

Reflections and Practices on the Development of the German National Library of Science and Technology

Authors: Jan Brase, Uwe Rosemann, Irina Sens, Gu Liping

Date: 2016-06-04T00:00:00+00:00

Abstract

Introduces new developments in research data applications and new challenges in research data rights management. Describes the scalable architecture of the German National Library of Science and Technology and the practical achievements of its research data portal GetInfo, permanent identification service DOI, global data citation service DataCite, among others, and discusses the need for training data librarians. These services can enhance retrieval of non-textual resources, identify digital data resources, promote formal citation of research data, and enhance library value.

Full Text

Volume 57, Issue 20, October 2013

Development Thinking and Practice on the German National Library of Science and Technology

Jan Brase, Uwe Rosemann, Irina Sens (Authors)

Gu Liping (Translator)

Abstract This paper briefly introduces the new developments in research data application and the emerging challenges in research data rights management. It describes the scalable framework of the German National Library of Science and Technology (TIB) and its practical achievements in the research data portal GetInfo, the permanent identification service DOI, and the global data citation service DataCite, and discusses the need for data librarian training. These services can enhance the retrieval of non-textual resources, identify digital data resources, enable formal citation of research data, and improve library value.

[Keywords] data citation; digital object identifier; research data integrated portal; rights and management of research data; metadata schema

[Classification Number] G250.76

DOI: 10.7536/j.issn.0252-3116.2013.20.009

The original article, titled “Kommentar aus Bibliothekssicht,” was published in Zeitschrift für Bibliothekswesen und Bibliographie. With the authors’ permission, it has been translated and slightly adapted for publication in Chinese. This paper is one of the research outcomes of the Chinese Academy of Sciences National Science Library’s Directed Talent Research Project “Research and Consultation on Science and Technology Information Policy” (Project No. Library 1203).

1. Challenges and Future of Library Services

Scientific information is becoming increasingly vast and complex, gradually evolving toward the integration of massive datasets. Although knowledge dissemination still relies on the form of scientific papers, from the perspective of the entire research process, this merely represents the final step in research data management workflows.

The U.S. National Science Foundation notes in its report “Cyberinfrastructure Vision for 21st Century Discovery”: “Science and engineering research and education are becoming increasingly data-intensive: digital technologies, instruments, and networks are used to collect, generate, share, and analyze data, and to disseminate findings. Scientists and engineers worldwide produce, acquire, analyze, integrate, and store terabytes of digital data daily through experiments, observations, and simulations. Consequently, new scientific methods dynamically integrate observational and simulation-generated data, intelligently adapting to changing conditions to uncover new knowledge” [1]. These new developments in research data application and the emerging challenges in research data rights management are guiding libraries toward new directions.

Microsoft Research’s open publication *The Fourth Paradigm: Data-Intensive Scientific Discovery* particularly observes how science itself shifted from empiricism to data-intensiveness at the end of the 20th century. Scientific communication, including peer review, is also undergoing fundamental transformation. Due to cost, timeliness, and the need to integrate experimental data with scholarly literature, digital libraries’ public sharing services will comprehensively replace traditional libraries’ role in collecting publications [2].

Scientific information is now presented not merely as a single article or monograph. Library portals can integrate information and provide development services, thereby continuing to serve as an important source for users to access information [3]. Library users are primarily concerned with whether they can obtain satisfactory answers to their queries—regardless of whether the information they seek takes the form of papers, books, or other content types such as

research data, 3D models, images, videos, etc. As providers of knowledge and information services, librarians' task is to deliver user services, offering appropriate scientific information as the most ideal response to their questions.

If the library community wishes to continue fulfilling this information provision responsibility, it must release bibliographic data for diverse content formats. Therefore, the key considerations are:

- The discovery of knowledge objects depends on appropriate metadata. This means enriching existing metadata for each object through various types of information as much as possible to satisfy user searches. Consequently, traditional keyword-based search methods in catalogs need to be expanded into new search approaches that can accommodate different content formats.
- New methods for resource acquisition and collection development must be established to ensure content quality.
- The ability to support referencing and citation is a prerequisite for publications containing scientific content to achieve permanent availability.

To this end, in recent years the German National Library of Science and Technology (Technische Informationsbibliothek, TIB) has established a DOI registration agency and the related international association (DataCite). Scalability means that libraries can directly handle massive amounts of information as well as traditionally familiar content carriers without needing to hold all the content themselves. On the one hand, libraries must consider the technical requirements for storing massive research datasets; on the other hand, they must consider the certification mechanisms and measures for specialized storage and maintenance of this content.

Library catalogs should not serve as mere windows displaying their own resources; rather, they should be developed into portals for distributed extended content. The term "extended content" refers to large datasets, cataloged videos, learning objects, and similar materials that provide results relevant to users' research. After successfully accessing and obtaining content, readers can, through stable links, further reach certified archives and data banks that specialize in storing relevant materials.

This continuously evolving digital library architecture is illustrated in Figure 1 [Figure 1: see original paper].

2. GetInfo Portal for Research Data: A Case Study

The German National Library of Science and Technology (TIB) is a national specialized library serving disciplines such as architecture, chemistry, informetrics, mathematics, and physics. Its GetInfo portal is a gateway specifically established for scientific research information in natural sciences and engineering, encompassing access points to advanced data banks, publishing aggregators, and

library catalogs, as well as integrated full-text delivery systems. Behind GetInfo lies a complete suite of scientific and technical information services provided by TIB.

As one of Europe's major library gateways, GetInfo provides users with access to research data. This new type of content service includes two approaches:

- Searching within GetInfo and the Union Catalog (Gemeinsamer Bibliotheksverbund, GBV). When users search for information on specific topics, they are often interested in datasets collected by certain scientists. They can directly use permanent identifiers (here referring to DOI digital object identifiers) to view research data as independent objects directly in the library catalog.

In GetInfo's search results, research data occupies a central position. By resolving the permanent identifier, users can see the data center from the metadata on the page and obtain datasets from the data center, much like downloading a record or a link to the entire dataset. This workflow follows the utilization patterns and functional effects of DOI in scientific journals.

- Visualization of research data. GetInfo provides visual maps for geological research data, among others. After resolving the DOI name in GetInfo, users can enter web pages of geological data centers such as PANGAEA, accessing data from within the data center.

3. Permanent Identification for Scientific Data: DOI Services

The German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) commissioned TIB to implement the "Publication and Citation of Research Data" project, which has successfully tested infrastructure for handling scientific data using DOI names and URNs [4]. In TIB's system, over 650,000 records from geosciences can already be seen with permanent identifiers.

Integrating data with publications is an important component of scientific collaboration, allowing research results to be verified and actively exchanged. Although the scientific community generally agrees that data should be utilized across disciplines, it recognizes that few people currently perform additional work on data preprocessing, literature organization, and ensuring data quality.

The traditional form of publishing scientific results is through specialized journals, but typically without publishing the underlying data. When these traditional publications enter "citation indexes," scientists can receive performance evaluations. However, these indexes have not, to date, included data publications.

The application of DOI as an identifier can link a scientific paper with the research data analyzed in that paper. Articles and datasets can be jointly cited

through their respective DOI names. This form of citation and linking provides research data as supplementary data citations directly pointing to scientific papers. Examples are as follows:

- Citation identifier for supplementary data:

Kuhlmann, H. et al. (2009): Age models, iron intensity, magnetic susceptibility records and dry bulk density of sediment cores from around the Canary Islands. PANGAEA, Bremen
doi:10.1594/PANGAEA.727522

- Linking to related research paper:

Kuhlmann, Holger; Freudenthal, Tim; Helmke, Peer; Meggers, Helge (2004): Reconstruction of paleoceanography off NW Africa during the last 40,000 years: Influence of local and regional factors on sediment accumulation. In: Marine Geology, 207(1-4), 209
doi:10.1016/j.margeo.2004.03.017

4. Global Citation of Scientific Data: DataCite

Based on TIB's experience as a DOI registration agency handling research data, numerous science and technology libraries and information centers jointly established the DataCite international consortium in December 2009. Its goal is to enable scientists to access research data on the Internet, accept research data as independent and citable research objects, and thereby follow the Principles of Good Scientific Practice [5]. Fifteen partners from ten countries have already built the DataCite framework: the British Library, the French Scientific and Technical Information Center (Institut de l' Information Scientifique et Technique, INIST), the Danish Technical Information Center, Delft University of Technology Library in the Netherlands, the Canada Institute for Scientific and Technical Information (CISTI), the Office of Scientific and Technical Information in the U.S. Department of Energy (OSTI), the California Digital Library, Purdue University, the Australian National Data Service (ANDS), the Swedish National Data Service (SND), and the Swiss Federal Institute of Technology. German partners come from TIB's Goporis partners: the German Central Library of Medicine (Deutsche Zentralbibliothek für Medizin, ZBMED), the German Library of Economics and Science, and the Leibniz Institute for Social Sciences GESIS.

Since its inception, DataCite has positioned itself as an intermediary between data centers, scientists, publishers, and libraries. It provides services in three areas:

- DataCite is the official registration agency for Digital Object Identifiers (DOI) and a member of the International DOI Foundation (IDF). As of January 2011, DataCite had registered DOI names for over one million digital objects.

- DataCite actively develops exchanges on standards and best practices and establishes workflows. Among DataCite partners, a working group is responsible for communicating a common metadata schema to the community that encompasses all research data citation information [6]. DataCite continuously participates in the work of the newly established CODATA Task Group on Data Citation Standards and Practices (www.codata.org/taskgroups/TGdatacitation/index.html), developing recommendations for research data citation. It constantly discusses cooperation possibilities with publishers: encouraging researchers to make their data available by establishing data citation indexes.
- All data registered with DataCite has been centralized in an accessible central metadata database since 2011. The content of this database is freely available and can be accessed through third-party sites such as Thomson Reuters' Web of Science.

All these activities follow DataCite's founding principle: science is global, and therefore requires global partnerships; it is willing to work with other global institutions, such as publishers, to jointly develop appropriate standards. Many scientists among the member partners are enthusiastic about local promotion activities, establishing networks within their research fields, libraries, and funding agencies. DataCite benefits from the service initiatives of many local scientists and receives support from local partner institutions.

5. Future-Oriented Library and Information Services

The diverse content displayed in library catalogs will create new significance for libraries and librarians [7]. Academic and specialized libraries can position themselves as service facilities to enhance research data access capabilities for universities or research institutes; alternatively, they can collaborate with regional research centers to establish data centers, importing resource links into library catalogs.

Library science is also important for storing and maintaining research datasets. In English-speaking countries, there are publications such as "The Data Librarian," and Germany's Potsdam University of Applied Sciences has already pioneered this course [7]. Library staff can actively participate in training provided by such institutions for knowledge objects and publishing domains, enabling them to appropriately handle research data when necessary.

References

- [1] National Science Foundation. Cyberinfrastructure vision for 21st century discovery [OL]. [2013-03-08]. <http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf>.

- [2] Hey T, Tansley S, Tolle K. The fourth paradigm: Data-intensive scientific discovery [M]. Redmond, WA: Microsoft Research, 2009: ixx. [Translator's note: Chinese translation available: Pan Jiaofeng, Zhang Xiaolin, et al., trans. The Fourth Paradigm: Data-Intensive Scientific Research [M]. Beijing: Science Press, 2012.]
- [3] Inger S, Gardner T. How readers navigate to scholarly content [OL]. [2013-03-08]. <http://www.sic.ox14.com/howreadersnavigatetoscholarlycontent.pdf>.
- [4] Heery R, Lyon L. Research and advanced technology for digital libraries [M]//Brase J. Using Digital Library Techniques-Registration of Scientific Primary Data. Berlin: Springer, 2004: 488-494.
- [5] Brase J. DataCite-A global registration agency for research data [OL]. [2013-03-08]. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5361881>.
- [6] DataCite. DataCite metadata schema for the publication and citation of research data [OL]. [2013-03-08]. <http://schema.datacite.org/>.
- [7] Pampel H, Hobohm HC, Bertelmann R. Datalibrarianship-Roles, tasks, competencies [OL]. [2013-03-08]. <http://hdl.handle.net/10760/14896>.

Call for Papers

Knowledge Management Forum (ISSN 2095-5472, Beijing Publication License No. 058) publishes research on knowledge production, creation, organization, integration, mining, sharing, analysis, utilization, and innovation. The journal covers knowledge management issues in government, enterprise, university, library, and other physical and virtual organizations, including theories, methods, tools, technologies, applications, policies, solutions, and best practices.

The journal practices article-by-article publishing. Accepted manuscripts enter a rapid publication process and are immediately made fully open access.

We invite submissions from scholars and practitioners worldwide:

1. Manuscripts should address knowledge-related topics, discussing issues in knowledge management, knowledge services, knowledge innovation, etc. Articles may focus on theory, application, technology, methods, models, or best practices.
2. Articles must be substantive, integrate theory with practice, have clear research purposes and appropriate methods, demonstrate academic insight, and offer reference or guidance value for theory or practice.
3. All submissions undergo similarity detection and peer review, including initial, secondary, and final editorial reviews.

4. Article length is flexible, but generally should be between 4,000-20,000 words.
5. Authors will be notified of acceptance decisions within one month.
6. Articles are primarily published online through our website (www.kmf.ac.cn) and authorized databases, with print-on-demand availability.

Please submit manuscripts at www.lis.ac.cn, indicating “Submission to Knowledge Management Forum.” Contact: Liu Yuanying, Tel: 010-82626611-6638.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv –Machine translation. Verify with original.