



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

Digital Curation of Research Data

Experiences of a Baseline Study in Germany

Chapter 3 Status of Discussion and Current Activities: The International Perspective

**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>

Die digitale Version dieses Werkes ist unter Creative Commons Namensnennung 3.0 lizenziert / The digital version of this work is licensed under a Creative Commons Attribution 3.0 Unported License <http://creativecommons.org/licenses/by/3.0/deed.en>

CC - BY 

Einfache Nutzungsrechte liegen beim Verlag Werner Hülsbusch, Glückstadt. The Verlag Werner Hülsbusch, Glückstadt, owns rights of use for the printed version of this work.

vwh Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

© Verlag Werner Hülsbusch, Glückstadt, 2013 · <http://www.vwh-verlag.de>

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

Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen, Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und als solche den gesetzlichen Bestimmungen unterliegen.

All trademarks used in this work are the property of their respective owners.

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

Content

Foreword	7
<i>Heike Neuroth, Stefan Strathmann, Achim Oßwald, Jens Ludwig</i>	
1 Digital Curation of Research Data: An Introduction	9
<i>Achim Oßwald, Heike Neuroth, Regine Scheffel</i>	
2 Status of Discussion and Current Activities: National Developments	18
<i>Stefan Winkler-Nees</i>	
2.1 Research Organizations	19
2.2 Recommendations and Policies	22
2.3 Information Infrastructure Institutions	28
2.4 Funding Organizations	33
3 Status of Discussion and Current Activities: The International Perspective	37
<i>Stefan Strathmann</i>	
3.1 International Organizations	37
3.1.1 United Nations Educational, Scientific and Cultural Organization (UNESCO)	38
3.1.2 Organisation for Economic Co-Operation and Development (OECD)	38
3.1.3 European Union (EU)	40
3.1.4 World Health Organization (WHO)	41
3.1.5 Knowledge Exchange	41
3.2 Model Realizations	42
3.2.1 National Science Foundation (NSF)	42
3.2.2 Australian National Data Service (ANDS)	43
4 Methodology: Subject of the Study	46
<i>Heike Neuroth</i>	
4.1 Structure of this Volume	47
4.2 Key questions for mapping research disciplines	48

4.3	Introduction to the Research Area	48
4.3.1	Background	49
4.3.2	Cooperative Structures	49
4.3.3	Data and Metadata	49
4.3.4	Internal Organization	51
4.3.5	Perspectives and Visions	52
5	Summary and Interpretation	54
	<i>Jens Ludwig</i>	
5.1	Cooperative Structures	55
5.2	Data and Metadata	58
5.3	Internal organization	65
5.4	Perspectives and Visions	67
6	Implications and Recommendations on Research Data Curation	69
	<i>Heike Neuroth, Achim Oßwald, Uwe Schwiegelshohn</i>	
	References	79
	Abbrevations	87
	Directory of Authors	91

3 Status of Discussion and Current Activities: The International Perspective

Stefan Strathmann

German issues of research data management are essentially identical to international issues in this field: It is often organizational adjustments to the structural framework that are lacking in order to address the changing requirements of a research which is based on the extensive use of information technology.⁸³ As in Germany, it is primarily grassroots projects and a few flagships that are actively addressing the challenges of sustainable research data management. However, there are some organizations and institutions that have been trying to guarantee long-term archiving and availability for several years. The organizations and institutions discussed briefly below represent only a selection of those working in this area.⁸⁴

3.1 International Organizations

The choice of institutions and organizations presented here takes account of the wide variety of different approaches concerning the digital curation of research data. Therefore, the selected institutions are to be seen representatively for a larger group of similar acting facilities.

83 See Chapter 2.1.

84 In addition to the institutionally-based activities and discussions outlined in this chapter, there are naturally international discussions taking place. In addition to the community-based discussions, the e-mail list Research-Dataman is particularly worthy of mention (<https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=RESEARCHDATAMAN>). It was initiated by the Digital Curation Center (DCC) in Britain on behalf of the Joint Information Systems Committee (JISC). <http://www.jisc.ac.uk>.

3.1.1 United Nations Educational, Scientific and Cultural Organization (UNESCO)

The “Charter on the Preservation of Digital Heritage,”⁸⁵ which was adopted on October 17, 2003, at the 32th UNESCO General Conference,⁸⁶ also includes scholarly research data:

The digital heritage consists of unique resources of human knowledge and expression. It embraces cultural, educational, scientific and administrative resources, as well as technical, legal, medical and other kinds of information created digitally, or converted into digital form from existing analogue resources. Where resources are “born digital,” there is no other format but the digital object.⁸⁷

Digital research data are considered as part of the digital cultural heritage and the UNESCO member states are directed to preserve this heritage in order to “[...] ensure that it remains accessible to the public.”⁸⁸ In reference to research data, the charter states that “Measures should be taken to [...] encourage universities and other research organizations, both public and private, to ensure preservation of research data.”⁸⁹ With this charter, the United Nations called attention at an early stage to the necessity of comprehensive measures for the preservation of the cultural and scholarly heritage and committed member states to conserving this heritage.

3.1.2 Organisation for Economic Co-Operation and Development (OECD)

Considering the enormous costs involved in the creation of research data, research data management is a topic for the OECD as well.⁹⁰ As early as 2004 archiving and providing access to publicly funded research data were

85 UNESCO (2003).

86 See UNESCO Homepage: <http://www.unesco.org>.

87 UNESCO (2003).

88 UNESCO (2003).

89 UNESCO (2003).

90 See Organisation for Economic Co-Operation and Development Homepage: <http://www.oecd.org>.

the subject of the final document⁹¹ of the “OECD Committee for Scientific and Technological Policy at Ministerial Level.” The research ministers came to the following conclusion:

Co-ordinated efforts at national and international levels are needed to broaden access to data from publicly funded research and contribute to the advancement of scientific research and innovation. To this effect, Ministers adopted a Declaration entrusting the OECD to work towards commonly agreed Principles and Guidelines on Access to Research Data from Public Funding.⁹²

The access to publicly funded research data is also the subject of an annex (Annex I: “Declaration on Access to Research Data from Public Funding”) to this document. On the basis of these ministerial decisions, recommendations for managing publicly funded research data were developed.⁹³ These “Principles and Guidelines for Access to Research Data from Public Funding” were adopted and published by the OECD. OECD recommendations are not legally binding, but “are considered to have a great moral force” nevertheless.⁹⁴ The final section is dedicated to sustainability:

Due consideration should be given to the sustainability of access to publicly funded research data as a key element of the research infrastructure. This means taking administrative responsibility for the measures to guarantee permanent access to data that have been determined to require long-term retention. This can be a difficult task, given that most research projects, and the public funding provided, have a limited duration, whereas ensuring access to the data produced is a long-term undertaking. Research funding agencies and research institutions, therefore, should consider the long-term preservation of data at the outset of each new project, and in particular, determine the most appropriate archival facilities for the data.⁹⁵

91 See Organisation for Economic Co-Operation and Development (2004).

92 See Organisation for Economic Co-Operation and Development (2004).

93 See Organisation for Economic Co-Operation and Development (2007).

94 This type of instrument is often referred to as “soft law”; see Organisation for Economic Co-Operation and Development (2007), p. 7.

95 Organisation for Economic Co-Operation and Development (2007), p. 22.

It is becoming more and more evident that especially the recommendation to prepare for research data curation from the very beginning of data creation – both on the part of the researchers and of those providing research funding – is one key to successful research data management.

3.1.3 European Union (EU)

In the research framework programs of the EU, the issue of research data infrastructures has been of special interest for several years now. Just as projects for the preservation of cultural objects, projects for the digital curation of research data have been, and continue to be, promoted. Examples include the completed project PARSE.Insight⁹⁶ or the ongoing project APARSEN⁹⁷, both of which were initiated by the “Alliance for Permanent Access.”⁹⁸ Currently, the development of research infrastructures, which includes data infrastructures, is one of the main priorities of the EU Seventh Framework Programme (FP7).⁹⁹ That is partly due to the implementation of the guidelines of the European Strategy Forum on Research Infrastructures (ESFRI).¹⁰⁰ The forum is a strategic instrument for the development of research infrastructures in Europe. The integration and strategically defined development of policies for research infrastructures are the main focus of this group, whose members are nominated by the research ministers of the member states.

A vision of an EU strategy for dealing with research data can be found in the report “Riding the Wave: How Europe can gain from the rising tide of scientific data” from the High-Level Group on Scientific Data.¹⁰¹ A central component of the vision outlined in this report is a collaborative

96 See PARSE.Insight Homepage: <http://www.parse-insight.eu>.

97 See European Commission, http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri.

98 See Alliance for Permanent Access Homepage: <http://www.alliancepermanent-access.org>.

99 See European Commission (2011).

100 See European Commission, http://cordis.europa.eu/fp7/capacities/research-infrastructures_en.html.

101 See High Level Expert Group on Scientific Data (2010).

data infrastructure in which data represent the actual infrastructure and the physical and technical infrastructure is in the background.

In addition the report of the “Comité des Sages”¹⁰² of January 2011 draws a direct connection between increasing digitization and the corresponding need for research data curation.

3.1.4 World Health Organization (WHO)

With the participation of the World Health Organization (WHO)¹⁰³ and the Wellcome Trust,¹⁰⁴ a number of organizations supporting proposals in the field of health care have agreed on common funding goals, visions, and principles in recent years and in early 2011 published a code of conduct in a joint statement.¹⁰⁵ In particular, this statement describes important requirements concerning data management and access to data created by means of this funding. The long-term goals include the following:

- Data collected for health research are made available to the scientific community for analysis which adds value to existing knowledge and which leads to improvements in health [...]
- To the extent possible, datasets underpinning research papers in peer-reviewed journals are archived and made available to other researchers in a clear and transparent manner¹⁰⁶

The long-term availability of research data in the medical sector is thus a clearly stated goal of the funding organizations involved.

3.1.5 Knowledge Exchange

Knowledge Exchange¹⁰⁷ is a collaborative consortium of four European funding institutions to achieve their funding goals more efficiently through

102 See Niggemann; De Decker; Levy (2011).

103 See World Health Organization (WHO), <http://www.who.int>.

104 See Wellcome Trust Homepage: <http://www.welcome.ac.uk>.

105 See Wellcome Trust (2011b).

106 See Wellcome Trust (2011a).

107 See Knowledge Exchange Homepage: <http://www.knowledge-exchange.info>.

the coordinated use of resources. The partners are Denmark's Electronic Research Library (DEFF),¹⁰⁸ the German Research Foundation (Deutsche Forschungsgemeinschaft [DFG]),¹⁰⁹ the British Joint Information Systems Committee (JISC),¹¹⁰ and the SURFfoundation¹¹¹ in the Netherlands.

Stable and long-term access to research data is one focus of their joint activities.¹¹² The website of the cooperation states:

In the future of academic and scholarly communication compound publications in the means of article and research data will play an ever increasing role. Research data have to be accessible both as open access and in the long term but also in different environments and tools.¹¹³

With the requirement that research data must not only be openly available in the long term, but also made accessible for various environments and with different tools, Knowledge Exchange continues the discussion about a life cycle model in research data curation.¹¹⁴

3.2 Model Realizations

3.2.1 National Science Foundation (NSF)

Since the beginning of 2011, the NSF¹¹⁵ in the United States has extensively revised the regulations for submitting a funding proposal.

Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labelled "Data Man-

108 See Denmark's Electronic Research Library (DEFF), <http://www.deff.dk/english/>.

109 See Deutsche Forschungsgemeinschaft (DFG); Chapter 2.1.

110 See Joint Information Systems Committee (JISC), <http://www.jisc.ac.uk>.

111 See SURFfoundation, <http://www.surffoundation.nl>.

112 Among other things, Knowledge Exchange established a separate working group for research data (see Knowledge Exchange (2011b)).

113 Knowledge Exchange (2011a).

114 See, e. g., Digital Curation Centre (2011a) and Australian National Data Service (ANDS): <http://ands.org.au/about-ands.html>.

115 See NSF Homepage: <http://www.nsf.gov>.

agement Plan”. This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results.¹¹⁶

All applications for funding now require a “data management plan (DMP),” in which the data that will be produced during the funding period must be described and which explains the way in which data will be handled.¹¹⁷ This includes the publication as well as the archiving of data:

The DMP should describe data formats, media, and dissemination approaches that will be used to make data and metadata available to others. Policies for public access and sharing should be described, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements. [...]

The DMP should describe physical and cyber resources and facilities that will be used for the effective preservation and storage of research data. These can include third party facilities and repositories.¹¹⁸

A proposed plan for ensuring sustainability beyond the duration of the project is emphasized by information about the plans for research data curation as well as by the inclusion of a report in all subsequent proposals and applications regarding the obligation to maintain and manage the data inherited from previously funded projects.¹¹⁹

3.2.2 Australian National Data Service (ANDS)

The ANDS¹²⁰ is establishing an Australian Research Data Commons (ARDC).¹²¹ This commons will include research data from all disciplines,

116 NSF Data Management Plan (2011).

117 This requirement applies generally to all applications. The specific requirements for each directorate, funding line, and so on, are regulated individually; see NSF Data Management for NSF SBE Directorate Proposals and Awards (2010).

118 NSF (2010), p. 3.

119 Specifically, this means that “data management must be reported in subsequent proposals by the PI and Co-PIs under ‘Results of prior NSF support’.”; see NSF (2010), p. 4.

120 See ANDS Homepage: <http://www.ands.org.au>.

121 See Australian National Data Service (2007).

all universities and all research institutions that are supported by public funding in Australia. Australian research data are to be transformed as a whole into a strategic national infrastructure. To achieve these goals, the ANDS is meeting various organizational and technical challenges and promoting the development of an infrastructure. Its objectives include the following:

- Support for research data managers
- Promotion of the transfer of research data in stable and accessible research data curation environments
- Provision of training opportunities in the field of data management that are independent of institutions or communities.
- The ability of researchers to access the Australian data commons and to work with it.
- Support for the integration of Australian research data in international as well as national and multi-disciplinary research groups.¹²²

For a practical implementation of these goals, a number of infrastructure components are currently being created:

- Data Capture Infrastructure (especially the integration of existing infrastructure)
- Research Metadata Store Infrastructure (the goal is a combination of “data stores” and “metadata stores”)
- Automatic Public Data Publication Infrastructure (for the publication of descriptions of data collections e. g. from public authorities [meteorology, statistics, etc.] or other providers of research data aggregations [libraries, museums, etc.]).
- Australian Research Data Commons Core Infrastructure (persistent identification, authority files, controlled vocabulary, retrieval options, etc.).
- Australian Research Data Commons Applications Infrastructure (possibilities for data integration, data visualization, and data analysis).¹²³

122 See Australian National Data Service: <http://ands.org.au/about-ands.html>.

123 See Australian National Data Service: <http://ands.org.au/ardc.html>.

The current budget (2012/2013 financial year) for these activities from the National Collaborative Research Infrastructure Strategy (NCRIS) and the Education Investment Fund (EIF) sums up to \$71,037,566.45¹²⁴.

As part of the efforts to build this comprehensive research data infrastructure, some of the functions of the Australian Research Data Commons (ARDC) have already been implemented and are accessible via the ANDS website.¹²⁵ For instance, it is already possible to register research data with accompanying metadata and persistent identifiers, and to make data accessible. The actual storage and archiving is the responsibility of the institutions and researchers who provide the data. The ANDS serves as a “metadata store,” meaning that it does not provide a way to archive research data. However, it does provide a search function and access to the extensive data resources which have already been registered. These options are supplemented by a series of helpful guides to best practices in research data management, research data curation, and the legal implications for the publication of research data. In addition, there are a variety of information and training sessions that familiarize (potential) users with the services of the ARDC and explain the basics of research data management. Such an ambitious national approach is just one way to ensure long-term access to research data. In addition, there are several other approaches, by far more common, to ensure the digital curation of research data. They may be based on individual projects, on national or international efforts or on community-specific or institutional activities.

124 See ANDS (2012–13), p. 14.

125 See ANDS Homepage: <http://www.ands.org.au>.