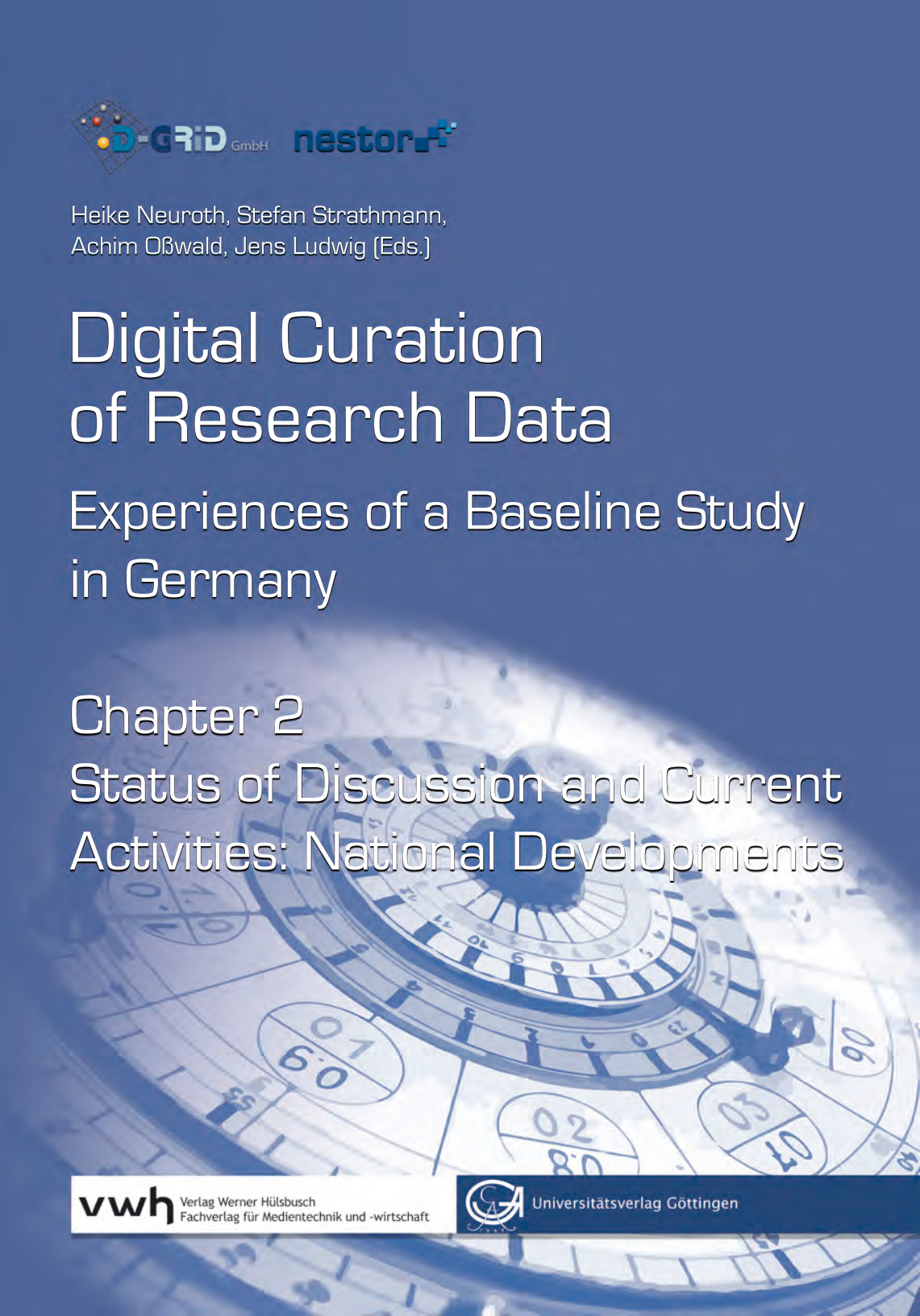
Heike Neuroth, Stefan Strathmann,
Achim Oßwald, Jens Ludwig (Eds.)

Digital Curation of Research Data

Experiences of a Baseline Study in Germany

Chapter 2
Status of Discussion and Current Activities: National Developments



vwh Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

Universitätsverlag Göttingen

**Heike Neuroth, Stefan Strathmann,
Achim Oßwald, Jens Ludwig (Eds.)**

Digital Curation of Research Data

Experiences of a Baseline Study in Germany

vwh
Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

Digital Curation of Research Data

Herausgegeben von Heike Neuroth, Stefan Strathmann, Achim Oßwald und Jens Ludwig · im Rahmen des Kooperationsverbundes nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland · http://www.langzeitarchivierung.de/

Edited by Heike Neuroth, Stefan Strathmann, Achim Oßwald and Jens Ludwig · within the context of nestor – Network of Expertise in the Long-Term Storage of Digital Resources for Germany · http://www.langzeitarchivierung.de/

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter http://www.d-nb.de abrufbar.

Bibliographic information of the German National Library
The German National Library lists this publication in the German National Bibliography; detailed bibliographic data is available online at http://www.d-nb.de.

Die Inhalte dieses Buches stehen auch als Onlineversion über die Website von nestor zur Verfügung / This work is available as an Open Access version at the nestor website: http://nestor.sub.uni-goettingen.de/bestandsaufnahme/index.php?lang=en

vwh Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

in Kooperation mit dem Universitätsverlag Göttingen
in cooperation with the Universitätsverlag Göttingen

Printed in Poland · ISBN: 978-3-86488-054-4

Content

2 Status of Discussion and Current Activities: National Developments

Stefan Winkler-Nees

The discussion in Germany about the handling und reuse of scholarly data from publicly funded projects is shaped by researchers who are relatively autonomous and independent. The basic principle of academic freedom, as laid out in article 5, paragraph 3 of the German constitution ("… Art and scholarship, research, and teaching shall be free. …" e.g. this freedom may include the right, not to publish), also determines attitudes about ways in which to deal with individual or collaboratively acquired scholarly findings, and, correspondingly, the willingness of researchers to make their data available. Claims, both existing and perceived, to knowledge produced by him or her lead to uncertainty and concerns regarding the unmanageable nature of subsequent re-use and the potential misuse of data once it has been provided. A constructive discussion about the potential and possibilities of the standardized preparation of research data is hindered by these uncertainties.

At the same time, it is generally acknowledged that a sustainable approach to project results was not adequately considered in the past, especially in publicly funded research in the light of the "virtualization of research."[40] In addition to the accelerating pace of transformation with simultaneous fundamental changes in academic research processes, and the potential value offered by professional information management, require organizational modifications to the current structural framework.

In the recent past, and in some cases already in the course of the last decade, several activities were developed independently of each other. These activities led to the development of infrastructures that were highly regarded on disciplinary levels. These "grassroots projects" were, however, unaccompanied by any universal interdisciplinary discussion, nor were they embedded in an overarching, consensually-agreed upon concept

40 See Horlings et al. (2006).

for data management. The heterogeneous, diverse research landscape in Germany, with its various research performing organizations, universities, federal structure and funding streams[41] and frequently ambiguously undefined responsibilities, has meant that the early impulses that were implemented were not coordinated with each other, and structural measures that planned for the future were not discussed comprehensively. The German Research Foundation (Deutsche Forschungsgemeinschaft [DFG]) report "Safeguarding Good Scientific Practice" was the first to contain a general, cross-disciplinary requirement to preserve data beyond a fixed period of time.[42] Correspondingly, there was originally no requirement to provide research data for scholarly purposes and for reuse. Research data first came to the forefront beginning with the intensifying debate about Open Access publishing and the corresponding change in the awareness of the necessity of access to digital information. In addition to academic researchers, stakeholders from the research organizations, from science policy, from the information infrastructure organizations, and from funding agencies began to take on more responsibility for developing, coordinating, and implementing appropriate measures.

2.1 Research Organizations

The strongest motivation to secure, archive, and make research data available according to professional criteria lies in recognizing and making use of the potential benefits that can be thus realized. Academic disciplines that work closely with digital data and, while doing so, engage in international collaboration, have in many cases built systems and infrastructures that suit their specific requirements. Due to the high acceptance rate on the part of the scholarly community, these systems represent successful flagship projects for the meaningful and effective use of research data.

41 See http://www.research-in-germany.de/main/2866/research-landscape.html for further information.

42 See DFG (1998).

One example of this engagement is the situation in the field of marine and environmental sciences, which led to the creation of the PANGAEA information system more than two decades ago. The network of the ICSU[43] World Data Center has established an internationally recognized information infrastructure through cooperation between one of the leading German marine research facilities, the Alfred Wegener Institute for Polar and Marine Research in Bremerhaven, and the Marum Center for Marine Environmental Sciences at the University of Bremen. In addition to the sustained commitment of those involved, the increasing integration of PANGAEA by scientists in their research activities as a source of information and as a data repository contributed to its success. Furthermore, funding organizations were successfully convinced of the significance and use of this system, so that a long-term and sustainable operation became feasible. The involvement of the Helmholtz Association of German Research Centres (Helmholtz-Gemeinschaft Deutscher Forschungszentren e.V.) in the discussion about access to research data also contributed to this development, particularly in the context of their open access activities. Thus, many research centers in the association are committed to the principle of free access to research data and an additional ICSU World Data Center (WDC) could be established (the WDC for Remote Sensing of the Atmosphere).[44]

Until now, the research institutions within the Leibniz Association (Leibniz-Gemeinschaft) have been reviewing individual approaches to deal with research data that demonstrate success in several initiatives at the disciplinary level. For example, the Leibnitz Institute for Psychology Information and Documentation (Leibniz-Zentrum für Psychologische Information und Dokumentation [ZPID])[45] has built a unique reference database for psychological literature, testing procedures, and various other materials. As a service provider for the Leibniz Association, the German National Library of Science and Technology (Technische Informations-

43 See ICSU, http://www.icsu.org.

44 See World Data Center for Remote Sensing of the Atmosphere, http://wdc.dlr.de.

45 See Leibniz Institute for Psychology Information (ZPID) http://www.zpid.de.

bibliothek Hannover [TIB])[46] in Hannover assumed the task of making research data available in a more efficient way for research. Research data, which form the basis of publications, are not hosted by the TIB itself as a general rule, but are registered and made searchable by assigning a DOI in order to create lasting citability. This service is intended to be generally available for all disciplines and institutions, including those outside the Leibniz Association.

The issue of dealing with research data at universities has as yet no particular nationwide resonance in Germany. In contrast to international approaches, German universities do not view themselves as under the obligation or even capable of initiating measures to improve the situation. However, several university libraries have seized the initiative, such as the Göttingen State and University Library in Göttingen (Niedersächsische Staats- und Universitätsbibliothek Göttingen [SUB]). The nestor project,[47] conducted under the aegis of the German National Library and carried out in cooperation with other libraries, museums, and institutions of higher education, has laid the basic groundwork for the ongoing discussion with a series of studies. On a disciplinary level, however, researchers predominantly use existing options available from research organizations or they attempt to make digital content available using solutions from their own institutes. In many cases, this type of approach does not possess the required sustainability and availability on a supra-regional level for research data repositories. Researchers at German research academies are in a comparable situation. The TELOTA system, which was established at the Berlin-Brandenburg Academy of Sciences and Humanities (Berlin-Brandenburgische Akademie der Wissenschaften [BBAW]) in 2002, is one example.[48] Here, the research data in particular, such as editions, dictionaries, bibliographies and documentation, which are produced in large quantities by academy programs, are intended to be preserved and made available in a suitable manner. The wide range of subjects, from ancient inscriptions to medieval glass painting and records about silk road docu-

46 See German National Library of Science and Technology, http://www.tib.uni-hannover.de/en.html.

47 See nestor, http://www.langzeitarchivierung.de.

48 See TELOTA (2011).

ments, from ancient inscriptions to the treatment of electronic "heritage assets," seems to be well-suited for the development of a system with services for the entire range of disciplines in the humanities.

In 2007, the Max Planck Society (Max-Planck-Gesellschaft [MPG]) established a new service unit, the Max Planck Digital Library [MPDL]), which "is intended to assist researchers at the Max Planck Society in organizing the research information process" („… den Forschern der Max-Planck-Gesellschaft helfen soll, den wissenschaftlichen Informationsablauf zu organisieren …").[49] The primary focus of the MPDL is on providing research information, the dissemination of information, eScience services, and providing support for the Max Planck Society in implementing open access. With the help of internally funded projects and the engagement of externally funded projects, researchers, particularly those of the Max Planck Society, should be able to have the best possible access to digital resources.

2.2 Recommendations and Policies

Before 1998, the handling of research data, according to the current understanding, did not play a significant role in shaping research policies in Germany. It was only with the increasing prominence of cases of academic misconduct within the bodies of the DFGDFG, that more attention was paid to the necessity of access to research results. Paramount was not the aspect of scholarly reuse and future use of data, but instead the possibility of verifying the accuracy of data after publication. The implementation of this commitment to accessibility was intended to be carried out by all recipients of funding from the DFG who would pledge themselves to "ensure good scholarly practices." In a memoranda published by the DFG in 1998, recommendation seven stated that: "primary data, as a basis of publications, are to be saved on durable and safe media at the same institu-

49 Max Planck Digital Library, http://www.mpdl.mpg.de.

tion in which they arose for at least ten years".[50] The motivation for this recommendation was to make results in publications more reproducible with the help of the data, and to provide evidence of scholarly misconduct, retrospectively as well. The form, formats, and responsibilities were not further specified, so this practice does not satisfy the requirements of modern data management.

The "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities" ("Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen")[51], which was signed by many German and international research organizations on October 22, 2003, can be considered an integral milestone on the path to improved accessibility to scholarly knowledge and research results. Although the focus lay on open access to research literature, the Berlin Declaration expanded the definition of open access to research information "as a comprehensive source of human knowledge and cultural heritage" as a response to the diverse possibilities of information access via the internet.

The significant changes in research and the provision of research information led to intensive discussions about strategic modifications to funding options for research information infrastructures. Accordingly, the DFG Committee on Scientific Libraries and Information Systems (Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme [AWBI]) prepared a position paper in 2006 that recommended modified priorities for funding measures.[52] This paper focused on the changing research requirements caused by increasing digital networking. With respect to research data, it is essential to develop new "structures for storing, referencing and making data available"[53] („Strukturen zur Speicherung, Referenzierung und Verfügbarmachung"). Established information institutions like libraries, archives, and museums are considered major actors

50 DFG (1998): „Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden."

51 See Website "Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen" (2006).

52 DFG (2006).

53 Ibid.

whose technical expertise in the field of information management and preservation should be connected to the new requirements for providing access to information. At the same time, it is necessary to factor in the varying requirements in different academic disciplines in this discussion and to create opportunities for referencing and making research data available. It is of primary importance to achieve increased willingness on the part of researchers in working together on the establishment and use of information infrastructures. This need was also highlighted in a study commissioned by the German Federal Ministry for Education and Research (Bundesministerium für Bildung und Forschung [BMBF]).[54] According to this study, researchers for the most part decide for themselves what they want to use for their work, which is why information infrastructure options must be attractive.

In 2008, the Alliance of German Research Organizations (Allianz der deutschen Wissenschaftsorganisationen), which includes some of the biggest research organizations in Germany, established the priority initiative "Digital Information" with a funding period until 2012.[55] One of the foundations of this initiative was the awareness that there are no systematic approaches and methods and no sustainable infrastructures for the backup, provision, and archiving of research data. Organizational and technical aspects as well as legal and financial aspects are largely unclear. Therefore, the main objective in the area of research data is to develop and implement coordinated measures that take discipline-specific differences into account to ensure the efficient and professional handling of research data in Germany. These activities should focus on three areas: (1) development of a common policy agreed upon by all alliance and partner organizations; (2) encouragement for the development of individual information infrastructures, which should be designed and developed as pilot projects by subject experts and information specialists working in close cooperation; and (3) at a later stage, the definition and characterization of the different use scenarios in the various academic disciplines. These measures should include all groups of stakeholders: researchers as data producers and users, research institutions and universities, infrastructure institutions as well as

54 See Horlings et al. (2006).

55 See Alliance of German Science Organisations (2008).

relevant parties in the state and federal government. The Alliance Initiative working group on "research data" produced a paper on "Grundsätze zum Umgang mit Forschungsdaten" ("Principles for the Handling of Research Data"), which was adopted on July 24, 2010, by the board of directors of all Alliance partner organizations.[56] This paper supported as a basic principle the free and open access to data from publicly-funded research in accordance with legal requirements. The rights of researchers are to be respected at the same time. Equally, the differences between academic disciplines and their respective requirements should be considered. This paper also recommended placing a higher value on making research data available, along with the necessary investment of resources that are connected, and to establish them as an integral part of a scholarly reputation. Furthermore, the recommendations stressed the necessity of integrating the management of research data and their methods and mechanisms into certification programs for infrastructure experts as well as into academic curricula. Research data should be collected, saved, and archived according to standards, both those currently existing and those yet to be developed. Finally, the document recommended developing suitable infrastructure in combination with sustainable research data management. After these recommendations are signed, appropriate measures between partner organizations should be coordinated to ensure that these suggestions are implemented.

Independently of the Alliance Initiative "Digital Information," the DFG subcommittee on information management, part of the committee for scientific libraries and information systems, published the "Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsdaten" ("Recommendations for the secure preservation and provision of digital research data").[57] These recommendations were the result of various workshops and roundtable discussions among experts at the initiative of the DFG. They included a definition of primary data for research,[58] as

56 See Alliance of German Science Organisations (2010).

57 See DFG (2009b).

58 In these recommendations, the term "primary research data" ("Forschungsprimärdaten"), which was used in earlier papers in this form, is used. Since this usage

well as the recommendation to adequately factor in subject-specific aspects. Research data should be preserved and secured under widely accepted standards and identification of the data producer should be provided. In this paper, researchers are requested to allocate research data make access free, unrestricted by local or national borders, if possible. Generally, data are to be provided with metadata under existing standards or those currently under development. Representatives of all academic disciplines are called upon to develop mechanisms and methods for suitable quality control.

In 2009, the Joint Science Conference (Gemeinsame Wissenschaftskonferenz des Bundes und der Länder [GWK]) asked the Leibniz Association to develop a comprehensive concept for subject-specific information infrastructures in Germany.[59] The commission Future of Information Infrastructure (Zukunft der Informationsinfrastruktur [KII]) that was created as a result, prepared recommendations during 2010 in eight thematically-oriented working groups on the most important aspects of future information infrastructure. Most of these working groups were closely connected in content and organization with the complementary working groups of the Alliance Initiative "Digital Information." In contrast to the functional nature of the Alliance working groups, who were not limited by time restraints, the KII commission had a temporary and strategically limited appointment that lasted only until the completion of its assignment. The working group "Forschungsdaten" ("Research Data") stated that there is a strong need for action despite numerous individual activities. This overlaps with the tasks identified by the Alliance working group "Research Data": they listed organizational, technical, legal and in particular financial challenges. They explicitly recommended considering the disciplinary differences. This is also reflected in the recommendations that were developed, which were addressed to the key players in the management of research data. The fact that in those academic disciplines with well-developed international networks, already institutional structures are in place, demonstrate that accredited and trustworthy institutions are in a

combines an unnecessary restriction with a definition of primary data that has not been generally agreed upon, the term "research data" is preferable.

59 See Kommission Zukunft der Informationsinfrastruktur (2011).

position to take a leadership role in the process of developing a sustainable way of providing research data. The working group stated that universities and research institutions are responsible for participating in raising awareness. This type of outreach activity could be achieved by measures such as the establishment of central structures, the implementation of data management plans, and providing options for securing data while following good scientific practice. Moreover, it is necessary to make the legal requirements and framework for dealing with data clearer. The research funding agencies are called upon to offer programs to support pilot projects and research projects in need of professional research data management, and to provide means for the development of discipline-specific organizational forms. Information infrastructures have the duty to establish local services and advising. Corresponding activities should be cross-linked nationally in close consultation with representatives from the individual disciplines. Likewise, both researchers and employees of research institutions should be able to take advantage of the existing expertise in information management with a focus on research data in the form of training courses and continuing education. The working group viewed as another essential item the linking of research data repositories with appropriate publication databases from publishers. Federal and state governments, as institutions funding research with public money, were advised to view research data as a national cultural asset and therefore to create possibilities for its sustainable preservation and future access. This must be accompanied by a clear definition of responsibilities and the installation of appropriate organizational structures as well as cooperation in clarifying legal frameworks. In addition, they expressed the need to provide immediate resources for a fundamental establishment of an appropriate research data infrastructure.

The paper "Comprehensive approach for information infrastructure in Germany" ("Gesamtkonzept für die Informationsinfrastruktur in Deutschland"), prepared by the commission, was presented to the GWK in May 2011. The concept as a comprehensive planning document and the process of structural cooperation between all key players in information infrastructure was welcomed and the German Council of Science and Humanities (Wissenschaftsrat [WR]) was asked to incorporate the results by mid-2012 in its paper "Recommendations for research infrastructures."

In an extensive position paper that was published in January 2011, the German Council of Science and Humanities addressed the significance of information infrastructures for research in Germany.[60] In this paper, research collections, libraries, archives, and data collections in the broader sense were subsumed under the term information infrastructures. In line with the recommendations of other research policy organizations, the German Council of Science and Humanities attributed great importance to the development of information infrastructures for research in Germany. Special emphasis was placed on the role of universities, which needs to be significantly expanded, funding on the state level, meticulously coordinated planning, and the close integration of the academic community into the design process. The German Council of Science and Humanities advocated developing a comprehensive national strategy for information infrastructure ("nationale Gesamtstrategie für Informationsinfrastrukturen") for Germany by 2020.

2.3 Information Infrastructure Institutions

In 1996, the German Rectors' Conference (Hochschulrektorenkonferenz [HRK]) declared in its recommendations from the 179th plenum session that „in der Informationsgesellschaft […] Methoden und Techniken der Erzeugung, Verbreitung und Vermittlung von Wissen grundlegend verändern [werden]" ("there [will be] a fundamental change […] in the methods and techniques of generating, distributing and transmitting knowledge in the information society").[61] As an immediate need for action, „zentrale Einrichtungen […] innerhalb der Hochschulen […] Rechenzentren, Medienzentren und Bibliotheken verstärkt Dienstleistungsfunktionen für die Fachbereiche übernehmen [sollten]" ("data centers, media centers, and libraries as central institutions […] within the universities […] [should]

60 See Wissenschaftsrat (2011b).

61 German Rectors' Conference (1996).

take on more service functions for the individual subject areas").[62] In principle, this need still exists today. Fifteen years ago, the emphasis was on the use of information technology in instruction and the dissemination of set course material, rather than on the provision of research data. However, the requirements in information science and organizational matters hardly differ in both areas.[63] The increasing use of digital information systems in research is an irreversible process[64] accompanied by specific challenges. The operators of information infrastructures have a special responsibility. Discipline-specific information infrastructures were discussed above and in certain ways, they reflect the self-organized response to a subject-specific need for digital information services. The needs of researchers are foregrounded here, which in the past, owing to insufficient consideration of information science/technical requirements, led to solitary structures that were not networked in many cases.

The need for action on the part of the "central institutions" – in the sense of assuming responsibility – as identified by the HRK in 1996, is now increasingly being taken up by individual shareholders. The TIB is one example, as already mentioned. In addition to the basic functions of a library in the field of providing access to literature, the active design and development of systems for the information supply of digital content is becoming increasingly important. Crucial at this point is the employment of internal information/technical expertise in information services that reflects research requirements. In addition to organizational and information/technical aspects, the user's perspective must be examined and considered above all. Of great importance in this context is the creation and communication of added value in the sense of providing incentives not only to use digital information systems through *data retrieval*, but also to enrich these digital information systems by providing information. The TIB in Hannover, for example, established itself as an agency for the awarding of Digital Object Identifiers (DOI) and is one of the founders of

62 Ibid.

63 See Alliance of German Science Organisations (2008); Alliance of German Science Organisations (2010); and Kommission Zukunft der Informationsinfrastruktur (2011).

64 See Horlings et al. (2006).

the international DataCite Consortium.[65] On the federal level, the German National Library of Medicine (Deutsche Zentralbibliothek für Medizin [ZB MED]), the Leibniz Institute for the Social Sciences (Leibniz-Institut für Sozialwissenschaften [GESIS]) and the German National Library of Economics (Deutsche Zentralbibliothek für Wirtschaftswissenschaften [ZBW]) are also members of DataCite. The use of DOI allows for persistent referencing to digital content on the internet. This is a vital prerequisite not only for the reliable retrieval of specific content, but also of potential benefit to the reputation of the original data producer. Contents can be linked to the data creators in this way. In addition, the system supports interlinking academic publications and research data stored in individual repositories. At the same time, persistent identification enables the development and progress of innovative publishing activities for research data[66] and has in the meantime led to new cooperations between research infrastructure organizations and publishers.[67]

Such initiatives are examples of how "traditional" information infrastructure institutions could integrate themselves into the discussion about the improved handling of research data. Institutions above and beyond research libraries should also be involved in this discussion because the long-term preservation of research data carries diverse challenges that cannot be solved by individual actors. The discussion about preservation periods, for example, is not about a set, clearly-defined period of time, but in fact should be guided by the need to archive research data for the foreseeable future.[68] Concerning research data and the decision which data should be preserved and how, general questions must be considered about

65 See DataCite (2011), http://datacite.org.

66 See, for example, Earth System Science Data (ESSD).

67 See, for example, that when using DOI, there is alternation between Elsevier and the information system PANGAEA and between data sets at PANGAEA and digitally available publications at Elsevier.

68 Stefan Luther from the German Pension Fund (Rentenversicherung) referred to, for example, in his presentation in 2010 at the 11th Oracle Bibliotheken Summit in Weimar about the "Interaction between High-Volume Archives and Storage Platforms" („Zusammenspiel von hochvolumigen Archiven und Storageplattformen") that the problem is not the data volume in itself, but the deletion of data. In this concrete case, the unclear lifecycle of the data was the setting.

long-term data curation and, linked to that, sustainable long-term availability. In this domain, the archives with their subject expertise should be more involved since there is a need to develop appropriate measures and standards which comply with research requirements. Several state and university archives are already interested in addressing this challenge. In addition, the need to professionally preserve, archive, and provide research data in a professional way requires an IT environment which is matched with these tasks. At this point, the expertise from data centers could doubtlessly be incorporated. The primary task of data centers is in ensuring the undisrupted operation of the IT infrastructure of an organization such as a university. However, the tasks of data centers are changing. Increasingly, providing services for users, in addition to the purely technical maintenance of the IT infrastructure, plays a role.[69] It is still a topic for discussion as to which of these services could be offered by data centers concerning research data management. However, it seems to be obvious that the professional expertise of data centers is needed, for example, in dealing with high-throughput technologies in life sciences.

Research collections and museums with research departments have only recently begun to participate in the discussion about the provision of comprehensive digital information. The installation and the systematic use of digital systems in such institutions, above and beyond the strictly management and inventory duties of collections, are still in the early stages. On an international level, the need for a coordinated approach has been a topic of discussion among major natural science museums within the framework of the Scientific Collection International (SciColl) Initiative since 2006.[70] At the national level, this topic has already been addressed, and, for example, the DFG's available funding options, which previously focused on manuscript and printed documents, were expanded in 2011 with the announcement of a call for pilot projects dealing with the "classification and digitization of object-based research collections" („Erschließung und Digitalisierung von objektbezogenen wissenschaftlichen Samm-

69 See, for example, the description of tasks of the computing center at the University of Stuttgart: http://www.hlrs.de/.

70 See Scientific Collections International (SciColl), http://scicoll.org.

lungen“).[71] The intention of this broad call for applications was to investigate the need for and interest in this field, and to gain experience for the development of future funding measures. In addition to the specific technical and organizational challenges in the field of developing and digitization of objects, with the backup, archiving, and retrieval of digital information in this context, information specific technical requirements arise, that also play a role in research data. Therefore, all measures should be accompanied by the appropriate relevant professional expertise in the field of information management. Accordingly, in the recommendations of the German Council of Science and Humanities (“Wissenschaftsrat”) from January 28, 2011, it is stated that these measures require a high degree of initiative and self-organization on the part of the collections and their host institutions.[72]

As the basis of most research results, research data are an integral part of scholarly publications. Therefore, creating connections between scholarly articles in digital form and their underlying research data is not only useful but is virtually obligatory. This interlinking is made possible by the assignment of DOI’s and in some cases has already been implemented.[73] There is still no standard process for creating this link, however. Similarly, there is no clear agreement regarding the corresponding responsibilities. In this area, it is obvious that academic publishers could play an important role and should have the responsibility for actively supporting this interlinking. As a prerequisite for independent research and for the best possible access to research results, a free and unrestricted access to relevant data – linked via research publications or research data repositories – needs to be guaranteed.

71 See DFG (2010a).

72 See Wissenschaftsrat (2011a).

73 See, for example, ESSD (2011).

2.4 Funding Organizations

As major sources of financial support for research, funding organizations play a prominent role in this general discussion. The requirement that funding recipients deal in a professional manner with research data and participate actively in measures to develop suitable systems must be accompanied directly by the funding necessary for doing so. Funding agencies have already recognized, as far as possible, the need for action and their corresponding duties. However, there is no consensus about which specific and long-term supporting measures should be taken since there are numerous, still unresolved questions about the details. It is imperative to first develop basic frameworks, particularly against the background of not insignificant financial needs, of which the concrete amount is difficult to ascertain at present, and the overlap of institutional funding independent of third-party sources. The actual needs of the individual academic disciplines and interdisciplinary fields take precedence here, as do the accompanying responsibilities and the implementation of the widest possible acceptance of a sustainable handling of research data and the associated changes to scholarly research processes. These profound changes in the culture of research make the participation of researchers at an early stage in this process mandatory.

As part of its funding priorities up to 2015, the DFG has announced a call in 2010 for "information infrastructures for research data"[74] for the development and optimization of information infrastructures, trying to achieve the efficient and effective handling of research data.[75] The aim of this announcement was to encourage new measures for discipline-specific forms of organizations and to provide options for existing research data repositories to expand their services or to professionalize them. As a result of this call, projects in which information infrastructure institutions are cooperating closely with researchers emerged from a variety of academic disciplines. At the same time, since April 2012 the application procedure for research projects has required applicants to include a plan for handling

74 See DFG (2010b).

75 See DFG (2006).

research data in their research project proposal. The application guidelines state:

> If data measurements that are suitable for reuse are to be collected systematically using project resources, please explain what measures have been taken, or will be taken, during the duration of the project, to secure the data sustainably and to make them available for possible reuse. Please also consider, if possible, the existing standards in your subject discipline and the options available from data repositories.[76]

This requirement is intended to support the result that applicants – if necessary – will formulate concrete plans for the handling of digital research findings in projects funded by the DFG, and possibly align themselves with appropriate partners from the field of information infrastructures. This approach also creates the possibility for existing research data repositories to adapt or expand their services specifically in accordance with the requirements of the academic community and, in doing so, achieve greater acceptance and sustainability. With a similar objective in the context of funding Priority Programs (Sonderforschungsbereiche [SFB]), in 2007 it became possible to apply for funding for central subprojects that deal, in cooperation with appropriate infrastructure institutions, with data management and the sustainable availability of research data compiled in the SFB.[77] Currently, more than twenty Priority Programs have made use of this option. A similar adjustment in the application process took place in the second round of the German Excellence Initiative[78], in which the application form for excellence clusters was

76 DFG (2010c), p. 32: „Wenn aus Projektmitteln systematisch (Mess-) Daten erhoben werden, die für die Nachnutzung geeignet sind, legen Sie bitte dar, welche Maßnahmen ergriffen wurden bzw. während der Laufzeit des Projektes getroffen werden, um die Daten nachhaltig zu sichern und ggf. für eine erneute Nutzung bereit zu stellen. Bitte berücksichtigen Sie dabei auch – sofern vorhanden – die in Ihrer Fachdisziplin existierenden Standards und die Angebote bestehender Datenrepositorien."

77 See DFG (2009a).

78 The German Excellence Initiative aims to support top-class research and to advance the quality of universities and research institutions in Germany. The Excellence Initiative was initiated by the German federal and state governments and the DFG

amended with the addition of the requirement to provide information about data management. This addition was designed to ensure that the handling of data should be covered as a central issue and appropriate measures considered at an early stage.

The German Federal Ministry for Education and Research is also pursuing the goal of ensuring the future (re)usability of project results with the requirement for funding applications to provide information in the context of a "plan for realization." In addition to a description of the scholarly, economic and technical prospects for success, it is about the project's adaptability in terms of the sustainable use of project results in subsequent academic projects, in applied research, or in the context of commercial applications.[79] For the preparation of projects, the BMBF provides an overview of subject information centers and nationwide information centers, which generally differ from research data repositories.[80] At present, a description of essential measures for securing, archiving, and re-using research data and other digital research outcomes is not included. This approach is strongly influenced by different strategies in the individual research areas and it is exclusively project-orientated.

With the D-GRID initiative[81], the BMBF has supported a national IT infrastructure since 2005 that has the goal of providing a high-performance computing and storage structure for academic research as well as for industry. Numerous individual projects within the D-GRID initiative address scholarly objectives as well as operational and commercial applications in addition to projects to guarantee the operation of the grid on different levels. D-Grid is characterized by intensive cooperation between industrial partners and academic and research institutions. In the scholarly context, a number of disciplines created specific grid-based initiatives that

---

was commissioned to run the initiative in cooperation with the German Council of Science and Humanities (Wissenschaftsrat).

79 See BMBF (2011a).

80 See BMBF (2011b).

81 The D-Grid Initiative (German Grid Initiative) was a federally funded project with the objective to develop computer infrastructure for research and education. Using and implementing grid computing technology the initiative started in 2005 with seven projects including an integrational project and several partner projects.

addressed the characteristic requirements for data processing and made available corresponding options. However, the focus has often been placed on applications and services that are comparable to virtual research environments. The subjects of sustained data and information management, data preservation and the support for and provision of research data have only been emphasized in a few projects up to this point (such as AstroGrid, C3Grid, TextGrid, and WissGrid).[82]

At present [2012; Ed.], there is no systematic financial support from other funding agencies, foundations, and other donors for projects dealing with the management of research data.

82 See D-GRID, http://www.d-grid-gmbh.de.