

**Committee on Scientific Library Services
and Information Systems
Subcommittee on Information Management**

**Recommendations for Secure Storage and Availability of
Digital Primary Research Data**

January 2009

At the invitation of the LIS Division, a roundtable discussion on the topic of "primary research data" was held at the Head Office of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) on 17 January 2008 (for protocol – in German only - , see: http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/forschungsprim_erdaten_0108.pdf). The objective of the roundtable discussion was to discuss the problem area "storage and provision of primary research data" in order to specify key elements that can serve as the basis for future funding measures of the DFG in the information infrastructure. These key elements are to serve as an interdisciplinary framework, within which additional, discipline-specific regulations and measures are to be developed.

Building on the results of the roundtable discussion, the Subcommittee on Information Management (UA-INF) drafted "Recommendations for Secure Storage and Provision of Digital Primary Research Data" and approved it at the meeting on 21 May 2008. The Committee on Scientific Library Services and Information Systems (AWBI) confirmed the recommendations of the subcommittee in the meeting on 25 and 26 September 2008.

Recommendations for Secure Storage and Availability of Digital Primary Research Data

Status 26 June 2008

Primary research data represent a valuable pool of information acquired at great financial expense. Depending on research area and method, the data are reproducible or are based on non-reproducible observations or measurements. In any case, the acquired data should be made accessible and freely available to the public upon completion of the research. This is the fundamental prerequisite for making data reusable within the scope of new research questions as well as for utilising the data in the re-examination of published results in the event of doubts in the publication.

In 1997, the DFG published "Safeguarding Good Scientific Practice", which included 16 recommendations. Recommendation 7 states: "Primary data as the basis for publications shall be securely stored for ten years in a durable form in the institution of their origin." Thus, the demand to keep data available has been a topic for the DFG for more than ten years.

Technical and organisational prerequisites are necessary for the storage and provision of primary research data and must be developed from within the individual scientific disciplines. This has already been accomplished in exemplary manner in some disciplines; other disciplines are still at the start of this task. Among the key objectives is orienting these structures towards international benchmarks and standards and embedding them in existing, transregional, international structures. In order to ensure that they are effective in the long term, care must be taken from the start.

For the establishment of the necessary organisational prerequisites and structures, discipline-specific working groups should be established to develop suggestions for the respective communities for determining which institutions are to be involved in the storage and provision as well as for the processes and structures that are to be used. Orientation towards already practised models may be helpful here. If possible, an effort should be made to cite the stored data sets and their sources. This option would give the scientist or academic incentive to publish data on his or her own initiative. There are already successful examples of this. The objective should be the voluntary passing on of the data by the scientist or academic. Another key criterion that should be considered is the quality control of the data.

These recommendations are directed toward both the scientists and academics as well as toward the information specialists. It should awaken awareness of the urgency of this issue. Furthermore, the objectives are named that should be taken into account during the implementation and the adherence of which is of fundamental importance. Lastly, the scientists and academics should be motivated to perform this task as a kind of voluntary commitment and to properly implement it within the scope of the specified structures on their own initiative.

On the basis of these recommendations, the DFG will make financial resources available as part of its funding opportunities to support the expansion of the storage and cataloguing structures. In particular, it is encouraging exploration projects to collect further experiences for the individual disciplines. Regulation of the storage and provision of the data should be specified right from the start in the scientist's work plan and should be stated in the funding proposals within the framework of the funding programme.

The 7 recommendations in detail:

1. *Primary research data are data that result during the course of scientific research, experiments, measurements, surveys or polls. They serve as the basis for scientific publications.*

Depending on the respective discipline, the primary research data are to be defined differently. The scientists and academics should, thus, decide themselves within their respective communities whether this is to include raw data or at what level of aggregation the data are to be stored for the long term. Furthermore, the granularity should be agreed upon within a rough framework: how much data constitute a data set that is assigned a stable address (persistent identifier).

2. *A discipline-specific organisational concept that regulates effective and sustainable storage of the data is to be defined.*

The scientists and academics determine within their respective disciplines an organisation or organisational structure that ensures both the fast, convenient provision of the data in ongoing research processes and their long-term availability. Furthermore, an agreement is to be made as to whether there is to be a maximum storage period after which the data are returned to the researcher or deleted.

3. *The research data are stored within the framework of defined standards.*

The adherence to standards is of utmost importance. Appropriate regulations are to be established by the scientists and academics in cooperation with information specialists. Attention is to be given here, above all, to international regulations in order to also ensure compatibility on an international level.

4. *The data are personally labelled and are stored under the name of the researcher.*

Storage of the data must encompass permission management. Regulations should correspond to the principles of Open Access.

5. *If possible, each scientist or academic makes his or her primary research data freely available on a transregional level.*

Depending on the culture in the respective discipline, the scientists and academics should come to a consensus as to whether the data are to be made freely available to the public immediately upon completion of the research or after expiration of a moratorium period. If the data were acquired through projects that were financed with public funds, they are, in principle, freely available to the public. Special regulations are to be established for research data subject to patents or used for other applications with direct economic use. This should occur in consensus with the affected scholars.

6. *The data are described by metadata.*

With the metadata (at least: Dublin Core), the bibliographic facts, among other information, are recorded. This includes the name of the researcher who collected the data, the name of the data set, the place and year of publication, as well as technical data (format, etc.). In the contextual metadata, the primary data are comprehensively described. Also included here are details on the basic conditions under which they were collected or measured. Here, the author also describes the research question that led to the origination of the data. All information necessary for reusing the data to address other research questions should be included here. The criteria of Information Life Cycle Management should be taken into account.

7. *The data are quality controlled.*

Criteria and processes that ensure the quality assurance of the data are to be developed in the individual scientific communities. Also to be defined in this regard is the degree of scientific qualification above which the research data are to be stored and made available to the public.