results, and of data sharing as an opportunity and catalyst for the collaborative generation of knowledge (in contrast to the business world). Due to digital technology, the basic legal and economic conditions that allow the sciences their typical scientific freedoms have begun to shift. Access to research data must therefore be organised and ensured in favour of education and the sciences by expanding the legal limitations, e. g. by restricting exploitative

use of data or, if necessary, compensating for use according to the level of public interest. One method of achieving this is to have researchers, libraries, etc., proactively license the data or publications as 'open' to the public domain.
An essential question is whether or not to allow the accessible data and content to be expanded, analysed, combined, or even improved. In some cases, access alone is adequate, but in other cases, and especially in the case of open educational resources and research data, rights to further use are essential (for example in the form of Creative Commons licenses).

- References
  → Accessibility, availability

RESEARCH DATA, RESEARCH DATA MANAGEMENT

Research data is data created in the course of scientific activity, e. g. through observations, experiments, simulations, surveys, interviews, the study of sources, records, digitisation, or evaluations. In actual research, one differentiates, although not always clearly, between primary research data and secondary research data, which documents and contextualises the process of creating primary data. In the research process, secondary data can turn into primary data again, which is significant in terms of the life cycle of research data.
Research data management includes all measures – even organisational measures extending beyond research activity in the narrow sense – that need to be taken in order to obtain high quality data, to follow good scientific practice within the data life cycle, to make results reproducible, make data available for reuse, and possibly fulfil existing documentation requirements (e. g. in the healthcare sector). Increasingly, funding organisations are requiring the project leaders to create a data management plan, and scientific institutions to be prepared for data management systems that guarantee sustainability. Data management plans for the start of a project or a research task are suitable for describing the data to be used and generated as well as the documentation, metadata, and standards required, for stating the potential legal restrictions (e. g. data protection) early on, for planning the necessary storage resources, and for specifying the criteria to determine which data should be made available externally in which form and how long it must be stored. At the organisational level, research facilities (e. g. universities) must ensure access to the corresponding infrastructure services within the facility (e. g. by creating new capacities or expanding the existing capacities) or in cooperation with external partners (through cooperation agreements, etc.).

- References
  → Data → Data life cycle → Research infrastructures → Metadata

RESEARCH INFRASTRUCTURES

Research infrastructures are scientific infrastructures which serve research (plants, resources, facilities and services). Such infrastructures include:

a. Large devices or instruments used for research purposes (e. g. research vessels, satellite and tracking stations, telescopes, particle accelerators),
b. Information and knowledge resources such as (digital or non-digital) collections, archives, libraries, databases,
c. Information and communication technology infrastructures such as computers, computer networks (grid, cloud computing),

d. Software,
e. Any other system, resource, facility, or service used in scientific research to perform a comparable function.

- References
  → Research data, research data management → Information infrastructures

STANDARDS, STANDARDISATION

Standards and rules are used in the processing of data, information, and knowledge with the goal of standardisation, and are thus a form of quality assurance (quality in the sense of reusability) in terms of content and in formal, legal, organisational, and technical terms. Formats for information media as well as rules for their registration (metadata) and storage had already been defined in the early phases of the classic library and archiving sciences. The German Institute for Standardization (DIN) founded in 1917 organises in the public interest standards for following generally accepted good engineering practices in business, the sciences, and administration. Their goal is to assure quality and the practical usability of innovations. One of the best known German standards is the DIN standard for paper sizes published in 1922. The International Organization for Standardization (ISO) was founded in 1947 in Geneva, Switzerland.

Digitisation in all areas of science and the high dynamics of digital technology require standardisation, especially for data and metadata, data exchange formats, interfaces, data models, markup languages, and vocabularies. There are a number of document formats such as PDF or HTML and an even larger number of media-specific metadata formats such as Dublin Core for object descriptions on the Internet, EAD for manuscripts and archives, or MARC, MAB, and MODS for library data exchange formats, etc., but there are no generally binding standard formats for metadata. This allows customisations for media-specific descriptions while simultaneously making the interoperability and compatibility of data and documents from different origins more difficult.

In the meantime, there is a whole series of important ISO standards relating to data, software programmes and processes, documentation, and project management. The application, observation, and coordinated development of defined international norms and standards are prerequisites for integrity and trustworthiness as well as for the creation of sustainable and efficient information infrastructures that guarantee usability, interoperability, accessibility, connectivity, and long-term availability.

Standardisations and the observance of standards for data management entail extensive yet scientifically and economically sensible investments in quality assurance and sustainability. Scientific information infrastructures should be integrated into international standardisation processes whenever it makes sense and the opportunity is available.

- References
  → Data → Data quality → Integration, interoperability → Metadata

CONTRIBUTORS

THE COUNCIL

Representatives of the scientific users

Prof. Dr. LARS BERNARD
TU Dresden – Department of Geosciences
Prof. Dr. Dr. FRIEDERIKE FLESS
DAI – German Archaeological Institute and Freie Universität Berlin
Prof. Dr. FRANK OLIVER GLÖCKNER
Max Planck Institut for Marine Microbiology and Jacobs University Bremen
Prof. Dr. STEFAN LIEBIG
Bielefeld University – Department of Sociology
Prof. Dr. WOLFGANG MARQUARDT
Forschungszentrum Jülich
Prof. Dr. OTTO RIENHOFF
University of Göttingen – Department of Medical Informatics
Prof. Dr. JOACHIM WAMBSGANß
Centre for Astronomy of Heidelberg University (ZAH)
Prof. Dr. DORIS WEDLICH
KIT – Karlsruhe Institute of Technology: Division I – Biology, Chemistry and Process Engineering

Representatives of the information facilities

SABINE BRÜNGER-WEILANDT
FIZ Karlsruhe – Leibniz Institute of Information Infrastructure
Prof. Dr. THOMAS BÜRGER
Saxon State and University Library Dresden (SLUB)
Prof. Dr. PETRA GEHRING
TU Darmstadt – Department of Philosophy
Dr. GREGOR HAGEDORN
Museum für Naturkunde – Leibniz Institute of Evolution and Biodiversity Science
Prof. Dr. MICHAEL JÄCKEL
Trier University
Dr. MARGIT KSOLL-MARCON
The Archives of the Bavarian State
Prof. Dr. KLAUS TOCHTERMANN
German National Library of Economics and Kiel University
Prof. Dr. RAMIN YAHYAPOUR
GWDG – IT in science and University of Göttingen

Representatives of the federal and state governments

RÜDIGER EICHEL
Ministry of Science and Culture of Lower Saxony

Dr. THOMAS GRÜNEWALD
Ministry of Education, Science and Research of North Rhine-Westphalia

Dr. STEFAN LUTHER
Federal Ministry of Education and Research

Dr. DIETRICH NELLE
Federal Ministry of Education and Research

Representatives of the public

Dr. habil. REINHARD BREUER
Science Journalist

Dr. h.c. ALBRECHT HAUFF
Thieme Group

Dr. SIMONE REHM
TRUMPF Group (until 12/2015)

ANDREA VOßHOFF
BFDI – German Federal Data Protection Authority

Chairpersons

Prof. Dr. OTTO RIENHOFF
Chair

SABINE BRÜNGER-WEILANDT
Deputy Chair

GUESTS AND EXTERNAL EXPERTS

Guests

PETER BÜTTGEN
BFDI – German Federal Data Protection Authority
Dr. HANS-JOSEF LINKENS
Federal Ministry of Education and Research
Prof. Dr. NORBERT LOSSAU
University of Göttingen
Dr. STEFAN WINKLER-NEES
German Research Foundation (DFG)
Dr. h.c. PETER WITTENBURG
Max Planck Computing and Data Facility

Consultation of the Working Group International Orientation, April 2016

Prof. Dr. JAN CARLSTEDT-DUKE
Karolinska Institutet, Sweden
Dr. RON DEKKER
Netherlands Organisation for Scientific Research, Netherlands
Prof. RNDr. LUDĚK MATYSKA
Masaryk University, Czech Republic
WALTER STEWART
Research Data Canada / Walter Stewart & Associates, Canada
Dr. PAUL WONG
Australian National Data Service, Australia

HEAD OFFICE

Dr. BARBARA EBERT
Director
Dr. STEPHANIE HAGEMANN-WILHOLT
Dr. SVEN RANK
Dr. ILJA ZEITLIN
Scientific Officers
SARAH BARUFFALDI
Team Assistant