WORLD LIBRARY
AND INFORMATION
CONGRESS:
78TH IFLA GENERAL
CONFERENCE
AND ASSEMBLY

IFLA

2012 Helsinki

LIBRARIES NOW!
INSPIRING
SURPRISING
EMPOWERING

The Role of Libraries in Curation and Preservation of Research Data in Germany: Findings of a survey

Achim Osswald
Professor
Cologne University of
Applied Sciences
Institute of Information Science
50968 Cologne, Germany
E-mail: achim.osswald@fh-koeln.de

Stefan Strathmann
Head, Competence Center on
Digital Preservation and Research Data
Goettingen State and
University Library
37073 Goettingen, Germany
E-mail: strathmann@sub.uni-goettingen.de

Abstract:

In the last few years the issue of curation of research data has become a topic of enhanced interest in scientific communities. But there is no clear understanding of how to deal with curation of research data. Regarding the situation in Germany, there was no clear picture of the methods that different academic disciplines use to preserve and curate their research data. Even more there was no clear understanding and consensus which role libraries could or should play in this issue. A baseline study was conducted in cooperation with nestor and the D-Grid GmbH in Germany. The study gives more stable data to scientists, service infrastructure experts and politicians to foster strategic concepts for digital curation and preservation in and between the disciplines. The survey addressed eleven disciplines including the humanities, social sciences, psycholinguistics, pedagogics, classical studies, geoscience, climate research, biodiversity, particle physics, astronomy and medicine. Results can be seen as more or less representative for the situation in Germany.

Findings show that libraries and data centers are involved but there is no clear result on the role libraries should or will play in the field. Libraries which are in close connection to scientists have an advantage in being addressed as institutions responsible for digital curation and preservation.

Librarian's competences regarding research data still have to be proofed. Nevertheless individual scientists and research data centers might need librarian´s expertise regarding the application of metadata standards and the provision of services in connection with the research data curated by institutions. Qualification and training regarding curation of research data is still at its beginning and should be addressed by programs of universities.

INTRODUCTION

In recent years, researchers, research organizations and political task forces have emphasized that the growing relevance of research data for research and development will influence not only processes dealing with this data but also the qualifications needed to fulfill these tasks (see e.g. ARL 2006; NSB 2005; RIN 2008; OECD 2007 or – on behalf of the European Commission – the High Level Expert Group on Scientific Data 2010). The general description of competences in collecting, describing and processing research data has resulted in some recommendations for tasks that libraries and librarians should take on (e.g. Donnelly 2008; Gold 2007a; Gold 2007b; Gold 2010; Jones 2008; Ruempel 2010; Swan & Brown 2008). Until now, discussions about these suggestions have not provoked a broad movement to new or revised LIS curricula. It seems that this applies to Germany as well as to most English-speaking countries – where some minor exceptions prove the rule (Pampel, Bertelmann & Hobohm 2010; Piorun et al. 2012).

When specifying those qualifications relevant to the curation of research data, the difference between generic and domain-related competences and qualifications should be kept in mind. Until now, mostly generic qualifications related to dealing with and the curation of research data have been assessed and described. On a basic level the curation and preservation of research data are links of a chain of generic activities that are influenced by domain-related aspects like type and volume of data, formats, etc. All this has become a topic of research and (case) studies challenging the field of data curation.

A baseline study which assessed the current status of activities in the field of digital curation and digital preservation of research data in eleven research-related disciplines in Germany may be an example for such kind of studies. The study was conducted during the first half of 2011.[1] It was guided by a research group from nestor[2], the German network of competence for digital preservation, in cooperation with the D-Grid GmbH, a non-profit limited liability company founded by the German Ministry of Education and Research in 2008.[3] With the support of scientists in the addressed disciplines, the study covers eleven disciplines including the humanities, social sciences, psycholinguistics, pedagogics, classical studies, geoscience, climate research, biodiversity, particle physics, astronomy and medicine. These disciplines have been selected because nearly every type of research data is relevant to these fields (see for comparison NSB 2005).

---

1 Editors of the study are Heike Neuroth, Stefan Strathmann, Achim Osswald, Regine Scheffel, Jens Klump and Jens Ludwig; see Neuroth et al. 2012.

2 In Germany libraries, archives, museums and leading experts work together in *nestor* to ensure the long-term preservation and accessibility of digital sources. *nestor* is a cooperation association including partners from different fields, but all connected in some way with the subject of "digital preservation". See http://www.langzeitarchivierung.de/eng/index.htm

3 D-Grid has the goal to ensure efficient collaboration and cooperation between different projects in the field of a sustainable grid infrastructure in Germany. See http://www.d-grid-gmbh.de/index.php?id=1&L=1

Results of the survey have been published (Neuroth et al. 2012) reporting on the state of affairs in these eleven disciplines. They were summarized and suggestions were provided for action in fields such as politics, education or continuing education for librarians and services to be offered by libraries.

RESULTS OF THE NESTOR BASELINE STUDY ON DIGITAL CURATION AND PRESERVATION OF RESEARCH DATA IN GERMANY

The baseline study gives stable data to scientists, service infrastructure experts and politicians to foster strategic concepts for digital curation and preservation in and among the disciplines. As a side effect it exemplifies processes and qualification needs related to these activities.

Some relevant questions for librarians and the LIS community are:

- What types of cooperative structures do already exist and which role do libraries play in these cooperations?

- Are there any generic concepts and role models for libraries and librarians in the field of data curation?

- Which qualifications and skills of librarians are needed to enhance the quality of data curation activities?

Results of the baseline study show that at the very least:

- The relevance of research data is emphasized by experts of all disciplines involved.
- Data sharing and the investment of time and resources in processing research data has not been common sense until now. Awareness about the relevance of data sharing in science, society and infrastructural institutions is a precondition for further development.
- Data management has domain-related and generic aspects. Cooperation between the groups involved will improve the effectiveness and efficiency of data management.
- Frequently infrastructural institutions like libraries and computer centers are cooperating partners in such activities. However, until now the role of both has not yet been clearly defined.
- An important part of data management is the back-up and storage of research data. While the technical part of this procedure is mostly a well-organized service carried out by computer centers, metadata-based description of the processes and the data saved has not been a priority until now. Nevertheless, both activities are a precondition for data sharing and use of data by third parties.
- The application of measures for data curation (including those already mentioned) by the institutions involved will reduce technical and organizational investment in the long run.
- Metadata and persistent identifiers are important for referencing and citing research data. To realize and support these tasks domain-related knowledge as well as generic competences of LIS specialists are needed.

A STATE-OF-THE-ART REPORT OR
HOW DO THESE RESULTS MATCH RESEARCH RESULTS OF OTHER COUNTRIES?

To get a broader view on the topic the findings of the German study (Neuroth et al. 2012) should be compared with findings of selected studies. This will give an international picture of the topic. In addition it will help to exemplify which tasks related to research data libraries and librarians could or should fulfill in the future and how librarians should be qualified to enable this.

In 2008 Swan & Brown (Swan & Brown 2008) prepared a report to JISC on "The Skills, role and career structure of data scientists and curators: An assessment of current practice and future needs". Gold (Gold 2010) published

a background paper for a talk presented to the American Geophysical Union 2009 on "Data & Libraries" which provides a comprehensive state-of-the-art report on the topic from a US perspective. Rümpel and her colleagues (Rümpel 2010; Büttner, Rümpel & Hobohm 2011) adopted and transferred the results of the international discussion to the German situation. Their conclusions and suggestions refer to the concept of data management and the qualification tracks universities provide to information scientists. Complementary, updated information about developments in the US and UK can be taken from Digital Curation Exchange[4] and the preliminary results of the project DigCurV (Digital Curator Vocational Education Europe)[5], a project funded by the European Commission's Leonardo da Vinci program to establish a curriculum framework for vocational training in digital curation. An overview on the situation in four European countries (Denmark, Germany, the Netherlands and the UK) was provided by the Knowledge Exchange project (van der Graaf & Waaijers 2011).

The results reported by the nestor baseline survey in Germany correspond to those of the European surveys of DigCurV and Knowledge Exchange. The nestor survey like the Knowledge Exchange report gives a clear indication that an increasing demand for activities in curation of research data will arise from research projects and results by the scientific community. Researchers and funding agencies are becoming aware of the opportunities that saving and curation of research data will have for them, their institutions and science in general in the future.

WHAT TYPES OF COOPERATIVE STRUCTURES DO ALREADY EXIST AND WHICH ROLE DO LIBRARIES PLAY IN THESE COOPERATIONS?

Cooperation is a driving principle of science. This applies not only in general to the eleven disciplines in the focus in the German survey but also to digital curation and preservation activities in these disciplines. Based on the division of labor in most cases curation and preservation activities are assigned to specialized institutions like domain-related or subject-specific research and information centers or World Data Centers. In some cases the task is assigned to special departments of research institutions which have an LIS focus and LIS-skilled staff members in this field. Others realize a federated cooperation of specialized teams/persons within research groups. Their activities are carried out with or without agreed codes of practice (q.v. Swan & Brown 2008).

In Germany funding agencies and science organizations have promoted the cooperation of libraries and computing centers in data management. Nevertheless, this role is neither clear nor well established. Until now, the role of libraries seems to be less accepted than the role of computing centers – perhaps because saving and long-term storage of data are the primary issue under discussion. Motivation for the curation of research data is still not as well perceived as it should be in general and from a strategic point of view. As a consequence libraries and librarians might expose their competence in this field by cooperating with researchers and research departments. This should be accompanied by conceptual discussions regarding the role and competences of LIS specialists.

A political initiative stimulated by a funded action program might result from the Knowledge Exchange cooperation. The report "A Surfboard for Riding the Wave" (van der Graaf & Waaijers 2011) gives some outlines for such an initiative.

---

4 "A space for all things 'digital curation'"; see http://digitalcurationexchange.org

5 Digital Curator Vocational Education Europe; DigCurV, a project funded by the European Commission's Leonardo da Vinci program to establish a curriculum framework for vocational training in digital curation "… brings together a network of partners to address the availability of vocational training for digital curators in the library, archive, museum and cultural heritage sectors needed to develop new skills that are essential for the long-term management of digital collections" (http://www.digcur-education.org/).

ARE THERE ANY GENERIC CONCEPTS AND ROLE MODELS FOR LIBRARIES AND LIBRARIANS IN THE FIELD OF DATA CURATION?

With regard to this question Corall (2008) suggested a model in the context of "e-content and digital library specialists" describing roles of content specialists (LIS-based), context specialists (discipline- or domain-based) and conduit specialists (IT- & media-based).[6] She emphasized the "boundary-spanning roles" (ibid.) of the responsible persons.
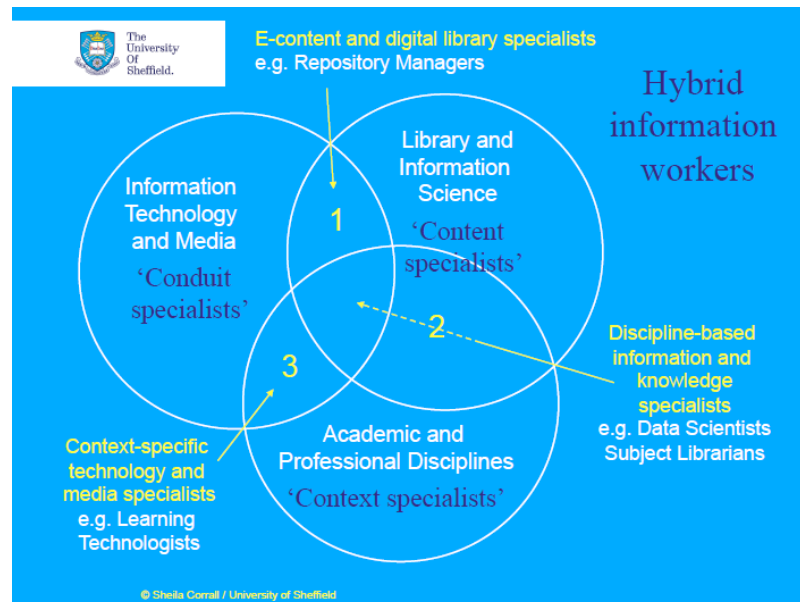

Figure 1: Profiles of hybrid information workers (Corall 2008; slide 6)

Nevertheless, there is no clear consensus on the job profiles emerging from these new activities. Until now, research data-related skills have been necessary for persons acting as data creators (mostly scientists), data scientists (persons focusing on the analysis of data), data managers (persons responsible for organizing, securing of and access to data) and data librarians (persons collecting and curating data). It will depend on the conditions and the framework of projects and activities if these activities are performed by one or up to four different persons. From a practioner's point of view it might seem a little bit sophisticated to differentiate these roles and to expect them in reality. Nevertheless, it can be helpful to differentiate and exemplify them theoretically helping order to help to define the skills and responsibilities relevant within the whole environment of curation of research data (see e.g. Buettner, Ruempel & Hobohm 2011).

These four roles in data management have been visualized and specified by Donnelly (Donnelly 2008) referring to the core skills needed (see figure 2). The roles (to be) taken by librarians or other specialists will depend on the personal profile of these persons, their interest, the policy of the library or the research institution, the openness of scientists and researchers and, finally, the allocation of resources. And it is still open whether there will be a generic solution to the question at all.

---

6 This model has been adopted by Pampel, Bertelmann & Hobohm 2008, Rümpel 2010 as well as by Büttner, Rümpel & Hobohm 2011 in publications in German.
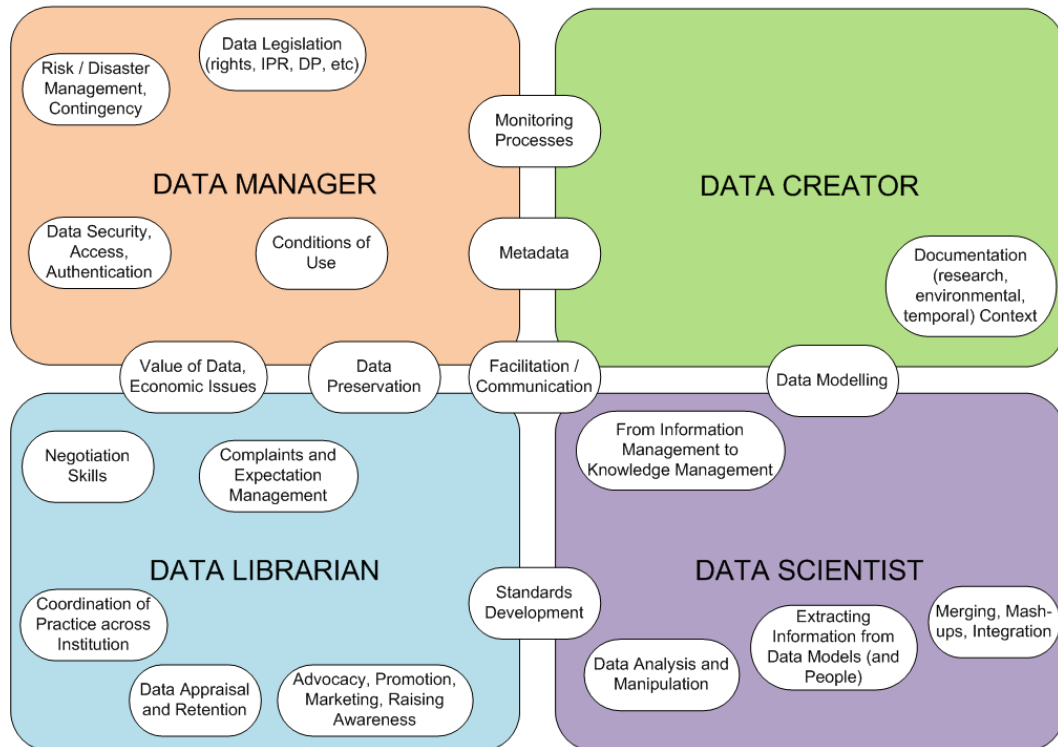
Figure 2: Core skills for data management (Donnelly 2008)

Swan & Brown (Swan & Brown 2008) take a practical approach to the topic by emphasizing that "there are generic data handling and management skills that are native to librarians and can be taught as part of the basic research skills training in an institution" (ibid; 24). They suggest three areas of action in which librarians could get influence on developments:

- "Training researchers to be more data-aware" (ibid.; 25).

- "Adopting a data archiving and preservation role" (ibid.; 25): Until now, librarians focus on downstream activities delivering information literacy programs to undergraduates. In the future they will face more demand by researchers and their interest to get support with and involvement in research activities and the curation of data generated in this context. This "offers libraries a strategic opportunity to reposition themselves with respect to research" (ibid.; 25).

- "The training and supply of data librarian" (ibid.; 25): While awareness and demand are rising skilled data librarians are missing.[7]

---

7 The Project "Digital Preservation Outreach & Education" (http://www.digitalpreservation.gov/education/reviewcurricula.html) conducted a curricula review in 2010 to check if the statement of Swan & Brown 2008 is still relevant: "Yet very few library and information science schools currently teach the skills that future data librarians will need. Of the 55 institutions that teach Information and Library Science (ILS) in the United States, for example, only a handful include any digital curation content in their courses." (ibid.; 25f)

WHICH QUALIFICATIONS AND SKILLS OF LIBRARIANS ARE NEEDED TO ENHANCE THE QUALITY OF DATA CURATION ACTIVITIES?

Skills associated with data curation activities are based on scenarios where librarians consult research groups and their activities, especially in data management. Those so-called "embedded librarians" are aware of the special discipline and science-related regulations and will be able to contribute to the projects that are involved. An important contribution will be the provision of interoperable metadata based on metadata standards and applying/enhancing these standards with regard to complex research projects and the generated data. Demonstration of data use cases and a pro-active role in data curation of data left as a legacy of projects could be included in the activities of libraries and librarians.

The European DigCurV (Digital Curator Vocational Education Europe) project recently provided the results of a survey (DigCurV 2012a; DigCurV 2012b) and is giving a broader view on the topic. DigCurV indicates that general knowledge about digital preservation issues is relevant to all levels of staff when providing a general introduction and explaining the key needs and challenges in this area. Beside this general knowledge the following skills and qualifications have been covered in training events and – as a consequence – might be most relevant to specialists involved or trying to get involved in that business (visualized in figure 3):[8]

- "Digital curation standards (metadata, OAIS etc.)"
- "Strategic planning (assist practitioners in identifying issues and goals necessary to plan and manage digital curation initiatives)"
- "Technical issues (assist in understanding and applying digital curation techniques)"
- "Legal aspects/Digital curation policies (assist in making legal decisions: copyright, freedom of information legislation, data protection requirements etc.)"
- "Digital curation and preservation tools (DAF, DRAMBORA, DMP, PLATO, PLATTER etc.)"
- "Digital repository audit and certification"
- Knowledge about "Trusted repositories (techniques and criteria for trusted digital repositories)" (all citations: DigCurV, 2012b; 13).

Further research is needed to compare these European findings with results from other continents.

---

8 Statements are presented in decreasing order of relevance. All quotes are extracted from DigCurV, 2012b; 13.
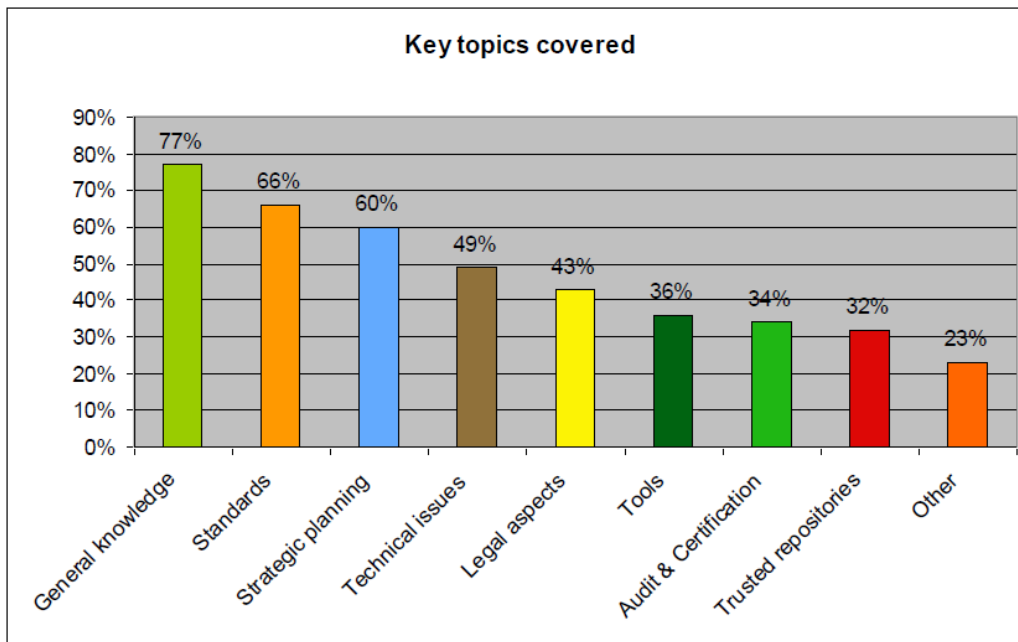
Figure 3: Key topics covered in training events (DigCurV, 2012b; 13)

CONCLUSION

The findings of the German survey on the curation and preservation of research data (Neuroth et al. 2012) reported here correspond with the resent research results provided by DigCurV and Knowledge Exchange:

There is a broad spectrum of tasks related to the handling and curation of research data which results in a demand for special qualifications to achieve them. Those qualifications are relevant e.g. in the field of metadata creation and management, format specifications and knowledge in connection with special formats, tools and techniques to process and deliver research data as well as procedures and processes to enable the sharing and usage of research data in future. Until now, these kinds of skills and services have rarely been offered by librarians because the curricula offered by most LIS programs did not cover the skills and the special kind of qualification needed for delivering such services. As a result every day practice in the various disciplines which produce and share data with others is – in some ways – accidental and relies on the personal interest of researchers and information specialists involved. Systematic and conceptual planning is necessary to change this situation by qualifying data librarians/data curators who will be able to generate processes and procedures based on a combination of generic and domain-related or subject-specific knowledge. Accordingly, scientists' discipline- or domain-specific awareness and competences of curation of research data should be fostered.

While there are tasks with an obvious connection to library skills a librarian role model related to research data or any best practice experiences have not been described until now in the eleven disciplines in focus. The role of libraries has not been defined either. Yet, there is a broad understanding that methodological expertise of libraries and librarians would be helpful to improve research data-related activities of scientists and research institutions.

It will take some more time to come to a clear understanding of the educational consequences that are to be drawn from these new emerging needs. Political factors on the national level are influencing the considerations whether undergraduate, postgraduate or continuing professional development programs might be the best way to serve the upcoming needs related to the curation of research data.

It is time to start a redesign of or realizing amendments to LIS curricula to open up the opportunity of extending the job and competence profile of LIS experts to digital curation and preservation of research data.

REFERENCES (checked at 4-29-2012)

ARL 2006

To Stand the Test of Time. Long-term Stewardship of Digital Data Sets in Science and Engineering. A Report to the National Science Foundation from the ARL Workshop on New Collaborative Relationships: The Role of Academic Libraries in the Digital Data Universe. Arlington, September 26-27, 2006

Büttner, Rümpel & Hobohm 2011

Büttner, S.; Rümpel, S.; Hobohm, H.-C. (2011). „Informationswissenschaftler im Forschungsdatenmanagement". In Büttner, S; Hobohm, H.-C.; Müller, L. (Ed.) (2011): Handbuch Forschungsdatenmanagement, 203-218. http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:kobv:525-opus-2405

Corall 2008

Corall, S. (2008). Research Data Management: Professional Education and Training Perspectives. Research Data Management Forum. Manchester, 2008 http://www.dcc.ac.uk/sites/default/files/documents/RDMF/RDMF2/07%20Corrall.pdf

DigCurV 2012a

Engelhardt, C.; Strathmann, S.; McKadden, K. (2012). Report and analysis of the survey of Training Needs. http://www.digcur-education.org/eng/Resources/Report-and-analysis-on-the-training-needs-survey

DigCurV 2012b

Karvelyte, V. et al. (2012). D2.1 Report on baseline survey into training opportunities and evaluation framework. Section 1: Training opportunities survey. http://www.digcur-education.org/eng/Resources/D2.1.1-Survey-of-existing-training-opportunities

Donnelly 2008

Donnelly, M. (2008). RDMF2. Core Skills Diagram. Research Data Management Forum, 17.12.2008. http://data-forum.blogspot.com/2008/12/rdmf2-core-skills-diagram.html

Gold 2007a

Gold, A. (2007). Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians. D-Lib 13 (9/10). http://www.dlib.org/dlib/september07/gold/09gold-pt1.html

Gold 2007b

Gold, A. (2007). Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries. D-Lib 13 (9/10). http://www.dlib.org/dlib/september07/gold/09gold-pt2.html

Gold 2010

Gold, A. (2010). Data Curation and Libraries: Short-Term Developments, Long-Term Prospects. San Luis Obispo. http://digitalcommons.calpoly.edu/lib_dean/27/

High Level Expert Group on Scientific Data 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 (2010). http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=6204

Jones 2008

Jones, E. (2008). E-Science Talking Points for ARL Deans and Directors. Washington. http://www.arl.org/bm~doc/e-science-talking-points.pdf

Neuroth et al. 2012

Neuroth, H. et al. (Ed.) (2012). Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Verlag W. Hülsbusch: Boizenburg & Universitätsverlag Göttingen. http://nestor.sub.uni-goettingen.de/bestandsaufnahme/index.php

NSB 2005

National Science Board (2005). Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century. Arlington; http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf

OECD 2007

Organisation for Economic Cooperation and Development (2007). OECD Principles and Guidelines for Access to Research Data from Public Funding. Paris. http://www.oecd.org/dataoecd/9/61/38500813.pdf

Pampel, Bertelmann & Hobohm 2010

Pampel, H.; Bertelmann, R.; Hobohm, H. (2010). Data Librarianship – Rollen, Aufgaben, Kompetenzen. In: Ein neuer Blick auf Bibliotheken (pp 169–176). 98. Deutscher Bibliothekartag in Erfurt 2009, Hildesheim: Olms.

Piorun et al. 2012

Piorun, M.; Kafel, D.; Leger-Hornby, T.; Najafi, S.; Martin, E.; Colombo, P.; LaPelle, N. (2012). Teaching Research Data Management: An Undergraduate/Graduate Curriculum, *Journal of eScience Librarianship*: Vol. 1: Iss. 1, Article 8. http://escholarship.umassmed.edu/jeslib/vol1/iss1/8 or doi:10.7191/jeslib.2012.1003

RIN 2008

Research Information Network (2008). Stewardship of digital research data. A framework of principles and guidelines. London. http://www.rin.ac.uk/system/files/attachments/Stewardship-data-guidelines.pdf

Rümpel 2010

Rümpel, S. (2010). Data Librarianship – Anforderungen an Bibliothekare im Forschungsdatenmanagement. Potsdam. http://opus.kobv.de/fhpotsdam/volltexte/2010/163/pdf/10600.pdf

Swan & Brown 2008

Swan, A.; Brown, S. (2008). The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs. Report to the JISC.

van der Graaf & Waaijers 2011

Van der Graaf, M.; Waaijers, L. (2011). A Surfboard for Riding the Wave. Towards a four country action programme on Research Data. http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fPrimary+Research+Data%2fSurfboard+for+Riding+the+Wave%2fKE_Surfboard_Riding_the_Wave_Screen.pdf