Curation and Preservation of Research Data in Germany: A survey across different academic disciplines

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

Heike Neuroth
Goettingen State and
University Library
37073 Goettingen, Germany
neuroth@sub.uni-goettingen.de

Stefan Strathmann
Goettingen State and
University Library
37073 Goettingen, Germany
strathmann@sub.uni-goettingen.de

ABSTRACT

This paper gives an overview of the design and purpose of a survey on the curation of research data in Germany. Eleven disciplines including, among others, the humanities, social sciences, and medicine are addressed. Issues and preliminary findings are summarized. At iPRES2011 findings of this survey will be presented to an international audience for the first time.

Categories and Subject Descriptors

A.1 INTRODUCTORY AND SURVEY

Keywords

Digital curation of research data, metadata, cooperative structures, research archives, cost and funding, training, perspectives and visions, scientific communities, survey, Germany.

1. INTRODUCTION

In the last few years the issue of curation of research data has become a topic of enhanced interest in scientific communities. Awareness of long-term availability, re-usability and integrity of research data has been stimulated by international reports (e.g. by the High Level Expert Group on Scientific Data (October 2010), on behalf of the European Commission) [1]. But even now there is no clear understanding of how to deal with research data curation. Concepts suggested range from attempting to find one promising approach that works for all disciplines to developing specific approaches for every single discipline.

For some disciplines, like astronomy or climate research, international cooperation among research institutions in several countries has already been established (see for example the World Data Systems). Others have not yet begun to address the problem.

On the national level, supporting or stimulating activities like implementing a data management plan have been realized in some countries, e.g. by the NSF (National Science Foundation, USA), where the submission of such a plan is required, or the DFG (German Research Foundation), where a data management plan is

strongly recommended. The ANDS (Australian National Data Service) has gone a step further and initiated the Australian Research Data Commons (ARDC).

2. BASELINE STUDY: RESEARCH DATA IN GERMANY

In Germany libraries, archives, museums and leading experts in the field of digital curation and digital preservation work together in *nestor*, the German competence network for digital preservation. Their objective is to ensure the long-term preservation and accessibility of digital resources [2]. In 2006 a group of experts published the first edition of "nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung" (nestor handbook. A small encyclopedia of digital preservation) [3] which is a comprehensive state-of-the-art documentation on digital curation and digital preservation in German language. The special issues of curation and preservation of digital research data are a topic of some small chapters only because activities in most disciplines have been in a very early state of development. The editors like to complement the encyclopedia with a survey about the curation of digital research data.

Regarding the situation in Germany, there is no clear picture of the methods that different academic disciplines use to preserve and curate their research data. There is a need to address the issue with a baseline study. This will give more stable data to scientists, service infrastructure experts and politicians to foster strategic concepts for digital curation and preservation in and between the disciplines.

To broaden and promote access to researchers and disciplines the editors started a cooperation with D-Grid GmbH [4], a non-profit Development and Operating Company founded by the German Ministry of Education and Research (BMBF) in 2008. D-Grid has the goal to ensure efficient collaboration and cooperation between different projects in the field of a sustainable grid infrastructure in Germany.

2.1 Study design

With the support of scientists in the addressed disciplines, the authors and editors conducted a detailed survey across 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 cover nearly all types of research data. Moreover, it seems that these disciplines have

already started with digital curation of data or have more experience in this field. Therefore, they can be seen as more or less representative for the situation in Germany and will help to get a deeper insight into digital curation and preservation of research data.

Well known and accepted experts from the different disciplines could be gained to serve as authors and contributors to the survey.

A workshop with scientists and scholars in these disciplines in March 2011 showed that there are several issues to be addressed in common. Nevertheless concepts and solutions addressed vary from discipline to discipline. Currently, the results of the survey and the workshop are being evaluated, normalized and a detailed report is being prepared.

2.2 Issues addressed

The report will be published at the beginning of 2012 and will address at least the following issues:

- What types of cooperative structures do already exist? How are they stimulated? What makes them successful?

- What are the types and the amount of data relevant for digital curation and preservation activities?

- What kinds of metadata standards are used? Are there international standards etc., relevant to the discipline in focus?

- Are there any research data archives already dealing with the curation of data in this specific academic discipline? If so, how are they organized and financed? How expensive is the initial funding of such archives and what are the operating expenses per year and in the long run?

- What are the perspectives and visions of data curation and preservation in the different scientific communities?

These issues are part of a larger sample of questions deduced by the editors within a preliminary study of the topic. They have been collected in a detailed survey given to the experts in the different disciplines. This enabled a structured and comparable view to the results collected.

2.3 Preliminary findings

Preliminary findings show that the 11 academic disciplines involved have produced a variety of solutions to the issues. Up to this point, the development of infrastructures to assure the quality of data, to archive it and to assure its long-term availability as well as re-usability has been influenced primarily by traditions and independent infrastructures in the different academic disciplines. Issues of multidisciplinarity and international interoperability are reflected in differing degrees. Scientists seem to be aware that there is a need to find a balance between community-driven approaches and standardization, common policies working for all disciplines versus domain-specific isolated solutions. Although not yet published, the survey has gained some attention inside and outside of Germany. While still in its preparation phase, it initiated an exchange of ideas among the different disciplines. The authors expect that the report will not only illustrate the situation of curation of research data in Germany, but will also stimulate a broader discussion among the different disciplines on an international level.

Major findings of this interdisciplinary survey will be presented at iPRES2011.

3. ACKNOWLEDGEMENTS

4. REFERENCES

[1] 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 - October 2010, http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=6204

[2] See http://www.langzeitarchivierung.de/eng/index.htm.

[3] See the latest edition of Neuroth, Heike; Oßwald, Achim; Scheffel, Regine; Strathmann, Stefan; Jehn, Matthias (Ed.) (2010): *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*: edition 2.3; http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-2010071949

[4] D-Grid GmbH; http://www.d-grid-gmbh.de/index.php?id=1&L=1