

Informationsinfrastruktur gestalten: Bericht von der 10. Internationalen Bielefeld Konferenz, 24. bis 26. April 2012

Viola Stüven

Die Studierenden von heute sind mit dem Internet aufgewachsen. Sie nutzen es wie selbstverständlich jederzeit, überall, mit unterschiedlichsten Geräten und im regen Austausch mit Gleichgesinnten, Freunden und Kommilitonen. Die Studierenden von heute sind die Wissenschaftlerinnen und Forscher von morgen. Wie werden sie auf die Arbeitsergebnisse anderer Wissenschaftler zugreifen? Welche Möglichkeiten werden sich ihnen eröffnen, eigene Forschungsergebnisse zu publizieren? Welche Dienstleistungen werden sie von ihren Bibliotheken vor diesem Hintergrund erwarten?

Alte Formate wie das gedruckte Buch aber auch der elektronische Zeitschriftenartikel werden zunehmend durch neue Arten des wissenschaftlichen Publizierens ergänzt und erweitert. Es entstehen heute schon angereicherte elektronische Artikel mit Verbindungen zu vielen weiteren Daten und Informationen sowohl textueller als auch visueller Art. Aber auch kleinere Einheiten des Wissens, auf die man sich in der eigenen Arbeit beziehen kann, die also eindeutig referenzierbar sind und dadurch zitierbar werden, spielen in Zukunft eine Rolle¹. Welche Infrastruktur ist nötig, Wissenschaftler zukünftig mit Daten, Informationen und Wissen zu versorgen? Mit welchen Dienstleistungen können Universitätsbibliotheken sowohl die Publikation der wissenschaftlichen Ergebnisse selbst als auch die Veröffentlichung der zugrunde liegenden Forschungsdaten unterstützen? Wie wird sich das Aufgabengebiet der Bibliothekare verändern?

Ein Forum für diese Fragen bot die 10. Internationale Bielefeld Konferenz mit 33 Referenten aus 9 Ländern und gut 300 Teilnehmern aus 24 Ländern im April dieses Jahres. Da mit der 10. Konferenz in 20 Jahren gleich zwei Jubiläen zu feiern waren, versprach *Michael Höppner*, Direktor der Universitätsbibliothek Bielefeld, in seinem Grußwort nicht nur eine spannende Konferenz mit einem vielfältigen Programm, renommierten Rednern und lebhaften Diskussionen, sondern auch denjenigen, die an allen Konferenzen teilgenommen hatten, ein Jubiläumstrunk beim Konferenzdinner.

Stefan Gradmann (Institut für Bibliotheks- und Informationswissenschaft, Humboldt-Universität Berlin) skizzierte im eröffnenden Vortrag den Veränderungsdruck, dem bibliothekarisches Arbeiten in Zeiten des Semantischen Webs unterliegt. Wenn Bibliothekare hier zukünftig eine Rolle spielen wollten, müssten sie

1 siehe hierzu: Mons, Barend; Velterop, Jan: Nano-Publication in the e-science era, [http://www.w3.org/wiki/images/4/4a/HCLSS\\$ISWC2009\\$Workshop\\$Mons.pdf](http://www.w3.org/wiki/images/4/4a/HCLSS$ISWC2009$Workshop$Mons.pdf) (zuletzt besucht am 15.05.2012)

ihren Fokus wieder stärker auf die Inhalte ihrer Informationsobjekte richten und zu Wissensarbeitern werden. Davon werde die Arbeit in Bibliotheken, laut Gradmann, so stark beeinflusst sein, dass er am Ende seines Vortrags einen neuen Terminus für „Bibliothek“ forderte.

Auch *Birte Christensen-Dalsgaard* (Royal Library Copenhagen) legte den Fokus ihres Beitrags auf neue Aufgabenbereiche der Bibliothekare. Dienstleistungen im Bereich der Forschungsdaten spielen hier eine große Rolle, denn Bibliothekare stimulieren die Nachnutzung, indem sie in geeigneter Weise strukturierte Metadaten bereitstellen. Der gute Ruf, den sich Bibliotheken in der Vergangenheit erworben haben, sei eine Chance, sich auch bei der langfristigen Speicherung der Forschungsdaten zu bewähren und darüber hinaus den Zugang über persistente Adressierungsmechanismen sicherzustellen.

Rahmenbedingungen für den Aufbau von Informationsinfrastrukturen

Eine wichtige Unterstützerin beim Aufbau der Informationsinfrastruktur der Zukunft ist die Deutsche Forschungsgemeinschaft (DFG), deren Aktivitäten im Bereich „Wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS)“ *Anne Lipp* präsentierte. Viele der im Lauf der Konferenz vorgestellten Projekte finden mit Unterstützung der DFG statt.

Hans Geleijnse (Tilburg Universität) forderte eine noch stärkere Kooperation zwischen Bibliotheken auf nationaler und internationaler Ebene. Mit dem Rückgang traditioneller Dienstleistungen und vor dem Hintergrund des steigenden Kostendrucks sind Koordination und Arbeitsteilung im bibliothekarischen Verbundsystem Deutschlands eine Voraussetzung, um effizienter auf neue Herausforderungen zu reagieren.

Vor demselben Hintergrund fand an der ETH-Bibliothek Zürich ein Reorganisationsprozess statt, den der Leiter Wolfram Neubauer vorstellte. Ausgehend von den strategischen Zielen fand auf Basis einer Produktanalyse eine organisatorische Umstrukturierung statt, die den Weg von der analogen zur digitalen Informationsversorgung der Angehörigen der ETH Zürich ebnen soll.

Open Science und Open Access

Wahre und verlässliche wissenschaftliche Erkenntnisse können nur da entstehen, wo ein wissenschaftlicher Austausch herrscht, der auf gegenseitigem Vertrauen beruht. *Peter Weingart* (Universität Bielefeld) führte in seinem Vortrag eindrücklich vor Augen, dass das Vertrauen in der wissenschaftlichen Kommunikation auf Offenheit beruht und politische sowie wirtschaftliche Interessen ausschließt.

Alma Swan (Enabling Open Scholarship and Key Perspectives Ltd.) veranschaulichte anschließend den Nutzen von Open Access (OA). Neben den Forschern selbst profitieren vor allem auch ihre Institutionen. Ein Repositorium mit vielen Publikationen kann zu einer erhöhten Präsenz der eigenen Institution im Internet führen. Aufgabe der Universitäten ist nicht nur Erkenntnis selbst zu erbringen, sondern auch das gewonnene Wissen in effizienter Weise möglichst weit in die Gesellschaft zu tragen. Open Access unterstützt diese Aufgabe in wirkungsvoller Weise. Mit dem EU-Projekt OAPEN (Open Access Publishing in European Networks) soll der OA-Gedanke auch auf wissenschaftliche Monografien ausgeweitet werden. *Eelco Ferwerda* (OAPEN Foundation, Den Haag) stellte das hybride Modell mit kostenfreien aber auch bezahlpflichtigen Publikationen der Geistes- und Sozialwissenschaften, die den Peer-Review-Prozess durchlaufen haben, vor.

Die Datenflut bewältigen

Daan Broeder (Max Planck Institut für Psycholinguistik, Nimwegen) präsentierte das EU-Projekt EUDAT (European Data), das im Oktober 2011 initiiert wurde. Das Projekt hat das Ziel, eine effiziente, hochwertige sowie gesamteuropäische Dateninfrastruktur bereitzustellen, die einen Umgang mit der Datenflut über geografische und disziplinäre Grenzen hinweg ermöglicht.

Erstmalig gab es in diesem Jahr auch öffentliche Vorträge bei der Bielefeld Konferenz. Angehörige der Universität und Bürger der Stadt waren hierzu eingeladen.

Anne Trefethen (University of Oxford) schilderte in ihrem Beitrag „Drowning in Data?“ eindrücklich das rasante Wachstum der Daten auch im öffentlichen Bereich. Sowohl beim Einkaufen mit Karte als auch beim Einkaufen im Internet entstehen immer neue Daten. Plattformen für Fotos und Videos und Soziale Netzwerke tragen darüber hinaus zu einem stetigen Wachstum der Datenflut bei.

Ziel der Arbeit von *Anne Trefethen* ist die Schaffung eines Radioteleskops der nächsten Generation, das Datenmengen in der Größenordnung von Petabytes produzieren wird. Für die Bewältigung dieser enormen Datenmenge müssen nicht nur geeignete Algorithmen entwickelt werden, sondern auch die entsprechende Hardware.

Es ist eine Herausforderung bei dem zunehmenden Wachstum der wissenschaftlichen Publikationen diejenigen zu finden, die von Interesse und Relevanz für die eigenen Fragestellungen sind. Eine Möglichkeit, diesen Prozess zu erleichtern, stellte *Johanna McEntyre* (EMBL-European Bioinformatics Institute, Cambridge) vor: Die Verbindung zwischen Artikel und Datenbanken, die den Leser zu Daten aber auch Visualisierungen von z.B. Proteinen oder Molekülen führen. Diese Verbindung stellt entweder der Autor selbst her oder sie wird von Experten aus dem

Bereich der Datenpflege (Data Curator) geknüpft. Automatisierte Verfahren wie das Text Mining können diesen Prozess unterstützen.

Anreize und Unterstützung im Forschungsdatenmanagement

Der klassische Weg der wissenschaftlichen Publikation, der auf Basis der erhobenen Daten wissenschaftliche Erkenntnis generiert und nur letztere publiziert, wird zunehmend obsolet. Im Interesse einer guten wissenschaftlichen Praxis sollen auch die Daten offen zugänglich sein, damit die daraus abgeleiteten Erkenntnisse transparenter werden und die Daten nachgenutzt werden können. Hierzu sind geeignete Anreize und Unterstützung gerade auch für kleine Forschungsgruppen erforderlich.

Jan Brase (Technische Informationsbibliothek Hannover (TIB), DataCite) präsentierte die Aktivitäten des globalen Konsortiums DataCite. DataCite hat das Ziel, die Zitierbarkeit von Datensätzen zu ermöglichen, um Wissenschaftler zu motivieren, ihre Daten offen zur Verfügung zu stellen. Datensätze werden über eine persistente Adressierung (Digital Object Identifier) zitierbar und sind über einen von DataCite bereitgestellten MetadataStore verfügbar.

Daten zu publizieren ist ein aufwendiger Prozess, daher sucht *Philipp Cimiano* (Universität Bielefeld) bei seiner Arbeit nach einer Methodologie, die es gerade auch kleinen Forschungsgruppen ermöglicht, ihre Daten zu veröffentlichen. Eine zentrale Rolle schrieb Cimiano in seinem Vortrag dem Data Curator zu. Wenn dieser in der Lage ist, sich schnell einen Überblick über die Forschung des Wissenschaftlers inklusive der dabei anfallenden Daten zu verschaffen sowie über IT-Kenntnisse verfügt, ist er ein idealer Partner für die Veröffentlichung von Forschungsdaten im Semantischen Web.

Kooperationen zwischen Bibliotheken, Einbeziehung der Bibliotheksbenutzer

Patrick Hochstenbach (Universitätsbibliothek Gent) beschrieb die Zusammenarbeit der Universitätsbibliotheken Lund, Gent und Bielefeld in den Projekten Catmandu und LibreCat. Ersteres hat das Ziel, aufbauend auf einer bestehenden Repositoriumslösung, ein modulares System zu entwickeln, das von anderen Bibliotheken nachgenutzt werden kann; aus letzterem soll ein Repositoryum der nächsten Generation hervorgehen.

Heiner Stuckenschmidt (Universität Mannheim) bereicherte die Konferenz mit seiner Einschätzung aus der Perspektive eines CIOs. Mit mehreren DFG-Projekten hat die UB Mannheim den Versuch unternommen, neue Wege durch Einbeziehung der Bibliotheksbenutzer zu gehen. Die Möglichkeit, Medien im Katalog um Rezensionen zu ergänzen und zu bewerten, kann wertvolle Hinweise auf relevante

Medien liefern (DFG-Projekt „Weblogs als Steuerungsinstrument in Hochschulbibliotheken“). Darüber hinaus werden Bibliotheksbenutzer auch in die inhaltliche Erschließung einbezogen, indem sie Publikationen mit eigenen Stichworten annotieren können (DFG-Projekt „Collaborative Tagging als neuer Service von Hochschulbibliotheken“).

Aus dem Projekt FreeSearch, vorgestellt von *Wolfgang Nejd* (Universität Hannover), ging eine Suchmaschine für wissenschaftliche Publikationen hervor, die sich derzeit aus den Quellen DBLP, TIBKat, CiteSeer und Bibsonomy speist. Bei der Frage, wie Duplikate aus den unterschiedlichen Quellen bereinigt werden können, greift das Projekt FreeSearch sowohl auf automatische Duplikaterkennung als auch auf menschliche Unterstützung zurück. Duplikate, die nicht sicher als solche bestätigt werden können, werden über den Amazon-Dienst Mechanicalturk in Form von Microjobs an menschliche Interessenten weitergegeben mit der Besonderheit, dass der Algorithmus von dem Ergebnis profitiert und sich verbessert.

Liauw Toong Tjiek (Aditya Nugraha), Bibliotheksdirektor aus Indonesien, präsentierte das Repositorium der Petra Christian Universität in Surabaya. Die Inhalte sind sowohl textueller als auch visueller Art und schließen historische Fotos, Poster und farbenprächtige traditionelle Muster der unterschiedlichen ethnischen Gruppen Indonesiens ein. Ausgehend von dem Wunsch, dieses Angebot an möglichst viele Interessenten heranzubringen und der Beobachtung, dass Studierende heute ihre Information über Netzwerke mit Freunden und Bekannten teilen, stellen die indonesischen Bibliothekare ihre Inhalte über Applikationen in Sozialen Netzwerken bereit.

Das eigene Angebot kann auch durch Inhalte anderer angereichert werden. Hier zeigte *Liauw Toong Tjiek* eine Möglichkeit, die Buchcover anderer Seiten in die eigene Internetseite einzubeziehen, um z.B. die Aufmerksamkeit auf Neuerwerbungen zu lenken.

Linked Open Library Data

Wenn die gesamte Information nicht mehr vor Ort vorgehalten werden kann, wird die Zusammenarbeit der Bibliotheken zunehmend wichtiger. Das betonte bereits Heiner Stuckenschmidt in seinem Vortrag. Für den Austausch von Informationen müssen daher entsprechende Voraussetzungen geschaffen werden. Linked Data sind hier ein Ansatz, Informationsressourcen zu verbinden. Die Daten werden technisch so aufbereitet, dass eine semantische Verknüpfung möglich ist. Wenn sie darüber hinaus ohne Einschränkungen von jedermann nutzbar sind, dann handelt es sich um Linked Open Data.

Klaus Tochtermann (ZBW – Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg) präsentierte die Aktivitäten der ZBW. Bereits 2008 begann man dort

damit, das Retrieval durch Einbeziehung semantischer Technologien zu verbessern. Inzwischen liefern semantisch ähnliche Suchanfragen gleiche Ergebnisse. Durch eine Verbindung der Thesauri über Linked Open Data werden semantisch ähnliche Dokumente auch bei übergreifenden Suchanfragen in Repositorien anderer Institutionen gefunden. Für Informationssuchende an der Schnittstelle zwischen zwei Gebieten ergeben sich hier neue Services. Aktuell arbeitet die ZBW an Lösungen, die semantische Suche auch auf den Bereich der Forschungsdaten auszuweiten.

Das Hochschulbibliothekszentrum (HBZ) in Nordrhein-Westfalen stellt seine Daten ebenfalls als Linked Open Data zur Verfügung. *Adrian Pohl* vom HBZ betonte, dass die Daten dazu auch als Gesamtheit herunterladbar sein müssen. Im Anschluss präsentierte Pohl die Projekte *lobid.org* (Linking Open Bibliographic Data) und *culturegraph.org* des HBZ. Mit dem Dienst *lobid.org* sollen bibliografische Daten und verwandte Daten zu Linked Open Data konvertiert werden. *Culturegraph.org* ist ein Service, der dazu dient, gemeinsame Identifikatoren (Uniform Resource Identifier) für kulturelle Erzeugnisse – von Büchern und anderen Texten, über Gemälde, Skulpturen bis hin zu Musikstücken – zu verwenden, um eine verlässliche und persistente Referenzierbarkeit zu gewährleisten.

Wissenschaftliches Arbeiten in der Cloud

Jan Reichelt zeigte das Literaturverwaltungsprogramm *Mendeley* mit Netzwerkfunktionalität für Wissenschaftler. *Mendeley* ist ein Dienst, der das Sammeln von Artikeln und anderer Dateien im Internet unterstützt. Bei der Aufnahme erkennt das System zuverlässig bibliografische Angaben und übernimmt diese. Die Artikel können um Hervorhebungen ergänzt, annotiert und mit anderen geteilt werden. Weitere Funktionalitäten sind ein Empfehlungsdienst und die Bereitstellung statistischer Daten. Obgleich noch relativ jung verfügt *Mendeley* bereits über viele Nutzer und viele Angebote. Durch die Zusammenarbeit mit *Swets*, die *David Main* vorstellte, haben Forscher und Endnutzer die direkte Sicht auf eigene Bestände.

Der zweite öffentliche Vortrag von *Marc van den Berg* (Tilburg Universität) beschäftigte sich mit Chancen und Risiken des Cloud Computing für Bibliotheken. Bei der Auslagerung von Anwendungen, Daten und Rechenvorgängen in das Internet sind Fragen technischer und rechtlicher Art abzuwägen. Forschungsdaten der eigenen Institution sollten innerhalb nationaler oder europäischer Grenzen bleiben.

Fazit

Die Vortragenden aus Belgien, Dänemark, Deutschland, Indonesien, den Niederlanden, Norwegen, der Schweiz, dem Vereinigten Königreich und den Vereinigten Staaten boten einen ausgewogenen Blick aus unterschiedlichen Perspektiven auf aktuelle Aktivitäten, Informationsinfrastrukturen zu verbessern und neu zu schaffen. Es konnte gezeigt werden, dass es dafür Koordination und Kooperation auf internationaler, europäischer und nationaler Ebene bedarf. Die Zusammenarbeit zwischen einzelnen Bibliotheken und eine stärkere Einbeziehung der Bibliotheksbenutzer bieten darüber hinaus Synergien auf Mikroebene.

Das große wissenschaftliche, aber auch gesellschaftliche Interesse, Forschungsdaten für die Nachnutzung zugänglich zu machen, wurde in vielen Vorträgen betont. Im Bereich des Forschungsdatenmanagements entstehen neue Aufgabenfelder und Chancen für Bibliotheken. Gleichzeitig verfügen Bibliotheken auch selbst über hochwertige Inhalte wie Metadaten und Thesauri, die sie dem Semantischen Web zu bieten haben.

Das gesamte Programm der 10. Internationalen Bielefeld Konferenz mit Audiomitschnitten, Vortragsfolien und Kurzprofil der Vortragenden steht auf den Seiten im Internet unter <http://conference.ub.uni-bielefeld.de> bereit. Der Tagungsband wird auf dem Publikationsserver der Universitätsbibliothek Bielefeld veröffentlicht.

Ein Vortragender nutzte die Redewendung „the proof of the pudding is in the eating“. Welche der Visionen, Projekte und Diskussionen der 10. Bielefeld Konferenz zu dauerhaften Lösungen in der Praxis führen, darüber kann man sich während der nächsten Konferenz austauschen: in Bielefeld.