

Kommentar

JAN BRASE, UWE ROSEMANN, IRINA SENS

Kommentar aus Bibliothekssicht

Foto: privat



Jan Brase

Foto: privat



Uwe Rosemann

Foto: privat



Irina Sens

EINLEITUNG

Wissenschaftliche Information wird immer komplexer und basiert immer häufiger auf großen Datensammlungen. Die übliche Verbreitungsform für Wissen ist immer noch der wissenschaftliche Artikel, der aber nur den letzten Schritt in einem Prozess darstellt, an dessen Anfang Forschungsdaten gestanden haben.

In ihrem Bericht »Cyberinfrastructure Vision for 21st Century Discovery« schrieb die National Science Foundation (USA):

»Science and engineering research and education have become increasingly data-intensive as a result of the proliferation of digital technologies, instrumentation, and pervasive networks through which data are collected, generated, shared and analyzed. Worldwide, scientists and engineers are producing, accessing, analyzing, integrating and storing terabytes of digital data daily through experimentation, observation and simulation. Moreover, the dynamic integration of data generated through observation and simulation is enabling the development of new scientific methods that adapt intelligently to evolving conditions to reveal new understanding.« ([5])

Im dem kürzlich veröffentlichten Buch »The Fourth Paradigm: Data-Intensive Scientific Discovery« [3] von Microsoft Research wird ebenfalls darauf hingewiesen, wie sich die Wissenschaft generell in den letzten Jahrhunderten von einer empirischen zu einer datenintensiven Disziplin gewandelt hat. Aus der Sicht von Bibliotheken werden dort ebenfalls neue Aspekte identifiziert:

»Scientific communication, including peer review, is also undergoing fundamental changes. Public digital libraries are taking over the role of holding publications from conventional libraries because of the expense, the need for timeliness, and the need to keep experimental data and documents about the data together.«

Wissenschaftliche Information tritt heutzutage nicht mehr nur in der Form eines Artikels oder eines Buches in Erscheinung. Bibliotheksportale sind jedoch auch weiterhin eine klassische Quelle für Information [4].

Ein Nutzer einer Bibliothek ist in erster Linie an einer zufriedenstellenden Antwort auf seine Suchanfrage interessiert, dabei ist auch für den Nutzer nicht entscheidend, ob die Information, die er sucht, in Form eines Artikels oder Buches angeboten wird, oder eine ganz andere Form, wie beispielsweise For-

schungsdaten, 3D-Modelle, Fotos, Filme oder ähnliches hat. Um ihrer Aufgabe als Wissens- und Informationsdienstleister gerecht zu werden und dem Nutzer die wissenschaftliche Information zu bieten, die die ideale Antwort auf seine Anfrage darstellt, müssen Bibliotheken ihre Kataloge für andere Inhaltsformen öffnen, wenn sie weiterhin ihren Aufgaben der Informationsversorgung gerecht werden wollen.

Entscheidend für diese Aufgaben ist unter anderem:

- Der Nachweis der Wissensobjekte durch adäquate Metadaten. Dies bedeutet auch die Anreicherung vorhandener Metadaten durch Informationen, die es ermöglichen, gezielt diejenigen Objekte anzubieten, die der Suche des Nutzers entsprechen. Dazu muss es ggf. nötig sein, auch die klassischen Methoden der Suche mittels Schlag- und Stichworten im Katalog durch neue Suchmethoden zu erweitern, die den verschiedenen Inhaltsformen entsprechen.
- In der Akquise und Erschließung neue Methoden zu entwickeln, um die Qualität der nachgewiesenen Inhalte sicherstellen zu können.
- Eine Prämisse für die Publikation und dauerhafte Verfügbarkeit von wissenschaftlichen Inhalten / Datensätzen ist die persistente Referenzierbarkeit und Zitierfähigkeit. Dafür wurde in den vergangenen Jahren an der TIB eine auf DOI basierende Registrierungsagentur und ein entsprechender internationaler Verein (DataCite) gegründet.

Weiterhin bedeutet aber auch gerade im Umgang mit teilweise sehr großen Datensätzen oder auch einer klassischen Bibliothek sehr fremden Inhaltsformen, dass Bibliotheken nicht mehr alle Inhalte selber vorhalten müssen. Teilweise sind die technischen Anforderungen an die Speicherung sehr großer Forschungsdatensätze immens, andererseits gibt es aber auch vertrauenswürdige Einrichtungen, die sich gerade auf die Speicherung und Pflege dieser Inhaltstypen spezialisiert haben.

Einen Bibliothekskatalog der Zukunft sollte man also nicht mehr als reines Schaufenster auf die eigenen Bestände verstehen, sondern als ein Portal zu verteilten externen Inhalten. Externe Inhalte wie große Datensätze, aber auch Filme oder beliebige Lernobjekte werden katalogisiert, um dem Nutzer passende Ergebnisse zu seinen Recherchen anzubieten. Der Zugang zu den Inhalten erfolgt dann über eine sta-

»Portal« statt »Schaufenster«

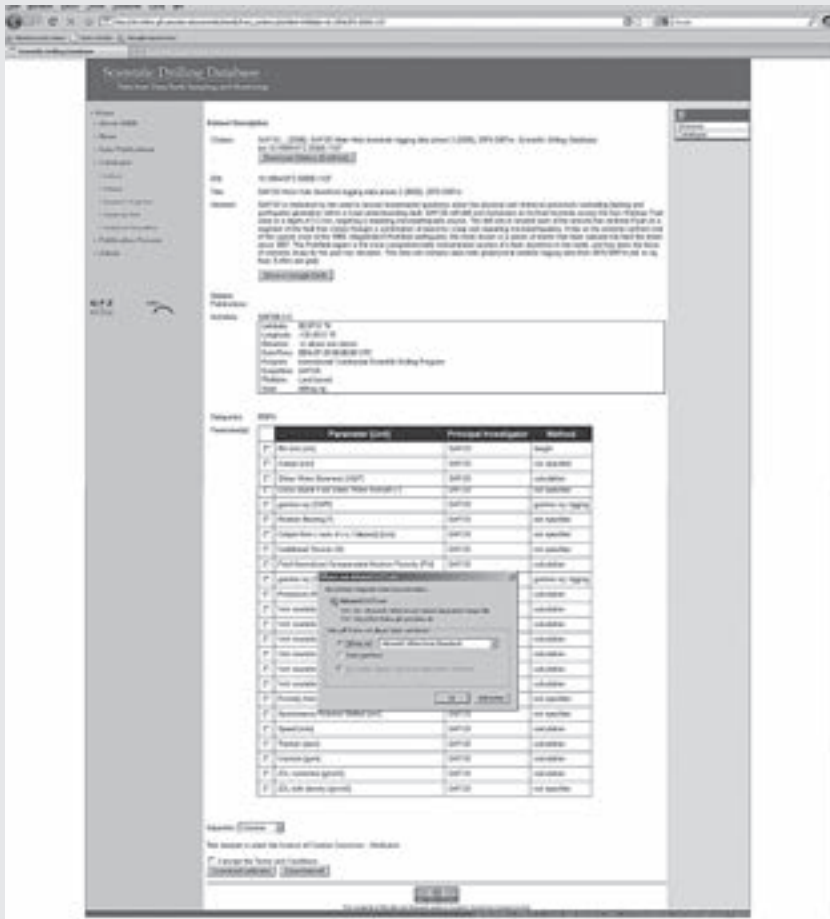


Abb. 3: Präsentation des Datensatzes beim Datenzentrum mit entsprechendem Download-Link



Abb. 4: Vorschau einer geologischen Karte beim Datenzentrum PANGAEA

dem Thema haben, sondern auch an den Datensätzen, die ein bestimmter Wissenschaftler gesammelt hat. Durch die Vergabe von persistenten Identifiern (hier DOI-Namen) werden diese Daten direkt über den Katalog zugänglich. Derzeit sind bereits über 3.500 Datensätze, die Grundlage einer wissenschaftlichen Publikation sind, direkt als eigenständige Objekte über GetInfo und den Gemeinsamen Bibliotheksverbund (GBV) zugänglich [1]. Abbildung 2 zeigt Daten zu einem Bohrkern als Ergebnis einer Suche in GetInfo. Durch Auflösen des persistenten Identifiers gelangt man auf eine Vorschauseite des Datenzentrums, die weitere Metadaten zu dem Datensatz enthält, sowie Download-Links zu einzelnen Teilen des Datensatzes oder den gesamten Daten (Abb. 3). Dieser Workflow entspricht der traditionellen Verwendung von DOI-Namen bei wissenschaftlichen Zeitschriften.

Ein weiteres Beispiel neuer Inhaltsformen sind geologische Karten, sozusagen visualisierte Forschungsdaten. Auch solche sind über GetInfo verfügbar. Die Auflösung des DOI-Namen führt hier ebenfalls auf eine Vorschauseite beim Datenzentrum (Abb. 4), von der dann die eigentliche Karte zugänglich ist.

PERSISTENTE IDENTIFIZIERUNG

Die Integration von Daten in Publikationen ist ein wichtiger Bestandteil wissenschaftlicher Zusammenarbeit, sie ermöglicht die Verifizierung wissenschaftlicher Ergebnisse und aktiven Wissensaustausch von Forschern.

Im wissenschaftlichen Bereich besteht zwar grundsätzlich Bereitschaft, Daten für eine interdisziplinäre Nutzung zur Verfügung zu stellen, aber es ist zurzeit unüblich, dass die erforderliche Mehrarbeit für Aufbereitung, Kontextdokumentation und Qualitätssicherung im Wissenschaftsbetrieb anerkannt wird.

Die klassische Form der Verbreitung wissenschaftlicher Ergebnisse ist ihre Veröffentlichung in Fachzeitschriften, normalerweise ohne Veröffentlichung der zugrunde liegenden Daten. Diese klassische Publikation wird im »Citation Index« erfasst. Dieser Index wird zur Leistungsbewertung von Wissenschaftlern herangezogen. Datenveröffentlichungen werden darin bisher nicht berücksichtigt.

In dem an der Technischen Informationsbibliothek (TIB) Hannover durchgeführten DFG-Projekt »Publikation und Zitierfähigkeit wissenschaftlicher Forschungsdaten« wurde eine Infrastruktur zur Registrierung von DOI-Namen und URNs für wissenschaftliche Datensätze geschaffen und erfolgreich getestet. Mit dem System wurden an der TIB bereits über 650.000 Datensätze aus dem Bereich der Geowissenschaften mit persistenten Identifiern versehen.

Die Verwendung von DOI-Namen als Identifier ermöglicht eine elegante Verlinkung zwischen einem wissenschaftlichen Artikel und den im Artikel analysierten Forschungsdaten. Artikel und Datensatz sind durch ihre jeweiligen DOI-Namen in gleicher Weise eigenständig zitierbar. Diese Form der Zitierung und Verlinkung bietet sich insbesondere bei Forschungsdaten an, die in direkter Beziehung zu wissenschaftlichen Artikeln stehen, sogenannte *supplementary data*. So wird beispielsweise der Datensatz:

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

in folgendem Artikel verwendet:

Kuhlmann, Holger; Freudenthal, Tim; Helmke, Peer; Meggers, Helge (2004): Reconstruction of paleoenvironment off NW Africa during the last 40,000 years: influence of local and regional factors on sediment accumulation. In: *Marine Geology*, 207(1–4), 209–224
doi:10.1016/j.margeo.2004.03.017

DATA CITE

Basierend auf der Erfahrung der TIB als DOI-Registrierungsagentur für Forschungsdaten gründeten mehrere technische Bibliotheken und Informationseinrichtungen im Dezember 2009 den Verein DataCite [2], der sich zum Ziel gesetzt hat, Wissenschaftlern den Zugang zu Forschungsdaten über das Internet zu erleichtern, die Akzeptanz von Forschungsdaten als eigenständige, zitierfähige wissenschaftliche Objekte zu steigern und somit die Einhaltung der Regeln guter wissenschaftlicher Praxis zu gewährleisten. Bis heute haben sich 15 Partner aus zehn Ländern unter dem Dach von DataCite zusammengefunden: die British Library, das französische Institut de l'Information Scientifique et Technique (INIST), das Technical Information Center of Denmark, die TU Delft Library aus den Niederlanden, das Canada Institute for Scientific and Technical Information (CISTI), das Office of Scientific and Technical Information (OSTI) des US Department of Energy, die California Digital Library (USA), die Purdue University (USA), der Australian National Data Service (ANDS), der Swedish National Data Service (SND) und die Eidgenössische Technische Hochschule in Zürich. Deutsche Mitglieder sind neben der Technischen Informationsbibliothek (TIB) die Goportis-Partner Deutsche Zentralbibliothek für Medizin (ZB MED), die Deutsche Zentralbibliothek für Wirtschaftswissen-

schaften sowie das Leibniz-Institut für Sozialwissenschaften GESIS.

Seit seiner Gründung versteht sich DataCite als Mittler zwischen Datenzentren, Wissenschaftlern, Verlagen und Bibliotheken. Die Aktivitäten von DataCite sind in drei Bereiche aufgeteilt:

1. DataCite ist eine offizielle DOI-Registrierungsagentur und Mitglied in der International DOI Foundation (IDF). Zum Stand: 1/2011 hat DataCite bereits über eine Millionen Objekte mit DOI-Namen registriert.

2. DataCite ist aktiv an der Entwicklung von Standards und dem Austausch von Best-Practice-Beispielen und der Definition von Workflows beteiligt. So wurde in einer eigens dafür eingerichteten Arbeitsgruppe unter Teilnahme aller DataCite-Mitglieder und im Austausch mit der Community ein gemeinsames Metadatenschema entwickelt, das alle Informationen zur Zitierung von Forschungsdaten enthält.

Weiterhin ist DataCite aktiv an der neu gegründeten CODATA Task Group Data Citation Standards and Practices (www.codata.org/taskgroups/TGdatacitation/index.html) beteiligt, die Empfehlungen zur Zitierung von Forschungsdaten definieren wird. In Kooperation mit Verlagen werden derzeit Möglichkeiten diskutiert, einen eigenen Data-Citation-Index zu erstellen, um Forschern noch mehr Anreize zu geben, ihre Daten verfügbar zu machen.

3. Alle von DataCite registrierten Daten werden ab 2011 über eine zentrale Metadatenbank zugänglich gemacht werden. Die Inhalte dieser Datenbank werden frei verfügbar sein und auch über Drittanbieter wie beispielsweise Thomson Reuters Web of Science nachgewiesen werden.

Alle diese Aktivitäten folgen dem Gründungsprinzip von DataCite: Wissenschaft ist global und daher braucht es einen global aktiven Ansprechpartner, der in Kooperation mit anderen global tätigen Einrichtungen wie Verlagen gültige Standards und Empfehlungen entwickeln kann. Der Wissenschaftler als solcher ist aber lokal tätig, eingebunden in sein eigenes Netz von Einrichtungen, Bibliotheken und Forschungsförderung. Daher wird DataCite von lokalen Mitgliedern getragen, die vor Ort dem Wissenschaftler zur Verfügung stehen.

WEITERE AUFGABEN FÜR BIBLIOTHEKEN

Neben dem Nachweis der Inhaltsformen im Katalog ergeben sich noch weitere neue Aufgabenfelder für Bibliotheken und Bibliothekare (siehe auch [7]). Universitätsbibliotheken können sich als Service-Einrichtungen verstehen, um den Zugang zu Forschungsda-

Mittler zwischen Datenzentren, Wissenschaftlern, Verlagen und Bibliotheken

lokale Wissenschaftler vs. globale Wissenschaft

ten universitätsweit zu verbessern. Dazu kann in Kooperation mit dem regionalen Rechenzentrum ein Datenzentrum aufgebaut werden, dessen Inhalte dann über den Bibliothekskatalog nachgewiesen werden. Bibliothekarische Fähigkeiten sind auch in der Speicherung und Pflege von Forschungsdatensätzen wichtig. In den angelsächsischen Ländern hat die Etablierung sogenannter »Data librarians« Einzug gehalten, an der Fachhochschule Potsdam gibt es bereits einen ersten solchen Ausbildungsgang in Deutschland. Die Mitarbeit von Universitätsbibliotheken kann hier aktiv in einzelnen Instituten einen neuen Umgang mit Forschungsdaten als eigenständigen Wissensobjekten und Publikationen protegiere und ggf. Schulungen zum richtigen Umgang mit Forschungsdaten durchführen.

LITERATUR

- [1] BRASE, J. (2004), Using Digital Library Techniques – Registration of Scientific Primary Data. Lecture Notes in Computer Science, No 3232, pp 488–494.
- [2] BRASE, J. (2010), DataCite – A global registration agency for research data, Working Paper 149 / 2010, German Council for Social and Economic Data (RatSWD).
- [3] HEY, T.; TANSLEY, S.; TOLLE, K. EDS. (2009), The Fourth Paradigm: Data-Intensive Scientific Discovery, Microsoft Research, ISBN 978-0-9825442-0-4.
- [4] INGER, S.; GARDNER, T. (2008), How Readers Navigate to Scholarly Content, www.sic.ox14.com/howreadersnavigatetoscholarlycontent.pdf

[5] NSF (2007), Cyberinfrastructure Vision for 21st Century Discovery, Arlington / VA: National Science Foundation (NSF), Cyberinfrastructure Council (CIC).

[6] DataCite (2011), DataCite Metadata Scheme for the Publication and Citation of Research Data, Available at <http://schema.datacite.org/>

[7] PAMPEL, H.; BERTELMANN, R.; HOBHOHM, H.-C., Data Librarianship – Rollen, Aufgaben, Kompetenzen, Tagungsband des 98. Deutschen Bibliothekartags, 2009.

DIE VERFASSER

Dr. Jan Brase ist Leiter der DOI-Registrierungsagentur an der Technischen Informationsbibliothek und Geschäftsführer von DataCite, DataCite – Technische Informationsbibliothek (TIB), Welfengarten 1b, 30167 Hannover, Mail: jan.brased@tib.uni-hannover.de

Uwe Rosemann ist Direktor der Technischen Informationsbibliothek und Universitätsbibliothek Hannover, Technische Informationsbibliothek (TIB), Welfengarten 1b, 30167 Hannover, Mail: uwe.rosemann@tib.uni-hannover.de

Dr. Irina Sens ist Stellvertretende Direktorin der Technischen Informationsbibliothek und Abteilungsleiterin Digitale Bibliothek und wissenschaftliche Dienste, Technische Informationsbibliothek (TIB), Welfengarten 1b, 30167 Hannover, Mail: irina.sens@tib.uni-hannover.de