

Research Data – a Topic of LIS Education?

Achim Osswald

Cologne University of Applied Sciences – Institute of Information Science
50968 Cologne, Germany, Gustav-Heinemann-Ufer 54
achim.osswald@fh-koeln.de

ABSTRACT

In recent years, access to and curation of research data have become a topic of discussion in national and international think tanks and advisory groups. There are various concepts and suggestions to stimulate and improve competences in dealing with research data. In addition, semantic web technologies and Linked Open Data are raising the awareness for access to data and environments in which data can be processed, retrieved, reused and preserved. Competences in dealing with these challenges will extend the scope of work and the tasks not only of researchers but of librarians as well. In the context of a survey on digital preservation of research data in Germany, research data-related tasks have been identified and consequences for the scope of library activities and qualification needs of librarians have been suggested. Competences in collecting, describing, and processing domain-related data in connection with other activities in digital curation of research data gain in importance. LIS (Library and Information Science) curricula should cover these extended qualification needs.

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, the difference between generic and domain-related competences and qualifications should be kept in mind. Until now, mostly generic qualifications related to dealing and curation of research data have been assessed and described. Nevertheless skills and competences needed are among the many aspects related to the topic. In some cases they are influenced by domain-related aspects like type and volume of data, formats etc. In other cases curation and preservation of research data are just links of a chain of generic activities. All this has become a topic of research and (case) studies endeavoring the field and its challenges.

The 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 nestor², the German network of competence for digital preservation, in cooperation with D-Grid GmbH,

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 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>

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 the type of research data relevant to these fields covers nearly every type of research data (see for comparison NSB, 2005).

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 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:

- Are there any generic concepts and role models for libraries and librarians in the field of data processing?
- Are there any best practice experiences which could be recommended to the LIS community?
- What kind of qualifications of librarians are needed to enhance the quality of data processing activities?

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

- The relevance of research data is emphasized by the 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 mentioned above) by the institutions involved will reduce technical and organizational efforts in the long run.
- Metadata and persistent identifiers are an important precondition to enable referencing and citing research data. These tasks need domain-related knowledge as well as generic competences of LIS specialists.

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 disciplines under focus. The role of libraries has not been set up 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

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>

research institutions.

HOW DO THESE RESULTS MATCH RESEARCH RESULTS IN OTHER COUNTRIES?

Osswald and Strathmann (Osswald & Strathmann, 2012) compare the findings of this German study with findings from an international perspective from the US (basically Gold, 2010), the UK (basically Swan & Brown, 2008), Germany (basically Ruempel, 2010 and Buettner, Ruempel & Hobohm, 2011) and the EU (DigCurV, 2012a; DigCurV, 2012b and DigCurV, 2012c) in order to get a broader and international picture of the topic. In addition the study helps to exemplify which tasks related to research data librarians and libraries could or should fulfill in future and how they should be qualified for these tasks.

POTENTIAL QUALIFICATION PROFILES OF LIBRARIANS

The findings of Osswald & Strathmann (2012) show that there will be various roles of librarians dealing with research data. 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).⁴ She emphasizes the “boundary-spanning roles” (ibid.) of those persons taking new responsibilities and tasks.

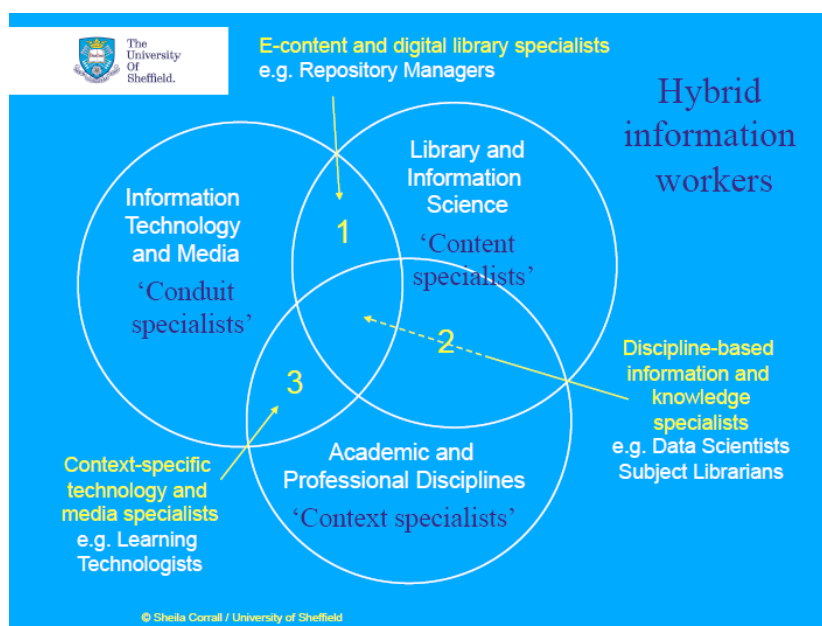


Figure 1: Profiles of hybrid information workers (Corall, 2008; slide 6)

Until now, there has not been a clear consensus on the job profiles emerging from these new activities and skills. Research data-related skills are necessary for persons acting as data creators (scientists by the majority), 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 practitioners 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 because this will help to define the skills and responsibilities relevant within the whole environment of curation of research data (see e.g. Buettner, Ruempel & Hobohm 2011).

⁴ This model has been adopted by Pampel, Bertelmann & Hobohm 2008, Ruempel 2010 as well as by Buettner, Ruempel & Hobohm 2011 in publications in German.

These four roles in data management have been visualized 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 their personal profile, 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.

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 if undergraduate, postgraduate or continuing professional development programs might be the best way to serve the upcoming needs related to the curation of research data.⁵

Nevertheless, LIS organizations and educational institutions should be aware of the development (e.g. related to Open Notebook Science or the so-called Citizen Science) and make up their mind on the consequences for the scope of library related activities. They should continue to survey demands of the R&D community in connection with the political arena while testing concepts for qualification already available (see e.g. <http://digitalcurationexchange.org>).

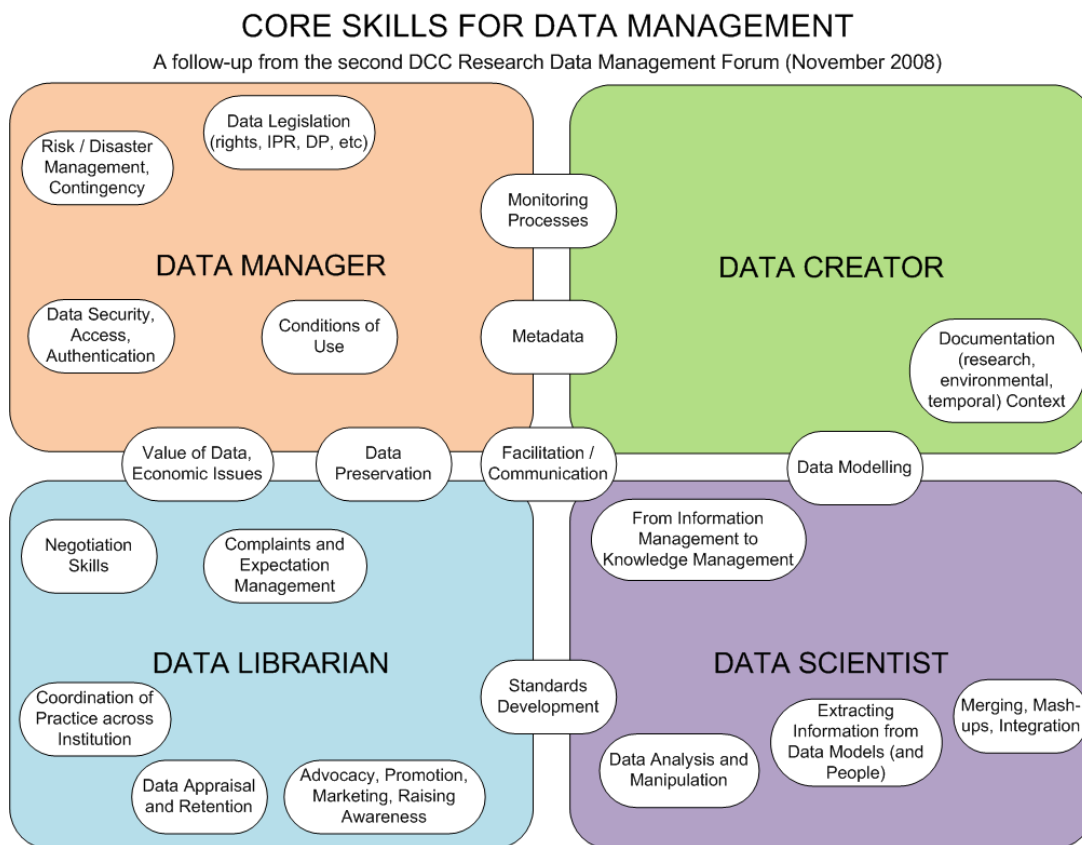


Figure 2: Core skills for data management (Donnelly, 2008)

CONCLUSIONS

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

⁵ In addition see Corall 2008 as well as Swan & Brown, 2008 for UK, Gold 2010 for considerations related to the US, DigCurv 2012b for the EU or Buettner, Ruempel & Hobohm 2011, 211-215 for a German perspective.

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 have not been covering 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 the researchers and information specialists involved. Systematic and conceptual activities are 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 knowledge.

It is time to improve an awareness of these tasks of librarians and libraries and to start a redesign of or amendments to LIS curricula to open up the opportunity of extending the job and competence profile of LIS experts. In the meantime, further educational activities related to the qualifications needed should be fostered.

ACKNOWLEDGMENTS

The author likes to thank his colleagues Heike Neuroth and Stefan Strathmann from Goettingen State and University Library, for their cooperation in the nestor and DigCurV context.

REFERENCES

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

Buettner, Ruempel & 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

Karvelyte, V. et al. (2012). D2.1 Report on baseline survey into training opportunities and evaluation framework. Executive Summary. <http://www.digcur-education.org/eng/Resources/D2.1-Executive-Summary>

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>

DigCurV, 2012c

Engelhardt, C.; Strathmann, S. (2012). D3.2 Report on focus groups meetings. DigCurV Deliverable (internal)

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.

Osswald: Research Data Contribution to the Festschrift for Professor Ramesh Babu (15/03/2012)

D-Lib 13(9/10)

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)

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/

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. Hülbusch: Boizenburg & Universitätsverlag Göttingen.

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>

Osswald & Strathmann, 2012

Osswald, A.; Strathmann, S. (2012). Research Data – Tasks and Challenges for Data Librarians. Paper submitted to International Conference on Trends in Knowledge and Information Dynamics (ICTK – 2012), Bangalore

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>

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

Ruempel, 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.

ao 11.3.12

Osswald: *Research Data Contribution to the Festschrift for Professor Ramesh Babu (15/03/2012)*