hbz

Wissen. Information. Innovation.

Digitale Wissenschaft

Stand und Entwicklung digital vernetzter Forschung in Deutschland

Ministerium für Innovation,
Wissenschaft und Forschung
des Landes Nordrhein-Westfalen

Digitale Wissenschaft

Stand und Entwicklung digital vernetzter Forschung in Deutschland

20./21. September 2010, Köln

Beiträge der Tagung

2., ergänzte Fassung

Herausgegeben von Silke Schomburg, Claus Leggewie, Henning Lobin und Cornelius Puschmann

2011, Köln

www.digitalewissenschaft.de

Inhaltsverzeichnis

Vorwort

Silke Schomburg

Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz)
Jülicher Str. 6
D-50674 Köln
schomburg@hbz-nrw.de

Das Hochschulbibliothekszentrum Nordrhein-Westfalen (hbz) hat als Einrichtung des Landes für Wissenschaftler und Bibliothekare die Tagung „Digitale Wissenschaft“ initiiert, maßgeblich organisiert und gemeinsam mit dem Kulturwissenschaftlichen Institut Essen und dem Zentrum für Medien und Interaktivität der Justus-Liebig-Universität Gießen ausgerichtet.[1] Als eine von sechs deutschen Verbundzentralen ist das hbz eine Einrichtung zur Erbringung von Dienstleistungen für die Hochschulen in Nordrhein-Westfalen und den größten Teil von Rheinland-Pfalz. Eine große Zahl von hbz-Produkten kommt den Wissenschaftlern unmittelbar für Forschung und Lehre zugute: Recherchieren sie nach Publikationen und neuesten Zeitschriftenartikeln, geben sie eine Fernleihbestellung auf, wollen sie eine neue elektronische Zeitschrift begründen oder Digitalisierungsprojekte aufsetzen – für all diese Dienste bietet das hbz vermittelt durch die Hochschulbibliotheken zuverlässige Unterstützung und innovative Lösungen an.

Hintergrund und Motivation

Was bewegt einen Informationsvermittler wie das hbz eine solche Tagung zu organisieren, die sich mit „Digitaler Wissenschaft“, also mit der Wissenschaft im digitalen Zeitalter, befasst?

Im Sommer 2009 haben sich die Kollegen im hbz mit der Planung von zukünftigen Diensten beschäftigt und eine Reihe von Thesen entwickelt sowie Tendenzen im Bereich der wissenschaftlichen Kommunikation herausgearbeitet. Intensivere Diskussionen dazu stehen noch aus. Das in diesem Zusammenhang erarbeitete „Perspektivpapier“ stellt für das hbz den Einstieg in die Diskussion über die Hauptfelder unserer zukünftigen Produkte und Dienstleistungen dar. Auch diese Konferenz bietet dem hbz und allen an ihr Beteiligten – seien es Vortragende oder Besucher – die Möglichkeit, aktiv an der Diskussion um die „Digitale Wissenschaft“ mitzuwirken.

[1] Dieser Text ist eine für diese Publikation leicht überarbeitete Fassung meiner Eröffnungsrede der Tagung „Digitale Wissenschaft 2010“. Ich möchte mich hier auch im Namen meiner beiden Chair-Kollegen, Claus Leggewie und Henning Lobin, herzlich bei Cornelius Puschmann und Adrian Pohl für ihren Einsatz bei der Vorbereitung und Organisation der Tagung bedanken. Agnes Maika ist zudem an dieser Stelle für die redaktionelle Bearbeitung der vorliegenden Beiträge zu danken.

Editorische Anmerkung: Die Herausgeber waren wie die Autoren an einer zeitnahen Publikation des Tagungsbandes interessiert, um die Diskussion weiter zu befördern. Aus diesem Grund wurden die einzelen Texte nicht in allen formalen und stilistischen Aspekten vereinheitlicht. Aus Gründen der besseren Lesbarkeit wurde jedoch einheitlich bei der Nennung von Personen nur die männliche Form verwendet. Wir weisen an dieser Stelle ausdrücklich darauf hin, dass diese Form auch weibliche Personen umfasst.

Wo kommen die Verbundzentralen als Dienstleister der Wissenschaft und ihrer Bibliotheken fachlich gesehen her? Wo stehen sie derzeit und wohin wollen sie sich entwickeln? Alle Dienste, die wissenschaftliche Bibliotheken der Wissenschaft in Kooperation mit Verbundzentralen leisten, drehen sich um zitierfähige wissenschaftliche Schriften, um Bücher, Zeitschriften und Artikel, die als Druck oder in elektronischer Form vorliegen können. Das hbz und seine Kunden bewegen sich somit im Bereich der schriftbasierten wissenschaftlichen Kommunikation und befassen sich hauptsächlich mit jenen Texten, die als formale, d. h. allgemein institutionell gebilligte wissenschaftliche Schriften kategorisiert werden können. Blogbeiträge, Vortragsfolien oder Wikieinträge gehören – aufgrund der ihnen oft nicht zugestandenen Zitierfähigkeit – (noch) nicht zu unserem Metier. Genauso wenig wie die Verfügbarmachung von Forschungsdaten, da es bisher in der Regel nicht üblich war, diese zu publizieren und langfristig zu archivieren. Dieses aus unserer Sicht wichtige Thema und die überfälligen Veränderungen im Bereich der Forschungsdaten sind nicht ohne Grund Gegenstand einer eigenen Session dieser Tagung.

So ist die intensive Auseinandersetzung mit dem Fortgang der Entwicklung des digitalen Zeitalters im Allgemeinen und den Perspektiven wissenschaftlicher Bibliotheken und von Bibliotheksverbünden im Speziellen als Desiderat anzusehen. Unter diesem Vorzeichen sehen wir auch die derzeit stattfindende Evaluierung der Verbundsysteme durch den Wissenschaftsrat.

Aus unserer Sicht sind die Rolle und die Aufgaben von Informationseinrichtungen untrennbar an die Art und Weise wissenschaftlicher Kommunikation und Publikation gekoppelt, wie auch die Effektivität wissenschaftlichen Arbeitens von intakten Informationseinrichtungen abhängt. Deshalb ist der Erfahrungsaustausch zwischen Wissenschaft, Bibliothek und Dienstleistern ebenso geboten wie auch die Koordination ihrer Aktivitäten.

Das digitale Zeitalter

Digitale Wissenschaft. Wissenschaft im Digitalen Zeitalter. Der Ausdruck „Digitales Zeitalter“ wird häufig zur Beschreibung des grundlegenden Wandels benutzt, den wir zeitlich spätestens seit den 1990er Jahren – mit der Ausbreitung des World Wide Webs – ansetzen. Wie lässt sich das „Digitale Zeitalter“ näher fassen?

Die Ursachen des stattfindenden Wandels liegen in der Ubiquität von Computer und Internet. Information ist heutzutage nicht mehr an körperliche Trägermedien gebunden und kann beliebig und nahezu kostenlos kopiert und distribuiert werden. Medien- und kommunikationsintensive Bereiche sind dementsprechend als erste von der Transformation erfasst worden. Klassische Distributionsbereiche wie die Musikindustrie, Verlage und auch Bibliotheken spüren die Auswirkungen seit Langem.

Die Entfaltung des digitalen Zeitalters stellt einen Großteil bestehender funktionierender Praktiken in Frage und rührt am Selbstverständnis von Informationseinrichtungen wie auch von ganzen wissenschaftlichen Disziplinen. Gleichzeitig bietet ein solcher Wandel nahezu unbegrenzte Chancen der Gestaltung, Weiterentwicklung und Neuorientierung.

Eine verbreitete Folge dieses Wandels ist bei vielen Akteuren eine Trauer um überkommene Praktiken, die nun zu erodieren drohen, verbunden mit der Suche nach Perspektiven,

nach Optimierungsmöglichkeiten und Innovationspotential. Und eben diesen Zwecken soll die Tagung „Digitale Wissenschaft“ dienen: den Beteiligten durch gegenseitigen Austausch Orientierung geben, potentielle Kooperationspartner zusammenführen und mögliche gemeinsame Wege und vielversprechende Handlungsbereiche aufzeigen.

Bereits stattfindende Veränderungen

Bevor thesenhaft Zukunftsperspektiven aufgezeigt werden, soll zunächst der Blick auf fünf wichtige Bereiche gerichtet werden, in denen bereits signifikante Entwicklungen stattgefunden haben:

1. ***Die Proliferation digitaler Medien.*** Wikis, Weblogs, Microblogging, elektronische Volltexte und einiges mehr erweitern das Feld der potentiell für Informationseinrichtungen relevanten Informationsressourcen. Dazu kommt die ständig wachsende Menge digitalisierter Printressourcen. Recherchierende wünschen sich in Folge dieser Entwicklung als Ergebnis ihrer Informationssuche den direkten Zugriff auf Volltexte. Gedächtnisinstitutionen führen seit einigen Jahren mit der Unterstützung der Deutschen Forschungsgemeinschaft (DFG) umfangreiche Digitalisierungsprojekte durch und auch Google stellt mit seinem Google-Books-Programm bereits kostengünstig digitale Inhalte in großem Umfang zur Recherche und zum Download bereit.

2. ***E-Learning.*** Mit dem E-Learning findet eine Weiterentwicklung der Lehre, die Veränderung der Lehr- und Lernpraktiken in der digitalen Sphäre statt. Als Beispiele sind hier elektronische Semesterapparate, Videomitschnitte von Vorlesungen und Wikis zu nennen. Studentische Kollaboration findet zudem zunehmend über das Internet – in virtuellen Lernzusammenhängen – statt.

3. ***Neue Wege der Informationsversorgung.*** Im Bereich von Informationsaggregation und Recherchedienstleistungen findet sich mittlerweile eine Vielzahl von Akteuren. In grossen Teilen werden Wissenschaftler und Studierende inzwischen durch Suchmaschinen wie Google, Online-Buchhändler wie Amazon oder die Online- Enzyklopädie Wikipedia versorgt. Sie nutzen Google häufig als Rechercheeinstieg und Wikipedia lässt sich aus ihrer täglichen Arbeit nicht mehr wegdenken. Auch fachspezifische von den Wissenschaftlern selbst geschaffene Wissenspools wie z. B. *arxiv.org* sind in diesem Zusammenhang zu nennen. Dazu gekommen sind in den letzten Jahren kollaborative Literaturverwaltungsanwendungen und soziale Netzwerke für Akademiker wie *Zotero, Mendeley, ResearchGate* oder *academia.edu.*

4. ***Urheberrecht.*** Die Content-Industrie, zu der die Verlage, die Film- und die Musikindustrie zählen, reagiert auf die stattfindende Vervielfältigung und Verbreitung von – auch urheberrechtlich geschützten – Inhalten bisher häufig mit Forderungen an die Legislative nach einem *verschärften Urheberrecht* sowie mit der Implementierung von Maßnahmen des *Digital Rights Managements,* die die Nutzung elektronischer Inhalte komplizierter und teurer machen. Die durch den technischen Fortschritt verbesserten Möglichkeiten der Informationsverbreitung stehen somit in direktem Gegensatz zu den rechtlichen Restriktionen der Weitergabe kultureller und wissenschaftlicher Inhalte.

5. ***Open-Access.*** Urheberrechtsproblematik und stark steigende Zeitschriftenpreise haben als Gegenreaktion in den letzten zehn Jahren die Open-Access-Bewegung entstehen lassen. Während dieses Publikationsmodell in vielen Naturwissenschaften schon weit verbreitet ist, stellen zunehmend auch Sozial- und Geisteswissenschaftler ihre Inhalte frei zugänglich zur Verfügung, sei es auf ihrer Homepage, in Open-Access-Journals oder als Pre- bzw. Postprint in Online-Repositorien. Förderorganisationen wie die DFG knüpfen mittlerweile Projektförderungen an die Bereitschaft der Forscher zur Open-Access-Publikation ihrer Ergebnisse.

Vier Thesen zur weiteren Entwicklung

In Ergänzung der genannten Entwicklungen, die massiv Wissenschaft, Bibliotheken und Informationseinrichtungen beeinflussen werden, seien abschließend vier Thesen zur Zukunft der digitalen Wissenschaft formuliert. Denn so viel ist sicher: Die Wissenschaft von morgen wird nicht so aussehen wie die printbasierte Wissenschaft von gestern. Es ist aus unserer Sicht vorgezeichnet, dass Monographien und Zeitschriften nicht einfach durch E-Books und E-Journals ersetzt werden, ohne dass die Praxis der wissenschaftlichen Kommunikation unverändert bleibt.

These 1: Die Dominanz von Zeitschriften und Monographien wird aufgelöst werden. Neue digitale Medien und Publikationsformen werden zunehmend institutionell anerkannt werden.

Die Stellung von Zeitschriften und Monographien als einzige Kanäle der formalen Wissenschaftskommunikation ist nicht länger unangefochten, schriftliche Kommunikation in Blogs, Wikis und anderen Medien wird immer mehr an Relevanz gewinnen. Erste erfolgreiche Initiativen zur Berücksichtigung neuer Formate wie Blogs und Wikis bei Berufungsverfahren hat es in den USA bereits vereinzelt gegeben. Auch zeichnet sich ab, dass Forschungsdaten zukünftig als Publikationsform bzw. als Teil von Publikationen eine wichtige Rolle in einer Reihe von Wissenschaftszweigen spielen werden.

These 2: Die Lösung des Problems der Langzeitarchivierung digitaler Inhalte wird die Hauptaufgabe für Informationseinrichtungen sein.

Die Verbreitung der digitalen Medien und die zunehmende Publikation von Forschungsdaten führten zum Entstehen des neuen Problemfelds der digitalen Langzeitarchivierung, der langfristigen Speicherung und Verfügbarhaltung von digitalen Inhalten, das in den letzten Jahren zunehmende Aufmerksamkeit erhalten hat. Ein großer Teil der Wissensproduktion und kultureller Erzeugnisse der vergangenen zwanzig Jahre ist vom Verschwinden und Vergessen bedroht, wenn nicht nachhaltige Konzepte digitaler Langzeitarchivierung entwickelt und umgesetzt werden. Auch für die in vielen Bereichen in großen Mengen anfallenden Forschungsdaten muss eine nachhaltige Lösung gefunden werden. Publikationen in völlig neuen Medienformen (Daten, Wikis etc.) müssen von Informationseinrichtungen gesammelt, archiviert und zugänglich gemacht werden. Die Lösung des Langzeitarchivierungsproblems von Webinhalten wird im besten Fall auf der Basis allgemeiner, offener Standards angegangen werden.

These 3: Linked-Data-Standards für die Publikation von Daten werden sich durchsetzen. Die Integration von anderswo generierten Daten in eigene Zusammenhänge und die Kombination von Daten aus verteilten Quellen wird durch die Etablierung allgemeiner Standards der Datenpublikation erleichtert werden. Insbesondere für Informationseinrichtungen, aber auch für datenintensive Forschungsbereiche bieten Linked-Data-Standards enorme Möglichkeiten der Datenintegration aus verschiedensten Quellen.

These 4: Die Gesetzgebung wird keine zufriedenstellende Lösung der Urheberrechtsproblematik liefern. Die Verbreitung standardisierter offener Lizenzierungspraktiken – zusammengefasst unter dem Begriff „Open Knowledge" – wird zu neuen Möglichkeiten der Erkenntnisgewinnung und wissenschaftlichen Kommunikation führen.
Damit wissenschaftliche Inhalte und Daten schneller und leichter verbreitet und weiterverwendet werden können, ist für die Stärkung der Rechte von Urhebern und der Öffentlichkeit einzutreten. Das langfristige Ziel ist, die Gesellschaft von einem freien Fluss von Wissen profitieren zu lassen. Zwar tut sich mit den Verhandlungen zum Dritten Korb bei der Gesetzgebung etwas im Bereich des Urheberrechts. Allerdings sollten – gerade auf der Basis vergangener Erfahrungen – die Erwartungen an den Gesetzgeber nicht zu hoch sein. Anstatt auf eine zukunftsfähige Gesetzgebung zu hoffen, bieten sich bereits auf der Basis des bestehenden Urheberrechts vielversprechende Handlungsperspektiven, die sich unter dem Stichwort „Open Knowledge" zusammenfassen lassen. Das hbz ist in der Open Knowledge Foundation aktiv, die sich für die offene Lizenzierung von Inhalten und Daten einsetzt, um deren Nachnutzbarkeit und Kompatibilität rechtlich zu gewährleisten. Mit der Freigabe eines großen Teils des hbz-Verbundkatalogs in die Public Domain in Zusammenarbeit mit den entsprechenden Bibliotheken im März 2010 haben das hbz und einige kooperierende Verbundbibliotheken Neuland betreten und einen großen Schritt in diese Richtung gemacht.

In Zukunft werden – so die hoffnungsfreudige These – andere Informationseinrichtungen sowie Wissenschaftler und Forschungsinstitutionen diesem Schritt folgen und ebenfalls Daten und Inhalte der Allgemeinheit zur freien Nutzung zur Verfügung stellen.

Ebenso wie die technische Kompatibilität dezentral produzierten Wissens mit Linked-Data-Standards hergestellt werden wird, werden Open-Data-Standards zunehmend die rechtliche Nachnutzbarkeit und Kompatibilität von Webpublikationen – seien es Daten oder Texte – sicherstellen. Open Knowledge im allgemeinen und Open Data im speziellen werden zunehmend – gerade im Bereich öffentlich geförderter Forschung und öffentlicher Informationseinrichtungen – an Bedeutung gewinnen und die rechtliche Kompatibilität verteilt produzierten, publizierten und archivierten Wissens sicherstellen. Dadurch wird eine offene und durchlässige Publikationslandschaft entstehen, die – auf der Basis des bestehenden Urheberrechts – Bestrebungen nach Zugangsbeschränkungen und Urheberrechtsverschärfungen eine Praxis offenen Wissens entgegensetzt.

Ich schließe mit dem Hinweis darauf, dass uns bei all den Diskussionen im Haus zum Thema „Digitale Wissenschaft" derzeit noch keine echte, d. h. virtuelle Alternative zu einer Tagung eingefallen ist, kein Tool und keine Technik, die das gemeinsame Diskutieren und Erleben, die vielzitierten Gespräche in der Kaffeepause ersetzen können – deswegen die Durchführung der Tagung „Digitale Wissenschaft".

Digital Humanities

Die Digitalisierung des Verstehens

Hanno Birken-Bertsch

birkenbertsch@gmail.com

Zusammenfassung

Um Orientierung zu gewinnen, werden drei Aspekte der Herausforderung der Geisteswissenschaften durch das Aufkommen und Vordringen digitaler Techniken der Datenverarbeitung exemplarisch vorgestellt. In praktischer Hinsicht verändert sich die Forschung durch die Googlebuchsuche und ähnliche Angebote. Allerdings war mit Roberto Busa schon sehr viel früher ein Geisteswissenschaftler am Geschehen beteiligt als man zu glauben gewöhnt ist. Und drittens ergeben sich aus Verfahren wie der LSA Fragen nach der Digitalisierung des Verstehens.

Drei Aspekte digitaler Wissenschaft und die Geisteswissenschaften

Der Slogan lautet „digitale Wissenschaft“. Dabei steht „Wissenschaft“ für jene rationale Form der Erkenntnis, die innerhalb der letzten zweitausend Jahre in der westlichen Welt entstanden ist. „Digital“ kommt von lateinisch „digitus“, deutsch „Finger“, und die Idee dahinter scheint das Zählen mit den Fingern zu sein. Auch messen kann man als zählen verstehen, denn wenn wir etwas messen, können wir die gemessenen Einheiten abzählen. Also können wir „digital“ in seiner ersten Bedeutung als „quantitativ“ oder „quantifizierend“ übersetzen. In diesem Sinne sind die Naturwissenschaften seit Galilei und Newton digital, die Geisteswissenschaften dagegen nicht.

In einer zweiten Bedeutung steht „digital“ für einen Modus der Datenverarbeitung. Daten werden entweder analog oder digital verarbeitet. Geschieht es analog, so besteht eine Analogie, eine Entsprechung, zwischen dem Gegenstand der Daten und der Art der Datenspeicherung bzw. zwischen dem realen Vorgang und seiner Verarbeitung. Umso tiefer die Rillen in eine Schallplatte geschnitten werden, um so lauter ist der Ton zu hören. Bewegt sich der Stift auf dem Zylinder eines analogen Gezeitenrechners nach unten, sinkt auch der Wasserspiegel, bis Ebbe eingetreten ist. In einem digitalen Speichermedium wie der CD gibt es dagegen keine der Rillentiefe vergleichbare körperliche Entsprechung zu den physikalischen Eigenschaften des Klangs, und ein digitaler Computer verarbeitet nur Folgen von Nullen und Einsen; sein Weg zum Ergebnis hat, anders als der des analogen Computers, keine Ähnlichkeit mit den realen Prozessen, über die das Ergebnis Erkenntnis verschafft.

Es ist zweifellos diese zweite Bedeutung von „digital“, die im Slogan von der digitalen Wissenschaft herangezogen wird. Es geht um eine Wissenschaft, die von den Techniken der digitalen Datenverarbeitung Gebrauch macht. Dennoch hilft es, die erste Bedeutung von „digital“ im Blick zu behalten, denn wenn eine Wissenschaft sich bereits ohnehin auf quantitative Daten stützt, kann sie diese auch leichter digital verarbeiten. Ihr wird die digitale Datenverarbeitung willkommen sein.

Für die Geisteswissenschaften dagegen stellt das Aufkommen und Vordringen digitaler Techniken der Datenverarbeitung seit den ersten digitalen Computern im Jahr 1941 eine große Herausforderung dar. Ich möchte drei Aspekte der Herausforderung vorstellen. Der erste ist eher *praktischer* Natur: Die Googlebuchsuche und ähnliche Angebote im Internet verändern auch die geisteswissenschaftliche Arbeit (1). Der zweite Aspekt ist die wenig bekannte, aber tatsächlich *vorhandene Tradition* geisteswissenschaftlicher Arbeit mit digitalen Computern – eine Tradition, die auf das Jahr 1949 zurückgeht und deren Geschichte damit fast so weit zurückreicht wie die des digitalen Computers selbst (2). Der dritte Aspekt ist theoretischer Natur. Es ist die Frage, ob nicht mit einer Rückwirkung der Digitalisierung auf unsere *Auffassungen* vom Verstehen und von geisteswissenschaftlicher Arbeit zu rechnen ist (3).

Auch die Geisteswissenschaften: Digitale Kataloge und allgemeine digitale Suchmaschinen

Seit 1991 gibt es das World Wide Web (WWW), seit 1998 durchsuchen wir dieses World Wide Web mit Google und seit 2005 können wir auch in digitalisierten Büchern suchen. Diese und ähnliche Entwicklungen machen den ersten Aspekt der Herausforderung der Geisteswissenschaften durch die digitale Datenverarbeitung aus, auf den ich eingehen möchte.

Ich weiß nicht, ob es empirische Forschung zu der Revolution gibt, die da stattfindet. Es wäre zu rekonstruieren, wie ein geisteswissenschaftlicher Forscher noch in den 80er Jahren des 20. Jahrhunderts gearbeitet hat, und sich dann ansehen, wie ein Forscher dieselben Aufgaben heute bewältigt. Körperliche Nähe zu den Datenquellen war damals vonnöten. Man war gezwungen sich in der Bibliothek aufzuhalten, die dortigen Nachschlagewerke nutzen und Bücher per Leihschein mit einer Signatur aus dem Zettelkatalog bestellen. Wie schnell sich eine kleine Frage beantworten ließ, hing ab von dem Bestand der jeweiligen Bibliothek, von der Geschwindigkeit, mit der Ausleihwünsche bearbeitet wurden und von der eigenen Findigkeit. Für speziellere Fragen waren Bibliographien zu konsultieren, deren Benutzung einen als fachkundig auswies.

Heute lassen sich sehr viele Fragen schon vom eigenen Schreibtisch aus beantworten. Manche Fachfrage ist durch die Eingabe einiger weniger Schlüsselausdrücke in eine der gängigen Suchmaschinen zu klären. Bei bibliographischen Fragen nutzt man die großen Bibliothekskataloge oder eine Metasuchmaschine wie den Karlsruher Virtuellen Katalog, mit der in einer Suchanfrage viele Kataloge gleichzeitig durchsucht werden. Aber sehr häufig wird man vor allem bei älteren Titeln auf die Bibliothekskataloge gar nicht mehr zuerst zugreifen, sondern nur dann, wenn keine elektronische Version – meist eine PDF- oder DjVu-Datei – zu finden war. Solche digitalisierten Bücher gibt es nicht nur bei Google, sondern auch im Internet-Archiv, bei Wikisource sowie in vielen Sammlungen von Bibliotheken oder Forschungsinstituten, die über das Netz verstreut sind. Die Qualität der Digitalisate, die Google ins Netz stellt, ist meist bescheiden, und die Metadaten zu den Büchern – also in etwa das, was im Bibliothekskatalog über das Buch vermerkt ist – sind immer wieder unvollständig oder fehlerhaft. Sehr viel besser steht es in dieser Hinsicht um die Digitalisate einer Sammlung wie beispielsweise der des Virtuellen Labors des Max-Planck-Instituts für Wissenschaftsgeschichte in Berlin. Allerdings kommt mit diesen separaten Sammlungen wieder

jene Fachkenntnis ins Spiel, die durch die hohen Erfolgschancen der allgemeinen Suchverfahren neutralisiert worden war.

Was für Auswirkungen haben diese Veränderungen mit sich gebracht? Birgt das Aufkommen digitaler Bibliothekskataloge, der Internet-Recherche und digitalisierter Bücher *überhaupt* eine Herausforderung? Ich werde sehr vorsichtig sein, wenn es darum geht, die Auswirkungen der skizzierten Veränderungen einzuschätzen, denn ins Prinzipielle reichende Veränderungen sind selten und die, die meinen, es gebe nichts Neues unter der Sonne, haben, wenn die Sonne nur steil genug vom Himmel sticht, immer recht.

Sicher scheint mir aber zu sein: Ein empirischer Vergleich zwischen der Arbeit eines Geisteswissenschaftlers vor zwanzig oder dreißig Jahren und heute würde zeigen: Der Aufwand, an ein Buch oder eine Information zu gelangen, ist gesunken, Mühen sind weggefallen, und der physisch-praktische Anteil an der Arbeit hat abgenommen. Umgekehrt heißt das: Wenn es weniger Mühe macht, Bücher oder Informationen zu finden, kann man damit auch nicht mehr brillieren. Das Anforderungsprofil für einen Geisteswissenschaftler verschiebt sich von Fleiß und Erfahrung zu Argumentationsfähigkeit und Originalität.

Jede Aussage ist in einer vorher nicht dagewesenen Weise überprüfbar. Wenn man etwa bei Peter Sloterdijk liest, der Ausdruck „Anthropotechnik“ sei zum ersten Mal 1926 in einer sowjetischen Enzyklopädie aufgetaucht (Sloterdijk, 2010, 24 Anm. 6), dann stellt sich diese Aussage nach einer kurzen Recherche mit Googlebuchsuche als Irrtum heraus. Der Ausdruck „Anthropotechnik“ kommt bereits 1907 in wissenschaftlichen Zeitschriften wie der „Zeitschrift für experimentelle Pädagogik“ vor. Diese neue Überprüfbarkeit betrifft alle Aussagen über das erste Vorkommen eines Wortes. Was man in einem Wörterbuch diesbezüglich findet, steht gewissermaßen unter dem Vorbehalt, es könne via Googlebuchsuche ein früheres Auftreten ausfindig gemacht werden.

Auch orthographische Fragen lassen sich elektronisch klären, indem man die Häufigkeit der konkurrierenden Schreibungen miteinander vergleicht. Man braucht nur, wenn man sich in Erinnerung rufen möchte, welche Schreibung von „selbstständig“ man einmal gelernt hat, die beiden Alternativen – „selb-ständig“ und „selbst-ständig“ – in eine Suchmaschine einzugeben und die Zahl der Treffer miteinander zu vergleichen.

Vielleicht ist für die Geisteswissenschaften dennoch in prinzipieller Hinsicht nichts passiert. Da aber die geisteswissenschaftliche Recherche leichter geworden ist, brauchen ihre Ergebnisse viel weniger auf Treu und Glauben angenommen werden. Sie können vielmehr in vielen Fällen auch von Laien überprüft werden, ja, es kann sogar auf sie verzichtet werden – warum ein Wörterbuch benutzen, dessen Aussagen dann ohnehin noch zu überprüfen sind?

(Die Rechtschreibreform von 1996 kam also just in dem Moment, in dem auch orthographische Lexika ihren Nimbus zu verlieren begannen. Gleichzeitig wurden die Textverarbeitungsprogramme mit immer besseren Rechtschreibprüfungen ausgestattet, was die Schwierigkeit der Orthographie für den Schreibenden ganz erheblich entschärft hat. Die Rechtschreibreform kam, als sie überflüssig geworden war.)

Wie aber kann der Effekt, den die Veränderungen – vom World Wide Web bis zur Googlebuchsuche – für die Geisteswissenschaften gebracht haben, auf den Punkt gebracht werden? Ich glaube, dieser Effekt ähnelt dem der Globalisierung. Eigentlich hat sich, sagen die Fachleute in beiden Fällen, prinzipiell nichts geändert. Die Dinge sind jedoch näher herangerückt, durch die Globalisierung wie durch die Digitalisierung. Was vorher räumlich weit weg war, ist nun elektronisch greifbar – aber nicht nur für mich, sondern auch für alle anderen.

Der spezielle Weg der Geisteswissenschaften: Computerphilologie und Digital Humanities

Während der erste Aspekt der Herausforderung der Geisteswissenschaften aus der Anwendung von allgemein zugänglichen Instrumenten erwächst, besteht der zweite in einer Entwicklung innerhalb der Geisteswissenschaften: Es sind spezielle Instrumente geschaffen und hochwertige Datenbanken aufgebaut worden.

Ein Pionier – oder gar der Pionier – dieser Entwicklung war der italienische Philosoph Roberto Busa (Hockey, 2006, 4; Winter, 1999). Busa wollte untersuchen, was Thomas von Aquin zum Begriff der Gegenwart, lateinisch „praesens“ bzw. „praesentia“, sagt. Dazu genügt es, wie er feststellte, nicht, die Stellen zu sammeln, an denen dieser die Worte „praesens“ und „praesentia“ benutzt. Ebenso heranzuziehen sind die Stellen, an denen Thomas von Aquin die Präposition „in“ gebraucht (Busa, 1980).

Bemerkenswerterweise richtete sich also das Bedürfnis nach digitaler Textverarbeitung nicht zuerst auf die seltenen Worte, sondern auf die gewöhnlichen. Es galt nicht primär dem selbständig bedeutungstragenden Substantiv „praesentia“, das auch der fleißige Leser leicht erkennt und zu dem konventionell angelegte philosophische Wörterbücher Auskunft geben, sondern der erst in einer Konstellation von Wörtern bedeutsamen Präposition „in“.

Roberto Busa hat sich die Stellen mit „in“ für seine Dissertation noch per Hand heraussuchen müssen. Nachdem er seine Dissertation 1946 verteidigt hatte, begann er mit dem Projekt des Index Thomisticus, eines Index aller Wörter in den Schriften Thomas von Aquins. Von 1949 an wurde er von der amerikanischen Firma IBM unterstützt, mit Lochkarten, Lesemaschinen und Computern. In den 70er Jahren wurde der Index in 56 gedruckten Bänden veröffentlicht. Die Werke des Thomas von Aquin gab Busa zuerst 1980 gedruckt und dann 1991 als CD-ROM heraus. Heute sind sie als Corpus Thomisticum im World Wide Web frei zugänglich.

Die Bemühungen Roberto Busas um den Text der Werke des Thomas von Aquin erstrecken sich also über mehr als 50 Jahre und illustrieren die technologische Entwicklung von den ersten digitalen Rechnern bis zum Internet. In diesen Jahrzehnten wurde auch eine Fülle von anderen Projekten in verschiedenen geisteswissenschaftlichen Disziplinen durchgeführt. Der Platz dieser Projekte blieb jedoch immer am Rande, fernab der großen Lehrstühle der geisteswissenschaftlichen Fächer: „[…] literary computing has, right from the very beginning, never really made an impact on mainstream scholarship“ (Rommel, 2006, 92). Auch wenn Busa 1952 so kühn war, von einer „Mechanisierung der philologischen Analyse“ zu sprechen (Busa, 1952), blieb der Einsatz der digitalen Datenverarbeitung den Geisteswis-

senschaften äußerlich: Es ist damit nach allgemeiner Auffassung das Eigentliche geisteswissenschaftlichen Bemühens noch nicht erreicht. Digital sind nur die Hilfswissenschaften, die durch das Bereitstellen von Texten, Indices und Bildsammlungen die großen Interpretationen vorbereiten.

Dieser zweite Aspekt, die digitale Tradition im eigenen Haus, wird für die Geisteswissenschaften allein noch nicht zur Herausforderung. Diese digitale Tradition könnte jedoch Teil einer Tendenz werden – der Tendenz, den Akzent geisteswissenschaftlicher Aufmerksamkeit vom Individuellen aufs Generelle zu verlagern. Vielleicht hat diese Verlagerung sogar schon stattgefunden. Der Aufwand, einen individuellen Text, mit den bisherigen Interpretationen als nicht zu unterschreitender Vorgabe, neu zu verstehen, hat sich vergrößert. Zu einem immer höheren Anteil besteht die geisteswissenschaftliche Arbeit daraus, die generellen Bedingungen zu bestimmen, die zu einer bestimmten Zeit geherrscht haben oder denen die Akteure an einem bestimmten Ort unterworfen sind.

Das ist nur eine erste Beobachtung, die zu erhärten wäre. Träfe sie zu, würden sich die Geisteswissenschaften in einer Weise wandeln, die sie für ihre eigenen digitalen Subdisziplinen öffnen würde, denn in diesen nähert man sich dem Einzelnen vom Allgemeinen her. Die Herausforderung erwüchse aus der genannten Verschiebung der Aufmerksamkeit, nicht aus der Anwendung digitaler Datenverarbeitung allein. Allerdings dürfte diese Anwendung einer der Faktoren sein, welche die Verschiebung der Aufmerksamkeit aufs Generelle hervorgebracht haben.

Die digitale Simulation des Verstehens

Der dritte Aspekt der Herausforderung der Geisteswissenschaften durch die Digitalisierung besteht aus den Rückwirkungen, die die bisher skizzierten Entwicklungen auf die Geisteswissenschaften haben können. Alle diese Entwicklungen bringen ein Vordringen statistischer Betrachtungsweisen mit sich. Eine Gruppe von Programmen sticht aber heraus. Aus dieser Gruppe nehme ich die sogenannte Latent Semantische Analyse als Beispiel. Die Bezeichnung ist allerdings irreführend, weil der Clue des Programms gerade darin besteht, das Semantische zu umgehen. Ich werde daher die Abkürzung „LSA“ als Bezeichnung verwenden.

Ausgangspunkt war für die Entwickler der LSA, zu denen u. a. Susan T. Dumais und Thomas K. Landauer gehörten, eine durchaus alltägliche Erfahrung: Wir suchen nicht nach Wörtern, sondern nach Inhalten (Deerwester et al., 1990, 391). Die Wörter, mit denen wir notgedrungen nach Inhalten suchen, stehen zu diesen Inhalten jedoch nicht in einem eineindeutigen Verhältnis. Im Gegenteil, für einen Inhalt kann es mehrere Wörter geben, und ein Wort kann mehrdeutig sein und sich auf verschiedene Inhalte beziehen.

Die Annahme, auf der LSA aufbaut, ist eine schlichte: Es gibt trotz fehlender Eineindeutigkeit hinreichend stabile Verhältnisse zwischen den Wörtern, um diese aufklären zu können. Diese Relationen bezeichnen die Autoren des ersten Aufsatzes zum Thema im Jahr 1990 als „semantic structure“ (Deerwester et al., 1990, 391 Anm. 1). Zu diesen Verhältnissen gehört zuerst das gemeinsame Vorkommen von Wörtern, dann aber auch so etwas wie die Gemeinsamkeiten von zwei Wörtern, die, obwohl sie vielleicht nie zusammen vorkom-

men, häufig gemeinsam mit bestimmten anderen Wörtern anzutreffen sind. Kommt das Wort „Hund“ in einem Text vor, dann findet man darin seltener das Wort „Tier“; die Wörter um „Hund“ und „Tier“ herum sind jedoch zu einem Gutteil die gleichen (Louwerse & Ventura, 2005, 302).

Diese Verhältnisse, diese sogenannte „semantische Struktur“ wird mathematisch dargestellt. Dumais, Landauer und Kollegen spekulieren nicht über den möglichen Status dieser Verhältnisse, sondern konzentrieren sich auf die Anwendung ihres Programms. Dabei stammen die markantesten Resultate vielleicht gar nicht von ihnen, sondern von den Psychologen Max Louwerse und Rolf Zwaan. Sie haben einen Korpus von amerikanischen Zeitungsartikeln auf die Namen der größeren Städte hin untersucht (Louwerse & Zwaan, 2009). Es handelt sich um gewöhnliche Texte, in denen die Namen von Städten erwähnt werden, weil von Ereignissen in diesen Städten berichtet wird, aber die geographischen Relationen der Städte untereinander nicht thematisch sind. Genau diese geographischen Relationen aber konnten den Texten entnommen werden. Die Texte machten keine Aussage darüber, wieweit New York und San Fransisco auseinanderliegen, sie handeln bloß von dem, was in diesen Städten passiert, mit LSA konnte ihnen jedoch solche geographischen Information abgerungen werden – nicht in absoluten Größen, sondern als Relationen untereinander und zu weiteren Städten.

Erstaunlich ist meines Erachtens vor allem dies: Texten – also Gegenständen geisteswissenschaftlicher Forschung – können Informationen entnommen werden, die sich gewöhnlicher Lektüre nicht erschließen. Es ist, als stünde man zu nahe vor dem Berg, um dessen Umrisse erkennen zu können. Der statistische Zugriff der LSA wirkt dagegen wie ein Weitwinkelobjektiv, mit dem ich die Kontur des Ganzen in den Blick bekomme.

Die Sprache enthält in dieser Betrachtung also Informationen über die Realität oder, wie Louwerse selbst sagt, „Weltwissen“, „world knowledge“ (Louwerse & Ventura 2005, 301), und zwar Wissen, für uns, wie im geschilderten Fall, unsichtbar und ohne Hilfsmittel vielleicht sogar dauerhaft unzugänglich sein kann. Warum sollte unser eigenes Verstehen nicht auch auf Wissen dieses Typs beruhen? Mit „Verstehen“ wäre dann an dieser Stelle ein interpretativer Zugriff gemeint, über den der Zugreifende nicht zur Gänze explizit Rechenschaft abgeben kann. Genau der Erfahrungsanteil, der dem Verstehen das Intuitive gibt, könnte durch statistische Verfahren unseres eigenen Geistes nach der Art der LSA gewonnen worden sein.

Verstehen wäre dann durch LSA oder ähnliche Programme simulierbar. Damit soll nicht gesagt sein, man könne Philologen und Philosophen demnächst durch Programme ersetzen. Nein, es scheint mir nur diese Möglichkeit nicht mehr prinzipiell ausgeschlossen. Es erscheint mir zumindest denkbar, wir könnten Verstehen, und damit grundsätzlich auch geisteswissenschaftliches Verstehen, durch Computer simulieren. Es würden dann die Geisteswissenschaften als ganze digitalisiert. Busas Formulierung von der „Mechanisierung der philologischen Analyse“ bekäme so eine sehr viel radikalere Bedeutung.

Wie dem auch sei. Ich habe zuerst zwei Bedeutungen von „digital“ unterschieden, einerseits „quantitativ“, andererseits „nicht analog“. Ich habe dann versucht, auf drei Aspekte der Veränderungen durch digitale Datenverarbeitung aufmerksam zu machen – drei

Aspekte, die ein gewisses Potential dazu haben, die Geisteswissenschaften in ihrer bisherigen Gestalt – und vielleicht noch mehr das Bild der Geisteswissenschaften von sich selbst – herauszufordern. Es war dies erstens die Verbreitung allgemeinzugänglicher digitaler Techniken wie der Suchmaschinen und der Googlebuchsuche, die auf die Geisteswissenschaften einen der Globalisierung ähnlichen Effekt haben. Es war dies zweitens die Tradition der digitalen Hilfswissenschaften, die sich mit ihrem Akzent auf dem Generellen als Trendsetter erweisen könnten. Und es war dies drittens – etwas spekulativer, wie ich zugebe – die neue Sicht auf das Verstehen als geisteswissenschaftlicher Urhandlung, die von der LSA aus möglich wird.

Literaturverzeichnis

Busa, Roberto (1952). Mechanisierung der philologischen Analyse. Nachrichten für Dokumentation. Korrespondenzblatt für die Technik und Wissenschaft, Wirtschaft und Verwaltung 3 (1952) 1, 14-19.

Busa, Roberto (1980). The Index Thomisticus. Computers and the Humanities 14 (1980), 83-90.

Deerwester, Scott, Dumais, Susan T., Furnas, George W., Landauer, Thomas K., Harshman, Richard (1990). Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science 41 (1990) 6, 391-407.

Hockey, Susan (2006). The History of Humanities Computing. A Companion to Digital Humanities, hrsg. von Susan Schreibman, Ray Siemens und John Unsworth, Madden, MA: Blackwell 2006, 3-19.

Louwerse, Max M., Ventura, Matthew (2005). How Children Learn the Meaning of Words and How LSA Does it (too). The Journal of the Learning Science 14 (2005), 301-309.

Louwerse, Max M., Zwaan, Rolf A. (2009). Language Encodes Geographical Information. Cognitive Science 33 (2009), 51-73.

Rommel, Thomas (2006). Literary Studies. A Companion to Digital Humanities, hrsg. von Susan Schreibman, Ray Siemens und John Unsworth, Madden, MA: Blackwell 2006, 88-96.

Sloterdijk, Peter (2010). Scheintod im Denken. Von Philosophie und Wissenschaft als Übung, Frankfurt am Main: Suhrkamp.

Winter, Thomas Nelson (1999). Roberto Busa, S.J., and the Invention of the Machine-Generated Concordance. The Classical Bulletin 75 (1999) 1, 3-20.

Geschichte schreiben im digitalen Zeitalter

Peter Haber

Historisches Seminar der Universität Basel
Hirschgässlein 21
CH-4051 Basel
peter.haber@unibas.ch

Zusammenfassung

Der Beitrag thematisiert den Wandel des Schreibens in der Geschichtswissenschaft unter dem Einfluss digitaler Systeme. Dabei geht es einerseits um die medialen Bedingtheiten des Schreibens im digitalen Kontext und – damit verbunden – um die Frage nach Narrativitität in der Geschichte. Zum anderen werden gemeinschaftliche Schreibprozesse und ihre Auswirkungen auf das Schreiben der Geschichte dargestellt.

Einleitung

Geschichtsschreibung, so stellte Rudolf Vierhaus vor Jahren einmal fest, ist die sprachliche Darstellung von komplexen diachronen und synchronen Zusammenhängen in der Vergangenheit (Vierhaus 1982). Wenn wir also über das Schreiben der Geschichte nachdenken, so bedeutet dies, dass wir über die Möglichkeiten dieser sprachlichen Darstellung unter den Bedingungen digitaler Schreibprozesse und Verbreitungswege nachdenken.

Konkret lassen sich dabei zwei Fragekomplexe ausmachen: Zum einen geht es um die medialen Bedingtheiten des Schreibens und, daraus abgeleitet, um die Frage des historischen Narrativs. Im Leitmedium des Digitalen, im World Wide Web, ist durch die Hypertextualität des Mediums das grundsätzlich lineare Moment des Narrativs in Frage stellt. Zum anderen geht es um neue Formen des gemeinschaftlichen Produktionsprozesses, um das sogenannte Collaborative writing, und damit verbunden um die Frage der Autorschaft von Geschichtsschreibung.

Beide Aspekte berühren nicht ausschließlich die Geschichtswissenschaften oder die Geistes- und Kulturwissenschaften, aber beide Aspekte rühren – gerade in den Geistes- und Kulturwissenschaften und ganz speziell auch in der Geschichtswissenschaft – an den Fundamenten des wissenschaftlichen Selbstverständnisses.

Narrativistisches Paradigma

In der Geschichtsschreibung beobachten wir seit einigen Jahrzehnten ein neues, narrativistisches Paradigma. Ohne Zweifel lässt sich hier der Einfluss der Textwissenschaften und insbesondere der Narratologie erkennen und es ist letztlich auf den Einfluss des „linguistic turn“ zurückzuführen, dass sich die Erkenntnis Raum geschaffen hat, dass das Schreiben von Geschichte an Sprachlichkeit und Textlichkeit gebunden ist.

Damit ist aber auch die Frage gestellt nach der Medialität der Geschichte und nach den sprachlich-narrativen Implikationen der Art und Weise, wie Geschichte geschrieben wird. Der narrative Modus der Geschichtsschreibung – die „erzählende Darstellung" wie es Johann Gustav Droysen schon vor über 140 Jahren genannt hat – gründet auf der Absicht, Kenntnis über die Vergangenheit zu erlangen und sie dadurch, dass sie rekonstruiert und erzählt wird, auch zu verstehen.

Dabei lassen sich das *Was* des Erzählten – *l'histoire* – vom *Wie* des Erzählens – dem *récit* – unterscheiden (Schönert, S. 142). Der mediale Wandel hat, so meine These, den Blickwechsel hin zum *récit* verstärkt. Angelika Epple hat im Kontext einer digitalen Historiographie auf die Implikationen einer Narratisierung der Geschichtsschreibung hingewiesen und drei Punkte genannt (Epple, S. 20): Zum einen verliert die Geschichtsschreibung mit ihrer Narratisierung den Anspruch auf Universalität, denn im Narrativ kann nicht alles gleichzeitig dargestellt werden. Die Gegenstände der Geschichtswissenschaft nehmen – zweitens – einen abgeschlossenen Charakter an und jede Erzählung verweist nur symbolisch auf ein Ganzes. Damit wird es für den Historiker möglich, aus der Materialfülle online wie offline auszuwählen. Und drittens perspektiviert Narratisierung die Geschichtsschreibung, denn jedes Narrativ wird von einem Autor mit einem je spezifischen Blickwinkel verfasst.

Im Mittelpunkt dieser Diskussionen steht die Frage nach Linearität und Hypertextualität. Hypertexte sind, so eine griffige Definition, netzwerkartig angeordnete, nichtlineare Texte (Krameritsch und Gasteiner, S. 145): Sie haben keinen definierten Anfang und keinen Hauptteil und entsprechend fehlt auch ein zusammenfassender Schluss. In einer hypertextuellen Struktur sind zudem alle Textelemente gleichwertig, das heißt, es gibt keine über- und untergeordneten Ebenen. Nichtlinearität ist im Kontext wissenschaftlicher Texte allerdings kein Phänomen, das erst mit dem Aufkommen von hypertextuellen Medien aktuell geworden ist.

Die selektive, assoziative und folglich nicht lineare Lektüre zum Beispiel gehört zu den gängigen Lesetechniken wissenschaftlicher Texte: Fußnoten, Verzeichnisse, Register oder Bibliographien sind heute Hilfsmittel beim nichtlinearen Lesen und deshalb fester Bestandteil wissenschaftlicher Textapparate. Zudem blickt die nichtlineare Darstellung von Texten auf eine mehrere Jahrhunderte alte Tradition zurück, wenn wir uns etwa den *Talmud Bavli*, den babylonischen Talmud, vor Augen führen. Die um den Haupttext angeordneten Anmerkungen und Verweise bilden mehrere Schichten des Textes, der aus der Mischna, der Gemara und den später hinzugefügten Kommentaren besteht. Damit stellt der Talmud einen in höchstem Grade assoziativ verknüpften, hypertextuellen Wissensraum dar (s. Börner-Klein).

Der Talmud, aber auch alle gedruckten Nachschlagewerke profitieren vom Umstand, dass das Medium Buch – etwa im Unterschied zu einer Filmrolle – ein zwar linear angelegter, aber nicht zwingend linear zu rezipierender Informationsträger ist. Dem Buch ist das Imperativ der linearen Lektüre jedoch eingeschrieben: Seitenzahlen, Vorwort und Nachwort sind Signale dafür, dass es der Autor gerne hätte, wenn das Buch vom Anfang bis zum Schluss gelesen würde. Mit textuellen Einsprengseln in der Art von „wie bereits oben ausgeführt" lässt sich der Druck auf den Leser verstärken. Viele Bücher sind zudem so angelegt, dass

eine nicht-lineare Lektüre wenig Sinn macht, Romane oder Krimis gehören ohne Zweifel dazu, ebenso die meisten geschichtlichen Einführungsbücher und viele wissenschaftliche Monographien. Sie bauen konzeptionell auf Linearität auf, können aber von ihrer Medialität her eine nicht-lineare Rezeption nicht verhindern. Im Anschluss an Krameritsch und Gasteiner soll dieser Texttyp als „monosequenziert“ bezeichnet werden. Die Textsegmente lassen sich zwar austauschen, aber der Leser läuft Gefahr, dass der Text dadurch unverständlich wird.

Davon lassen sich mehrfachsequenzierte Texte unterscheiden, die nicht mehr für eine eindeutig definierte Lektüre konzipiert sind, sondern die den Anforderungen des Lesers entsprechend mehrere Lesepfade anbieten. Typischerweise sind wissenschaftliche Handbücher auf diese Weise aufgebaut, ebenso Reiseführer oder Lexika. Auch Sammelbände können dieser Kategorie zugeschlagen werden. Zu den Merkmalen mehrfachsequenzierter Texte gehört, dass sie den unterschiedlichen Lektüreoptionen zum Trotz eine lineare, umfassende Lektüre vom Anfang bis zum Schluss ermöglichen.

Dies unterscheidet mehrfachsequenzierte Texte von unsequenzierten Texten. Diese können ohne Verlust von Verständnis in einer beliebigen Art und Weise gelesen werden. Es gibt weder einen definierten Anfang noch vorbestimmte Lesepfade, sondern jeder Leser erstellt sich seinen Interessen gemäß eine eigene Abfolge von Textelementen (sogenannten informationellen Einheiten), die mit *Links* untereinander verbunden sind. Unsequenzierte Texte produzieren – und das ist für den Kontext der Geschichtsschreibung entscheidend – immer eine Leerstelle: Der Leser weiß nicht, ob er alles gelesen hat, was er hätte lesen können, obwohl es ihm eigentlich zur Verfügung gestanden wäre.

Zwei Fragen stellen sich dabei der Geschichtswissenschaft: Wie wirkt sich diese Leerstelle auf das Verständnis historischer Zusammenhänge aus? Und wie gehen wir, die „Produzenten“ von Geschichtsschreibung, mit dieser Situation um?

Formen des gemeinschaftlichen Schreibens

Im Bereich des Schreibens lassen sich eine Vielzahl von strukturellen und habituellen Unterschieden zwischen den verschiedenen Wissenschaftsdisziplinen beobachten, die gerade auch im Hinblick auf das Potential und die Chancen gemeinschaftlicher Schreibprozesse von epistemologischer Bedeutung sind.

In der sehr häufig experimentell angelegten Forschung der Naturwissenschaften hat der Akt des Schreibens die Funktion, das Vorgehen und die Ergebnisse der Versuche nachträglich zu dokumentieren. Dabei zeichnen in der Regel eine Reihe von Autoren für den entsprechenden Text verantwortlich, womit aber weniger eine Mitwirkung am Akt des Schreibens bezeugt wird, sondern die Mitwirkung am gesamten Forschungsprojekt in einer bestimmten, oft auch aufschlüsselbaren Funktion. Qualifikationsarbeiten werden oft in größere Vorhaben eingebettet, wodurch nicht die Formulierung der Fragestellung, sondern die Durchführung und Dokumentation von in den Grundrissen bereits definierten Experimenten im Vordergrund steht.

In den Geisteswissenschaften hingegen steht der Prozess des Schreibens im Mittelpunkt der epistemologischen Anordnung. Die Formulierung der Forschungsfrage, die methodische

Konzeption und dann insbesondere die Arbeit am Text haben zentrale Bedeutungen und lassen sich, so zumindest das gängige Bild, schlecht kollektiv durchführen. Gemeinschaftliche Schreibpraktiken werden deshalb vor allem zwei Anwendungstypen zugeschrieben: Zum einen für Faktensammlungen wie zum Beispiel die Erarbeitung einer Enzyklopädie, zum anderen für den Einsatz bei neuen Unterrichtsformen. Das gemeinsame Schreiben eines Textes setzt voraus, dass die Schreibenden sich über Konzept, Arbeitsschritte und über Formulierungen austauschen: Gemeinsames Schreiben bedeutet demnach, normalerweise stillschweigend verlaufende Aktivitäten zu verbalisieren. Dies ist nicht grundsätzlich neu und gemeinsame Schreibprozesse gab es auch unabhängig von Computern und dem Internet. Doch mit dem gemeinsamen Schreiben am Netz erhält dieser Prozess eine Unmittelbarkeit, die es so zuvor nicht gab.

Die Möglichkeit, synchron am gleichen Text zu schreiben und dabei alle Arbeitsschritte transparent zu halten, kann auch das hierarchische Gefüge in einem Team verändern und neuartige Konflikte auslösen. Dürfen, um nur einige Beispiele zu nennen, schlechte Formulierungen des Vorgesetzten von allen korrigiert werden und wenn ja, wie sind sie zu kommentieren? Ist es wünschenswert, wenn alle Anläufe, einen Gedanken zu formulieren, mit Zeitstempel dokumentiert und für alle einsehbar sind? Und wer entscheidet zu welchem Zeitpunkt, wann die Arbeit am Text abzuschließen ist?

Der Terminologie der angelsächsischen Literatur folgend, lassen sich zwei Hauptformen des gemeinschaftlichen Schreibens unterscheiden: Interactive writing und Group writing (Lehnen, S. 150). Beim Interactive writing handelt es sich im Grunde genommen nicht um ein gemeinsames Schreiben, denn es geht darum, die Expertise anderer Personen beim Schreiben als ein zentrales Moment einzubeziehen. Die eigentliche Arbeit am Text verbleibt bei einem einzigen Autor. Entscheidend aber ist, dass die Textrevisionen nicht erst am Schluss erfolgen, sondern laufend vorgenommen werden und so zu einem Teil des Schreibprozesses werden. Moderne Textverarbeitungsprogramme wie Microsoft Word unterstützen solche Prozesse mit vielfältigen Hilfsfunktionalitäten wie zum Beispiel „Änderungen nachverfolgen“ oder „Dokumente vergleichen und zusammenführen“, welche die Arbeitsprozesse auch visualisieren. Da die Texte nicht simultan, sondern konsekutiv bearbeitet werden, ist eine Vernetzung des Computers nicht nötig, da die Textversionen auch auf einem beliebigen Datenträger ausgetauscht werden können. Beim *Group Writing* werden die beim *Interactive writing* zeitlich und räumlich noch versetzt verlaufenden Aktivitäten in einem gemeinsamen Produktionsprozess verdichtet und synchronisiert. Bei diesem Vorgehen verstärkt sich der Effekt, dass verschiedenes Wissen aller Beteiligten und unterschiedliche Perspektiven auf das Thema in den Text einfließen.

Im traditionellen Setting steht bei der gemeinsamen Produktion von Texten das Gespräch im Mittelpunkt. Heute lässt sich das direkte Gespräch durch Chat oder Internet-Telephonie ersetzen, das vielleicht sogar mit Videoübertragung ergänzt wird. Zusätzlich und gleichzeitig ist es möglich, ein im Netz abgespeichertes Dokument gemeinsam zu bearbeiten. Moderne Systeme lassen verschiedene Szenarien der Zusammenarbeit zu: So können mehrere Autoren gleichzeitig am Text arbeiten, wobei jeweils auf dem Bildschirm eingeblendet wird, wer aktiv am Text arbeitet. Um Revisionskonflikte zu verhindern, werden Textabsätze, die von einem der Autoren bearbeitet werden, für die anderen Autoren automatisch gesperrt.

Möglich ist aber auch eine asynchrone Zusammenarbeit mehrerer Autoren, wobei die registrierten Autoren jedes Mal benachrichtigt werden, wenn einer der Autoren am Text Änderungen vorgenommen hat. Jede Änderung, Ergänzung oder Streichung wird dokumentiert und in einer Versionshistorie abgebildet; diese ist für alle Mitarbeitenden einsehbar, verschiedene Versionen können miteinander verglichen und ältere Versionen wieder hergestellt werden.

Ob sich kollaborative Arbeitsweisen in der Geschichtswissenschaft wirklich etablieren werden, lässt sich heute schwer beurteilen. Es ist zu vermuten, dass eine ausdifferenzierte Praxis entstehen wird, bei der das *Interactive writing* an Bedeutung gewinnen wird, ohne dabei das Konzept des Autors verschwinden zu lassen. Kollaborative Schreibpraktiken werden zudem bei der Erstellung von historischen Enzyklopädien und Wörterbüchern eine zunehmende Bedeutung erlangen. Fraglich ist, ob in absehbarer Zeit die dominante Rolle der einem Autor eindeutig zuzuschreibenden Monographie an Bedeutung verlieren wird und in diesem Bereich gemeinschaftliche Schreibprozesse mehr Bedeutung erhalten werden. Für die Historiker der Zukunft heißt das nichts weniger, als beide Schreibpraxen – das gemeinschaftliche ebenso wie das individuelle – einüben und in ihren unterschiedlichen Ausprägungen beherrschen zu müssen. In den geschichtswissenschaftlichen Hochschulcurricula muss deshalb das wissenschaftliche Schreiben einen wesentlich größeren Stellenwert erhalten, als dies bisher der Fall ist. Der dynamische Wandel bei den entsprechenden Arbeitsumgebungen wird in den nächsten Jahren kaum nachlassen, was von Dozierenden und Studierenden eine hohe Flexibilität und eine kontinuierliche Weiterbildung in diesem Bereich verlangt.

Die Herausforderung wird nicht sein, ein entsprechendes Tool oder eine spezielle Schreibtechnik besonders virtuos zu beherrschen, sondern der jeweiligen Situation angepasst die „richtige“ Arbeitsumgebung – zum Beispiel ein *Wiki*-System, ein *Weblog* oder ein anderes Hilfsmittel – auszuwählen und die dazu passende Schreibtechnik anwenden zu können. Dabei geht es um die Frage, wer mit welchen Rechten an welchen Inhalten was verändern darf. An die Frage nach den Rollenmodellen koppelt sich sehr häufig die Frage nach der Definitionsmacht und der Kontrolle innerhalb eines Publikationsprojektes, weshalb es sich hier nicht primär um technische, sondern um soziale und gruppendynamische Fragen handelt, die zu einem frühen Zeitpunkt gelöst werden müssen.

Fazit

Das Thema „Geschichte Schreiben im digitalen Zeitalter“ präsentiert sich, so das Fazit, als ein Bündel unterschiedlicher Fragen, die weit über das historiographische Kerngeschäft hinaus reichen. Auch wenn die Fragen heute noch vage zu sein scheinen, dürfte es unbestritten sein, dass der digitale Wandel längst über das Suchen und Finden hinaus geht und dass der gesamte wissenschaftliche Workflow betroffen ist.

Zum Autor

Peter Haber ist Privatdozent für Allgemeine Geschichte der Neuzeit am Historischen Seminar der Universität Basel. Er hat sich an der Philosophisch-Historischen Fakultät der Universität Basel mit einer Studie über die Geschichtswissenschaft im digitalen Zeitalter habilitiert. Im Sommersemester 2010 war er Gastprofessor für „Geschichte, Didaktik und digitale Medien“ am Institut für Geschichte sowie am Institut für Wirtschafts- und Sozialgeschichte der Universität Wien. Zuletzt erschien der mit Martin Gasteiner zusammen herausgegebene UTB-Band „Digitale Arbeitstechniken für die Geistes- und Kulturwissenschaften“. Im Netz ist er unter http://hist.net/haber zu finden.

Literatur

Börner-Klein, Dagmar (2002): Assoziation mit System: Der Talmud, die „andere“ Enzyklopädie, in: Pompe, Hedwig / Scholz, Leander (Hrsg.): Archivprozesse. Die Kommunikation der Aufbewahrung, Köln 2002 (= Mediologie; 5).

Epple, Angelika (2005): Verlinkt, vernetzt, verführt – verloren? Innovative Kraft und Gefahren der Online-Historiographie, in: Epple, Angelika / Haber, Peter (Hrsg.): Vom Nutzen und Nachteil des Internets für die historische Erkenntnis. Version 1.0, Zürich 2005 (= Geschichte und Informatik; 15), 15-32.

Krameritsch, Jakob / Gasteiner, Martin (2006): Schreiben für das WWW: Bloggen und Hypertexten, in: Schmale, Wolfgang (Hrsg.): Schreib-Guide Geschichte. Schritt für Schritt wissenschaftliches Schreiben lernen, Wien 2006 (2. Auflage), 231-271.

Lehnen, Katrin (2000): Kooperative Textproduktion. Zur gemeinsamen Herstellung wissenschaftlicher Texte im Vergleich von ungeübten, fortgeschrittenen und sehr geübten SchreiberInnen (Dissertationsschrift), Bielefeld 2000.

Schönert, Jörg (2004): Zum Status und zur disziplinären Reichweite von Narratologie, in: Borsò, Vittoria / Kann, Christoph: Geschichtsdarstellungen. Medien – Methoden – Strategien, Köln 2004 (= Europäische Geschichtsdarstellungen; 6), 131-143.

Vierhaus, Rudolf (1982): Wie erzählt man Geschichte? Die Perspektive des Historiographen, in: Quandt, Siegfried / Süssmuth, Hans (Hrsg.): Historisches Erzählen. Formen und Funktionen, Göttingen 1982, 49-56.

Semantic Web Techniken im explorativ geisteswissenschaftlichen Forschungskontext

Niels-Oliver Walkowski

Berlin-Brandenburgische Akademie der Wissenschaften
DFG-Projekt Personendaten-Repositorium
Jägerstr. 22/23
D-10117 Berlin
walkowski@bbaw.de

Zusammenfassung

Das Projekt des Semantic Web geht mit der Hoffnung einher menschliches Wissen nicht nur digital verarbeitbar zu machen, sondern auch den Prozess der Generierung von Wissen an den Computer delegieren zu können. Der Artikel möchte das dahinterliegende Konzept von Wissensdynamik aus geisteswissenschaftlicher Sicht problematisieren, ohne jedoch den formulierten Anspruch völlig aufzugeben. Alternativ zum informationstechnologisch geprägten Ansatz automatischer Inferenz wird die Adaption von Semantic Web Techniken in einer explorativen Forschungssituation erörtert.

„Neues Wissen“ und das Semantic Web

„[…] Semantic Web technology could provide a substrate for the discovery of new knowledge that is not contained in any one source, and the solution of problems that were not anticipated by the creators, […]“ (Gruber, 2008)

Mit diesem Satz pointiert Gruber eine Utopie, die im Zusammenhang mit dem Semantic Web einmal mehr reaktiviert wurde. Die Idee, den Computer in die Lage zu versetzen Wissen automatisch zu prozessieren oder wie im Zitat zumindest die Grundlage dafür zu liefern, ist so schillernd, dass sie den Blick auf entscheidende Fragen versperrt. Was gilt überhaupt als new knowledge, und wie sehen die Operationen aus, die mit Hilfe des „alten“ Wissens dieses neue Wissen erzeugen sollen. Ist neues Wissen die „Entdeckung“ neuer Paradigmen (Kuhn, 1978) oder sind es Perspektivenverschiebungen (Bachmann-Medick, 2009). Der Diskurs über das Semantic Web klammert häufig aus, dass aus wissenschaftstheoretischer Perspektive keinesfalls Konsens darüber besteht, auf welcher Grundlage wissenschaftliche Innovation stattfindet. Es ist daher wesentlich sinnvoller zu fragen, welche Operationen der Wissensverarbeitung das Semantic Web beschreibt. Tim Berners-Lee charakterisiert das Generieren von Wissen in einem Beispiel wie folgt:

„If a city code is associated with a state code, and an address uses that city code, then that address has the associated state code. A program could then readily deduce, for instance, that a Cornell University address, being in Ithaca, must be in New York State, which is in the U.S., and therefore should be formatted to U.S. Standards.“ (Berners-Lee, 2001)

Auf Grund der Einfachheit des gewählten Beispiels wird schnell deutlich, dass das, was als neues Wissen bezeichnet wird, genauer beschrieben werden kann als implizites Wissen, das

explizit gemacht wird. Mittels einer Inferenz wird aus einer Klassenzugehörigkeit (durch ein Prädikat) eines konkreten Subjekts eine andere Eigenschaft dieses Subjekts – in diesem Fall transitiv – deduziert. Dieser Tatbestand verändert sich auch dann nicht wesentlich, wenn die Wissensbasis und die Inferenzen komplexer werden. Was an Wissen generiert werden soll, muss bei der semantischen Beschreibung einer Entität immer schon vorausgesetzt werden, denn die Strukturen der formalen Logik sind deterministisch (Beierle & Kern-Isberner, 2008, S. 74). Der Ansatz mit den Strukturen der Logik zu neuem Wissen zu gelangen basiert auf dem von Wittgenstein als sprachphilosophisches Missverständnis entlarvten Paradigma, die Struktur der Sprache und der Welt als identisch anzusehen (Wittgenstein, 1997).

Einen Versuch der Abgrenzung im Prozess der Wissensgenerierung nimmt Hans Reichenbach vor, indem er zwischen einem context of discovery und einem context of justification unterscheidet (Reichenbach, 1938). Während im Letzteren strenge Regeln der Wissenschaft gelten, zu denen je nach Wissenschaftsmilieu verschieden ausgeprägt auch die formale Logik gehört, wird der Ursprung von Wissen durch Theorie jenseits dieser Regeln und Normen verortet. Tatsächlich wurde die Grundlage der Theoriebildung lange Zeit in den Bereich des Irrationalen verlegt, so z. B., wenn Popper sie als intuitiv und einer logischen Analyse weder fähig noch bedürftig bezeichnet (Popper, 1989). In den letzten Jahrzehnten ist jedoch häufig verknüpft mit dem Begriff der Explorativität der Versuch unternommen worden, diesen Bereich selbst wissenschaftlich zu durchdringen. Als Beispiel kann die Grounded Theory genannt werden. Auch wenn Exploration mit Sicherheit in verschiedenen Wissenschaftsmethodiken unterschiedlich definiert wird, kann sie dennoch für jedweden Forschungsprozess einen fruchtbaren Ausgangspunkt bieten.

Um ein genaueres Bild davon zu bekommen, wie Semantic Web Technologien in einem explorativen Forschungsprozess eingebunden werden können, lohnt sich der Rückgriff auf Bruno Latours anschauliche Beschreibung des Experimentierens, mit dem er allgemeine Aspekte des Forschens herausarbeiten will. Latour charakterisiert das Organisieren eines Experiments als einen Prozess, in dem eine spezifische Situation durch Abtasten von Weltlichkeit sowie des Isolierens und Arrangierens von Bestandteilen dieser Weltlichkeit konstruiert wird. Es handelt sich im explorativen Prozess um eine Phase, in der das Material noch nicht tragfähig genug identifiziert wurde und die innere Struktur im Hinblick auf das Ausgangsproblem erst noch entwickelt werden muss. Daher erscheint der Forschungsgegenstand während des Experimentierens zunächst immer als kontingent. Das Arrangieren ist dabei ein spielerischer Prozess mit ungewissem Ausgang. Da ein Try immer auch ein Error nach sich ziehen kann, bedeutet Explorieren ein ständiges Wechseln zwischen Gestalten und Verwerfen. Dies erfordert ein hohes Maß an Flexibilität von der Technologie, die den Forschungsprozess digital unterstützen soll.

Die aufgezeigte Charakterisierung zielt unter anderem auf die Mittel ab, mit denen Forschungsgegenstände und Weltlichkeit zur Darstellung kommen und wodurch sie repräsentiert werden. Im Kontext des Semantic Web werden Entitäten mittels einer Beschreibungssprache, des Resource Description Frameworks (RDF) erschlossen. Unter Zuhilfenahme einer durch RDF(S) oder der Web Ontology Language (OWL) repräsentierten Wissensbasis kann abgeleitet werden, ob eine RDF-Beschreibung Gültigkeit besitzt. Üblicherweise

bezeichnet die informationstechnologische Literatur den konkreten Graphen einer RDF-Beschreibung als die Interpretation einer Ontologie, da aus dieser mehrere mögliche RDF-Graphen folgen können (Hitzler, Krötzsch & Sure, 2007). Entsprechend geht der konkreten Beschreibung von Ressourcen im Workflow zumeist die Formulierung einer domainspezifischen Ontologie voraus. Die explorative Forschungssituation stellt sich genau umgekehrt zu dieser Situation dar. Eine Theorie auf deren Grundlage eine Ontologie gebildet werden könnte, mit deren Hilfe konkrete Forschungsgegenstände geordnet und sortiert werden, liegt noch nicht vor, sondern soll Ergebnis des Forschungsprozesses sein. Ebenso stellt sich in einem derartigen Forschungsprozess die Generierung einer Theorie als die eigentliche Interpretationsleistung dar, wohingegen die konkreten Beschreibungen im Sinne einer in der Praxis verwurzelten Vernunft den harten Boden für Systematisierungen und Hypothesenbildungen bilden. Die Beschreibung von Wissensbasis und Interpretation wird also auf den Kopf gestellt und dennoch bieten gerade die Semantic Web Technologien eine gegenüber anderen Technologien hervorragende Ausgangslage zur Unterstützung des Explorationsprozesses. Der Grund hierfür ist, dass Wissensbasis und Interpretation überhaupt unabhängig voneinander beschrieben werden können. Dies ist ein erheblicher Vorteil gegenüber Wissensrepräsentationen in Datenbanken, bei denen die Definition eines Inhaltsmodells, z. B. in Form von Objektklassen oder eines Entity-Relationship Modells die Grundlage für eine Dateneingabe ist. Im Semantic Web Kontext wird so die weitgehend voraussetzungsfreie Beschreibung von Entitäten und ein induktives Arbeiten mit Ontologien ermöglicht.

Die Umkehrung der Priorisierung von Beschreibung und Ontologisierung kann dafür genutzt werden, die vier Ansätze für exploratives Arbeiten nach Bortz (2006) entlang eines Workflows anhand von Techniken des Semantic Web anzuordnen und in diese zu integrieren. Dabei findet methoden- und theoriebasierte Exploration durch die Generierung von assertionalem Wissen in RDF statt. Quantitative Exploration kann unter Zuhilfenahme von automatisierten Mining-Techniken zur Anwendung kommen. Da dies keine speziell Semantic Web-orientierte Problematik ist, findet sie an dieser Stelle keine Beachtung. Schließlich kann mittels der Ontology-Based-Data-Alignment Strategie eine qualitative Exploration unternommen werden.

Dichte Beschreibungen und induktive Ontologien

Um dem zu betrachtenden Forschungsgegenstand „auf den Grund zu kommen“, bedarf es in der methoden- und theoriebasierten explorativen Situation keines logisch konsistenten Modells, sondern einer möglichst dichten Beschreibung. Ziel ist es, möglichst viele Daten aus unterschiedlichen theoretischen Perspektiven und Methodenansätzen zusammenzubringen. Eine derartig dichte Beschreibung erfasst ihren Gegenstand in der Fülle und den Kontexten seiner Erscheinung. Sie erzeugt ein vielschichtiges Material mit Aussagen, die sich logisch widersprechen können, weil sie den Forschungsgegenstand in unterschiedlichen Situationen erfassen. Die Identifizierung einer Beziehung eines Forschungsgegenstandes zu anderen Entitäten kann sich als schwierig erweisen, weil eine zu beschreibende Situation etwas Ganzheitliches darstellt, das im Blick des Forschers erst in Bestandteile gegliedert werden muss. Ohne die Forderungen eines vorformulierten Inhaltsmodells durch

eine formale Wissensrepräsentation in OWL oder RDF-Schema, ist es dem Wissenschaftler möglich, eine derartige dichte Beschreibung und ihre zugrunde liegende Kontingenz in RDF abzubilden. Sie ermöglicht damit die Repräsentation von komplexeren Strukturen und eine präzisere Abbildung realweltlicher Sachverhalte und ihrer Repräsentation in Forschungsobjekten. Vokabulare und Erschließungsmuster können, wo neue Perspektiven mit einbezogen werden, on the fly variiert und angepasst werden. Ein solches Vorgehen wird dadurch ermöglicht, dass die zur Hypothesenbildung benötigte Systematisierung und Formalisierung in einem später folgenden Prozess der Ontologisierung externalisiert werden kann. RDF ermöglicht Entitäten mittels des selben Prädikats mehrfach zu erschließen und so kontingente Aussagen im zuvor beschriebenen Sinne zuzulassen. Auch wird die Erschließungstiefe, also die Anzahl von Tripples mit gleichem Subjekt, nicht durch ein Inhaltsmodell beschränkt.

Neben diesen Vorteilen, die sich aus dem graphenorientierten Datenmodell von RDF ergeben, besitzt RDF noch eine Reihe von Ausdrucksmitteln, die sich zur Abbildung von sich als kontingent darstellenden Forschungsgegenständen auf der Ebene der assertionalen Beschreibung nutzen lassen. Dazu gehören z. B. die Möglichkeiten, die Blank Nodes und verschiedene Listentypen bieten. Listen, die als Subjekt oder Objekt eines Tripples dienen, können weiterführend sein, wenn der Forscher in einem gegebenen Kontext nicht in der Lage ist, sich auf ein Objekt festzulegen z. B. weil eine solche Festlegung erst stattfinden kann, wenn das ganze Material gesichtet ist. Die Bestimmung von Entitäten offen zu halten ist sinnvoll, wo sie beim Zeitpunkt der Datenaufnahme unentscheidbar ist, um nicht auf eine Beschreibung vollständig verzichten zu müssen. Blank Nodes können dazu genutzt werden Problemstellungen zu behandeln, bei denen es gerade um die Identifikation von monolithischen und komplexen Gegenständen in einem Forschungsdiskurs geht. Als „Metaknoten“ sind sie in der Lage einzelne URI's zu komplexen Gebilden zusammenzufassen.

Für den Zweck der Inventarisierung und des Brainstorming, die von Bortz der qualitativen Datenanalyse zugerechnet werden, bietet neben der einfachen Abfrage durch SELECT vor allem der DESCRIBE Operator von SPARQL herausragende Anknüpfungspunkte. Da das Semantic Web ein unabgeschlossener Raum mit verteilten Ressourcen ist und deshalb für den Benutzer nicht vorausgesetzt werden kann, dass er weiß, wie eine Ressource sinnvoll mit SELECT abgefragt werden kann, hat das W3C eigens einen Abfragemechanismus bereitgestellt, der das sich Vertrautmachen mit einer Ressource ermöglichen soll. Zwar liegt die explorativ erzeugte Datenbasis in den meisten Fällen im Gegensatz zum Web nicht verteilt vor, doch gleichen sich die beiden Szenarien durch die Heterogenität der Beschreibungen. Der DESCRIBE Befehl reagiert also auf ein ähnliches Bedürfnis. Bisher wird beim W3C noch recht widersprüchlich darüber diskutiert, wie DESCRIBE diese Aufgabe lösen soll (W3C, 2008). Gerade in dieser Offenheit liegt aber auch die Chance für eine Implementierung innerhalb einer spezifischen digitalen Forschungsinfrastruktur.

Für die Typenbildung und die Strukturanalyse stehen zwei Ansatzpunkte zur Verfügung. Zum einen die Fähigkeit des CONSTRCUT Operators Datenstrukturen zu verändern. Eine explorative Forschungssituation profitiert von dieser Möglichkeit erheblich, weil eine in einer dichten Beschreibung vorliegenden Datenbasis mittels einer Anfrage in unterschiedlichen Inhaltsmodellen getestet werden kann. Im Prinzip wird es auch möglich die Datenstrukturen parallel zum Forschungsprozess und der in ihm stattfindenden Prozesse und

unter Verwendung situativ gültiger Thesen zu verändern. Der andere Weg ist, den Datenbestand, wie oben angeschnitten, unter Verwendung von Ontolgiesprachen zu harmonisieren. In einem derartigen Vorgehen werden z. B. owl:Class oder owl:samAs Tripples verwendet, um heterogene Daten auf einer weiteren Abstraktionsebene zu harmonisieren. Ähnliche Strategien formieren in anderen Zusammenhängen unter dem Begriff des Ontology-Based-Data-Alignments (Euzenat & Valtchev, 2004).

Die Kombination aus einzelnen spekulativ angelegten „Ontologieschnipseln“ und SPARQL kann die Methode des sogenannten Pretesting unterstützen. Hierbei geht es noch nicht darum, ein systematisches Bild der Daten zu erzeugen, sondern mögliche Ansätze im Kleinen zu testen, um deren Tragfähigkeit im Hinblick auf eine theoretische Ausarbeitung zu prüfen. Z. B. kann mittels einer Abfrage geguckt werden, welche Konsequenzen die Veränderung der Definition zweier Klassen auf die Beziehung zwischen ihnen hat. Ontologien werden zum Werkzeug in der Arbeit mit dem Material, welches viele verschiedene Interpretationen zulässt. Die so gewonnene Flexibilität zwischen dichter Beschreibung in RDF und spekulativen Ontologien entspricht dem spielerischen Charakter, der als Voraussetzung des explorativen Forschungsprozesses beschrieben wurde.

Literatur

Bachmann-Medick, D. (2009). *Cultural turns : Neuorientierungen in den Kulturwissenschaften*. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verl.

Berners-Lee, T., Hendler, J., Lassila, O. et al (2001). The semantic web. *Scientific american*, *284*(5), 28–37.

Bortz, J. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler: mit 156 Abbildungen und 87 Tabellen*. Berlin, Heidelberg, New York, Barcelona, Hongkong, London, Mailand, Paris, Tokio: Springer.

Euzenat, J., & Valtchev, P. (2004). Similarity-based ontology alignment in OWL-lite. *Ecai 2004 (Frontiers in Artificial Intelligence and Applications)*. Valencia: Proceeding, 333-343.

Gruber, T. (2008). Collective knowledge systems: Where the social web meets the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, *6*(1), 4–13.

Kuhn, T. S., & Krüger, L. (1978). *Die Entstehung des Neuen: Studien zur Struktur der Wissenschaftsgeschichte*. Frankfurt: Suhrkamp.

Latour, B., & Roßler, G. (2002). *Die Hoffnung der Pandora: Untersuchungen zur Wirklichkeit der Wissenschaft*. Frankfurt: Suhrkamp.

Popper, K. (1989). *Logik der Forschung* (9. Aufl.). Tübingen: Mohr.

Reichenbach, H. (1938). *Experience and prediction an analysis of the foundations and the structure of knowledge*. Chicago: The University of Chicago Press.

W3C (2008, 08. Februar). *SPARQL/Extensions/Describe*. Retrieved from http://esw.w3.org/SPARQL/Extensions/Describe.

Die Anwendung von Ontologien zur Wissensrepräsentation und -kommunikation im Bereich des kulturellen Erbes

Georg Hohmann

Germanisches Nationalmuseum
Referat für Museums- und Kulturinformatik
Kartäusergasse 1
D-90402 Nürnberg
g.hohmann@gnm.de

Zusammenfassung

Die wissenschaftliche Forschung an sog. Gedächtnisinstitutionen produziert kontinuierlich eine große Menge an digitalen Daten. Für den Datenaustausch zwischen diesen Institutionen werden in der Regel Datenaustauschformate genutzt, deren Verwendung fast immer einen Daten- und Granularitätsverlust verursacht. Eine Alternative bieten Ontologien im Sinne des *Semantic Web* an. Mit dem *CIDOC Conceptual Reference Model* liegt bereits eine umfassende und erprobte Ontologie für das kulturelle Erbe vor. Sie kann als Referenzontologie genutzt werden, auf deren Basis lokale Applikationsontologien erstellt werden können. Diese Herangehensweise garantiert eine vollständige semantische Integrität und ermöglicht einen vollautomatisierten Datenaustausch ohne *Mapping* und Datenverluste.

Einleitung

Die wissenschaftliche Forschung an den Gedächtnisinstitutionen Museum, Archiv und Bibliothek produziert kontinuierlich eine große Menge an digitalen Daten. Der Heterogenität der bearbeiteten Materialien, des kulturellen Erbes, ist es geschuldet, dass sich bis heute kein einheitliches Speicher- und Verwaltungssystem etablieren konnte. Stattdessen existiert eine Vielzahl von Insellösungen, die jeweils auf einen spezifischen inhaltlichen und organisatorischen Zweck zugeschnitten sind. Die lokal vorliegenden Daten in einer Gedächtnisinstitution sind daher in der Regel nicht mit andernorts vorliegenden Daten ohne Weiteres kompatibel. Nicht zuletzt das Internet hat aber den Druck erhöht, diese Daten verfügbar zu machen und auszutauschen, um Synergieeffekte zu erzeugen und das eigene gespeicherte Wissen zu erweitern. Dieser Datenaustausch wird heute hauptsächlich unter Nutzung von Metadatenschemata realisiert.

Wissensrepräsentation und -kommunikation mit Metadatenschemata

Metadatenschemata sind der Versuch, vorhandenes Wissen über Gegenstände der realen Welt verbal zu kategorisieren, zu schematisieren und zu organisieren. Mit ihrer Hilfe lässt sich das Wissen in leicht verwaltbaren Einheiten ablegen, die als „Dokument“ oder „Metadatendokument“ bezeichnet werden. Die Einheitlichkeit dieser Dokumente dient dem Zweck der strukturierten Ablage und der Vergleichbarkeit des gespeicherten Wissens.

Im Bereich des kulturellen Erbes existiert eine ganze Reihe von Metadatenschemata. Für die Erfassung von Sammlungsinformationen in Kunstmuseen sind etwa die *Categories for the Description of Works of Art* (CDWA) bzw. die vereinfachte Variante CDWA-*Lite* vorgesehen. CDWA wurde seit Mitte der 1990er Jahre vom *Getty Research Institute* entwickelt und findet heute breite Anwendung. Aus dem großen Fundus von Metadatenstandards aus dem Bibliothekswesen sei beispielhaft das *Machine Readable Cataloging* (MARC) genannt, das in verschiedenen Ausformungen für unterschiedliche Medientypen und Institutionen existiert. MARC wird seit den 1970er Jahren federführend von der *Library of Congress* gepflegt. Im Archivwesen hat *Encoded Archival Description* (EAD) breite Anwendung gefunden, das ebenfalls von der *Library of Congress* entwickelt wurde. Daneben existieren noch zahlreiche andere Schemata, und es werden auch weiterhin Versuche unternommen, neue Schemata zu etablieren.

Um ein solches Metadatenschema zu nutzen, müssen zunächst die Inhalte des lokalen Datenbestands auf diese abgebildet werden (*Mapping*). Handelt es sich dabei um eine relationale Datenbank, müssen bei diesem Vorgang die Spalten der in der Datenbank enthaltenen Tabellen auf die Elemente des Austauschformats bezogen werden. Oft ist dieser Vorgang mit einer Bearbeitung der Daten verbunden, da nicht immer genau eine Spalte genau einem Element des Metadatenschemas entspricht. In vielen Fällen müssen Daten konkateniert, substituiert oder segmentiert werden, um den Anforderungen der gewählten Schemata gerecht zu werden (Baca, 2003, S. 49ff.). Dieser Effekt wird noch potenziert, wenn unterschiedliche Schemata aufeinander abgebildet werden, um die damit abgebildeten Daten interoperabel zu machen (Abbildung 1).

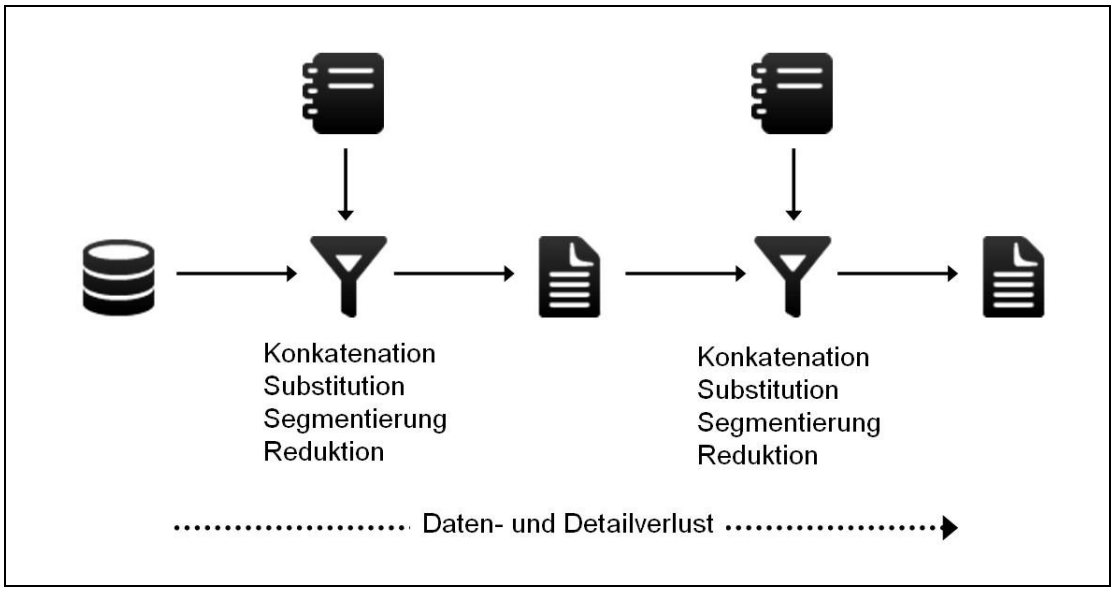


Abbildung 1: Daten- und Detailverlust beim Mapping.

Durch die Zweckgebundenheit von Metadatenschemata und die damit verbundene Fokussierung auf einen Teilaspekt einer Wissensdomäne ist es zudem in der Regel nie möglich, alle in der lokalen Datenbank vorgehaltenen Daten in der gleichen Granularität in einem gewählten Schema unterzubringen. Ein solches Schema ist im Normalfall inhaltlich wesentlich eingeschränkter und detailärmer als die Speicherstruktur des Ursprungssystems. Der

Ausgangsdatensatz lässt sich aus dem Zielformat nicht mehr vollständig rekonstruieren. Zusammengefasst lässt sich sagen, dass die Verwendung von Metadatenschemata fast immer zu einem Daten- und Detailverlust führt.

Hinzu kommt, dass Metadatenschemata eine dokumentenorientierte Wissensrepräsentation implizieren. Analog zur traditionellen Speicherung und Strukturierung von Wissen – etwa auf Karteikarten – wird ein Metadatendokument als eine Einheit begriffen, die oft den einzigen Wissenszugang darstellt. Das Metadatendokument ist der einzige Repräsentant des Wissens und die einzige Basis für eine weitere, auch maschinelle Verarbeitung und Kommunikation dieses Wissens. Ein Zugriff auf das Wissen im Ursprungssystem ist nicht vorgesehen.

Wissensrepräsentation im *Semantic Web*

Eine Alternative zu diesem dokumentenbasierten Ansatz zur Wissensrepräsentation und -kommunikation stellt das *Semantic Web* dar. In der Vision des *Semantic Web* wird die ursprüngliche Datenquelle selbst zum Bestandteil des Netzes und muss nicht durch eine andere Form repräsentiert werden. Ein solcher *Semantic Web* Knoten exponiert seinen Datenbestand in dem gleichen Detailgrad, den die interne Datenhaltung besitzt. Jedes einzelne Datum wird dabei separat als eigene Einheit repräsentiert, die mit anderen Einheiten interagieren kann. Grundlage für die Art der Datenrepräsentation ist das *Resource Description Framework* (RDF), das bereits seit 1999 vom *World Wide Web Consortium* (W3C) entwickelt und standardisiert wird. RDF erlaubt es, die einzelnen Dateneinheiten in Form rudimentärer Satzkonstruktionen miteinander in Verbindung zu setzen. Ein sog. *Triple* besteht aus einem Subjekt (S), einem Prädikat (P) und einem Objekt (O), das wiederum als Subjekt weiterer Aussagen dienen kann. Ein konkretes Beispiel für ein *Triple* wäre „Die Adlerfibel des Germanischen Nationalmuseums (S) hat die Inventarnummer (P) 'FG1608' (O)“.

Doch RDF gibt nur die syntaktischen Grundlagen vor. Welche Sprache gesprochen, was mit dieser Sprache ausgedrückt werden kann und welchen Regeln diese Sprache inhaltlich unterliegt, definiert RDF nicht. Diese Aufgabe übernimmt eine Ontologie. Eine Ontologie definiert eine Wissensdomäne und legt die darin verwendeten Konzepte und ihre Relationen untereinander detailliert fest. Im Beispiel könnte sie eine Klasse für „Museumsobjekt“ und eine Klasse für „Inventarnummer“ definieren, die über die Eigenschaft „hat Inventarnummer“ miteinander verbunden werden können. Zur Definition von Ontologien hat ebenfalls das *World Wide Web Consortium* (W3C) die *Web Ontology Language* (OWL) entwickelt, die inzwischen in der Version 2 vorliegt und sich als Standard etabliert hat. Eine besondere Bedeutung kommt der Sprachausformung OWL 2 DL (*Description Logics*) zu, die zu einer entscheidbaren Untermenge der Prädikatenlogik erster Stufe äquivalent ist. Das bedeutet, dass bereits vorhandene und erprobte Berechnungsverfahren aus diesem Bereich auf OWL 2 DL Ontologien angewandt werden können.

Mit RDF und OWL liegen also bereits geeignete Instrumente vor, um Daten in Form von Aussagesätzen zu repräsentieren als auch die dazu verwendeten Sprachkonstrukte so zu definieren, dass beide Ebenen einheitlich maschinell zu verarbeiten sind (Allemang & Hendler, 2008).

Wissenskommunikation mit Ontologien

Doch wie ist es nun um die Kommunikation zwischen zwei oder mehreren unabhängigen Wissensnetzen bestellt? Um die Dateneinheiten des einen Wissensnetzes im anderen Wissensnetz verfügbar zu machen, müssen diese zueinander in Beziehung gesetzt werden.

Ein Weg, dies zu erreichen, liegt auf der Hand: Das *Mapping* von Ontologien untereinander. Dieser häufig verwendete Ansatz birgt jedoch die gleichen Probleme wie die Abbildung von Metadatenschemata aufeinander. Der einzige Vorteil ist, dass dadurch, dass Ontologien in OWL DL durch erprobte Algorithmen verarbeitet werden können, ein automatischer Prozess Auskunft darüber geben kann, ob ein *Mapping* formal korrekt durchgeführt wurde oder Widersprüche erzeugt.

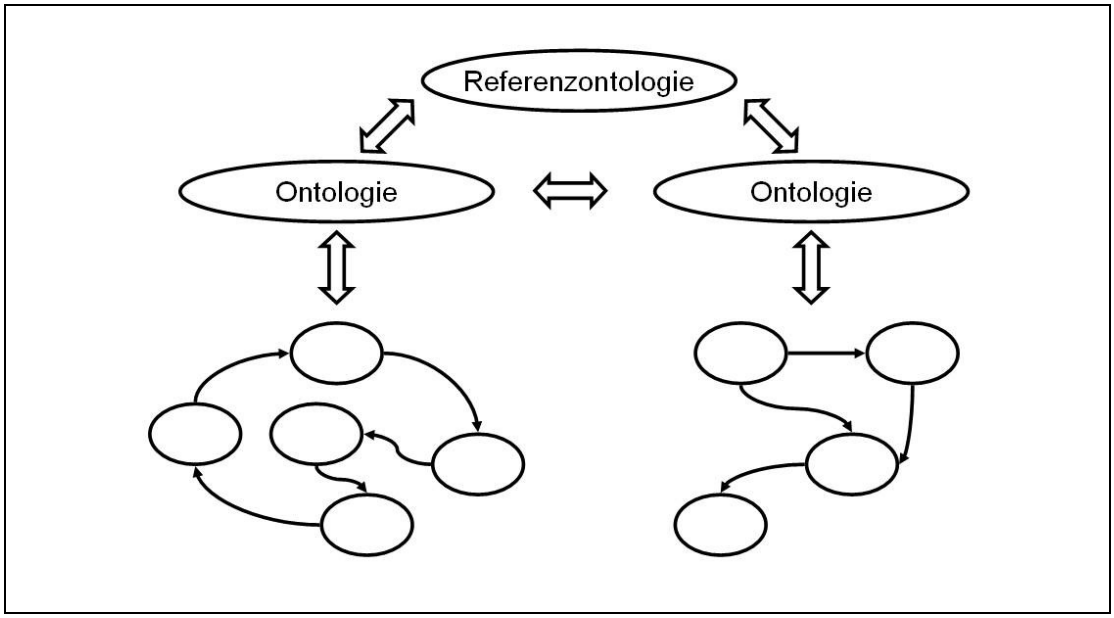


Abbildung 2: Verwendung einer Referenzontologie für den Datenabgleich.

Eine geeignetere Methode ist die Verwendung von Referenzontologien (Abbildung 2). Diese zeichnen sich dadurch aus, dass sie das Wissen einer Domäne auf einer sehr abstrakten Ebene beschreiben. Die Konzepte und Eigenschaften anwendungsbezogener Ontologien werden als Subkonzepte und -eigenschaften einer Referenzontologie gebildet (Hohmann, 2008). Sie erben damit alle Merkmale des jeweiligen Konzepts oder der Eigenschaft und können weitere Merkmale hinzufügen. Bedienen sich nun beide Ontologien derselben Referenzontologie, ist zwischen den Wissensnetzen kein *Mapping* mehr notwendig, um mit beiden parallel arbeiten zu können.

Grundsätzliche Voraussetzung ist natürlich das Vorhandensein einer Referenzontologie, die innerhalb einer Wissensdomäne konsensfähig ist, einen hohen Reifegrad besitzt und im Rahmen eines Standardisierungsprozesses festgehalten wurde. Erstaunlicherweise gehört der Bereich des kulturellen Erbes zu den wenigen Wissensdomänen, die bereits eine solche Ontologie vorweisen können.

CIDOC Conceptual Reference Model

Eine Arbeitsgruppe des *Comité international pour la documentation* (CIDOC) des *International Council of Museums* (ICOM) arbeitet seit den 1990er Jahren an einer einheitlichen Wissensrepräsentation für die Dokumentation des kulturellen Erbes. Sie veröffentlichte bereits 1999 die erste Version des *CIDOC Conceptual Reference Model* (CRM). Das CRM ist eine exakte Definition von Klassen und Eigenschaften, die die grundlegende Konstellation von Personen, Orten, Zeiten und Ereignissen im Hinblick auf die Dokumentation historischer Sachverhalte beschreibt.

Jede Klasse im CRM repräsentiert ein Konzept der realen Welt, das über seine Eigenschaften mit anderen Konzepten des Modells verknüpft ist. Eine Klasse kann dabei jeweils die Ober- und/oder Unterklasse weiterer Klassen bilden. Die oberste Klasse bildet „E1 CRM Entity“ von der sich alle anderen Klassen ableiten. Nach dem Prinzip der Vererbung übernehmen Unterklassen dabei die Eigenschaften der jeweiligen Oberklassen und können diese weiter ausspezifizieren oder einschränken. Ebenso verhält es sich bei den Eigenschaften, die die jeweiligen Relationen von einer Klasse zu einer anderen Klasse darstellen. Insgesamt definiert das CRM 86 Klassen und 137 Eigenschaften. Ein bedeutender Schritt für die Etablierung des CRM im Bereich des kulturellen Erbes war 2006 die Standardisierung durch die *International Organization for Standardization* (ISO). Die Arbeiten am CRM wurden damit aber nicht beendet, sondern stetig weitergeführt. Inzwischen liegt das CRM weitreichend überarbeitet und erweitert in der Version 5.0.2 vor (Crofts, Doerr, Gill, Stead, and Stiff, 2010) und wird erneut der ISO zur Aktualisierung des Standards vorgelegt.

Das CRM versteht sich als eine rein abstrakte Ontologie und macht keine Aussagen über eine mögliche Implementierung in spezifischen Systemen. Aus diesem Grund existiert auch keine offizielle OWL-Variante des CRM. Diese Lücke füllt das Erlangen CRM / OWL (http://erlangen-crm.org), welches das CRM als OWL DL-Ontologie implementiert und damit die Grundlage für die Verwendung des CRM in auf *Semantic Web* Technologien basierenden Systemen bildet.

Die Verwendung einer Referenzontologie

Anhand der bereits gezeigten Beispielontologie lässt sich die Verwendung des CRM als Referenzontologie demonstrieren. Es wurde definiert, dass ein Museumsobjekt über eine Inventarnummer identifiziert wird. Als Entsprechung im CRM bietet sich das „*E22 Man-Made Object*“ an, das über die Eigenschaft „*P1 is identified by*“ mit einem „*E42 Identifier*“ verbunden werden kann. Die Klassen und auch die Eigenschaft der Anwendungsontologie werden nun als Subklassen und -eigenschaft dieses CRM-Konstrukts modelliert und erben damit alle ihre Merkmale (Abbildung 3).

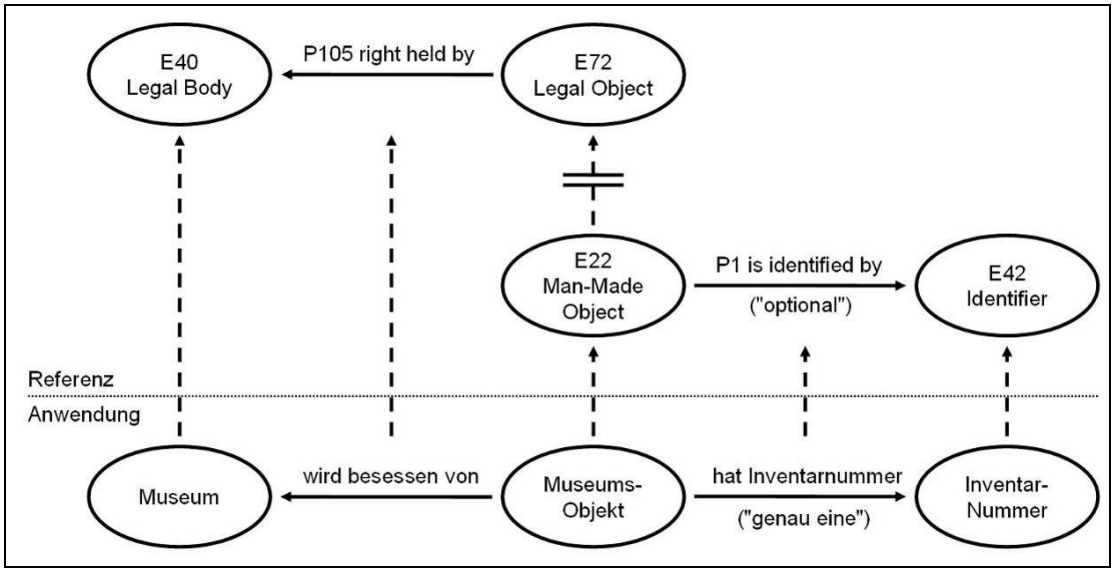


Abbildung 3: Referenzontologie und Anwendungsontologie.

Typisch ist dabei die nähere Spezifizierung des Konzepts. Aus einem „*E22 Man-Made Object*“ wird ein Museumsobjekt, aus einem „*E42 Identifier*“ eine Inventarnummer. Die Verwendung der Eigenschaft kann dabei zusätzlich eingeschränkt werden. Während die CRM-Eigenschaft optional ist, kann in der Anwendungsontologie definiert werden, dass ein Museumsobjekt immer exakt eine Inventarnummer haben muss. Dass mit „*E22 Man-Made Object*“ eine geeignete Klasse gewählt wurde, zeigt sich auch, wenn man sich die Vererbung innerhalb des CRM anschaut. Über mehrere Stufen hinweg ist ein „*E22 Man-Made Object*“ auch eine Subklasse von „*E72 Legal Object*“. Als solches kann es Rechten unterliegen, die eine Körperschaft besitzt. Aufgrund der Vererbung kann also auch das Museumsobjekt einem Recht unterliegen, hier dem Besitzrecht eines Museums.

Schlussbemerkung

Mit der Verwendung von Ontologien eröffnen sich vollkommen neue Möglichkeiten der Wissensrepräsentation und -kommunikation nicht nur im Bereich des kulturellen Erbes. Das Wissen ist dabei nicht mehr an die Form des Dokuments gebunden, sondern wird in granularen Dateneinheiten vorgehalten, die je nach Verwendungszweck ad hoc aggregiert werden können. Eine Maschine hat über mehrere Abstraktionsschichten variabel Zugriff auf diese Daten, die in einem beliebigen Detailgrad vorliegen können. Es findet keine Reduktion oder Modifikation der Daten in Hinblick auf einem spezifischen Zweck mehr statt.

Gedächtnisinstitutionen können in diesem Bereich eine Vorreiterrolle spielen. Die notwendigen Vorarbeiten sind geleistet, die Technologie vorhanden. Die wichtigste Voraussetzung ist jedoch die möglichst freie Verfügbarkeit von Daten, weshalb den entsprechenden Institutionen anzuraten ist, getreu dem Motto der *Semantic Web* Bewegung „*Open your data*“ ihr Wissen zur Verfügung zu stellen.

Literatur

Allemang, D., & Hendler, J. (2008). Semantic Web for the Working Ontologist. Modeling in RDF, RDFS and OWL. Burlington: Morgan Kaufmann Publishers.

Baca, M. (2003). Metadata Schemas And Controlled Vocabularies For Art, Architecture, And Material Culture. *Cataloging & Classification Quarterly*, 36(3/4), 47–55.

Crofts, N., Doerr, M., Gill, T., Stead, S., & Stiff, M. (Hrsg.). (2010). Definition of the CIDOC Conceptual Reference Model, January 2010 (5.0.2). http://cidoc-crm.org/docs/cidoc_crm_version_5.0.2.pdf

Hohmann, G. (2008). Die Anwendung des CIDOC CRM für die semantische Wissensrepräsentation in den Kulturwissenschaften. In P. Ohly & J. Sieglerschmidt (Hrsg.), *Wissensspeicherung in digitalen Räumen. Nachhaltigkeit, Verfügbarkeit, semantische Interoperabilität*. *Proceedings of the 11th Conference of ISKO German Chapter Konstanz 2008*. Würzburg: Ergon, 210–222.

Der Leipziger Professorenkatalog. Ein Anwendungsbeispiel für kollaboratives Strukturieren von Daten und zeitnahes Publizieren von Ergebnissen basierend auf Technologien des Semantic Web und einer agilen Methode des Wissensmanagements

Christian Augustin, Ulf Morgenstern & Thomas Riechert

Universität Leipzig

Institut für Informatik
Abteilung Betriebliche Informationssysteme
Johannisgasse 26
D-04103 Leipzig
riechert@informatik.uni.leipzig.de

Historisches Seminar
Lehrstuhl für Neuere und Neueste Geschichte
Beethovenstraße 15
D-04017 Leipzig
professorenkatalog@uni-leipzig.de

Zusammenfassung

Der Beitrag gibt einen Überblick über das interdisziplinäre Forschungsprojekt Catalogus Professorum Lipsiensis, das kollektivbiographische Zugänge der Wissenschafts- und Universitätsgeschichte auf innovative Weise mit Technologien des Semantic Web verbindet. Erläutert werden die geschichtswissenschaftlichen Implikationen sowie die Architektur und Funktionsweise des Professorenkatalogs.

Universitätsgeschichte und Professorenkataloge

Professorenkataloge zählen seit der Blütephase des deutschen Hochschulwesens in der zweiten Hälfte des 19. Jahrhunderts zum festen Bestandteil der akademischen Memorialkultur. Indem der Lehrkörper einer Hochschule in seiner Gesamtheit erfasst wird, folglich also vom frisch habilitierten Privatdozenten bis zum Großordinarius vergangener Zeiten alle Lehrenden Eingang finden, dienen derartige Kataloge sowohl der repräsentativen Außendarstellung, wie der Identitätsstiftung. Die versammelten biographischen Informationen umfassen zumeist allgemeine Angaben wie etwa Lebensdaten, Eckpunkte der akademischen Karriere, wissenschaftliche Verdienste und bedeutende Publikationen. Je nach Umfang enthalten detailliertere Kataloge aber auch weitere Daten privater oder gesellschaftlicher Art, wie etwa Affiliationen oder Parteimitgliedschaften.

Für die Wissenschaftsgeschichte sind derartige Materialsammlungen natürlich von kaum zu überschätzendem Wert, geben sie doch bei systematischer Auswertung Auskunft über verschiedenste überindividuelle biographische Muster einer fest umrissenen Gruppe Hochschulangehöriger. Während der inhaltliche Zugang zu diesen teils mehrbändigen Nachschlagewerken früher jedoch über Register oft mühsam war, lässt sich der prosopographische

Reichtum an individuellen Merkmalen im Zeitalter der Online-Kataloge vollständiger und schneller erschließen.

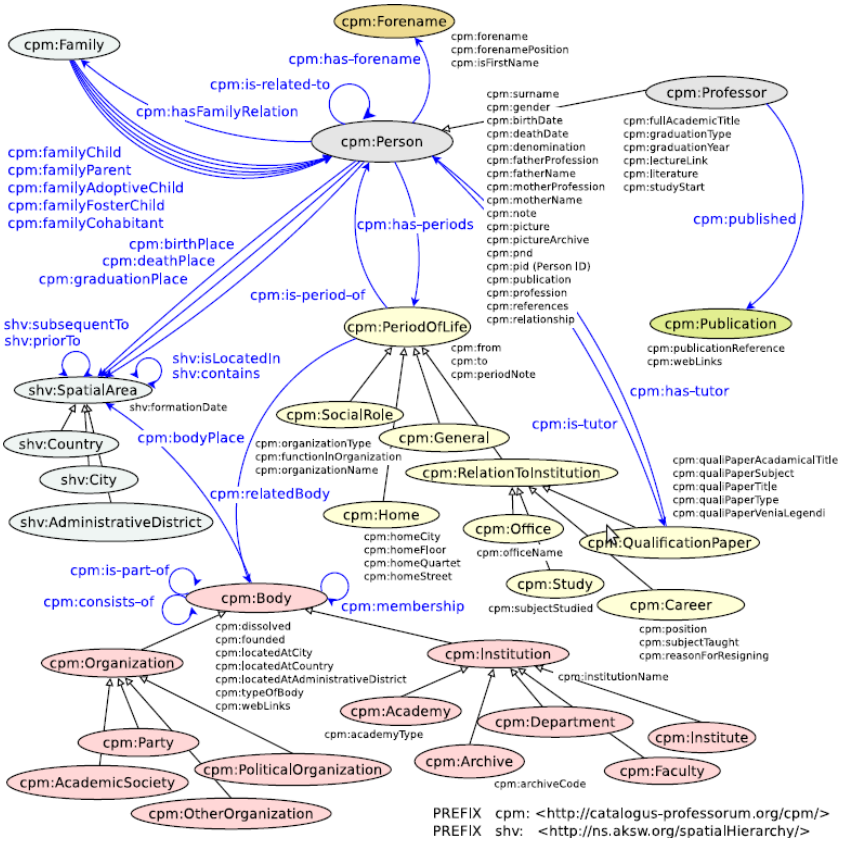


Abbildung 1: Catalogus Professorum Model

Technische Umsetzung

Wie die meisten ähnlich gelagerten Projekte ist auch der Leipziger Professorenkatalog[1] als elektronische Ressource konzipiert. Er basiert auf einer Forschungsdatenbank, die mit Technologien des Semantic Web realisiert wird. Im Mittelpunkt steht die Einzelbiographie des Hochschullehrers, wobei eine maximale Datenerfassungstiefe angestrebt wird. Einen Überblick über die in der Datenbank umgesetzten Konzepte liefert Abbildung 1. Zum gegenwärtigen Zeitpunkt umfasst der Katalog ca. 1600 solcher Biographien. Mit Hilfe des semantischen Daten-Wikis OntoWiki[2] erfolgt die Ergänzung des Datenmaterials kollaborativ aus Bibliotheken und Archiven.

Anwendung findet die Forschungsdatenbank einerseits als wissenschaftsgeschichtliches Analyseinstrument, andererseits als Grundlage für den online-Professorenkatalog, der über eine live-Abfrage ein festgelegtes Datenspektrum als Kurzbiographie für die Öffentlichkeit bereitstellt. Die hierdurch ermöglichte Interaktion mit einem breiten Nutzerkreis führt zu einem ständig wachsenden Zufluss weiterer Detailinformationen.

[1] Verfügbar im Web unter http://www.uni-leipzig.de/unigeschichte/professorenkatalog (Zone 3).
[2] OntoWiki ist quelloffen und steht unter http://ontowiki.net als Download zur Verfügung.

Die Architektur der Forschungsdatenbank stellt eine Kombination aus verschiedenen Applikationen dar. Verschiedene Anwendergruppen greifen über angepasste Benutzerschnittstellen auf die Inhalte des Katalogs zu. Zusätzlich verfügt die Datenbank über generische, durch Maschinen interpretierbare Schnittstellen. Diese basieren auf dem RDF-Format, welches zur Beschreibung von Informationen im Semantic Web genutzt wird. Die Katalogdaten sind so als Linked Data verfügbar und können via SPARQL-Query abgefragt werden.

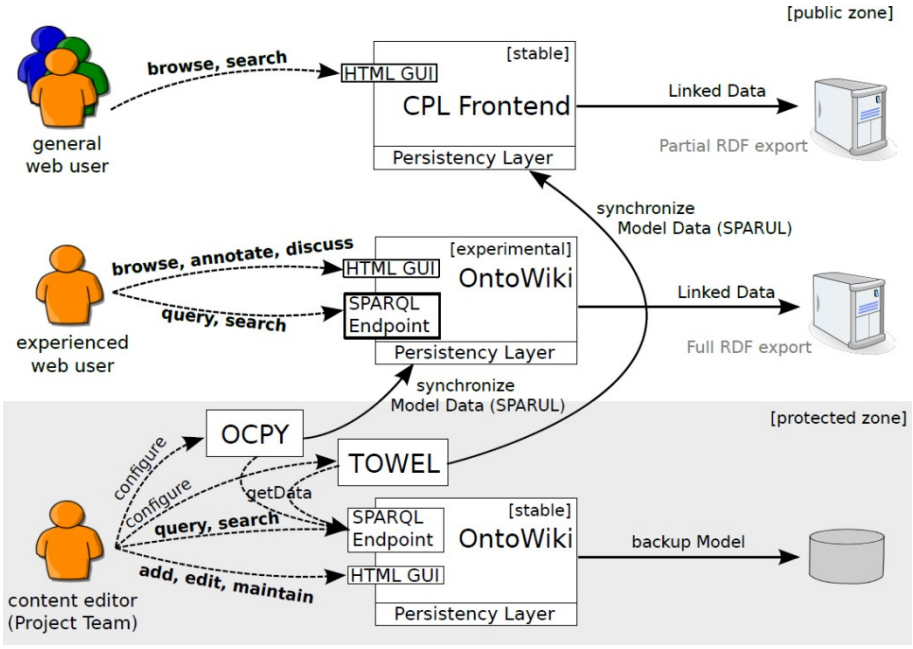


Abbildung 2: Catalogus Professorum Architecture

Abbildung 2 zeigt die Architektur der Datenbank, welche ausgehend von den Sichtweisen der verschiedenen Anwendergruppen auf den Katalog, drei Schichten unterscheidet: Die (1) geschützte Zone zum kollaborativen und räumlich unabhängigen der Erfassung, Strukturierung und Validierung der Katalogdaten, die (2) öffentliche Zone für die wissenschaftliche Nutzung[3] und die (3) allgemein zugängliche öffentliche Zone. Letztere ist auch in die Webpräsentation der Universität Leipzig integriert. Dieser Zugang bietet einfache Recherchemöglichkeiten und stellt Inhalte in verschiedenen Repräsentationsformen, u. a. als Linked Data im RDF-Format, dar. Der wissenschaftliche Zugang erfolgt über eine experimentelle Installation des Daten-Wikis OntoWiki. Die Anwender profitieren hier von den jeweils neuesten Funktionen und Erweiterungen, wie z. B. einer graphischen Abfrage.

Anwendung

Dass ein weiterer Vorteil in der Kostenersparnis gegenüber mehrbändigen Lexika liegt, spricht zudem für das Format. So sind dem Wunsch, alle erreichbaren Informationen zu einem Professor zu erfassen, keine finanzielle Grenzen gesetzt, die im Zusammenhang mit dem Umfang eines Buches stünden. Schwierigkeiten ergeben sich aber – je weiter die erfassten Daten in die Gegenwart reichen – aus datenschutzrechtlichen Gründen. Während dieses Problem für die Epoche des Nationalsozialismus mehr und mehr verschwindet, bleibt

[3] Wissenschaftliche Schnittestelle des Kataloges verfügbar unter: http://catalogus-professorum.org/ (Zone 2).

es für die Jahre der DDR virulent. Aus forschungspragmatischen Gründen wurde eine einfache und eben auch erst durch die elektronische Form des Katalogs möglich gewordene Regelung getroffen. Alle Daten zu politischen Aktivitäten, die als gesicherte Belege, z. B. in Archiven, gesammelt werden, finden Eingang in die Datenbank. Welche Einzelinformationen welchem Nutzer zugänglich gemacht werden, hängt nun davon ab, welche Regelung das jeweilige Archivgesetz vorsieht. Verstreichen Veröffentlichungsfristen des Persönlichkeitsrechts (zehn Jahre nach dem Tod bzw. zehn Jahre nach dem 100. Geburtstag), werden alle Daten für jeden frei zugänglich. Bis dahin bleibt der Zugriff von der Genehmigung der Projektleiter abhängig. Auf der Website des Katalogs, die für eine breite Öffentlichkeit gedacht ist (Zone 3), finden sich daher in vielen Fällen nur ausgewählte Daten, die aber nach Verstreichen der genannten Frist automatisch ergänzt werden.

Literatur

Morgenstern, U., & Riechert, T., (Eds.). (2010). Catalogus Professorum Lispsiensis. Konzeption, technische Umsetzung und Anwendungen für Professorenkataloge im Semantic Web. *Leipziger Beiträge zur Informatik*, (XXI), Leipzig.

Thomas Riechert, Ulf Morgenstern, Sören Auer, Sebastian Tramp, and Michael Martin. (2006). Knowledge Engineering for Historians on the Example of the Catalogus Professorum Lipsiensis. In Proceedings of The 9th International Semantic Web Conference – ISWC 2010, Shanghai China, November 7-11, 2010, Berlin / Heidelberg, 2006. Springer. (Best In-Use Paper) Online: http://iswc2010.semanticweb.org/accepted-papers/386

Ein interaktiver Multitouch-Tisch als multimediale Arbeitsumgebung

Karin Herrmann, Max Möllers, Moritz Wittenhagen & Stephan Deininghaus

RWTH Aachen University
Human Technology Centre
Theaterplatz 14
D-52062 Aachen
herrmann@humtec.rwth-aachen.de

Zusammenfassung

Der Beitrag stellt ein interdisziplinäres Projekt vor, in dem Literaturwissenschaftler und Informatiker einen interaktiven Multitouch-Tisch entwickeln, um eine neuartige Umgebung für komplexe literaturwissenschaftliche Arbeitsvorhaben zu schaffen. Insbesondere Arbeitssituationen, die das parallele Arbeiten mit Büchern und Datenbanken erfordern, sollen unterstützt werden, indem digitale Informationen mit physisch präsenten Materialien verknüpft werden.

Einführung

Das interdisziplinäre Projekt *Brain/Concept/Writing* an der RWTH Aachen University verfolgt u. a. das Ziel, Instrumente zu entwickeln, die den Anforderungen netzbasierten Arbeitens in den Geisteswissenschaften entsprechen, und zwar durch die Entwicklung interaktiver Umgebungen und intelligenter Benutzeroberflächen. In unserem Projekt nutzen Literaturwissenschaft und Editionswissenschaft daher das Know-how der Human-Computer Interaction (HCI), um den Zugriff auf die archivierten Wissensbestände den unterschiedlichen Bedürfnissen der Benutzer anzupassen. Die Innovationen im Bereich der HCI beinhalten neue Hardware-Schnittstellen wie beispielsweise interaktive Tischoberflächen, aber auch neue Interaktionsmetaphern wie das „Multitouch“-Prinzip der mehrhändigen Interaktion mit Informationen auf einem Display. Ziel ist es, in Form eines Multitouch-Tisches einen neuartigen Arbeitsplatz für Literaturwissenschaftler zu entwickeln, der die Stärken der konventionellen Arbeitsweise mit gedruckten Materialien mit technologischen Neuerungen wie etwa digitaler Annotation verknüpft.

Nutzer-Orientierung

Unser Ansatz besteht darin, die Bedürfnisse des Nutzers von vorneherein in den Entwicklungsprozess einzubeziehen. Weil Digitalisierung kein Wert an sich ist, kommen die technischen Möglichkeiten erst nach intensiven Nutzerstudien ins Spiel. Wir haben Spezifika literaturwissenschaftlichen Arbeitens ausführlich analysiert, haben Literaturwissenschaftler beim Arbeiten beobachtet, gefilmt, qualitative Interviews geführt sowie Nutzertests mit Beispielaufgaben entworfen und durchgeführt. Die daraus resultierenden Erkenntnisse wurden in Beziehung zu den Ergebnissen allgemeinerer Studien über den Umgang mit

physischen und digitalen Dokumenten in verschiedenen Modalitäten gesetzt (vgl. Adler et al., 1998; Marshall & Bly, 2005; Morris et al., 2007; Sellen & Harper, 1997; Sellen & Harper, 2003).

Die Vorzüge digitalen Materials gegenüber Papier in der literaturwissenschaftlichen Praxis sind demzufolge im Wesentlichen funktionaler Natur und liegen in den Möglichkeiten, Text computergestützt zu durchsuchen, zu analysieren und für die weitere Verarbeitung aufzubereiten. Der Zugriff auf diese Daten muss heute jedoch meist durch herkömmliche Desktop-Computer geschehen. Dies führt zu einer scharfen Abgrenzung gegenüber der Interaktion mit papierbasierten Dokumenten, welche sich durch natürliche, raumgreifende und durch haptische Präsenz des Materials unterstützte Arbeitsabläufe auszeichnet.

Ziel ist es, dass die Nutzer durch digitale Unterstützung etwas gewinnen, ohne die Vorteile des Arbeitens mit Papier und Büchern aufgeben zu müssen. Zu den Leitfragen zählt also nicht nur: Wo kann man durch Digitalisierung unterstützen und entlasten? Sondern auch: Was ist beim Arbeiten mit und auf Papier gut, was soll auf jeden Fall erhalten werden? So haben wir den Aspekt der Nutzerakzeptanz von vorneherein einbezogen.

Konstruktion und Funktionsweise

Der Tisch ist eine Kombination aus Ein- und Ausgabegerät. Er besteht aus einem Aluminiumgerüst, an dem mehrere Projektoren angebracht sind. Diese Projektoren werfen ihr Bild über eine Spiegelkonstruktion von unten auf die Tischplatte aus Acryl. Diese Acrylplatte ist mit einem Diffusor versehen und fungiert somit als Bildschirm. Die Tischplatte ist 1,40 m breit und 80 cm tief und befindet sich etwa auf Höhe „normaler“ Tische.

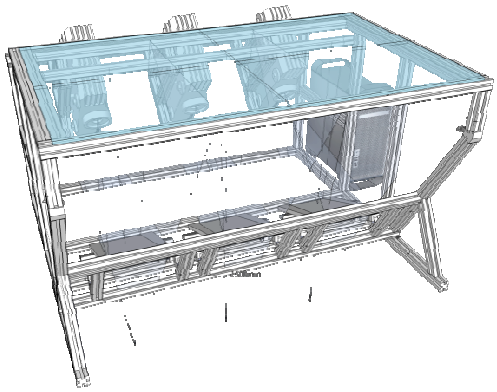


Abbildung 1: Zeichnung des interaktiven Tisches. RWTH Aachen Universitv.

Um die Oberfläche als Eingabemedium zu nutzen, wird eine Kombination aus Frustrated Total Internal Reflection (vgl. Han, 2005) und Diffuse Illumination genutzt: Infrarot-LEDs fluten die Acrylplatte von den Seiten und von unten mit Licht. Sobald sich Hände, Buchrücken etc. der Platte von oben nähern oder sie berühren, wird das LED-Licht nach unten reflektiert. Dort, am Boden des Tisches, befinden sich zwei Infrarot-Kameras, die diese Objekte „sehen“ und z. B. einen Fingerabdruck als Benutzereingabe interpretieren können. Außerdem wird der Tisch in der Lage sein, spezielle Marker auf Buchrücken zu erkennen. Der Tisch „weiß“ also, welches Buch sich auf ihm an welcher Stelle befindet. Eine über dem Tisch befindliche hochauflösende Kamera kann die Bücher aufnehmen (vgl. Everitt et al., 2008) und die aufgeschlagene Seite identifizieren. Somit kann die Anzeige von Zusatzinformationen auf die aufgeschlagene Seite abgestimmt werden. Zudem macht es dieser Ansatz möglich, aus physischen Dokumenten direkt am Tisch digitale Exzerpte zu erstellen

oder einzelne Seiten aus Büchern einzublenden, die sich momentan nicht auf dem Tisch befinden, aber in der Vergangenheit vom System erfasst wurden.

Die Tischoberfläche fungiert als großes Display, das Inhalte anzeigen kann. Außerdem kann der Tisch Objekte, die auf der Platte abgelegt werden, erkennen, und deren Bewegungen registrieren. So ist es möglich, digitale Informationen mit physisch präsenten Materialen wie Büchern zu verknüpfen und sie in Relation zu diesen anzuzeigen (vgl. Koike et al., 2000; Wu et al., 2008). Das gedruckte Buch bleibt damit Hauptbezugspunkt; die Arbeit mit Büchern soll durch digitale Informationen etwa aus Datenbanken oder digitalen Editionen unterstützt werden.

Die Verfügbarkeit digitaler Zusatzinformation wird durch digitale Karteikartenreiter entlang des physischen Textes angezeigt. Mithilfe dieser Reiter kann der Benutzer weitere Dokumentenseiten mit entsprechendem digitalen Inhalt unter dem Buch hervorziehen und somit Art und Umfang der Zusatzinformation flexibel an seine Bedürfnisse anpassen.

Auch die Notizen des Benutzers sollen integriert werden: Sie werden digital gespeichert, und der Nutzer kann sie auf der Oberfläche hin- und herziehen und den Objekten, mit denen er gerade arbeitet, hinzufügen. Dies wird möglich durch eine neue Technologie, die es erlaubt, Schreibprozesse in bislang nicht gekannter Weise aufzuzeichnen (vgl. Mackay et al., 2002; Yeh et al., 2006). Diese technologische Innovation soll es ermöglichen, unmittelbar auf die Tischplatte zu schreiben und damit Notizen in digitalen Materialien zu machen.

Der Computer im Tisch hat nicht nur die Aufgabe, die beschriebenen Funktionen zu steuern, sondern hat auch klassische Datenbankfunktionen, d. h. er speichert die eigenen Eingaben, sorgt für den Zugang zu vorhandenen Datenbanken, etwa digitalen Archiven, und ist mit dem Internet verbunden. Darüber hinaus ermöglicht er eine kontextsensitive Bedienung. Von zentraler Bedeutung ist hier die spezielle Software, die alle Teilfunktionen steuert und nutzergerecht zusammenführt.

Leistung der Technologie

Ziel ist die Entwicklung einer interaktiven Oberfläche für literaturwissenschaftliches Arbeiten, speziell für die Nutzung von Editionen. Kennzeichen dieses Arbeitens sind häufige nicht-lineare Interaktion wie etwa das Nachschlagen in Apparat, Zeugenverzeichnis, Kommentar, das Aufrufen von digitalisierten Faksimila, Recherche in Datenbanken und anderes mehr. Dabei gehen wir davon aus, dass die Benutzer meist auch mit Büchern arbeiten; darüber hinaus soll der Tisch ihnen kontextrelevante Informationen anzeigen, z. B. Handschriften-Scans, Sekundärliteratur, Lexika etc. An den physisch präsenten Büchern, die der Tisch dank spezieller Labels erkennt, können digitale Fähnchen angebracht werden: Vermerke, Markierungen, Buchzeiger, Notizen. Der Nutzer hat die Möglichkeit, Dokumente einfach durch Berühren zu verschieben und so hin- und herzuziehen – und die digitale Kontextinformation wandert mit (vgl. Steimle et al., 2010).

Der Multitouch-Tisch soll also ein neuartiges Bindeglied zwischen greifbaren und digitalen Medien darstellen. Voraussetzung ist nicht, dass alle Inhalte bereits in digitaler Form vorliegen; auch soll der Tisch alle Vorteile erhalten, die das Arbeiten mit Büchern und Papier

bietet. Und er soll literaturwissenschaftliches Arbeiten unterstützen, ohne dass gewohnte Arbeitsabläufe aufgegeben werden müssen.

Nutzerstudien

Der Planung des Tisches gingen umfassende Nutzerstudien voraus: Auf Basis des ausgewerteten Beobachtungsmaterials wurde eine erste Nutzerstudie generiert, und parallel dazu wurde ein erster Papierprototyp des Systems entworfen, an dem diese Studie durchgeführt werden sollte. Zuvor jedoch wurden das Studiendesign und auch der Prototyp durch einen informellen Testlauf validiert.

Der Nutzertest bestand aus drei Teilen:

einem Eingangsinterview, das nach Vorerfahrungen mit Editionen und interaktiven Oberflächen fragt und die Zielsetzung des Projekts kurz vorstellt, ohne auf die genaue Funktionsweise des Systems einzugehen,

dem Hauptteil: einem interaktiven Test mit mehreren Teil-Aufgaben, bei dem der Versuchsleiter das Verhalten des Prototypen simulierte. Hier sollten die Tester einige einfache Aufgaben lösen, die gängige Arbeitsschritte unserer Zielgruppe (Informationen suchen, Quellen miteinander abgleichen, zusammenfassende Notizen machen) nachbilden,

und einem teilstrukturierten Abschlussinterview, in dem Auffälligkeiten im Test noch einmal aufgegriffen, Probleme mit den Testern nachvollzogen und ihr während des Experiments gebildetes mentales Modell des Systemverhaltens sowie mögliche Designalternativen ergründet wurden.

Anhand der Testaufgaben wurden einfachste Arbeitsabläufe mitverfolgt:

Wann benutzt jemand ein Buch?

Wann zieht jemand weitere Materialien hinzu?

Wie arrangiert jemand Material auf dem Tisch?

Eine besondere Rolle in den Interviews spielten zudem die persönlichen Strategien des Notizen-Machens: Schreiben die Testpersonen häufig in Bücher, auf separate Blätter, in eine Kladde? Wie werden die Notizen organisiert? Gibt es eine Hierarchie der unterschiedlichen Arten von Notizen? Wie eng ist die Verknüpfung eigener Notizen mit dem bearbeiteten Text bzw. den verwendeten Büchern? Die Ergebnisse sind in die Optimierung des Prototypen eingeflossen, und weitere Iterationen von Nutzerstudien folgten; das Design wurde also stetig evaluiert und weiterentwickelt.

Anwendungsbereiche

Wir zielen mit dem Multitouch-Tisch in erster Linie darauf ab, eine neuartige Umgebung für komplexe literaturwissenschaftliche Arbeitsvorhaben zu entwickeln, vor allem für das Arbeiten mit Editionen, bei dem häufig eine Vielzahl von Büchern und Datenbanken gleichzeitig verwendet wird. Gleichzeitig haben wir festgestellt, dass auch die Arbeit des Editors durch die neue Technologie unterstützt werden kann, wenn etwa mehrere Handschriften-Digitalisate gleichzeitig nebeneinander angezeigt und mit den Händen einfach hin- und her-

geschoben werden können. Insbesondere Arbeitssituationen, die das parallele Arbeiten mit Büchern und Datenbanken erfordern, sollen unterstützt werden; wir denken hier in erster Linie an Literaturwissenschaftler, doch die Anwendungsbereiche sind durchaus auf andere Felder wie die Rechtswissenschaften erweiterbar, in denen die genaue Textanalyse eine hohe Bedeutung hat (vgl. Marshall et al., 2001). Auch Nutzungen für interaktive Ausstellungen in Bibliotheken, Archiven und Museen sind für die Zukunft denkbar.

Literaturverzeichnis

Adler, A., Gujar, A., Harrison, B. L., O'Hara, K., & Sellen, A. (1998). A Diary Study of Work-Related Reading: Design Implications for Digital Reading Devices. *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*, 241-248. doi:10.1145/274644.274679.

Deininghaus, S. (2010). *An Interactive Surface for Literary Criticism*. (Diploma thesis). Theses database, Media Computing Group, RWTH Aachen University, Germany.

Everitt, K. M., Morris, M. R., Brush, A. J. B., & Wilson, A. D. (2008). DocuDesk: An Interactive Surface for Creating and Rehydrating Many-to-Many Linkages among Paper and Digital Documents. *TABLETOP 2008: 3rd IEEE International Workshop on Horizontal Interactive Human Computer Systems*, 25-28. doi:10.1109/TABLETOP.2008.4660179.

Han, J. Y. (2005). Low-Cost Multi-Touch Sensing Through Frustrated Total Internal Reflection. *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*, 115-118. doi:10.1145/1095034.1095054.

Koike, H., Sato, Y., Kobayashi, Y., Tobita, H., & Kobayashi, M. (2000). Interactive Textbook and Interactive Venn Diagram: Natural and Intuitive Interfaces on Augmented Desk System. *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, 121-128. doi:10.1145/332040.332415.

Mackay, W. E., Pothier, G., Letondal, C., Bøegh, K., & Sørensen, H. E. (2002). The Missing Link: Augmenting Biology Laboratory Notebooks. *UIST '02: Proceedings of the 15th annual ACM symposium on User interface software and technology*, 41-50. doi:10.1145/571985.571992.

Marshall, C. C., Price, M. N., Golovchinsky, G., & Schilit, B. N. (2001). Designing E-Books for Legal Research. *JCDL '01: Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries*, 41-48. doi:10.1145/379437.379445.

Marshall, C. C. & Bly, S. (2005). Turning the Page on Navigation. *JCDL '05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*, 225-234. doi:10.1145/1065385.1065438.

Morris, M. R., Brush, A. J. B., & Meyers, B. R. (2007). Reading Revisited: Evaluating the Usability of Digital Display Surfaces for Active Reading Tasks. *TABLETOP 2007: 2nd Annual*

IEEE International Workshop on Horizontal Interactive Human-Computer Systems, 79-86. doi:10.1109/TABLETOP.2007.12.

Sellen, A. J. & Harper, R. H. R. (1997). Paper as an Analytic Resource for the Design of New Technologies. *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, 319-326. doi:10.1145/258549.258780.

Sellen, A. J. & Harper, R. H. R. (2003). *The Myth of the Paperless Office*. Cambridge, MA, USA: MIT Press.

Steimle, J., Khalilbeigi, M., & Mühlhäuser, M. (2010). Hybrid Groups of Printed and Digital Documents on Tabletops: A Study. *CHI EA '10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, 3271-3276. doi:10.1145/1753846.1753970.

Wu, C., Robinson, S. J., & Mazalek, A. (2008). Turning a Page on the Digital Annotation of Physical Books. *TEI '08: Proceedings of the 2nd international conference on Tangible and embedded interaction*, 109-116. doi:10.1145/1347390.1347414.

Yeh, R., Liao, C., Klemmer, S., Guimbretière, F., Lee, B., Kakaradov, B., Stamberger, J., & Paepcke, A. (2006). ButterflyNet: A Mobile Capture and Access System for Field Biology Research. *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, 571-580. doi:10.1145/1124772.1124859.

Digitale Geisteswissenschaften: Studieren!

Patrick Sahle

Cologne Center for eHumanities (CCeH)
Universität zu Köln
D-50923 Köln
sahle@uni-koeln.de

Zusammenfassung

Digitale Wissenschaften können sich nur entwickeln und durchsetzen, wenn sie auch in der Lehre vermittelt werden. Wie für andere Zweige auch bestehen dabei für die Geisteswissenschaften grundsätzlich zwei Optionen: Die Einbettung von speziellen Lehrinhalten in die bestehenden Curricula und die Einrichtung eigener Studiengänge. In dem international schon länger als „Digital Humanities" etablierten Forschungsbereich sind auch im deutschsprachigen Raum inzwischen einige Studiengänge entstanden, die eine Spezialisierung auf digitale Geisteswissenschaften verschiedener „Geschmacksrichtungen" ermöglichen.

Digitale Wissenschaft?

Die verschiedenen Zweige der Wissenschaft zeigen unterschiedliche Entwicklungen im Übergang zu „digitalen Wissenschaften". Die Situation in den Geisteswissenschaften ist dadurch geprägt, dass hier Schrift und Buch in allen Bereichen des wissenschaftlichen Prozesses von zentraler Bedeutung waren und immer noch sind. Während den auf Empirie, Abstraktion, Experiment, Modellierung und Berechnung ausgerichteten Wissenschaften computergestützte Verfahren gut zu entsprechen scheinen, haben die Geisteswissenschaften immer noch größere Probleme, ihre Arbeitsweisen in einer digitalen Medienumwelt abzubilden und unter ihren neuen Bedingungen weiter zu entwickeln.

Die Ausbildung digitaler Geisteswissenschaften wird maßgeblich auf der Ebene methodeninnovativer Forschungsprojekte vorbereitet. Die dabei langsam zu etablierenden konzeptionellen Grundlagen und praktischen Arbeitsweisen müssen dann auch Eingang in die alltägliche Forschungspraxis der Gelehrten finden. Vor allem müssen sie aber bereits in der Lehre vermittelt werden, um zu einer nachhaltigen Entwicklung zu führen. Hier bestehen eine ganze Reihe struktureller Probleme. Die Curricula und Lehrpläne sind immer schon „voll", eine Ergänzung um neue Module würde deshalb zu Lasten anderer, etablierter Inhalte gehen. Der Umstieg auf neue Arbeitsweisen wird oft als Aufgabe für die jüngeren, nachwachsenden Wissenschaftler bezeichnet, während die Lehre von den älteren, etablierten Forschern bestimmt wird. Zudem werden methodische und technische Bereiche der Wissenschaft vielfach als spezialistisch und nicht zum Kern des Faches gehörig betrachtet.

Die Lehre digitaler Forschung kann in den Geisteswissenschaften auf verschiedenen Ebenen angesiedelt werden. Innerhalb der regulären Fachstudien kann eine „eingebettete" Vermittlung technischer, konzeptioneller und methodischer Kompetenzen erfolgen. Dabei werden an inhaltlich exemplarischen Stoffen die allgemeinen Werkzeuge und Methoden zur Arbeit mit diesen Stoffen direkt mit gelehrt. Sinnvoller und wünschenswert erscheint aber

die gebündelte Vermittlung solcher Kompetenzen in expliziten Lehrveranstaltungen. Diese führen nicht zu einer Belastung bestehender Lehreinheiten und erlauben eine tiefere und übergreifende Beschäftigung mit einzelnen Aspekten moderner Forschung, erfordern aber zusätzliche Ressourcen für neues Lehrpersonal.

Digital Humanities (DH)

Ein dritter Weg wird dann beschritten, wenn die neuen Themenfelder und Lehrinhalte zu eigenständigen Studiengängen führen, der die verschiedenen geisteswissenschaftlichen Studien in Mehr-Fach-Programmen begleiten und bereichern können. Seit vielen Jahrzehnten ist dazu der Bereich der „Digital Humanities“ (früher: „Humanities Computing“; neuerdings auch „eHumanities“) einschlägig und z. B. durch gemeinsame Verbände, Fachzeitschriften, Handbücher, Mailinglisten und Kongressserien als Forschungsfeld konturiert und gefestigt.[1] Die Digital Humanities beschäftigen sich auf der praktischen Seite mit der Bearbeitung geisteswissenschaftlicher Fragestellungen mit digitalen Mitteln, vor allem auf der Projekt- inzwischen aber auch der Infrastrukturebene.[2] Dahinter steht aber eine mindestens ebenso wichtige Dimension, in der spezifische Methoden entwickelt und die konzeptionellen Grundlagen der Geisteswissenschaften unter den Bedingungen des Rechnereinsatzes und einer digitalen Medienumwelt reflektiert und in ein theoretisches Rahmenwerk eingefügt werden. Die Digital Humanities umgreifen dabei alle geisteswissenschaftlichen Fächer und beziehen sich auf alle ihre Gegenstände, Fragen, Methoden und Publikationsformen. Sie haben zugleich aber starke Anknüpfungspunkte an benachbarte Disziplinen wie die „Library and Information Sciences“ (u. a. mit den Arbeitsfeldern Bibliothek, Archiv und Museum), die allgemeine Informationswissenschaft und die Informatik. Dabei ist zu unterstreichen, dass es den Digital Humanities grundsätzlich weniger um die Anwendung bestimmter Techniken und Methoden, als um die Entwicklung und theoretische Reflektion neuer Lösungsansätze geht. Das Feld hat in den letzten Jahrzehnten einen so breiten Kanon an Konzepten, Verfahren und Standards hervorgebracht, dass er ein eigenes Studienfach nicht nur trägt, sondern hier sogar wiederum die Fokussierung auf bestimmte Bereiche nahe legt.

Die Zusammenführung der Entwicklungen in einem eigenständigen Bereich „Digital Humanities“ hat weitere positive Nebeneffekte: Erstens sind die Probleme auf der technischen und methodischen Ebene ihrer Natur nach fast immer fachübergreifend, sehr häufig aber spezifisch geisteswissenschaftlich. Zweitens entsteht in den einzelnen Fächern oft nicht die kritische Masse an interessierten Forschern, die für eine lebendige Methodendiskussion nötig ist. Und drittens verhindert die interdisziplinäre Zusammenarbeit die Ausbildung von Insellösungen und Parallelentwicklungen, während sie die gegenseitige Nutzbarkeit der Forschungsleistungen stärkt.

[1] Einen systematischen Einblick in die „Landschaft“ der Digital Humanities versucht zuletzt Svensson (2010). Einen guten Einblick in die aktuellen Themen und Tendenzen geben die Abstract-Bände der jährlichen Zentralkonferenz „Digital Humanities“, siehe hier zuletzt den Band zur Tagung in London, Digital Humanities 2010 (2010).

[2] Für die Bemühungen im Infrastrukturbereich siehe vor allem das EU-weite Projekt DARIAH, dem sich ab 2011 ein deutsches DARIAH-DE-Projekt anschließen wird.

Die gegenwärtige Situation

Im Bereich der lehrenden Vermittlung von Ansätzen der Digital Humanities waren lange Zeit nur einige internationale Zentren führend. Hier wurden vor allem unregelmäßige „Schools“, aber auch schon spezialisierende MA- und PhD-Programme angeboten.[3] Erst in der letzten Dekade hat die Lehre allmählich an Breite gewonnen. Und erst in diesen Jahren wird das Thema der Curricula und Studiengänge systematisch angegangen, und es entstehen an vielen Orten neue Studienmöglichkeiten.[4] In diesem Zusammenhang hat die Historisch-Kulturwissenschaftliche Informationsverarbeitung (HKI) und das Cologne Center for eHumanities (CCeH) an der Universität zu Köln 2009 eine Initiative gestartet, um die verschiedenen Studiengänge und Programme an einen Tisch zu bringen und die weiteren Entwicklungen in diesem Feld durch stärkere Abstimmung und durch die Verständigung auf gemeinsame Grundlagen zu fördern.[5] Dabei ist die Identifikation von Akteuren im Bereich der Digital Humanities ebenso schwierig, wie die Definition von Lehrangeboten als zum Bereich der digitalen Geisteswissenschaften gehörig. Eine der Ursachen hierfür liegt darin, dass alle Wissenschaftler, die hier engagiert sind, einen Hintergrund entweder in einem der traditionellen geisteswissenschaftlichen Fächer, den „Library and Information Sciences“ oder der Informatik haben (müssen) und sich erst nach ihrer Ausbildung und während ihrer wissenschaftlichen Karriere den Digital Humanities zugewandt und dabei ihre ursprüngliche Identifikation durchaus nicht aufgegeben haben.

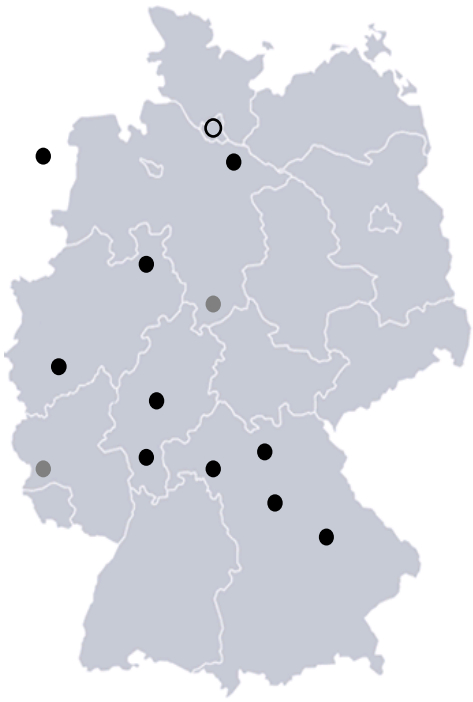


Abbildung 1: Digital-Humanities-Studienstandorte 2010. (Graue Kreise: Studiengänge in Vorbereitung; nicht ausgefüllte Kreise: nur studienbegleitende Module)

Eine eindeutige Zuordnung über die Benennung des eigenen Arbeitsfeldes und der Lehrangebote ist ebenfalls kaum möglich, da der Begriff „Digital Humanities“ erst langsam anfängt, sich als Sammelbegriff zu etablieren. Er wird immer noch begleitet von teilweise schon älteren spezielleren Bezeichnungen wie Computerlinguistik, Computerphilologie, Texttechnologie, Kulturerbeinformatik, (z. B. „historische“) Fachinformatik, Medieninformatik, Informationsverarbeitung oder Archäoinformatik.

Diese historisch bedingten Konfigurationen und Identifikationen haben dazu geführt, dass zwar an verschiedenen Standorten Lehrangebote entstanden sind, die sich selbst den „Digital Humanities“ im weiteren Sinne zuordnen, die Studienmöglichkeiten zugleich aber äußerst individuell sind. Es lassen sich derzeit kaum zwei Standorte finden, die ein auch nur

[3] Siehe exemplarisch Sinclair und Gouglas (2002).

[4] Einen größeren Überblick wird Hirsch (2011) geben.

[5] Siehe den Nachweis der Treffen auf http://www.cceh.uni-koeln.de/veranstaltungen.

annähernd gleiches Angebot machen würden. Die Beschreibung des gesamten Feldes kann deshalb nur über die Unterschiede erfolgen, die sich aus verschiedenen Betrachtungsweisen heraus ergeben.

Beschreibungsdimensionen

Zur Auslotung der Studienmöglichkeiten kann die Frage gestellt werden, wie sich die Digital Humanities an den verschiedenen Standorten zu den Humanities verhalten und wie sie in ihre geisteswissenschaftliche Umgebung eingebettet sind. Zur genaueren Beschreibung und zum Vergleich der Ausbildungsgänge lassen sich vier Betrachtungsrichtungen untersuchen.

Wann studiert man die digitalen Geisteswissenschaften?

Die Digital Humanities stellen in Bezug auf die Humanities eine Spezialisierung dar oder bilden einen methodischen Überbau. Insofern kann auf einen allgemeinen BA aus den Geisteswissenschaften ein reiner Digital-Humanities-MA folgen. Häufiger vertreten ist inzwischen aber ein kontinuierliches Studium. Dies kann die Belegung von speziellen Modulen über den gesamten Studienverlauf hinweg bedeuten oder das Studium eines Zwei-Fach-BA/MA, bei dem eine Geisteswissenschaft mit einem Digital-Humanities-Fach kombiniert wird. Theoretisch möglich ist außerdem die Kombination eines DH-BA mit einem anschließend rein inhaltlich ausgerichteten MA.

Welchen Anteil haben digitale Ansätze innerhalb der geisteswissenschaftlichen Studien?

Auf der Ebene von Modulen und Zertifikaten findet eine Ausbildung zusätzlich zu den eigentlichen Fachstudien statt. Als explizites Teilfach kann z. B. Informatik manchmal nur als Nebenfach zum geisteswissenschaftlichen Hauptfach belegt werden. In der Regel ist aber auch eine klare Schwerpunktsetzung möglich, bei der in einem Digital-Humanities-Studiengang ein geisteswissenschaftliches Nebenfach gewählt wird.

Ist die Ausrichtung eher technisch oder methodisch? Eher theoretisch oder eher praktisch?

Die gründliche Kenntnis verschiedener Technologien bis hin zu Programmiersprachen ist in allen Studiengängen ein wichtiger Bestandteil. Dagegen wird den methodischen Rückwirkungen von Technologien auf die Fragestellungen und Forschungspraktiken in den Geisteswissenschaften ein sehr unterschiedliches Gewicht gegeben. Absolventen der in Frage stehenden Studiengänge sollten auf jeden Fall in der Lage sein, digitale Projekte mit geisteswissenschaftlichem Hintergrund zu realisieren. Es bleibt dann der persönlichen Ausrichtung überlassen, ob und in welchem Maße die eigene Arbeit sich auch auf die Weiterentwicklung der Methoden und die Theoriebildung bezieht.

Welcher Bereich der Geisteswissenschaften soll im Zentrum stehen?

Vielfach sind die einzelnen Digital-Humanities-Standorte aus einer bestimmten Fachperspektive herausgewachsen. Schon traditionell kann deshalb zwischen verschiedenen „Geschmacksrichtungen“ unterschieden werden. Hierbei ist neben den großen Lagern der Linguistik, der Literaturwissenschaften und der historischen Wissenschaften vor allem auch noch der Bereich des Kulturerbes (Bibliotheken, Archive, Museen) zu nennen. Weniger

fachorientiert ist die Ausrichtung dann, wenn z. B. die allgemeine Informationswissenschaft oder die Informatik institutioneller oder personeller Träger der Ausbildung ist.

Perspektiven für die Zukunft

Immer noch besteht das Problem, dass vielen Akteuren auf der wissenschaftlichen und der wissenschaftspolitischen Ebene unbekannt ist, dass es auf dem Feld der Geisteswissenschaften eine Spezialdisziplin gibt, die sich fast ausschließlich mit der Lösung geisteswissenschaftlicher Probleme in der angewandten Forschung beschäftigt. Ebenso ist den wenigsten Studienanfängern klar, dass für sie die Möglichkeit besteht, ihre fachlichen Interessen mit einem Studiengang zu verbinden, der ihnen nicht nur neue Werkzeuge und methodische Optionen an die Hand gibt, sondern zugleich ihre Berufsaussichten deutlich verbessert. Diese Probleme können nur behoben werden, wenn das Fach insgesamt besser sichtbar gemacht wird. Bausteine dazu sind u. a. eine gemeinsame Basisdefinition der Digital Humanities, ein gemeinsames Informationsportal und eine gemeinsame Werbung für die verschiedenen Studiengänge. Im Rahmen der oben angesprochenen Initiative soll außerdem ein Kern- und ein Referenzcurriculum entwickelt werden. Dabei kann das Kerncurriculum deutlich machen, was allen Programmen trotz ihrer abweichenden Betitelung gemein ist, während das Referenzcurriculum die ganze Breite der Lehrinhalte aufzeigen kann, zu denen sich ein bestimmter Studiengang auf eine bestimmte Weise positioniert.

Bibliografie

Digital Humanities 2010 Conference Abstracts (2010). King's College London, July 7th-10th, 2010. Office for Humanities Communication. Retrieved from http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/pdf/book-final.pdf.

DARIAH – Digital Research Infrastructure for the Arts and Humanities. (2006). Retrieved Dezember 30, 2010, from http://dariah.eu/.

Hirsch, B.D. (Ed.). (2011). Teaching digital humanities: Where are we now? Ann Arbor: University of Michigan Press.

Sinclair, S., & Gouglas, S.W. (2002). Theory into practice: A case study of the humanities computing master of arts programme at the University of Alberta. Arts and Humanities in Higher Education 1/2, 167-183. Retrieved from http://ahh.sagepub.com/content/1/2/167.full.pdf+html.

Svensson, P. (2010). The landscape of digital humanities. Digital Humanities Quarterly, 4(1). Retrieved from http://www.digitalhumanities.org/dhq/vol/4/1/000080/000080.html.

Anhang: Übersicht über die Studienmöglichkeiten

Bamberg: (1.) „Computing in the Humanities" als Master of Science (MSc); (2.) „Angewandte Informatik" als Nebenfach in geisteswissenschaftlichen BA-Studiengängen.

Bielefeld: (1.) „Texttechnologie" als BA-Nebenfach; (2.) Texttechnologische Module in anderen Studiengängen.

Darmstadt: (1.) MA "Linguistic and Literary Computing";(2.) Informatik im BA „Joint Bachelor of Arts".

Erlangen: (1.) BA an der Philosophischen Fakultät mit Informatik als erstem Fach; (2.) geisteswissenschaftlich ausgerichtete Informatik-Module für andere MA-Programme.

Gießen: „Computerlinguistik und Texttechnologie" als Haupt- oder Nebenfach im MA „Sprache, Literatur, Kultur".

Göttingen: BA/MA in Kombination aus Informatik und Geisteswissenschaftlichem Fach (in Vorbereitung).

Graz: (1.) Digital-Humanities-Module und Zertifikat „Informationsmodellierung in den Geisteswissenschaften" in verschiedenen Fächern. (2.) MA „EuroMACHS" (Media, Arts and Cultural Heritage Studies).

Groningen: BA „Informatiekunde" (Informatik für Geisteswissenschaftler).

Hamburg: Module und Zertifikat „Computerphilologie" für verschiedene Studiengänge.

Köln: (1.) BA/MA „Informationsverarbeitung"; (2.) „Medieninformatik" im BA/MA Medienwissenschaft; (3.) MA „EuroMACHS" (Media, Arts and Cultural Heritage Studies); (4) „IT-Zertifikat der Philosophischen Fakultät" als Begleitstudium im Rahmen des Studium Integrale.

Lüneburg: BA-Minor „Digitale Medien / Kulturinformatik" (z. B. mit Major „Kulturwissenschaften").

Regensburg: (1.) BA/MA „Informationswissenschaft"; (2.) BA „Medieninformatik"; (3.) Digital-Humanities-Studieneinheiten in anderen Fächern.

Trier: Digital-Humanities-BA (in Planung).

Würzburg: BA/MA „Digital Humanities".

Wissenschaftskommunikation und Web 2.0

Social Network Sites in der Wissenschaft

Michael Nentwich

Österreichische Akademie der Wissenschaften
Institut für Technikfolgen-Abschätzung
Strohgasse 45/5
A-1030 Wien
mnent@oeaw.ac.at

Zusammenfassung

Auch Wissenschaftler und deren Institutionen sind in den bekannten social network sites (SNS) wie LinkedIn oder Facebook präsent und teilweise hoch aktiv. Seit ca. 2008 haben sich daneben speziell auf Wissenschaftler abzielende SNS entwickelt, die bemerkenswerten Zulauf haben und spezielle Dienste für diese besondere Zielgruppe anbieten (etwa in Hinblick auf die Suche und Verwaltung von Publikationen, gemeinsame Schreibumgebungen, Jobbörsen oder institutionelle Verzeichnisbäume). Der Beitrag beschreibt Typen und Funktionalitäten von SNS und geht speziell der Frage nach, welche Rolle SNS in der Wissenschaft bereits spielen und in Zukunft spielen könnten.

Einleitung

Dass Wissenschaftler heute zum größten Teil digital kommunizieren und kooperieren, im Internet recherchieren und elektronisch publizieren, ist ein Befund, der sich bereits Anfang dieses Jahrhunderts abzeichnete (vgl. Nentwich, 2003). Der digitale Kommunikationsraum, der durch das Internet aufgespannt wird, ist ständigen Veränderungen unterworfen, nicht nur was die Netzwerkknoten (Personen, Institutionen, Datenbanken etc.), sondern auch und insbesondere, was die Dienste und Anwendungen anlangt, die diese Kommunikation über das Internet mitgestalten. Gerade jene Plattformen, die unter dem Begriff des Web 2.0 zusammengefasst werden, haben in den letzten Jahren zu einem großen Innovationsschub geführt, der sogar völlig neue Kommunikationsformen hervorgebracht hat, etwa Microblogging. Diese Entwicklung ist auch an der Wissenschaft nicht spurlos vorbei gegangen, im Gegenteil: Viele Forscher „twittern“, beteiligen sich an Wikipedia, „bloggen“ oder vernetzen sich über Facebook und ähnliche Plattformen (sogenannte *social network sites* – SNS) – vgl. dazu Nentwich (2009). Der folgende Beitrag geht speziell der Frage nach, welche Rolle SNS in der Wissenschaft bereits spielen und in Zukunft spielen könnten. Empirische Basis dieser Untersuchung ist eine Internet- und Literaturrecherche sowie insbesondere die teilnehmende Beobachtung in etlichen solcher Netzwerke seit Mitte 2009. Es werden rein wissenschaftliche und allgemeine SNS, die auch wissenschaftlich genutzten werden, mit ihren typischen Funktionalitäten vorgestellt; Nutzungsintensitäten und -arten diskutiert; sowie darauf aufbauend erste Überlegungen zur Funktionalität von SNS im Wissenschaftsbetrieb angestellt.

Dieser Beitrag stellt einige vorläufige Ergebnisse des Projektverbundes „Interactive Science“ (2008–2011), gefördert von der VW-Stiftung, zum Thema „Interne Wissenschaftskommuni-

kation über digitale Medien“[1] dar. Dessen Teilprojekt 1 befasst sich unter dem Titel „Kollaboratives Wissensmanagement und Demokratisierung von Wissenschaft“ in erster Linie mit neuen Diensten und Anwendungen, die im weitesten Sinne dem Web 2.0 zuzurechnen sind. Bisher sind das unter anderem: Second Life (König & Nentwich, 2008), Wikipedia/ Wikiversity/Wikibooks (König & Nentwich, 2009), Microblogging (Herwig, Kittenberger, Nentwich, & Schmirmund, 2009), Google/Google Scholar/Google Books (König & Nentwich, 2010) und eben SNS (Nentwich & König, 2011).

Typen und Funktionalitäten von SNS

Trotz oder gerade wegen der Neuheit des Phänomens, gibt es in der Literatur bereits zahlreiche mehr oder weniger enge Definitionsvarianten für SNS. Viele der in diesem Zusammenhang genannten Web 2.0-Anwendungen *dienen* dazu Menschen im weitesten Sinne *sozial* zu *vernetzen*. Da selbst E-Mailing im Prinzip diese Funktion erfüllt, erscheint es sinnvoll, die sogenannten „Profile“ zu fokussieren, also jene Selbstbeschreibungen mit mehr oder weniger detaillierten persönlichen Informationen, die über das bloße Anlegen eines User-Accounts hinausgehen. Bei SNS stellen die Profile die zentralen Knotenpunkte dar und nicht etwa die kommunikativen Inhalte (z. B. Microblogs in der Timeline) oder sonstiger Content (Videoclips, Fotos oder Links). Als Kernfunktionen von SNS sind dementsprechend das Identitätsmanagement, die Kontaktverwaltung und das Networking bestimmend. SNS können sowohl dazu dienen, (bestehende offline) Netzwerke abzubilden oder zur Bildung neuer (online) Netzwerke beizutragen.

Es gibt mittlerweile eine beeindruckend hohe Anzahl von unterschiedlichen SNS, die sich nicht nur in der Anzahl der Mitglieder (von unter 100 bis zu 500 Millionen) unterscheiden, sondern die man auch nach Nutzungsart, Zugangsvoraussetzung und Funktionalitäten grob typisieren kann. Nach den *Nutzungsformen* kann man etwa zwischen rein privat genutzten SNS (z. B. FamilyLounge), rein beruflichen (z. B. LinkedIn) oder gemischt verwendeten SNS (z. B. Facebook) unterscheiden. Auch die *Zugangsvoraussetzungen* unterscheiden sich stark: Während einige prinzipiell offen sind (z. B. ResearchGATE), sind manche zumindest teilweise kostenpflichtig (z. B. Premium Xing) und andere insofern geschlossen, als man ihnen nur per Einladung durch ein Mitglied beitreten kann (z. B. Science 3.0).

SNS unterscheiden sich weiterhin nach den angebotenen *Funktionalitäten*. Während manche praktisch bei allen vorzufinden sind, gibt es bei einigen auch zielgruppenspezifische Funktionen. Wie bereits oben erwähnt bilden *Profile* den Kern eines SNS, das sind quasi ausführliche Visitenkarten mit weiterführenden Hyperlinks, die nicht nur für Personen, sondern teilweise auch für Institutionen oder Gruppen erstellt werden können. Ebenfalls Kernbestand jedes SNS sind die Funktionen zur *Vernetzung*, also der Herstellung von zweiseitigen Kontakten („Befreunden“) oder eines einseitigen Beobachtungsstatus („Folgen“ oder „Fan-Status“), die Darstellung, zum Teil aufwändige Visualisierung des Netzwerks und die Suche nach anderen Netzwerkteilnehmern. Alle SNS bieten verschiedene *Kommunikationskanäle* an, von internem Mail, Chat, Foren bis zu Microblogging (Statusmeldungen). Wesentlicher Bestandteil sind Funktionen zur Aufmerksamkeitslenkung, damit die Nutzer

[1] http://wissenschaftskommunikation.info.

die teils vielfältigen Aktivitäten in ihrem persönlichen Netzwerk effizient verarbeiten können sollen. Dies sind Aktivitäten, die teilweise durch andere im SNS aktiv ausgelöst und dann ausgewählten Netzwerkmitgliedern bekannt gegeben werden, etwa Kommentare, das Klicken auf „Gefällt-mir-Buttons“, das sogenannte Teilen von Ressourcen, das Bewerten (Rating) von Aktivitäten anderer usw. Teilweise wird die Aufmerksamkeitslenkung auch automatisiert durch die Plattform hergestellt, etwa durch die externe Benachrichtigung zu bestimmten Aktivitäten per E-Mail, durch automatisch generierte Hinweise auf „verwandte“ Themen oder Personen (mit ähnlichen Profilangaben oder vergleichbaren Aktivitäten). Viele SNS bieten auch die Gründung von *Gruppen* an (geschlossen oder offen), in denen sich die Gruppenmitglieder halböffentlich oder unter sich in einem Forum austauschen können; teilweise wird dies durch Spezialfunktionen wie ein gemeinsames Dokumentenarchiv u. ä. unterstützt. Funktionen rund um *Veranstaltungen*, wie die Unterstützung bei der Terminvereinbarung einschließlich Zu- und Absagen oder das gemeinsame Führen eines Kalenders ergänzen die Palette. Darüber hinaus gibt es, wie erwähnt, *Spezialfunktionen*, etwa literaturbezogene in wissenschaftsspezifischen SNS zur Unterstützung bei der Suche nach wissenschaftlichen Veröffentlichungen oder zu passenden Publikationsorganen („similar abstract search“); auch Neuerscheinungsdienste oder der Volltext-Upload von Publikationen werden mitunter angeboten. Einzelne SNS stellen auch eine Jobbörse oder eine eigene Blogging-Plattform zur Verfügung.

Allgemeine und wissenschaftsspezifische SNS

Wie einleitend angedeutet, haben auch bereits viele Wissenschaftler Profile in SNS. Abstrakt kann man zwischen Plattformen, die allgemeinerer Natur sind, aber nicht nur privat, sondern auch (teilweise) wissenschaftlich genutzt werden, einerseits, und rein wissenschaftlichen SNS, andererseits, unterscheiden. Bei letzteren findet man solche, die prinzipiell universell und disziplinenübergreifend angelegt sind, und solche, die lokal bzw. disziplinen- oder themenspezifisch orientiert sind. Die folgende Übersicht zeigt einen Ausschnitt aus der Vielzahl solcher SNS:

Übersicht: Ausgewählte wissenschaftliche bzw. wissenschaftlich genutzte SNS

Name	Anzahl der Mitglieder	Gründungs-jahr	Name	Anzahl der Mitglieder	Gründungs-jahr
Allgemeine, tw. auch wissens. SNS			research. iversity	9.400	2008
Facebook	500.000.000	2004	LabRoots	4.000	2008
MySpace	100.000.000	2004	Epernicus	20.000	2008
LinkedIn	80.000.000	2003	Research Cooperative	NA	2008
VZ-Gruppe	17.000.000	2005	ScholarZ	393	2008
Xing	10.000.000	2003	Ways.org	NA	2004
Wissenschaftliche SNS			Scispace.net	NA	2007
ResearchGATE	700.000	2008	myExperiment	3.500	2007
Mendeley	676.000	2009	Vivo 40.841	2010	2004
Sciencestage	270.000	2008	Science 3.0	230	2010
Academia.edu	211.000	2008	AtmosPeer	125	2010
Nature Network	25.000	2007	EPTA Ning	83	2008

Quelle: Recherche Nentwich und König, Stand: Dezember 2010

Diese Übersicht ist noch in Bearbeitung (vorläufige Endfassung in: Nentwich & König, 2011). Auffällig ist, dass die allgemeinen SNS bereits um das Jahr 2003 auftauchen, während die wissenschaftsspezifischen erst um das Jahr 2007 und später gegründet werden (vielleicht mit der Ausnahme von Ways.org, das 2003 gegründet wurde und sich kontinuierlich ab 2004 in ein SNS weiterentwickelt hat). Seit 2009 kann man auch für den Wissenschaftsbereich von einem Boom sprechen: So ist die Zahl der Mitglieder in ResearchGATE beispielsweise ab Sommer 2009 von 150.000 auf über 500.000 (Stand: Oktober 2010) gestiegen.

Trotz der teils beachtlichen Mitgliederzahlen, sind die beobachteten *Nutzungsintensitäten* höchst unterschiedlich. Nicht nur im wissenschaftlichen Bereich kann man zumindest folgende fünf Nutzungstypen unterscheiden: Viele, möglicherweise die Mehrzahl der Mitglieder sind praktisch „Karteileichen“, die lediglich einen Account angelegt und das absolute Minimum an erforderlichen Informationen eingegeben haben, um das Netzwerk zu testen („Me-too-Präsenz“). Eine ebenfalls größere Gruppe nutzen SNS gleichsam lediglich als „Virtuelle Visitenkarte“, geben also weit mehr als das Minimum von sich preis, um damit auch in diesem Netzwerk gefunden zu werden, verwenden aber die weitergehenden Funktionen nicht. Dieser Typ Nutzer pflegt das Profil nur rudimentär und loggt sich selten wieder auf der Plattform ein. Demgegenüber gibt es eine Gruppe Nutzer, die zumindest Vernet-

zungsanfragen positiv beantworten (sog. „passive Vernetzung“), aber sonst nur wenig aktiv sind. Eine vermutlich wachsende Anzahl von Nutzern von SNS, aber nach unserer Beobachtung eine Minderheit, vernetzt sich aktiv und kommuniziert regelmäßig. Dies sind gleichsam die Early Adopters, die ausprobieren, wie man diese Instrumente einsetzen kann. Deren Netzwerk ist in der Regel deutlich größer (mehr als zehn Kontakte). Die aktivste Gruppe ist jene der „Cyber-UnternehmerInnen“ (Nentwich, 2003), die sich auch an der Weiterentwicklung und Gestaltung (zumindest per Feedback an die Entwickler) beteiligen, Gruppen gründen, als Moderatoren von Foren auftreten, viele Statusmeldungen absetzen usw. Die Vertreter dieser kleinen Gruppe versuchen das neue Medium nicht nur aktiv zu nutzen, sondern mitzugestalten und andere zu motivieren sich einzubringen.

Insbesondere im wissenschaftlichen Feld sind folgende *Nutzungsarten* beobachtbar: Wissenschaftler pflegen in SNS ihre Kontaktliste. Sie verwenden sie als multilateralen Kommunikationskanal und Forum zum Austausch, ähnlich wie E-Mail-Diskussionslisten. Des Weiteren bieten sich SNS als bilateraler Kommunikationskanal (insb. internes Mail oder Chat) an, um mit Personen Kontakt aufzunehmen. SNS dienen weiter als Microblogging-Plattform, allerdings bislang unter Wissenschaftlern oft als Ort der Zweitveröffentlichung von Meldungen, die bereits über andere Kanäle (z. B. Twitter) verschickt wurden. Potenziell eignen sich manche SNS auch zur Zusammenarbeit, wenn und soweit sie Groupware-Funktionen anbieten, sowie auch als E-Learning-Plattform für die Kommunikation zwischen Lehrenden und Studierenden. Nicht zu unterschätzen ist schließlich, dass sich allgemeine SNS aufgrund der steigenden Nutzungszahlen auch als Kanal für die Öffentlichkeitsarbeit eignen, sowohl für Individuen (Selbstmarketing) als auch für Forschungseinrichtungen.

Zur potenziellen Rolle von SNS in der Wissenschaft

Um abschätzen zu können, ob SNS zukünftig eine bedeutsame Rolle im Wissenschaftsbetrieb spielen werden, muss die Frage gestellt werden, ob das Angebot für die Wissenschaft funktional ist und welche Potenziale und Hemmnisse im Diffusionsverlauf bestehen. Die folgenden Analysebausteine werden bei dieser Abschätzung eine Rolle spielen (genauer in Nentwich & König, 2011):

Anonymität und Pseudonymität: Anonymität ist bislang in SNS nicht implementiert, wäre aber durchaus funktional bei wissenschaftlicher Nutzung, etwa wenn es um Qualitätssicherung (peer review) oder um Brainstorming geht, bei dem die nachträgliche Zuordnung von einzelnen Ideenbausteinen vorab hemmend sein könnte. Demgegenüber ist Pseudonymität in allgemeinen SNS durchaus üblich, aber im wissenschaftlichen Kontext kontraproduktiv: Einerseits wollen Wissenschaftler in aller Regel nur mit realen Kollegen kommunizieren und kooperieren, andererseits gibt es einen Zusammenhang zwischen dem Status, den man sich innerhalb eines SNS erarbeitet, und jenem in der realen Welt, freilich nur, wenn die Identität der Personen eindeutig ist.

Multikanalität: Die Vielfalt der Kommunikationskanäle stellte schon bisher die Wissenschaftler vor die Herausforderung, den Überblick zu bewahren. Mit dem Aufkommen von SNS ist diese zumindest vorübergehend noch größer geworden, da insbesondere nicht alle Kollegen dasselbe SNS nutzen und man daher genötigt ist, in mehreren SNS gleichzeitig präsent zu sein. Dieses Problem könnte entweder durch (disziplinspezifische) Monopolbildung

(alle Kollegen beim selben Netzwerk) oder durch Schnittstellenharmonisierung gelöst werden (Nutzer abonnieren einen Metadienst, der die Aktivitäten aus den verschiedenen Netzwerken in einem Interface zusammenführt). Weiter könnten sich multifunktionale SNS herausbilden, die alle Funktionalitäten (von der Kommunikation bis zur Kollaboration) in einer Plattform vereinen (One-Stop-Service).

Virtuelle Zusammenarbeit: Der hohe Bedarf an technisch ausgereiften Diensten zur einfachen, hoch effizienten Kollaboration unter örtlich verteilten Wissenschaftlern, könnte prinzipiell durch SNS gedeckt werden. Das Potenzial scheint hoch, was nicht zuletzt an der Vielzahl wissenschaftsspezifischer SNS erkennbar ist. Es wird aber bislang gering genutzt, was einerseits auf noch bestehende technische Probleme, andererseits auf mangelnde Erfahrung mit Groupware bzw. SNS, aber auch auf das weitgehende Fehlen einer entsprechenden „Kultur der verteilten Zusammenarbeit“ zurückzuführen ist. Das oben erwähnte Problem der Multikanalität, also der Aufmerksamkeitsstreuung, trägt ebenfalls dazu bei.

Privacy: SNS bringen den Mitgliedern dann am meisten Nutzen, wenn sie möglichst viel von sich preisgeben und besonders aktiv sind. Da diese Aktivitäten notwendigerweise digital festgehalten und zur Erfüllung des Zwecks von SNS laufend verknüpft werden, entstehen besonders aussagekräftige Profile, die unter Datenschutz und Privacy-Gesichtspunkten prinzipiell bedenklich sind. Solange es zu keiner Vermischung von privaten und beruflichen (hier: wissenschaftlichen) Aktivitäten kommt, erscheint dies im beruflichen Kontext weniger problematisch, auch wenn die Angaben über die berufliche Aktivitäten zweifellos transparenter werden und damit weniger zielgruppenspezifisch steuerbar sind. In Zukunft könnte sich die Problematik freilich verschärfen, wenn etwa das Leseverhalten direkt ausgewertet würde („use tracking“), ganz zu schweigen von der von Atkinson (1996) erdachten „control zone“ mit nach Nutzergruppen geschichteten Ratings.

Vorläufiges Fazit

Angesichts der erst seit ein bis zwei Jahren einsetzenden, verstärkten Nutzung und der damit verbundenen noch mangelnden Verfügbarkeit von aussagekräftigen (qualitativen und quantitativen) Nutzungsstudien, ist es für eine umfassende Einschätzung noch zu früh. Die Nutzungszahlen steigen, aber wir befinden uns noch am Anfang der möglichen Diffusions-S-Kurve, deren genauer Verlauf noch unabsehbar ist. Es handelt sich um einen hochgradig dynamischen Bereich, in dem sich bereits sehr viele unterschiedliche SNS mit einer großen Bandbreite, auch und gerade im engeren wissenschaftlichen Feld, entwickelt haben. Die Funktionalitäten dieser SNS sind demgegenüber noch keineswegs konsolidiert. Funktionale Potenziale für die Wissenschaft sind zweifellos vorhanden, da Kommunikation, Kooperation und Vernetzung zu deren zentralen Aktivitätsformen zählen. Es ist somit zu erwarten, dass sich langfristig SNS in irgendeiner Form für die Wissenschaftskommunikation etablieren werden.

Literatur

Atkinson, R. (1996). Library Functions, Scholarly Communication, and the Foundation of the Digital Library: Laying Claim to the Control Zone. *The Library Quarterly, 66*(3), 239-265.

Herwig, J., Kittenberger, A., Nentwich, M. & Schmirmund, J. (2009). *Twitter und die Wissenschaft. Steckbrief 4 im Rahmen des Projekts "Interactive Science"*. Dezember, http://epub.oeaw.ac.at/ita/ita-projektberichte/d2-2a52-4.pdf.

König, R. & Nentwich, M. (2008). *Wissenschaft in "Second Life". Steckbrief 1 im Rahmen des Projekts Interactive Science*. November, http://epub.oeaw.ac.at/ita/ita-projektberichte/d2-2a52-1.pdf.

König, R. & Nentwich, M. (2009). *Wissenschaft in Wikipedia und anderen Wikimedia-Projekten. Steckbrief 2 im Rahmen des Projekts Interactive Science*. April, http://epub.oeaw.ac.at/ita/ita-projektberichte/d2-2a52-2.pdf.

König, R. & Nentwich, M. (2010). *Google, Google Scholar und Google Books in der Wissenschaft. Steckbrief 3 im Rahmen des Projekts Interactive Science*. Mai, http://epub.oeaw.ac.at/ita/ita-projektberichte/d2-2a52-4.pdf.

Nentwich, M. (2003). *Cyberscience: Research in the Age of the Internet*. Vienna: Austrian Academy of Sciences Press. http://hw.oeaw.ac.at/3188-7.

Nentwich, M. (2009). *Cyberscience 2.0 oder 1.2? Das Web 2.0 und die Zukunft der Wissenschaft*. November, http://epub.oeaw.ac.at/ita/ita-manuscript/ita_09_02.pdf.

Nentwich, M. & König, R. (2011). *Wissenschaft und Social Network Sites. Steckbrief 5 im Rahmen des Projekts Interactive Science*. Jänner, http://epub.oeaw.ac.at/ita/ita-projektberichte/d2-2a52-5.pdf.

Von der Schulbank in die Wissenschaft. Einsatz der Virtuellen Arbeits- und Forschungsumgebung von Edumeres.net in der internationalen Bildungsmedienforschung

Sylvia Brink, Christian Frey, Andreas L. Fuchs, Roderich Henrÿ, Kathleen Reiß & Robert Strötgen

Georg-Eckert-Institut für internationale Schulbuchforschung
Celler Straße 3
D-38114 Braunschweig
edumeres@gei.de

Zusammenfassung

Informationen sammeln, Wissen generieren und Ergebnisse publizieren sind klassische Schritte im Forschungsprozess von Natur- und Geisteswissenschaften. Das Informations- und Kommunikationsportal Edumeres.net mit seiner Virtuellen Forschungsumgebung unterstützt Wissenschaftler dabei von der Ideenfindung über die kollaborative Textgenerierung bis hin zur fertigen, zitierbaren Publikation. Ein Fallbeispiel aus der internationalen Schulbuchforschung stellt die Plattform vor.

Die Bildungsmedienforschung und das Georg-Eckert-Institut

Bildungsmedienforschung ist ein Disziplinen und Ländergrenzen übergreifendes Forschungsfeld, dessen Arbeit sich an der Schnittstelle von Wissenschaft, Bildungspolitik und Praxis bewegt. Ein so heterogen ausgerichtetes Forschungsgebiet erfordert für kollaboratives Forschen und Arbeiten eine Infrastruktur, die durch Konferenzen und traditionelle Formen der Kommunikation nur teilweise abgedeckt werden kann.[1]

Dem Georg-Eckert-Institut für internationale Schulbuchforschung mit Sitz in Braunschweig ist es deshalb daran gelegen fehlende Kommunikationsformen zu ergänzen und bestehende zu verbessern. Die Möglichkeiten des Internets sowie die Entwicklungen der letzten Jahre auf dem Gebiet des Social-Networkings und der kollaborativen Wissensgenerierung bieten dafür gute Voraussetzungen.[2]

[1] „Der Wissenschaftler der Zukunft greift von einem beliebigen Standort auf (s)eine virtuelle Umgebung zu und findet dort Programme, Forschungsdaten und Sekundärquellen (wie Publikationen, Datenbanken und Dienste) vor, die er für seine aktuelle Arbeit braucht.“ Neuroth, H., Aschenbrenner, A., & Lohmeier, F. (2007). e-Humanities – eine Virtuelle Forschungsumgebung für die Geistes-, Kultur- und Sozialwissenschaften. *Bibliothek. Forschung und Praxis*, 31(3), 272-279. S. 273.

[2] „Die Naturwissenschaften haben seit einigen Jahren eine neue Tradition der kollaborativen Arbeitsweise entwickelt und sind es heutzutage schon eher gewohnt (international) vernetzt, unter Einbeziehung neuester Technologien und Infrastrukturen, zu arbeiten. Die Geisteswissenschaften haben diese Herausforderungen und Chancen meist noch vor sich.“ Neuroth, et. al., (2007), S. 276.

Ein erster Schritt des Instituts in diese Richtung war der Onlinegang des Informationsteils von Edumeres.net. Edumeres steht für „Educational Media Research“ und soll zum virtuellen Knotenpunkt der Bildungsmedienforschung werden.

Das Informations- und Kommunikationsportal Edumeres.net

Seit 2009 findet der Nutzer unter www.edumeres.net Informationen und Aktuelles aus und für das Fachgebiet der Schulbuch- und Bildungsmedienforschung. In einem ersten Modul folgen nach den obligatorischen Mitteilungen über Konferenzen, Calls for Paper und Stellenausschreibungen Möglichkeiten zur Literaturrecherche in Onlinekatalogen und Datenbanken. Eine besondere Datenbank entsteht dabei unter dem Navigationspunkt „Netzwerk“, hinter dem sich Angaben zu Forschungseinrichtungen, Institutionen und einzelnen Forschern finden lassen. Texte und Grafiken zu Bildungs- und Schulbuchsystemen sowie ein Glossar mit Fachbegriffen runden das Angebot an Informationen ab.

Ein zweites Modul, überschrieben mit „Publikationen“, beherbergt an erster Stelle das Projekt der Schulbuchrezensionen. Hier wurden bereits über hundert aktuelle Schulbücher aus den Fächern Erdkunde, Geschichte und Sozialkunde von Wissenschaftlern, Lehrern sowie Schüler- und Studentengruppen rezensiert. Es schließen sich Analysen, Dossiers und Beiträge zu Themen der Bildungsmedienforschung an. Diese sind mit einer URN versehen und werden im pdf-Format zum Download angeboten. Im Gegensatz zu diesen reinen Onlinepublikationen sind die letzten beiden Veröffentlichungsreihen in diesem Modul bereits in einer Printversion erschienen. Es handelt sich um die Schriftenreihe und die Zeitschrift des Georg-Eckert-Instituts. Beide werden sukzessive und nach Ablauf vertraglich vereinbarter Fristen hier online zugänglich.

Ein Großteil der bisher genannten Informationen wird in einem dritten Modul noch einmal aufgegriffen und thematisch sortiert. Diese Neugruppierung bietet in machen Fällen einen leichteren Zugang zu gesuchten Inhalten.

Die Virtuelle Forschungsumgebung von Edumeres.net

Zu diesen drei genannten Modulen hat sich im Sommer 2010 ein viertes hinzugesellt, das das Portal zu einem Informations- und Kommunikationsportal erweitert. Es steht für die Virtuelle Forschungsumgebung. Gefördert von der Deutschen Forschungsgemeinschaft (DFG) soll sie Arbeitsvorgänge und Forschungsprozesse der Bildungsmedienforschung virtuell abbilden und so die oben angesprochenen Defizite beheben sowie Verbesserungen und Erleichterungen in der wissenschaftlichen Kommunikation bieten.

Aufgeteilt in zwei sich ergänzende Bereiche, dient die Virtuelle Forschungsumgebung zum einen der Kommunikation und Kontaktbildung der Nutzer untereinander, zum anderen stellt sie ein Set an Werkzeugen zur Verfügung, das den Prozess der Wissensgenerierung und anschließenden Publikation unter einer Oberfläche zusammenfasst und ermöglicht.

Community-Funktionen wie etwa ein eigenes Online-Profil, eine persönliche Kontakteliste und ein Nachrichtenmodul dienen dem direkten Austausch mit anderen registrierten Nutzern. Der Mehrwert gegenüber sozialen Netzwerken, die ähnliche Funktionen anbieten, be-

steht in der Verknüpfung mit dem Projekt-Bereich. Hier findet die gemeinsame Ideenfindung, die Konzeptphase, die Textarbeit, die Edition und die Veröffentlichung statt.

Ein Fallbeispiel: Schulbuchanalyse

Ein konkretes Beispiel aus dem Bereich der Schulbuchforschung soll dies näher erläutern.

Ein Wissenschaftler hegt den Gedanken eine Schulbuchanalyse zum Thema „Integrationspolitik in deutschen Sozialkundebüchern der Sekundarstufe“ zu verfassen. Da er diese gerne mit auswärtigen Kollegen erstellen möchte, registriert er sich unter www.edumeres.net und bittet auch die anderen, gleiches zu tun. Im Community-Bereich richtet er sich eine Kontakteliste ein, um so eine schnelle Kontaktaufnahme zu ermöglichen. Der nächste Schritt führt ihn in den Projekt-Bereich. Ein Blick in die Liste der aktuellen Projekte zeigt, dass noch niemand am selben Thema arbeitet. Er legt also ein neues Projekt an und lädt seine Kollegen dazu ein. Als Projektleiter hat er nun Zugriff auf alle Funktionen in seinem Projekt. Doch auch die eingeladenen Projektmitglieder können jetzt ohne Beschränkung am Projekt mitarbeiten. Für die Diskussion steht Ihnen ein Forum zur Verfügung, zur Materialsammlung ein Dokumentenmanager und für die Protokollierung des Arbeitsfortschrittes sowie der Dokumentation gegenüber anderen Nutzern von Edumeres.net ein Weblog-System. Die eigentliche Textarbeit jedoch geschieht mit der Dokumentenfunktion. Hier können Textdokumente in verschiedenen Formaten angelegt und kollaborativ, das heißt gemeinsam und gleichzeitig, bearbeitet werden. Um Zwischenstände sichern zu können, ist es jederzeit möglich Sicherungskopien anzulegen. Ist ein erster Endstand erreicht der von externen Gutachtern beurteilt werden soll, können weitere Mitglieder dem Projekt hinzugefügt werden, wenn gewünscht in der Funktion eines Beobachters, dem nur Lese- und Kommentarrechte zugeteilt sind. Ganz am Ende steht dann die Publikation. Diese wird vom Projektleiter angestoßen und erfolgt über die Redaktion von Edumeres.net. Diese begutachtet und veröffentlicht den Beitrag, versehen mit einer zitierbaren URN, im Publikations-Bereich von Edumeres.net.

Dieser Entstehungsprozess einer wissenschaftlichen Schulbuchanalyse steht exemplarisch für das Verfassen von Texten in Edumeres.net. Da es drei Öffentlichkeitsstufen gibt, kann der Projektleiter die Sichtbarkeit seines Projektes bestimmen: von „öffentlich“ bei Projekten an denen die Mitarbeit möglichst vieler gewünscht wird, über „eingeschränkt“ für eine feststehende Gruppe, hin zu „privaten“ Projekten, die nur für die Projektgruppen einsehbar sind. In jedem dieser drei Fälle bleiben die Dokumente auf einem Server von Edumeres.net. Auch wenn für die Textverarbeitung auf den externen Dienstleister Zoho und dessen Weboberfläche zugegriffen wird, so werden die Dateien dennoch auf einem eigenen Server abgelegt.

Das „Begleitete Projekt“

Um besonders neue Nutzer nicht mit den Funktionen der Virtuellen Forschungsumgebung alleine zu lassen, verfolgt Edumeres.net das Konzept des „Begleiteten Projektes“. Auf Wunsch wird jedem neuen Projekt ein Mitglied der Redaktion zur Seite gestellt. Dieses arbeitet eng mit dem Projektleiter zusammen und hilft bei Fragen und Problemen. Auf diese

Weise ist es möglich auch die Vorbereitung, die Durchführung und die Nachbereitung von Konferenzen oder Tagungen über die Virtuelle Forschungsumgebung laufen zu lassen.

Ausblick

Der auf der Konferenz „Digitale Wissenschaften 2010“ in Köln vorgestellte Stand ist die erste öffentlich zugängliche Version der Virtuellen Forschungsumgebung. Zurzeit werden die Erfahrungen aus den bisherigen Projekten und die Ergebnisse von vor kurzem durchgeführten Nutzer- und Usabilitytests ausgewertet. Die Ergebnisse fließen in die Weiterentwicklung ein, die sich im Laufe der nächsten Wochen Stück für Stück online bemerkbar machen wird. Ziel ist eine einfach zu bedienende Arbeitsumgebung für die Forschung und Wissenschaft, damit in Zukunft noch viele Schulbücher und andere Projekte mit Hilfe der Virtuellen Forschungsumgebung von Edumeres.net den Weg in die digitale Wissenschaft finden.

Typisierung intertextueller Referenzen in RDF

Adrian Pohl

Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz)
Jülicher Str. 6
D-50674 Köln
pohl@hbz-nrw.de

Zusammenfassung

Dieser Beitrag beleuchtet, in welcher Weise die maschinenlesbare Dokumentation und Typisierung intertextueller Referenzen unter Nutzung des Resource Description Framework (RDF) für die Wissenschaft von Nutzen sein können. Der Fokus liegt dabei auf der Typisierung expliziter Referenzen (etwa in Form von Zitationen und Literaturverzeichnissen) in wissenschaftlichen Texten. Es werden bestehende Vokabulare und erste Anwendungsfälle vorgestellt sowie Perspektiven aufgezeigt.

Einleitung

Intertextuelle Bezüge in Form von Zitaten, Verweisen, Zitationen usw. sind ein konstitutiver Bestandteil wissenschaftlicher Praxis in sämtlichen Fachbereichen. In den Literaturwissenschaften erhält das Phänomen der Intertextualität seit vier Jahrzehnten viel Aufmerksamkeit. Dabei kommt auch dem Typus des wissenschaftlichen Texts durchaus Beachtung zu (vgl. Jakobs, 1999, Adamzik, 2007), gleichwohl die Betrachtung sich auf literarische Texte im engeren Sinne konzentriert.

Im Bereich der Bibliotheks- und Informationswissenschaft oder der Erforschung dessen, was am treffendsten mit „Scholarly Communication“ bezeichnet wird, findet der Ausdruck 'Intertextualität' zwar selten Verwendung, allerdings spielt das Phänomen durchaus eine wichtige Rolle: Zum Beispiel stellen Datenbanken, die intertextuelle Bezüge in Form von Referenzangaben wissenschaftlicher Texte sammeln und recherchierbar machen (sogenannte Zitationsdatenbanken wie der *ISI Citation Index* oder *Scopus*), in vielen Fächern ein wichtiges Rechercheangebot für Wissenschaftler dar und sind Grundlage bibliometrischer Verfahren. Allerdings sind die Möglichkeiten von Anwendungen über die bestehenden Daten beschränkt; so findet eben keine Typisierung von Zitationen statt, d. h. die Art der jeweiligen Bezugnahme ist nicht weiter spezifiziert. Da zudem der Großteil der Zitationsdaten weder offen lizenziert noch frei zugänglich ist, ist die Nachnutzung und Anreicherung der gesammelten Daten meist nicht möglich.[1]

In diesem Text geht es letztlich darum zu skizzieren, wie Linked-Data-Praktiken für den Einsatz in der Wissenschaft fruchtbar gemacht werden können – sei es eben zur Bestimmung der Relevanz und Qualität wissenschaftlicher Beiträge, zur Verbesserung von Rechercheidiensten oder aber zur standardisierten maschinenlesbaren Dokumentation literaturwissenschaftlicher Erkenntnisse.

[1] Zur Definition von „offen“ siehe http://www.opendefinition.org. Vgl. auch O'Steen, 2010.

Eingangs findet eine Klärung grundlegender Begrifflichkeiten statt und Möglichkeiten zur Typisierung intertextueller Referenzen in der Wissenschaft werden angesprochen. Nach einer knappen Einführung in RDF werden Vokabulare im Bereich *Semantic Publishing* zur Typisierung von Referenzen wissenschaftlicher Zeitschriftenartikel und Monographien vorgestellt. Abschließend werden erste Anwendungsbeispiele skizziert und Zukunftsperspektiven aufgezeigt.

Intertextuelle Referenz

Intertextualitätstheorien handeln allgemein gesprochen von der Bezugnahme eines Textes auf andere Texte. Der Ausdruck 'Intertextualität' geht zurück auf Julia Kristeva, deren Text *Bachtin, das Wort, der Dialog und der Roman* den Ursprung der Intertextualitätstheorie markiert (Kristeva, 1972). Kristeva nennt darin *jeden* Text eine „Absorption und Transformation eines anderen Textes" (Kristeva, 1972, 348) und vertritt damit einen *weiten Intertextualitätsbegriff*, der eben allgemein jeden Text als eine Rekomposition aus Elementen anderer Texte ansieht. Andreas Böhn schreibt dazu:

„Dieser weite Intertextualitätsbegriff, der insbesondere im Umfeld des Poststrukturalismus und Dekonstruktivismus intensiv rezipiert wurde, richtet sich gegen als ideologisch belastet angesehene Begriffe wie 'Autorschaft' und 'Geschlossenheit' des literarischen Werks zugunsten der Konzeption eines alles umfassenden Textuniversums, in dem die Einzeltexte nur noch Knotenpunkte vielfältiger Bezugslinien darstellen." (Böhn, 2007, 204)

Heinrich Plett bezeichnet Vertreter des weiten Intertextualitätsbegriff als *Progressives* (vgl. Plett, 1991, 3f).

Der *enge Intertextualitätsbegriff* beschränkt sich auf Fälle konkreter Bezugnahme individueller Texte auf ihnen vorgängige Texte, wodurch der Intertextualitätsbegriff für die Textanalyse operationalisierbar gemacht wird (vgl. Böhn, 2007, 204). Vertreter einer solchen engeren Intertextualitätstheorie ist Gérard Genette, der mit seinem Werk *Palimpseste* (Genette, 2004) eine ausgereifte Konzeption von – wie er es nennt[2] – Transtextualität und ihren Spielarten vorlegte, die in der Intertextualitätstheorie großen Einfluss gehabt hat. Vertreter eines engen Intertextualitätsbegriffs werden von Plett als *Traditionalists* bezeichnet (vgl. Plett, 1991, 4).

Auch in diesem Beitrag findet der engere Intertextualitätsbegriff Verwendung wie ihn etwa Genette vertritt. Das heißt, Gegenstand sind konkrete Bezüge zwischen Einzeltexten und nicht die Bezugnahme eines einzelnen Textes zur Gesamtheit des „Textuniversums".

Explizite vs. Nicht-explizite Referenzen

Eine grundlegende Unterscheidung zweier Formen intertextueller Referenz ist jene zwischen expliziter und nicht-expliziter Referenz.

Explizite Referenz zwischen zwei Texten A und B findet dann statt, wenn innerhalb des bezugnehmenden Textes A seine Bezugnahme auf Text B deklariert wird. Dies kann etwa

[2] Gérard Genette gebraucht in *Palimpseste* 'Intertextualität' nicht in der gängigen Weise, stattdessen benutzt er als allgemeinen Begriff *Transtextualität*, wovon die Intertextualität eine von fünf Spielarten ist.

durch die Nennung von identifizierenden Angaben zu Text B (Titel, Autor, Veröffentlichungsjahr, Identifikatoren wie DOI, URN etc.) in Fußnoten und Literaturverzeichnissen, durch die Adressierung von Text B mittels eines Hyperlinks oder durch ähnliche Verfahren geschehen.

Nicht-explizite Referenzen zwischen zwei Texten A und B zeichnen sich hingegen dadurch aus, dass sie im Text nicht deklariert werden und somit durch den Leser „entdeckt" oder gar konstituiert werden müssen. Es ist die nicht-explizite intertextuelle Referenz in Form der Imitation, des Plagiats, der Parodie etc., die in den Literaturwissenschaften – etwa in Gérard Genettes Buch *Palimpseste* – bisher am meisten Aufmerksamkeit erhalten hat.[3]

Abbildung 1 stellt die grundlegende Unterscheidung zwischen expliziter und nicht-expliziter intertextueller Referenz dar.

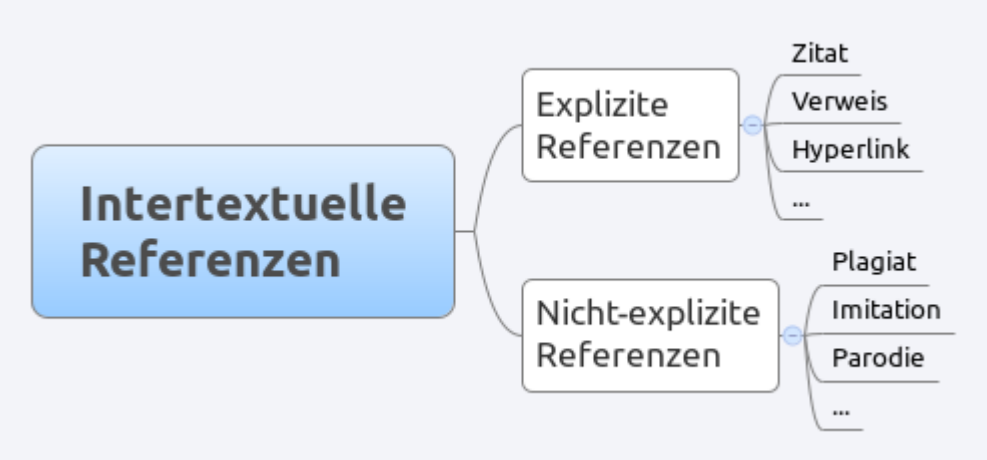


Abbildung 1: Zwei Grundformen intertextueller Referenz

Typisierung intertextueller Referenzen in der Wissenschaft

In den Literaturwissenschaften könnte eine standardisierte, maschinenlesbare Dokumentation und Sammlung aufgefundener nicht-expliziter intertextueller Referenzen interessantes Datenmaterial und neue quantitative Erkenntnisse zutage fördern. Zum Beispiel ließen sich, die von Literaturwissenschaftlern dokumentierten nicht-expliziten intertextuellen Verweise – wie sie etwa Gérard Genette in *Palimpseste* aufführt – in eine standardisierte, maschinenlesbare Form bringen und auf dieser Basis „Intertextualitätsgraphen" aufbauen, die das Verweisungsnetzwerk der Literatur widerspiegeln. Im Folgenden werden diese Möglichkeiten RDF-basierter Typisierung nicht-expliziter intertextueller Referenzen nicht näher betrachtet, vor allem, weil bisher keine konkreten Ansätze in diese Richtung bekannt sind. Zuvorderst geht es in diesem Text darum, Perspektiven einer Auszeichnung der bestehenden expliziten Referenzen in wissenschaftlichen Texten aller Fachrichtungen zu beleuchten.

[3] Die von Genette so bezeichnete 'Architextualität', d. h. die Referenz eines Textes auf eine Textgattung, stellt eine spezifische Form der Intertextualität dar, die nicht in das Explizit-vs.-Nicht-Explizit-Schema hineinpasst. Im Kontext dieses Beitrags ist die Architextualität ohnehin nicht relevant und wird außen vor gelassen.

Ausdrückliche Bezugnahme auf andere Texte ist *die* zentrale Handlung einer jeden Wissenschaftlerin ganz gleich welchen Fachbereichs. In jeder Einführung in das wissenschaftliche Arbeiten nimmt die deklarierte Bezugnahme auf andere Texte – d. h. das Zitieren, der Verweis oder die Erstellung eines Literaturverzeichnisses – eine wichtige Rolle ein. Diese expliziten Referenzen konstituieren ein Netz wissenschaftlicher Texte, Distribution und Häufigkeit der Referenzen geben weitere wichtige Informationen. Nur sind Informationen über Textverknüpfungen derzeit meist nur recht mühselig oder kostspielig zu gewinnen. Die Bezüge durch Zitationen in Fußnoten und Literaturverzeichnissen sind eben in erster Linie für menschliche Leser geschaffen. Allerdings sind sie mehr oder weniger standardisiert, so dass sie sich parsen und – für Maschinen lesbar – verlinken lassen.[4] Wie Verlinkung unter Berücksichtigung der Linked-Data-Best-Practices stattfindet wird im nächsten Abschnitt erläutert.

RDF und Linked Open Data

Linked Data hat in den letzten Jahren zunehmend an Aufmerksamkeit gewonnen und immer mehr Datensammlungen werden unter Berücksichtigung der Linked-Data-Best-Practices[5] veröffentlicht.

Die Essenz von Linked Data ist die Identifikation von Entitäten mittels eindeutiger, öffentlicher Identifikatoren (HTTP-URIs[6]) sowie die Verknüpfung dieser Entitäten mittels typisierter Links. Zum Beispiel lässt sich die Aussage „Platon ist der Autor des *Sophistes*“ in RDF wie folgt ausdrücken:

<http://dbpedia.org/resource/Sophist_(dialogue)>
<http://purl.org/dc/elements/1.1/creator> <http://dbpedia.org/resource/Plato> .

Solcherlei Aussagen in RDF bilden die Grundlage des Linked-Data-Netzes. Man nennt sie auch RDF-Tripel, weil sie aus drei Teilen bestehen, die als ‚Subjekt', ‚Prädikat' und ‚Objekt' bezeichnet werden. Im vorliegenden Beispiel nimmt der Subjekt-URI auf das Werk *Sophistes* Bezug, der Prädikat-URI auf die Relation *erstellt von* und der Objekt-URI auf die Person *Platon*.

[4] In Frankreich etwa entwickelt derzeit ein von Google finanziell unterstütztes Projekt im Kontext der Digital-Humanities-Forschung „Robust and Language Independent Machine Learning Approaches for Automatic Annotation of Bibliographical References in DH Books, Articles and Blogs“, siehe Dacos, 2011.

[5] Die grundlegenden vier „Linked Data Principles“ finden sich in Tim Berners-Lees „Linked Data Design Issues“, Berners-Lee, 2006 ff. Einen Überblick über bisher gemäß diesen Prinzipien publizierte Daten gibt die Linked-Data-Cloud, siehe http://richard.cyganiak.de/2007/10/lod/.

[6] HTTP steht für „Hypertext Transfer Protocol“ und URI für „Uniform Resource Identifier“. HTTP-URIs haben dieselbe Form wie die vom WWW allseits bekannten URLs (Uniform Resource Locator).

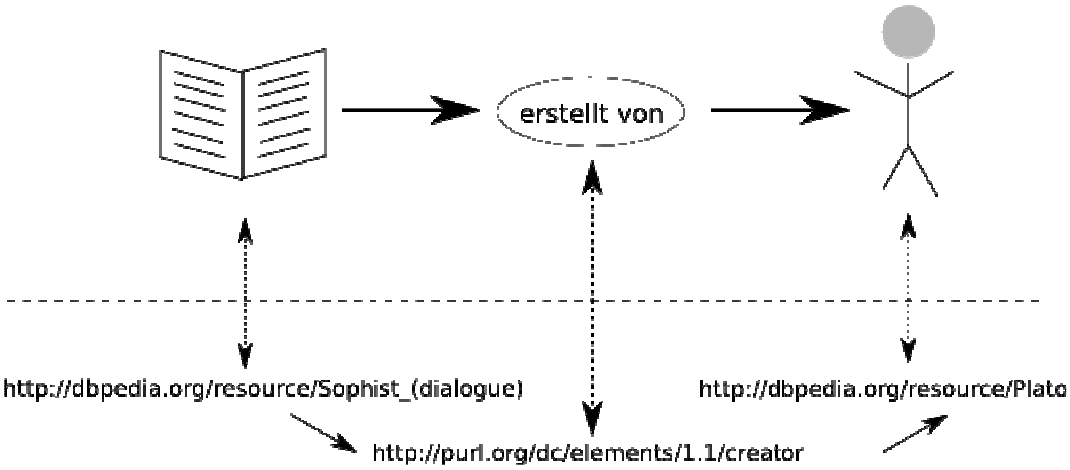


Abbildung 2: Ein RDF-Tripel[7]

Jedes RDF-Tripel hat diese Struktur: Subjekt-Prädikat-Objekt. Der Witz eines RDF-Tripels im Vergleich zum klassischen Hyperlink des World Wide Web ist, dass Subjekt und Prädikat nicht einfach unkommentiert miteinander verbunden (verlinkt) sind, sondern dass diese Verbindung typisiert ist, indem sie durch einen HTTP-URI identifiziert wird. Wenn dezentral Aussagen über dieselbe Entität gemacht werden bzw. Entitäten mittels desselben URI – desselben Prädikats – verknüpft werden, entstehen vernetzte Informationen: Linked Data.[8] Dadurch, dass HTTP-URIs öffentliche Identifikatoren sind, können sie von jedermann im webweiten Kontext genutzt werden.

Wenn in diesem Beitrag von typisierten intertextuellen Referenzen in RDF gesprochen wird, dann ist schlicht folgendes damit gemeint: Wir haben zwei Texte, die durch URIs identifiziert werden und durch ein RDF-Prädikat – also einen weiteren URI – miteinander in Relation gesetzt werden.

Bestehende Vokabulare

In den Lebenswissenschaften – dem Bereich der Wissenschaft, in dem der jährliche Publikationszuwachs am größten[9] und damit der Wunsch nach schnellerem Überblick, besseren Orientierungs- und Recherchemöglichkeiten am stärksten ist – entwickeln sich seit einigen Jahren erste Ansätze für eine zukünftige Praxis der maschinenlesbaren Annotierung von Referenzen. Hier sollen zunächst zwei Vokabulare (auch „Ontologien“ genannt) vorgestellt werden, die eine Menge von RDF-Prädikaten zur Typisierung intertextueller Referenzen, also zur Klassifikation von Relationen zwischen Texten zur Verfügung stellen.[10]

[7] Abbildung übernommen aus Pohl / Ostrowski, 2010.
[8] Für eine detaillierte Einführung in Linked Data siehe den Preprint Pohl, 2011.
[9] MEDLINE/PubMed verzeichnet einen jährlichen Zuwachs von ca. 700 000 Nachweisen in den letzten vier Jahren (2007-2010), siehe http://www.nlm.nih.gov/bsd/index_stats_comp.html.
[10] Teile dieses Abschnitts sind übernommen aus den Blogbeitrag Pohl, 2010.

Scientific Discourse Relationships Ontology

Bereits 2008 entwickelten Wissenschaftler aus der biomedizinischen Alzheimerforschung die *Scientific Discourse Relationships Ontology*[11] im Kontext des SWAN-Projekts *Semantic Web Applications in Neuromedicine*, „a project to develop knowledge bases for the neurodegenerative disease research communities, using the energy and self-organization of that community enabled by Semantic Web technology".[12] Diese Ontologie umfasst die in Abbildung 3 gezeigten RDF-Prädikate zur Bestimmung von intertextuellen Referenzen.

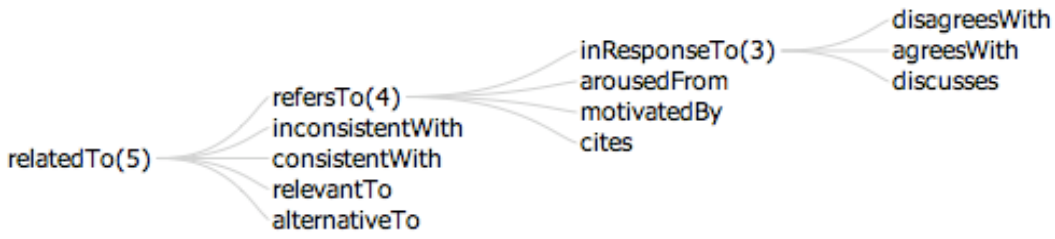


Abbildung 3: Aufbau der Scientific Discourse Relationships Ontology[13]

Diese Ontologie ist sehr übersichtlich, weil sie sich auf allgemeine Relationen beschränkt und eine Spezifizierung unterlässt. Auffällig ist, dass der Großteil des Vokabulars sich auf die Typisierung bereits expliziter Referenzen bezieht während vier Prädikate (inconsistentWith, consistentWith, relevantTo, alternativeTo) auch nicht-explizite Referenzen spezifizieren können.

Citation Typing Ontology (CiTO)

Die *Citation Typing Ontology* (CiTO) wurde ursprünglich vom Oxforder Biologen David Shotton entwickelt und diente zunächst neben der Typisierung von Referenzen auch der Beschreibung bibliographischer Ressourcen.[14] Mit der Publikation der SPAR-Ontologien (Semantic Publishing and Referencing) durch David Shotton und Silvio Peroni[15] wurde die Beschreibung bibliographischer Ressourcen in die *FRBR-Aligned Bibliographic Ontology* (FaBiO) ausgelagert, so dass CiTO sich nunmehr auf die Typisierung intertextueller Referenzen beschränkt.[16] Auch ist die CiTO mittlerweile mit der *Scientific Discourse Ontology* in Einklang gebracht worden.

In der Selbstbeschreibung der CiTO heißt es:

„The purpose of the CiTO Ontology is to enable characterization of the nature or type of citations, both factually and rhetorically.

[11] http://swan.mindinformatics.org/spec/1.2/discourserelationships.html.
[12] http://swan.mindinformatics.org/.
[13] Abbildung übernommen aus http://swan.mindinformatics.org/spec/1.2/discourserelationships.html.
[14] Die ursprüngliche und mittlerweile überholte Fassung von CiTO ist hier einzusehen: http://imageweb.zoo.ox.ac.uk/pub/2009/citobase/cito-20090311/cito-content/owldoc/. Vgl. auch. Shotton 2009.
[15] http://purl.org/spar/ Siehe auch Shotton / Peroni, 2010.
[16] Die aktuelle Fassung der CiTO: http://purl.org/spar/cito/.

The citations characterized may be either direct and explicit (as in the reference list of a journal article), indirect (e.g. a citation to a more recent paper by the same research group on the same topic), or implicit (e.g. as in artistic quotations or parodies, or in cases of plagiarism).“[17]

Die CiTO ist also mit dieser neuen Version nicht mehr nur auf die Typisierung expliziter Referenzen ausgelegt, sondern soll auch der Ausweisung nicht-expliziter Referenzen dienen, wenngleich explizite Referenzen den Kern der CiTO ausmachen.

Interessant ist im oben genannten Zitat auch die Unterscheidung zwischen direkter und indirekter Referenz. David Shotton beschreibt in einem Blogpost ein solches Beispiel indirekter Referenz: Nimmt ein Text A explizit auf einen Text B Bezug, der wiederum eine überarbeitete Fassung von Text C ist, so referenziert Text A indirekt auch Text C (siehe Abbildung 4).[18] Es handelt sich mit anderen Worten um eine vermittelte Referenz, die zwar nicht-explizit ist aber gleichwohl unter Umständen aus bestehenden RDF-Aussagen inferiert werden kann.

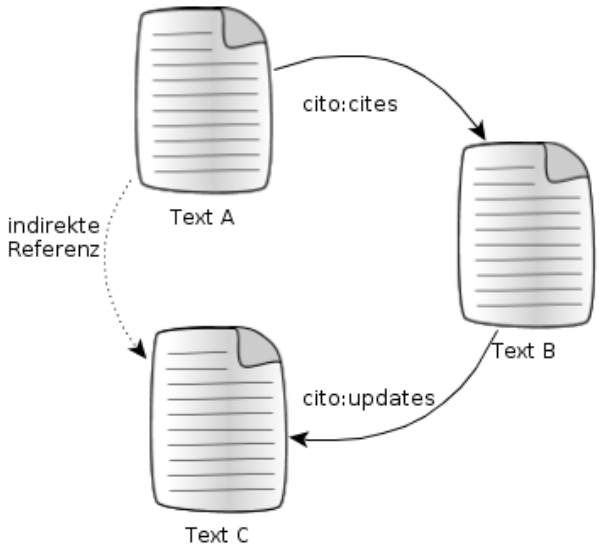


Abbildung 4: Indirekte Referenz

Die Typisierungsmöglichkeiten der CiTO seien hier kurz erläutert: Das oberste Prädikat ist cito:cites, welches derzeit 30 Unterprädikate hat. Zudem ist zu diesen Prädikaten jeweils eine Umkehrung definiert, d. h. es gibt auch das Prädikat cito:isCitedBy mit ebenfalls 30 Unterprädikaten. In Tabelle 1 werden die Unterprädikate von cito:cites dargestellt.

[17] http:/purl.org/spar/cito.
[18] Vgl. Shotton, 2010.

Spezifizierung expliziter Referenz	Annotation nicht-expliziter Referenz	Unbestimmt
agrees with	parodies	confirms
cites as authority	plagiarizes	refutes
cites as data source	ridicules	
cites as evidence		
cites as metadata document		
cites as related		
cites as source document		
cites for information		
corrects		
credits		
critiques		
disagrees with		
discusses		
disputes		
extends		
includes excerpt from		
includes quotation from		
obtains background from		
obtains support from		
qualifies		
refutes		
reviews		
supports		
updates		
uses data from		
uses method in		

Tabelle 1: Die CiTO-Prädikate im Überblick

Ein Blick auf die Tabelle macht deutlich, dass sich die CiTO auf die Klassifizierung expliziter Referenzen fokussiert. Lediglich drei Prädikate beziehen sich auf nicht-explizite Referenzen und zwei Prädikate lassen sich unter gegebenen Umständen auf beide Arten anwenden.

Erste Anwendungsbeispiele

Die *Scientific Discourse Relationships Ontology* ist neben anderen Ontologien des SWAN-Projekts Grundlage für die *AlzSWAN Knowledge Base*[19] „a community-driven knowledge-base of Alzheimer disease, in which researchers can annotate scientific claims, data, and information, putting these into the context of testable hypotheses and treatment discovery".[20] Diese Semantic-Web-basierte Community-Plattform für die Alzheimerforschung befindet sich derzeit noch im Beta-Stadium.

Die *CiTO* hat bereits erste Anwendungsfälle gefunden. Zum Beispiel ermöglicht der soziale Referenzverwaltungsdienst *CiteULike*[21] die Verknüpfung gespeicherter Referenzen durch CiTO-Prädikate (vgl. Schneider, 2010). Damit werden die Mittel für einen sozial generierten Zitationsindex zur Verfügung gestellt, wobei eben auch – im Gegensatz zu anderen Zitationsdatenbanken – die Typisierung von Referenzen ermöglicht wird.

Ein weiteres Anwendungsbeispiel hat jüngst Martin Fenner geliefert, der ein Wordpress-Plugin entwickelt hat, das es ermöglicht Hyperlinks mit CiTO-Prädikaten zu typisieren (siehe Fenner, 2011). Dieses Plugin bietet Wissenschaftlern die Möglichkeit mit innovativen, zukunftsträchtigen Formen des wissenschaftlichen Schreibens und Publizierens „beyond the pdf"[22] zu experimentieren.

Perspektiven

Wie sich zeigt, haben bereits einige Entwicklungen im Bereich der Typisierung intertextueller Referenzen in RDF stattgefunden, allerdings hat sich noch keine breite Anwendung ergeben. Dies liegt sicher zu einem großen Teil daran, dass die Aufnahme von Linked Data und Semantic-Web-Technologien in der Wissenschaft wie auch in anderen Bereichen noch keine kritische Masse erreicht hat, vor allem auch, weil es bisher an entsprechenden einfach zu bedienenden Werkzeugen fehlt. Welche zukünftigen Entwicklungen sind wünschenswert? Welche Schritte sollten als nächstes gegangen werden?

Wichtige Voraussetzung für die Typisierung intertextueller Referenzen sind HTTP-URIs für die in Beziehung gesetzten Texte. Deswegen ist es wichtig, solche URIs zu prägen, etwa indem bibliographische Daten aus Bibliothekskatalogen und Zeitschriftendatenbanken in RDF zur Verfügung gestellt werden.[23]

[19] http://hypothesis.alzforum.org.

[20] http://hypothesis.alzforum.org/swan/do!getFAQ.action.

[21] http://www.citeulike.org/.

[22] „Beyond the PDF" ist der Titel eines – für an digitaler Wissenschaft Interessierte sehr relevanten – Workshops, der im Januar 2011 in San Diego, Kalifornien stattfand, siehe https://sites.google.com/site/beyondthepdf/.

[23] Erste Schritte in diese Richtung machen zum Beispiel die Open Library (http://openlibrary.org/), das JISC-Projekt OpenBib (http://openbiblio.net/p/jiscopenbib/), in dessen Rahmen Daten der British Library als Linked Open Data veröffentlicht wurden, sowie in Deutschland das hbz mit dem Dienst lobid-resources

Für eine „semantische“ Aufwertung bestehender Literaturverweise sind zuverlässige Anwendungen zum Parsen von Zitationen sowie die Nutzung gemeinsamer Vokabulare zur RDF-Repräsentation[24] von Zitationen wichtig.

Um die Typisierung intertextueller Referenzen bereits bei der Texterstellung zu ermöglichen sind neue Tools für die wissenschaftliche Textproduktion nötig, wie z. B. Texteditoren zur leichten Erstellung von mit RDF angereichertem HTML, LaTeX-Erweiterungen oder auch entsprechende Plugins für Microsoft Word und Open Office.

Eine andere Entwicklungslinie könnte sich auf die Typisierung nicht-expliziter intertextueller Bezugnahmen konzentrieren. Vokabulare zur Dokumentation und Klassifikation nicht-expliziter Referenzen könnten Literaturwissenschaftler und andere für die Dokumentation und Auswertung ihrer Erkenntnisse über intertextuelle Bezugnahmen nutzen.

Referenzen

Adamzik, Kirsten (2007): Textsortenvernetzung im akademischen Bereich. Preprint unter http://www.unige.ch/lettres/alman/adamzik/alman/adamzik%20vernetzung%20preprint.pdf. Erscheint gedruckt in: Baumann, Klaus-Dieter / Kalverkämper, Hartwig (Hg.): Fachtextsorten-in-Vernetzung, Forum für Fachsprachenforschung, Berlin: Frank & Timme.

Berners-Lee, Tim (2006 ff.): Linked Data Design Issues. Zugänglich unter http://www.w3.org/DesignIssues/LinkedData.html.

Böhn, Andreas (2007): Intertextualitätsanalyse. In: Handbuch Literaturwissenschaft, Band 2: Methoden und Theorien, Stuttgart: Metzler, 204-216.

Dacos, Marin (2011): Le Cléo lauréat du Grant Google pour les Digital Humanities, http://leo.hypotheses.org/5942.

Fenner, Martin (2011): How to use Citation Typing Ontology (CiTO) in your blog posts, http://blogs.plos.org/mfenner/2011/02/14/how-to-use-citation-typing-ontology-cito-in-your-blog-posts/.

Jakobs, Eva-Maria (1999): Textvernetzung in den Wissenschaften, Tübingen: Niemeyer.

Kristeva, Julia (1972): Bachtin, das Wort, der Dialog und der Roman. In: Ihwe, Jens (Hg.): Literaturwissenschaft und Linguistik. Ergebnisse und Perspektiven. Bd. 3: Zur linguistischen Basis der Literaturwissenschaft II. Frankfurt/M.: Suhrkamp, 345-375.

O'Steen, Ben (2010): JISC Open Citations: Wider Benefits to Sector & Achievements for Host Institution, http://opencitations.wordpress.com/2010/07/15/jisc-open-citations-wider-benefits-to-sector-achievements-for-host-institution/.

(http://lobid.org/de/resource.html) und die Bibliothek der Universität Mannheim (http://data.bib.uni-mannheim.de/).

[24] Hier bietet sich etwa die Bibliographic Reference Ontology (BiRO) an, siehe http:/purl.org/spar/biro.

Plett, Heinrich (1991): Intertextualities. In: Ders. (Hg.): Intertextuality. Berlin/New York: de Gruyter, 3-29.

Pohl, Adrian (2010): Partizipativer Katalog, Intertextualität und Linked Data, einsehbar unter http://www.uebertext.org/2010/02/partizipativer-katalog-intertextualitat.html.

Pohl, Adrian (2011): Linked Data und die Bibliothekswelt. Preprint, erscheint im Konferenzband zur ODOK 2010, http://hdl.handle.net/10760/15324.

Pohl, Adrian / Ostrowski, Felix (2010): ‚Linked Data' – und warum wir uns im hbz-Verbund damit beschäftigen, B.I.T. Online 13 (3) , 259-268. Preprint einsehbar unter http://hdl.handle.net/10760/14836.

Schneider, Jodi (2010): CiTO in the wild. Blogbeitrag einsehbar unter http://jodischneider.com/blog/2010/10/18/cito-in-the-wild/.

Shotton, David (2009): CiTO, the Citation Typing Ontology, and its use for annotation of reference lists and visualization of citation networks. Einsehbar unter http://imageweb.zoo.ox.ac.uk/pub/2008/publications/Shotton_ISMB_BioOntology_CiTO_final_postprint.pdf.

Shotton, David / Peroni, Silvio (2010): Introducing the Semantic Publishing and Referencing (SPAR) Ontologies. Einsehbar unter http://opencitations.wordpress.com/2010/10/14/introducing-the-semantic-publishing-and-referencing-spar-ontologies/.

Twitteranalyse zwischen Selbstreflexion und Forschung

Matthias Rohs

2headz
Seestr. 13
CH-8800 Thalwil
matthias.rohs@2headz.ch

Thomas Bernhardt

Universität Bremen
AB Medienpädagogik (FB12)
Bibliothekstr. 1
D-28359 Bremen
th.bernhardt@uni-bremen.de

Zusammenfassung

Der Microblogging-Dienst Twitter ist aus der öffentlichen Kommunikation nicht mehr wegzudenken. Die täglich 80 Mio. Tweets (Mitteilungen über Twitter) bieten dabei auch für die Forschung ein enormes Potential. In diesem Beitrag wird auf die Möglichkeiten des Einsatzes von Twitter-Analysetools eingegangen. Dazu wird eine Kategorisierung dieser Tools vorgenommen und Einsatzformen, aber auch Grenzen in der Forschung diskutiert.

Einführung

Mit dem Aufkommen des so genannten Web 2.0 wurden und werden damit verbundene Anwendungen auch in Wissenschaft und Forschung diskutiert (Nentwich, 2009). Wissenschaftsblogs und Social Networks haben in diesem Zusammenhang viel Aufmerksamkeit erfahren. Ob sie sich in der Wissenschaftskommunikation etablieren werden, ist indes noch fraglich. In ähnlicher Weise werden auch die Möglichkeiten für die Online-Forschung ausgelotet (Zerfaß, Welker & Schmidt, 2008). Noch befindet sich dieser Bereich aber weitestgehend in einem Experimentierstadium.

Eine der Anwendungen des Web 2.0 ist das Microblogging. Bekannt ist hier vor allem die Anwendung Twitter (siehe u. a. Herwig et al. 2009, S. 1ff), die binnen kürzester Zeit zu einer der am häufigsten genutzten Anwendungen des Web 2.0 avanciert. Einer aktuellen Statistik ist zu entnehmen, dass es über 105 Mio. Benutzerprofile gibt (Twitter-Welt, 2010) und täglich 90 Mio. Tweets verschickt werden – Tendenz steigend (flickr, 2010).

Die große Anzahl von Daten macht die Twitter-Kommunikation für viele Forschungsbereiche wie Markt-, Kommunikations- und Sozialforschung sehr interessant. Von entscheidender Bedeutung für die Forschung ist jedoch, dass die Twitter-Kommunikation per default öffentlich ist und Twitter eine Schnittstelle zur Datenauswertung bietet. Dies hat dazu geführt, dass es im Internet eine Vielzahl von Statistik-Anwendungen gibt, die eine Analyse

der Twitter-Nutzung ermöglichen. In diesem Beitrag wird eine Kategorisierung und Analyse dieser Anwendungen vorgenommen und deren Potentiale für Wissenschaft und Forschung kritisch hinterfragt.

Analyse-Tools für Twitter

Nach vorsichtigen Schätzungen dürfte es einige hundert Web-Anwendungen geben, die eine zumeist kostenlose Analyse der Twitter-Daten ermöglicht. Betreiber dieser Web-Services reichen von Privatpersonen, die den Dienst auf eigenen Servern betreiben, über Start-Ups bis hin zu größeren Unternehmen im Bereich Öffentlichkeitsarbeit oder Onlinemarketing. Während die Motive für die Entwicklung solcher Anwendungen im Marketingbereich sehr naheliegend sind, muss bei Privatpersonen in erster Linie von einer intrinsischen Motivation ausgegangen werden. Diese Vermutung wird u. a. dadurch unterstützt, dass bei den meisten Diensten kein konkretes Geschäftsmodell erkennbar ist. Vereinzelt wird das Modell des „Featured Users“ verwendet, wonach man gegen einen bestimmten Betrag bei diesen Diensten als Empfehlung für anderen Nutzer angezeigt wird. Ein wissenschaftlicher Hintergrund von Anwendungen ist nur in Einzelfällen auszumachen.

API-Schnittstelle

Grundlage für alle Analysetools bildet die API-Schnittstelle (application programming interface) von Twitter. In der Webentwicklung steht API üblicherweise für eine Gruppe aus HTTP-Anfragen mit den dazugehörigen Definitionen für die meist in Extensible Markup Language (XML) oder JavaScript Object Notation (JSON) formatierten Antworten. Im Web 2.0 ermöglichen Web-APIs durch die Kombination mehrerer oder den Zugriff auf einzelne Web-Services die Programmierung neuer Anwendungen, sogenannter Mashups (Wikipedia, 2010).

Twitter verfügt zurzeit über drei verschiedene Web-APIs:

Die Twitter-REST (Representational State Transfer)-API bietet Entwicklern Zugriff auf zentrale Twitter-Daten, wie Update der Timeline, Statusdaten und Nutzerinformationen.

Die Such-API bietet die Möglichkeit zur Interaktion mit der Twitter-Suche (http://search.twitter.com/) sowie den Trenddaten.

Als letztes bietet die Streaming-API einen umfangreichen und annähernden Echtzeit-Zugriff auf öffentliche und geschützte Twitter-Daten in zusammengefasster und gefilterter Form an (Twitter, 2010a).[1]

Klassifizierung der Tools nach Verwendungszweck

Die offenen Schnittstellen von Twitter ermöglichen es Entwicklern, neue Anwendungen auf Basis des Microblogging-Dienstes zu entwickeln[2]. Bisher gibt es jedoch keine klare Klassifizierung der Anwendungen. Vor diesem Hintergrund wurden zunächst Web-Anwendungen zusammengetragen, die sich im weitesten Sinne zur Analyse von Twitter eignen. Anwen-

[1] Einen umfangreichen Überblick über die Interaktionsmöglichkeiten bietet hierbei das Twitter-API-Wiki http://apiwiki.twitter.com/.

[2] Auf www.onefourty.com bekommt man einen sehr guten Eindruck über die Vielzahl an unterschiedlichen Werkzeugen.

dungen, die sich zwar die Twitter-API zu Nutze machen, aber lediglich als Client zum Einsatz kommen (z. B. Tweetdeck oder Tweetie) wurden hierbei nicht berücksichtigt. Diese Sammlung wurde so lange erweitert, bis keine neuen Verwendungszwecke mehr erkennbar waren. Eine Analyse der dann vorliegenden Tools ergab fünf Kategorien:

Report Friends & Fans – Netzwerkanalyse

Anwendungen dieser Kategorie bieten eine quantitative Auswertung aller Kontakte eines Accounts zu einem Zeitpunkt oder im Verlauf an. Beispiel hierfür ist http://friendorfollow.com/, bei dem man nach Eingabe (s)eines Twitter-Accounts auf einem Blick sieht, (1) wem man folgt, wer einem aber nicht zurück folgt; (2) wer einem folgt, man selbst ihm aber nicht zurück folgt und (3) wem man folgt und die einem auch folgen.

Monitoring Twitter-Account

Anwendungen dieser Kategorie ermöglichen eine sehr ausführliche Analyse eines Twitter-Accounts. So kann die Intensität der Twitter-Nutzung analysiert werden, zu welchen Zeitpunkten getwittert wurde oder wer die Adressaten der Nutzung waren. Neben einer reinen quantitativen Analyse bieten manche Anwendungen eine weitergehende Berechnung bzw. qualitative Bewertung an. So liefern manche Dienste die Bewertung der eigenen Twitter-Aktivität, wie z. B. der „klout score“, der eine Aussage über den Onlineeinfluss treffen soll (siehe http://klout.com/).

Tracking Conversations

In diese Kategorie lassen sich Anwendungen zuordnen, die eine qualitative oder quantitative Darstellung der Konversation zwischen zwei oder mehreren Accounts ermöglichen. Ein Beispiel hierfür ist der Service http://www.bettween.com.

Monitoring Hashtags / Tracking Topics

Tools in dieser Kategorie dienen der Analyse von bestimmten Themen, die bei Twitter über Hashtags, also durch die Verwendung einer Raute vor einem zentralen Schlagwort oder einen bestimmten Kürzel (z. B. #btw für Bundestagswahl), gekennzeichnet werden. Auf diese Weise lassen sich auch Trends identifizieren, wie beim Service „What the trend?“.

Search-Engine

Die Dienste dieser Kategorie dienen vornehmlich dem Durchsuchen und Verfolgen von bestimmten Themen. Als Beispiel sei hier der Dienst http://www.topsy.com angeführt.

Der Funktionsumfang bei allen Anwendungen ist sehr heterogen, wo bei Diensten wie Xefer (http://www.xefer.com/twitter/) nur eine bestimmte Analyse angeboten wird (an welchem Wochentag zu welcher Uhrzeit getwittert wird), gibt es Dienste wie TwitterAnalyzer (http://www.twitteranalyzer.com/), die umfangreiche Analyseverfahren anbieten, die sich eigentlich zu unterschiedlichen Kategorien zuordnen ließen.

Twitter-Analysetools als Forschungsinstrumente

Forschungsbereiche

Es ist nicht verwunderlich, dass vor allem im Bereich der Markt- und Kommunikationsforschung die Analyse von Twitter breit angewandt und diskutiert wird (Schenk, Taddicken & Welker, 2008). Die Möglichkeit Kundenmeinungen, schnell zu finden und zu analysieren, ist für viele Unternehmen und Marketingabteilungen von großer Wichtigkeit. Auch für die Kommunikationsforschung liegt es auf der Hand, dass eine so umfassende Dokumentation von Kommunikationsprozessen einen großen Schatz für die Forschung darstellt.

Aber auch in anderen Bereichen, wie z. B. der Lehr-Lernforschung, wird die Nutzung von Twitter analysiert (z. B. Ebner et al., 2010). Allgemein ist wohl davon auszugehen, dass die Analyse von Twitter überall dort von Interesse werden könnte, wo Twitter zur Kommunikation eingesetzt wird. Dies ist Prinzipiell in einer Vielzahl von Kontexten möglich, die an dieser Stelle noch nicht abzusehen sind und mit der weiteren Entwicklung und Verbreitung der Anwendung zusammenhängt.

Möglichkeiten und Grenzen

Die Untersuchung von Twitter-Kommunikation ist qualitativ wie quantitativ möglich. In diesem Artikel liegt der Fokus jedoch auf den Anwendungen, bei denen eine quantitative Auswertung im Mittelpunkt steht. Eine umfassende Darstellung ist an dieser Stelle nicht möglich, dennoch sollen einige Beispiele die Bandbreite der Anwendungsmöglichkeiten illustrieren.

Eine Möglichkeit der Anwendung besteht darin, bei Forschungsvorhaben zu Twitter die Accounts der Teilnehmenden zu analysieren und so umfassende Daten zu den Nutzungsgewohnheiten zu erhalten, die dann durch eine Online-Befragung ergänzt werden können. Ein Beispiel dafür ist die Ultimate Twitter Study (http://www.ultimatetwitterstudy.com/).

Eine weitere Möglichkeit besteht in soziometrischen Analysen, wie sie mit einer Reihe von Tools aus der Kategorie Tracking Conversation erstellt werden können. Dadurch können Beziehungen von Mitgliedern einer Gruppe ermittelt werden, so dass Aussagen über Strukturen oder der Stellung einzelner Teilnehmer/Nutzer möglich werden. Dazu werden Soziogramme erstellt, die die Beziehungen zwischen unterschiedlichen Twitternutzern graphisch darstellen und Strukturen leicht ablesbar machen.

Eine erste Sichtung von Publikationen zur Twitterforschung zeigt jedoch, dass kaum auf bestehende Anwendungen, wie den hier beschriebenen, zurückgegriffen wird. In den meisten Fällen wird nur darauf verwiesen, dass die API als Grundlage für die Auswertung diente. Konkretere Ausführungen werden jedoch nicht gemacht.

Gleichzeitig zeigen sich aber auch Grenzen, die sich in verschiedene Aspekte gliedern lassen:

Inhalte: Die Analyse, wie hier vorgestellt, beschränkt sich auf Textmitteilungen. Bilder, Foliensätze, Sprachnachrichten und Videos, die ebenfalls über Twitter kommuniziert werden können, sind hingegen nicht der Analyse mit den hier vorgestellten Tools zugänglich.

Flüchtigkeit: Twitter ist eine Anwendung, die ihre Attraktivität vor allem durch die Aktualität der Meldungen besitzt. So verwundert es auch nicht, dass Twitterbeiträge nur etwa zehn Tage über die Suche auffindbar sind. „Zwar werden Tweets auch von Suchmaschinen wie Google und Bing indiziert, über die Twittersuche selbst sind sie jedoch nicht mehr verfügbar." (vgl. Herwig et al., 2009, S. 3). Auch werden nur etwa 3.200 Updates eines Profils angezeigt (ebd., S. 8). Damit zeigen sich deutliche Grenzen für retrospektiv ausgerichtete Forschung.

Verfügbarkeit: Anonyme Dienste können maximal 150 und über OAuth autorisierte Services maximal 350 Anfragen pro Stunde an die Server von Twitter stellen (Twitter, 2010b). Darüber hinaus kann Twitter ohne große Ankündigung Spezifikationen in ihrer API ändern und damit einen Zugriff auf die Daten beschränken oder ganz und gar verhindern (wie z. B. durch die Einführung des OAuth-Verfahrens 2010).

Darüber hinaus werden allgemeine Kritikpunkte wie die Qualität der Daten oder die Repräsentativität und Verallgemeinerbarkeit der Ergebnisse genannt, die aus methodischer Perspektive für das spezifische Forschungsanliegen kritisch hinterfragt werden müssen.

Fazit

Nach dieser ersten Annäherung an die Möglichkeiten und Grenzen der hier vorgestellten Dienste lässt sich sagen, dass sie erstaunliche Möglichkeiten bieten, für die seriöse Onlineforschung aber allenfalls eine Ergänzung darstellen. Wesentlich scheint jedoch zu sein, auf die damit verbundenen forschungsethischen Fragen hinzuweisen. Im Gegensatz zu anderen Web-2.0-Diensten ist gerade durch die hier angesprochenen Tools die Verfügbarkeit und Aufbereitung von Daten so einfach geworden, dass aufwändige Analysen ohne forschungsmethodische Kenntnisse möglich sind. Eine „Jedermannsforschung" ohne fundierte wissenschaftsmethodische Kenntnisse birgt die Gefahr, bestehende Tendenzen der Verflachung wissenschaftlicher Arbeiten zu verstärken. Gleichzeitig lädt Twitter durch die per default offene Kommunikation dazu ein, ungefragt Nutzerdaten zu erheben. Folgt man dem Grundsatz „Internet researchers should make sure that the individuals they study are aware of their role in the research process" (Barnes 2004, S. 220), wäre es zum einen notwendig, Nutzer der hier vorgestellten Tools auf diese Grundsätze hinzuweisen, als auch Twitter-Nutzer stärker über die Möglichkeiten der existierenden Anwendungen als auch die sich daraus ergebende Konsequenzen aufzuklären.

Literatur

Barnes, S. B. (2004). Issues of Attribution and Identification in Online Social Research, In: M. D. Johns, S.-L. Chen & G. J. Hall (Hrsg.), *Online Social Research: Methods, Issues & Ethics*, New York: Peter Lang Verlag, 203-222.

Ebner, M., Lienhardt, C., Rohs, M. & Meyer, I. (2010): Microblogs in Higher Education – A chance to facilitate informal and process-oriented learning? *Computers & Education*, 55 (1), 92-100.

Flickr (2010). Tweets per day von @Twitter. Online: http://www.flickr.com/photos/twitteroffice/4990581534/ [29. September 2010]

Herwig, J., Kittenberger, A., Nentwich, M. & Schmirmund, J. (2009). *Microblogging für die Wissenschaft. Das Beispiel Twitter*. Steckbrief 4 im Rahmen des Projekts „Interactive Science“, Institut für Technikfolgen-Abschätzung der Österreichischen Akademie der Wissenschaften. Wien, Online: http://www.webcitation.org/5mYU63UlG

Nentwich, M. (2009). Cyberscience 2.0 oder 1.2? Das Web 2.0 und die Wissenschaft. *IAT manu script* II/2009, Institut für Technikfolgen-Abschätzung, Wien. Online: http://www.webcitation.org/5maAcdOi0

Twitter (2010a). API Overview. Online: http://dev.twitter.com/pages/api_overview [29. September 2010].

Twitter (2010b). Rate Limiting. Online: http://dev.twitter.com/pages/rate-limiting [29. September 2010].

Twitter-Welt (2010). Twitter Statistik. Online: http://www.twitter-welt.ch/2010/06/twitter-statistik/ [30. September 2010]

Schenk, M., Taddicken, M. & Welker, M. (2008). Web 2.0 als Chance für die Markt- und Sozialforschung? In A. Zerfaß, M. Welker & J. Schmidt (Hrsg.), a.a.O., 243-266.

Wikipedia (2010). Stichwort: „Web Service“, Version vom 29. September 2010, 20:51 Uhr. Online: http://en.wikipedia.org/w/index.php?title=Web_ service&oldid=387793758

Zerfaß, A., Welker, M. & Schmidt, J. (Hrsg.) (2008). *Kommunikation, Partizipation und Wirkungen im Social Web. Band 1: Grundlagen und Methoden: Von der Gesellschaft zum Individuum. Neue Schriften zur Online-Forschung 2*, Köln: Herbert von Halem Verlag.

Social Software in Forschung und Lehre: Drei Studien zum Einsatz von Web 2.0 Diensten

Katrin Weller, Ramona Dornstädter, Raimonds Freimanis, Raphael N. Klein & Maredith Perez

Heinrich-Heine-Universität Düsseldorf
Abteilung für Informationswissenschaft
Universitätsstraße 1
D-40225 Düsseldorf
weller@uni-duesseldorf.de

Zusammenfassung

Dieser Beitrag fasst Ergebnisse aus drei unabhängigen Studien zusammen, in denen verschiedene Nutzergruppen (Studenten und Hochschulmitarbeiter) zu ihrem Umgang mit Social Software in universitären Kontexten befragt wurden.

Einleitung

Unter Social Software bzw. Web 2.0 Diensten verstehen wir Internetdienste, welche einer Gruppe von Nutzern erlauben, Informationen online zu veröffentlichen, auszutauschen oder zu verwalten. Insbesondere zählen dazu Wikis, Social Bookmarking, Blogs, Microblogging und Soziale Netzwerke (Peters, 2009). In der Regel dienen diese Dienste im Internet vor allem Unterhaltungszwecken. Sie lassen sich aber auch für Aufgaben wie Wissensmanagement oder Information Retrieval und als Kommunikationsmittel nutzen. Als solche haben sie ein hohes Einsatzpotential, um akademische (aber z. B. auch innerbetriebliche) Arbeitsabläufe zu unterstützen (Waldrop, 2008; Weller et al., 2007). Unsere zentrale Fragestellung war, inwiefern derartige Potentiale in akademischen Zusammenhängen erkannt und angenommen werden. Um uns einer Antwort an diese Frage anzunähern, haben wir zunächst in drei Online-Umfragen Studenten und Wissenschaftler zu ihrem Nutzungsverhalten befragt (Weller et al., 2010). Die Ergebnisse sollten erste Impulse für weitere Forschungen zu diesem Thema liefern. In diesem Beitrag werden wir auf einzelne Ergebnisse der drei Umfragen beispielhaft eingehen, insbesondere in Bezug auf die Nutzung von Wikis bzw. Wikipedia. Verwandte Untersuchungen lieferten beispielsweise Head & Heisenberg (2010).

Methode

Wir haben in den vergangenen Jahren die folgenden Umfragen unter verschiedenen Zielgruppen durchgeführt: Gruppe A) deutsche Studenten aus verschiedenen Studiengebieten (1024 Teilnehmer, Winter 2007/08) (Klein et al., 2009), Gruppe B) deutsche Studenten der Informationswissenschaft und verwandter Disziplinen (346 Teilnehmer, Sommer 2009) (Dornstädter & Freimanis, 2010), und Gruppe C) deutsche Wissenschaftler aus verschiedenen Fachbereichen (136 Teilnehmer, Sommer 2009) (Perez, 2010). Alle drei Umfragen wurden als Online-Fragebögen durchgeführt.

Die Studien wurden unabhängig voneinander durchgeführt und hatten jeweils unterschiedliche Ausrichtungen und individuelle Fragestellungen. Alle berücksichtigen dabei jedoch in einem gewissen Rahmen, inwiefern bestimmte Web 2.0 Dienste bekannt sind und wie sie genutzt werden.

Studie A

Im Mittelpunkt dieser Umfrage standen Informationskompetenz und Verhalten bei Informationssuchen. Dazu wurde nach der Bekanntheit und Nutzung verschiedener Recherchetools und Informationsquellen gefragt, darunter auch verschiedene Web 2.0-Angebote. Dieser Fragebogen war über einen Zeitraum von sechs Wochen im Winter 2007/08 online zugänglich. Es konnten generell Studierende aller Hochschulen teilnehmen, aber da der Fragebogen vor allem über einen internen Verteiler der Heinrich-Heine-Universität Düsseldorf verbreitet wurde, waren 95% der Umfrageteilnehmer auch Angehörige dieser Universität. Es nahmen insgesamt 1043 Studierende an der Umfrage teil[1]. Die Teilnehmer kamen aus verschiedenen Studiengängen (darunter: 38% im Bereich Geisteswissenschaft, 29% Naturwissenschaft, 6% Rechtswissenschaft, 5% Sozialwissenschaft, 5% Medizin, 3% Wirtschaftswissenschaft, 14% andere) und befanden sich in verschiedenen Studienphasen.

Studie B

Diese Studie betrachtet vor allem Aspekte der Informationskompetenz bei Studierenden der Informationswissenschaft und verwandten Disziplinen. Nur ein kleinerer Teilbereich behandelt in diesem Kontext auch den Umgang mit Web 2.0-Diensten (darunter Wikipedia und Wikis, Blogs, Pod- und Vodcasts) sowie mit Web-Suchmaschinen und Internetforen. Der Fragebogen war für fünf Wochen im Sommer 2009 online zugänglich. 346[2] Studenten aus insgesamt zehn verschiedenen Hochschulen im deutschsprachigen Raum nahmen an der Umfrage Teil (davon 132 von der Fachhochschule Köln, 46 von der Fachhochschule Chur, 44 von der Heinrich-Heine-Universität Düsseldorf, 38 von der Universität Hildesheim, 35 von der HTWK Leipzig, und 46 von anderen Hochschulen, 5 ohne Angaben).

Studie C

Studie C schließlich befasste sich mit der Fragestellung, inwiefern Soziale Software bei Hochschulmitarbeitern bekannt ist und von diesen genutzt wird. Zu verschiedenen Web 2.0 Diensten wurden daher aufeinander aufbauend die folgenden Fragen gestellt: 1. Kennen Sie Dienst x? Wenn ja: Benutzen Sie Dienst x? Wenn ja: Für welchen Zweck benutzen Sie Dienst x? Sind Sie passiver oder aktiver Nutzer (mit Definition)? Wie hoch ist der Stellenwert von Dienst x für Ihre tägliche Arbeit? Der Online-Fragebogen war für fünf Wochen im Sommer 2009 frei zugänglich und wurde insbesondere an Mitarbeiter der Universität Düsseldorf direkt verschickt, die diesen weiter verbreiten konnten. Daher waren 89% der Umfrageteilnehmer Mitarbeiter der Universität Düsseldorf. Insgesamt nahmen 136 Teil-

[1] Davon konnten in der Regel 931 verwendet werden. Nicht alle Fragen mussten beantwortet werden, um den Fragebogen abzuschließen. Daher wurden alle Fragebögen ausgewertet, auch wenn nicht alle Fragen beantwortet waren. Das gleiche gilt für die Studien B und C. In diesem Artikel wird daher für jede Frage die Anzahl der eingegangenen Antworten separat als „n“ angegeben.

[2] Vergleiche Fußnote 1; 214 Teilnehmer haben alle Fragen komplett beantwortet. Die jeweilige Zahl der Antworten pro Fragestellung wird im Ergebnisteil als „n“ mit angegeben.

nehmer[3] an der Umfrage teil, darunter 16 Professoren, 60 wissenschaftliche Mitarbeiter, 15 Lehrbeauftragte, 20 studentische Hilfskräfte, 13 Doktoranden, 8 andere Hochschulmitarbeiter und 4 ohne Angaben. Die Teilnehmer stammen aus verschiedenen Fachbereichen, darunter Sprach- und Geisteswissenschaft, Sozialwissenschaft, Naturwissenschaft, Lebenswissenschaft und Medizin und Informatik – allerdings machten 50% der Teilnehmer aus Grund des Schutzes der Privatsphäre keine Angaben zum jeweiligen Fachbereich.

Ausgewählte Ergebnisse: Die Nutzung von Wikis und Wikipedia

Es zeigte sich insgesamt eine große Bekanntheit einzelner Dienste, insbesondere solcher, die vor allem zu Unterhaltungszwecken dienen (Soziale Netzwerke wie Facebook oder Videoportale wie YouTube) oder die einen schnellen und kostenlosen Zugang zu Informationen ermöglichen (wie Wikis). Insgesamt am bekanntesten ist klar Wikipedia; diese Community-basierte Enzyklopädie kannten sämtliche Teilnehmer aller drei Studien (siehe Abbildung 1 für Umfrage C, Abbildung 2 für Umfrage A, Abbildung 3 für Umfrage B). Andere Wikis sind ebenfalls vielfach bekannt, werden aber deutlich weniger genutzt als Wikipedia. Nur 2,4% der Studierenden in Umfrage B kennen keine anderen Wikis außer Wikipedia (Abbildung 3), bei den Hochschulmitarbeitern in Umfrage C gaben 38,1% an, keine anderen Wikis zu kennen. Wegen der großen Popularität werden wir uns in dieser Ergebnisdarstellung vor allem mit den Studienergebnissen zu Wikipedia befassen. Die Abbildungen geben aber auch Auskünfte zu anderen Diensten. So sind beispielsweise Social Bookmarking Services (wie Delicious) ein deutlich weniger bekannter Web 2.0-Dienst: 74% der Gruppe A und 76% der Gruppe C kennen Social Bookmarking nicht (Abbildungen 1 und 2). Bei den Hochschulmitarbeitern gaben jedoch 44,4% derjenigen, die Social Bookmarking kennen an, diese Dienste auch zu nutzen (Abbildung 4), und zwar überwiegend aktiv (Abbildung 6) und auch zu beruflichen Zwecken (Abbildung 5).

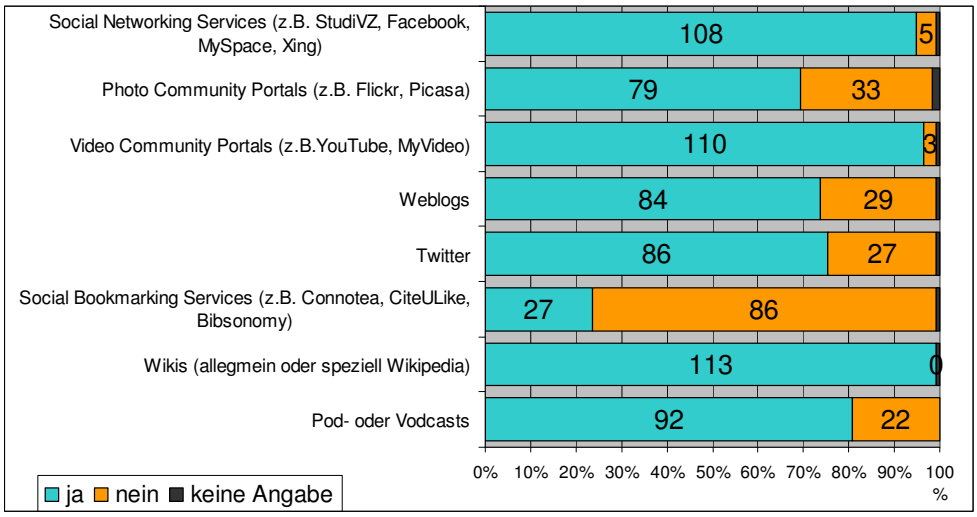


Abbildung 1. Umfrage C (n=114): Kennen Sie die folgenden Web 2.0-Dienste?

[3] Vergleiche Fußnote 1.

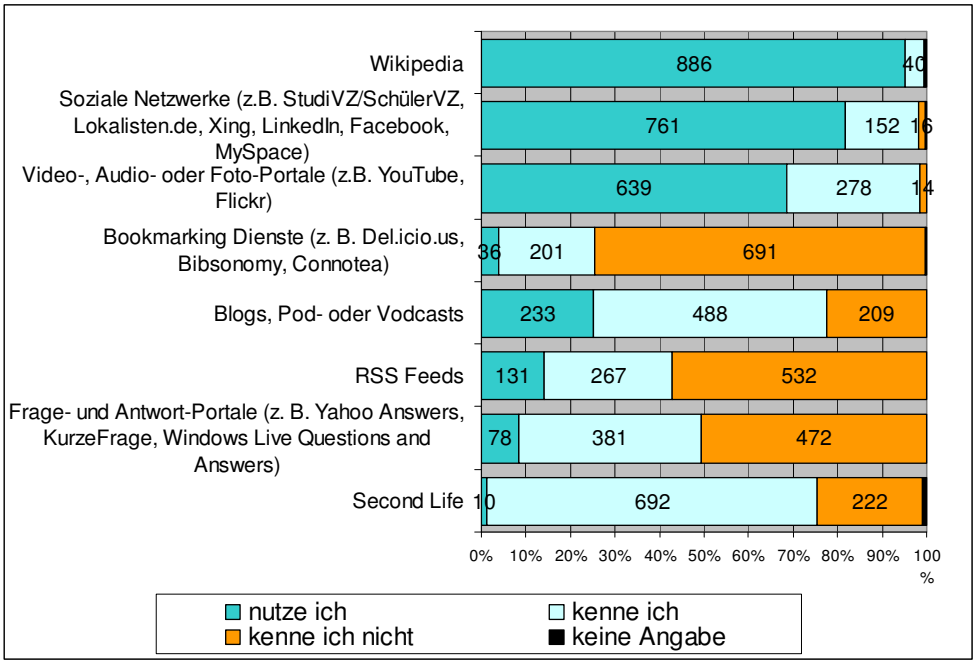


Abbildung 2. Umfrage A (n=931): Nutzen bzw. kennen Sie die folgenden Anwendungen?

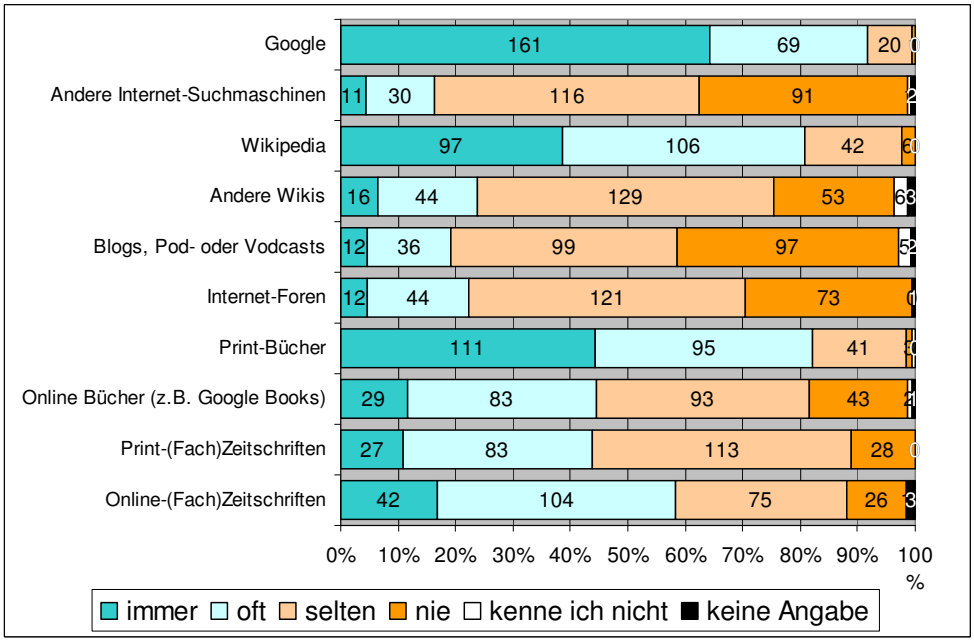


Abbildung 3. Umfrage B (n=251): Wie oft nutzen Sie folgende Informationsquellen bzw. Hilfsmittel für wissenschaftliche Recherchen (z. B. Hausarbeiten, Hausaufgaben, Referate, Abschlussarbeiten)?

Wikis und Wikipedia bei Hochschulmitarbeitern

Teilweise wurde auch näher erfragt, ob und wie einzelne Dienste genutzt werden. Dies soll am Beispiel von Wikis und Wikipedia kurz veranschaulicht werden: Nur 6% der befragten Wissenschaftler (Gruppe C) gaben an, dass Sie überhaupt keine Wikis nutzen (Abbildung 4), 49% nutzen Wikipedia und 45% Wikipedia sowie andere Wikis. Von denjenigen, die Wikis benutzen, gaben 73% an, dies auch im Rahmen ihrer wissenschaftlichen Arbeit zu tun (Abbildung 5). Wikipedia wird vor allem als Nachschlagewerk genutzt, Wikis dienen den

Wissenschaftlern aber auch als Mittel zur Wissensorganisation in Arbeitsgruppen oder als Hilfsmittel, um gemeinsam Publikationen zu verfassen. 30% der Wiki-Nutzer gaben außerdem an, dass sie Wikipedia dazu verwenden, studentische Texte auf Plagiate zu kontrollieren. Die meisten Teilnehmer aus Gruppe C gaben an, passive Wiki-Nutzer zu sein, d. h. sie schreiben in der Regel selbst keine Wiki-Artikel (Abbildung 6).

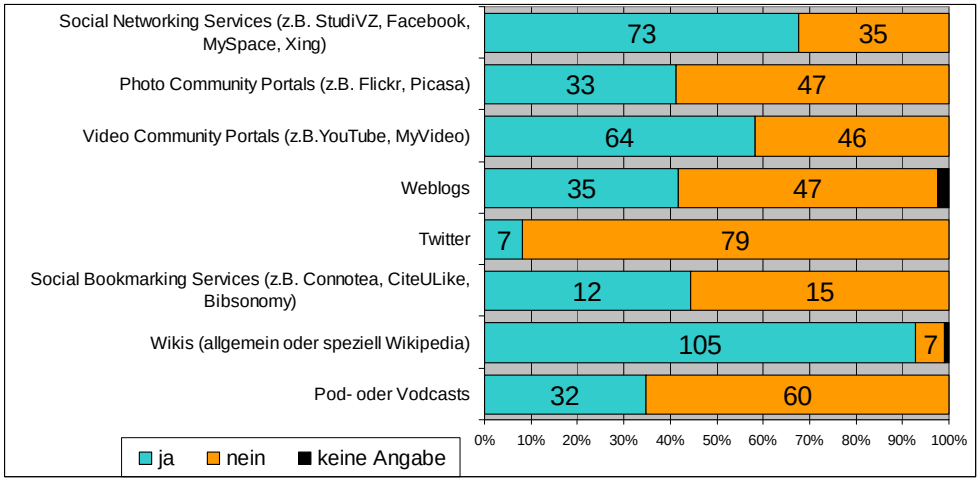


Abbildung 4. Umfrage C: Nutzen Sie die folgenden Web 2.0-Dienste (nur für die Teilnehmer, die zuvor angegeben haben, dass sie die entsprechenden Dienste kennen, vgl. Abb. 1)?

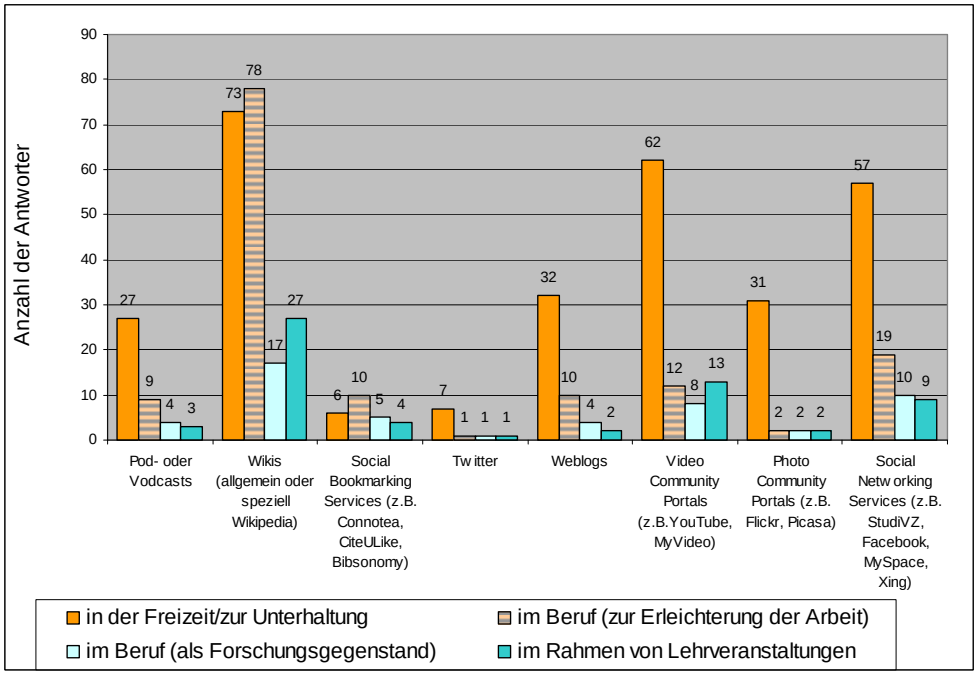


Abbildung 5. Umfrage C: Wozu nutzen Sie die folgenden Web 2.0-Dienste (Mehrfachantworten möglich, nur für die Teilnehmer, die zuvor angegeben haben, dass sie die entsprechenden Dienste nutzen, vgl. Abb. 4)?

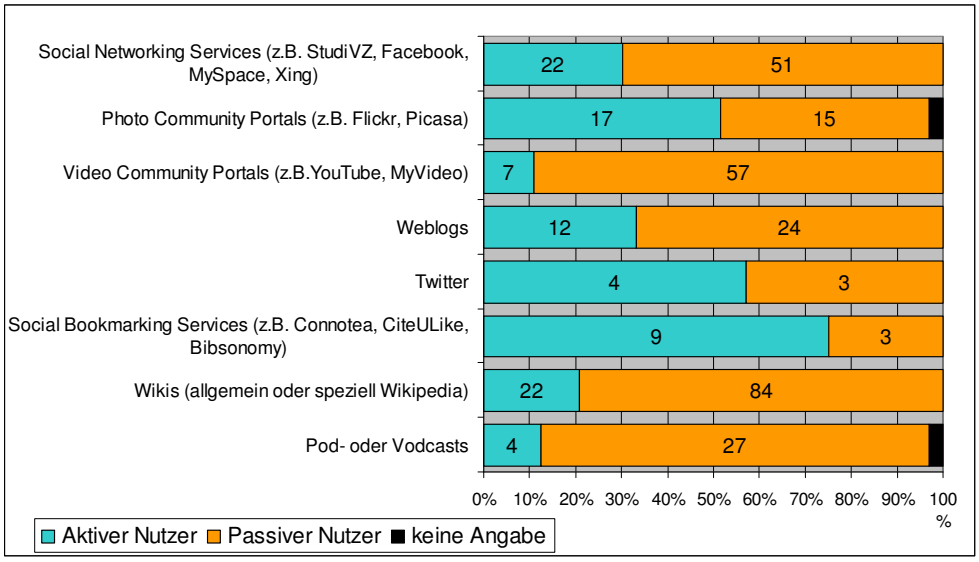


Abbildung 6. Umfrage C: Sind Sie bei den folgenden Web 2.0-Angeboten aktiver oder passiver Nutzer (nur für die Teilnehmer, die zuvor angegeben haben, dass sie die entsprechenden Dienste nutzen, vgl. Abb. 4)?

Wikis und Wikipedia bei Studierenden

Auch für die Studenten in Gruppe A wurde erfragt, ob und zu welchem Zweck sie Wikis nutzen. Es zeigt sich, dass Studenten Wikipedia regelmäßig für studiumsbezogene Recherchen nutzen (Abbildung 7 für Umfrage A, Abbildung 3 für Umfrage B). Nur 7,6% der Teilnehmer aus Umfrage A und nur 2,4% aus Umfrage B gaben an, Wikipedia nie für wissenschaftliche Recherchen zu nutzen.

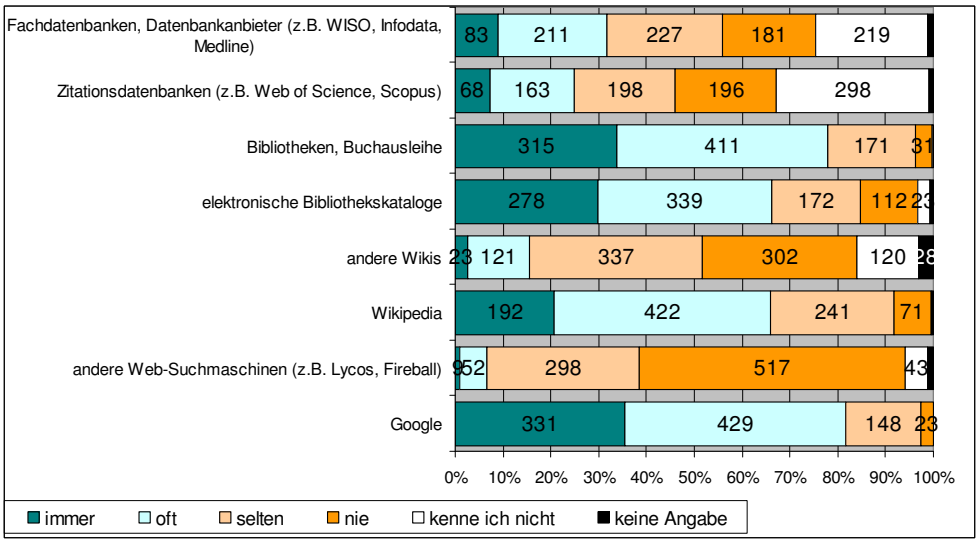


Abbildung 7. Umfrage A (n=931): Wie oft benutzen Sie folgende Mittel für die wissenschaftliche Recherchearbeit?

Auf die Frage in Studie A „Sie suchen eine Begriffsdefinition, wo schauen Sie als erstes nach“ (ohne vorgegebene Antwortmöglichkeiten), wurde Wikipedia mit 377 Nennungen am häufigsten angegeben. 25% der Studenten in Studie A gaben außerdem an, Wikipedia schon einmal in einer Seminararbeit oder Abschlussarbeit zitiert zu haben; 47% haben außerdem die Erfahrung gemacht, dass die Nutzung von Wikipedia von Dozenten verboten

wurde. Die Studierenden der Informationswissenschaft in Umfrage B wurden zudem befragt, inwiefern der angemessene Umgang mit bestimmten Informationsquellen im Studium thematisiert wird (Abbildung 8). Dabei gaben 19,0% an, dass die Nutzungsmöglichkeiten von Wikipedia „sehr gut“ thematisiert werden (44,4% „gut“, n=216).

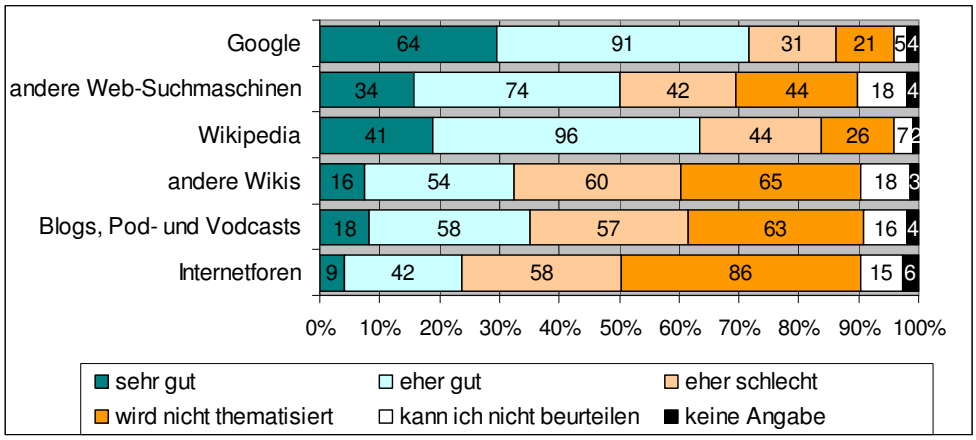


Abbildung 8. Umfrage B (n=216): Wie wird Ihrer Meinung nach der Umgang mit folgenden Informationsquellen bzw. Hilfsmitteln in Ihrem jetzigen Studium vermittelt?

Insgesamt ist bei Studierenden die Nutzung von Wikipedia eher passiv ausgerichtet. 18,7% der Teilnehmer aus Studie A gaben an, schon einmal einen Wikipedia-Artikel bearbeitet zu haben. Erweiterte Funktionen werden nur wenig genutzt: 43,3% haben noch nie die Diskussionsseite eines Wikipedia-Artikels aufgerufen, 48,2% haben noch nie die Versionsgeschichte eines Wikipedia-Artikels angesehen (Abbildung 9).

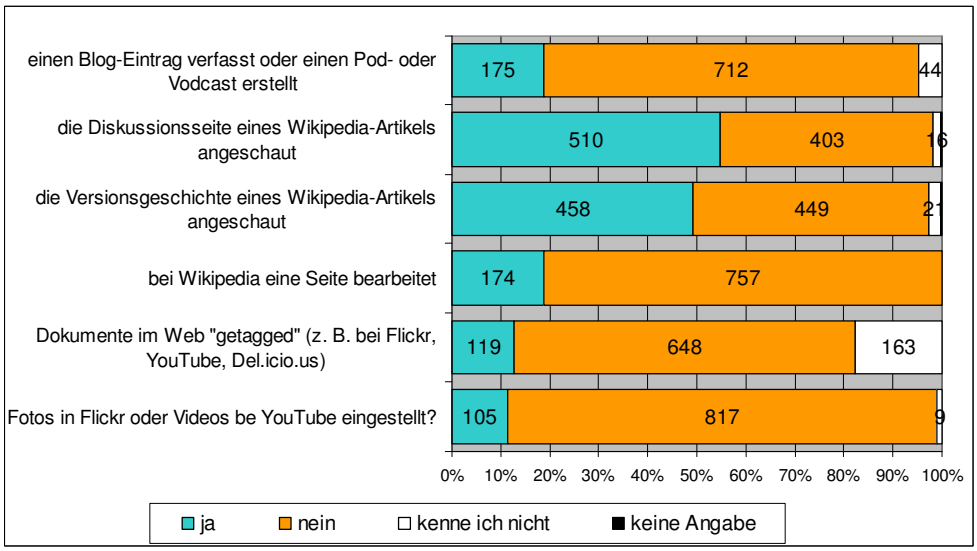


Abbildung 9. Umfrage A (n=931): Haben Sie schon einmal selbst die folgenden Aktionen durchgeführt?

Tendenziell sind die Studierenden aus Studie B etwas skeptischer gegenüber Wikipedia-Inhalten als die Studierenden in Studie A (Abbildung 10). Insgesamt trauen sich Studierende aber generell zu, die Zuverlässigkeit von Wikipedia-Inhalten selbst zu beurteilen (in Studie A antworteten nur drei Teilnehmer, dass sie die Qualität nicht beurteilen können, in Studie B keiner).

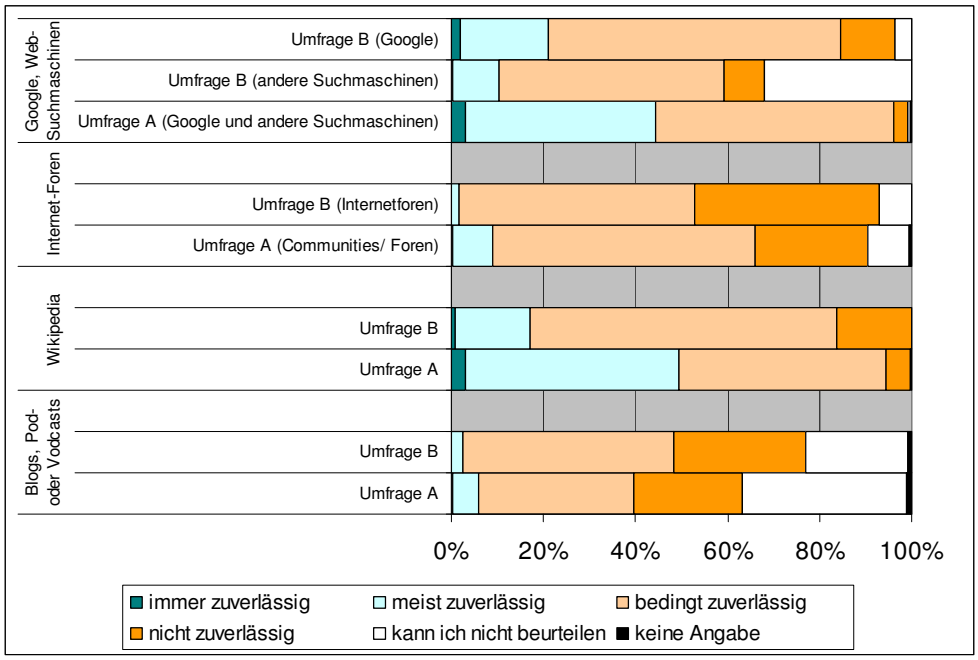


Abbildung 10. Umfrage A (n=929) und Umfrage B (n=238). Als wie zuverlässig schätzen Sie die Informationen ein, die Sie über die folgenden Wege bekommen?

Fazit

Die hier zusammengefassten Ergebnisse bilden nur einen Teilbereich der durchgeführten Studien ab, für vollständige Informationen verweisen wir auf Klein et al. (2009), Freimanis und Dornstädter (2010), Perez (2010) sowie Weller et al. (2010). Die Ergebnisse der drei Studien zeigen insgesamt, dass die Einsatzmöglichkeiten von Web 2.0-Angeboten sowohl von Studenten als auch von Wissenschaftlern bisher nur teilweise erkannt und ausgeschöpft werden. Wikis bzw. speziell Wikipedia stellen sich dabei als die am bekanntesten und am häufigsten genutzten Angebote heraus. Wikipedia hat bei allen befragten Gruppen einen hohen Stellenwert als Nachschlagewerk, insbesondere in frühen Phasen der wissenschaftlichen Recherche. Die Nutzung bleibt dabei aber eher passiver Natur. Weniger als ein Fünftel der Studierenden in Umfrage A haben bereits eine Wikipedia-Seite bearbeitet und auch nur 21% der Hochschulmitarbeiter in Umfrage C betrachten sich selbst als aktive Wiki-Nutzer.

Die hier gewonnenen Erkenntnisse lassen verschiedene vertiefende Anschlussstudien sinnvoll erscheinen. Insbesondere sind weitere Studien notwendig, die (anders als im Fall unserer drei voneinander unabhängigen Umfragen) aufeinander aufbauend und standardisiert verschiedene Zielgruppen befragen. Neben quantitativen wären auch qualitative Studien wünschenswert. Unsere Studien haben zudem bislang einen weiten Bereich verfügbarer

Social Software Dienste noch unberücksichtigt gelassen (z. B. Wikiversity, Nature Networks oder Slideshare).

In Anlehnung daran wäre das abschließende Ziel die Erarbeitung von Leitlinien für die effektive und angemessene Nutzung von Social Software Angeboten in akademischen Zusammenhängen.

Danksagung

Dank gilt allen Teilnehmern, die bei einer der drei Umfragen mitgemacht haben. Dank geht auch an Lisa Beutelspacher, Katharina Hauk, Christina Terp, Denis Anuschewski, Christoph Zensen and Violeta Trkulja für ihre Mitarbeit an Studie A. Darüber hinaus danken wir den Kollegen der Abteilung für Informationswissenschaft, Heinrich-Heine-Universität Düsseldorf, besonders Isabella Peters und Wolfgang G. Stock. Und wir danken Bernd Klingsporn für eine zusätzliche Überprüfung der Ergebnisse.

Referenzen

Freimanis, R., & Dornstädter, R. (2010). Informationskompetenz junger Information Professionals: Stand und Entwicklung. *Information – Wissenschaft und Praxis*, 61 (2), 123-128.

Head, A.J., & Heisenberg, M.B. (2010). How today's college students use Wikipedia for course-related research. *First Monday*, 13(3), abgerufen am 22.10.2010 unter http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2830/2476.

Klein, R.N., Beutelspacher, L., Hauk, K., Terp, C., Anuschewski, D., Zensen, C., Trkulja, V., & Weller, K. (2009): Informationskompetenz in Zeiten des Web 2.0. Chancen und Herausforderungen im Umgang mit Social Software. *Information – Wissenschaft und Praxis*, 60 (3), 129-142.

Perez, M. (2010). Web 2.0 im Einsatz für die Wissenschaft. *Information – Wissenschaft und Praxis*, 61 (2), 128-134.

Peters, I. (2009). *Folksonomies. Indexing and Retrieval in Web 2.0*. De Gruyter, Saur: Berlin.

Waldrop, M.M. (2008). Science 2.0: Great New Tool or Great Risk? *Scientific American*, abgerufen am 22.10.2010 unter http://www.scientificamerican.com/article.cfm?id=science-2-point-0-great-new-tool-or-great-risk.

Weller, K., Mainz, D., Mainz, I., & Paulsen, I. (2007). Wissenschaft 2.0? Social Software im Einsatz für die Wissenschaft. In M. Ockenfeld (Ed.), *Proceedings der 29. Online Tagung der DGI*. Frankfurt a.M.: DGI, 121-136.

Weller, K., Dornstädter, R., Freimanis, R., Klein, R.N., & Perez, M. (2010). Social Software in Academia: Three Studies on Users' Acceptance of Web 2.0 Services. In *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, Raleigh, NC, USA, abgerufen am 22.10.2010 unter http://journal.webscience.org/360/.

E-Science und Forschungsdatenmanagement

Die Bibliothek als Dienstleister für den Umgang mit Forschungsdaten

Johanna Vompras, Jochen Schirrwagen & Wolfram Horstmann

Universitätsbibliothek Bielefeld
Universitätsstr. 25
D-33615 Bielefeld
{johanna.vompras, jochen.schirrwagen, wolfram.horstmann}@uni-bielefeld.de

Zusammenfassung

Der Bedarf an Dienstleistungen für ein effektives Forschungsdaten-Management stellt die Wissenschaft und Infrastruktureinrichtungen vor große Herausforderungen. Für die Erstellung eines allgemeingültigen Konzeptes zur Verwaltung von Forschungsdaten werden an der Universitätsbibliothek Bielefeld fachspezifische Pilotprojekte gestartet, in denen Anforderungen und wissenschaftliche Arbeitsabläufe in den verschiedenen Disziplinen analysiert werden. Somit können geeignete Infrastrukturen geplant und realisiert werden.

Einleitung

Forschungsdaten stellen eine wesentliche Grundlage für den wissenschaftlichen Erkenntnisprozess dar. Aber wie gestaltet man einen effizienten und nachhaltigen Umgang mit ihnen? Dieser Frage begegnet die Universitätsbibliothek Bielefeld mit der Zielsetzung, ein Dienstleistungskonzept für den Umgang mit Forschungsdaten für Bielefelder Wissenschaftler zu erstellen – eine Idee, die aus konkreten Anfragen von Seiten der Forschenden entstanden ist. Dazu werden spezialisierte Pilotprojekte mit fachlichen Initiativen gestartet. Indem Forschende ihre Anforderungen selbst definieren, kann die Vielfalt wissenschaftlicher Arbeitsabläufe und Typen von Forschungsdaten optimal berücksichtigt werden. Als Ergebnis der Anforderungsanalyse lassen sich geeignete Infrastrukturen planen und realisieren. In einem weiteren Schritt können Generalisierungen abgeleitet werden, um ein allgemeingültiges Dienstleistungskonzept für die Verwaltung von Forschungsdaten zu erstellen. Dazu gehören unter anderem die Bereitstellung von Speicher- und Rechenkapazität in Kooperation mit dem Rechenzentrum sowie die Beratung und technische Unterstützung bei der Nutzung und Verarbeitung geeigneter Daten- und Metadatenformate.

Die Motivation für das Forschungsdatenmanagement baut auf der Tatsache auf, dass die mühselige und teure Erhebung sowie die Aufbereitung und Dokumentation von Forschungsdaten in der Wissenschaft nicht gerade als eine vorrangige Aufgabe angesehen wird. Vor allem in kleineren Projekten stellt die nachhaltige Bereitstellung und Archivierung der gewonnenen Daten einen erheblichen Mehraufwand dar, so dass dieser Schritt oftmals unterbleibt und die Daten der Öffentlichkeit verwehrt bleiben. Stattdessen werden diese undokumentiert auf privaten PCs, Speichermedien oder in der Schreibtischschublade aufbewahrt.

Die Universitätsbibliothek ist sich bewusst, dass zur guten wissenschaftlichen Praxis ebenso ein korrekter Umgang mit Forschungsdaten gehört. Dazu gehört die Datenaufbereitung,

z. B. Erhebung und ihre Dokumentation, wobei die Datenbeschreibung mittels von für die jeweilige Wissenschaftsdisziplin gängigen Formatstandards durchgeführt werden muss, um einen direkten Austausch zu gewährleisten. Als nächster Aspekt steht die Sicherung und die systematische Langzeitarchivierung im Mittelpunkt und zu allerletzt die Publikation der Daten, die diese in eigenständige wissenschaftliche Objekte überführt. Publizierte Forschungsdaten liefern der Wissenschaft neue Use-Cases: die wissenschaftlichen Arbeiten werden nach außen sichtbar und nachvollziehbar und dienen als Motivation für neue Experimente und Forschungsfragen. Darüber hinaus wird durch die eindeutige Identifikation der Datensätze die Zitierbarkeit, Verlinkung und Katalogisierung der Daten möglich gemacht.

Das Pilotprojekt „Datenservice-Zentrum für Betriebs- und Organisationsdaten“

Ein erstes Pilotprojekt wurde mit der Initiative zum Aufbau eines Datenservice-Zentrums in der Soziologie gestartet. Im Mittelpunkt steht dabei die Schaffung eines Archivs für quantitative und qualitative Betriebs- und Organisationsdaten, mit dem Ziel diese zentral zu archivieren und für sekundäranalytische Zwecke zur Verfügung zu stellen. Diese sogenannten Mikrodaten stellen eine Form von Primärdaten dar, die im Rahmen von Studien aus Erhebungen, Tests oder Befragungen gesammelt werden und für die speziell ein Beschreibungsstandard durch die Data Documentation Initiative (DDI) entwickelt wurde. DDI ist ein XML-basiertes Metadaten-Format für die Dokumentation und Austausch von sozialwissenschaftlichen Datensätzen. Ein solcher Datensatz im DDI-Format (sog. DDI Instanz) beinhaltet unter anderem Informationen darüber, wie das Projekt entstanden ist, wie die Daten erhoben, aufbereitet und analysiert wurden und welchem Themenbereich das Projekt einzugliedern ist. Auch detailliertere Angaben zu den in den Befragungen benutzten Fragebögen, der internen Variablenkodierung oder mit dem Datensatz verknüpften verantwortlichen Personen werden in einer solchen Datei dokumentiert. Abbildung 1 skizziert beispielhaft eine mögliche Datennutzung in einem Archiv-basiertem Forschungsprozess. Als erstes werden relevante Studieninformationen – nach einer vorhergehenden Datenbereinigung (Data Cleaning) – in das DDI Format überführt und in einer zentralen Datenbank archiviert. Durch das standardisierte Format kann mittels Anwendungssoftware eine Recherche in den Daten selbst und in ihren Dokumentationen durchgeführt werden. Dabei können Studien nicht nur durchsucht und ihre Ergebnisse analysiert werden, sondern bereits existierende Methoden (z. B. Variablenschemas, Konzepte) für Replikationsstudien wiederverwendet werden.

Die DDI-Spezifikation der aktuellen Version 3.1 basiert auf XML Schemas und hat das Data Life Cycle Modell zur Grundlage. DDI-Module, d. h. die einzelnen Komponenten einer DDI-kodierten Datei, entsprechen einzelnen Stadien des Datenlebenszyklus den sozialwissenschaftliche Daten typischerweise durchlaufen. Dazu gehört z. B. der Entwurf eines Fragebogens, Methodologie der Datensammlung und Analyse, aber auch die Indexierung, Katalogisierung und Langzeitarchivierung der erhobenen Daten. Die in den einzelnen Modulen enthaltenen Metadaten können in der Definition späterer Stadien durch Referenzierung wiederbenutzt werden. DDI erlaubt ebenfalls die Strukturierung von Informationen in

Längsschnittstudien (z. B. Panelstudien). Hier gibt es ein spezielles Modul zur Vergleichbarkeit der einzelnen zeitlich versetzen Erhebungen.

Die Benutzung eines normierten, persistenten Adressierungsschemas wie das der Uniform Resource Names (URNs) stellt eine Basis zur Identifizierung und Versionierung von DDI-Elementen dar. Durch die strikte Trennung zwischen Daten und Schematas, wird die Referenzierung und somit die Wiederverwendbarkeit (re-use) der Daten selbst (z. B. Variablen) und der Metadatenschemata (z. B. Codeschemata) innerhalb einer DDI Instanz ermöglicht. Darüber hinaus ist ein Austausch dieser Daten zwischen kooperierenden Forschungsinstituten möglich. Ein weiteres Merkmal von DDI besteht aus der Definition eines Profils, das eine spezielle Teilmenge von verwendeten/nicht verwendeten DDI-Elementen und Attributen darstellt, die für die Studienbeschreibungen innerhalb einer Anwendung oder Organisation verwendet werden. Dieses Profil ermöglicht eine automatisierte Verarbeitung der Daten und dient ebenfalls als eine Art Dokumentation sowohl für den Datenproduzenten als auch für den Remote-User.

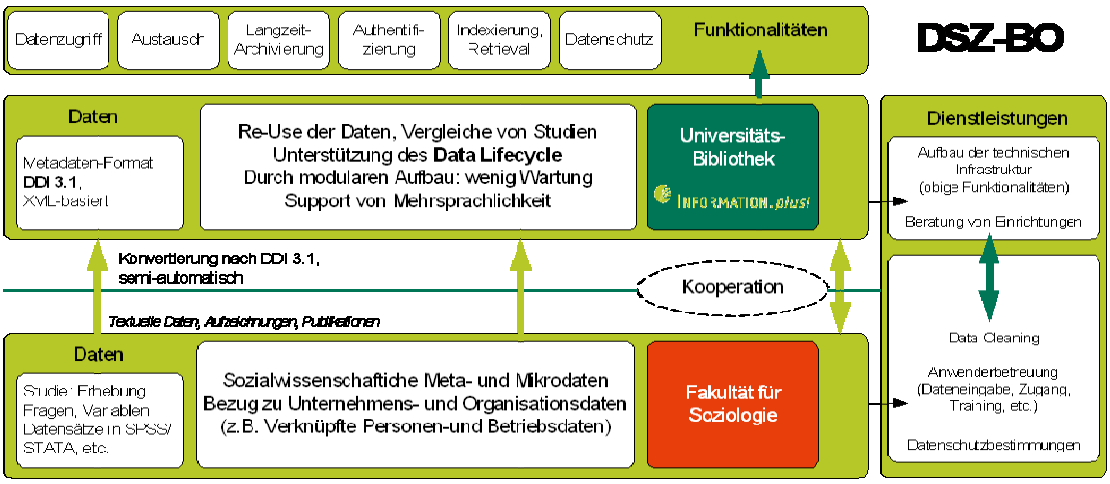


Abbildung 1: Datenservice-Zentrum für Unternehmens- und Organisationsdaten (DSZ-BO)

Abbildung 1 stellt einen groben Überblick über das Zusammenspiel der beiden Kooperationspartner Soziologie und Bibliothek dar und präsentiert die einzelnen Komponenten der technischen Infrastruktur für das Datenservicezentrum.

Während der Vorphase des Projektes wurde beispielhaft erprobt, inwieweit sich Studien aus dem Bereich der Betriebs- und Organisationsforschung, z. B. ALLBUS Betriebsbefragung, in Abbildung 1: Datenservice-Zentrum für Unternehmens- und Organisationsdaten (DSZ-BO)DDI-Format kodieren lassen. Da die Daten einen Bezug zu Organisationen aufweisen (z. B. zu Betrieben, Unternehmen, Vereinen, Verbänden, Kindergärten, Schulen oder Einrichtungen des Gesundheitswesens) und hinterher für Fragestellungen der Organisationsforschung nutzbar sein sollen, bedürfen sie einer speziell darauf ausgerichteten Dokumentation. Als Beispiel dafür sei die Bereitstellung von Vokabularien für diverse Betriebsformen oder die Dokumentation von räumlichen Gegebenheiten, wie z. B. Angabe des Bundeslandes, in dem die Studie durchgeführt wurde, genannt.

Im nächsten Schritt geht es um die Implementierung des Prototypen. Dieser umfasst die in Abbildung 1 dargestellten Funktionalitäten, von denen hauptsächlich die effiziente Speiche-

rung, Indexierung der DDI-Files und die webbasierte Rechercheschnittstelle im ersten Teil des Pilotprojektes eine Rolle spielen. Die technische Infrastruktur wird die folgenden Funktionalitäten bieten:

Retrieval: Auf der Basis der in das DDI-Format überführten Studiendokumentationen wird eine Suche nach Studien durch Filterung möglich (z. B. Filterung nach Erhebungsjahr, Bundesland oder Thema), in der Kriterien und Vergleichsoperatoren angegeben werden können. Anders als bei der Suche wird beim Filtern nur eingeschränkt, welche Datensätze angezeigt werden. Innerhalb von Studien wird ebenfalls eine Metadaten-basierte (machine-actionable) und Volltextsuche auf deskriptiven Elementen (human-readable) möglich sein, wie zum Beispiel in den Fragetexten oder zusätzlichen Studienbeschreibungen.

Beispiel für eine Filterung/Suche könnte wie folgt aussehen: „Finde alle Studien zum Thema Betriebliche Ungleichheit, bei denen sich die Ergebnisse nach Bundesländern auswerten lassen“.

Die Suche wird in den folgenden Ebenen ausgeführt:

1. Studienebene: z. B. Projektname, Datensatzbezeichnung, Autoren, Kurzbeschreibung, Finanzierung, Kontaktangaben und Publikationen

2. Erhebungsebene: Informationen zum Forschungsdesign, z. B. Erhebungszeitraum, Methoden der Datenerhebung, verfügbares Dokumentationsmaterial und zugehörige Datensätze

3. Frageebene: Fragetext, Interviewerанweisungen, Filterführungen, Fragehinweise, Konstrukte, Stichwörter, Vorlagen aus anderen Studien

4. Variablenebene: Variablen-ID, Variablenname und Label, Ausprägungsnummer und Label, gültige und fehlende Werte, Skalenniveau und entsprechende Parameter (Mittelwerte etc.), weitere deskriptive Statistiken

Die graphische Benutzerschnittstelle erlaubt eine übersichtliche und leicht bedienbare Retrievalmöglichkeit in den Studieninformationen und den dazugehörigen Metadaten. Das Layout der einzelnen Masken orientiert sich inhaltlich an Informationen, die für die Darstellung der Dokumentation von Organisations- und Betriebsdaten notwendig sind. Die Konzeption der Datenbank, die Realisierung von Abfragen und Visualisierung der Ergebnisse werden gezielt im Hinblick auf die leichte Pflege der Datenbank, flexible Erweiterbarkeit der Metadaten und die zukünftige Aufnahme weiterer Studien mit Bezug zur Betriebs- und Organisationsebene erstellt.

Datenschutz/Authentifizierung: Die Fakultät für Soziologie wird im Rahmen des DSZ-BO-Services die von den Datenproduzenten übersandten Mikrodaten für ihre sekundäranalytische Nutzung aufbereiten. Einen für den Datenschutz wichtigen Aspekt stellt hierbei die Anonymisierung der Daten dar. In Kooperation mit den Datenproduzenten werden hierfür spezielle Datenschutzrichtlinien für die Herausgabe der Daten herausgearbeitet. Insbesondere sind damit Anonymisierungsverfahren verbunden, bei denen Informationen aggregiert, bzw. gelöscht werden, um den Detaillierungsgrad der Variablen zu reduzieren. Damit

wird verhindert, dass bei einer Weitergabe an Dritte bestimmte Individuen (z. B. Personen oder Betriebe) durch Verknüpfung von Informationen identifiziert werden.

Der Zugang für registrierte Personen unterscheidet sich sowohl hinsichtlich der Anonymität der nutzbaren Daten als auch in der Art der Datenbereitstellung. Es werden auch Unterschiede im Personenkreis gemacht, dem die Daten zugänglich gemacht werden dürfen (Nutzerkreis). Das geschieht auf der Basis von mit Datenproduzenten abgestimmten Nutzungsverträgen. Je nach Zugangsrecht des Benutzers werden die Datensätze zur Off-Site-Nutzung (z. B. Public Use Files (PUF), Scientific Use Files (SUF)) oder zur On-Site-Nutzung (z. B. kontrollierte Datenfernverarbeitung oder Nutzung an speziell dafür eingerichteten Arbeitsplätzen) verfügbar gemacht.

Austausch: Ziel der Ausbaustufen ist ein Archiv, das basierend auf DDI den direkten Austausch von Studien zwischen Forschenden und Datenarchiven ermöglicht und somit die Wiederverwendung von Datensätzen unter neuen Fragestellungen in Form von Sekundäranalysen erlaubt. Ein übergreifender Austausch von Forschungsdaten setzt voraus, dass sie in standardisierter Form dokumentiert (hier im DDI-Metadatenformat) und bereitgestellt werden.

Für kooperierende Datenproduzenten bedeutet die Nutzung der Infrastruktur einen erleichterten Zugang zum Bestand an Betriebs- und Organisationsdaten, die Inhalte und die verwendeten Methoden mittels einer webbasierten Oberfläche recherchierbar macht.

Langzeitarchivierung: Durch die Langzeitarchivierung wird der Erhalt der Forschungsdaten sichergestellt und ihre Langzeitverfügbarkeit gewährleistet. Durch die Benutzung eines standardisierten Formats zur Dokumentation bleibt ebenfalls die eindeutige Interpretierbarkeit der Inhalte über eine lange Zeit bestehen. In der ersten Phase des Pilotprojektes wird das Datenservicezentrum für die komplexen und umfangreichen Metadaten der Betriebs- und Organisationsforschung eine effiziente und sichere Informationserfassung ermöglichen. Das zukünftige Konzept sieht einen webbasierten Editor für die zentrale Erfassung der Daten vor, die dann gemeinsam mit ihren logischen Verknüpfungen zu Zusatzmaterialien automatisiert in das DDI Format überführt und langfristig in einer Datenbank gesichert werden.

Fazit

Das Forschungsdatenmanagement steigert vor allem die Arbeitsbedingungen für Wissenschaftler, in dem es neue Recherchemöglichkeiten eröffnet und Forschungsergebnisse mit einer neuen Qualität versieht. Das erste Pilotprojekt soll als Initiator dienen, eine disziplinübergreifende Infrastruktur zu schaffen, die es Wissenschaftlern ermöglicht, ihre Forschungsdaten standardisiert zu dokumentieren, zu versionieren und der Öffentlichkeit zugänglich zu machen. Dem erhöhten Aufwand, der mit der Dokumentation verbunden ist, stehen zahlreiche Vorteile gegenüber: Die mit Forschungsdaten verknüpften Publikationen erhalten einen erheblichen Mehrwert, der sich in der erhöhten Aufmerksamkeit gegenüber den Forschungsergebnissen oder in einer möglichen erhöhten Zitationsrate der Publikation bemerkbar macht und somit als Qualitätsnachweis für die wissenschaftliche Tätigkeit dient.

Durch den möglichen Austausch der Daten werden Replikationsstudien und Sekundäranalysen ermöglicht und das Aufleben neuer wissenschaftlicher Fragestellungen gefördert.

Es ist offensichtlich, dass die bibliothekarische Expertise, die es versteht, Publikationsdaten so aufzubereiten und zu erschließen, dass sie über Suchmaschinen und Bibliothekskataloge optimal recherchierbar sind, im Bereich des Forschungsdatenmanagements nicht fehlen darf. Eine aktive Gestaltung der Infrastruktur für die Bereitstellung von Serviceangeboten im Bereich des Forschungsdatenmanagements bedeutet für die Bibliothek eine Positionierung während des gesamten Forschungsprozesses, d. h. von der Schaffung von Datenstrukturen für die Datenerhebung und Aggregation bis hin zur Publikation der Forschungsdaten. In Kooperationen mit den verschiedenen Forschungsinstituten und dem Hochschulrechenzentrum wird es gelingen, den Wissenschaftler auf organisatorischer und technischer Ebene bei der Datenfreigabe zu unterstützen, damit die mühselig erhobenen Forschungsdaten langfristig erhalten bleiben.

Bibliotheken und Bibliothekare im Forschungsdatenmanagement

Stefanie Rümpel & Stephan Büttner

Fachhochschule Potsdam
Fachbereich Informationswissenschaften
Friedrich-Ebert-Straße 4
D-14467 Potsdam
stefanie.ruempel@gmx.de, st.buettner@fh-potsdam.de

Zusammenfassung

Der Beitrag gibt einen Überblick über die Rolle der Informationsinfrastruktur in Wissenschaft und Forschung beim Umgang mit Forschungsdaten und zeigt Desiderate informationswissenschaftlicher Forschung. Es wird der Frage nachgegangen, welche Kompetenzen für das Berufsbild „datenorientierte Bibliothekare“ die zurzeit ausgebildeten Diplombibliothekare bzw. BA bereits vorweisen können und inwieweit neue Kompetenzen gefordert werden um Tätigkeiten im Forschungsdatenmanagement zu übernehmen.

Einleitung

Wissenschaftliche Bibliotheken und Bibliothekare stehen traditionell am Ende des geistigen Schaffensprozesses. Sie konzentrieren sich auf die Ergebnisse der Wissenschaftler, die Publikationen, erschließen diese und stellen sie der wissenschaftlichen Community zur Verfügung. Doch was ist mit den Daten aus dem Forschungsprozess? Diesbezüglich wurde, zumindest bisher, billigend in Kauf genommen, dass die Forschungsdaten bspw. erhoben bei Messungen, oft für immer verloren gingen. Gegenwärtig entwickeln sich die Forschungsdaten zu einem zentralen Thema. Mit teilweisen oder komplett virtuellen Wissens- und Forschungsumgebungen ändert sich das Aufgabengebiet der in die Informationsinfrastruktur eingebundenen Experten gravierend. Von der Ideengenerierung über die experimentelle Datenerhebung, der Aggregation und der Kollaboration, bis zur Publikation begleiten diese Forschungsumgebungen den Forschungsprozess von Anfang bis Ende. Bereits 1998 wurde mit der DFG-Denkschrift „Sicherung guter wissenschaftlicher Praxis“ (DFG, 1998) in der Empfehlung 7 angeraten: „Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden.“ (ebd., S 12). Dennoch hat es noch Jahre gebraucht, bis diese Empfehlung in der wissenschaftlichen Community als auch bei der Informationsinfrastruktur angekommen ist. Andererseits gab und gibt es bei den einzelnen Akteuren durchaus berechtigte Hemmschwellen. Den Wissenschaftlern fehlen Anreize, um z. B. Daten als Publikation zu werten. Es gibt Angst vor „Missinterpretation“ der Daten durch Dritte. Zudem ist eine Bereitschaft zur Beschreibung der Daten mit Metadaten kaum ausgeprägt. Es fehlen Vorgaben (Policies) von Seiten der Verwaltung, den Forschungsförderern und den Verlagen, zum Umgang mit Forschungsdaten. Zunehmend greift die Erkenntnis, dass auch die Informationstechnologie nicht Probleme löst, sondern ein Tool zur Problemlösung ist.

Aktuelle Ergebnisse von Forschungsprojekten sowie einer Diplomarbeit zeigen deutlich, dass Datenmanagement eine neue Ausprägung des Wissensmanagements darstellt, mithin ein originäres Thema der Informationswissenschaften ist. Es geht um Bewertung und Einordnung in Kontexte, um Metadaten, um Ontologien. Und, essentiell geht es um die Verknüpfung der Daten mit den Experten.

Bibliotheken als Teil der Informationsinfrastruktur werden deutlich als ein Akteur im Forschungsdatenmanagement positioniert bzw. angesehen. Betont wurde dies auf dem Bibliothekskongress in Leipzig 2010 durch die Session „Bibliotheken und Forschungsdaten". Durch E-Science wächst die Erkenntnis, dass der Wert der Forschung insbesondere in den Daten steckt und sich daher das Arbeitsspektrum auf das Primärobjekt, die Forschungsdaten, erweitern muss.

„[…] data is the currency of science, even if publications are still the currency of tenure. To be able to exchange data, communicate it, mine it, reuse it, and review it is essential to scientific productivity, collaboration, and to discovery itself." (Gold, 2007)

Um einen Mehrwert von Daten zu erhalten, muss ein funktionierendes Forschungsdatenmanagement geschaffen werden. Die gespeicherten und gepflegten Daten können erneut interpretiert werden, ermöglichen die Nachvollziehbarkeit der Forschungsergebnisse (Qualitätskontrolle) und dienen der Vermeidung von kostenintensiven Messwiederholungen. (Dallmeier-Tiessen, 2010, S. 3-4, 7). Dies verlangt nach Veränderungen der Arbeiten aller Akteure, die im Wissenschaftsprozess eingebunden sind. Konkretisiert auf Bibliotheken heißt dies, sie stehen nicht mehr am Ende der Forschung, sondern müssen bereits von Beginn an agieren. Was bedeutet dies konkret für Bibliothekare? Diese Frage wurde weder in Leipzig noch in den existierenden deutschen Publikationen beantwortet. Die Diplomarbeit „Data Librarianship – zu Anforderungen an Bibliothekare im Forschungsdatenmanagement" sollte Klarheit in den Diskussionen zu E-Science bringen. Im Folgenden wird auf wesentliche Erkenntnisse der Arbeit eingegangen, jedoch in stark verkürzter Form (Rümpel, 2010).

In den USA, Kanada und Großbritannien gibt es bereits eine Anzahl von Bibliothekaren, die intensiv mit Forschungsdaten arbeiten – die sog. „Data Librarians". Eine Berufsgruppe, die zwar sehr klein ist, da sich nur die größeren Universitätsbibliotheken einen Data Librarian in ihrer Einrichtung leisten, aber Dienstleistungen in Bezug auf Forschungsdaten anbietet. Die Recherche nach Data Librarians in Deutschland hingegen ergab kaum Ergebnisse. Es erfolgt keine Identifizierung mit diesem Begriff. Doch wenn bspw. US-amerikanische Bibliothekare datenorientiert arbeiten, sollte dies auch in Deutschland notwendig und möglich sein.

Ein wesentlicher Inhalt der Diplomarbeit ist die Aufführung möglicher Aufgabenfelder für Bibliothekare im Forschungsdatenmanagement aus der deutschen Perspektive gewidmet. Die vorhandenen bibliothekarischen Kompetenzen bildeten die Ausgangslage. Ziel und Sinn war nicht die Schaffung eines neuen Berufsbildes, sondern zu untersuchen, ob das existierende sich dem neuen Informationsobjekt, den Forschungsdaten, annehmen kann. Als Folge dieser Analyse mussten auch Konsequenzen für die Ausbildung sowie die gegenwärtige Situation von Bibliothekaren im Forschungsdatenmanagement betrachtet werden.

Eine erste Analyse, ermöglicht durch die Existenz von Stellenausschreibung zu Data Librarians, ließ notwendige Kompetenzen und Aufgabenfelder bei der Arbeit mit Forschungsdaten erkennen. Da es im Wesentlichen um die deutsche Perspektive ging, reichte dies jedoch nicht aus. Zur Vervollständigung wurden deutsche Experten aus dem Bereich Bibliothek und Forschungsdaten interviewt, um die notwendige Perspektive und Aktualität zu erreichen.

Kompetenzen und Aufgabenfelder

Bibliothekare erwerben im Studium Kompetenzen, die für die bibliothekarischen Aufgaben wie das Sammeln, Bewahren, Ordnen, Bereitstellen und Vermitteln von veröffentlichten Informationsquellen aller Art notwendig sind (vgl. Plassmann, 2010, S. 10). Diese Aufgabenfelder können adäquat auf das Forschungsdatenmanagement übertragen werden. Auch Daten werden gesammelt (Aufbau von Datensammlungen im Forschungsdatenrepository), bewahrt (Curation, Preservation), geordnet (mit Metadaten versehen), bereitgestellt (zugänglich in Repositorien) und vermittelt (Beratung, Vermittlung von Datenkompetenz). Der wesentliche Unterschied zu den Tätigkeitsfeldern mit klassischen Medien besteht darin, dass datenorientierte Bibliothekare nicht Informationsquellen nach der Veröffentlichung vermitteln, sondern bereits ab der Entstehung. Doch unabhängig von diesen Differenzen sind es auch immer Informationsobjekte. Diese Aussage wird ebenfalls durch die Betrachtung der notwendigen Kompetenzen bestärkt. Durch die Stellenausschreibungen von Data Librarians wurde erkannt, dass Kommunikation, Kundenservice, Zwischenmenschlichkeit, Erschliessung und Vermittlung besonders gefordert werden. Dies sind Fähig- und Fertigkeiten, die jeder Bibliothekar nach seinem Studium besitzen sollte. Weiterhin sind Kenntnisse der Datenformate, Analysesoftware, Urheberrecht und Bewertung von Daten notwendig. Dies sind ebenfalls Kompetenzen, die Bibliothekare seit jeher haben. Kernaufgabe von Informationswissenschaftlern ist die Bewertung von Informationen, und auch in Fragen zum Urheberrecht sind Bibliothekare firm oder sogar involviert. Die Kompetenzen sind daher nicht neu, aber sie erfordern eine Spezialisierung auf Daten. Als Fazit wird deutlich, dass Bibliothekare nicht nur geeignet, sondern geradezu dazu prädestiniert sind, um mit Forschungsdaten zu arbeiten.

Basierend auf dieser Erkenntnis war es möglich Aufgabenfelder zusammenzutragen, bei denen eine Beteiligung von Bibliothekaren im Forschungsdatenmanagement wichtig oder sogar notwendig ist. Neben der Analyse von Stellenausschreibungen waren insbesondere die Experteninterviews wichtig, um die deutsche Perspektive widerzuspiegeln.

Um eine strukturierte Zusammenstellung und Zuordnung der Aufgaben zu erhalten, wurde sich am Curation Lifecycle Model des Digital Curation Centres orientiert.

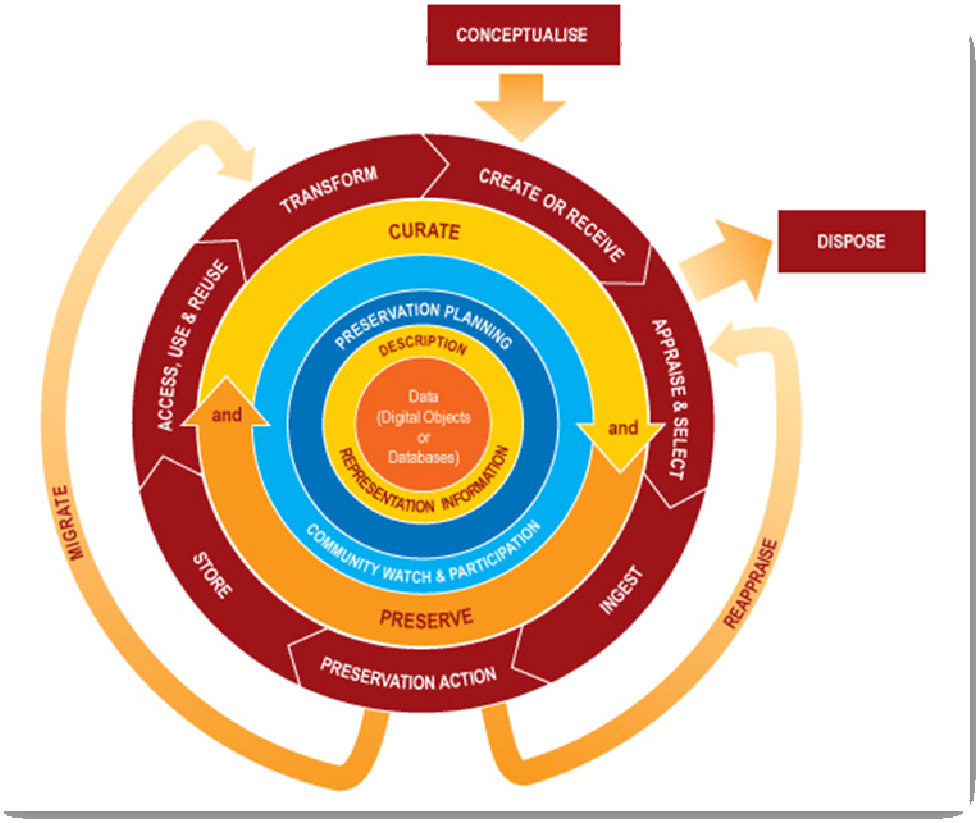


Abbildung 1: Curation Lifecycle Model (Quelle: Digital Curation Centre, 2010)

Folgende Aufgabenfelder für Bibliothekare konnten den verschiedenen Phasen im Curation Lifecycle zugeordnet werden:

Konzeption:
Beteiligung an der Erstellung des Forschungsdatenmanagementplans, Aufklärungsarbeit

Datenerstellung:
Unterstützung bei der Erschließung der Daten

Datenübernahmen:
Identifikation des Bedarfs, Auswahl und Recherche der Daten, Beschaffung und Verhandlungen bezüglich der Datennutzung, Erschließung der erworbenen Daten (diese Arbeit kann jedoch nur erfolgen, wenn Daten sichtbar gemacht wurden)

***Bewertung, Einarbeitung, Aktivitäten der Bewahrung, Speicherung*:**
Unterstützung bei der Auswahl speicherwürdiger Daten, Betreibung von Repositorien (für Dokumente wie auch Daten), Unterstützung bei der Curation, Durchführung von Zufriedenheitsmessungen

Zugang, Nutzung, Wiederverwendung:
Durchführung von Benutzerschulungen und Auskünften, Unterstützung bei der Transformation von Daten

Ein Aufgabenbereich, der nicht direkt im Forschungsdatenmanagement stattfindet, aber relevant ist, um dessen Akzeptanz und Bekanntheit zu steigern, ist das Marketing. Auch in diesem Bereich wurden Bibliothekaren viele Aufgaben zugesprochen. Data Librarians werden häufig in Stellenausschreibungen dazu aufgefordert, an bspw. Kongressen teilzunehmen. Denn durch deren Präsenz auf Veranstaltungen können sie nicht nur für ihre Einrichtung werben, sondern einen Gewinn für das gesamte Forschungsdatenmanagement durch Aufklärung, Darlegung des Nutzens und mögliche Optionen, bspw. wie Daten gespeichert werden, bringen. Eine weitere Aufgabe ist die Vermittlung von Datenkompetenz, um bei Studenten in einem frühen Stadium ihr Bewusstsein für den richtigen Umgang mit Forschungsdaten zu stärken und ihnen Kompetenzen bezüglich des Nutzens von Daten zu vermitteln. Von den befragten Experten wurde besonders die Vermittlerrolle von Bibliothekaren betont, die bspw. als neutrale Instanz zwischen allen Akteuren agieren könnten.

Konsequenzen für die Ausbildung

Es gibt demnach viele Möglichkeiten der Beteiligung von Bibliothekaren im Forschungsdatenmanagement. Was bedeutet dies nun für die Ausbildung? Tatsache ist, dass in den gegenwärtigen Ausbildungsmöglichkeiten, Bachelor, Master und weiterbildender Master, das Thema E-Science bzw. Forschungsdaten nicht explizit behandelt wird. Um Bibliothekare zu motivieren, sich der neuen Rolle anzunehmen, müssen diese aber ein Bewusstsein für Forschungsdaten erhalten. Dies kann durch die Thematisierung im Unterricht erfolgen. Des Weiteren muss die Rolle der datenorientierten Bibliothekare gestärkt werden. Ein Blick in die USA zeigt, dass die Data Librarians kein datenorientiertes Studium erhalten, sondern sie sich das spezifische Know How durch Workshops wie „Providing Social Science Data Services: Strategies for Design and Operation“ (Jacobs, 2010) aneignen. Diese Möglichkeit der Weiterbildung kann ohne weiteres auf die datenorientierten Bibliothekare in Deutschland übertragen werden. Weiterbildungsmöglichkeiten würden Bibliothekare selbst in ihrer Arbeit stärken, aber auch die Sicht auf sie als Akteure im Forschungsdatenmanagement. Eine dritte Option ist die Schaffung eines datenorientierten Masters für Bibliothekare. Diese erscheint jedoch sehr diskussionswürdig. Derzeitig erlangen Bibliothekare grundsätzlich einen konsekutiven Master auf ihren bereits bibliotheksorientierten Bachelor. Somit wäre für einen geschaffenen datenorientierten Master die Studentenzahl sehr begrenzt. Und auch die Frage nach den disziplinspezifischen Kenntnissen kann keine Option eingeräumt werden. Durch Schaffung eines nicht-konsekutiven Masters könnte jeder Interessierte sich zum „datenorientierten Bibliothekar“ ausbilden lassen. Dieses bietet zwar Vorteile, wie einen großen Wissensaustausch zwischen Bibliothekaren und Wissenschaftlern aus verschiedensten Bereichen und höhere Studentenzahlen. Doch ergeben sich bei dieser Aufführung auch parallel Gegenargumente. Ist es möglich und wünschenswert ein Curriculum zu erstellen, dass innerhalb einer Masterperiode sowohl Studenten mit Vorwissen wie auch Laien zu vollwertigen Bibliothekaren ausbildet? Grundlegend besteht dieser Anspruch an einen

nicht-konsekutiven Master (vgl. Arbeitsgemeinschaft Bildung für Deutschland, 2009), doch inwieweit dies umgesetzt werden kann, bedarf noch weiterer Diskussionen.

Gegenwärtiger Einsatz von Bibliothekaren im Forschungsdatenmanagement

Bei der Recherche nach datenorientierten Bibliothekaren in Deutschland ist das Ergebnis schnell dargelegt. Es gibt lediglich einen Datenbibliothekar, Hannes Grobe von Pangaea (vgl. AWI, 2010). Darüber hinaus wurde eine Stellenausschreibung der Universitätsbibliothek Bielefeld identifiziert. In dieser wird ein wissenschaftlicher Mitarbeiter für die Entwicklung ihrer Forschungsdatenservices mit einem Abschluss aus der Informatik oder den Informationswissenschaften gesucht. Damit wird klar, dass Bibliothekare noch nicht so aktiv im Forschungsdatenmanagement arbeiten, wie bspw. in den USA. Ein Ergebnis, das auf Grund der noch anhaltenden Entwicklungen von E-Science absehbar war. Die befragten deutschen Experten hatten dazu differenzierte Meinungen. Manche vertraten den Standpunkt, dass es durchaus Bibliothekare gibt, die schon im Forschungsdatenmanagement arbeiten. Andere wiederum sehen den gegenwärtigen Einsatz von Bibliothekaren eher in der Erstellung von Konzepten und Lösungen zu Umsetzung des Forschungsdatenmanagements. Auch ist erkennbar, dass durch die Stellenausschreibung nicht nur Bibliothekare zum Bewerben aufgefordert wurden, sondern auch Informatiker. Mit anderen Worten, die Aufgaben könnten auch von anderen erfüllt werden. Wer zuerst kommt, wird die Aufgaben übernehmen – First come – first serve.

„Es ist die Zeit, in der Bibliothekare sich einbringen und aktiv werden sollten. Wird der richtige Zeitpunkt verpasst, werden andere Akteure ihre Aufgaben übernehmen und die Chance von Bibliothekaren verstreicht.“ (Rümpel, 2010, S. 70)

Bibliothekare müssen somit jetzt den Mut und Willen aufbringen diese Aufgaben anzunehmen. Gab es während der Bearbeitungszeit der Diplomarbeit nur eine Stellenausschreibung, konnten mittlerweile vier weitere innerhalb eines Monates verzeichnet werden. Der Zeitpunkt ist also gekommen. Aber nicht nur für Bibliothekare, sondern auch für Entscheidungsträger die Chance für Bibliotheken zu erkennen und Angebote zu schaffen, wodurch Bibliothekare gestärkt werden.

Literatur

Arbeitsgemeinschaft Bildung für Deutschland (2009). Bachelor-Studium.net: Master-programme; Altbewährtes oder Neuorientierung im Masterstudium. Köln. http://www.bachelor-studium.net.

Dallmeier-Tiessen, S. (2010). Forschungsdaten 2010: Relevanz, Positionen und Akteure. Bibliothekskongress 2010. Leipzig. http://www.opus-bayern.de/bib-info/volltexte/2010/841.

Deutsche Forschungsgemeinschaft (1998). Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft". Weinheim. http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf.

Digital Curation Center (2010). DDC: because good research needs good data: DDC Digital Curation Lifecycle Model. Edinburgh. http://www.dcc.ac.uk/resources/curation-lifecycle-model.

Gold, A. (2007). Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians. In D-Lib Magazine, 13(9/10). http://www.dlib.org/dlib/september07/gold/09gold-pt1.html.

Jacobs, J. (2010). IASSIST Workshop: Providing Social Science Data Services: Strategies for Design and Operation. Michigan. http://www.iassistdata.org/blog/workshop-providing-social-science-data-services-strategies-design-and-operation.

Plassmann, E., Rösch, H., Seefeldt, J., & Umlauf, K. (2006). Bibliotheken und Informationsgesellschaft in Deutschland: Eine Einführung. Wiesbaden: Harrassowitz.

Rümpel, S. (2010). Data Librarianship – Anforderungen an Bibliothekare im Forschungsdatenmanagement. Diplomarbeit, Fachhochschule Potsdam, Potsdam.

Stiftung Alfred-Wegener-Institut für Polar- und Meeresforschung (AWI) (2010). AWI. Alfred-Wegener-Institut für Polar- und Meeresforschung: AWI Staff Members. Bremerhaven. http://www.awi.de/en/home/.

Usability-Probleme bei der Nutzung bibliothekarischer Webangebote und ihre Ursachen

Malgorzata Dynkowska
Justus-Liebig-Universität Gießen
Zentrum für Medien und Interaktivität
Ludwigstr. 34
D-35390 Gießen
Malgorzata.K.Dynkowska@zmi.uni-giessen.de

Zusammenfassung

Die Beschäftigung mit der Qualität von Webangeboten im Kontext der Web-Usability-Forschung gewinnt seit einigen Jahren immer mehr Aufmerksamkeit, nicht zuletzt aufgrund der gesellschaftlichen Relevanz dieses Stranges der Qualitätsforschung. Das Interesse an und die Nachfrage nach benutzerfreundlichen Webangeboten in Zeiten des „medialen Umbruchs“ und der Verlagerung vieler Kommunikationsprozesse und -aufgaben auf das Internet wächst kontinuierlich. Dabei wird der Anspruch der Benutzerfreundlichkeit nicht nur an kommerzielle, sondern auch zunehmend an öffentliche Webangebote gestellt. In dem Beitrag werden aktuelle empirische Erkenntnisse zur Nutzung moderner Bibliothekswebangebote präsentiert, die aus einer linguistisch motivierten Usability-Untersuchung hervorgegangen sind.

Einführung

„Bibliotheken sind Orte des Zugangs zu Wissen in vielfältiger Form“ (Schneider, 2010, 165). Eine der Zugangsformen sind Bibliothekswebangebote. Webangebote wissenschaftlicher Bibliotheken stellen dabei einen sehr spezifischen, hochkomplexen Typ dar: Sie integrieren heute „eindrucksvolle Informationsschätze“ – sowohl qualitativ als auch quantitativ gesehen – und stellen sie einer breiten Nutzerschaft zur Verfügung (Fritz in Dynkowska, 2006, 2). Qualitativ noch so hochwertige „wissenschaftliche Informationen“ nutzen jedoch nichts, wenn sie nicht (auf-)gefunden und nicht genutzt werden. Der vorliegende Beitrag präsentiert einige zentrale Befunde zur Usability (dt. zumeist *Benutzerfreundlichkeit*)[1] von Bibliothekswebangeboten, die aus einer umfangreichen Usability-Studie zum Gießener Webangebot des Bibliothekssystems hervorgegangen sind. Die Besonderheit der Studie lag in der methodischen Erweiterung der Untersuchungskategorien der gängigen Usabilityforschung (z. B. Nielsen, 2001) mit Mitteln einer funktional-dynamischen Texttheorie, der neuen Hypertextlinguistik und der sprachwissenschaftlich orientierten Rezeptionsforschung

[1] Für den Terminus technicus ‚Usability‘ werden im deutschsprachigen Raum in der Regel Ausdrücke Gebrauchstauglichkeit, Nutzbarkeit, Brauchbarkeit und (Be-)Nutzerfreundlichkeit verwendet. In dem Beitrag wird der Ausdruck (Web-)Usability bevorzugt, da zum einen die Bezeichnungen nicht als Synonyme für Usability verstanden werden können und zum anderen Ausdrücke wie Brauchbarkeit zum Teil andere theoretische Ansätze benennen (etwa Brauchbarkeitsforschung in der Linguistik).

(ausführliche Darstellung siehe Dynkowska, 2010).[2] Die Untersuchung des Gießener Bibliothekswebangebots hatte dabei einen exemplarischen Charakter.

Usability-Evaluation des Gießener Bibliothekswebangebots: linguistisch motivierter Mehrmethodenansatz

Die Notwendigkeit und Relevanz einer umfassenden Beschäftigung mit dem Thema der Usability von Bibliothekswebangeboten wurde in bibliothekarischen Fachkreisen in Deutschland zwar weitgehend erkannt und auch zunehmend gefordert (u. a. Wissenschaftsrat, 2001; siehe auch Homann, 2002; Schulz, 2002; Bieselin, 2004), systematische empirische Untersuchungen befinden sich jedoch noch in den Anfängen. Usability-Evaluationen von wissenschaftlichen oder öffentlichen Bibliothekswebangeboten werden häufig im Rahmen akademischer Abschlussarbeiten durchgeführt oder konzentrieren sich nur auf Teilaspekte des umfangreichen elektronischen Angebots von Bibliotheken, zum Beispiel auf die Benutzerfreundlichkeit von Online-Katalogen. Dabei werden zum einen unterschiedliche Methoden eingesetzt und zum anderen zum Teil divergierende Qualitätskriterien bei der Bewertung herangezogen, so dass die jeweiligen Ergebnisse nur bedingt vergleichbar sind. Im Kontext bibliothekarischer Benutzerforschung durchgeführte Untersuchungen sind darüber hinaus hauptsächlich auf Erhebungen zum Bekanntheitsgrad, zur Nutzungshäufigkeit und zur Zufriedenheit mit dem Webangebot oder seinen Teilangeboten ausgerichtet. Methodisch beschränken sie sich auf den Einsatz und die Auswertung von Fragebögen, mit denen Bibliotheksnutzer schriftlich oder online befragt werden.

Neuere Arbeiten zeigen einen unmittelbaren Handlungsbedarf auf: Die für Forschung und Lehre speziell entwickelten Anwendungen werden von Studierenden „oft schwer überblickt und suboptimal in das Studium integriert […] – hoch differenzierte und hochleistungsfähige Instrumente werden zuweilen mit den relativ einfachen Bedürfnissen der Informationssuchenden konfrontiert“ (Schröter, 2006, 1287; vgl. Klatt & Gavriilidis et al., 2001). Bibliotheksnutzer scheitern bei der Nutzung von komplexen Bibliothekswebangeboten nicht selten. Was bedeutet das aber genau? Wie werden Webangebote von Universitäts- bzw. Hochschulbibliotheken tatsächlich genutzt? Wie konstruieren Nutzer Schritt für Schritt einen für sie zusammenhängenden Pfad durch das Webangebot? Wie verstehen sie die einzelnen Textbausteine? Welche Probleme haben sie und wo liegen die Ursachen einer erfolglosen Nutzung? Um diese Fragen zufriedenstellend zu beantworten, wurden in der Gießener Usability-Studie unterschiedliche Methoden in einem differenzierten qualitativen Untersuchungsdesign integriert (vgl. Richter, 2002; Bucher, 2001). Abbildung 1 vermittelt einen grafischen Überblick über den verfolgten Mehrmethodenansatz.

[2] Die Usability-Studie – und zugleich eine Dissertation – entstand im Rahmen des Projekts „Web-Usability des Informations- und Interaktionsangebots von Hochschulbibliotheken“, das in den Jahren 2004 bis 2006 an der Universität Gießen durchgeführt wurde.

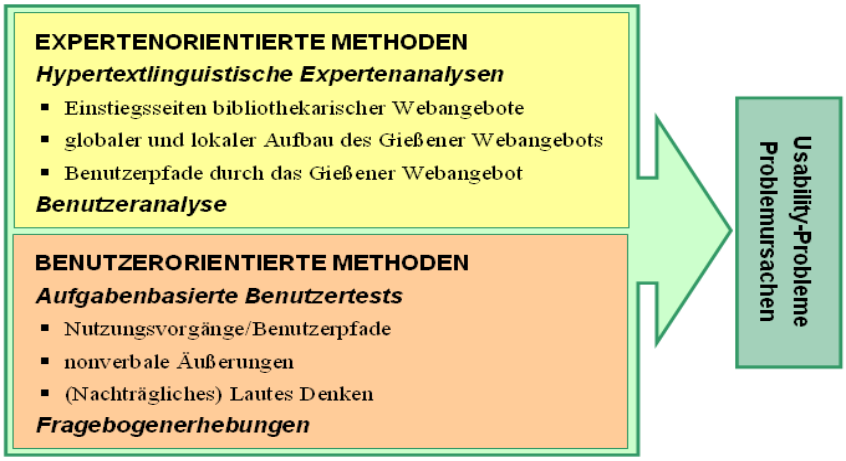


Abbildung 1: Mehrmethodenansatz zur Usability-Evaluation

Die Studie bediente sich sowohl expertenorientierter als auch benutzerorientierter Methoden zur Usability-Evaluation. Sie schloss eine hypertextlinguistische Expertenanalyse des globalen Aufbaus des Gießener Webangebots sowie eine Expertenanalyse der in sogenannten Benutzertests empirisch erhobenen tatsächlichen Benutzerpfade durch dieses Webangebot ein. Benutzertests bzw. Usability-Tests werden als „Mittel der Wahl“ bezeichnet (Sarodnick & Brau, 2006, 188), wenn es um die Aufdeckung von Nutzungsproblemen realer Nutzer geht. In 44 Benutzertests wurden Erhebungen der Navigationshandlungen der Tester im untersuchten Webangebot, der nonverbalen Äußerungen (Gestik, Mimik) und

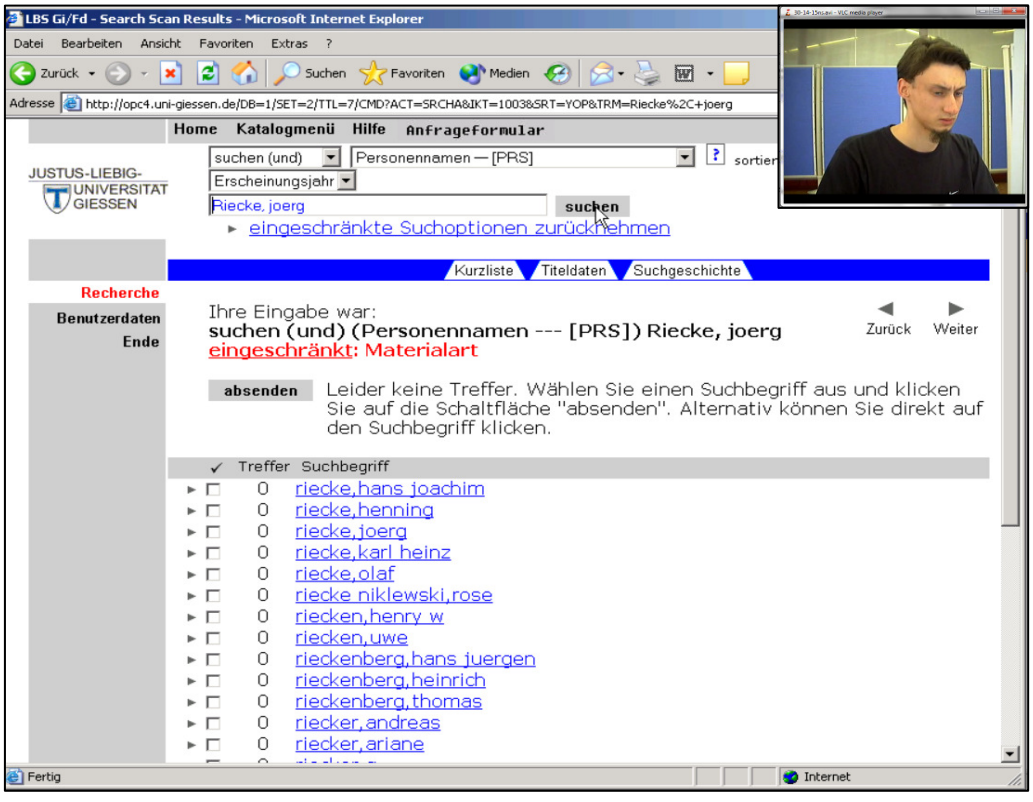


Abbildung 2: Analysevideo eines Benutzertests

spontanen Benutzerkommentare (sog. Lautes Denken) vorgenommen. Die gewonnenen Daten wurden dabei zu integrierten, mehrkanaligen Analysevideos zusammengefasst (siehe Abb. 2). Vor den Tests wurde eine Pretest-Befragung mittels Fragebogen (u. a. soziodemografische Daten und Einschätzung der relevanten Wissensbestände) und nach den Tests eine retrospektive Befragung in problemzentrierten Interviews durchgeführt. Allen Benutzertests lag ein festgelegtes, benutzer- und webangebotsspezifisches Testszenario zugrunde. Flankiert wurde die Untersuchung durch eine Expertenanalyse der Einstiegsseiten mehrerer Bibliothekswebangebote sowie durch eine Benutzeranalyse.

Das Kernstück der Untersuchung bildete die Analyse der dokumentierten individuellen Benutzerpfade durch das Gießener Bibliothekswebangebot mithilfe des auf die spezifische Rezeptionssituation von Hypertexten angepassten Analyseinstrumentariums der funktionaldynamischen Texttheorie (u. a. Fritz, 1994, 1999; vgl. Bucher, 2004). Dieser Beschreibungsapparat erlaubt nicht nur eine hohe Trennschärfe bei der Analyse von konkreten Nutzungsschwierigkeiten. Er ermöglicht vor allem eine differenzierte Betrachtung der bisher in Usability-Studien nur pauschal und häufig unzureichend berücksichtigten sprachlichen Form (Syntax, Wortschatz), der textlichen und gestalterischen Realisierung (lokale und globale Sequenzierung, thematische Struktur, Wissensaufbau), der kommunikativen Handlungsmuster sowie der Adressatenorientiertheit in den Webangeboten.

Zentrale Befunde

Das wesentliche Ziel der durchgeführten Analysen war die Identifizierung der Faktoren problematischer bzw. missglückter Nutzungsvorgänge. Die Befunde wurden in Form von zwei Systematisierungen dargestellt: einer Typologie von potenziellen Usability-Problemen und einer Systematisierung von Problemursachen bei der Nutzung von Bibliothekswebangeboten. Usability-Probleme wurden dabei als Probleme bei der individuellen Konstruktion von zusammenhängenden, sinnvollen Benutzerpfaden aufgefasst und folgende sechs Problemtypen differenziert (ausführlich siehe Dynkowska, 2010, 277ff.):

Einstiegsprobleme,

Entscheidungsprobleme,

Fortsetzungsprobleme,

Orientierungsprobleme,

Einordnungsprobleme und

Sackgassenprobleme.

Sackgassenprobleme treten unter anderem dann auf, wenn bestimmte (sinnvolle) Schritte auf dem Benutzerpfad – in der Regel vom System – verhindert werden. Abbildung 3 zeigt ein Beispiel für eine Sackgasse nach dem Klick auf den Link *Abkürzungsverzeichnisse*.

Die eruierten Probleme treten nicht unabhängig voneinander auf, sie können sich auch gegenseitig bedingen. Sie kommen häufig nicht nur einzeln, sondern auch in Problem-Ketten vor. Es lässt sich ferner kein schematischer Zusammenhang zwischen den Problemen und bestimmten Nutzertypen aufdecken, allenfalls kann man gewisse Tendenzen erkennen. So

haben Bibliotheksnovizen tendenziell viel häufiger Einstiegsprobleme als Bibliotheksexperten. Entscheidungs- und Fortsetzungsprobleme lassen sich dagegen häufiger bei den Nutzern beobachten, die über verhältnismäßig wenig Interneterfahrung und -wissen verfügen. Sackgassenprobleme sind insbesondere durch Systemmängel bedingt, daher kann ihr Auftreten grundsätzlich nicht in Verbindung mit bestimmten (mangelnden) Wissensbeständen gebracht werden. Im Hinblick auf die Sackgassenprobleme kann man jedoch davon ausgehen, dass Poweruser, die zugleich Bibliotheksexperten sind, schneller einen Ausweg aus einer virtuellen Sackgasse finden können.

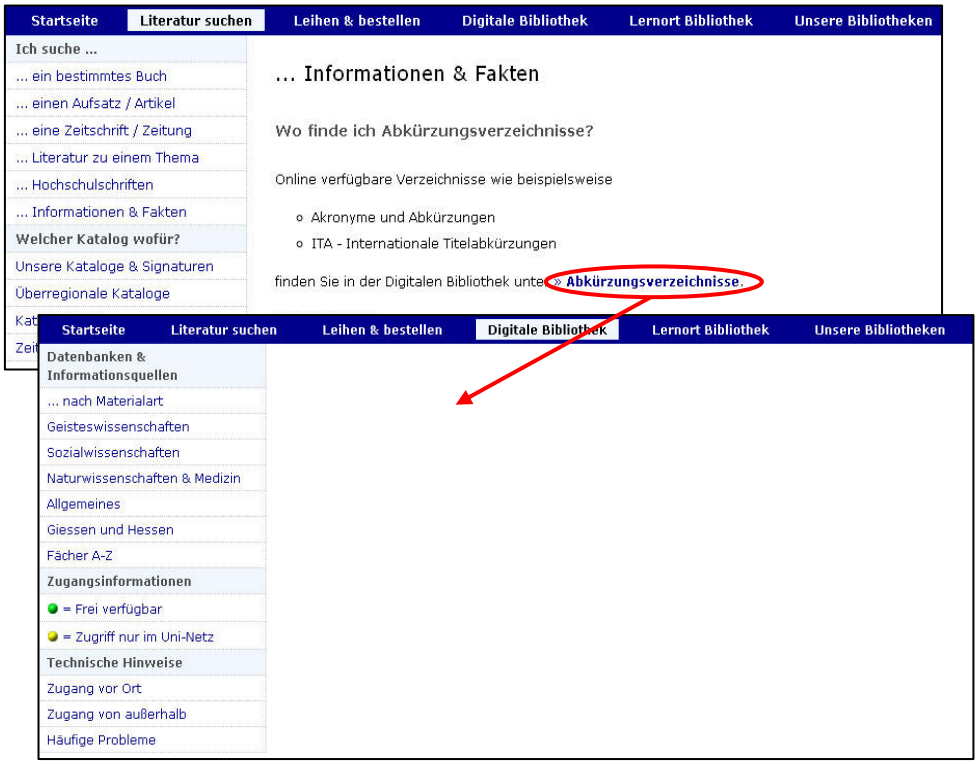


Abbildung 3: Sackgasse nach dem Klick auf den Link Abkürzungsverzeichnisse

Ursachen für die Probleme bei der Nutzung von bibliothekarischen Webangeboten lassen sich zunächst generell als Barrieren beschreiben, die den nächsten idealtypischen oder einen sinnvollen Schritt[3] auf dem Benutzerpfad – bei der Erreichung eines bestimmten (Teil-)Ziels – behindern oder verhindern. Sie bilden in der Regel ein Geflecht aus korrelierenden Faktoren, die zusammenwirken, wie zum Beispiel Wissensstand des Nutzers bei einem bestimmten Nutzungsschritt in Verbindung mit einem bestimmten Linkausdruck und der Stellung eines Textbausteins in der Nutzungssequenz. Die Problemursachen sind darüber hinaus vielfältig und heterogen. Wissensdefizite der Nutzer, zum Beispiel fehlendes und für die Nutzung von Bibliothekswebangeboten relevantes Handlungswissen über den Ablauf von Ausleihvorgängen, stellen eine andere Ursachenart dar als unverständliche Icons (siehe die Ursachenbereiche „Benutzerwissen“ und „sprachliche und bildliche Zeichen“ in Abb. 4).

[3] Zum Instrument eines sogenannten idealtypischen Pfades, das in der Analyse der Benutzerpfade angewendet wurde, siehe Dynkowska, 2010, 199ff.

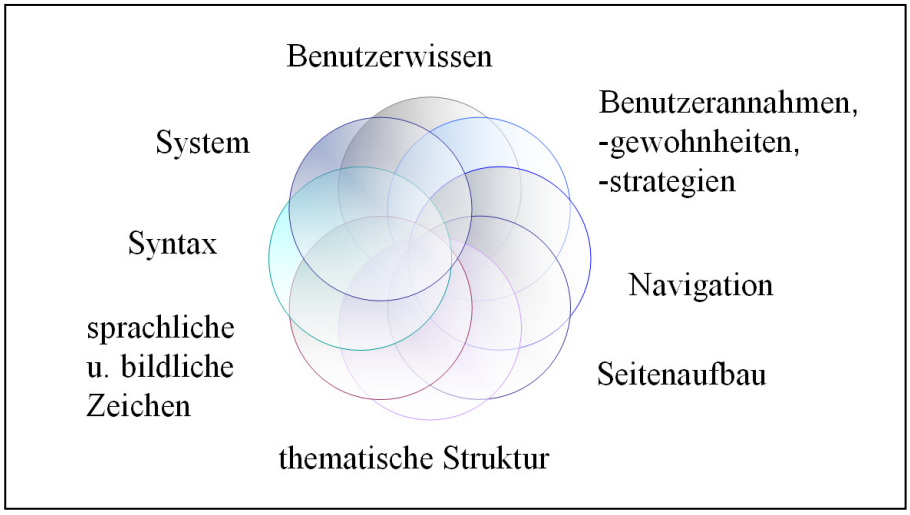


Abbildung 4: Ursachenbereiche für Usability-Probleme

Die Ursachen können grundsätzlich unterschiedliche Trag- und Reichweiten haben: Einige haben tendenziell eine eher lokale Auswirkung (z. B. unverständliche Icons), andere beeinträchtigen deutlich den globalen Aufbau von Benutzerpfaden (z. B. unzutreffende Annahmen über Nutzungsmöglichkeiten des Webangebots oder inkonsistente Navigationsstruktur). Die tatsächliche Tragweite einer bestimmten Ursache kann jedoch nur jeweils in einem bestimmten sequenziellen lokalen und globalen Zusammenhang bestimmt werden, d. h. häufig anhand des Zusammenspiels von mehreren Ursachen und den Nutzungsproblemen bzw. Problem-Ketten. In Abbildung 4 werden die insgesamt acht identifizierten Ursachenbereiche dargestellt.

Eine(r) der bedeutendsten Ursachen(-typen) für Usability-Probleme sind Wissensdefizite der Nutzer insbesondere im Bereich des Bibliothekswissens (siehe Bereich „Benutzerwissen" in Abb. 4). Was die Ausdrücke *Fernleihe*, *Freihandbereich* oder *bibliografische Datenbanken* bedeuten, wissen die Nutzer häufig nicht. Weitere wichtige Problemursachen sind die so genannte Amazon- und die so genannte Google-Strategie, d. h. die Übertragung von bestimmten, für die Nutzer sinnvollen Handlungsabfolgen bzw. Handlungsmustern aus dem Umgang mit der Suchmaschine Google und dem Online-Shop Amazon auf den Umgang mit den (im Vergleich selten genutzten) Bibliothekswebangeboten. Auf weitere Ursachentypen kann hier nicht eingegangen werden; zu detaillierten Ergebnissen siehe Dynkowska, 2010, Kapitel 6.3.

Fazit und Ausblick

Die in diesem Beitrag kurz skizzierten empirischen Erkenntnisse zur Nutzung moderner Bibliothekswebangebote erweitern nicht nur die Kenntnis einschlägiger Usability-Probleme, sondern verdeutlichen auch die Komplexität einer Usability-Optimierung. Dabei haben beide Systematisierungen, die Typologie der Usability-Probleme und die der Problemursachen, insofern eine praktische Relevanz, als sie für konkrete Aufgaben bei Usability-Evaluationen nutzbar gemacht werden können. Insbesondere die Formulierung von Problemursachen bzw. Ursachenbereichen liefert Ansatzpunkte für konkrete Optimierungen von Bibliothekswebangeboten. Wie sich von den Ursachenbereichen (Abb. 4) eindeutige und auf die Web-

angebote der Universitätsbibliotheken zugeschnittene Empfehlungen zur benutzerfreundlichen Gestaltung ableiten lassen, wurde exemplarisch in Dynkowska 2006 gezeigt. Die explorative Durchsicht aktueller Webangebote wissenschaftlicher Bibliotheken zeigt darüber hinaus, dass zwar deutliche Fortschritte im Hinblick auf das Qualitätskriterium der Usability zu verzeichnen sind, die differenzierten Usability-Probleme und deren Ursachen jedoch keine einmaligen Phänomene darstellen.

Die Optimierungspraxis in bibliothekarischen Webredaktionen, zum Beispiel im Rahmen von größeren Relaunch-Projekten, ist mit unterschiedlichen Schwierigkeiten verbunden, die in der einschlägigen Literatur bisher nur relativ knapp oder gar nicht thematisiert wurden. Sie haben eine unmittelbare Auswirkung auf die optimierten Webangebote insofern, als sie – unbeabsichtigt – alte Unzulänglichkeiten oder Mängel verlagern oder gar neue herbeiführen können. So führen häufig diverse organisatorische, technische und redaktionelle bzw. personelle Engpässe, die den sehr komplexen Prozess der Optimierung begleiten, zu Einschränkungen und Kürzungen bei den durchzuführenden Optimierungsaufgaben. Diese Faktoren sollten in übergreifenden Optimierungskonzepten, die noch weitgehend ausstehen, berücksichtigt werden. Die Erstellung derartiger Konzepte und ferner eine entsprechende Integration derselben in bestehende Verfahren zur Qualitätssicherung redaktioneller und bibliothekarischer Arbeit stellt nicht zuletzt vor dem Hintergrund der fortdauernden Entwicklung der Internettechnologie und der damit verbundenen Entwicklung der Gewohnheiten, Erwartungen, Interessen und Bedürfnissen der Web- und Bibliotheksnutzer eine Herausforderung dar. Dieser vielschichtigen Entwicklungsdynamik muss Rechnung getragen werden, indem gewonnene Erkenntnisse über Anforderungen an die Usability bibliothekarischer Webangebote systematisch validiert werden.

Literatur

Bieselin, T.-B. (2004). Nur Dabeisein ist nicht alles: Anregungen für Bibliotheks-Websites in der Praxis. *Information, Wissenschaft & Praxis*, 55 (4), 230–233.

Bucher, H.-J. (2001). Wie interaktiv sind die neuen Medien? Grundlagen einer Theorie der Rezeption nicht-linearer Medien. In H.-J. Bucher & U. Püschel (Hrsg.), *Die Zeitung zwischen Print und Digitalisierung*. Wiesbaden: Westdteutscher Verlag, 139–171.

Bucher, H.-J. (2004). Online-Interaktivität – Ein hybrider Begriff für eine hybride Kommunikationsform. In C. Bieber & C. Leggewie (Hrsg.), *Interaktivität: ein transdisziplinärer Schlüsselbegriff*. Frankfurt am Main: Campus, 132–167.

Dynkowska, M. (2006). *Gestaltung von benutzerfreundlichen Online-Angeboten wissenschaftlicher Bibliotheken: Empfehlungskatalog*. Justus-Liebig-Universität Gießen. Online: http://www.uni-giessen.de/usability/downloads/Empfehlungskatalog_.pdf [20.12.2010].

Dynkowska, M. (2010). *Web-Usability aus linguistischer Sicht am Beispiel von bibliothekarischen Webangeboten*. Gießener Elektronische Bibliothek (= Linguistische Untersuchungen, Band 2). Online: http://geb.uni-giessen.de/geb/volltexte/2010/7910/ [17.12.2010].

Fritz, G. (1994). Grundlagen der Dialogorganisation. In G. Fritz & F. Hundsnurscher (Hrsg.), *Handbuch der Dialoganalyse*. Tübingen: Niemeyer, 177–201.

Fritz, G. (1999). Coherence in hypertext. In W. Bublitz, U. Lenk & E. Ventola (Hrsg.), *Coherence in spoken and written discourse: How to create it and how to describe it*. Amsterdam: Benjamins, 221–232.

Homann, B. (2002). Bibliothekarische Web-Sites: 1. Defizite bibliothekarischer Websites. Ergebnisse eines Round Tables. *Bibliotheksdienst*, 36 (10), 1323–1335.

Klatt, R. & Gavriilidis, K. et al. (2001). *Nutzung elektronischer wissenschaftlicher Information in der Hochschulausbildung. Barrieren und Potenziale in der innovativen Mediennutzung im Lernalltag der Hochschulen*. Endbericht, Kurzfassung, Fragebögen und Tabellenband. Online: http://www.stefi.de [13.11.2010].

Nielsen, J. (2001). *Designing Web Usability*. München: Markt+Technik.

Richter, G. (2002). Web-Usability. Oder: Wie man die Benutzbarkeit von Web-Seiten untersuchen kann. *Spiegel der Forschung*, 19 (2), 75–78.

Sarodnick, F. & Brau, H. (2006). *Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendung*. Bern: Huber.

Schneider, U. J. (2010). Über die Rolle von Bibliotheken in einer digitalisierten Welt. In H. Burda, M. Döpfner, B. Hombach & J. Rüttgers (Hrsg.), *2020 – Gedanken zur Zukunft des Internets*. Essen: Klartext, 165–170.

Schröter, M. (2006). Fünf Jahre nach SteFi oder: Auf der Suche nach Informationskompetenz im Studienalltag. „Von der ‚Ware' Information zur ‚wahren' Information – Erstellen einer Fachinformationsseite Geschichte von Studierenden für Studierende". *Bibliotheksdienst*, 40 (11), 1286–1295.

Schulz, U. (2002). „Das stiehlt meine Zeit." Über die Nutzungsqualität von Bibliothekswebsites. *BuB*, 54 (4), 224–228.

Wissenschaftsrat (2001). *Empfehlungen zur digitalen Informationsversorgung durch Hochschulbibliotheken* (Drs. 4935/01). Online: http://www.wissenschaftsrat.de/download/archiv/4935-01.pdf [13.11.2010].

E-Science-Interfaces – ein Forschungsentwurf

Sonja Palfner

Technische Universität Berlin
Zentrum Technik und Gesellschaft
Sekretariat WF1
Hardenbergstraße 36 A
D-10623 Berlin
sonja.palfner@tu-berlin.de

Zusammenfassung

Die Ordnung der Wissenschaft unterscheidet Forschung von Service. Diese Ordnung ist nicht natürlich – sie wird hergestellt. Kommt es im Zuge von Wissenschaftsentwicklungen, die unter dem Begriff E-Science aktuell stattfinden, zu einer Neu-Ordnung der Wissenschaft oder bleibt alles beim Alten? Lösen sich beispielsweise inhaltliche oder institutionelle Grenzen auf, weil sich Service- und Forschungspraktiken vermischen? In diesem Beitrag wird die Frage nach der Ordnung der Wissenschaft in Forschung und Service behandelt und dabei werden institutionelle und technische Interfaces als neuralgische Orte des Ordnung-Schaffens im E-Science-Kontext definiert.

Einführung

E-Science, Digitale Wissenschaften, E-Infrastrukturen – es tut sich etwas in den Wissenschaften und zwar über Fächergrenzen hinweg. Das Versprechen auf eine neue Wissenschaft liegt in der Luft: „The outcome from such an undertaking is clear and substantial: new, faster, better and different science than has been possible before“ (Coveney & Atkinson, 2009). Diese radikale Vision wird flankiert durch Äußerungen, die stärker den unterstützenden und ergänzenden – und weniger den transformierenden – Charakter neuer Technologien hervorheben. Hier soll die Technologie eine stille Begleiterin sein, die den Forschenden „nicht mit Unzulänglichkeiten oder Problemen auf der Ebene der Basisinfrastruktur belästigt“ (Neuroth, Aschenbrenner & Lohmeier, 2007). Ein wohlbekanntes Bild von Wissenschaft taucht auf: Technologien prägen zwar den Möglichkeitsraum wissenschaftlicher Praxis, ja sie besitzen sogar das Potenzial ihn zu revolutionieren, und dennoch ist der Wissenschaftler scheinbar geschützt vor diesem Objektbereich beziehungsweise gilt es, ihn nicht mit diesem zu belästigen. Der Forschende ist und bleibt in diesem Bild das handelnde Subjekt welches sich der (neuen) E-Infrastrukturen bedient und mit Hilfe dieser neues Wissen generieren kann. Zwar werden durchaus Überlappungen der verschiedenen Akteursgruppen (wissenschaftliche Nutzer, Inhalteanbieter und Software-Entwickler) konstatiert, aber die Existenz dieser Gruppen inklusive der Zuschreibungen von Handlungsfeldern bleibt unhinterfragt. Zwar wird Diensten, Daten und Nutzern zugesprochen, „gleichermaßen essentielle Teile“ (Küster, Ludwig & Aschenbrenner, 2009) eines „Digitalen Ökosystems“ (ebd.) zu sein, aber die Ordnung der Wissenschaft zwischen Forschung (E-Science)

und Service (E-Infrastruktur), zwischen Subjekten (Menschen) und Objekten (Daten) scheint, folgt man Publikationen zum Thema, bestehen zu bleiben.

Die Ordnung der Wissenschaft: Service und Forschung

Was also ändert sich? Ist es „nur“ die Arbeitsweise und damit verbunden die Effizienz und Qualität der Forschung, wie folgender Entwurf nahelegt: „Der Wissenschaftler der Zukunft greift von einem beliebigen Standort auf (s)eine virtuelle Umgebung zu und findet dort Programme, Forschungsdaten und Sekundärquellen […] vor, die er für seine aktuelle Arbeit braucht“ (Neuroth et al., 2007)? Oder ist die Ordnung der Wissenschaft selbst betroffen? Finden Neuverhandlungen oder Grenzauflösungen zwischen Forschung und Service statt? In der Literatur aus dem Kontext der E-Sciences (und ich beziehe mich in diesem Artikel ausschließlich auf den Bereich der Digital Humanities) ist von Grenzerosionen nicht oft die Rede, doch finden sich durchaus Hinweise, wie bspw. diesen: „Die Verzahnung zwischen Wissenschaft/Forschung und Infrastrukturen hat in diesem Zusammenhang [hier ist die Rede von E-Science, S.P.] einen neuen Grad erreicht. Eine wichtige Rolle spielen dabei Virtuelle Forschungsumgebungen […] deren Ziel es ist, für Wissenschaftler in allen Bereichen Bedingungen zu schaffen, die ihre Forschungsprozesse maximal unterstützen“ (Jannidis, Lohmeier, Neuroth & Rapp, 2009). Verzahnung ist nicht gleichbedeutend mit Erosion – aber wie bemisst sich dieser „neue Grad“ und wo wird die Grenze zwischen Forschung und Infrastruktur von wem gezogen?

Das Beispiel der Virtuellen Forschungsumgebung kann ergänzt werden durch einen weiteren Bereich, in dem man die Frage nach Grenzauflösung/Hybridisierung und/oder Trennung zwischen Forschung und Service stellen kann: gemeint ist die Wissenschaftsorganisation. In den Geisteswissenschaften wird Bibliotheken, Archiven und Rechenzentren als Service-Institutionen eine besondere Rolle im Aufbau von E-Infrastrukturen für die Wissenschaften zugesprochen (Neuroth et al., 2007; Rapp, 2007). Gleichzeitig wird der Aufbau von Kompetenzzentren gefordert, die dazu beitragen sollen, „innovative Virtuelle Forschungsumgebungen zu entwickeln und damit die Potenziale der computergestützten Forschung in den e-Humanities zu entfalten und auszuschöpfen“ (Jannidis et al., 2009). Sind diese Kompetenzzentren nun gedacht als Serviceeinrichtungen oder als Forschungseinrichtungen? Oder findet in ihnen sowohl Dienstleistung als auch Forschung oder womöglich eine Vermischung statt?

Wissenschaft im Wandel?

Wissenschaft verändert sich. Viel ist in den letzten Jahren in der interdisziplinären Wissenschafts- und Technikforschung geschrieben worden zum Wandel der Wissenschaft und zwar lange bevor die „E-Konjunktur“ einsetzte (siehe etwa Gibbons, Limoges, Nowotny, Schwartzman, Scott & Trow, 1994; Nowotny, Scott & Gibbons, 2001; Weingart, 2001). Ein zentraler Befund lautet, dass wissenschaftlicher Wandel als Erosion von Grenzen oder als Entdifferenzierung von gesellschaftlichen Teilbereichen gefasst werden kann (prominent die Grenze zwischen Wissenschaft und Öffentlichkeit oder die zwischen den Disziplinen). Eine Lektion, die sich aus diesen Auseinandersetzungen lernen lässt und die auch für die Diskussionen um das „Neue“ der E-Sciences hilfreich sein könnte, ist, dass der Befund der

Erosion von Grenzen nicht unwidersprochen blieb (siehe etwa Krücken, 2003). Ähnliches lässt sich mit Blick auf die Organisation von Wissenschaft sagen. Auch hier treffen Befunde zur Auflösung „alter“ akademischer Strukturen und Befunde zur Trägheit und Wandlungsresistenz der Institution Wissenschaft aufeinander. An anderer Stelle wird die These der Entdifferenzierung gegen die These der Ausdifferenzierung von Teilbereichen gestellt: „Fast jede Organisation kann als hybrid in dem Sinne beschrieben werden, dass Organisationen generell die Leistungserwartungen verschiedener Funktionsbereiche der Gesellschaft wie Wissenschaft, Politik, Wirtschaft oder Recht koordinieren müssen. […] Einer solchen ‚Hybridisierung‘ von Funktionen innerhalb von Organisationen entspricht nun nicht eine Entdifferenzierung der Funktionssysteme der Gesellschaft“ (Halfmann & Schützenmeister, 2009). Hybrid scheint hier keineswegs für den Moment der Vermischung zu stehen, sondern für die Praktiken der Koordination und Kommunikation zwischen Teilsystemen.

Betrachtet man die Literatur aus den E-Sciences, dann scheint vieles dafür zu sprechen, der vorschnellen Rede von der Erosion der Grenzen und der Entdifferenzierung von Funktionsbereichen mit Skepsis zu begegnen. Die Digitale Revolution scheint mit der Überwindung von Trennungen nichts zu tun haben zu wollen. Oder anders formuliert: Die Digitale Wissenschaft lebt in der Grenzziehung. Und doch ist damit die Vermischung nicht aus der Welt. Sie deutet sich in solchen Begriffen wie „Verzahnung“ genauso an, wie in Auseinandersetzungen, in denen unerwartet heftig auf der Trennung von Forschung und Service bestanden wird (als Beispiel hierfür kann ich das Deutsche Klimarechenzentrum anführen, welches als „Serviceeinrichtung“ gegründet wurde und gleichzeitig von Beginn an ein Ort war, an dem sich Service und Forschung mischten).

Wissenschaft ist eine soziale Praxis

Aus der Perspektive der Wissenschafts- und Technikforschung (ich nenne hier nur als Beispiel den Ansatz der Akteur-Netzwerk-Theorie; siehe etwa Law, 2006) wird nun genau das Ziehen der Grenze, also der Prozess verstanden als soziale Praxis, interessant. Wie kommt die Möglichkeit beispielsweise zwischen Forschung und Service zu unterscheiden, überhaupt zustande? Diese Fragedimension basiert auf der Annahme, dass jede Unterscheidung und die damit verbundenen Zuschreibungen und Verteilungen von Rollen und Aufgaben erklärungsbedürftig sind. Aufgabe ist es somit zu analysieren, welche Mechanismen dazu führen, dass eine solche Trennung vorgenommen werden kann, wie basierend auf dieser Trennung Aufgaben verteilt werden und welche Effekte dies erzeugt; kurz: es geht um „die Analyse des Ordnungskampfes“ (Law, 2006). Denn eines ist sicher: Wissenschaft basiert nicht nur auf Rollenzuschreibungen, sondern diese sind verbunden mit Prozessen der In- und Exklusion. Und mit Zygmunt Bauman gesprochen, fahren Tag für Tag von den Hallen der Wissenschaften zwei Arten von Lastwagen – die einen steuern die Bibliotheken, die anderen die Mülldeponien an (siehe Bauman, 2005). Ein Informatiker, der E-Infrastrukturentwicklung für die Klimaforschung betreibt, schilderte mir folgendes Problem: Er kann hierzu weder im Bereich der Informatik gut publizieren („Das ist keine Informatik“) noch im Bereich der Klimaforschung („Das ist keine Klimaforschung“). Seine Arbeit ist weder das eine noch ist sie das andere und doch ist sie beides; aber für diese Mitte/Mischung scheint es

keinen sicheren Ort zu geben. Wie verteilt wird und wer durch das Raster fällt, ist eine soziale Frage.

Das Bild von einem Kampf deutet darauf hin, dass Ordnung kein fixer Zustand ist und dass ein Umschlagen in Unordnung durchaus möglich ist. Damit sind wir wieder bei der Unterscheidung zwischen Vermischung/Erosion der Grenze und Säuberung/Trennung. Ich möchte an dieser Stelle Bruno Latours Überlegungen aus seinem Buch „Wir sind nie modern gewesen“ (2002) einfügen. Er beschreibt hier zwei verschiedene Ensembles von Praktiken: Übersetzung und Reinigung. Übersetzungspraktiken sind solche, die Hybride zwischen Natur und Kultur hervorbringen. Durch Reinigung dagegen werden getrennte Zonen von Menschen auf der einen Seite und nicht-menschlichen Wesen auf der anderen Seite geschaffen. Seiner Meinung nach sind diese Ensembles nicht voneinander zu trennen; vielmehr schafft die Reinigung erst die Hybride: „Je mehr man sich verbietet, die Hybride zu denken, desto mehr wird ihre Kreuzung möglich“ (Latour, 2002). Ihm geht es also gerade nicht darum, sich für eine Seite zu entscheiden, sondern Erosion und Trennung als zwei Seiten einer Medaille zu begreifen, die sich produktiv aufeinander beziehen.

Bezogen auf die E-Infrastrukturentwicklung in den Wissenschaften könnte dieser Gedanke helfen, die Vermischung, Entdifferenzierung oder Erosion der Grenze als Effekt der Trennung zwischen Teilbereichen (Forschung und Service) im Blick zu behalten. Dies – und damit komme ich zur Beschreibung eines Forschungsvorhabens – muss empirisch passieren.

E-Science-Interfaces

Mir scheint, dass es zur Untersuchung von Fragen des Grenzziehens und Vermischens keine besseren Orte gibt als Interfaces. In den Naturwissenschaften findet man den Begriff zur Beschreibung der physikalischen Phasengrenze zweier Zustände eines Mediums (flüssig, gasförmig). Ein Medium kann Verschiedenes sein, je nachdem in welcher Umgebung es ist. Zwischen unterschiedlichen Teilsystemen wiederum wird das Interface zur Bedingung der Möglichkeit für das Gelingen von Kommunikation erklärt. Hier ist es nicht das Medium, welches seinen Ort wechselt, sondern die an der Kommunikation Beteiligten können bleiben, wo sie sind; das Interface dient ihrer Verständigung (bspw. von verschiedenen Computerprogrammen). Kurz: Interfaces sind Grenzräume oder Passagen-Zonen. Sie ermöglichen Übergänge und Trennungen, wobei es eines ihrer wesentlichen Merkmale ist, dass diese Funktionalität den Nutzern verborgen bleibt und nur bei Störungen in das Bewusstsein dringt. Es ist die Aufgabe des Verbindens, welche die Interfaces für Paul Wouters zu einem zentralen Bestandteil der E-Sciences macht: „If we want the humanities to profit from investments in large informational infrastructures such as the grid, developing the right kinds of interfaces is the critical issue. This means interfaces between different humanities disciplines, as well as between humanities, the social sciences, the sciences, and the public at large“ (Wouters, 2009).

Ich möchte die Aufmerksamkeit auf zwei Interfaces richten, die für die Digitalen Wissenschaften relevant sind: Virtuelle Forschungsumgebungen und Institutionen, die Expertise im Zusammenhang mit E-Science bündeln (sollen). Ich habe beide bereits oben kurz angesprochen. Setzt man nun, wie ebenfalls bereits ausgeführt, voraus, dass es keine immer schon vorhandenen Teilsysteme gibt, die lediglich miteinander über ein Interface gekoppelt

werden müssen, dann stellt sich die Frage, wie die Interfaces konstruiert werden und wie sie funktionieren müssen, so dass in und mit ihnen die Teilsysteme als solche entstehen und miteinander kommunizieren können. Ihr Innenleben wird also interessant. Dieses Innenleben wiederum ist – wie der Name schon sagt – lebendig. Sprich, Interfaces sind keine leeren Hüllen, sie sind ebenso von menschlichen und nicht-menschlichen Akteuren bevölkert wie ihre Umgebung und agieren auch in beide Richtungen, nach innen und nach außen. Das bedeutet auch, dass man nicht a priori davon ausgehen sollte, dass sie rein technisch funktionieren. Wolf-Dieter Narr konstatiert, dass Institutionen nur dynamisch zu verstehen sind, auch wenn es genau diese Dynamik ist, die gleichsam verschwindet und die Institution als das „Objektive“ erscheinen lässt (1988). Theodore R. Schatzki versteht ähnlich die Analyse von Organisationen als eine Analyse „on organizations as they happen“ (2006).

Kurzum, Interfaces prägen wissenschaftliche Praxis und geben den Möglichkeitsraum in hohem Maße hierfür vor. Ihr Innenleben und das, was durch sie in der Vermittlung hindurchgeht, müssen zueinander passen.

Das Projekt „E-Science-Interfaces“, das im Herbst 2010 beginnen soll, stellt sich nun die Aufgabe zu analysieren, wie die Ordnung der E-Sciences erzeugt wird und welche Effekte dies auf die Beteiligten hat. Der empirische Ort hierfür sind zwei Interfaces in ihrem Werden (Virtuelle Forschungsumgebungen und Kompetenzzentren) in wiederum zwei unterschiedlichen Wissenschaftskontexten. Der erste Kontext ist das TextGrid, eine E-Infrastruktur in den Geisteswissenschaften, und der zweite Kontext ist das C3-Grid, eine E-Infrastruktur in der Klimaforschung.

Gleichzeitig scheint es mir aber auch wichtig zu sein, der Frage nachzugehen, was mit der Möglichkeit zur Erosion der Grenze, mit der Vermischung oder Entdifferenzierung der Teilsysteme – schlicht, mit der Unordnung – passiert, die ja, laut Latour, gerade ein Ergebnis des Trennens und Reinigens darstellt. Wie sind Erosion und Trennung aufeinander bezogen? Ich vermute, dass für dieses Projekt eine Sprache gefordert ist, die nicht mit der „Begriffsklatsche“ (Narr, 2006) der Wissenschaften voraneilt, sondern ihre Kraft zu Ordnen immer wieder selbstkritisch reflektiert. Ebenso scheint mir ein möglichst teilnehmend beobachtendes Verfahren angebracht, um „soziale Wirklichkeit nicht von außen wahrzunehmen und zu be-, gar abzuurteilen, sondern sie im immanenten Nachvollzug zu verstehen“ (Narr, 2006).

Literaturverzeichnis

Baumann, Z. (2005). Verworfenes Leben. Die Ausgegrenzten der Moderne. Bonn: Hamburger Edition.

Coveney, P. V. & Atkinson, M. P. (2009). Crossing boundaries: computational science, e-Science and global e-Infrastructure. Philosophical Transactions of the *Royal Society*, 367, 2425-2427. http://rsta.royalsocietypublishing.org/content/367/1897/2425.full.pdf+html [download 15.10.2010]

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P. & Trow, M. (1994). The New Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies. London: SAGE.

Jannidis, F., Lohmeier, F., Neuroth, H. & Rapp, A. (2009). Virtuelle Forschungsumgebungen für e-Humanities. Maßnahmen zur optimalen Unterstützung von Forschungsprozessen in den Geisteswissenschaften. Bibliothek. Forschung und Praxis, 33(2), 161-169.

Krücken, G. (2003). Helga Nowotny, Peter Scott, Michael Gibbons: Re-Thinking Science. Knowledge and the Public in an Age of Uncertainty. Polity Press, Oxford 2001. 278 Seiten, ISBN 0-7456-2608-4, £14.99 [Rezension]. die hochschule. journal für wissenschaft und bildung, 1, 237-241. www.hof.uni-halle.de/journal/texte/03_1/dhs2003_1.pdf [download 15.10.2010].

Küster, M.W., Ludwig, C. & Aschenbrenner, A. (2009). TextGrid: eScholarship und vernetzte Angebote. it – Information Technology, 51(4), 183-190. www.textgrid.de/fileadmin/TextGrid/veroeffentlichungen/kuester-ludwig-aschenbrenner_textgrid-escholarship.pdf [download 15.10.2010]

Latour, B. (2002). Wir sind nie modern gewesen. Versuch einer symmetrischen Anthropologie. Frankfurt am Main: Fischer Taschenbuch Verlag.

Law, J. (2006). Notizen zur Akteur-Netzwerk-Theorie: Ordnung, Strategie und Heterogenität. In: A. Bellinger & D.J. Krieger (Hg.), ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie. Bielefeld: transcript Verlag, 429-446.

Narr, W.-D. (1988). Das Herz der Institution oder strukturelle Unbewußtheit – Konturen einer Politischen Psychologie als Psychologie staatlich-kapitalistischer Herrschaft. Politische Psychologie heute, Sonderheft 9, 111-146.

Narr, W.-D. (2006). Vom Foucault lernen – Erkennen heißt erfahren riskant experimentieren. In: B. Kerchner & S. Schneider (Hg.), Foucault: Diskursanalyse der Politik. Eine Einführung.Wiesbaden: VS Verlag, 345-363.

Neuroth, H., Aschenbrenner, A. & Lohmeier, F. (2007). e-Humanities – eine virtuelle Forschungsumgebung für die Geistes-, Kultur- und Sozialwissenschaften. Bibliothek. Forschung und Praxis, 31(3), 272-279. www.bibliothek-saur.de/2007_3/272-279.pdf [download 15.10.2010].

Nowotny, H., Scott, P. & Gibbons, M. (2001). Re-Thinking Science. Knowledge and the Public in an Age of Uncertainty. Oxford: Polity Press.

Rapp, A. (2007). Das Projekt „TextGrid. Modulare Plattform für verteilte und kooperative wissenschaftliche Textverarbeitung – ein Community-Grid für die Geisteswissenschaften“. Chancen und Perspektiven für eine neue Wissenschaftskultur in den Geisteswissenschaften. In: Arbeitsgemeinschaft historischer Forschungseinrichtungen in der Bundesrepublik Deutschland e.V. (Hg.), Jahrbuch der historischen Forschung in der Bundesrepublik Deutschland 2006. München: Oldenbourg Wissenschaftsverlag, 61-68.

Schatzki, T.R. (2006). On Organizations as they Happen. Organization Studies, 27(12), 1863-1873.

Weingart, P. (2001). Die Stunde der Wahrheit? Zum Verhältnis der Wissenschaft zu Politik, Wirtschaft und Medien in der Wissensgesellschaft. Weilerswist: Velbrück.

Wouters, P. (2009). [without title]. Grid Talk, www.gridtalk.org/Documents/Grids-and-eHumanities.pdf [download 12.08.2010].

An der Schwelle zum Vierten Paradigma – Datenmanagement in der Klimaplattform

Lars Müller

FH Potsdam
Fachbereich Informationswissenschaften
Friedrich-Ebert-Straße 4
D-14467 Potsdam
lars.mueller@fh-potsdam.de

Zusammenfassung

Angesichts der informationstechnischen Entwicklungen befinden wir uns am Übergang zur Daten getriebenen Wissenschaft. Die empirische Untersuchung des Forschungsdatenmanagements in der Klimaplattform ergab als wesentliche Anforderung für diesen Prozess die Entwicklung einer Forschungsdateninfrastruktur, die Übergänge zwischen Disziplinen, Small und Big Science, Institutionen sowie technischen Systemen schafft. Praktische Ansatzpunkte hierfür sind die Einführung von Forschungsdatenpolicies, Semantic Web Technologien und Servicestrukturen für das Forschungsdatenmanagement.

Das Vierte Paradigma

Das „Vierte Paradigma“ bezeichnet nach Jim Gray von Microsoft Research die Vision von einem neuen Wissenschaftsparadigma: nach Empirie, Theorie und Simulation befinden wir uns am Übergang zu einer „Daten getriebenen Wissenschaft“. (Gray, 2009. S. XVIII f.) Die informationstechnologische Verknüpfung der heterogenen Datenmengen eröffnet neue Dimensionen wissenschaftlicher Erkenntnis. Veranschaulichen lässt sich dies anhand „Virtueller Observatorien“: Astronomische Messdaten aus unterschiedlichsten Quellen werden unter einer Oberfläche zusammengeführt und dienen für sich als Datenbasis weiterer Forschung. Voraussetzung für diese eScience (enhanced Science) ist ein systematisches Datenmanagement: die Datenkuratierung. (Vgl. Bell, Hey, & Szalay, 2009; Lynch, 2009. S. 181.) Das heißt, die im Forschungsprozess angefallenen relevanten Daten werden gesichert, archiviert und nachnutzbar gemacht. Neben den Aspekten der Auffindbarkeit und Wiederverwertbarkeit von Datensätzen ist die Sicherung der Nachweiskette ein zentrales Motiv für diese Anstrengungen. (PARSE.Insight, 2009, S. 7.)

Datenmanagement in der Klimaplattform

Forschungsdomäne Klimawandel – Projekt Wibaklidama

An der Fachhochschule Potsdam wurde im Projekt Wibaklidama (Wissensbasiertes Klimadatenmanagement) die Praxis des Forschungsdatenmanagements untersucht, um anhand der Ergebnisse den Wandel mit zu gestalten. Im Fokus standen dabei die Einrichtungen der „Forschungsplattform Klimawandel“, kurz: Klimaplattform. Sie ist ein eingetragener Verein, bestehend aus 23 Wissenschaftseinrichtungen, mit dem Ziel, „Brandenburg-Berlin als Modellregion für das wissenschaftliche Verständnis und den Umgang mit den Folgen des Kli-

mawandels im nationalen und internationalen Kontext zu platzieren.“ (Forschungsplattform Klimawandel, 2008) Die beteiligten Einrichtungen setzen sich zusammen aus Forschungsinstituten, Hochschulen aber auch dem Deutschen Wetterdienst als staatliche Behörde. Das Themenspektrum der Forschungs- und Arbeitsfelder ist ähnlich breit gefächert wie die Art der Institutionen: von Agrarforschung über Gewässerforschung bis zu globalen Fragestellungen der Klimafolgenforschung. Allen Einrichtungen ist gemeinsam, dass sie klimaforschungsrelevante Daten erfassen, über sie verfügen oder mit ihnen arbeiten. Die Fachhochschule Potsdam betreut in der Klimaplattform das Querschnittsforum „Datenaustausch und -verfügbarmachung.“ (Forschungsplattform Klimawandel, 2008)

Mit qualitativen Erhebungen und der Durchführung von Workshops wurde im Projekt Wibaklidama die Praxis des Datenmenagements und der Bedarf an Forschungsdaten ermittelt. Im Dialog mit Datenmanagern wurden Ansätze zum Ausbau einer Dateninfrastruktur entwickelt. (Grossmann, 2010. S. 3.)

Daten in den Einrichtungen Klimaplattform

Bereits in dem begrenzten Forschungsfeld „Klimawandel“ gibt es eine große Vielfalt an Forschungsdaten. Das Spektrum reicht von Daten aus limnologischen Instrumenten über Gravitationsmessungen durch Satelliten bis hin zu Wetteraufzeichnungen. Manche Einrichtungen erheben selbst keine Daten, sondern greifen auf die Bestände anderer zu und berechnen auf dieser Basis komplexe Modelle. (Vgl. Interviews Nr. 1,2,3,4,9,10) Wesentliches Kennzeichen klimaforschungsrelevanter Daten ist, dass sie großvolumig und heterogen sind (Data-Practices-Befragung, Frage 11.).

Bei der Analyse der Klimaplattform-Einrichtungen haben sich vier charakteristische Formen im Umgang mit Forschungsdaten herauskristallisiert (Müller, 2010. S. 12f.). 1) Datenrepositorien für Messinstrumente: Daten werden automatisch in Repositorien eingespielt und über Internetportale zugänglich gemacht. 2) Disziplinspezifische Repositorien-Projekte und Datenzentren: zu speziellen Teildisziplinen werden Datenrepositorien unterhalten. 3) Projektbezogene Datenpublikation (Publikation auf Websites): Zwischenform der Datenpublikation durch Veröffentlichung auf Internetseiten der Forschungsprojekte, die Forschungsdaten sind damit zugänglich, aber nicht im Sinne von zuverlässigen Publikationsservern archiviert und auch nicht systematisch mit standardisierten Metadaten erschlossen. 4) Einzelprojekte mit interner Ablage/Speicherung: die anfallenden Daten werden auf internen Speichersystemen abgelegt, ein Zugang von außen ist nicht vorgesehen.

Vereinzelt existieren in den Einrichtungen Abteilungen oder Arbeitsgruppen, mit der Aufgabe Datenservices für ihre Einrichtung zu Entwickeln und anzubieten.

Teilen von Daten: Bedarf und Vorbehalte

Eindeutig zeigt sich ein breiter Bedarf an Zugangsmöglichkeiten zu vorhandenen Forschungsdaten. „Wir erzeugen eigentlich gar nicht selber Daten, sondern nehmen die Daten Anderer“, (Interview Nr. 4.) so ein Datenmanager. Wichtig für die Nutzung fremder Daten sind „[g]eeichte und korrigierte, vergleichbare Messwerte (Saubere Rohdaten ohne Fehler und prozessierte Daten).“ (Onlinebefragung Informationsbedarf, Antwort 17 v. 11.10.2009) Um gezielt und unmittelbar auf Forschungsdaten zugreifen zu können, wurde in der Online-

Befragung mehrfach der Wunsch nach Verzeichnissen zu Datenbeständen sowie zu datenrelevanten Projekten und Publikationen geäußert. Der Bedarf an solchen Verzeichnissen verweist wiederum auf die Bedeutung hochwertiger Metadaten. Der mehrfach geäußerte Wunsch nach einem Expertenverzeichnis macht deutlich, dass der über Personen vermittelte Weg zu den Daten von großer Relevanz ist. Dies deckt sich auch mit den Ergebnissen aus unseren anderen Befragungen, dass Zugriff auf fremde Forschungsdaten oft über die persönliche Ansprache der Wissenschaftler führt, die über „ihre" Daten verfügen. (Vgl. Grossmann, 2010. S. 7.)

Die Bereitschaft, eigene Daten zu publizieren, ist trotz des bekannten Bedarfs nicht sehr ausgeprägt. Folgende Aussage ist kein Einzelfall: „Metadaten werden nicht automatisch generiert, sondern das muss händisch gemacht werden." (Interview Nr. 4). Die Publikation von Daten wäre mit erheblichem Aufwand für die Forscher verbunden. Zudem wurde die Furcht vor Missinterpretation publizierter Daten mehrfach hervorgehoben. Forscher wollen sich also die Interpretationshoheit sichern und damit wohl auch Wettbewerbsnachteile vermeiden. (Vgl. Interviews Nr. 1, 4, 6)

Auch auf Seiten der Datennutzer gibt es große Vorbehalte gegenüber fremden Daten. Dem grundsätzlichen Bedarf steht die Skepsis hinsichtlich der Qualität dieser Daten gegenüber: „Ich selber bin vorsichtig mit Daten, die ich nicht kenne" (Interview Nr. 1). Für die Qualitätsbewertung bei der Nachnutzung von Daten ist der Persönliche Kontakt zu den Datenerzeugern bzw. -inhabern von großer Bedeutung. Somit gilt für Datenerzeuger wie -nutzer: „der Zugriff auf Daten muss im Dialog erfolgen" (Interview Nr. 3., vgl. auch Grossmann, 2010. 7f.).

Es ist deutlich geworden, dass gegenseitiges Vertrauen eine Grundbedingung für die Nachnutzung von Daten ist. Die Herausbildung von allgemein anerkannten Publikationsformen für Forschungsdaten, die den üblichen Maßstäben für wissenschaftliche Veröffentlichungen entsprechen, ist deshalb fundamental wichtig, weil die qualitätsgeprüfte Publikation das Vertrauensverhältnis zwischen Datenerzeuger und Nutzer vermitteln kann.

Small Science

In den untersuchten Einrichtungen dominiert nach wie vor die „Small Science." (Vgl. Grossmann, 2010) Überschaubare Forschungsprojekte werden zu spezifischen Fragen von wenigen Einzelpersonen durchgeführt. Die anfallenden Daten befinden sich dann am Arbeitsplatz der Wissenschaftler, einen Austausch gibt es innerhalb der Arbeitsgruppe, darüber hinaus findet nur wenig Austausch statt. Nach Projektabschluss ist der Verbleib der Daten ungewiss, oft gehen sie verloren. Policies und Datenmanagementpläne existieren nicht, sind nicht bekannt oder werden nicht eingehalten.[1] Selbstverständlich findet in den Arbeitsgruppen ein spezifisches Datenmanagement statt. Allerdings dominieren projektbezogene Lösungen, wie die Ablage von Daten in einem einfachen Dateisystem. Faktisch sind bei „selbstgestrickten" Lösungen die Daten ohne das Wissen der beteiligten Forscher wertlos. Selbst wenn zentrale Services wie ein Datenrepository in der Institution existieren, werden

[1] Ergebnis des Wibaklidama-Sommerworkshop, 22.6.2010; vgl. hierzu auch die Präsentation von von Jens Klump: http://wibaklidama.fh-potsdam.de/fileadmin/downloads/klump.2010-06022-wibaklidama.pdf.

sie nicht zwangsläufig genutzt. Als besonders problematisch erweist sich die sorgfältige Erschließung mit Metadaten, die für Nachnutzung von Forschungsdaten unabdingbar, aber in der Praxis nur schwer umzusetzen ist, denn sie ist mühsam und zeitintensiv.

Dreh- und Angelpunkt für Nachnutzung von Daten sind die Datenerzeuger selbst. Sie prüfen, ob sie Zugang gewähren können bzw. dürfen und verfügen über das erforderliche Kontextwissen. Das Auffinden erfolgt mittelbar über Projektinformationen und Publikationen. (Grossmann, 2010. S. 7.)

Big Science

Die Situation sieht in datenintensiven Großprojekten deutlich anders aus. „Data Center" werden von einigen Klimaplattform-Mitgliedern selbst oder unter ihrer Beteiligung betrieben. Sie sind disziplinär ausgerichtet, hoch entwickelt und ermöglichen Erfassung nach Metadatenstandards, Online-Recherche und Download von Datensets. Die Daten werden laufend in die Systeme eingespielt und stehen der wissenschaftlichen Analyse zur Verfügung. In den Organisationsbereichen, in denen diese Großprojekte durchgeführt werden, existieren ausgearbeitete Datenpolicies, die Metadatenerfassung findet statt und erfolgt unter Einhaltung internationaler Standards. Der Zugang für externe Forscher zu den Daten ist meist eindeutig geregelt und wird vielfach auch ohne Einschränkungen über Online-Portale ermöglicht. Angesichts hoher Kosten bei der Datengewinnung bzw. der Unwiederholbarkeit von Messungen wurde von Anfang an darauf geachtet, geeignete Standards und Formate zur Speicherung zu nutzen. Auch die digitale Langzeiterhaltung der Daten wurde geplant und ist weitgehend gesichert. Allerdings handelt es sich in vielen Fällen um Insellösungen und der Ausbau zu einer institutionen- und disziplinübergreifenden Infrastruktur, die auch die Small Science einschließt, steht aus. (Grossmann, 2010, S. 4. Müller 2010, S. 11f.)

Einrichtungstypen

Projekte im Sinne systematischer Forschungsdatenkuratierung gibt es innerhalb der Klimaplattform nur in den reinen Forschungseinrichtungen. Dort existiert zudem ein hohes Maß an internationaler Vernetzung, Datenrepositorien werden entwickelt und betrieben. In der Regel gibt es auch Richtlinien zum Umgang mit Forschungsdaten. Allerdings gilt auch in diesen Forschungszentren die Unterscheidung zwischen Big Science und Small Science: Die Small Science existiert parallel zu den großen Daten-Infrastrukturprojekten in denselben Einrichtungen. Es zeigt sich dennoch ein deutlicher Unterschied zwischen diesen großen und fachlich hoch spezialisierten Forschungsinstituten und den Hochschulen.

Forschungsdaten als Gegenstand oder Publikationsform eigener Arbeit haben wir an Hochschulen nicht gefunden. Zentrale Dienstleistungen der Rechenzentren beschränken sich dort auf Bereitstellung von IT-Infrastruktur und Gewährleistung der technischen Datensicherung. Da Forschungsdatenrepositorien disziplinspezifisch aufgebaut und betrieben werden, können die fachlich breit ausgerichteten Hochschulen diesen Service nicht leisten.

Ein Sonderfall in unserer Untersuchung war der Deutsche Wetterdienst. Er unterscheidet sich im Umgang mit seinen Daten sowohl von Forschungseinrichtungen als auch Hochschulen und wird deshalb als eigener Einrichtungstyp betrachtet. Die Daten, ihre Speicherung

und Weitergabe gehören zum Kerngeschäft dieser Behörde. Die Handhabung der Daten leitet sich aus ihrem gesetzlichen Auftrag ab und ist genau geregelt. Datenspeicherung und Metadatenerfassung erfolgt gemäß internationaler Standards und stellt kein organisatorisches Problem dar. Auch Weitergabe und Verkauf von Daten ist eindeutig geregelt. (Vgl. Müller, 2010. S. 14f.) Bei der Entwicklung von Forschungsdateninfrastrukturen müssen folglich nicht nur die Wissenschaftsdisziplinen, sondern auch die institutionellen Bedingungen, unter denen sie ausgebaut werden, Beachtung finden.

Ergebnisse

Vier Brücken zur Forschungsdateninfrastruktur

Die Entwicklung einer Infrastruktur für die Data-Driven-Science im Sinne des „4. Paradigmas“ steht am Anfang. Es existieren zahlreiche Insellösungen. Für die Ausdehnung dieser Anätze bedarf es weiterer Anstrengungen. Bezogen auf die Untersuchung der Klimaplattform lässt sich sagen, dass in vier Feldern Brücken geschlagen werden müssen, um Forschungsdatenaustausch und -versorgung nachhaltig zu verbessern:

Zwischen den Disziplinen

Gerade im interdisziplinär geprägten Forschungsfeld „Klimawandel“ bilden Daten aus Nachbardisziplinen häufig die wesentliche Basis für neue Arbeiten. Übergreifende technische wie qualitative Standards werden die Integration von Daten aus unterschiedlichen Systemen verbessern. Die Verwendung von verbreiteten Metadatenstandards muss den disziplinübergreifenden Transfer von Forschungsdaten unterstützen.

Von Big Science zu Small Science

Große, projektübergreifende Systeme müssen auch die Forschungsdaten aus den Small Sciences aufnehmen und ganze Wissenschaftsdisziplinen abdecken.

Zwischen den Institutionen

Aktive Kooperationen zwischen den drei Einrichtungstypen sind erforderlich, um die gesamte Wissenschaftslandschaft mit einer guten Forschungsdateninfrastruktur zu versorgen.

Zwischen technischen Systemen

Die Wissenschaftlerarbeitsplätze müssen mit den Forschungsdatenspeichern/-repositorien unmittelbar gekoppelt werden können. Das gilt gleichermaßen für Datenerzeuger wie -nutzer. Es geht dabei darum, bereits im Entstehungsprozess Daten systematisch zu verwalten, automatisch mit Metadaten zu versehen und in archivtauglichen Formaten zu speichern. Für die Nachnutzung heißt dies, den direkten Zugriff mit Analyse- oder Visualisierungstools auf die Repositorien zu ermöglichen.

Ansatzpunkte

Für den Ausbau einer Daten-Infrastruktur zeichnen sich drei zentrale Aufgabenfelder ab.

1) Eine breite Einführung von Forschungsdatenpolicies muss auf die Institutsorganisation sowie die Förderung einer Publikationskultur für Forschungsdaten zielen. Damit werden die Sensibilisierung für das Thema Datenmanagement vorangetrieben und die institutionellen Rahmenbedingungen verbessert. Es geht dabei beispielsweise um die Festlegung von ver-

bindlichen Standardformaten, um wissenschaftliche Anerkennung für publizierte Datensätze und nicht zuletzt um ausreichende technische wie personelle Ausstattung des Datenmanagements. Auf dieser organisatorischen Basis kann die technische Infrastruktur aufsetzen. 2) Die Einführung von Semantic-Web-Technologien zum Datenmanagement und -austausch zeichnet sich als geeignet ab, um bestehende Einzellösungen zu einem Forschungsdatennetz zu verweben. Vielfalt und Ungleichzeitigkeit von Systementwicklungen erlauben kein zentral administriertes System. Durch die Verwendung eines offenen Standards wie Linked Data und möglichst offenen Schnittstellen sollen sich Verknüpfungen und Übergänge zwischen Disziplinen und Systemen selbst generieren. (Vgl. Anschlussprojekt: Borchert und Rümpel, 2010) 3) Es bedarf der Entwicklung von Servicestrukturen, die Datenanbieter wie Nutzer von Verwaltungsaufgaben entlasten.

Fazit

Angesichts der eingangs erläuterten Vision vom Vierten Paradigma wirken die genannten Maßnahmen profan. Es zeigt sich, dass „Data-driven-Science“ ein Infrastrukturprojekt auf organisatorischer wie technischer Ebene ist. Die Entwicklungen stehen am Anfang und sind eine komplexe und langfristiges Aufgabe. Dennoch: Es gibt beeindruckende Technik und tragfähige Konzepte für datenbasierte Wissenschaften, die ein großes Potential für die Schaffung neuen Wissens bergen. Wir stehen an der Schwelle, diese Ansätze zu einer umfassenden Infrastruktur auszubauen.

Quellen

Interviews

Die qualitativen Interviews wurden mit Datenmanager/inne/n aus jeweils unterschiedlichen Einrichtung der Klimaplattform geführt. Der Interviewleitfaden beinhaltete Art, Handhabung und Nutzung der Daten in der Einrichtung sowie künftigen Bedarf.

Interview -Nummer	Datum	Interview er/in
Interview Nr. 1	3.6. 2009	Grossmann, Silke
Interview Nr. 2	5.6.2009	Grossmann, Silke
Interview Nr. 2b	24.7.2009	Grossmann, Silke
Interview Nr. 3	15.6.2009	Grossmann, Silke
Interview Nr. 4	17.6.2009 / 2.7.2009	Grossmann, Silke; Büttner, Stephan; Hobohm, Hans-Christoph
Interview Nr. 5	6.7.2009	Grossmann, Silke
Interview Nr. 6	13.7.2000	Grossmann, Silke; Büttner, Stephan
Interview Nr. 7	13.7.2009	Grossmann, Silke
Interview Nr. 8	30.7.2009	Grossmann, Silke
Interview Nr. 9	28.9.2009	Grossmann, Silke; Büttner, Stephan
Interview Nr. 10	14.9.2009	Grossmann, Silke

Onlinebefragung Informationsbedarf

Im Zeitraum vom 16.10.2009 bis 24.11.2009 wurde in der Klimaplattform eine Onlinebefragung zur Bedarfsermittlung im Forschungsdatenmanagement durchgeführt.

Data-Practices-Erhebung

Im Dezember 2009 wurden mit einem standardisierten Fragebogen die „Data-Practices“ in der Klimaplattform erhoben.

Literatur

Bell, G., Hey, T., & Szalay, A. (2009). Computer Science: Beyond the Data Deluge. Science, 323(5919), 1297–1298. doi:10.1126/science.1170411.

Borchert, Friederike & Rümpel, Stefanie (2010). Semantisches Datenmanagement – Werkzeuge und Selbstlernangebote für den nachhaltigen Einsatz in der Klimaforschung, AP1. Projektbericht, unveröffentlicht. Potsdam. Onlinepublikation voraussichtlich Januar 2011.

Forschungsplattform Klimawandel (2008). Motivation. http://www.klimaplattform.de/motivation-zur-klimaplattform.html [22.12.2010]

Gray, J. (2009). Jim Gray on eScience: a Transformed Scientific Method. In T. Hey, S. Tansley, & K. Tolle (Eds.), The Fourth Paradigm, Data-Intensive Scientific Discovery. Redmond: Microsoft Research, xvii–xxxi. http://research.microsoft.com/enus/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf.

Grossmann, S. (2010). Datenmanagement in der Klimaforschung; Akteure, Bedarfe, Practices: Projektbericht B1. Potsdam. urn:nbn:de:kobv:525-opus-1704.

Lynch, C. (2009). Jim Gray's Fourth Paradigm and the Construction of the Scientific Record. In T. Hey, S. Tansley, & K. Tolle (Eds.), The Fourth Paradigm, Data-Intensive Scientific Discovery. Redmond: Microsoft Research, 177–183. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part4_lynch.pdf.

Müller, L. (2010). Umfeldanalyse zum Klimadaten-Management: Projektbericht B2. Potsdam. urn:nbn:de:kobv:525-opus-1712.

PARSE.Insight (2009). Road Map. Deliverable D2.1. from http://www.parse-insight.eu/downloads/PARSE-Insight_D2-1_DraftRoadmap_v1-1_final.pdf.

da|ra – Ein Service der GESIS für die Zitation sozialwissenschaftlicher Daten

Brigitte Hausstein & Wolfgang Zenk-Möltgen

GESIS – Leibniz Institut für Sozialwissenschaften
Bachemer Straße 40
Köln
brigitte.hausstein@gesis.org, wolfgang.zenk-moeltgen@gesis.org

Zusammenfassung

Während für Forschungspublikationen der freie Zugang (Open Access) immer mehr zur gängigen Praxis wird, sind die Bemühungen hinsichtlich allgemein zugänglicher Datenpublikationen erst am Anfang. Darüber hinaus ist eine verlässliche Identifikation und Zitation eines speziell für die Beantwortung einer Forschungsfrage verwendeten sozialwissenschaftlichen Forschungsdatensatzes nur begrenzt möglich. Ein Weg zur Lösung der Problematik ist der Einsatz von speziellen Persistenten Identifikatoren. Das von der International DOI Foundation (IDF) verwaltete System der DOIs hat ausgezeichnete Aussichten auf Verbreitung und Langlebigkeit. Durch die Mitgliedschaft in DataCite hat GESIS die Voraussetzungen für die Vergabe von DOIs geschaffen. Im Februar 2010 wurde ein Pilotprojekt zur Entwicklung eines Registrierungsservice für die von GESIS selbst vorgehaltenen Forschungsdaten begonnen. Der Beitrag beschreibt die Kernkomponenten des entwickelten Services.

Hintergrund

Nicht nur der Zugang zu Literatur, sondern auch allgemein zugängliche und langfristig verfügbare Forschungsdaten sind für eine Wissenschaft, die exzellente Forschungsergebnisse erbringen will, essentiell. Die schnelle Entwicklung der digitalen Technologien und Netzwerke der letzten Jahre hat die Produktion, Verbreitung und Nutzung von Forschungsdaten jedoch radikal verändert. Neue Verfahren und Messinstrumente bringen wachsende, aber auch komplexere Datenmengen und -typen hervor. Entsprechend entwickelte Software-Tools helfen die Fülle der gesammelten Primärdaten zu verwalten, zu interpretieren und in Informations- und Wissenssammlungen zu transformieren. Das wichtigste und allseits präsente Forschungswerkzeug, das Internet, hat die Art und Weise, wie Daten und Informationen ausgetauscht und verfügbar gemacht werden, stark verändert (Uhlir 2003).

Während für Forschungspublikationen neben den traditionellen Angeboten der freie Zugang (Open Access) immer mehr zur gängigen Praxis wird, sind die Bemühungen hinsichtlich allgemein zugänglicher Datenpublikationen erst am Anfang. Obwohl grundsätzlich die Bereitschaft zur Weitergabe der Primärdaten existiert, scheitert dies oft an den fehlenden Kapazitäten, die für die Aufbereitung und Metadatenbeschreibung notwendig sind. Dies gilt auch für die Sozialwissenschaften, die im Vergleich zu anderen Disziplinen bereits eine ausgeprägtere Kultur des Data Sharings kennen.

Die Verbreitung der Forschungsergebnisse erfolgt daher fast ausschließlich über Publikationen in Fachzeitschriften. Die „Allianz der deutschen Wissenschaftsorganisationen“ hat jedoch Ende Juni 2010 in den „Grundsätze(n) zum Umgang mit Forschungsdaten“ eine Primärdaten-Policy gefordert, um bei Wissenschaftlern das Bewusstsein für den Handlungsbedarf und für den Nutzen von Primärdaten-Infrastrukturen zu schärfen (Allianz der deutschen Wissenschaftsorganisationen, 2010). Und von Seiten der Forschungsfinanzierer wird zunehmend gefordert, nicht nur die Forschungspublikationen, sondern auch die entstandenen Primärdaten im Sinne von Good Scientific Practice öffentlich zugänglich zu machen. Daraus ergibt sich die besondere Bedeutung einer reinen Datenpublikation, mit allen Möglichkeiten der eindeutigen Identifikation und kompakten Zitierung, die für Textpublikationen bereits Standard sind. So wird auch im Rahmenkonzept für die Fachinformationsinfrastruktur in Deutschland konstatiert: „Es wird immer wichtiger nicht nur die endgültigen Ergebnisse der Forschungsarbeit verfügbar zu machen, sondern auch die im Verlauf des Forschungsprozesses generierten und veränderten Daten. Hier kommen eine Reihe neuer Anforderungen, z. B. im Rahmen der ‚Sicherung guter wissenschaftlicher Praxis’ auf die Wissenschaftler, ihre Institute ebenso wie auf die Informationseinrichtungen zu.“ (Arbeitsgruppe Fachinformation, 2009: 6)

Darüber hinaus zeigt sich in allen Wissenschaftsbereichen auch die Dringlichkeit der Kopplung von Datenarchivierung und wissenschaftlicher Literaturpublikation. Die Trennung von Forschungspublikation und zugrundeliegenden Daten erschwert die Evaluation der Publikation und schränkt die Nachvollziehbarkeit der dargestellten Ergebnisse ein. Bislang erfolgt die Verbindung von Daten- und Forschungspublikation nur punktuell. Voraussetzung für die Verbindung von Forschungsprimärdaten und wissenschaftlicher Publikation ist jedoch neben der Langzeitarchivierung der Daten und entsprechender Qualitätssicherung, die Möglichkeit zur Publikation von Daten mit eindeutiger Identifizierbarkeit und Referenzierbarkeit.

Persistent Identifier und Forschungsdatenregistrierung

Ein Weg zur Lösung der geschilderten Problematik ist der Einsatz von speziellen Persistent Identifiern (PID). Deren Funktion entspricht in etwa einer ISBN-Nummer bei gedruckten Werken, die lediglich ein einziges Mal vergeben wird, so dass sie nur identische Kopien einer Publikation bezeichnen kann. Hinzu kommt die Unterscheidung zwischen dem Identifier und der Lokation eines Objekts, die es ermöglicht, das Objekt unabhängig von seinem Speicherort zu identifizieren. Dies unterscheidet die PID von einer URL. Zur Sicherstellung der eindeutigen Vergabe und der Zuweisung von Kennung und Speicherort bedarf es eines automatisierten Dienstes. Jeder PID werden dabei Adressinformationen, z. B. ein Universal Ressource Locator (URL) zugewiesen. Von zentraler Bedeutung sind hier geeignete organisatorische Maßnahmen, die Verweise auf die tatsächlichen Speicherorte der Ressourcen aktuell halten. Programme können dann über einen sog. Resolver die zitierte PID auflösen, so dass ein Zugang zu den zitierten Forschungsdaten möglich wird.

Es existieren mittlerweile für die Identifikation von elektronischen Textpublikationen diverse Systeme von PIDs, die technisch die Basis für einen Service auch zur Identifizierung von Daten leisten können: Archival Research Key (ARK), Digital Object Identifier (DOI), Handle,

Library of Congress Control Number (LCCN), Life Science Identifiers (LSID), Persistent URL (PURL), Uniform Resource Name (URN) und weitere. Auf einen gemeinsamen Standard haben sich die verschiedenen Nutzergemeinden jedoch noch nicht geeinigt, da die Systeme im Prinzip gut ineinander überführbar sind. Um die langfristige Eignung zu beurteilen, ist hier weniger die technische als die organisatorische Ausgestaltung relevant.

Im Februar 2010 wurde GESIS Mitglied in DataCite. Damit wurden die Voraussetzungen (Vergaberecht) für die Etablierung eines Registrierungsservices für sozialwissenschaftliche Primärdaten auf der Basis von Digital Object Identifiers (DOI) geschaffen. DataCite ist ein 2009 in London gegründetes internationales Konsortium mit zwölf Mitgliedern aus neun verschiedenen Ländern, die gemeinsam das Ziel verfolgen, die Akzeptanz von Forschungsdaten als eigenständige, zitierfähige wissenschaftliche Objekte zu fördern. DataCite bietet neben der Mitgliedschaft in der International DOI Foundation (IDF) eine international abgestimmte Vorgehensweise bei technischen Lösungen, Standards, Best Practises und Workflows sowie ein gemeinsames Metadatenschema.

da|ra Registrierungsagentur für sozialwissenschaftliche Daten

GESIS hat im Frühjahr 2010 ein Pilotprojekt unter dem Titel da|ra (Registrierungsagentur für sozialwissenschaftliche Daten) zur Etablierung eines Registrierungssystems für Forschungsdaten der Sozialforschung im deutschsprachigen Raum gestartet. Ziel ist es, eine Infrastruktur zu entwickeln, die es ermöglicht, Forschungsdatenbestände mit DOIs zu versehen und sie mit ihren Titeln, Themen, Autoren, Provenienzen, Methoden und Zugangsmöglichkeiten so umfassend wie möglich nachzuweisen, findbar und zitierbar zu machen.

In Kooperation mit DataCite wurde dazu mit der technischen Implementierung eines Registrierungstools begonnen und ein spezifisches Metadatenmodell entwickelt. Gegenwärtig erfolgt mit der Registrierung der Forschungsdaten des GESIS-Bestandes der Test der technischen und organisatorischen Lösung. In einer weiteren Projektphase (2011-2012) wird der Service weiteren datenhaltenden Organisationen, wie z. B. die Forschungsdatenzentren, Datenservicezentren oder institutionalisierten Forschungsprojekten angeboten und die Einbeziehung von Wirtschaftsdaten vorbereitet. Schwerpunkt in dieser Phase ist die Weiterentwicklung des Registrierungsservices, die Zertifizierung von Publikationsagenten, die Entwicklung von Standards und Guidelines sowie die Entwicklung eines erweiterten Nachweissystems. Die Einbeziehung weiterer Datenrepositorien und Datenkuratoren als Publikationsagenten ist für die Ausbauphase (ab 2013) geplant. In dieser Phase wird der Dienst in Abhängigkeit von der Entwicklung eines Self-archiving Services bei GESIS auch Einzeldatenproduzenten angeboten.

Organisation und Verwaltung der Registrierung

Die DOI Registrierung erfolgt bei da|ra in enger Kooperation mit den datenhaltenden Organisationen, den so genannten Publikationsagenten, die für die Pflege und Speicherung der Forschungsdaten sowie für die Metadatenpflege zuständig sind. Die Datensätze verbleiben bei den Datenzentren, da|ra speichert die Metadaten und macht die registrierten Inhalte über eine Datenbank recherchierbar.

Die Rahmenbedingungen und Voraussetzungen für die DOI Registrierung sind in einer Policy festgehalten. Darüber hinaus werden der Workflow und sämtliche Verantwortlichkeiten im Registrierungsprozess in einem Service Level Agreement vereinbart. Hierunter fallen Fragen zur Qualitäts- und Persistenzsicherung (Daten und Metadaten), zu Urheberrechten, Versionierungen, Kosten, zur Verfügbarkeit des Services sowie deren Funktionalitäten etc. Hierzu werden die Erfahrungen der DataCite Mitglieder einbezogen.

Technische Implementation

Die technische Implementation erfolgt über eine serviceorientierte Architektur (vgl. Abb. 1). Zentrale Komponenten, wie beispielsweise das Registrieren einer DOI oder das Indexieren, wurden in separate Services ausgelagert. Entsprechend verfügt auch das zentrale Informationssystem über Schnittstellen, die als Webservice angesprochen werden können. Die Daten werden in einer Datenbank gemäß DDI Standard (http://www.ddialliance.org) abgespeichert. Dadurch werden ein Im- und Export der Daten im XML-Format auf einfache Weise ermöglicht und ein Austausch mit den Publikationsagenten direkt unterstützt.

Für die Funktionalität der Suche wird ein Indexierungsserver eingesetzt (SOLR). Der Bereich der Nutzerzugänge umfasst darüber hinaus Browsing und Visualisierungen (z. B. Matrixdarstellungen der nachgewiesenen Studien nach Zeit und Themenklassifikation). In diesem Zuge wird eine Anreicherung der angezeigten Daten mit semantischen Auszeichnungen sowie eine Suche auf der Basis von SPARQL-Anfragen realisiert.

Die Oberfläche zum Editieren der Metadaten wird von Projektmitarbeitern oder den Publikationsagenten benutzt, um die Metadaten zu pflegen. Die Edition wird durch die Verwendung von kontrollierten Vokabularen unterstützt, damit möglichst standardisierte Metadaten erstellt werden.

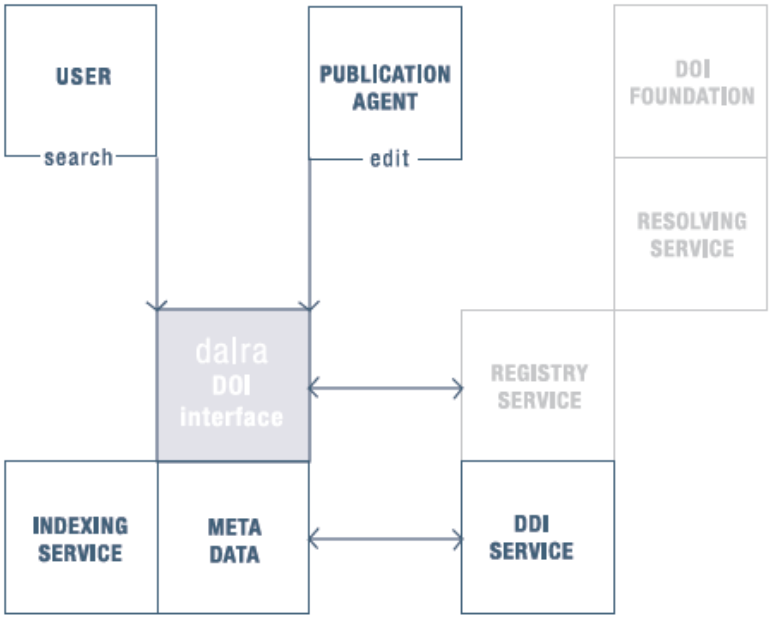


Abbildung 1: Technische Implementation da|ra

Metadatenmodell

Neben dem technischen Registrierungsservice leistet da|ra auch die Übernahme von Metadatenbeschreibungen in sein Datenbanksystem. Damit wird eine differenzierte Suche aller registrierten Datensätze ermöglicht, eine inhaltliche Beschreibung zur Verfügung gestellt und die Voraussetzung zur einheitlichen Zitation der Daten geschaffen. Unter Berücksichtigung der besonderen Herausforderungen der Dokumentation digitaler Datenobjekte wurde für das Informationssystem ein eigenes Beschreibungsschema entwickelt. Dieses basiert auf dem Metadatenschema des GESIS-Datenbestandskatalogs und wurde in Anlehnung an das DataCite-Kernschema (basierend auf ISO 690-2, dem Dublin Core Metadata Element Set, Version 1.1, und dem DOI-Metadata Kernel der International DOI Foundation) erweitert. Durch das Mapping mit DDI 2.0 und die geplante Ausweitung auf DDI 3.1, die Passfähigkeit mit von DataCite verwendeten Standards sowie durch die Verwendung kontrollierter Vokabulare ist die Interoperabilität der erfassten Ressourcen mit anderen Datenbeständen auch im internationalen Rahmen gewährleistet.

Ein wesentliches Element der Datenbeschreibung ist die Möglichkeit der Versionierung, die gerade im Bereich Daten unerlässlich ist für die eindeutige Referenzierung eines Datensatzes. Da jede Version mit einem eigenen DOI und eigenen Metadaten versehen ist, wird die eindeutige Bezugnahme auf den einer Analyse zugrundeliegenden Datensatz und damit z. B. auch die Verifizierung von Analyseergebnissen ermöglicht.

Ein Schwerpunkt der Vereinbarung mit dem Publikationsagenten ist die Sicherstellung der Qualität der Metadaten. Bei der Registrierung neu anfallende Metadaten werden durch da|ra zunächst geprüft und bei Bedarf entsprechend dem da|ra-Metadatenschema nachgearbeitet. In der Folge ist der Publikationsagent verantwortlich für die Aktualität der Metadaten, insbesondere die korrekte Auflösung des DOI, während da|ra für die permanente technische Auflösbarkeit zuständig ist. Hierdurch wird die Persistenz des Angebotes gewährleistet.

Beispiel: Kurzansicht nach Schnellsuche in da|ra

Fazit

Die Testphase des Projektes konnte mit der Registrierung von 4465 Studien des GESIS Datenbestandes erfolgreich abgeschlossen werden. Sowohl die technischen wie auch die organisatorisch-administrativen Basislösungen haben sich bewährt. Damit sind die Voraussetzungen für die Zitation eines wichtigen sozialwissenschaftlichen Datenbestandes geschaffen worden. Zukünftig gilt es darauf aufbauend Mehrwertdienste zu entwickeln und weitere datenhaltende Organisationen einzubeziehen.

Für weitere Informationen zum da|ra Registrierungsservice besuchen Sie die Website: www.gesis.org/dara oder kontaktieren die Autoren.

Literaturverzeichnis

Allianz der deutschen Wissenschaftsorganisationen (2010): Grundsätze zum Umgang mit Forschungsdaten. Retrieved from http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaetze/

Askitas, N. (2010): What Makes Persistent Identifiers Persistent? Working Paper No. 147. RatSWD Working Paper Series June 2010. Retrieved from http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_147.pdf

Arbeitsgruppe Fachinformation (2009): Rahmenkonzept für die Fachinformation. Vorlage zur Sitzung des Ausschusses der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder (GWK) am 29.09.2009.

Bellini, E., Cirinnà, C., Lunghi, M, Damiani, E., Fugazza, C. (2008): Persistent Identifiers distributed system for Cultural Heritage digital objects. Retrieved from http://www.bl.uk/ipres2008/presentations_day2/38_Lunghi.pdf.

Bleuel, J. (2000): Zitation von Internet-Quellen (Citing of Internet sources). In: Hug, T. (Eds.): Wie kommt Wissenschaft zu Wissen? Band1: Einführung in das wissenschaftliche Arbeiten. Hohengehren: Schneider Verlag.

Blue Ribbon Task Force (2010): Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information. Retrieved from http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf.

Brase, J. (2010): DataCite – A global registration agency for research data Working Paper No. 149. RatSWD Working Paper Series July 2010. Retrieved from http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_149.pdf

Brase, J. & Klump, J. (2007): Zitierfähige Datensätze. Primärdaten-Management durch DOIs. In: Wissenschaftskommunikation der Zukunft. 4. Konferenz der Zentralbibliothek. Retrieved from http://dc110dmz.gfz-potsdam.de/contenido/std-doi/upload/pdf/Brase_Wisskomm_2007.pdf

DFG (1998): Recommendations of the Commission on Professional Self Regulation in Science. Proposals for Safeguarding Good Scientific Practice. Retrieved from http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/self_regulation_98.pdf

Diepenbroek, M. & Grobe, H. (2007): PANGAEA® als vernetztes Verlags- und Bibliotheks-system für wissenschaftliche Daten. In: Wissenschaftskommunikation der Zukunft. 4. Konferenz der Zentralbibliothek. Retrieved from http://www.fz-juelich.de/zb/datapool/page/1000/Diepenbroek_Abstract.pdf

Dittert, N., Diepenbroek, M. & Grobe, H. (2002): Data and information management for the CMTT synthesis. Manuscript. Retrieved from http://epic.awi.de/Publications/Dit2002b.pdf

European Union (2010): 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. A submission to the European Commission. Retrieved from http://goo.gl/WrxO

GESIS Report 4/10 (2010): Der Aktuelle Informationsdienst für die Sozialwissenschaften. GESIS – Leibniz Institute für Sozialwissenschaften. Mannheim September 2010. Retrieved from http://www.gesis.org/fileadmin/upload/institut/presse/gesis_report/gesis_report_10_04.pdf

Klump, J., Bertelmann, R. et al. (2006): Data publication in the open access initiative. Data Science Journal, 5(79).

Metadata for the Publication and Citation of Scientific Primary Data (Version 3.0). Retrieved from http://www.icdp-online.org/contenido/std-doi/upload/pdf/STD_metadata_kernel_v3.pdf.

Paskin, N. (2000): Digital Object Identifier: implementing a standard digital identifier as the key to effective digital rights management. The International DOI Foundation Kidlington, Oxfordshire, United Kingdom. Retrieved from http://www.doi.org/doi_presentations/aprilpaper.pdf

Parsons, M. A., Duerr, R. & Minster, J.-B. (2010): Data Citation and Peer Review. In: Eos, Transactions, American Geophysical Union, 91(34), 297–304. Retrieved from http://aurora.gmu.edu/spaceweather/images/2010EO340001.pdf.

Reilly, S. (2010): Digital Object Repository Server: A Component of the Digital Object Architecture. D-Lib Magazine, 16(1/2). Retrieved from http://www.dlib.org/dlib/january10/reilly/01reilly.print.html

Uhlir, P. F. (2003): Discussion Framework. In Esanu, J. M. & Uhlir, P. F. (Eds.): The Role of Scientific and Technical Data and Information in the Public Domain. Washington DC: The National Academies Press.

Uhlir, P. F. & Schroder, P. (2008): Chapter 8 – Open Data for Global Science. In Fitzgerald, B.: Legal Framework for e-Research: Realising the Potential. Sydney: Sydney University Press Law Books 39 (2008) 188. Retrieved from http://www.austlii.edu.au/au/journals/SydUPLawBk/2008/39.html.

Internet:

Allianz der deutschen Wissenschaftsorganisationen. Retrieved from http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaetze/.

California Digital Library. Retrieved from http://www.cdlib.org/inside/diglib/ark/.

Corporation for National Research Initiatives® (CNRI). Retrieved from http://www.cnri.reston.va.us.

da|ra. Retrieved from http://www.gesis.org/dara.

DataCite. Retrieved from http://www.datacite.org.

Data Documentation Initiative. Retrieved from http://www.ddialliance.org/.

Die Deutsche Zentralbibliothek für Medizin (ZB MED). Retrieved from http://www.zbmed.de/info.html

Digital Object Identifier (DOI®) System. Retrieved from http://www.doi.org/.

GESIS – Leibniz Institut für Sozialwissenschaften (Mannheim, Köln, Bonn, Berlin). Retrieved from http://www.gesis.org.

Technische Informationsbibliothek (TIB) Hannover. DOI-Registrierungsagentur. Retrieved from http://www.tib-hannover.de/de/die-tib/doi-registrierungsagentur/.

Technische Informationsbibliothek (TIB) Hannover. Projekt CODATA. Retrieved from http://www.tib-hannover.de/de/die-tib/projekte/codata/.

Library of Congress: Structure of the LC Control Number. Retrieved from http://www.loc.gov/marc/lccn_structure.html

Life Sciences Identifiers (LSID). Retrieved from http://lsids.sourceforge.net/.

Pangaea®: Publishing Network for Geoscientific & Environmental Data. Retrieved from http://www.pangaea.de/.

Publikation und Zitierfähigkeit wissenschaftlicher Primärdaten. Retrieved from http://www.std-doi.de/front_content.php.

Persistent Uniform Resource Locator (PURL). Retrieved from http://purl.oclc.org/docs/index.html.

The Handle System®. Retrieved from http://www.handle.net.

Uniform Resource Name (URN) Syntax. Retrieved from http://tools.ietf.org/html/rfc2141.

ZACAT – GESIS Online Study Catalogue. Retrieved from http://zacat.gesis.org/webview/index.jsp.

Elektronisches Laborbuch: Beweiswerterhaltung und Langzeitarchivierung in der Forschung

Jan Potthoff & Sebastian Rieger

Karlsruhe Institute of Technology (KIT)
[Steinbuch Centre for Computing]
Hermann-von-Helmholtz-Platz 1
D-76344 Eggenstein-Leopoldshafen
sebastian.rieger@kit.edu jan, potthoff@kit.edu

Paul C. Johannes

Universität Kassel [provet]
Wilhelmshöher Allee 64
D-34109 Kassel
paul.johannes@uni-kassel.de

Moaaz Madiesh

Physikalisch-Technische Bundesanstalt (PTB)
Bundesallee 100
D-38116 Braunschweig
moaaz.madiesh@ptb.de

Zusammenfassung

In den experimentellen Wissenschaften werden Forschungsprozesse, -daten und -ergebnisse immer häufiger nur noch in digitaler Form in einem elektronischen Laborbuch aufgezeichnet. Dabei müssen auch die Pflichten zur Archivierung und Sicherung nach der guten wissenschaftlichen Praxis eingehalten werden. Elektronischen Laborbüchern sollte deswegen auf technischem Weg eine Verbindlichkeit verliehen werden, welche mit der von herkömmlichen papiergebundenen Laborbüchern vergleichbar ist. Dazu sollen insbesondere bestehende Sicherheitstechniken zur digitalen Langzeitarchivierung und zur Beweiswerterhaltung von elektronischen Dokumenten nutzbar gemacht werden.

Einleitung: Laborbücher, elektronische Datenmengen und die gute wissenschaftliche Praxis

Das Laborbuch ist eng mit den experimentellen Wissenschaften verbunden. Es ist nach Ebel und Bliefert (2009, S. 5f.) ein Tagebuch, in dem der experimentierende Wissenschaftler in einer über den Tag hinausreichenden Form seine tägliche Arbeit und seine Forschungsergebnisse für die spätere Auswertung dokumentiert. Im Laborbuch werden Forschungsprojekte festgehalten und es enthält Informationen über experimentelle Ergebnisse, Methoden und die Beteiligung von Forschern. Unabhängig von der Bezeichnung (wahlweise Labor-

buch, Laborjournal, Protokoll-, Notiz- oder Tagebuch) sind diese Dokumentationen und Datensammlungen essentielle Werkzeuge in der Forschung. Es sind nach Ebel, Bliefert und Greulich (2006, S. 16) die Keimzellen der naturwissenschaftlichen Literatur, auf welchen Zwischenberichte, Berichte und Publikationen aller Art aufbauen. Laborbücher gehören zum Standard der wissenschaftlichen Praxis und haben eine jahrhundertealte Tradition (vgl. Nachweise bei Holmes, Renn und Rheinberger (Hrsg.) (2003, S. VIII ff. und S. 25f.)). Über die Forschung hinaus werden Laborbücher auch in der Analytik, z. B. in medizinischen Laboren, der Lebensmittelkontrolle und anderen praktischen Anwendungsbereichen (z. B. staatliche Zulassungsverfahren), zur Dokumentation von Untersuchungsergebnissen verwendet.

Das Laborbuch in Papierform hat jedoch in vielen Bereichen ausgedient. Die in Forschungsprozessen zu verarbeitenden (elektronischen) Datenmengen haben stark zugenommen. Durch elektronische Laborgeräte und automatisierte Laborprozesse werden digitale Daten erzeugt, die sehr umfangreich sind. Diese stetig anwachsende Datenmenge ist mit herkömmlichen Laborbüchern in Papierform nicht länger zu beherrschen, z. B. produzieren allein die Forschungsprojekte am LHC (Teilchenbeschleuniger des CERN in Genf) ca. 15 Millionen Gigabyte jährlich. Solche Datenmengen werden nicht mehr in ein Laborbuch in Papierform eingetragen, sie existieren daneben in digitaler Form. So entstehen hybride Laborbücher, die teilweise auf Papier und teilweise am Computer geführt werden. Diese Hybriden entstehen aber auch bei kleinen Datenmengen, wenn durch elektronische Laborgeräte erzeugte digitale Daten aus Gründen der besseren Verarbeitungsmöglichkeiten nur elektronisch vorgehalten werden. Teilweise werden Laborbücher auch schon komplett elektronisch geführt, dann aber zum Zwecke der Langzeitarchivierung und Beweiswerterhaltung vollständig ausgedruckt und aufwendig gelagert. Ein medienbruchfrei am Computer geführtes elektronisches Laborbuch (eLab) dient damit nicht nur der Arbeitserleichterung, sondern ist auch ein notwendiges Werkzeug des Daten- und Dokumentenmanagements.

Denn eine nachhaltige (elektronische) Dokumentation von Forschungsergebnissen ist auch nach den Regeln der guten wissenschaftlichen Praxis (GWP), wie sie z. B. die Deutsche Forschungsgemeinschaft (DFG) und die Max-Planck-Gesellschaft verlangen, erforderlich. Nach diesen ist eine Dokumentation von Ansätzen, Erkenntnissen und wichtigen Ergebnissen während des Forschungsprozesses durchzuführen. Dabei soll auf eine eindeutige und nachvollziehbare Dokumentationsweise geachtet werden. Die Regeln der GWP schreiben eine Archivierung der Forschungsdaten für einen Zeitraum von mindestens zehn Jahren vor. Dabei muss eine langfristige Interpretierbarkeit der Daten realisiert werden. Darüber hinaus ist die Integrität und Authentizität der Forschungsdaten über den gesamten Prozess und die anschließende Archivierung sicherzustellen.

In diesem Zusammenhang erarbeiten die Universität Kassel, die Physikalisch Technische Bundesanstalt Braunschweig und die Gesellschaft für wissenschaftliche Datenverarbeitung mbH, Göttingen im Rahmen des DFG-geförderten Verbundprojekts „Beweissicheres elektronisches Laborbuch“ (BeLab)[1], wie eine beweiswerterhaltende Langzeitarchivierung von digitalen Forschungsprimärdaten und dazugehörigen Metadaten erreicht werden kann. Ziel ist die Vollständigkeit der im Forschungsprozess entstandenen Daten zu erhalten.

[1] Siehe auch http://www.belab-forschung.de.

Dokumentation des Forschungsprozesses

Die Dokumentationsweise des wissenschaftlichen Forschungsprozesses ist im Bezug auf den Wissenschaftler sehr individuell. Dies liegt nicht nur an den unterschiedlichen Forschungsrichtungen und Forschungsverbünden, sondern auch an der Organisation und Arbeitsweise des einzelnen Wissenschaftlers selbst.

Der Forschungsprozess lässt sich grob in fünf Phasen einteilen, siehe Abbildung 1.

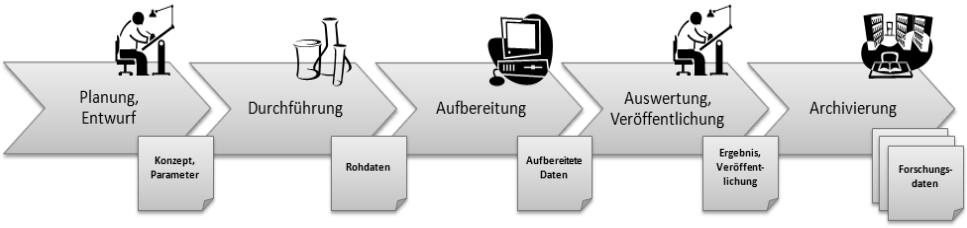


Abbildung 1. Phasen des Forschungsprozesses

Zu Beginn erfolgt eine Planungsphase, in der das Experiment geplant und Parameter festgelegt werden. Die Durchführung, Aufbereitung und Auswertung des Versuchs ist fachspezifisch. Es werden Messgeräte verwendet, deren digitaler Output (Rohdaten) in der Datenmenge stark variiert. Teilweise sind Daten, die von Messgeräten ausgegeben werden, bereits vorverarbeitet, i. d. R. werden jedoch die durch den Versuch erhaltenen Rohdaten vor der weiteren Betrachtung aufbereitet. Beispielsweise werden Bilder, die am Windkanal des Max-Planck-Instituts für Dynamik und Selbstorganisation aufgezeichnet werden, um Turbulenzen zu analysieren, durch High Performance Computer (HPC) mittels Simulationsberechnungen vorverarbeitet. Abschließend erfolgt die Archivierung der Forschungsdaten, so dass die Daten für eine spätere Nachnutzung oder zur Darlegung von Forschungsergebnissen weiterhin genutzt werden können.

Das klassische Laborbuch in Papierform ist auch noch heute im wissenschaftlichen Prozess aufzufinden. Als großer Vorteil gegenüber der elektronischen Erfassung kann die flexible Dokumentationsweise, wie das schnelle Schreiben von Anmerkungen, genannt werden. Das Führen eines Laborbuchs ausschließlich in Papierform ist jedoch kaum noch möglich. Ein hybrides Laborbuch, also ein Nebeneinander von Aufzeichnungen in papierener und in elektronischer Form, birgt aber Gefahren. So könnten beispielsweise Daten von Mitarbeitern nach ihrem Ausscheiden vom Forschungsinstitut verloren gehen, wenn nicht alle Daten zentral gespeichert wurden.

Ein großer Vorteil elektronischer Laborbücher ist aber gerade die zentrale Datenhaltung. So können Daten und Dokumentation einheitlich gepflegt und später für die Nachnutzung anderer Forscher freigegeben werden. Ist das Vorgehen der Dokumentation an den Forschungsprozess angepasst, lassen sich weitere Effizienzsteigerungen erzielen. Ein weiteres Problem besteht darin, dass großen Datenmengen relevante Metadaten (z. B. Zeit und Ort des Experiments) zugeordnet werden müssen. Elektronische Laborbücher lösen dieses Problem durch die Verknüpfung und vereinfachte Pflege dieser Daten.

Problemfeld (Langzeit-)Archivierung digitaler Forschungsdaten

Forschungsdaten bilden die Grundlage wissenschaftlichen Erkenntnisgewinns. Die nachhaltige Sicherung und Bereitstellung von Forschungsdaten dient nicht nur der Prüfung früherer Ergebnisse, sondern auch der Erzielung künftiger Ergebnisse (so auch Allianz der deutschen Wissenschaftsorganisationen [AdW] (2010, 24. Jun.)). Ein großer Teil dieser Daten wird jedoch von Forschenden oder Arbeitsgruppen nicht angemessen dauerhaft archiviert und ist daher einer späteren Wiederverwertung nicht mehr zugänglich. Untersuchungen ergaben, dass diese Daten beispielsweise in selbstdefinierten Verzeichnisstrukturen lokal oder auf Servern abgelegt werden. Berichtet wurde auch von Wechseldatenträgern, die während Experimenten, die außerhalb des Instituts durchgeführt wurden, Verwendung fanden. Hinzu kommt, dass mit Hilfe moderner wissenschaftlicher Methoden Daten in enormem Umfang erzeugt werden, aber adäquates informationsfachliches Knowhow sowie die erforderlichen Infrastrukturen für die digitale Langzeitarchivierung nicht ausreichend zur Verfügung stehen.

Neben dem Erhalt der Daten, d. h. der auf geeigneten Datenträgen gespeicherten Bits, wird im Rahmen der Langzeitarchivierung von einer möglichst langlebigen Interpretierbarkeit der Daten gesprochen. Trotz der Ungewissheit, welche Datenformate langfristig interpretierbar bleiben, existieren Kriterien, nach denen Formate hinsichtlich der Eignung für eine Langzeitarchivierung bewertet werden können. Ist zum Beispiel die Formatspezifikation frei verfügbar, kann eine Datei unabhängig von der erzeugenden Software interpretiert werden. Standards und die Möglichkeit der Angabe von Metadaten im Dateiformat fördern des Weiteren eine spätere Interpretation der Daten. Dagegen kann nach Ludwig (2010) die Komplexität eines Datenformats und evtl. Schutzmechanismen, wie Kopierschutz oder Verschlüsselung, die Interpretationsfähigkeit erschweren

Problemfeld Beweiswerterhaltung

Sind Datensicherung und Interpretierbarkeit gewährleistet, ist aus Sicht der Beweiswerterhaltung die Datenauthentizität und -integrität von großer Bedeutung. Gerade im digitalen Umfeld sind Manipulationen leicht durchzuführen und ohne besondere Maßnahmen nicht nachzuvollziehen. Gleiches gilt für die Verwendung von Metadaten, um die Authentizität eines Dokuments nachzuweisen. Allgemein kann festgehalten werden, dass durch den Nachweis eines lückenlosen und vollständig geführten Laborbuchs die Richtigkeit von Forschungsdaten und -ergebnissen untermauert werden kann.

Dies ist zu beachten, wenn die Archivierung der im eLab festgehaltenen Daten auch im juristischen Sinne Verwendung finden soll. So kann das Laborbuch in Rechtsstreitigkeiten als Beweismittel Einsatz finden und während eines Gerichtsverfahrens als Urkunde, Objekt des Augenscheins oder zur Untermauerung von Sachverständigen- und Zeugenaussagen dienen. Es sind verschiedene Szenarien denkbar, in denen ein Laborbuch als Beweismittel im Rechtsstreit eingeführt werden könnte. Auf das Laborbuch kann es z. B. in universitären Disziplinarverfahren, beim Patentrechtsstreit (so z. B. nach Entscheidungen des LG Düsseldorf (30.04.2002) und des Europäischen Patentamts (20.08.1997)), im Zulassungsbereich, beim Urheberrechtsstreit und, nach Tiedemann (2010), bei Verfahren um die Vergabe von Fördergeldern als Beweismittel ankommen. So könnte Streit um die Echtheit des Labor-

buchs und der darin eingetragenen Ergebnisse herrschen. Das Laborbuch könnte auch als Beweismittel für Forschungs- oder Messungsergebnisse dienen (so für Messungsergebnisse geschehen in einer des Entscheidung des Bundesfinanzhof (02.02.1984)).

Bei der Verwendung eines eLabs stellt sich dabei das Problem, dass im Zivilprozessrecht elektronische Dokumente nach § 371 Zivilprozessordnung (ZPO) nur als Objekte des Augenscheins gelten und der freien Beweiswürdigung durch das Gericht unterliegen. Dagegen können herkömmliche Laborbücher als Urkunden gewertet werden. Der Urkundsbeweis ist aber das stärkste Beweismittel im Prozessrecht, denn das Gericht ist bei seiner Entscheidung zum Beweiswert von Urkunden an die Beweiswertungsregeln der ZPO gebunden (vgl. §§ 415 ff. ZPO). Privaturkunden begründen danach den Beweis dafür, dass die in ihnen enthaltenen Erklärungen von den Ausstellern abgegeben worden sind. Öffentliche Urkunden, die z. B. von einer Behörde ausgestellt wurden, begründen sogar den vollen Beweis über die beurkundeten Tatsachen. In einem Gerichtsverfahren will man daher möglichst auf die Urkunde als Beweismittel zurückgreifen können.

Lösungsmöglichkeiten im BeLab-Projekt: Elektronische Signaturen und Zeitstempel

Den vorgenannten Problemfeldern der digitalen Langzeitarchivierung und Beweiswerterhaltung kann durch den Einsatz geeigneter Sicherheitstechniken und Verfahren entgegengewirkt werden.

Zur Sicherstellung der Datenauthentizität und -integrität von elektronischen Daten können fortgeschrittene oder qualifizierte elektronische Signaturen nach dem Signaturgesetz (SigG) verwendet werden. Denn nach § 371a ZPO sind elektronische Dokumente auch als Urkunden zu werten, wenn sie mit einer qualifizierten elektronischen Signatur versehen sind. Die Verknüpfung des eLabs mit qualifizierten elektronischen Signaturen könnte deswegen die Abhilfe eines der Probleme der Beweiswerterhaltung bringen. Nach den Vorstellungen des Gesetzgebers ersetzt also nur die qualifizierte Signatur die eigenhändige Unterschrift. Dort aber, wo es primär auf die Integrität der erhobenen Rohdaten ankommt, kann möglicherweise auf die qualifizierte elektronische Signatur zugunsten einer, meist kostengünstigeren, fortgeschrittenen Signatur verzichtet werden. Weiter ist es durch die Verwendung von Messgeräten, die bereits über eine Signaturkomponente verfügen, möglich die Daten über den gesamten wissenschaftlichen Forschungsprozess (von der Datenerhebung bis zur Archivierung) durch Signaturen zu sichern.

Ein weiteres Verfahren zur Steigerung des Beweiswertes sind elektronische Zeitstempel. Zeitstempel sind die authentische und unverfälschbare Verknüpfung von Daten mit einer Zeitaussage. Sie dienen dazu, den unveränderten Zustand eines elektronischen Dokumentes von einem bestimmten Zeitpunkt an nachweisen zu können. Im Forschungsprozess spielen Zeitstempel an mehreren Stellen eine besondere Rolle, z. B. beim Experiment selbst, beim Authentifizierungsprozess, bei Änderungen an den Daten (Korrigieren oder gar Stornieren), bei der Übersignatur und bei der Langzeitarchivierung.

Das im Rahmen des BeLab-Projektes entwickelte Konzept für eine beweissichere Langzeitarchivierung von Forschungsprimärdaten nutzt die im Vorfeld beschriebenen etablierten Si-

cherheitstechniken. Basierend auf dem entworfenen Konzept soll ein Prototyp (im folgenden BeLab-System genannt) entwickelt werden, der anschließend sowohl aus technischer als auch juristischer Sicht evaluiert wird.

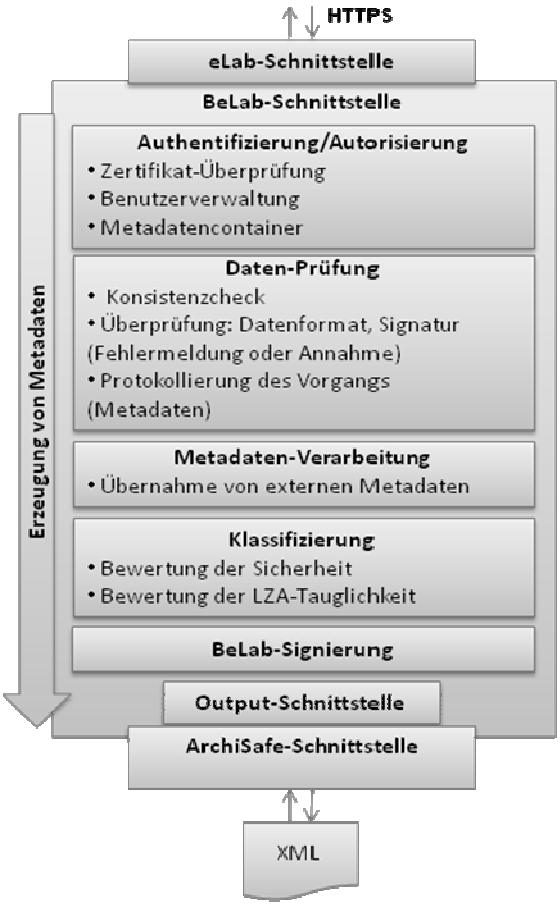


Abbildung 2. Arbeitsweise des BeLab-Systems

Das BeLab-System überprüft in mehreren Schritten die vom Benutzer übergebenen Daten, siehe Abbildung 2. Dabei werden sowohl bereits verwendete Signaturen auf Gültigkeit, Datensammlungen auf Konsistenz und das Datenformat auf die Eignung zur Langzeitarchivierung überprüft. Die abschließende Archivierung basiert auf der technischen Richtlinie des Bundesamtes für Sicherheit in der Informationstechnik [BSI] TR-03125. Ziel ist, die zu archivierenden Forschungsdaten in der Form aufzubereiten, dass sie dem ArchiSafe-Modul der BSI TR-03125 zur Ablage im Archiv übergeben werden können. Durch die Verwendung des ArchiSafe-Moduls in Kombination mit dem zur BSI TR-03125 gehörigen ArchiSig- und Krypto-Modul ist eine langfristige Archivierung inklusive der eingesetzten Signaturen sichergestellt. BSI TR-03125, ArchiSafe und ArchiSig übernehmen die Spezifikationen der RFC 4810 „Long-Term Archive Service Requirements“ und RFC 4998 „Evidence Record Syntax“, der Internet Engineering Task Force [IETF] und basieren damit auf international anerkannten Standards.

Fazit

In Forschung und Praxis liegen die Forschungs- und Messdaten in zunehmendem Maße digital vor. Dadurch werden eine (teilweise) elektronische Dokumentation und in jedem Fall eine digitale Archivierung notwendig. Maßnahmen, die für eine langfristige Authentizitäts- und Integritätssicherung für Laborbücher in Papierform entworfen wurden, können daher nur noch teilweise durchgesetzt werden.

Da die Anforderungen an eine elektronische Dokumentation im starken Maße vom Forschungsbereich und dem Wissenschaftler selbst abhängig sind, ist ein generischer Ansatz zur Gewährleistung einer beweissicheren Langzeitarchivierung von Forschungsprimärdaten sinnvoll.

Durch fortgeschrittene und qualifizierte Signaturen in Kombination mit Zeitstempeln kann eine elektronische Datenauthentizität und -integrität gewährleistet werden. Um einen möglichst hohen Beweiswert zu erzielen, muss der Forschungsprozess in allen Stufen betrachtet werden.

Ziel des BeLab-Projektes ist der Entwurf eines Prototypen, der die ganzheitliche Beweiswerterhaltung ausgehend vom Messgerät bis hin zum digitalen Archiv sicherstellt. Durch den Einsatz des Prototyps kann zukünftig das elektronische Laborbuch äquivalent zu herkömmlichen Laborbüchern als Beweismittel verwendet werden.

Literaturverzeichnis

Allianz der deutschen Wissenschaftsorganisationen [AdW]. (2010, 24. Jun.). *Grundsätze zum Umgang mit Forschungsdaten vom 24. Juni 2010.* Zu finden unter http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaetze/ – Stand vom 23.12.2010.

Bundesamt für Sicherheit in der Informationstechnik [BSI]. (2009). Technische Richtlinie 03125 – *Vertrauenswürdige elektronische Langzeitspeicherung*, zu finden unter https://www.bsi.bund.de/ContentBSI/Publikationen/TechnischeRichtlinien/tr03125/index_htm.html – Stand vom 23.12.2010.

Bundesfinanzhof [BFH]. Urteil vom 02.02.1984, Az. VII R 73/79.

Ebel, H.F., Bliefert, C. (2009). *Bachelor-, Master- und Doktorarbeit: Anleitungen für den naturwissenschaftlich-technischen Nachwuchs* (4. Aufl.). Weinheim : Wiley-VCH.

Ebel, H.F., Bliefert, C., Greulich, W. (2006). *Schreiben und Publizieren in den Naturwissenschaften* (5. Aufl.). Weinheim : Wiley-VCH.

Europäisches Patentamt [EPA]. Entscheidung vom 20.08.1997, Az. T0886/1993.

Holmes, F.L., Renn, J. & Rheinberger H.J. (Hrsg.). (2003). *Reworking the bench: research notebooks in the history of science*. Doderecht, NL : Kluwer Academic Publishers.

LG Düsseldorf. Urteil vom 30.04.2002, Az. 4aO95/01.

Ludwig, J. (2010). Formate – Auswahlkriterien. In Neuroth, H., Oßwald, A., Scheffel, R., Strathmann, S. & Jehn, M. (Hrsg.), *Nestor Handbuch – Eine kleine Enzyklopädie der digitalen Langzeitarchivierung* (Version 2.3) (Kap.7.3), zu finden unter http://nestor.sub.uni-goettingen.de/handbuch/ – Stand vom 23.12.2010.

Internet Engineering Task Force [IETF] (2007). RFC 4810. *Long-Term Archive Service Requirements*. Wallace, C., Pordesch, U. & Brandner. R., zu finden unter http://datatracker.ietf.org/doc/rfc4810/ – Stand vom 23.12.2010.

Internet Engineering Task Force [IETF] (2008). RFC 4998. *Evidence Record Syntax*. Gondrom, T., Brandner. R. & Pordesch, U., zu finden unter http://datatracker.ietf.org/doc/rfc4998/ – Stand vom 23.12.2010.

Signaturgesetz (SigG) vom 16. Mai 2001 (BGBl. I S. 876), das zuletzt durch Artikel 4 des Gesetzes vom 17. Juli 2009 (BGBl. I S. 2091) geändert worden ist.

Tiedemann, P. (2010), Entzug des Doktorgrades bei wissenschaftlicher Unlauterkeit, *Zeitschrift für Rechtspolitik*, 2, 53-55.

Zivilprozessordnung (ZPO) in der Fassung der Bekanntmachung vom 5. Dezember 2005 (BGBl. I S. 3202; 2006 I S. 431; 2007 I S. 1781), die zuletzt durch Artikel 3 des Gesetzes vom 24. September 2009 (BGBl. I S. 3145) geändert worden ist.

Das Deutsche Textarchiv: Vom historischen Korpus zum aktiven Archiv

Alexander Geyken, Susanne Haaf, Bryan Jurish, Matthias Schulz, Jakob Steinmann, Christian Thomas & Frank Wiegand

Berlin-Brandenburgische Akademie der Wissenschaften
Deutsches Textarchiv
Jägerstraße 22/23
D-10117 Berlin
redaktion@deutschestextarchiv.de

Zusammenfassung

Das Deutsche Textarchiv (DTA) macht ein linguistisch annotiertes Volltextkorpus über das Internet frei zugänglich. Die Auswahl beinhaltet sowohl belletristische als auch Fachtexte, die im 17., 18. und 19. Jahrhundert in deutscher Sprache erschienen sind. Das DTA bietet ein vielseitiges Textkorpus für sprachgeschichtliche Forschungen und computerlinguistische Analysen. Durch seine disziplinenübergreifende Zusammensetzung eignet es sich darüber hinaus beispielsweise für fachspezifische, literatur- oder buchwissenschaftliche Untersuchungen. Der Beitrag erläutert die Zusammensetzung des DTA-Korpus, stellt die inhaltlichen und technischen Hintergründe des Projekts vor und gibt einen Ausblick auf die geplante Entwicklung des DTA zu einem aktiven Archiv.

Einleitung

Das Deutsche Textarchiv (DTA, www.deutschestextarchiv.de), ein von der Deutschen Forschungsgemeinschaft (DFG) gefördertes Projekt an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW), stellt ein disziplinenübergreifendes, linguistisch annotiertes Volltextkorpus deutschsprachiger Texte bereit. Grundlage hierfür bilden digitale Faksimiles historischer Druckwerke, die im Zeitraum von ca. 1650 – 1900 erschienen sind. Die Bilddigitalisate und die darauf basierenden elektronischen Volltexte sind über die Website des Projekts öffentlich verfügbar. Die Nutzer des DTA können zwischen einer Bild- und einer Textansicht wählen, wobei letztere in HTML oder einem strukturell reicheren, auf den Empfehlungen der Text Encoding Initiative (TEI P5) beruhenden XML-Format angeboten wird.

In der ersten Projektphase (Juli 2007 bis November 2010) wurden etwa 700 Druckbände der Zeit zwischen 1780 und 1900 im Volltext digitalisiert (dies entspricht einem Umfang von etwa 500 Millionen Zeichen). In seiner zweiten Phase wird das Projekt bis zum Jahr 2014 weitere 650 Texte des 17. und 18. Jahrhunderts im Volltext digitalisieren. Diese werden auf der DTA-Website in regelmäßigen Abständen veröffentlicht. Zum Abschluss von Phase 2 wird das Projekt somit mehr als 1.300 historische Drucke eigendigitalisiert haben und als strukturierte elektronische Volltexte anbieten können.

Zusammensetzung des Korpus

Die Auswahl von Texten für das DTA fand unter sprachwissenschaftlich-lexikographischen Gesichtspunkten statt. In das Korpus wurden Werke aufgenommen, die in der Geschichte der (deutschsprachigen) Literatur oder für die Entwicklung wissenschaftlicher Disziplinen einflussreich waren und die intensiv rezipiert wurden. Neben solchen, als kanonisch geltenden Werken wurden auch einige weniger bekannte Texte berücksichtigt, um die Ausgewogenheit des Korpus zu erhöhen. Die Textauswahl bietet ein großes Spektrum an Genres der literarischen und wissenschaftlichen Produktion.

Zur Zusammenstellung des Korpus wurden im Vorfeld des Projekts zunächst einschlägige Literaturgeschichten und (Fach-)Bibliographien ausgewertet. Parallel dazu wurde das umfangreiche Quellenverzeichnis des *Deutschen Wörterbuchs* von Jacob und Wilhelm Grimm konsultiert.[1] Einzelne Titel, die sich bei der Erstellung bzw. Fortführung des *Deutschen Wörterbuchs* als lexikographisch besonders ergiebig erwiesen haben, wurden der Auswahl hinzugefügt. Ergebnis der Recherchen war eine umfangreiche ‚Basisliste' deutschsprachiger Werke des 17., 18. bzw. 19. Jahrhunderts. Diese Basisliste wurde kommentiert und ergänzt von Mitgliedern der BBAW sowie externen Spezialisten, die sie im Hinblick auf ihr jeweiliges Fachgebiet prüften.

Um den historischen Sprachstand möglichst unverfälscht zu erfassen, wurden die in deutscher Sprache herausgegebenen Erstausgaben der jeweiligen Werke herangezogen. Der Begriff „Erstausgabe" bezieht sich im Zusammenhang mit dem DTA in der Regel auf die *erste gedruckte, selbstständige Publikation eines Textes*. Von diesem Prinzip wurde nur in Einzelfällen abgewichen, etwa dann, wenn eine seltene Erstausgabe eines bestimmten Werks aus konservatorischen Gründen nicht digitalisiert werden konnte (vgl. auch Duntze & Hartwieg, 2010, pp. 63 – 64) oder wenn eine spätere Ausgabe von der Forschung als maßgebliche Fassung des betreffenden Werks angesehen wird – z. B. eine vom Autor überarbeitete und vermehrte Ausgabe oder eine Ausgabe letzter Hand.

Vom gedruckten Buch zum elektronischen Volltext

Das Deutsche Textarchiv arbeitete in der ersten Projektphase vornehmlich mit der Staatsbibliothek zu Berlin, der Herzog August Bibliothek Wolfenbüttel, der Niedersächsischen Staats- und Universitätsbibliothek Göttingen und weiteren, kleineren Bibliotheken zusammen, die Exemplare aus ihren Beständen zur Verfügung stellten. Neben diesen über die erste Projektphase hinaus bestehenden Kooperationen sollen weitere für Phase 2 etabliert werden.

Die Digitalisierung erfolgt in den Räumlichkeiten der Bibliotheken, entweder durch die Bibliotheken selbst oder durch Dienstleister. Über die eigens vom DTA beauftragten Digitalisierungen hinaus profitiert das Projekt von den wachsenden digitalen Sammlungen der Bibliotheken, deren Bilddigitalisate im Rahmen entsprechender Vereinbarungen übernommen werden können. Der im DTA erstellte Volltext steht anschließend auch den Kooperationspartnern zur Verfügung.

[1] Hierbei wurde eng mit der Arbeitsstelle „Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm" zusammengearbeitet, die ebenfalls an der BBAW beheimatet ist.

Die Bilddigitalisate der historischen Drucke bilden die Grundlage für die projektinterne Vorstrukturierung, die mit Hilfe eines vom DTA entwickelten ‚ZoningTools'[2] vorgenommen wird. Damit können Strukturcharakteristika der Seiten, wie beispielsweise Überschriften, Illustrationen oder Marginalien gekennzeichnet werden. Anschließend werden die Texte im sogenannten Double-Keying-Verfahren durch einen Dienstleister erfasst. Dabei werden strukturelle und typographische Merkmale der Vorlage in einem gegenüber dem TEI-Zielformat reduzierten Markup von den abtippenden Personen ausgezeichnet. Die Konvertierung in das Zielformat TEI P5 erfolgt weitestgehend automatisch durch Skripte, die vom DTA entwickelt wurden. Projektmitarbeiter des DTA führen abschließend eine Qualitätskontrolle der Dokumente durch.

Die Bereitstellung der Volltexte im TEI-P5-Format soll deren Nachnutzung erleichtern, beispielsweise als Basis für die Herstellung von textkritischen Ausgaben. Daneben erleichtern die standardisierten Metadaten die Verknüpfung der Texte mit anderen Korpusbeständen.

Bei der computerlinguistischen Aufbereitung, die auf den Texten im strukturierten TEI-P5-Format aufsetzt, werden die Texte in Worteinheiten zerlegt und diese auf ihre jeweilige Grundform zurückgeführt (lemmatisiert). Historische Schreibweisen, die von der heutigen Orthographie abweichen, werden auf die gegenwartssprachliche Variante abgebildet (vgl. Jurish, 2010). Diese Zuordnung ist nicht nur wortbasiert, sondern bezieht durch eine kontextsensitive Disambiguierung auf der Basis eines dynamischen Hidden Markov Models auch noch den engeren Satzkontext mit ein (Jurish, forthcoming). Dadurch wird die Genauigkeit dieser Zuordnung erhöht. Durch diese Analyseschritte ist das Korpus schreibweisentolerant und orthographieübergreifend durchsuchbar. Darüber hinaus wird das DTA-Korpus in einem automatisierten Verfahren hinsichtlich der Wortarten annotiert und für den Substantivbereich mit Thesaurusinformationen verknüpft. Grundlage hierfür bilden der Part-of-Speech Tagger moot (vgl. Jurish, 2003) und die konzeptbasierte lexikalische Begriffshierarchie LexikoNet.[3] Sämtliche linguistische Annotationen geschehen im Standoff-Verfahren. Auf ‚Normalisierungen' des historischen Ausgangstexts wird somit gänzlich verzichtet.

Die Suchmaschine DDC, die für das *Digitale Wörterbuch der deutschen Sprache* (www.dwds.de) entwickelt wurde, ermöglicht neben der gezielten Suche nach einer bestimmten Zeichenkette auch komplexe Suchanfragen sowie die Suche nach flektierten Formen. Die Suche im DTA kann über den gesamten verfügbaren Bestand ausgeführt oder wahlweise auf einzelne Strukturelemente wie Überschriften oder Fußnoten, auf bestimmte typographische Besonderheiten oder einzelne Bände begrenzt werden. Text und Bilddigitalisat wurden vorab in einem automatisierten Verfahren wortweise verknüpft, so dass Treffer von Suchanfragen auch auf dem Digitalisat sichtbar werden.

[2] ZOT – ZoningTool zur Makrostrukturierung von Bilddigitalisaten, vgl. www.deutschestextarchiv.de/documentation/resources/#part_3.

[3] Vgl. www.dwds.de/erschliessung/LexikoNet. In der zweiten Projektphase des DTA soll diese Erschließung durch die Nutzung von Normdaten wie der Gemeinsamen Normdatei (GND) der Deutschen Nationalbibliothek ausgebaut werden.

Ausblick: Das DTA als aktives Archiv

Unter dem Stichwort ‚aktives Archiv' wird das DTA in den kommenden Jahren weiterentwickelt werden. Die Nutzer des DTA sollen die Möglichkeit bekommen, noch intensiver mit dem Datenbestand zu interagieren.

Bereits jetzt ist es möglich, persistente Lesezeichen auf Bildausschnitte der Digitalisate zu setzen, um diese als Belegstellen in einem persönlichen Account zu speichern und mit eigenen Kommentaren zu versehen. Eine vergleichbare Funktion für die Volltexte ist vorgesehen: Markierte Passagen sollen mitsamt der seitengenauen bibliographischen Information im eigenen Bereich gespeichert, annotiert und kommentiert werden können. So bleiben die Fundstellen in dem umfangreichen Korpus überschaubar, und das kooperative Annotieren von Texten in Seminar- und Forschungsgruppen wird dadurch erleichtert. Belegstellen von Text und Bild sollen per E-Mail verschickbar und problemlos exportierbar sein.

Im Bereich der Metadaten bietet das DTA drei verschiedene Schnittstellen an. Literaturmanagementprogramme werden durch das Angebot von COinS (ContextObjects in Spans, http://www.ocoins.info/) unterstützt. Mit dieser Methode zur Einbindung von bibliographischen Metadaten in HTML-Seiten können die Metadaten des DTA automatisch von den weit verbreiteten Programmen Citavi oder Zotero gespeichert werden. Zum Austausch von Autoreninformationen bietet das DTA das relativ junge Format PND-BEACON an. Über diese Schnittstelle können alle im DTA vertretenen Personen zusammen mit der Anzahl ihrer Werke im DTA abgerufen werden. Durch PND-BEACON wird somit eine effiziente Vernetzung mit anderen Projekten möglich, die wie das DTA personenbezogene Daten erfassen. Schließlich bietet das DTA eine OAI-PMH-Schnittstelle an. Dadurch wird der DTA-Bestand demnächst auch über Portale wie Europeana (www.europeana.eu) oder die derzeit im Aufbau befindliche Deutsche Digitale Bibliothek (www.deutsche-digitale-bibliothek.de) recherchierbar sein.

Erweiterung des Korpus

In der zweiten Projektphase soll der aus Eigenmitteln digitalisierte Bestand, das ‚DTA-Basiskorpus', durch Kooperationen mit anderen Projekten angereichert werden, um dem Ziel, das DTA zu einem repräsentativen Querschnittskorpus der deutschen Sprache von 1650 – 1900 auszubauen, näher zu kommen. Hierfür wird derzeit eine umfangreiche Dokumentation des sogenannten DTA-Basisformats erstellt. Dieses enthält die vollständige Liste der TEI-P5-Elemente, die im Projekt verwendet werden, sowie die semantische Beschreibung der Verwendungsweisen dieser Elemente in den Texten des DTA. Zahlreiche Beispiele aus dem DTA-Projektkontext werden diese Dokumentation ergänzen. Darüber hinaus sollen Konvertierungsroutinen zur Verfügung gestellt werden, mit denen Dritte ihre bereits digitalisierten Texte in das DTA-Basisformat überführen können. Die bereits angesprochene Nachnutzung von Digitalisaten aus Bibliotheksbeständen wird die textuelle Anreicherung flankieren.

Literaturverzeichnis

Duntze, O., Hartwieg, U. (2010). Über den Kanon hinaus. Das Deutsche Textarchiv kooperiert mit der Staatsbibliothek zu Berlin. Bibliotheksmagazin H. 1/2010, 62 – 66.

Jurish, B. (2010). Comparing canonicalizations of historical German text. Proceedings of the 11th Meeting of the ACL Special Interest Group on Computational Morphology and Phonology (SIGMORPHON), 72-77, Uppsala, Sweden, 15 July 2010. Retrieved 20 December, 2010, from www.aclweb.org/anthology/W10-2209.

Jurish, B. (2003). A Hybrid Approach to Part-of-Speech Tagging. Final Report, Projekt Kollokationen im Wörterbuch, BBAW, Berlin. Retrieved 20 December, 2010, from http://www.ling.uni-potsdam.de/~moocow/pubs/dwdst-report.pdf.

Jurish, B. (forthcoming). More than words: Using token context to improve canonicalization of historical German. To appear in Journal for Language Technology and Computational Linguistics (JLCL), 2011. See http://www.jlcl.org.

Digitale Medienarchivierung und Lieddokumentation in Verbundsystemen

Digitalisierung, Publikation und Langzeitarchivierung historischer Tonaufzeichnungen des Deutschen Volksliedarchivs (DVA) im Projekt „Datenbank- und Archivierungsnetzwerk Oberrheinischer Kulturträger“ (DANOK)

Gangolf-T. Dachnowsky

Deutsches Volksliedarchiv
Silberbachstr. 13
D-79100 Freiburg i.Br.
gangolf.dachnowsky@dva.uni-freiburg.de

Zusammenfassung

Der Umgang mit digitalen Daten wird aus archivalischer Sicht als direkte Fortsetzung der klassischen Medienarchivierung betrachtet. Ein Blick in deren Historie verdeutlicht, dass üblicherweise keine Materialien – aufgrund fehlender Erfahrungen bezüglich deren Langzeitbeständigkeit – von der Archivierung ausgenommen wurden. Deshalb sollte heute ein Fehlen von fachlich schlüssigen und unter den gegebenen ökonomischen Bedingungen praktikablen digitalen Langzeitarchivierungskonzepten nicht als Argument für eine archivalische Sackgasse herhalten. Vielmehr ist dieses Problem als Teil der konservatorischen Bemühungen im Bereich der Bestandserhaltung zu betrachten und damit zeitlich nachgeordnet in Folge der Archivierung – im konkreten Bedarfsfall – zu bewältigen. Das Deutsche Volksliedarchiv hat sich innerhalb eines EU-Projekts diesem Aufgabenfeld angenommen. Die Ausgangslage, die erarbeiteten Konzepte und Ergebnisse sowie deren Auswirkungen auf die zukünftige Methodik der Medienarchivierung beschreibt der folgende Aufsatz.

DVA – Das Institut

Als Institut für internationale Popularliedforschung – eine freie und selbständige Forschungseinrichtung des Landes Baden-Württemberg – widmet sich das Deutsche Volksliedarchiv der Erforschung populärer Musikkulturen in Vergangenheit und Gegenwart. Dabei arbeitet ein multidisziplinäres Team aus u. a. Musikethnologen, Musikwissenschaftlern, Historikern, Germanisten sowie Informations- und Kommunikationswissenschaftlern an liedbezogenen Fragestellungen in den Bereichen Forschung und Dokumentation. Das Deutsche Volksliedarchiv wurde im Jahr 1914 von dem Germanisten und Volkskundler Prof. Dr. John Meier (1864–1953) gegründet. Von Beginn an lag ein Schwerpunkt der Forschungstätigkeit in der Edition populärer Lieder. Diese findet heute ihre digitale Fortsetzung in der wissenschaftlichen Popularlied-Edition „Populäre und traditionelle Lieder – Historisch-kritisches Liederlexikon“ unter www.liederlexikon.de. Das Institut verfügt über eine umfangreiche Fachbibliothek und verschiedene Sammlungen zur populären Musikkultur, vom traditionellen Volkslied bis zum Musical. Ebenfalls bedeutend ist die breitangelegte Lieddokumentation – zu großen Teilen die Projektdokumentation des historischen Editionsvorhabens. Die Dokumentation wird heute sukzessive auf eine Erschließung in die etablierten Verbund-

systeme umgestellt: die Katalogisierung in der Tonträger- und Liederbuchdokumentation in den OPAC des Südwestdeutschen Bibliotheksverbundes, sowie in den Kalliope-OPAC zum Nachweis bzw. zur Erschließung der Nachlässe und Autographen. Auf vorhandene Digitalisate wird gegebenenfalls direkt aus diesen Katalogen heraus verlinkt (Deutsches Volksliedarchiv, 2010b).

DANOK – Das Projekt

In den vergangenen Jahren konnte am Institut die Digitalisierung der gesamten Bestände musikethnologisch bedeutender Tonaufzeichnungen auf Magnetbändern abgeschlossen werden. Die wertvolle Sammlung mit ihren über 14.000 Liedbelegen auf ca. 600 Stunden Audiomaterial enthält eine Vielzahl einmaliger Tondokumente des deutschsprachigen Populariedes. Deren Digitalisierung hatte unter konservatorischen Gesichtspunkten höchste Priorität: Materialtypische Alterungsprozesse der Magnetbänder gefährden die dauerhafte Reproduzierbarkeit der akustischen Informationen auf diesen Primärquellen.

Ein Großteil der Digitalisierungsmaßnahme wurde mit Geldern der Europäischen Union im INTERREG-Projekt „Datenbank- und Archivierungsnetzwerk Oberrheinischer Kulturträger“ (DANOK) innerhalb einer Kooperation regionaler trinationaler Projektpartner realisiert. Der Fokus lag dabei – neben der Onlinepräsentation der singulären Liedaufzeichnungen – auf den archivspezifischen und informationstechnischen Besonderheiten von Audioaufnahmen (Datenbank- und Archivierungsnetzwerk Oberrheinischer Kulturträger – DANOK, 2007). Archivalische Digitalisierungsworkflows für andere Medientypen haben sich mittlerweile etabliert: Text- und/oder Bildinformationen werden in einem sogenannten Hybridverfahren zusätzlich als analoge Information auf geeigneten Medien – in der Regel Mikrofilm – langzeitarchiviert. Im Bereich der Audiodaten fehlt eine solche Möglichkeit der redundanten Sicherung auf ein physisch archivierbares Material mit darauf abgelegter analoger Information. Von besonderem Interesse für Archive erscheint deshalb ein innovativer Ansatz des am Projekt beteiligten Instituts für Physikalische Messtechnik IPM zur Archivierung von digitalen Audiodaten auf Farbmikrofilm (Hofmann, 2005; Riedel, Hofmann & Fraunhofer IPM Freiburg, 2007) – ein Forschungsvorhaben, das innerhalb der Projektlaufzeit umfassend weiterentwickelt werden konnte. Dabei werden „die digitalen Daten [...] mit moderner Lasertechnologie auf langzeitstabilem Polymerfilm ausbelichtet. Der Belichter schreibt die digitalen Daten als Helligkeitsstufen auf das analoge Filmmedium; die unterscheidbaren Abstufungen repräsentieren dabei die digitalen Null- und Einser-Codes.“ (Fraunhofer-Institut für Physikalische Messtechnik IPM, 2008).

Vom Findbuch in den Verbund-OPAC – von der Magnetaufzeichnung in das digitale Langzeitarchiv

Abbildung 1: Sichere Langzeitarchivierung – Digitale Daten auf Film (Fraunhofer-Institut für Physikalische Messtechnik IPM, 2008)

Nachdem der Ansatz des Fraunhofer-Instituts für Physikalische Messtechnik IPM zur Langzeitarchivierung digitaler Daten auf Farbmikrofilm innerhalb der Projektlaufzeit zwar umfänglich weiterentwickelt werden konnte, sich eine ökonomisch vertretbare Lösung aber noch nicht abzeichnete, wurden die Audiodigitalisate an das Bibliotheksservice-Zentrum Baden-Württemberg (BSZ), das auch den SWB-OPAC betreut, übergeben und dort in die digitalen Archivsysteme übernommen (Bibliotheksservice-Zentrum Baden-Württemberg, 2009; Wolf, Schweibenz & Mainberger, 2009; Wolf, 2010). Die Metadaten der Digitalisate erstellte ein Dienstleister durch das Transkribieren der tabellarischen Findbucheinträge in ein projektspezifisches XML-Format.

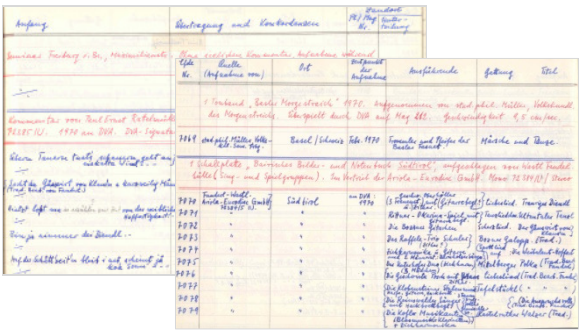


Abbildung 2: Doppelseite aus dem Findbuch der Tonträger

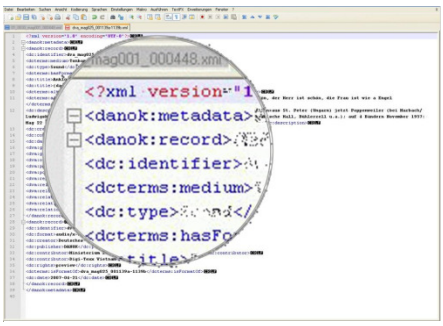


Abbildung 3: Transkription der Findbucheinträge im XML-Format des DANOK-Projekts

Die Publikation dieser Metadaten erfolgte über deren Migration in den OPAC des Südwestdeutschen Bibliotheksverbundes (SWB). Verlinkungen aus den einzelnen Katalogisaten heraus verweisen direkt auf die Downloadadressen der Audiofiles. Damit sind diese kulturhistorisch wichtigen Tondokumente über die gängigen Bibliotheksportale recherchierbar und in neuere nationale wie internationale Vorhaben – zu nennen sind hier derzeitig insbesondere die Deutsche Internetbibliothek sowie die Europeana – ohne zusätzlichen Aufwand integrierbar. In gleicher Art und Weise wird bei den Folgeprojekten verfahren, so z. B. bei der vom Ministerium für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg geförderten umfänglichen Digitalisierung von Mikrofilmausgaben historischer Liederbücher, Liedflugschriften und handschriftlicher Liederbücher aus den Beständen des DVA.

PPN:	316001082
	[Elektronische Ressource]
Aufsatz:	Hohenauer **Jodler**
In:	Feldforschungsaufnahmen des Deutschen Volksliedarchivs, Mag322 / Deutsches Volksliedarchiv <Freiburg, Breisgau>. - Freiburg i. Br. : Deutsches Volksliedarchiv. - Online-Ressource, Lied 10906
Anmerkung:	MP3. - Hrsg.: Univ. Freiburg. - Liedgattung: **Jodler** Liedanfang: Hohenauer **Jodler** Aufn.: Haufenreith [Steiermark, Gerichts-Bezirk Weiz], 1975-10-04
Link zum Volltext:	Elektronische Ressource: Zugang zum Digitalisat (OPAC-Hinweis: Tondokument. Schnitt aus **Magnettonband** 322)
Art und Inhalt:	Musiktondokument
RVK-Notation:	LV 51000 INFO *Ähnliche Literatur*
Lokale Notation:	[vergeben von Deutsches Volksliedarchiv: <Frei 99>]
	A 0211491

Abbildung 4: Darstellung der in den OPAC des Südwestdeutschen Bibliotheksverbundes migrierten Daten mit deren direkter Verlinkung auf das digitalisierte Tondokument (Deutsches Volksliedarchiv, 2010a).

Medienarchivierung gestern, heute und morgen

Weitere Digitalisierungsprojekte sind in Arbeit bzw. in der Planung, so aktuell die Digitalisierung der Feldforschungsaufnahmen auf Edison-Zylindern. Dabei wird zunehmend deutlich, dass die Problematik der Langzeitverfügbarkeit von Audioaufnahmen keine Besonderheit des digitalen Zeitalters ist, sondern sich schon bei den vielfältigen Medienbrüchen des vergangenen Jahrhunderts zeigte. So scheint zum heutigen Zeitpunkt das tonarchivalische Wissen speziell im Umgang mit Edison-Walzen und Schellackplatten akut gefährdet, in nicht allzu ferner Zukunft werden uns, aus der archivalischen Perspektive, diese Problem wohl im Umgang mit WAV- und MP3-Dateien ereilen.

So steht die Medienarchivierung durch die moderne Informationstechnik vor völlig neuen Herausforderungen und Möglichkeiten. Anzumerken ist dabei allerdings, dass Probleme wie die Heterogenität von Formaten, schnell aufeinander folgende Innovationen bei den Informationsträgertechnologien, das Überspielen (d. h. Migrieren) auf aktuelle Medien in die entsprechenden neuen Formate auf dem Gebiet der Tonträgerarchivierung schon vor dem digitalen Zeitalter existierten. Dabei zeigte sich bereits historisch eine unter dem Aspekt der Bestandserhaltung fortschreitende Komplexität: Waren Schellackplatten für die Archivierung noch nahezu ideal, so sind Magnetaufzeichnungen durch ihre materialspezifischen Zerfallsprozesse schon heute akut gefährdet, CDs traut man nur eine Haltbarkeit im Bereich von Jahrzehnten zu, DVDs eine noch geringere.

Damit kann das in mehr als einem Jahrhundert gewachsene tonarchivalische Wissen im Umgang mit heterogenen Medien sicherlich den Blick auf die Probleme der aktuellen Informationstechnik bereichern. Dabei ist die Herangehensweise in Schallarchiven zu Teilen geprägt von einem gewissen Pragmatismus, so im Bereich der Archivierung, beim Emulieren der Abspieltechnik und dem Erstellen von Benutzerkopien, d. h. der archivalische Geschäftsgang verbietet es Medien von der Archivierung auszunehmen, nur weil deren Langfristverfügbarkeit nicht gewährleistet werden kann. Gleichzeitig können wir selbstverständlich nur die zurzeit bestmögliche verfügbare Technik zum Abspielen verwenden und nicht die für die Zukunft beste denkbare. Weiterhin werden Benutzerkopien zur Schonung des

Originals angelegt und nicht um dieses entbehrlich zu machen. Genau hier endet der Pragmatismus des Archivars: Das Original ist nicht ersetzbar! Was im Bereich historischer Drucke als selbstverständlich gilt, sollte es ebenfalls bei multimedialen Medien – gleichgültig ob analog oder digital – sein. Auch wenn Magnetbänder Zerfallsprozesse zeigen, sind sie weiterhin geeigneten bestandserhaltenden Maßnahmen zuzuführen. Keiner mag heute prognostizieren, mit welchen technischen Mitteln und wissenschaftlichen Fragestellungen Forscher in der Zukunft an diese Materialien herantreten werden. Diese archivalische Selbstverständlichkeit bereichert die Diskussion um die Langzeitarchivierung digitaler Daten: Wir müssen die Daten in ihren Ursprungsformaten erhalten, migrierte Daten sind lediglich als Benutzerkopien geeignet, die Emulation geeigneter Techniken zum Zugriff auf die Originaldaten sollte zu Forschungszwecken jederzeit möglich sein.

Fazit

Die Medienarchivierung beschäftigt sich zum gegebenen Zeitpunkt, unter den bestehenden technischen Möglichkeiten, im ökonomisch realisierbaren Rahmen mit der bestmöglichen Sicherung von Information und Medium. Dabei war und ist die Langzeitarchivierungsperspektive analoger wie digitaler Medien zum Zeitpunkt ihrer Archivierung nicht zwingend geklärt, sondern als Teil der bestandserhaltenden Bemühungen zu betrachten.

Eine Migration von Daten ohne Erhalt der Primärquelle ist aus archivalischer Sicht abzulehnen. Einer Emulation technischer Möglichkeiten zur Reproduktion der Information auf den Primärquellen ist immer der Vorzug zu geben. Migrierte Daten können ein pragmatisches Mittel zur Erstellung von Benutzerkopien sein (aktuell z. B. der Onlinezugriff auf komprimierte Dateien). Die Verfügbarkeit der Primärquelle muss für wissenschaftliche Zwecke jedoch jederzeit gewährleistet bleiben.

Um diesen Zielsetzungen umfassend gerecht werden zu können, ist eine weitergehende Kooperation und Vernetzung zwischen den Archiven, Bibliotheken und Dokumentationseinrichtungen notwendig. Vieles ist schon auf den Weg gebracht: die Nutzung gemeinsamer Katalogisierungsstandards, die Verwendung internationaler Formate, die Katalogisierung in die Verbund-OPACs sowie deren Vernetzung über Portale und virtuelle Kataloge. Ein weiterer Bedarf besteht nunmehr im kooperativen Aufbau und Betrieb digitaler Langzeitarchive. Das „Kompetenznetzwerk Langzeitarchivierung – nestor“ (Deutsche Nationalbibliothek, 2010) und das Projekt „kopal – Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen“ (Niedersächsische Staats- und Universitätsbibliothek Göttingen, 2010) bieten schlüssige Konzepte, deren Umsetzungen sicherlich in absehbarer Zukunft den einzelnen Medienarchiven zur Verfügung stehen werden.

Literatur

Bibliotheksservice-Zentrum Baden-Württemberg (2009). Langzeitarchivierung von Liedbelegen. Zugriff am 11.10.2010 unter http://www2.bsz-bw.de/cms/digibib/pjktboa090128/.

Datenbank- und Archivierungsnetzwerk Oberrheinischer Kulturträger – DANOK (2007). Datenbank und ArchivierungsNetzwerk Oberrheinischer Kulturträger – Banque de données et réseau d'archivages culturels du Rhin Supérieur. Zugriff am 11.10.2010 unter http://www.danok.eu/.

Deutsche Nationalbibliothek (2010). Willkommen bei nestor. Zugriff am 11.10.2010 unter http://www.langzeitarchivierung.de/.

Deutsches Volksliedarchiv (2010a). Katalog des Deutschen Volksliedarchivs. Zugriff am 11.10.2010 unter http://pollux.bsz-bw.de/DB=2.316/CMD?ACT=SRCHA&IKT=12&SRT=YOP&TRM=316001082&MATCFILTER=N&MATCSET=N&NOABS=Y.

Deutsches Volksliedarchiv (2010b). Das Deutsche Volksliedarchiv – Institut für internationale Popularliedforschung. Zugriff am 11.10.2010 unter http://www.dva.uni-freiburg.de/.

Fraunhofer-Institut für Physikalische Messtechnik IPM (2008). Sichere Langzeitarchivierung: Digitale Daten auf Film. Zugriff am 11.10.2010 unter http://www.ipm.fhg.de/fhg/ipm/extra/bigimg/laserbelichtung/bits_on_film/digitsgraustufengr.jsp.

Hofmann, A. (2005). Digitale Farbmikrofilmbelichtung. In H. Walravens, Newspapers in Central and Eastern Europe, IFLA publications 110, 207-215.

Niedersächsische Staats- und Universitätsbibliothek Göttingen (2010). kopal – Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen. Zugriff am 11.10.2010 unter http://kopal.langzeitarchivierung.de/.

Riedel, W. J., Hofmann, A., Fraunhofer IPM Freiburg (2007). ArchivLaser® – Langzeitarchivierung auf Farbmikrofilm. In: Zeitschrift für Bibliothek, Archiv und Information in Norddeutschland, 27/1, 115-120.

Wolf, S., AG Langzeitarchivierung des Ministeriums für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg (2010). Langzeitarchivierung digitaler Objekte in Baden-Württemberg – Ein Schichtenmodell der Kompetenzen, Funktionen, Dienstleistungen und Schnittstellen. Zugriff am 11.10.2010 unter http://opus.bsz-bw.de/swop/volltexte/2010/818/.

Wolf, S., Schweibenz, W., Mainberger, Ch. (2009). Langzeitarchivierung am Bibliotheksservice-Zentrum Baden-Württemberg – Konzept, Aktivitäten und Perspektiven. Zugriff am 11.10.2010 unter http://www.zlb.de/aktivitaeten/bd_neu/heftinhalte2009/Techn020309BD.pdf.

Neue Forschungsfelder im Netz. Erhebung, Archivierung und Analyse von Online-Diskursen als digitale Daten

Olga Galanova & Vivien Sommer

Technische Universität Chemnitz
Institut für Medienforschung
Thüringer Weg 11
Chemnitz
olga.galanova@phil.tu-chemnitz.de, vivien.sommer@phil.tu-chemnitz.de

Zusammenfassung

Dieser Beitrag beschäftigt sich mit den Fragen, was die mediale Spezifität von Online-Diskursen als digitale Phänomene ausmacht sowie welche theoretischen und methodischen Zugänge ausgearbeitet werden sollen, um die Multimodalität, die Interaktivität und die Hypertextualität webbasierter Kommunikation medienadäquat zu erfassen.

Einleitung

In unserem Forscherteam an der TU Chemnitz untersuchen wir in einem von der DFG geförderten Projekt themenspezifische Diskussionen in ihrer besonderen Ausprägung im Internet. Dabei begreifen wir Online-Kommunikation über gesellschaftliche Themen als Diskurse. Mit diesem Vorhaben stellen sich zahlreiche neue methodologische und methodische Fragen, welche in der bisherigen Diskursanalyse noch keine Rolle gespielt haben, da man sich in der Regel an den ‚klassischen' Massenmedien als Leitmedien orientiert hat. Bei der Analyse von netzbasierter Kommunikation ist man jedoch mit einer spezifischen Medialität konfrontiert, wodurch es erforderlich wird ein Methodeninstrumentarium zu entwickeln, um Online-Diskurse untersuchen zu können. In unserem Beitrag behandeln wir einige der Fragen bzw. Probleme, die sich durch die Analyse von digitalen Daten als Diskurse ergeben und erläutern mögliche Lösungsstrategien, die wir im Rahmen des Projektes erarbeitet haben.

Was ist ein Online-Diskurs?

Online-Diskurse charakterisieren wir als internetbasierte Kommunikation über öffentliche Themen, die in gesamtgesellschaftliche Diskurse integriert sind. Das Web kann dabei als ein Resonanzraum eines Diskurses beschrieben werden, in dem Themen zumeist aus traditionellen Massenmedien aufgenommen und in Teilöffentlichkeiten des Netzes, wie etwa in Blogs, Wikis oder Social Networks diskursiv weiterverarbeitet werden. Öffentliche Themen werden so crossmedial, im Zusammenspiel von on- und offline-Medien konstituiert und diskursiv bearbeitet (Fraas, Pentzold 2008). Doch auch wenn die internetbasierte Kommunikation im gesamtgesellschaftlichen Diskurs integriert ist, betrachten wir sie aufgrund ihrer medialen Spezifik als einen eigenen Forschungsgegenstand, der ein besonderes diskursanalytisches Vorgehen erfordert (vgl. Fraas & Pentzold & Meier 2010).

Charakteristisch für Online-Diskurse ist, dass das spezifische Handeln von Individuen bzw. Gruppen fassbar ist. Während Akteure in massenmedialen Diskursen weniger als Individuen erkennbar sind, sondern in institutionalisierten Rollen hinter makrostrukturellen Medienorganisationen zurück treten, können sie im Rahmen von Online-Diskursen auf der Mikroebene in interpersonalen Interaktionsprozessen agieren und Wirkungen auf der makrostrukturellen Ebene erzielen (vgl. Fraas & Pentzold 2008; Schmidt 2006).

Ein weiteres Merkmal von Online-Kommunikation, welches in die Analyse miteinbezogen werden muss, ist die zunehmende Multimodalität. Multimodal wird im Sinne von Kress und van Leeuwen (2001) als soziokulturell geprägtes, kommunikatives Zusammenwirken unterschiedlicher Zeichenmodalitäten verstanden. Diskursive Praxis ist generell als multimodales Zeichenhandeln zu beschreiben. Durch die Digitalisierung der Medien lassen sich visuelle Zeichen freier und einfacher gestalten und darstellen. Dadurch ist die kommunikative Relevanz von Bildern und anderen visuellen Ausdrucksmöglichkeiten (Diagramme, Layout, Typographie) noch weiter vorangeschritten (vgl. Meier 2008).

Angesichts der genannten Eigenschaften von Online-Diskursen, ist es wichtig ein methodisches Instrumentarium zu konzipieren, welches die speziellen Verfahren der Inhaltsproduktion im Netz einbezieht. Ziel der Analyse ist es Deutungsmuster zu rekonstruieren. Deutungsmuster werden in Anlehnung an Keller (2005) als grundlegende bedeutungsgenerierende Schemata verstanden, die diskursiv verbreitet werden und dabei sowohl auf den gesellschaftlich verfügbaren Wissensvorrat rekurrieren als auch neue Muster generieren. Wir erarbeiten die Deutungsmuster eines Online-Diskurses mittels des Vergleichs der verschiedenen Diskursfragmente. Für die Erarbeitung der beschriebenen Deutungsmuster ist die Konzeption eines Analyse-Korpus daher grundlegend.

Wie lassen sich online-medienadäquate Analyse-Korpora erstellen?

Bei der Erstellung eines Untersuchungs-Korpus stellen sich erste methodische Herausforderungen. Weil es im Online-Bereich im Gegensatz zum Print-Bereich, neben der unregelmäßigen Publikationspraxis (noch) keine institutionalisierte Archivierungspraxis gibt, müssen bestehende Methoden der Datenerhebung erweitert werden. Das hat zur Folge, dass man als Forscher mit dem Problem der Dezentralität konfrontiert ist, d. h. Themen entwickeln sich weniger systematisch und entfalten sich in unterschiedlichen Teilöffentlichkeiten (vgl. Meier 2008). Daher ist es umso wichtiger, plausible Kriterien für die Entscheidung, welche Online-Diskursfragmente in den Analyse-Korpus aufgenommen werden sollen, zu entwickeln. Zudem ist es sinnvoll in der Online-Daten-Erhebung eine Auswahl typischer Fälle vorzunehmen. Ein sehr hilfreicher Ansatz für diese Forschungsstrategie bieten die Methoden der qualitativen Sozialforschung, insbesondere die Grounded Theory, welche in den 1960er Jahren von Barney Glaser und Anselm Strauss (2005) begründet wurde. Sie systematisierten die Auswahl relevanter Fallbeispiele nach dem Prinzip des theoretical samplings (Glaser & Strauss 2005; Strauss 1998; Strauss & Corbin 2010). Mittels dieses Prinzips werden zum einen Daten aus möglichst ähnlichen und zum anderen aus möglichst unterschiedlichen Quellen herangezogen, bis bei der Datenerhebung keine neuen inhaltlichen Aspekte mehr auftauchen. Um diese Fälle bzw. Diskursfragmente miteinander vergleichen zu können, ist die Bestimmung geeigneter Auswahlkriterien notwendig. Besonders zu Be-

ginn des Forschungsprozesses ist es jedoch schwierig, diese Kriterien zu entwickeln, da das Vorwissen über den Gegenstand nicht besonders differenziert ist, so dass Glaser und Strauss für das Sampling-Verfahren vorschlagen, mit einigen „lokalen Konzepten“ zu Eigenschaften und Konstitution des Forschungsgegenstandes zu beginnen (Glaser & Strauss 2005: 53). Diese Einstiegsstrategie ist nach unserem Verständnis mit Strategien frameorientierter Diskursanalysen vergleichbar, in denen Schlüsselausdrücke als thematische Referenzen auf zentrale Diskursstränge recherchiert werden.

Frames verstehen wir zunächst einmal als Repräsentationsformate kollektiver Wissensbestände, die in der Kommunikation die jeweilige Vertextung des Wissens steuern. Diese Begriffsbestimmung ist vergleichbar mit der der Deutungsmuster (siehe oben). Dennoch unterscheiden wir diese beiden Begriffe terminologisch: Frames sind nach unserem Verständnis als Instrumente der Datenauswertung bzw. -erhebung einzusetzen, während Deutungsmuster das Ziel der Analyse sind (vgl. Fraas, Meier & Pentzold 2010). Marvin Minsky (1975) folgend werden Frames als komplexe Strukturen aus Slots (konzeptuelle Leerstellen), Fillers (konkreten kontextbasierten Ausfüllungen dieser Leerstellen) und Default-Werten (inferierten Standartwerten) konzeptualisiert. Zudem können sie als Modelle für das Kontextualisierungspotential der diskursiv zentralen Schlüsselkonzepte verstanden werden. So hat Konerding (1993) aus dem lexikalischen Inventar der deutschen Sprache zehn sogenannte Matrixframes abgeleitet, aus deren Slot-Struktur die Slot-Struktur konkreter Konzeptframes entwickelt werden kann. Deren Slot-Struktur hat Konerding als einen Fragenkatalog konzipiert, durch den man die konkreten Filler ‚erfragt‘.

Das beschriebene Vorgehen möchten wir anhand einer Beispiel-Untersuchung des konkreten Online-Diskurses um die öffentliche Debatte über den Fall des vermeintlichen KZ-Aufsehers John Demjanjuk verdeutlichen. Ihm wird vorgeworfen in dem nationalsozialistischen Vernichtungslager Sobibor an der Ermordung von 27.900 Menschen mitgewirkt zu haben. Über die Auslieferung Demjanjuks von den USA nach Deutschland und den sich daran anschließenden Prozess entspannte sich im Netz eine Auseinandersetzung. Es ging dabei zum einen darum, ob es sich bei dem in Amerika Festgenommenen tatsächlich um den Sobibor-Wachmann handle und zum anderen, ob er aufgrund seines Alters und seines Gesundheitszustandes prozess- und straffähig ist.

Für einen ersten Einstieg in das Diskursfeld und für die Bestimmung möglicher Schlüsselkonzepte des Diskurses wurde ein sogenannter Einstiegstext gewählt. In unserem Beispiel-Diskurs haben wir uns für einen Videobeitrag aus der ARD Mediathek entschieden (Tagesschau, 2010). Ausschlaggebend für die Wahl ist seine Kommunikationsform. Mit dem journalistischen Videoausschnitt ist eine vermeintlich neutrale Perspektive auf den Diskurs gegeben, die wichtige Akteure benennt und relevante Positionen wiedergibt.

Demgemäß lässt sich an dem vorliegenden Beispiel als erstes Schlüsselkonzept DEMJANJUKS GESUNDHEITSZUSTAND bestimmen. Dieser wird im genannten Beitrag im Zusammenhang mit seiner Transportfähigkeit und seiner Auslieferung nach Deutschland thematisiert und das nicht nur auf der sprachlichen Ebene, sondern auch auf der visuellen, indem gezeigt wird, wie Demjanjuk im Rollstuhl vom Krankenwagen abgeholt wird.

Der dann hier in Frage kommende Matrixframe für die Bestimmung der Slotstruktur ist PERSON IN BESONDEREM ZUSTAND. Die Slot-Struktur des jeweiligen Matrixframes besteht wie bereits erläutert aus einer Reihe von Prädikationsfragen, durch deren Beantwortungen man die Leerstellen (Slots) durch konkrete Filler ausfüllen kann.

Für den Matrixframe PERSON IN BESONDEREM ZUSTAND ergeben sich zum Beispiel folgende Fragen (Konerding 1993):

1. Auf welche Art und Weise tritt der Zustand bei der Person auf?

2. Wie ergeht es der Person mit dem Zustand?

3. In welchen Handlungen oder (funktionalen) Zusammenhängen spielt der Zustand der Person eine bestimmte Rolle?

Um weitere potentielle Diskursfragmente für das Korpus auszuwählen orientierten wir uns neben der Slotstruktur des jeweiligen Matrixframes in der Auswahl auch an den unterschiedlichen online-medialen Kommunikationsformen (soziokulturell konventionalisierte Formen), d. h., dass wir sowohl interpersonale Online-Kommunikation anhand von Foren in das Korpus aufnehmen und analysieren als auch maximal kontrastierend dazu massenmediale und mehrfachadressierte Nachrichtentexte der Online-Magazine. Ebenfalls ausgewählt werden Diskursfragmente wie Weblogs, die zwar ähnlich wie die massenmedialen Texte mehrfachadressiert, jedoch weniger institutionell geprägt sind. Ein weiterer Auswahlfokus liegt auf den unterschiedlichen Zeichensystemen und den zeichenhaften Inszenierungspraktiken, die in den kommunikativen Handlungen zur Anwendung kommen. So werden neben sprachlich verfassten Daten auch damit layouttechnisch kombinierte statische Bilder und audiovisuelle Video-Daten analysiert. Ein viertes Auswahlkriterium sind die Akteure bzw. Akteursgruppen, die bestimmten Interessengruppen und Netzwerken angehören und so bestimmte Perspektiven in den Diskurs wiederholt einbringen.

Wie lassen sich Online-Diskurse angemessen archivieren?

Die ausgewählten Online-Diskursfragmente sollen abgespeichert und beständig für die Analyse zugänglich gemacht werden. Dies macht eine Archivierung der im Netz veröffentlichten multimodalen Texte notwendig. Das Vorhaben, die Online-Daten in ihrer ursprünglichen ‚Qualität' aufzubewahren, ist aber eine große Herausforderung. Bei der Speicherung müssen Animationen sowie das Layout einer Website und simultane, auditive Daten in ihrer Ursprungsqualität erhalten werden, denn sie bilden ebenso wie (Schrift-)Sprache diskursive Ausdrucksmöglichkeiten, derer sich die Akteure eines Diskurses bedienen können. Die Abspeicherungsbemühungen wie etwa das Kopieren von Online-Inhalten in Word-Dokumente bietet somit keine adäquate Archivierungsmöglichkeit. Vielmehr ist die Verwendung von spezieller Software zur Archivierung der Web-Inhalte unumgänglich.

Der am weitesten verbreitete Zugriff auf Online-Daten erfolgt über Spiegelungssoftware. Programme wie z. B. Win HTTrack, Offline Explorer und Web Dumper haben die Funktion von sogenannten Webcrawlern, die das manuelle Aufrufen von Webseiten simulieren und

dadurch automatisiert Online-Daten in ein Archiv speichern können.[1] In der konkreten Anwendung von diesen Spiegelungsprogrammen stellen sich für die Online-Diskursanalyse allerdings grundsätzliche Probleme. Denn mit diesen Programmen können keine P2P-Video- und Audiodateien (z. B. von der Videoplattform YouTube) abgespeichert werden. Alles, was im Netz nur ‚gestreamt' wird, müsste zusätzlich aufgenommen und erneut mit den gespeicherten Webseiten verknüpft werden. Auch wenn man diese nachträglichen Verknüpfungen durchführt, lässt sich doch der ursprüngliche Kontext einer Website mit P2P-Streams nicht wiederherstellen.

Ein weiteres Problem besteht in der Abspeicherung von Verlinkungen auf andere Websites. In dem Versuch, die Verlinkungen einer Website ebenfalls zu archivieren, besteht die Gefahr riesige Datenmengen, wenn nicht sogar das ganze Netz zu archivieren. Um dies zu vermeiden, bieten alle Spiegelungsprogramme den Einstellungsparameter der ‚Tiefe' an. Mittels dieser Einstellung gibt man an, bis zu welchem ‚Grad' verlinkte Seiten abgespeichert werden. Bei einer ‚Tiefe' von null wird nur die angegebene Webseite gespeichert, bei einer ‚Tiefe' von eins alle Seiten die auf der Ausgangsseite verlinkt sind und bei einer ‚Tiefe' von zwei auch wiederum die verlinkten Seiten der verlinkten Seiten der ‚Tiefe' eins. Bei der starken Vernetzung des World Wide Web kann dies bereits bei einer Tiefen-Angabe von zwei zu einer Speicherung erheblicher Datenmengen führen, die kaum zu überblicken sind und außerdem die Archivierungskapazitäten mit irrelevanten Daten belasten.

Diese Schwierigkeiten bei der Speicherung von Online-Daten als Objekte außerhalb ihres Nutzungskontextes weisen auf einen interessanten Umstand hin, nämlich dass Online-Daten keine fertigen, dinghaften Produkte sind. Eine der wichtigsten Eigenschaften von Online-Diskursen ist ihre Interaktivität, die dafür sorgt, dass die Bedeutungs- und Sinnkonstruktionen nur im Nutzungsprozess möglich sind. Dazu gehören unter anderem Prozesse des Wartens auf das Herunterladen, des Spielens mit dynamischen Menüs und Animationen der Aktivierung der Verlinkungen zu anderen Seiten sowie die Möglichkeit zu scrollen oder ein laufendes Video anzuhalten.

Als eine sehr gute Alternative zu den beschriebenen Spiegelungsprogrammen hat sich für uns die Archivierung der Daten mittels des Screencapturings erwiesen. Dabei zeichnet eine Software die visuellen Signale der Grafikkarte und die auditiven Signale der Soundkarte eines Computers auf. Also all das was ein Nutzer auf dem Bildschirm sehen und aus den Audiogeräten hören kann. Die Aufzeichnung kann in Form eines Videos mit der Screencapturing-Software aufgenommen und in unterschiedliche Formate exportiert werden. Durch die Methode des Abfilmens ist es also möglich visuelle Muster einer Webseite zu erhalten und die verschiedenen Modalitäten animiert und synchron zu speichern und darzustellen.

[1] Für den Austausch über die technischen Möglichkeiten von Webcrawlern bedanken wir uns bei Eric Kanold und Fabian Wörz.

Welche Analyse-Methode bietet sich für die Diskursanalyse von Online-Daten an?

Mit dem nächsten Schritt soll angedeutet werden, wie die abgespeicherten digitalen Online-Daten als Diskurse analysiert werden können. Für die Analyse von Online-Diskursen verwenden wir das Verfahren der Multimodalitätsanalyse, die Stefan Meier als Kombination von semiotischer Bild- und Sprachanalyse entwickelt und in seiner Studie über Online-Diskussionen zur zweiten deutschen Wehrmachtausstellung angewendet hat (vgl. Meier 2008). Dieses Analyse-Verfahren berücksichtigt den multimodalen Charakter von Online-Diskursen und ermöglicht es, die Bedeutungskonstruktion nicht nur als Produkt von Sprache, sondern im Zusammenhang mit unterschiedlichen visuellen Elementen zu erkennen. Die Fokussierung auf diese Merkmale soll es ermöglichen, die Diskursfragmente in ihrer Bedeutung nicht nur auf der sprachlichen Zeichenebene, sondern auch auf der visuellen Gestaltungsebene zu analysieren. Mittels des genannten Analyse-Verfahrens werden die verschiedenen Zeichenmodalitäten wie Bild und Sprache in ihrer multimodalen Kombination und der sich damit rekonstruierenden Aussagekraft untersucht.

Im Vergleich zu anderen Methoden, die meistens ausgehend von forschungspraktischen Fragestellungen ausgewählt werden, begründet sich das multimodalitätsanalytische Vorgehen aus der Datenqualität heraus, nämlich aus der Charakteristik des Mediums, der Komplexität und der Zeichenhaftigkeit der Daten selbst, noch bevor diese an ein bestimmtes analytisches Schema angepasst werden. Die Bestimmung von medialen Potenzialen bzw. Bedingungen der Kommunikation sind somit erste analytische Schritte, die den ‚Grad' der möglichen Multimodalität definieren.

Aus diesem Grund erfolgt unsere Beispiel-Analyse an einem einzelnen Diskursfragment des Analyse-Korpus. Obwohl es mit der Multimodalitätsanalyse immer möglich ist Standardformen und typische Darstellungsstrategien der diskursiven Praktiken sowie prototypische, verfestigte und sich rationalisierte Techniken zu untersuchen, sollte man mit der Eigenständigkeit eines Beispiels reflexiv umgehen.

Für die Demonstration haben wir eine Forumsdiskussion zum Thema die „Zehn meistgesuchten Kriegsverbrecher" ausgewählt, die sich im Forum „Community für Freunde Serbiens" entwickelt hat (Community4um, 2010).

Die Abspeicherung der zeichenhaften Komplexität der Kommunikation auf diesem politischen Forum ist eine zentrale Bedingung für die Vorbereitung der Daten für die Multimodalitätsanalyse. Denn bei der Gestaltung eines Beitrags steht den Diskutanten dieses Forums nicht nur eine Möglichkeit zur Verfügung, die eigenen Standpunkte sprachlich zu formulieren und Inhalte durch unterschiedliche Schriftarten hervorzuheben, sondern auch Bilder und Videos einzufügen.

So zum Beispiel bedankt sich der erste Teilnehmer bei dem zweiten für das Posting der Liste der „Kriegsverbrecher" und setzt die Diskussion fort, indem er die Porträts der genannten Personen einfügt. Interessant ist dabei, dass der zweite Teilnehmer die Porträts nicht in die Reihenfolge einordnet, die schon im ersten Beitrag vorgegeben ist. Zuerst werden die Porträts von den Personen eingefügt, dessen Namen im vorherigen Beitrag fett getippt

wurden. Die visuelle Hervorhebung der Namen durch eine fette Schrift und durch eine Einordnung in die Liste der Nazi-Kriegsverbrecher bildet dabei eine besonders wirksame Strategie, die Relevanz von bestimmten Inhalten hervorzuheben.

Die Art und Weise, wie diese Diskussion durch eine Kombination von Sprache und Bild gestaltet wird, lässt ein mögliches Deutungsmuster einer Schuldzuschreibung erkennen, das eine wesentliche Rolle für die Entwicklung weiterer Diskussionen um Demjanjuk im Deutungskontext spielt. Anhand dieser kommunikativen Sequenz lässt sich zudem zeigen, dass auch Farben und Formen als Ressourcen für die Sinnrekonstruktion wichtig sind. Sie sollen im Zusammenhang mit dem Kontext des Diskurses und dem Medium seiner Entfaltung beschrieben werden.

Die Entwicklung der Diskussion um die Schuldzuschreibung ist mit dem Thema der Strafe verknüpft. In seinem weiteren Beitrag drückt der zweite Teilnehmer seine Enttäuschung bezüglich dessen aus, dass die Kriegsverbrecher immer noch frei sind. Ein Smiley in seinem Beitrag zeigt ein Klopfen mit dem Kopf gegen eine Steinwand. Seine Enttäuschung erläutert er am Beispiel von Demjanjuk, wodurch das Thema um Demjanjuk in der weiteren Forumsdiskussion zentral wird. Anhand von zwei Videos über die Auslieferung Demjanjuks aus den USA nach Deutschland stellt er dar, wie schwierig es sein kann, ein Urteil über den vermeintlichen KZ-Aufseher zu fällen.

Das erste Video, in dem Demjanjuk als gesundheitlich schwach präsentiert wird, wird als „aua aua wie Krank und Transportunfähig doch Demjanjuk ist“ kommentiert. Als Antwort auf dieses Video steht das zweite Video welches die Funktion hat, die scheinbar tatsächliche Lage zu präsentieren, nämlich dass Demjanjuk sich relativ sicher und selbständig bewegen kann. Durch die Betitelung des zweiten Videos als „hier bewegt Demjanjuk sich recht gut …“ weist der Forumsteilnehmer auf die Diskrepanzen hin, zwischen dem was Demjanjuk scheinbar als schlechten Gesundheitszustand inszeniert und wie es ihm ‚tatsächlich‘ ginge. In der Begleitung eines lachenden Smiley bekommt diese Betitelung den Status einer moralischen Beurteilung von Demjanjuk als Lügner zu, der bei einer Unaufrichtigkeit ertappt wurde.

Anhand von diesem Beispiel ist zu demonstrieren, welche multimodalen Tools die Diskussionsteilnehmer wie nutzen, um Deutungsmuster der Schuldzuschreibung im Fall von Demjanjuk zu konstituieren und zu inszenieren. Die Erschließung von diesem multimodal gestalteten Deutungsmuster ermöglicht es, die Abstraktion von der Mikro- auf die Makro-Ebene des gesellschaftlichen Diskurses zu realisieren.

Zusammenfassung

Anhand der vorangegangenen Analyse des Online-Diskurses haben wir zu zeigen versucht, dass und wie Online-Daten als gesellschaftliche Diskurse konzeptualisiert und als spezifisches Objekt wissenschaftlicher Analyse erkannt werden können. Dabei haben wir gezeigt, was die mediale Spezifität von Online-Diskursen als digitale Phänomene ausmacht sowie welche methodischen Zugänge ausgearbeitet werden sollen, um die Multimodalität und die Interaktivität webbasierter Kommunikation adäquat zu erfassen, zu archivieren und zu analysieren.

Deutlich wurde, dass die wissenschaftliche Kontextualisierung von Online-Diskursen nicht bei der Frage der begrifflichen Konzeptualisierung stehen bleiben kann. Denn die Erhebung, Archivierung und Analyse von Online-Diskursen als digitalen Daten bekommen dabei eine ebenso entscheidende Rolle.

Literatur

Busse, Dietrich (2008): Diskurslinguistik als Epistemologie – Grundlagen und Verfahren einer Sprachwissenschaft jenseits textueller Grenzen. In: Warnke, Ingo H.; Spitzmüller, Jürgen (Hg.): Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene ; [Sektion … im September 2006 im Rahmen des 41. Linguistischen Kolloquiums an der Universität Mannheim]. Berlin: de Gruyter (Linguistik – Impulse & Tendenzen, 31), 57–87.

Community4um. (2010). Abgerufen am 8. Oktober, 2010, von http://cccc.community4um.de/f23t818-die-zehn-meistgesuchten-naz1-kriegsverbrecher.html.

Fraas, Claudia; Pentzold, Christian (2008): Online-Diskurse – Theoretische Prämissen, methodische Anforderungen und analytische Befunde. In: Warnke, Ingo H. / Spitzmüller, Jürgen (Hrsg.): Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene, 291-326.

Fraas, Claudia / Meier, Stefan / Pentzold, Christian (2010): Konvergenz an den Schnittstellen unterschiedlicher Kommunikationsformen – Ein Frame-basierter analytischer Zugriff. In: Lehnen, Katrin / Gloning, Thomas / Bucher, Hans-Jürgen (Hg.): Mediengattungen: Ausdifferenzierung und Konvergenz. Frankfurt a. M., New York: Campus. (In Vorbereitung).

Glaser, Barney G.; Strauss, Anselm L. (2005): Grounded theory. Strategien qualitativer Forschung. 2. Aufl. Bern: Huber (GesundheitswissenschaftenMethoden).

Keller, Reiner (Hg.) (2005): Die diskursive Konstruktion von Wirklichkeit. Zum Verhältnis von Wissenssoziologie und Diskursforschung. Konstanz: UVK (Erfahrung, Wissen, Imagination : Schriften zur Wissenssoziologie, Bd. 10).

Konerding, Klaus-Peter (1993): Frames und lexikalisches Bedeutungswissen. Untersuchungen zur linguistischen Grundlegung einer Frametheorie und zu ihrer Anwendung in der Lexikographie. Univ., Diss.-Heidelberg, 1992. Tübingen: Niemeyer (Reihe germanistische Linguistik, 142).

Kress, Gunther; van Leeuwen, Theo (2001): Multimodal discourse. The modes and media of contemporary communication. London: Arnold [u. a.].

Meier, Stefan (2008): (Bild-)Diskurse im Netz. Konzepte und Methode für eine semiotische Diskursanalyse im World Wide Web. Univ., Diss. Chemnitz. Köln: Halem.

Minsky, M. (1975): A framework for representing knowledge. In: Winston, P.H. (Hg.): The psychology of computer vision. New York: McGraw-Hill, 211-277.

Strauss, Anselm (1998): Grundlagen qualitativer Sozialforschung. Datenanalyse und Theoriebildung in der empirischen soziologischen Forschung. München: Fink (UTB für WissenschaftUni-Taschenbücher, 1776, Soziologie).

Strauss, Anselm; Corbin, Juliet (2010): Grounded theory. Grundlagen qualitativer Sozialforschung. Unveränd. Nachdr. der letzten Aufl. Weinheim: Beltz Psychologie Verl.-Union.

Tagesschau. (2010). Abgerufen am 2. November, 2010, von http://www.tagesschau.de/multimedia/video/video538038.html.

Elektronisches Publizieren und Open Access

Open Access – Von der Zugänglichkeit zur Nachnutzung

Heinz Pampel

Helmholtz-Gemeinschaft
Helmholtz Open Access Projekt – Koordinationsbüro
Deutsches GeoForschungsZentrum GFZ
Telegrafenberg
D-14473 Potsdam
pampel@gfz-potsdam.de

Zusammenfassung

Der Beitrag widmet sich, ausgehend von einer knappen Bestandsaufnahme, der Perspektive von Open Access im Kontext einer digitalen Wissensschaft. Dabei wird anhand von Beispielen die Bedeutung der Nachnutzung von Information und Wissen beschrieben.

Einleitung

Mit der Unterzeichnung der „Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen“ haben die deutschen Wissenschaftsorganisationen 2003 die Notwendigkeit einer Publikationsstrategie betont, die unter dem Begriff Open Access fordert, „eine umfassende Quelle menschlichen Wissens und kulturellen Erbes“ über das Internet frei zugänglich zu machen (Berliner Erklärung, 2003). Der Begriff Open Access umfasst heute vielfältige Aktivitäten, die darauf abzielen, Information und Wissen ohne finanzielle, rechtliche und technische Barrieren im Internet auf Basis vertrauenswürdiger Infrastrukturen zugänglich und nachnutzbar zu machen.

Open Access wird aktuell in zwei Strategien umgesetzt: Der sogenannte „Grüne Weg“ des Open Access ermöglicht, in Abhängigkeit der disziplinären Publikationskultur, die Veröffentlichung von Pre- und Postprints in sogenannten Repositorien. Diese Volltextdatenbanken garantieren in ihrer Form als institutionelle Repositorien den freien Zugriff auf die Publikationen einer Institution und in Form von disziplinären Repositorien den freien Zugriff auf Publikationen einer Disziplin.

Der „Goldene Weg“ des Open Access widmet sich der Erstveröffentlichung wissenschaftlicher Publikationen in einer frei zugänglichen Fachzeitschrift. Entsprechend der Herausforderung der Finanzierung dieses Modells haben sich in den letzten Jahren diverse Geschäftsmodelle etabliert, bei denen die Finanzierung auf die Institution des Publizierenden oder des Herausgebenden umgeschichtet wird.[1]

Nach Björk et al. (2010) waren im Jahr 2009 bereits über 20 Prozent aller Artikel, die 2008 in JCR-Zeitschriften[2] erschienen sind, frei zugänglich. Im Oktober 2010 ermöglichen 5.533

[1] Im Rahmen der dynamischen Entwicklung von Open Access wird die ehemals starke Differenzierung zwischen den beiden Strategien immer unbedeutender.

[2] Zeitschriften, die in der Datenbank Journal Citation Reports (JCR) gelistet sind.

Open-Access-Zeitschriften (DOAJ, 2010) und über 1.700 Open-Access-Repositorien (OpenDOAR, 2010) Wissenschaftlern die Publikation von Aufsätzen und anderen, mehrheitlich textuellen Materialien. Über einen Suchdienst wie die Bielefeld Academic Search Engine (BASE) sind 25.518.361 Dokumente frei zugänglich (BASE, 2010).

Die wachsende Bedeutung von Open Access zeigt sich nicht nur an den steigenden Zahlen der Publikationsorgane und Angebote kommerzieller Verlage[3], sie wird auch auf der politischen Ebene deutlich. Die Verankerung von Open Access in der „Digitalen Agenda für Europa“ zeigt die Relevanz des Themas: „Öffentlich finanzierte Forschungsarbeit muss [...] durch frei zugängliche Veröffentlichung wissenschaftlicher Daten und Unterlagen allgemein verbreitet werden.“ (Europäischen Kommission, 2010).

Open Access – Grundlage der digitalen Wissenschaft

Das Potenzial einer digital vernetzten Wissenschaft wird deutlich, wenn Informationsobjekte aller Art im Open Access nicht nur zugänglich („gratis“), sondern auch nachnutzbar („libre“) sind.[4] Konzepte, die unter dem Präfix "E" diskutiert werden, wie E-Science und eResearch, betonen das Potenzial der Vernetzung und Nachnutzung von Texten, Forschungsdaten und anderen Materialien. Ziel ist es, Wissenschaftlern auf Basis virtueller Forschungsumgebungen neue Herangehensweisen an wissenschaftliche Fragestellungen zu ermöglichen, die unter dem sich formierenden vierten Paradigma des wissenschaftlichen Arbeitens gefasst werden (Hey et al., 2009).

Der sich wandelnde Umgang mit Information und Wissen und die steigende Bedeutung der Nachnutzung und Vernetzung von Informationsobjekten soll im Folgenden an drei Beispielen knapp skizziert werden:

Offener Zugang zu Forschungsdaten

Die Forderung nach einem zeitgemäßen Umgang mit wissenschaftlichen Daten beschäftigt die internationale Wissenschaftsgemeinde. In Deutschland hat die Allianz der deutschen Wissenschaftsorganisationen im Rahmen der Schwerpunktinitiative „Digitale Information“ „Grundsätze zum Umgang mit Forschungsdaten“ verabschiedet, in denen sie die „langfristige Sicherung und den grundsätzlich offenen Zugang zu Daten aus öffentlich geförderter Forschung“ fordert (Allianz, 2010). Hintergrund ist das Bestreben der Wissenschaftsorganisationen, die Nachnutzung und die Nachprüfbarkeit wissenschaftlicher Daten zu verbessern, um – wie von der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) gefordert – die Steigerung des gesellschaftlichen Nutzens durch frei zugängliche Forschungsdaten zu ermöglichen (OECD). Aktivitäten wie die Open-Access-Zeitschrift Earth System Science Data (ESSD), die sich der qualitätsgesicherten Publikation von geowissenschaftlichen Forschungsdaten widmet, zeigen Bestrebungen, einzigartige wissenschaftliche Daten für alle Interessierten zur Nachnutzung zugänglich zu machen (Dallmeier-Tiessen & Pfeiffenberger, 2010).

[3] Beispielhaft sei hier das 2010 gestartete Open-Access-Programm SpringerOpen des Verlages Springer Science+Business Media genannt.

[4] Zur Definition von „gratis“ und „libre“ siehe Suber (2008).

Die Forderung nach der Nachnutzung geht in Abhängigkeit der Disziplinen Hand in Hand mit der Forderung nach Open Access. Dabei gilt es die Heterogenität der Disziplinen zu berücksichtigen.[5]

Semantic Publishing – Texte maschinenlesbar machen

Ein weiterer Trend, der die Relevanz der Nachnutzung von Informationsobjekten betont, wird unter dem Begriff Semantic Publishing gefasst. Nach Shotton (2009) beschreibt der Begriff vielfältige Aktivitäten und Techniken, die eine Erweiterung eines Artikels unterstützen: „[...] I define 'semantic publishing' as anything that enhances the meaning of a published journal article, facilitates its automated discovery, enables its linking to semantically related articles, provides access to data within the article in actionable form, or facilitates integration of data between papers."

Die Diskussion um das Semantic Publishing wird insbesondere in den Biowissenschaften diskutiert – eine Disziplin, die durch ihre Nähe zum technologischen Fortschritt als Trendsetter im Bereich des wissenschaftlichen Publikationswesens charakterisiert werden kann. Am Beispiel eines Artikels in der Zeitschrift PLoS Neglected Tropical Diseases (Reis et al., 2008) zeigen Shotton et al. (2009) die Chancen und Herausforderungen der semantischen Anreicherungen einer wissenschaftlichen Textpublikation.

Anreicherungen wie die von Shotton und seinen Co-Autoren sind nur möglich, wenn die verwendeten Ressourcen unter Lizenzen publiziert sind, die eine Nachnutzung ermöglichen.[6]

Aggregation und Mehrwerte

Ein weiteres Beispiel, das das Potenzial der Nachnutzung wissenschaftlicher Inhalte betont, ist das Konzept der „PLoS Hubs" der Public Library of Science (PLoS). Kuratoren aggregieren, vernetzen und erweitern Informationsobjekte im Rahmen des „Hubs": „The vision behind the creation of PLoS Hubs is to show how open-access literature can be reused and reorganized, filtered, and assessed to enable the exchange of research, opinion, and data between community members." (PLoS, 2010). In dem 2010 gestarteten PLoS Biodiversity Hub werden so bereits publizierte Artikel zum Thema Biodiversität, ähnlich dem Konzept des „Overlay Journal" (Brown, 2010), unter einer thematischen Plattform zusammengeführt und darüber hinaus durch zusätzliche Ressourcen wie beispielsweise Taxonomien und graphische Materialien (Fotos und Karten) ergänzt. Das Potenzial der Aggregation und der Schaffung von Mehrwerten rund um einen Artikel zum Thema Biodiversität beschreiben die Kuratoren des Hubs wie folgt: „Increasing linkages and synthesizing of biodiversity data will allow better analyses of the causes and consequences of large scale biodiversity change, as well as better understanding of the ways in which humans can adapt to a changing world" (Mindell, 2010).

[5] Beispielsweise ist der offene Zugang zu Forschungsdaten in den Sozialwissenschaften aufgrund ethnischer und rechtlicher Rahmenbedingungen nur eingeschränkt möglich.

[6] Im Falle des erwähnten Artikels von Reis et al. (2008) wurde der Originalartikel unter der Creative-Commons-Lizenz „Namensnennung" publiziert.

Fazit

Open Access fördert – über die Beseitigung finanzieller, rechtlicher und technischer Barrieren hinaus – die Entwicklung einer Wissenschaftskommunikation, die das Potenzial des Internets für Forschung und Lehre konsequent umsetzt. Die Möglichkeit der Nachnutzung und Vernetzung von Information und Wissen in digitalen Kommunikationsräumen ist Grundlage einer digital arbeitenden Wissenschaft.

Die Schaffung von Rahmenbedingungen, die den offenen Umgang mit Information und Wissen ermöglichen, ist die Herausforderung der kommenden Jahre.

Dank

Der Autor dankt den Kollegen im Koordinationsbüro des Helmholtz Open Access Projekts für anregende Diskussionen: Roland Bertelmann, Kathrin Gitmans, Hans Pfeiffenberger und Paul Schultze-Motel.

Literaturverzeichnis

Allianz der deutschen Wissenschaftsorganisationen. (2010). *Grundsätze zum Umgang mit Forschungsdaten*. Retrieved 20 October, 2010, from http://www.allianzinitiative.de.

Berliner Erklärung. (2003). *Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen*. Retrieved 20 October, 2010, from http://oa.mpg.de.

Björk, B., Welling, P., Laakso, M., Majlender, P., Hedlund, T., & Guðnason, G. (2010). Open Access to the scientific journal literature: situation 2009. *PLoS ONE, 5*(6), e11273. from doi:10.1371/journal.pone.0011273.

Brown, J. (2010). Overlay journals, repositories and the evolution of scholarly communication. *Open Repositories Conference 2010*, Madrid. Retrieved 20 October, 2010, from http://or2010.fecyt.es.

Dallmeier-Tiessen, S., & Pfeiffenberger, H. (2010). Peer Reviewed Data Publication in Earth System Sciences. In C. Puschmann & D. Stein (Eds.), *Towards Open Access Scholarship*. Berlin (6), Open Access Conference. Düsseldorf: Düsseldorf University Press, 77-84. Retrieved 20 October, 2010, from urn:nbn:de:hbz:061-20100722-142254-7.

Europäische Kommission. (2010). *Eine Digitale Agenda für Europa*. Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen (No. KOM(2010) 245). Brüssel: Europäische Kommission. Retrieved 20 October, 2010, from http://ec.europa.eu.

Hey, T., Tansley, S., & Tolle, K. (2009). Jim Gray on eScience: a transformed scientific method. In T. Hey, S. Tansley, & K. Tolle (Eds.), *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond: Microsoft Research, XVII-XXXI. Retrieved 20 October, 2010, from http://research.microsoft.com.

Mindell, D. (04.10.2010). *Aggregating, tagging and connecting biodiversity studies*. The Offical PLoS Blog. Retrieved 20 October, 2010, from http://blogs.plos.org.

Organisation for Economic Co-operation and Development (OECD). (2007). *Principles and Guidelines for Access to Research Data from Public Funding*. Paris: Principles and Guidelines for Access to Research Data from Public Funding. Retrieved 20 October, 2010, from http://www.oecd.org.

PLoS. (2010). *About PLoS Hubs: Biodiversity*. Retrieved 20 October, 2010, from http://hubs.plos.org.

Reis, R. B., Ribeiro, G. S., Felzemburgh, R. D. M., Santana, F. S., Mohr, S., Melendez, A. X. T. O., Queiroz, A., u. a. (2008). Impact of Environment and Social Gradient on Leptospira Infection in Urban Slums. *PLoS Negl Trop Dis*, *2*(4), e228. Retrieved 20 October, 2010, from doi:10.1371/journal.pntd.0000228.

Shotton, D. (2009). Semantic publishing: the coming revolution in scientific journal publishing. *Learned Publishing*, *22*(2), 85-94. Retrieved 20 October, 2010, from doi:10.1087/2009202.

Shotton, D., Portwin, K., Klyne, G., & Miles, A. (2009). Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article. *PLoS Comput Biol, 5*(4), e1000361. Retrieved 20 October, 2010, from doi:10.1371/journal.pcbi.1000361.

Suber, P. (2008). Gratis and libre open access. *SPARC Open Access Newsletter, 124*. Retrieved 20 October, 2010, from http://www.earlham.edu.

Open Access und die Konfiguration der Publikationslandschaften

Silke Bellanger

Zentral- und Hochschulbibliothek Luzern
Sempacherstr. 10
CH-6002 Luzern
silke.bellanger@zhbluzern.ch

Dirk Verdicchio

Institut für Soziologie, Universität Basel
Petersgraben 27
CH-4051 Basel
verdicchio@gmx.net

Zusammenfassung

Der Aufsatz geht von der Beobachtung aus, dass in den Geistes- und Sozialwissenschaften Closed Access als selbstverständliche Publikationsform gilt, Open Access dagegen als eher exotisch angesehen wird. Da das Prinzip des Closed Access den wissenschaftlichen Ansprüchen von Offenheit und Zugänglichkeit entgegen steht, ist der Eindruck der Selbstverständlichkeit von Open Access erklärungsbedürftig. Um die Herstellung dieser Selbstverständlichkeit nachzeichnen zu können, schlagen wir eine Perspektive auf das Feld wissenschaftlicher Publikationen vor, die mit Hilfe der Akteur-Netzwerk-Theorie, die Mittel und Praktiken in den Blick nimmt, mit denen Publikationen in Waren transformiert und deren Qualifizierung routinisiert werden. Eine solche Perspektive hat den Vorteil, dass sie auf Wert- und Moraldebatten verzichtet und stattdessen das Augenmerk auf die alltägliche Praxis von Autoren, Verlagen und Bibliotheken richtet.

Einleitung

Für die wissenschaftliche Praxis scheint Open Access nur Vorteile zu bringen. Open Access-Werke sind mit einem normalen Internetzugang von überall her zugänglich. Die Notwendigkeit auf Zeitschriften über bestimmte IP-Adressen zuzugreifen entfällt ebenso wie die Kosten und die Wartezeiten bei Artikelbeschaffungen. Darüber hinaus, so zeigen einige Studien, scheinen Open Access-Publikationen die Sichtbarkeit von Artikeln und die Anzahl der Zitationen eher zu erhöhen (siehe z. B. Antelmann, 2004). Dennoch sind Open Access-Publikationen in weiten Teilen der Geistes- und Sozialwissenschaften eine eher seltene Publikationsform, während Closed Access für viele Forschende die naheliegende Option für die eigenen Veröffentlichungen zu bleiben scheint (Koch, Mey, & Mruck, 2009). Diese Selbstverständlichkeit von Closed Access ist bemerkenswert, wird hier doch das Ideal einer allgemeinen Zugänglichkeit wissenschaftlichen Wissens durch dessen Verknappung und die Informationsdienstleistung von Bibliotheken durch Warenerwerb ersetzt.

Wir möchten in unserem Beitrag der Frage nachgehen, wie diese Selbstverständlichkeit des Closed Access (im Folgenden mit CA abgekürzt) perpetuiert wird. Wir schlagen dazu vor, Ansätze aus der Wirtschafts- und Wissenschaftssoziologie mit Erfahrungen aus der bibliothekarischen Praxis zu kombinieren. Dies geschieht in zwei Schritten. In einem ersten Schritt werden wir mit Hilfe der Systemtheorie eine Spannung beschreiben, die das Feld wissenschaftlicher Publikationen durchzieht. Diese Spannung entsteht dadurch, dass bei CA-Publikationen zwei Universalitätsansprüche aufeinander treffen: derjenige der Wissenschaft und derjenige der Wirtschaft. In einem zweiten Schritt werden wir dann in Anlehnung an die Akteur-Netzwerk-Theorie (ANT) eine Perspektive vorschlagen, die auf die Mittel und Praktiken aufmerksam macht, die zu einer Fraglosigkeit von CA-Publikationen führen und den Widerspruch zwischen Wirtschaft und Wissenschaft zum Verschwinden bringen.

Da eine vollständige Analyse des Feldes wissenschaftlicher Publikationen den Rahmen dieses Artikels sprengen würde, werden wir uns darauf beschränken, die theoretische Rahmung und die Argumentation eines solchen Vorgehens aufzuzeigen und einige Elemente sichtbar zu machen, die in eine solche Analyse einfließen würden. Dennoch hoffen wir, dass wir damit den Kontroversen über Open Access (OA) eine neue Sichtweise hinzufügen können. Diese würde sich dann dadurch auszeichnen, dass sie die Aufmerksamkeit von den prinzipiellen Diskussionen über Wissenschaftsideale und die Machbarkeit neuer Publikationsformen auf die Praktiken lenkt, die das Feld der wissenschaftlichen Publikationen in seiner gegenwärtigen Gestalt konstituieren.

CA zwischen Wirtschaft und Wissenschaft

Folgt man Rudolf Stichweh (2005) stellt die Universalität eines der wichtigsten Motive der Selbstbeschreibung der Wissenschaft dar. Diese Universalität weist mehrere Dimensionen auf. Eine zentrale Dimension ist die räumliche und zeitliche Unabhängigkeit wissenschaftlicher Wahrheiten, so dass weder der Ort noch das Alter eine als Wahrheit anerkannte Aussage zu beeinflussen vermag. Zudem weist die Universalität der Wissenschaft eine thematische Dimension auf. Diese bedingt, dass die Wissenschaft keine Sachverhalte akzeptiert, die sich ihrem Wahrheitsanspruch zu entziehen vermögen. Für den hier gegebenen Zusammenhang ist jedoch vor allem der Anspruch auf soziale Universalität von Bedeutung. Soziale Universalität besagt, dass sich niemand dem Wahrheitsanspruch der Wissenschaft entziehen kann. D. h., dass eine wissenschaftliche Wahrheit für alle Menschen wahr sein soll. Daher macht es in einer wissenschaftlichen Kommunikation auch keinen Sinn, eine wissenschaftliche Wahrheit anzuerkennen und zugleich auf eine alternative persönliche Wahrheit zu bestehen. Gerade für diesen Anspruch auf soziale Universalität ist die Offenheit von wissenschaftlichem Wissen essentiell: Dieses Wissen soll nicht geheim, sondern für jedermann und jedefrau zugänglich sein, denn nur so ist es möglich den Anspruch auf soziale Universalität aufrecht zu erhalten. Rudolf Stichweh (2003) schreibt dazu:

„If science can claim universality, especially social universality in the sense of presupposing validity of its truth claims for any individual whosoever in the world, then it follows with a certain consequence that access to these universal truths should not be denied to any one of those individuals for whom these truths are supposed to be valid on the first hand. And if

openness is the only standard acceptable in dealing with scientific knowledge then again this openness should be realized for a public of maximum social extension." (p. 212)

Der Anspruch einer allgemeinen Verfügbarkeit wissenschaftlichen Wissen wird vom Prinzip des CA konterkariert, indem es den Zugriff darauf beschränkt. Die Regulierung des Zugriffs auf wissenschaftliches Wissen durch Bezahlung führt ein ökonomisches Kalkül in die Zugänglichkeit dieses Wissens ein. Als Funktionssystem beansprucht zwar auch die Wirtschaft einen Universalismus, doch ist dieser an eine entsprechende Zahlungsfähigkeit geknüpft (Luhmann, 1999). Die Spannung, die dabei entsteht, ist ein Ergebnis zweier systemspezifischer Operationen: Wo die Wissenschaft Wahrheiten produziert, produziert die Wirtschaft Knappheit. Dieser Widerspruch zwischen Offenheit und Verknappung ist es, der uns hier interessiert, denn es ist erstaunlich, dass die Wissenschaft eine Praxis akzeptiert und darüber hinaus zu einem Standard macht, die dem grundlegenden Selbstverständnis entgegen steht. Unsere Ausführungen beruhen auf der These, dass diese Akzeptanz auf einer Form der Arbeitsteilung zwischen Forschenden und Bibliotheken beruht.

Konfiguration des Publikationsmarktes

Daher wollen wir einige Investitionen aufzeigen, die das Konzept des Closed Access von einem Widerspruch in eine Selbstverständlichkeit verwandeln. Mit Hilfe der Akteur-Netzwerk-Theorie lassen sich diejenigen Elemente beschreiben, die das Feld wissenschaftlicher Publikationen so konfigurieren, dass daraus eine privilegierte Position für CA-Inhalte erwächst.

Der Prozess der Kommodifizierung

Die Verwandlung von wissenschaftlichem Wissen in eine Ware beruht auf Prozessen, bei denen Objekte und Akteure definiert und stabilisiert werden. Denn, so Michel Callon (2006), „[m]an wird nicht als Ware geboren, man wird dazu" gemacht (p. 553). Diese Prozesse, die dafür verantwortlich sind, dass sowohl ein ökonomisches Gut als auch ein Markt entstehen können, werden in der ANT „Rahmung" und „Singularisierung" genannt. Die Operation der Rahmung besteht darin, distinkte Entitäten zu schaffen, die voneinander unabhängig und miteinander unverbunden sind und dabei diejenigen Elemente, die nicht in Betracht gezogen werden (sollen), außen vor zu lassen. Ein einfaches Beispiel für die Rahmung einer Markttransaktion sieht Callon beim Tausch von Eigentumsrechten an einem Auto: Dabei identifiziert er drei unabhängige und unterscheidbare Komponenten: den Käufer, den Produzent/Verkäufer und das Auto. Die Unabhängigkeit von Verkäufer und Käufer ist in diesem einfachen Beispiel evident und bereitet keine Probleme. Anders ist das jedoch bei dem Automobil. Indem dieses den Besitzer wechselt, wird auch das Wissen und die Technologie des Herstellers weitergegeben. Das macht es notwendig, dass es dekontextualisiert und jede Bindung zu seinem Ursprung gekappt wird. Erst dadurch kann es als ein ökonomisches Gut angesehen und behandelt werden (Callon, 1998, p. 16ff.).

Man sieht sofort, dass die Identifikation der relevanten Komponenten bei wissenschaftlichen Publikationen nicht ganz so einfach ist. Dies bereits schon dadurch, dass Produzent und Verkäufer von wissenschaftlichen Artikeln nicht identisch sind. Unserer Meinung nach müsste man, wenn man die einzelnen Komponenten einer Markttransaktion von wissen-

schaftlichen Artikeln identifizieren möchte, mindestens zwischen dem Produkt (dem Artikel), dem Produzent (dem Forscher), dem Verkäufer (dem Verlag) und dem Käufer (entweder einer Bibliothek oder einem Forscher) unterscheiden. Die notwendige Dekontextualisierung wird dadurch erreicht, dass wissenschaftliche Artikel in der Regel aus dem laufenden Forschungskontext herausgelöst und als eigenständige Publikationen betrachtet werden. Die Angaben, mit denen man dann einen Artikel identifiziert, sind der Name des Autors, das Jahr der Veröffentlichung und der Name der Zeitschrift – aber nicht die Universität, an der die betreffende Forschung stattfand und auch nicht der Name der Organisation, die die Forschung finanzierte. Hinzu kommt das Layout, das in Form eines Corporate Designs einen Artikel als den einer bestimmten Zeitschrift und eines bestimmten Verlags kennzeichnet und damit als ein Produkt dieser Zeitschrift ausweist.

Zu den sozialen, technischen, rechtlichen und ökonomischen Investitionen, die aus einem schriftlichen Bericht ein singuläres Gut entstehen lassen, gehören aber noch weitere Faktoren: Der Forschungsprozess und die Arbeit verschiedener Akteure verschwinden in dem Produkt Zeitschriftenartikel, Autoren treten Rechte ab, Herausgeber etablieren mit Begutachtungsverfahren Auswahlkriterien für Artikel und koppeln Zeitschriften an wissenschaftliche Kulturen, Kommunikationskonventionen und die professionellen Berufsverläufe. Zugleich unternehmen Verlage Anstrengungen, die dafür Sorge tragen sollen, dass eine Zeitschrift als ein Gut wahrgenommen und für die anvisierten Käufer (Bibliotheken und Forschungseinrichtungen) interessant wird. Dazu gehören Kauf-, Abonnement- und Lizenzbedingungen, gedruckte und elektronische Verfügbarkeit von Publikationen, Zugänglichmachung im Rahmen von Medienpaketen, Referenzierungen in Datenbanken und anderen intermediären Beratungs- und Suchinstrumenten.

Qualifizierungen und Bindungen

Mit den Begriffen ‚Qualifizierung' und ‚Bindung' möchten wir nun in einem weiteren Schritt deutlich machen, wie das gerahmte Produkt, der wissenschaftliche Artikel, in den Forschungsprozess reintegriert wird. Die Qualifizierung von Produkten spielt bei der Bindung von Produkten an Käufer bzw. Konsumenten eine entscheidende Rolle. Denn die etwas magisch wirkende Abstimmung von Angebot und Nachfrage vollzieht sich aufgrund der doppelten Bewegung von Singularisierung und Vergleichbarkeit von Produkten, die zu einer Hinwendung von Konsumenten zu einem Produkt und einer Bindung zwischen Produkt und Käufer führen (Callon & Muniesa, 2007).

In der Qualifizierung vollzieht sich der Prozess der Differenzierung und des Vergleichs von ähnlichen und doch verschiedenen Produkten. Mittels Unterscheidungskriterien für Qualität werden verschiedene Güter in Relation zu einander gestellt. Qualität und die Kriterien, an denen sie festgemacht wird, sind aus der Perspektive der ANT keine intrinsischen Eigenschaften von Produkten, sondern sie sind Ergebnisse von Beurteilungs-, Bewertungs- und Messverfahren (Callon, Méadel, & Rabeharisoa, 2002). Für wissenschaftliche Publikationen bedeutet dies, dass diese erst durch die Tätigkeit von Gutachtern und Qualitätszuschreibungen, wie bspw. ‚Peer Reviewed Article', eine erkennbare und unterscheidbare Qualität erhalten, die sie für die wissenschaftliche Leserschaft mit anderen Artikeln vergleichbar macht.

Die Bindung von Kunden an bestimmte Güter beschreibt die ANT als eine Folge von qualifizierenden Kalkulationen seitens der Konsumenten. Der Vergleich und die Differenzierung von Produkten werden mit den Effekten einer Produktwahl abgeglichen. Callon, Méadel und Rabeharisoa (2002) sprechen in diesem Zusammenhang von zwei prinzipiellen Modalitäten: Zum einen gibt es Kalkulationen, die aufgrund einer expliziten Bewertung bzw. Neubewertung von Gütern eine bestimmte Produktwahl hervorbringen. Zum anderen vollziehen sich Bewertungsprozesse auf der Basis vertrauter und unhinterfragter Bewertungskriterien, die eine routinehafte Auswahl und Nutzung von Gütern bewirken. Das Design von Produkten, die Werbung, besondere Verkaufsangebote oder aber die Positionierung von Waren in Geschäften sind Praktiken, mittels derer die Störung und die Unterbrechung einer routinisierten Bewertung und Produktwahl erfolgen und eine Neubewertung zugunsten anderer Produkte möglich werden soll. So zeigen Méadel und Rabeharisoa (2002) in ihrer Studie zur Bewertung/Neubewertung von Lebensmitteln, wie die veränderte Gestaltung von Orangensaftpackungen, auf denen z. B. bekannte Comicfiguren zu sehen sind, die Kaufroutinen von Eltern und Kindern aufbrechen und eine neue Bewertung von Produkten veranlassen können.

Bezieht man diese Überlegungen zur Qualifikation von Gütern und zur Bindung von Kunden auf den Publikationsmarkt, wird deutlich, dass Bibliotheken und Forschende auf verschiedene Art und Weise Publikationen qualifizieren und auswählen. Sie werden als verschiedene Agenten auf dem Markt aktiv: Bibliotheken bewerten Publikationen in der Regel als Käufer und Forschende nutzen sie als Konsumenten.

Bibliotheken als Käufer

Bibliotheken agieren als Käufer und changieren beim Erwerb zwischen qualifizierender Kalkulation und Routine: Kosten und Nutzen eines Medienerwerbs werden abgewogen, unterschiedliche Angebote von Verlagen evaluiert und kombiniert, Zeitschriften abbestellt oder neu lizensiert. Zugleich erfolgt der Medienerwerb aber auch routiniert auf der Basis etablierter, i. d. R. unhinterfragter Kriterien, wie z. B. dem Verlagsrenommee oder dem Renommee einer Zeitschrift. Denn die Bibliotheken agieren als Käufer und Dienstleister. Mit ihren Dienstleistungen – der Verfügbarmachung von Publikationen – antizipieren sie die Bedürfnisse der Nutzenden und perpetuieren deren Qualifizierungsverfahren. Sie machen Publikationen über Kataloge und Verzeichnisse auf ihren Homepages zugänglich, bieten Hilfestellungen und Empfehlungen bei der Auswahl von Publikationen an und leiten Angebote seitens der Verlage an die Konsumenten (Forschende und Studierende) weiter, um Forschungs- und Lernprozesse zu ermöglichen und zu erleichtern. Als Dienstleister, der seine Angebote genutzt wissen möchte, heben Bibliotheken mit ihren Dienstleistungen strukturell häufig CA-Publikationen prominent hervor.

Forschende als Konsumenten

Konsumenten begegnen Produkten nicht alleine, wie Franck Cochoy (2008) es für Kunden in Supermärkten analysiert hat. Materielle Hilfsmittel begleiten sie ebenso wie Marketing- und Designprozesse. In Supermärkten gehören zu diesen Begleitern das Design und die Verpackung, in Buchläden die Anordnung der Bücher in den Regalen und Auslagen und im Wissenschaftsbetrieb gehören dazu die Internetauftritte von Verlagen oder die Suchinstrumente und Medienpräsentationen von Bibliotheken. Zugleich sind Konsumenten in die kollektiven Evaluierungen ihrer sozialen Netzwerke eingebunden. Forschende sind Teil der wissenschaftlichen Kulturen und deren Konventionen der Mediennutzung und -bewertung. Zudem bieten intermediäre Instanzen – von den Sales-Managern von Verlagen über die wissenschaftlichen Bibliothekare, die IT-Abteilungen und Rechenzentren bis hin zur Graphik – Orientierungen, die die Qualifikationen der Forschenden unterstützen und eine wechselseitige Anpassungen des Guts (Publikation) und des Konsumenten (Wissenschaftler) ermöglichen. Durch das Zusammenspiel dieser Elemente erhalten bereits getroffene Qualifizierungen eine relative Stabilität, die eine routinierte Bewertung fördert. Dagegen kommen bei der Nutzung von Publikationen in der wissenschaftlichen Praxis ökonomische Kriterien der Kalkulation, wie sie die Bibliotheken anlegen müssen, selten zum Tragen.

Mit der Verbindung zwischen Forschenden als Konsumenten und Forschung als Produkt des Publikationsmarktes erfolgt eine erneute Einbindung und Aneignung der Publikationen in den Forschungskontext und -prozess. Die Aneignung von Produkten seitens der Konsumenten nivelliert jedoch nicht die wirtschaftlichen Strukturen der Nutzung oder die Marktförmigkeit von Publikationen. So betonen Callon und Muniesa (2005), dass zwischen der Anpassung der Produkte an Konsumenten und der Einbindung in die Welt des Konsumenten eine Asymmetrie bestehen kann:

„However, irrespective of how strong the consumer's calculative agency that evaluates the attachment of goods to his or her own world may be, it remains weak compared with the calculative power of supply […].“ (p. 1238)

Schluss

Die eingangs beschriebene Spannung, die sich aus der Differenz von Wissenschaft und Wirtschaft ergibt, wird in Anpassungsprozesse zwischen Publikationen als Gütern und Forschenden als Konsumenten übersetzt. Qualifizierungsprozesse, die zur Auswahl und Nutzung von Publikationen führen, spielen bei der Bindung von Produkt und Konsument eine entscheidende Rolle. Denn Produzenten und intermediäre Dienstleister wie bspw. Bibliotheken antizipieren die Beurteilungsverfahren der Nutzer und schaffen Umgebungen, die eine routinehafte Nutzung von Publikationen fördern.

Die vorgeschlagene Beschreibung bietet mehrere Anschlussstellen für weitergehende Überlegungen. So sollte deutlich geworden sein, dass die Strategien der Rahmung, Singularisierung und Qualifizierung, wie wir sie für CA-Publikationen beschrieben haben, auch für OA-Veröffentlichungen relevant sind. Indem wir das Problemfeld als Markt reformulieren, in dem ein Kollektiv von Akteuren kalkuliert und Qualifikationsweisen stabilisiert werden, lassen sich wesentliche Praktiken, die das Feld wissenschaftlicher Publikationen konstitu-

ieren, differenzieren und beschreiben: So wird deutlich, dass Kalkulationen heterogen sind. Die verschiedenen Marktakteure kalkulieren und qualifizieren auf unterschiedliche Weise – die Bibliotheken, die als Käufer fungieren, wenden andere Kriterien an als Forschende, die die Rolle von Konsumenten einnehmen. Diese differierenden Kalkulationsmodi werden in der vorgeschlagenen Perspektive beobachtbar und einer eingehenden Analyse zugänglich. Für OA-Initiativen ist eine solche Perspektive unserer Ansicht nach insofern von Bedeutung, als dass hier die Widerstände, jenseits von Wert- und Moraldebatten, in ihrer banalen Praxis und Routine fassbar werden.

Literatur

Antelmann, K. (2004). Do open access articles have a greater resarch impact? *College & Research Libraries News*, 65(5), 372-382.

Callon, M. (1998). The embeddedness of economic markets. In Callon, M. (Hg.), *The Laws of the Markets*. Oxford: Blackwell, 1-57.

Callon, M (2006). Akteur-Netzwerk-Theorie: Der Markttest. In Belliger, A. & Krieger, D.J. (Hg.), *ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie.* Bielefeld: transcript, 545-559.

Callon, M., Méadel, C., & Rabeharisoa, V. (2002). The economy of qualities. *Economy and Society,* 31 (2), 194-217.

Callon, M., & Muniesa, F. (2005). Economic markets as calculative collective devices. *Organization Studies*, 26 (8), 1229-1250.

Cochoy, F. (2008). Calculation, qualculation, calqulation: shopping cart arithmetic, equipped cognition and the clustered consumer. *Marketing Theory*, 8 (1), 15-44.

Koch, L., Mey, G., & Mruck, K. (2009). Erfahrungen mit Open Access – ausgewählte Ergebnisse aus der Befragung zu Nutzen und Nutzung von "Forum Qualitative Forschung / Forum: Qualitative Social Research" (FQS). *Information – Wissenschaft & Praxis*, 60 (5), 291-299. http://eprints.rclis.org/16860 (Zugriff 06.12.2010).

Luhmann, N. (1999). *Die Wirtschaft der Gesellschaft*. Frankfurt am Main: Suhrkamp.

Stichweh, R. (2003). The multiple publics of science: inclusion and popularization. *Soziale Systeme*, 9 (2), 210-220.

Stichweh, R. (2005). Die Universalität des wissenschaftlichen Wissens. In Gloy, K. & zur Lippe, R. (Hg.), *Weisheit – Wissen – Information*. Göttingen: V & R unipress, 177-191.

Kommunikation und Publikation in der digitalen Wissenschaft: Die Urheberrechtsplattform IUWIS

Elena Di Rosa, Valentina Djordjevic & Michaela Voigt

Projekt IUWIS
Humboldt-Universität zu Berlin
Institut für Bibliotheks- und Informationswissenschaft
Unter den Linden 6
D-10099 Berlin
elena.di.rosa@cms.hu-berlin.de, v.djordjevic@ibi.hu-berlin.de, voigtmic@cms.hu-berlin.de

Zusammenfassung

Der folgende Beitrag stellt die Urheberrechtsplattform IUWIS vor. Es handelt sich um ein Webangebot, welches umfassend die aktuellen Ereignisse und Diskussionen im Bereich Urheberrecht in Wissenschaft und Bildung abbilden will. Neben grundlegenden Informationen für die Allgemeinheit, welche zum Teil redaktionell erarbeitet werden, sollen Experten, Wissenschaftler oder andere Akteure des Themengebiets durch kollaborative und kommunikative Dienste aktiv die Inhalte auf der Plattform mitgestalten.

Was ist IUWIS?

Mit IUWIS entsteht eine webbasierte **I**nfrastruktur zum Thema **U**rheberrecht für **Wis**senschaft und Bildung, die die Orientierung in dem immer komplexer werdenden Feld Urheberrecht erleichtern möchte. Dabei dokumentiert IUWIS die verschiedenen, durchaus kontroversen Stimmen und Positionen zum Urheberrecht. IUWIS leistet keine verbindliche Rechtsberatung, sondern sammelt und vermittelt Antworten zu urheberrechtlichen Problemen für Wissenschaft und Bildung. Bei konkreten Problemen hilft IUWIS bei der Suche nach kompetenten Ansprechpartnern.

IUWIS orientiert sich am Web 2.0-Paradigma und nutzt die dort entwickelten Werkzeuge, so dass Nutzer auf bekannte Interaktionsformen zurückgreifen können. IUWIS versteht sich als sozialer Dienst, in dem Nutzer ihr Wissen mit anderen teilen und gemeinsam Lösungsvorschläge für Probleme erarbeiten können. Die IUWIS-Website bietet verschiedene Möglichkeiten der Kommunikation und Interaktion, so dass betroffene und interessierte Nutzer direkt miteinander in Kontakt treten können.

Angesiedelt ist das Projekt am Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin und wird gefördert von der Deutschen Forschungsgemeinschaft (DFG) im Rahmen des Programms „Wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS)“.

Weshalb ist IUWIS notwendig?

Die Digitalisierung hat die gesamten Produktions- und Rezeptionsbedingungen in der Informationsgesellschaft verändert. Damit ist das Rechtsgebiet Urheberrecht ins Zentrum der Debatten um Wissensproduktion und -dissemination gerückt. In Wissenschaft und Bildung haben urheberrechtliche Fragen großen Einfluss auf die Arbeitsbedingungen – für Lehrende, Forscher und Studenten. Dazu kommt, dass viele Regelungen des Urheberrechts für juristische Laien unverständlich sind, wodurch Unsicherheiten bei der Wissensproduktion und -rezeption entstehen. Das liegt sicherlich auch daran, dass z. B. den meisten Wissenschaftlern die Bedeutung des Urheberrechts für ihre Arbeit nicht bewusst ist. Eric Steinhauer, einer der bekanntesten deutschen Bibliotheksjuristen, formuliert es wie folgt:

„Eine mögliche Antwort kann im Stellenwert des Urheberrechts in der Wissenschaft selbst gesucht werden. Obwohl wissenschaftliches Publizieren und ein durch Publikationen geschaffenes Renommee für jeden ernsthaften Wissenschaftler sehr wichtig sind, ist das Interesse für die spezifisch juristischen Fragen des wissenschaftlichen Publizierens schwach bis gar nicht ausgeprägt.“ (Steinhauer, 2007)

Ausgehend von dieser Beobachtung muss es im ersten Schritt darum gehen, das Problembewusstsein zu schärfen. Das Ziel von IUWIS ist, daran anschließend eine zeitgemäße Lösung zu entwickeln, um dem gesteigerten Informations- und Diskussionsbedürfnis zu begegnen.

Fragen zu Open Access, Verlagsverträgen, Lizenzierung von digitalen Werken oder Nutzungsbedingungen von E-Books werden in der Öffentlichkeit kontrovers diskutiert (Kaube, 2010; Kuhlen, 2010; Bury, 2010; Braun, 2010). Dabei sind die Fronten der verschiedenen Interessengruppen relativ klar abgesteckt: Den Rechteinhabern – Buch- und Musikverlagen, Produktionsfirmen – stehen die Interessen der Nutzer wie Bibliotheken, Archive und Forschungsgruppen gegenüber. Wissenschaftler nehmen dabei eine Doppelrolle ein, da sie zum einen als Produzenten und zum anderen als Rezipienten im wissenschaftlichen Publikations- und Kommunikationsprozess auftreten.

Auch Lehrende sind mit urheberrechtlichen Fragen konfrontiert. Zu ihrem Arbeitsalltag gehört es, urheberrechtlich geschützte Werke in elektronischer und gedruckter Form zu verwenden. Es gibt dafür zwar Schrankenregelungen für Forschung und Lehre, die Lehrenden einen Spielraum geben sollen. Diese sind aber nicht eindeutig, so dass in der Praxis Fragen bleiben, die Lehrende oft ratlos zurücklassen: Welche Medien dürfen in welcher Form genutzt werden? Wann müssen Rechteinhaber um Erlaubnis gefragt werden? Wie muss die Nutzung vergütet werden? Vor allem bei der Nutzung von elektronischem Material ist es sehr schwer zu beurteilen, ob eine Handlung nach rechtlichen Maßstäben zulässig ist oder nicht.

Auf IUWIS können zu diesen und anderen komplexen Fragen, die sich in der Praxis stellen, Informationen gesammelt werden. Zusätzlich zu diesen praktischen Hinweisen zum Urheberrecht ist es aber notwendig, dass sich die wissenschaftlichen Gemeinschaften an der weiteren Ausgestaltung des Urheberrechts beteiligen, da sonst die Gefahr besteht, dass die Interessen der Wissenschaft nicht angemessen berücksichtigt werden. Seit einigen Jahren

finden mit den Reformen des Urheberrechts durch den Ersten, Zweiten und jetzt Dritten Korb Neuverhandlungen über den Interessenausgleich zwischen Urhebern, Verwertern und Nutzern statt. IUWIS möchte diesen Prozess nicht nur begleiten – durch Blogbeiträge, Artikelsammlungen und Analysen –, sondern auch eine Plattform bieten, auf der eine Diskussion über diese Fragen stattfinden kann.

Bündelung von Ressourcen

Die Kommunikation zum Thema Urheberrecht wird derzeit von dezentralen Strukturen unterschiedlicher Angebote beherrscht: Zahlreiche thematische Blogs, Mailinglisten und Foren diskutieren die neuesten Entwicklungen, da statische Informationsangebote die Ereignisse um die Veränderungen des Urheberrechts nicht einfangen können. Nicht nur direkt am Schöpfungs- und Nutzungsprozess Beteiligte haben Schwierigkeiten aufgrund der unsicheren Rechtslage, auch Dienstleister wie Bibliotheken wissen nicht immer, ob ihr Umgang mit urheberrechtlich geschütztem Material rechtskonform ist. Prominent hierbei ist eine gerichtliche Auseinandersetzung zwischen dem Ulmer-Verlag und der Universitätsbibliothek der TU Darmstadt: In dem Prozess geht es um die Nutzung von digitalisierten Büchern an elektronischen Leseplätzen in Bibliotheken. Hieran wird deutlich, dass alle Akteure in Wissenschaft und Bildung betroffen sind – sei es als Urheber, Nutzer oder Dienstleister.

IUWIS bündelt Informations- und Diskursangebote mit dem Ziel, den Bedürfnissen heterogener Zielgruppen gerecht zu werden und den unterschiedlichen Meinungen auch im Gesetzgebungsprozess Raum zu geben. Ausgehend von der Diversität der Nutzer soll mit IUWIS ein virtueller Diskursraum entstehen, der je nach Bedarf und Intention genutzt werden kann.

Das IUWIS-Angebot: Infopool und Dossier

Hierfür verfolgt IUWIS einen hybriden Ansatz: Über den **Infopool** sollen relevante Informationen zum Thema erschlossen und zugänglich gemacht werden; ergänzt wird dies durch Kommunikationsfunktionen im Bereich der **Dossiers**. Im Vordergrund steht dabei, die Diskussionsstränge, die derzeit im Internet ins Leere laufen, zusammenzufassen und festzuhalten.

Im **Infopool** werden interne und externe Inhalte gesammelt; es wurde ein Datenmodell entwickelt, um Inhalte strukturiert nach verschiedenen Kriterien (formal wie inhaltlich) erschließen zu können. Dadurch können die Informationsobjekte für den Diskurs nutzbar gemacht werden. Die Informationsobjekte bestehen aus internen Quellen, wie Blogbeiträgen oder Kommentartexten von IUWIS-Nutzern oder Aufsätzen, die bei IUWIS veröffentlicht wurden, und externen Quellen, die erschlossen und verlinkt sind. Dank der Erschließung der Quellen über Metadaten ist eine mehrschichtige Recherche möglich; zukünftig sollen die Metadaten auch exportierbar sein. Eine wichtige Rolle bei der Erschließung von Inhalten spielt die *Tagsonomy*, bei der Nutzer den Informationsobjekten freie Schlagwörter (sogenannte *Tags*) zuweisen können, über die eine intuitive und einfache Navigation stattfinden kann. Unterstützt wird diese nutzererstellte Verschlagwortung von einem kontrollierten Vokabular, um Redundanzen und Unschärfen einzuschränken.

In den **IUWIS-Themendossiers** werden die im Infopool gesammelten und erschlossenen Dokumente verknüpft und unter der Beteiligung der interessierten Nutzer kontextualisiert. Wenn ein Themengebiet entweder von den Nutzern oder der IUWIS-Redaktion als relevant identifiziert wurde, kann auf einfache Weise ein Dossier eröffnet werden. Dabei werden einschlägige Dokumente in dieses Dossier eingebettet, sowohl automatisiert anhand der Tags als auch manuell. Die Dossiers werden betreut von Dossier-Moderatoren, die aus der Community stammen und eine Mittlerfunktion übernehmen, damit etwa die Diskussionen in geregelten Bahnen verlaufen und relevante Dokumente mit dem Dossier verknüpft werden. In den Dossiers können die Dokumente und Beiträge kommentiert und diskutiert werden.

Neben der Abbildung von thematischen Diskursen sollen die Dossiers auch die Vernetzung von Akteuren und Institutionen ermöglichen, indem zum Beispiel geschlossene Dossiergruppen für einzelne Benutzergruppen eingerichtet werden können. Die Dossiers sind geplant als Verbindung der informativen und kommunikativen Dienste, insofern an dieser Stelle beide Stränge verknüpft werden können. Nutzer können je nach eigener Interessenlage eher lesen oder den Diskurs durch Kommentare, Dossier- oder Blogbeiträge ausgestalten.

Grundsätzlich versteht sich IUWIS als kollaborative Umgebung, indem es eine Infrastruktur für die Sammlung und Kommentierung der thematisch relevanten Informationen bietet. So sollen die Dossiers nicht allein von den IUWIS-Mitarbeitern angestoßen und gepflegt werden, sondern von allen registrierten Nutzern der Website angelegt werden können. Vor allem zu Beginn jedoch wird es eine redaktionelle Aufgabe sein, relevante Themen zu identifizieren und sie zielgruppengerecht aufzubereiten. Im Sinne einer nachhaltigen Entwicklung arbeitet das IUWIS-Team daran, eine Community um das Portal IUWIS zu bilden, die auch nach Projektende weiterhin diskutiert und thematisch relevante Inhalte erstellt.

Stand der Arbeit

IUWIS befindet sich immer noch im Aufbau. Die komplexen Community-Funktionen, die in der Konzeptionsphase angelegt wurden, müssen teilweise noch in die Praxis umgesetzt werden. Gegenwärtig liegt der Schwerpunkt auf der Erweiterung des Infopools und der Erschließung der Aufsätze, Essays, Kommentare, Gerichtsurteile und Gesetze, die tagtäglich in Print und Online erscheinen. Im November 2010 ist eine erste Dossierversion erscheinen, die im Folgenden schrittweise erweitert und mit Inhalten gefüllt wird und hoffentlich auch Interessierte aus Wissenschaft und Bildung zum Gespräch zusammenbringen wird.

Die Kommunikations- und Informationsstrukturen von IUWIS könnten zukünftig auch auf andere Fachgebiete übertragen werden, in denen verschiedene Zielgruppen angesprochen sind und andere thematische Diskurse dargestellt und vermittelt werden sollen.

Literaturangaben

Braun, Ilja (2010, 4. März). Was kauft man, wenn man ein Ebook kauft? In iRights.info, Zugriff am 17. Dezember 2010 unter http://www.irights.info/index.php?q=node/870.

Bury, Liz (2010, 25. März). Author Contracts 2.0: Putting Cash Before Copyright and Control. Publishing Perspectives, Zugriff am 17. Dezember 2010 unter http://publishingperspectives.com/2010/03/author-contracts-2-0-putting-cash-before-copyright-and-control/.

Kaube, Jürgen (2010). Chemiker über die Nachteile des „Open Access“. Frankfurter Allgemeine Zeitung, 08. Dezember 2010, N5.

Kuhlen, Rainer (2010, 23. September). Open Access – im Interesse aller: Produzenten, Nutzer und der publizierenden Informationswissenschaft. In Studienheft 7: Forschungs- und Technologiepolitik des Bund demokratischer WissenschaftlerInnen. Preprint, Zugriff am 17. Dezember 2010 unter http://www.kuhlen.name/MATERIALIEN/Publikationen2010/OA--RK230910-Word97-final.pdf.

Steinhauer, Eric W. (2010, 27. September). Urheberrechtsnovelle – Das Urheberrecht in der Wissenschaft, oder „The Dirty Way Of Information. In H-Soz-u-Kult, Zugriff am 17. Dezember 2010 unter http://hsozkult.geschichte.hu-berlin.de/forum/id=938&type=diskussionen.

Die „European Psychology Publication Platform“ zur Steigerung von Sichtbarkeit und Qualität europäischer psychologischer Forschung

Isabel Nündel, Erich Weichselgartner & Günter Krampen

Leibniz-Zentrum für Psychologische Information und Dokumentation
Universität Trier
Trier
nuendel@zpid.de

Zusammenfassung

Das Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) entwickelt mit weiteren Kooperationspartnern die *European Psychology Publication Platform*. Diese Open-Access-Publikationsinfrastruktur ist als Knotenpunkt für die wissenschaftliche Gemeinschaft gedacht und soll die Sichtbarkeit und Qualität europäischer psychologischer Forschung durch ihre Vielfalt an Sprachen, Publikationsvarianten und Mehrwertdiensten befördern.

Einleitung

Die fachübergreifende Verbreitung des Open Access Gedankens ist nicht zuletzt Ausdruck eines wachsenden Bedürfnisses nach ungehindertem Zugang zu Informationen innerhalb der wissenschaftlichen Gemeinschaft. Bisherige Modelle wie der grüne oder der goldene Weg weisen in ihrer Umsetzung und Akzeptanz, nicht zuletzt im Bereich der Psychologie noch Schwächen auf (Mey & Mruck, 2007), die die Verfügbarkeit der Fachliteratur beeinträchtigen. Doch gerade für die Forschung im inhomogenen europäischen Raum ist es essentiell, sowohl in ihrer inhaltlichen als auch sprachlichen Vielfalt präsent zu sein, da nur so ein Gegengewicht zu der in vielen Fachgebieten vorherrschenden dominanten Sichtbarkeit der anglo-amerikanischen Forschung geschaffen werden kann.

Das Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) hat sich daher die Entwicklung einer digitalen europäischen Publikationsplattform für die Psychologie zum Ziel gesetzt. Das Projekt „European Psychology Publication Platform“ soll durch seine mehrsprachige Gestaltung dazu beitragen, die Sprachenvielfalt Europas zu erhalten sowie den Austausch und die Zusammenarbeit zwischen Wissenschaftlern und Praktikern europaweit zu stärken.

Dem Open Access Gedanken gemäß sollen sowohl die Veröffentlichung eines wissenschaftlichen Beitrags als auch dessen Lesen und die weitere Nutzung der Plattform kostenfrei sein. Da im akademischen Wissenschaftsprozess in der Regel ohnehin eine öffentliche Förderung erfolgt, erscheint die Mitfinanzierung der Verbreitung der Ergebnisse als logische Konsequenz, um die Sichtbarkeit der Forschung und somit eine umfassende Forschungsförderung sicherzustellen.

Durch die Einbindung verschiedenster Publikationsformen soll die Vielfalt an wissenschaftlicher Information im Bereich der Psychologie adäquat abgebildet werden. Entsprechend ist die Aufnahme von Artikeln, Monographien und Sammelbänden geplant sowie das Angebot von Forschungsberichten, Konferenzbeiträgen und Postern. Des Weiteren sollen auch Rohdaten, Tests, Praxisleitfäden und multimediale Formate in den Informationsbestand aufgenommen werden. Peer-Review gewährleistet die Einhaltung höchster Qualitätsstandards.

Innovative Zusatzfunktionen sollen die Möglichkeiten der digitalen Wissenschaft voll ausschöpfen, zum Beispiel mehrsprachige Metadaten, Verlinkung von Zitaten, „lebende Artikel“, Kommentare, u. v. m. Die Erfassung der Viewer- beziehungsweise Downloadzahlen erlaubt darüber hinaus Rückschlüsse auf den „impact“ einzelner Publikationen (Herb, 2010).

Bedingt durch den Mehrwert der digitalen Plattform gegenüber dem Printmedium stellen sich hohe Ansprüche an die technische und inhaltliche Qualität, welche die „European Psychology Publication Platform“ zu einem Gemeinschaftsprojekt macht, für das es Interessenten in zahlreichen europäischen Ländern gibt (Uhl, 2009).

Neue Chancen und Herausforderungen durch Open Access

Die Open Access Bemühungen splitten sich momentan in zwei Lager auf: Die Vertreter des grünen Weges und die Vertreter des goldenen Weges. Allerdings weisen bisher beide Ansätze noch Schwächen auf.

So werden zwar über den grünen Weg und somit durch die Zweitveröffentlichung in institutionell oder fachlich gebundenen Repositorien größere Ansammlungen an wissenschaftlichen Informationen geschaffen und diese dadurch besser verknüpft. Es besteht allerdings trotz Initiativen wie der SHERPA/RoMEO-Liste häufig Verwirrung bezüglich der Rechteabklärung mit dem Verleger der Erstveröffentlichung, wodurch die allgemeine und freie Zugänglichkeit von Texten in Langzeitarchiven erschwert wird. Angebote wie der im Mai dieses Jahres veröffentlichte juristische Leitfaden „Zur Online-Bereitstellung älterer Publikationen“ von Dr. Till Kreutzer stellen eine Hilfestellung in diesem Bereich dar (Kreutzer, 2010).

Der zweite Ansatz umgeht diese Schwierigkeit, indem direkt eine Open-Access-Fachzeitschrift zur Erstveröffentlichung und damit der goldene Weg genutzt wird. Hier entsteht jedoch der Bedarf einer Verknüpfung der diversen Fachzeitschriften, um die einzelnen Beiträge in ihrer Gesamtheit für die wissenschaftliche Gemeinschaft sichtbar und leicht zugänglich zu machen, worum sich Projekte wie das Directory of Open Access Journals (DOAJ) bemühen. Zudem kommt dieser Weg nicht für umfangreichere Veröffentlichungen wie etwa Monographien in Frage und schließt auch andere Formen wissenschaftlicher Information wie Poster oder psychologische Tests weitestgehend aus. Daher mangelt es zurzeit noch an einer zentralen Anlaufstelle für Wissenschaftler und Praktiker, die eine Veröffentlichung von Forschungsergebnissen in jeglicher Publikationsform ermöglicht und dabei einen freien Zugriff auf diese gewährleistet.

Aus diesem Grund hat sich das Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) die Entwicklung einer europäischen Publikationsplattform für die Psychologie und ihre Nachbardisziplinen zum Ziel gesetzt. Das Projekt mit dem Arbeitstitel „Euro-

pean Psychology Publication Platform“ soll daher im Folgenden in seinen Grundzügen vorgestellt werden.

EINE Plattform für Psychologische Information mit europäischem Fokus

Den Bedarf für eine europäische Publikationsinfrastruktur, die sowohl das Recherchieren als auch das Veröffentlichen in Open Access Publikationen erleichtert, verdeutlichte eine 2008 vom ZPID durchgeführte Umfrage in der europäischen Psychologie, an der sich 493 Personen aus 24 Ländern beteiligten. 58% der Befragten kannten Open Access Zeitschriften, wobei von diesen lediglich 6% regelmäßig in derartigen Zeitschriften veröffentlichten. Im Gegensatz dazu gaben 80% der Befragten an, gerne in einer europäischen Open Access Zeitschrift veröffentlichen zu wollen (Uhl, 2009).

Unter dem vorläufigen Titel „European Psychology Publication Platform“ baut das ZPID daher zurzeit zusammen mit internationalen Partnern aus ganz Europa eine europäische Publikationsplattform für wissenschaftliche Informationen aus dem Bereich der Psychologie und ihren Nachbardisziplinen auf (Weichselgartner & Uhl, 2009). Die Beteiligung weiterer Länder ist geplant, da ein solch umfassendes Projekt nur durch eine gemeinsame europaweite Anstrengung verwirklicht werden und dadurch letztendlich der gesamten Wissenschaftsgemeinschaft zu Gute kommen kann.

Eine mehrsprachige Gestaltung der Plattform soll dazu beitragen, die Sprachenvielfalt Europas zu erhalten und den Austausch und damit verbunden die internationale Zusammenarbeit zwischen Wissenschaftlern und Praktikern zu stärken. Eine Fokussierung auf europäische Beiträge wird zudem den Ergebnissen europäischer Wissenschaft eine breite Darstellungsplattform bieten und damit entscheidend deren Sichtbarkeit fördern. Gleichzeitig ist das Angebot auch offen für Beiträge aus dem nichteuropäischen Raum und bietet so eine umfassende Sammlung aktueller Forschungsergebnisse.

Doch nicht nur für die Autoren wissenschaftlicher Publikationen soll mit diesem Projekt ein zentraler Anlaufpunkt geschaffen werden. Auch für weitere Nutzer wissenschaftlicher Informationen, seien es Studenten, Praktiker oder anderweitig interessierte Personen, soll die Plattform den Ausgangspunkt ihrer Recherchen darstellen.

Durch eine optimale Nutzung der vielfältigen Möglichkeiten der digitalen Umgebung einer elektronischen Plattform wird zudem ein aktuelles und umfassendes Medium geschaffen, das in dieser Weise für die Psychologie bisher nicht existiert.

Konkrete Umsetzung der Plattform

Auf der „European Psychology Publication Platform“ werden verschiedenste Publikationsformen vertreten sein, um so die vorhandene Vielfalt an verfügbarer wissenschaftlicher Information im Bereich der Psychologie adäquat abzubilden. Die Qualität der Beiträge wird dabei durch Peer-Review gewährleistet. Entsprechend ist die Aufnahme von Artikeln, Monographien, Sammelbänden und „grauer Literatur“ wie Forschungsberichten, Konferenzbeiträgen und Postern geplant. Des Weiteren sollen auch Rohdaten, Tests, Praxisleitfäden und multimediale Formate in den Informationsbestand aufgenommen werden. Dabei wäre auch eine Einbindung verschiedener Repositorien denkbar, wodurch sowohl dem Anspruch auf

Abbildung aktueller Forschung als auch auf langfristige Archivierung bisheriger wissenschaftlicher Informationen Rechnung getragen würde. Durch die Zuweisung von Digital Object Identifiern (DOI) wird die dauerhafte Zitierbarkeit der einzelnen Beiträge sichergestellt.

Die Plattform erhält zudem einen Mehrwert durch zahlreiche Zusatzfunktionen, die die neuesten technischen Entwicklungen im Bereich der digitalen Informationsverwaltung nutzen. So werden umfangreiche, mehrsprachige Metadaten verfügbar gemacht. Darüber hinaus sollen Harvesting und die Verlinkung von Zitaten möglich sein. Die Verknüpfung von Artikeln mit den für sie relevanten Forschungsdaten könnte zudem die komplexe Darstellung von Forschungsergebnissen unterstützen. Durch die Verfügbarkeit vielfältiger Bausteine derart publizierter Forschungsarbeiten bestünde eine verbesserte Ausgangslage für darauf aufbauende weiterführende Forschungen.

Das Projekt eröffnet die Möglichkeit von „lebenden Artikeln“, die von Autoren ergänzt und angereichert werden können. Dies würde ein interaktives Schreiben und Forschen auch auf internationaler Ebene erleichtern. Außerdem könnten sich die Leser durch Kommentare und Anmerkungen aktiv in die jeweilige wissenschaftliche Diskussion auf der Plattform einbringen. Ein reger Austausch innerhalb der Disziplin würde so unterstützt und der direkte Einfluss einer Publikation auf die wissenschaftliche Gemeinschaft sichtbar gemacht werden. Durch die Erfassung der Viewer- beziehungsweise Downloadzahlen könnte außerdem zusätzlich auf den „impact“ einzelner Publikationen geschlossen werden. Zudem soll dem jeweiligen Autor stets die Gelegenheit zur Reaktion auf die geäußerten Kommentare der übrigen Nutzer gegeben werden. So kann das Feedback der Forschungsgemeinschaft angenommen und dadurch eine permanente Weiterentwicklung einzelner Untersuchungspunkte und Forschungsunternehmen ermöglicht werden.

Durch die Integration von Open-Access-Zeitschriften wird die Beitragsvielfalt der Plattform bereichert und die Abbildung aktueller Diskussionen und Forschungsbestrebungen unterstützt. Das ZPID sammelt in diesem Bereich zurzeit praktische Erfahrungen in einer Kooperation mit dem International Journal of Internet Science (IJIS). Mit Systemen wie dem Open Journal Systems (OJS) des Public Knowledge Projects kann den Verlegern von Zeitschriften dabei eine unkomplizierte Hilfestellung zur Unterstützung ihres workflows von der Beitragseinreichung bis zum Peer-Review-Verfahren geboten werden. Unter Einhaltung höchster wissenschaftlicher Standards kann so der Zeitraum zwischen Beitragseinreichung und letztendlicher Publikation entscheidend verkürzt und die Verbreitung von Forschungsergebnissen beschleunigt werden. Durch offene Peer-Review-Prozesse könnte die Aufnahme von Artikeln zudem transparent gehalten werden.

Ein zentrales Anliegen bei der Entwicklung der „European Psychology Publication Platform“ ist zudem der Open Access Gedanke. Dabei soll sich die Kostenfreiheit allerdings nicht nur auf die Nutzer der Plattform, sondern auch auf die Autoren der einzelnen Publikationen beziehen. Die Veröffentlichung eines wissenschaftlichen Beitrags auf der Plattform soll daher mit keinerlei finanziellem Aufwand für die Verfasser verbunden sein. Wie eingangs ausgeführt, wird der Großteil der akademischen Forschung aus öffentlichen Mitteln finanziert und es ist nicht einsichtig, warum deren Ergebnisse mitunter von kommerziellen Verlagen

zurückgekauft werden müssen. Die Mittel, die dafür den Bibliotheken zur Verfügung stehen, könnten alternativ in den Betrieb einer Publikationsplattform wie der hier vorgestellten fließen, mit dem zusätzlichen Vorteil der Chance auf eine weitaus größere Verbreitung der Ergebnisse durch Open Access. Ziel der „European Psychology Publication Platform“ ist es entsprechend, europäische Forschung weitestgehend international sichtbar zu machen.

Die Schaffung einer europäischen Publikationsplattform für die Psychologie stellt ein großes Vorhaben dar, das sich durch seine hohen Ansprüche an technische und inhaltliche Qualität sowie die internationale Verknüpfung des Angebots auszeichnet. Durch dieses Projekt wird die Sichtbarkeit, Verfügbarkeit und Qualität europäischer psychologischer Forschung entscheidend befördert. Gleichzeitig wird es gerade dadurch zu einem Gemeinschaftsprojekt, welches nur durch Bemühungen der wissenschaftlichen Gemeinschaft getragen werden kann. Es ist demnach der Einsatz und die Beteiligung möglichst vieler europäischer Universitäten, Forschungseinrichtungen und Verleger gefragt, um die europäische Plattform für psychologische Information in ihrer angedachten Qualität ins Leben zu rufen und auch langfristig betreiben zu können.

Aus diesem Grund freut sich das ZPID jederzeit über weitere Unterstützung und steht auch gern für nähere Auskünfte zur Verfügung (für weitere Informationen: www.psychprints.eu). Denn nur das Beitragen zahlreicher Partner kann ermöglichen, dass diese Plattform auch tatsächlich zum Knotenpunkt für Psychologische Information in Europa und so eine zentrale Anlaufstelle für die wissenschaftliche Gemeinschaft wird.

Fazit

Der Aufbau der *European Psychology Publication Platform* ermöglicht eine optimale Ausschöpfung der elektronischen Publikationsmöglichkeiten und trägt zugleich dem Open Access Gedanken Rechnung. Durch die Bereitstellung dieses Knotenpunkts für die Psychologie und ihre Nachbardisziplinen wird die Sichtbarkeit, Verfügbarkeit und Qualität der europäischen psychologischen Forschung befördert und die Vernetzung von Wissenschaftlern und Praktikern gestärkt.

Literatur

Herb, U. (2010). OpenAccess Statistics: Alternative Impact Measures for Open Access documents? An examination how to generate interoperable usage information from distributed Open Access services“. In C. Boukacem-Zeghmouri (Hrsg.), *L'information scientifique et technique dans l'univers numérique. Mesures et usages*. Paris: L’association des professionnels de l’information et de la documentation, ADBS, 165-178.

Kreutzer, T. (2010): Zur Online-Bereitstellung älterer Publikationen. *http://www.allianzinitiative.de/fileadmin/leitfaden.pdf; doi*:10.2312/allianzoa.002.

Mey, G. & Mruck, K. (2007). Open Access – Auswirkungen einer Informationskrise ... als Chance für die Information. Journal *für Psychologie, 15 (2).*

Uhl, M. (2009). Survey on European Psychology Publication Issues. In E. Weichselgartner & M. Uhl (Eds.), Proceedings of the workshop on European psychology publication issues. *Psychology Science Quarterly, 51 (Supplement),* 17-24*.*

Weichselgartner, E. & Uhl, M. (Eds.). (2009). Proceedings of the workshop on European psychology publication issues. *Psychology Science Quarterly, 51 (Supplement). Lengerich: Pabst.*

Digitale Publikationen in Osteuropawissenschaften. Digitale Reihe hervorragender Abschlussarbeiten des Projektes OstDok

Arpine A. Maniero

Projektkoordinatorin OstDok
Collegium Carolinum
Hochstraße 8
D-81669 München
arpine.maniero@extern.lrz-muenchen.de

Zusammenfassung

Die Möglichkeit der Aufarbeitung und Bereitstellung wissenschaftlicher Materialien im Internet geht heute über den Rahmen „einfacher" Retrodigitalisierungsprojekte weit hinaus. Es geht vielmehr um die Entwicklung und Vernetzung fachlicher und fachübergreifender Repositorien und Datenbanken, um den weltweiten und unkomplizierten Zugang zu online Materialien, ja um den wissenschaftlichen Austausch auf der virtuellen Ebene. Während aber das Angebot an wissenschaftlicher Literatur und bereits gedruckten Publikationen dem Forscher durch die Retrodigitalisierung in immer größerem Umfang frei zur Verfügung steht, gestaltet sich der Weg für online Erstpublikationen erst einmal schwieriger. Zwar scheint sich diese Praxis immer mehr zu etablieren, ist aber dennoch fachspezifisch unterschiedlich stark ausgeprägt. So werden in den Fachkreisen immer wieder die Unterschiede etwa zwischen den Natur- und Geisteswissenschaften in der elektronischen Publikationskultur hervorgehoben, wobei den geisteswissenschaftlichen Texten nicht die Aktualität zugesprochen wird, die eine äußerst wichtige Rolle für die Naturwissenschaften spielt. Dementsprechend – so eine These – wären die Geisteswissenschaften auf die vor allem schnellere Verbreitung ihrer Forschungsergebnisse nicht in dem Maße angewiesen, wie es in den Naturwissenschaften der Fall wäre.[1] Des Weiteren werden die unterschiedlichen Publikationsformate wie die in den Naturwissenschaften maßgebenden Zeitschriftenartikel und in den Geisteswissenschaften eher üblichen Monographienformate als signifikante Unterschiede angesprochen.[2]

Im Folgenden wird am Beispiel des Projektes OstDok (Osteuropadokumente online) auf die Perspektiven des online Publizierens in den Osteuropawissenschaften eingegangen, wobei der Akzent auf der Veröffentlichung hervorragender Abschlussarbeiten liegt.

[1] Vgl. dazu u. a. Landes, L. (2009). Open Access und Geschichtswissenschaften – Notwendigkeit, Chancen, Probleme. In Libreas. Library Ideas. Elektronische Zeitschrift für Bibliotheks- und Informationswissenschaft, 14(1). Verfügbar unter http://www.libreas.eu/ausgabe14/024lan.htm (Zugriff auf diese sowie alle anderen Links zuletzt am 11.10.2010).

[2] Vgl. Gradmann, S. (2007). Open Access – einmal anders. Zum wissenschaftlichen Publizieren in den Geisteswissenschaften. Erstveröffentlichung in ZfBB, 54(4-5), S.170-173. Verfügbar unter http://edoc.hu-berlin.de/oa/articles/reWDQUdy2bil/PDF/209UiQvDa8SCA.pdf.

Einführung

Während die technischen Voraussetzungen zum Aufbau anspruchsvoller Fachrepositorien bereits weitgehend gegeben sind, stellt deren inhaltlicher Aufbau eine erst noch zu überwindende Hürde dar. Die Anschaffung wissenschaftlicher Materialien und ihre Darbietung in einer attraktiven online Umgebung, der Gedanke des offenen Zugangs sowohl zu wissenschaftlichen Forschungsergebnissen als auch zu Rohmaterialien wie Quellen und Primärdaten bedürfen nach wie vor noch intensiver Überzeugungsarbeit. Für eine erfolgreiche Konzeption eines solchen Fachrepositoriums setzt dies nachhaltige Werbung sowie eine intensive Kommunikation mit den einschlägigen Fachkreisen und einzelnen Wissenschaftlern voraus.

Ziel von Open Access Projekten sollte es jedoch nicht sein, die Fachrepositorien ausschließlich inhaltlich zu füllen. Weitaus entscheidender, wenngleich schwerer zu erreichen sind die intensive Nutzung, Zitierung und Weiterempfehlung der online Materialien, was ihre Vertrauenswürdigkeit und ihren wissenschaftlichen Wert voraussetzt, aber durchaus auch die Konzeption und der inhaltliche Aufbau des Portals selbst. Der Vertrauensaufbau gegenüber online Ressourcen, die Anerkennung ihres wissenschaftlichen Wertes, ja ihre Gleichsetzung mit den auf dem „traditionellen“ Wege entstandenen wissenschaftlichen Materialien sind die wichtigsten Schritte auf dem Weg zur Kultur des online Publizierens. Anders gesagt, für eine erfolgreiche Etablierung einer online Publikationskultur müssen die elektronischen Publikationen, vor allem aber die Erstpublikationen, als genuin wissenschaftliche Materialien die Fachwelt von der Notwendigkeit der neuen alternativen Publikationsmodelle überzeugen. Hierfür ist eine Vereinheitlichung der Beschaffungs-, Veröffentlichungs-, Erschließungs- und Archivierungsmethoden fachspezifisch wie fachübergreifend von besonderer Wichtigkeit. Die online Dokumente müssen inhaltlich entsprechend durchsuchbar und selbst als gesamtes Dokument im Netz auffindbar sowie sachlich erschlossen und langzeitarchiviert sein. Das prägt maßgeblich die Erwartungen an online Materialien, setzt aber darüber hinaus eine veränderte Arbeitsweise sowohl von den Autoren als auch von den Betreibern der Fachrepositorien voraus.

Auch in technischer Hinsicht werden bei der Erstellung von elektronischen Erstveröffentlichungen andere Prioritäten zu setzen sein. Die online Präsentation der Texte soll auf die Bedürfnisse der Bildschirmnutzung ausgerichtet sein. Dies bedeutet, dass die optisch attraktive und den etablierten Standards genügende Darbietung nicht minder wichtig ist als die Voraussetzungen für die Datenbankspeicherung und langfristige Archivierung.

Warum Osteuropa? Das Projekt OstDok[3]

Der Bereich der Osteuropastudien eignet sich wegen des regen wissenschaftlichen Austausches ganz besonders für den Aufbau eines fachspezifischen Repositoriums, wobei das Interesse gegenüber solchen Portalen gleichermaßen wie die Bereitschaft der individuellen und institutionellen Partizipation in den betroffenen Ländern hervorgehoben werden soll. In den osteuropäischen Ländern etablierten sich elektronische Publikationen in den vergan-

[3] Ausführlicher zum Projekt OstDok vgl. die Aufsätze u. a. von Johannes Gleixner und Norbert Kunz im Literaturanhang.

genen Jahren sehr stark. Dies führte zu einer größeren Bereitschaft osteuropäischer Autoren, das Angebot der Fachrepositorien zur Veröffentlichung ihrer Werke wahrzunehmen. Umgekehrt ermöglichen solche Portale den westlichen Osteuropawissenschaftlern, stärker im osteuropäischen Raum präsent zu sein. Werden die online Informationen dem Nutzer unabhängig von der örtlichen und/oder finanziellen Situation dauerhaft und unentgeltlich zur Verfügung gestellt, so wird hiermit auch dem Prinzip Open Access Rechnung getragen.

Freilich gestalten sich wegen des Mangelns bzw. Fehlens ähnlicher Fachrepositorien die Vergleichsmöglichkeiten und die Vereinheitlichung der Aufbaumethoden äußerst schwierig. Zwar bieten viele wissenschaftliche Institutionen einschlägige osteuropabezogene Materialien an, die wenigsten davon erfüllen jedoch die weithin akzeptierten Qualitätsstandards. Dazu gehören etwa technisch eine professionell aufgebaute elektronische Umgebung und inhaltlich einschlägige Fachservices wie der Zugriff auf aktuelle Dissertations- und Habilitationsprojekte, osteuropabezogene Aufsatzdatenbanken, elektronische Publikationen sowie umfangreiche Kataloge wissenschaftlich relevanter Webressourcen.[4]

Mit dem Projekt OstDok soll somit ein auf die ost-, ostmitteleuropäische Geschichte orientiertes Fachportal aufgebaut werden, das neben Publikationen der Projektpartner[5] auch ausgewählte stark rezipierte Standardwerke für einen digitalen Lesesaal, Fachzeitschriften wie die „Bohemia“ vom Collegium Carolinum und die „Zeitschrift für Ostmitteleuropa-Forschung“ vom Herder-Institut sowie ausschließlich Digitale Publikationen wie Qualifikationsarbeiten darbieten wird. Betont sei in Hinblick auf die immer noch mangelnde Zugänglichkeit zu aktueller westlicher Literatur in Osteuropa gerade das umfangreiche Retrodigitalisierungsangebot des Portals.

Mit dieser umfassenden Struktur und einschlägigen Inhalten will OstDok somit zum einen Nachwuchsforscher, zum anderen aber auch etablierte Osteuropawissenschaftler ansprechen und ihnen die Möglichkeit der umfangreichen Recherche und des wissenschaftlichen Austausches bieten, vor allem aber als adäquates Veröffentlichungsplattform dienen. Sowohl den Wissenschaftlern, die in einer fachlich kompetenten Umgebung publizieren wollen, als auch den Nutzern, die sich auf den wissenschaftlichen Wert des Fachrepositoriums und der darin enthaltenen Informationen verlassen wollen, wird dadurch die Nutzung der auf mehreren Ebenen technisch, inhaltlich wie rechtlich geprüften und dauerhaft gesicherten online Materialien ermöglicht.

[4] An dieser Stelle sei das Projekt ViFaOst erwähnt, das dem Projekt OstDok vorausgegangen war und nicht nur bereits die erwähnten Fachservices, sondern auch aufwendig redigierte Abschluss- und Qualifikationsarbeiten anbietet. Zum Angebot von ViFaOst siehe ausführlicher: www.vifaost.de.

[5] Als Projektpartner fungieren neben der Bayerischen Staatsbibliothek (www.bsb-muenchen.de) das Collegium Carolinum in München (www.collegium-carolinum.de), das Herder-Institut in Marburg (www.herder-institut.de) und das Osteuropa-Institut in Regensburg (www.osteuropa-institut.de).

Digitale Reihe hervorragender Abschlussarbeiten

Sowohl im öffentlichen Bewusstsein als auch in der wissenschaftlichen Debatte und Publizistik hat sich über die Entwicklung der online Publikationskultur die Ansicht etabliert, dass angesichts der sinkenden Erwerbsfähigkeit der Bibliotheken und dem damit verbundenen schrumpfenden Angebot diese Art des Publizierens eine Alternative zu den immer teurer werdenden Printpublikationen darstellt, wobei das herkömmliche wissenschaftliche Publikationswesen dadurch nicht unwesentlich beeinträchtigt wird. Gleichzeitig entwickelten Open Access Projekte willkommene Lösungen für die kostengünstige Veröffentlichung von für den Druck nicht „geeigneten" Materialien wie z. B. Hochschulschriften, deren wissenschaftlicher Wert mitunter unbestritten und deren Veröffentlichung mit einer maximalen Sichtbarkeit, ja weltweiten Verbreitung, von der Fachgemeinschaft ausdrücklich erwünscht ist. Wenn selbst die Verlage mittlerweile über eine Veröffentlichung von Hochschulschriften nachdenken, wenngleich dies sehr wenige ausgewählte Arbeiten betrifft und in kleinen Auflagen geschieht, so etabliert sich auch hier der Gedanke, diese Veröffentlichungen nach einiger Zeit auch online und auf dem Wege Printing on Demand zu vermarkten.

Die Notwendigkeit der online Bereitstellung solcher Abschlussarbeiten ist von vielen Universitäten durchaus anerkannt, auch wenn in diesem Bereich bisher kaum praktische Arbeit geleistet wurde. Blickt man auf konkrete Fachdisziplinen, so sind meist nur spärliche Ergebnisse verfügbar. Es liegt oft an den Universitäten, durch die Entwicklung fachorientierter und fachübergreifender Repositorien die elektronische Publikationskultur zu etablieren, wobei gerade die an den Hochschulen entstehenden wissenschaftlichen Materialien die Grundlage für solche Repositorien bilden können. Nicht zuletzt würde damit auf der Ebene qualitätsgesicherter und hoch spezialisierter Fachrepositorien der kostengünstigen Produktion solcher wissenschaftlicher Arbeiten Rechnung getragen, die oft gar nicht im Print erscheinen.

Dennoch dient die konsequente Veröffentlichung der Abschluss- und Qualifikationsarbeiten auf den Servern vieler Universitäten fast ausschließlich dem Zweck der Datensicherung. Diese werden nur sehr eingeschränkt in die überregional und fachspezifisch eigens für diesen Zweck geschaffenen Repositorien eingespeist. Zudem mangelt es an ausreichender Verschlagwortung, fachlicher Zuordnung und vor allem einer redaktionellen Begleitung, was diese Arbeiten für die Fachöffentlichkeit weitgehend unattraktiv macht. Auch wenn die Zahl der gespeicherten Materialien kontinuierlich steigt, führt dies zum Ergebnis, dass diese immer noch sehr unübersichtlich, fachlich, regional- und epochalübergreifend unsortiert vorliegen und dementsprechend unzureichend benutzt werden. Von einer Benutzung für wissenschaftliche Zwecke oder einer Zitation kann hier keine Rede sein. Erschwerend hinzu kommen die meist unausreichenden Kenntnisse der angehenden Wissenschaftler über Fragen der Langzeitarchivierung und Zugänglichkeit, aber vor allem über die Rechtslage bei online Publikationen, die zu einer deutlichen Zurückhaltung, um nicht zu sagen zum weitgehenden Desinteresse der Autoren gegenüber dem online Publizieren führen.

Somit ist es offensichtlich, dass OstDok angesichts der im deutschsprachigen Raum jährlich entstehenden Vielzahl hervorragender Master- und Diplomarbeiten, die der Fachöffentlich-

keit oft verborgen bleiben, mit deren Veröffentlichung auf ein in diesem Bereich deutlich vorhandenes Defizit reagiert.

Durch die Veröffentlichung hervorragender Abschlussarbeiten stellt OstDok den Nachwuchswissenschaftlern eine erste qualitätsgeprüfte Publikation mit Garantie der Langzeitarchivierung in Aussicht, was zugleich eine bessere Sichtbarkeit auch für die jeweilige tragende Einrichtung im Netz bedeutet. Kommt eine international ausgerichtete Nutzungsumgebung für die Vorstellung und Diskussion der Projekte hinzu, so wird den Nachwuchswissenschaftlern dadurch die Möglichkeit geboten, ihre Forschungsansätze der Fachöffentlichkeit vorzustellen, an Forschungsdiskussionen teilzunehmen und vom kompetenten wissenschaftlichen Meinungsaustausch zu profitieren.

Dabei werden solche Kriterien wie die Sichtbarkeit der wissenschaftlichen Materialien, deren Archivierung und dauerhafte Bereitstellung sowie die Verlässlichkeit des Zugangs und der Dialog mit der wissenschaftlichen Gemeinschaft als wichtige Voraussetzungen für die Qualitätssicherung konsequent umgesetzt. Ein aufwendig zusammengestelltes Layout sorgt des Weiteren für die Wiedererkennbarkeit und einheitliche Gestaltung sowie die professionelle visuelle Darbietung der Publikationen und soll einen hohen Komfort bei der Nutzung am Bildschirm garantieren. Die Gewährleistung weiterer Merkmale wie etwa die Unveränderlichkeit und Nachvollziehbarkeit der online Dokumente ist wiederum als Voraussetzung für ihre Zitierbarkeit unerlässlich. Strebt die Digitale Reihe an, den Ansprüchen einer eigenständigen digitalen Buchreihe zu genügen, so wird zudem großer Wert auf zeitliche, geographische und thematische Schwerpunkte gelegt (etwa die Osteuropäische Geschichte des 18. bis 20. Jahrhunderts).

Einer der größten Kritikpunkte an elektronischen Publikationen, nämlich das Begutachtungsverfahren, das eine für die Printpublikationen ja bereits etablierte Vorgehensweise darstellt, wird im Rahmen der Digitalen Reihe von einem fachlich kompetenten Herausgeberrat – jeweils die Lehrstuhlinhaber der ost-, ostmitteleuropäischen Geschichte in München, Jena, Freiburg, Berlin und Leipzig – geleistet. Die Herausgeber werden die Abschlussarbeiten für die Veröffentlichung empfehlen und für ihre wissenschaftlich inhaltliche Qualität Sorge tragen. Der Redaktion der Digitalen Reihe obliegt es sodann, die Arbeiten technisch und strukturell an die bestehenden Standards anzupassen, so dass thematisch, inhaltlich und technisch hoch professionelle, wissenschaftliche Werke erwartet werden dürfen. Durch die Anbindung an den Katalog des Bayerischen Bibliotheksverbundes wird des Weiteren die bessere Sichtbarkeit dieser Arbeiten in der Fachöffentlichkeit ermöglicht, wovon vor allen Dingen die jüngere Generation der Wissenschaftler profitieren wird.

Ausblick

Es ist nunmehr deutlich geworden, dass sich in den letzten Jahren ein wesentlicher Wandel in der wissenschaftlichen Kommunikation und im Veröffentlichungsprozess vollzogen hat. Je teurer und komplizierter der Weg zum Verlag und zu den traditionellen Printpublikationen wird, desto attraktiver erscheinen die Herstellung, Bereitstellung und Archivierung von wissenschaftlichen Materialien auf den Onlineservern. Die Verlage bleiben zwar weiterhin als unverzichtbare Partner für die Veröffentlichung wertvoller Schriften präsent, für solche Publikationen, deren Veröffentlichung auf dem traditionellen Wege aus verschiedenen

Gründen nicht finanzierbar wäre, ist die online Veröffentlichung jedoch eine wertvolle Alternative.

Freilich bringen Vorurteile über eine vermeintlich mangelnde wissenschaftliche Qualität der online Materialien, mögliche Verletzungen des Urheberrechtes, Bedenken über Sperrfristen, Plagiate, nicht kontinuierliche und langfristige Archivierung, mitunter auch die nicht ausreichende bzw. fehlende Selektion wissenschaftlich hochwertiger Texte in der Flut der online Ressourcen den Verdacht mit sich, dass die Produktivität der Wissenschaft insgesamt sinke. Werden die elektronischen Publikationen mithin als „im Vergleich zur konventionellen Publikation über Printmedien eine verbesserte Form der wissenschaftlichen Kommunikation“[7] dargestellt, so sind wiederum das von Experten mehrmals angesprochene Fehlen mediengerechter Geschäftsmodelle und adäquater rechtlicher Rahmenbedingungen bei der Veröffentlichung und Verbreitung der wissenschaftlichen Materialien gleichwohl entscheidende Kritikpunkte.[8] Im weitesten Sinne betrifft das nicht nur die online Erstpublikationen, sondern auch etwa Open Access Zeitschriften, deren wissenschaftliche Eignung genauso in Frage gestellt wird.

Um dauerhaft eine ernstzunehmende elektronische Publikationskultur zu etablieren, sollten vor allem eine aufwändige redaktionelle Betreuung sowie neue allgemein anerkannte Methoden der online Verarbeitung wissenschaftlicher Texte den Publikationsprozess fortwährend begleiten. Im Idealfall wird dies auch zu einer wachsenden Zitierbarkeit dieser Dokumente führen. Die Tendenz des online Publizierens verändert sich jedenfalls dahingehend, auch im Netz hochprofessionelle, durch den redaktionellen Aufwand hohes Qualitätsniveau garantierende und anspruchsvoll präsentierte Texte anzubieten.

Eine zentrale Frage der elektronischen Publikationskultur, die die Fachkreise und die Öffentlichkeit beschäftigt, ist somit einerseits die Zielrichtung, die die Open Access Projekte anschlagen, andererseits aber auch die zufrieden stellenden Lösungen solcher Fragen wie nach der Bewahrung der Autorenrechte und des geistigen Eigentums – namentlich sollen Open Access Projekte nicht mehr als Bedrohung, sondern vielmehr als eine gleichberechtigte Plattform zur u. a. wissenschaftlichen Verwirklichung parallel neben traditionellen Modellen angesehen werden – sowie nach der Realisierung der besseren Zugänglichkeit, der besseren Vernetzung im internationalen Kontext und nicht zuletzt der optischen Attraktivität.

An den verantwortlichen Fachrepositorieren liegt es schließlich – beanspruchen diese über die Grenzen eines reinen Informationsportals hinaus als ernstzunehmende wissenschaftliche Plattform wahrgenommen zu werden – die Sicherung dieser Kriterien dauerhaft zu gewährleisten.

[7] Vgl.: Deutsche Initiative für NetzwerkInformation: Elektronisches Publizieren an Hochschulen – Empfehlungen. Verfügbar unter http://edoc.hu-berlin.de/series/dini-schriften/1-de/PDF/1-de.pdf.

[8] Andermann, H. & Degkwitz, A. (2003). Neue Ansätze in der wissenschaftlichen Informationsversorgung. Ein Überblick über Initiativen und Unternehmungen auf dem Gebiet des elektronischen Publizierens. S. 4. Verfügbar unter http://www.epublications.de/AP.pdf.

Literaturverzeichnis

Monographien

Andermann, H. & Degkwitz, A. (2003). Neue Ansätze in der wissenschaftlichen Informationsversorgung. Ein Überblick über Initiativen und Unternehmungen auf dem Gebiet des elektronischen Publizierens. DFG-Projekt „Perspektiven für den Bezug elektronischer Fachinformation in der Bundesrepublik Deutschland“ an der Universitätsbibliothek Potsdam. Verfügbar unter http://www.epublications.de/AP.pdf

Blum, C. (2007). Open Access als alternatives Publikationsmodell der Wissenschaft. In U. Rautenberg & V. Titel (Hrsg.). Alles Buch. Studien der Erlanger Buchwissenschaft, XXI. Universität Erlangen-Nürnberg. Verfügbar unter http://www.buchwiss.uni-erlangen.de/forschung/publikationen/Blum.pdf

Spindler, G. (Hrsg.). (2006). Rechtliche Rahmenbedingungen von Open Access-Publikationen. In Göttinger Schriften zur Internetforschung, Bd. 2. Göttingen: Universitätsverlag. Verfügbar unter http://www.univerlag.uni-goettingen.de/OA-Leitfaden/oaleitfaden_web.pdf

Aufsätze

Elektronisches Publizieren an Hochschulen – Empfehlungen. In Deutsche Initiative für NetzwerkInformation: Verfügbar unter http://edoc.hu-berlin.de/series/dini-schriften/1-de/PDF/1-de.pdf

Gleixner, J. (2010). OstDok – Elektronische Erstpublikationen im Open Access in der Praxis. In O. Hamann (Hrsg.) *„Integration durch Information – We love to inform you…“ 38. ABDOS Tagung Martin/Slowakei, 18. bis 21. Mai 2009. Referate und Beiträge.* Berlin: Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, 40-43.

Gradmann, S. (2007). Open Access – einmal anders. Zum wissenschaftlichen Publizieren in den Geisteswissenschaften. Erstveröffentlichung in *ZfBB. Zeitschrift für Bibliothekswesen und Bibliographie, 54(4-5),* 170-173. Verfügbar unter http://edoc.hu-berlin.de/oa/articles/reWDQUdy2bil/PDF/209UiQvDa8SCA.pdf

Graf, K. (2004). Wissenschaftliches E-Publizieren mit "Open-Access" – Initiativen und Widerstände. In *Historical Social Research, 29(1),* 64-75. Verfügbar unter http://hsr-trans.zhsf.uni-koeln.de/hsrretro/docs/artikel/hsr/hsr2004_599.pdf

Kunz, N. (2009). Elektronisches Publizieren und Open Access – Auch in den Osteuropawissenschaften des deutschsprachigen Raumes? In *Die Osteuropabibliothek der Zukunft. Das Bibliotheks- und Informationswesen zu Osteuropa vor neuen Herausforderungen. 37. ABDOS-Tagung Marburg, 26. bis 28. Mai 2008. Referate und Beiträge* Berlin: Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, 106-115.

Landes, L. (2009). Open Access und Geschichtswissenschaften – Notwendigkeit, Chancen, Probleme. In *Libreas. Library Ideas. Elektronische Zeitschrift für Bibliotheks- und Informationswissenschaft, 14(1).* Verfügbar unter http://www.libreas.eu/ausgabe14/024lan.htm

Mruk, K., Gradmann, S. & Mey, G. (2004). Open Access: Wissenschaft als Öffentliches Gut. In *FQS. Forum Qualitative Sozialforschung, 5(2).* Verfügbar unter http://www.qualitative-research.net/index.php/fqs/article/view/624/1351

Prömel, Hans J. (2005). Editorial. In *CMS-Journal, 27(12.08.2005) Open Access und elektronisches Publizieren,* 1. Verfügbar unter http://edoc.hu-berlin.de/cmsj/27/proemel-hans-juergen-1/PDF/proemel.pdf

Seadle, M. (2009). Die Zukunft des Wissenschaftlichen Publizierens. Editoral. In *CMS-Journal, 32(01.06.2009). Wissenschaftliches Publizieren im digitalen Zeitalter,* 3-4. Verfügbar unter http://edoc.hu-berlin.de/cmsj/32/seadle-michael-3/PDF/seadle.pdf

Schirmbacher, P. (2005). Open Access – die Zukunft des wissenschaftlichen Publizierens. In *CMS-Journal, 27(12.08.2005). Open Access und elektronisches Publizieren,* 3-7. Verfügbar unter http://edoc.hu-berlin.de/cmsj/27/proemel-hans-juergen-1/PDF/proemel.pdf

Ders. (2005). Die neue Kultur des elektronischen Publizierens. In *CMS-Journal, 27(12.08.2005). Open Access und elektronisches Publizieren,* 19-22. Verfügbar unter http://edoc.hu-berlin.de/cmsj/27/schirmbacher-peter-19/PDF/schirmbacher.pdf

Schröder, K. (2005). Das Projekt „Dissertationen Online“. In *CMS-Journal, 27(12.08.2005). Open Access und elektronisches Publizieren,* 48-50. Verfügbar unter http://edoc.hu-berlin.de/cmsj/27/schroeder-karin-48/PDF/schroeder.pdf

Rezensieren im Zeitalter des Web 2.0 recensio.net – Rezensionsplattform für die europäische Geschichtswissenschaft

Lilian Landes

Bayerische Staatsbibliothek
Zentrum für Elektronisches Publizieren (ZEP)
lilian.landes@bsb-muenchen.de

Zusammenfassung

Die geschichtswissenschaftliche Buchrezension steht vor großen Herausforderungen, zu denen neben einem rasant wachsenden Publikationsaufkommen und veränderten Rezeptionsgewohnheiten der jüngeren Generation auch die Internationalisierung der Wahrnehmung von Neuerscheinungen gehört. Die europaweit ausgerichtete Plattform „recensio.net“ verfolgt neben einem „klassischen“ Ansatz auch das Ziel, innovative, web 2.0-basierte Formen des (wissenschaftlichen) Rezensierens zu etablieren.

Zur aktuellen Situation des Rezensionsgeschäfts

Die wissenschaftliche Buchrezension gehörte schon immer zu jenen Textgenres, die Entwicklungen des jeweils betroffenen Fachs unmittelbar nach außen trugen und für die die Frage nach Publikationsgeschwindigkeit und -flexibilität wichtiger waren als für andere – zumal in den Geschichtswissenschaften, wo die Erscheinungsgeschwindigkeit von Forschungsergebnissen traditionell weniger im Zentrum steht, als dies in anderen Disziplinen der Fall ist, wo sich die Diskussion derselben im Kreis der Fachwissenschaftler oft aber in ähnlich gemächlichen Bahnen vollzieht, so dass es mitunter Jahre dauert, bis nach dem Erscheinen einer Schrift die zugehörige Rezension in einem Fachjournal erscheint. Gerade die Rezension scheint prädestiniert für die intensive Nutzung jener Vorteile, die das elektronische Publizieren unter Bedingungen des Open Access bietet.

Mit der zunehmenden Digitalisierung wissenschaftlicher Kommunikations- und Publikationswege werden sich strukturelle Veränderungen für das Rezensionswesen in den Geschichtswissenschaften ergeben. Die Diskussion über Neuerscheinungen wird nicht nur schneller, sondern ihrem Wesen nach zugleich partikularer, detailorientierter, interdisziplinärer, flexibler und insbesondere auch internationaler werden; sie wird in mittelfristiger Sicht aus dem Raster der gewohnten Dramaturgie traditioneller Buchbesprechungen ausbrechen. Bislang aber fehlt für die Erprobung solcher Verfahren das notwendige Instrumentarium, das sich an den Möglichkeiten und Erfahrungen des sogenannten „Web 2.0“ orientiert.

Zugleich wird der Markt für „traditionell“ verfasste, online oder auf Papier publizierte Rezensionen zunehmend unübersichtlich. Oft unter großem Aufwand erarbeitete und sorgfältig redigierte Besprechungen erfahren zu wenig Beachtung angesichts des wachsenden Angebots, des Wahrnehmungsverlustes von Printrezensionen im Kreis der jüngeren Wis-

senschaftler und einer oft nicht gut sicht- und findbaren Onlinepublikation von Rezensionen kleinerer Fachzeitschriften, denen die Anbindung an die zentralen Suchinstrumente des Wissenschaftlers fehlt oder die eine Ausstattung der Rezensionen mit suchrelevanten Metadaten nicht leisten können.

Online-Rezensionsjournale verzeichnen stetig wachsende Nutzungsraten. Mit den sehepunkten ist etwa auf nationaler Ebene seit Jahren ein Erfolgsmodell für offenen und raschen Zugriff auf Buchbesprechungen etabliert worden. Dieses Modell nun schlicht auf eine internationale Ebene zu heben, würde dahingehend am Ziel vorbeigehen, als dass damit zwar der aktuell immer noch zu selten praktizierte grenzübergreifende Blick auf Neuerscheinungen gefördert, zugleich aber die Menge bestehender Rezensionsorgane grundsätzlich erhöht, die Unübersichtlichkeit des Rezensionsmarkts eher verstärkt denn reduziert, sowie die Erprobung innovativer Besprechungsmodi aufgeschoben würde.

recensio.net

Mit der Schaffung einer internationalen Plattform für geschichtswissenschaftliche Rezensionen reagieren die Bayerische Staatsbibliothek, das Deutsche Historische Institut Paris und das Institut für Europäische Geschichte Mainz auf die geschilderte Situation. Schon 2005 hat Manfred Hildermeier, ehemaliger Vorsitzender des „Verbandes der Historiker und Historikerinnen Deutschlands“, von einem nicht weiter erörterungsbedürftigen „Gemeinplatz“ gesprochen, „dass nationalstaatliche Grenzen in der Regel [...] von der Sache her als Gliederungsprinzip historischer Forschung und historiographischer Darstellungen ausgedient haben“ [1].

Mit *recensio.net* wird im Januar 2011 eine europaweite, mehrsprachige Open-Access-Plattform für Rezensionen geschichtswissenschaftlicher Literatur entstehen. *recensio.net* wird kein Rezensionsjournal sein, sondern als internationale Plattform „klassische Rezensionen“ bestehender Print- und E-Journale zusammenführen und parallel Web 2.0-basierte Formen wissenschaftlichen Rezensierens erproben.

Ansatz eins: die „klassische“ Rezension

Diese beiden Grundgedanken werden *recensio.net* zu einem Arbeitsinstrument machen, das dem traditionellen Textgenre „Buchrezension“ Raum gibt: Dabei bleibt das gewohnte Layout aller beteiligten Zeitschriften erhalten, indem diese ihre Rezensionsteile als PDF-Dateien auf *recensio.net* veröffentlichen. Zugleich sind die einzelnen Journale neben der übergreifenden Plattformsuche auch gezielt einzeln ansteuerbar. Dabei ist *recensio.net* als ein Ort der Zusammenführung, nicht als exklusiver Publikationsort „klassischer“ Rezensionen gedacht, so dass einer parallelen Onlinepublikation auf anderen Websites oder auf Papier nichts im Wege steht. Auch die Frage, ob *recensio.net* als Pre- oder Postprint-Publikationsort genutzt werden soll – ersteres würde die eingangs angesprochene Möglichkeit zur schnelleren Netzpublikation ausschöpfen – oder auch die Frage, ob der Rezensionsteil einzelner Zeitschriften gänzlich ins Netz ausgelagert wird und als HTML-Text auf der Platt-

[1] Manfred Hildermeier, Deutsche Geschichtswissenschaft: im Prozess der Europäisierung und Globalisierung, in: zeitenblicke 4 (2005), Nr. 1, [12.12.2010], http://www.zeitenblicke.de/2005/1/hildermeier/index.html.

form erscheint, bleibt den einzelnen Redaktionen überlassen. Diese arbeiten weiterhin vollständig autark. Ihr Zugewinn wird eine bessere Sichtbarkeit der Rezensionen sein, die *recensio.net* im Bibliothekskatalog des BVB verlinkt sowie im Volltext recherchierbar macht.

Immer mehr Verlage erkennen den positiven Einfluss, die im Open Access publizierte, qualitativ überzeugende Buchbesprechungen auf die Außenwirkung von Fachzeitschriften haben. Dies führt zu der angenehmen Situation, dass bezüglich der sonst oft heiß diskutierten Grundsatzfrage nach elektronischer OA-Publikation in Bezug auf Rezensionen alle Beteiligten an einem Strang ziehen und gleichermaßen profitieren: Der Verlag von einem Renommeegewinn, die Autoren (der besprochenen Schrift, wie der Rezension) von mehr Sichtbarkeit und Publikationsgeschwindigkeit, welche nicht zuletzt ebenso dem Leser zugute kommt.

Ansatz zwei: die „lebendige“ Rezension

Andererseits wird *recensio.net* zugleich jenen Veränderungen Rechnung tragen, die sich mit der Etablierung des Web 2.0 im Alltag der Netznutzer (vor allem der jüngeren) mittelfristig vom kommerziellen auch auf den (geschichts-)wissenschaftlichen Buchmarkt übertragen werden: Autoren erhalten die Möglichkeit, die Kernthesen ihrer Schriften auf *recensio.net* zu publizieren. Moderierte Nutzerkommentare lassen nach und nach „lebendige Rezensionen“ und Diskussionen rund um die angezeigte Veröffentlichung entstehen. Die Möglichkeiten, die das Netz zur raschen, diskussionsaffinen Publikation bietet, sollen auf diese Weise besser als zuvor ausgeschöpft werden. Es wird möglich sein, lediglich zu Einzelaspekten einer Schrift Stellung zu nehmen, wodurch letzthin auch fächerübergreifende Blicke auf Neuerscheinungen erleichtert werden. Zudem wird dem immer enger getakteten Arbeitsalltag des Historikers entsprochen, dem immer weniger Zeit zur Ausarbeitung vollständiger Rezensionen bleibt, der aber dennoch qualifiziert zu Entwicklungen seines Fachgebiets Stellung nehmen und umgekehrt seine eigenen Thesen durch Kollegen diskutiert sehen möchte. In den Redaktionen „klassischer“ Fachrezensionen wird diese Entwicklung in den letzten Jahren anhand der immer schwierigeren Gewinnung von Rezensenten deutlich spürbar. Dennoch stehen die beiden Grundideen von *recensio.net* nicht konkurrierend, sondern einander ergänzend nebeneinander: Die „klassische“ Rezension wird ihren Platz behalten und durch präsentationsbasierte „lebendige“ Rezensionen ergänzt. Status und Notwendigkeit von „living documents“, also von prozessbetonten Veröffentlichungsformen anstelle der statischen Qualität üblicher Veröffentlichungsformen, wurde im Hinblick auf die wissenschaftliche Nutzung des Internets bereits früh konstatiert.[2]

Auch die Frage nach der Art der Texte, die Autoren auf *recensio.net* präsentieren und kommentieren (lassen) können, ist eine zentrale: Nicht nur die traditionelle Monographie, sondern auch in Zeitschriften oder Sammelbänden publizierte Aufsätze sollen in der Fachcommunity diskutiert werden. Der Aufsatz verzeichnet als geschichtswissenschaftliches Textgenre in den letzten Jahrzehnten gegenüber dem traditionell dominierenden „Buch“ einen erheblichen Bedeutungsgewinn (was sicher ebenso sehr im Vorbild der Naturwis-

[2] Vgl. etwa Michael Nentwich, Cyberscience. Research in the age of the internet, Wien, Austrian Academy of Sciences Press 2003.

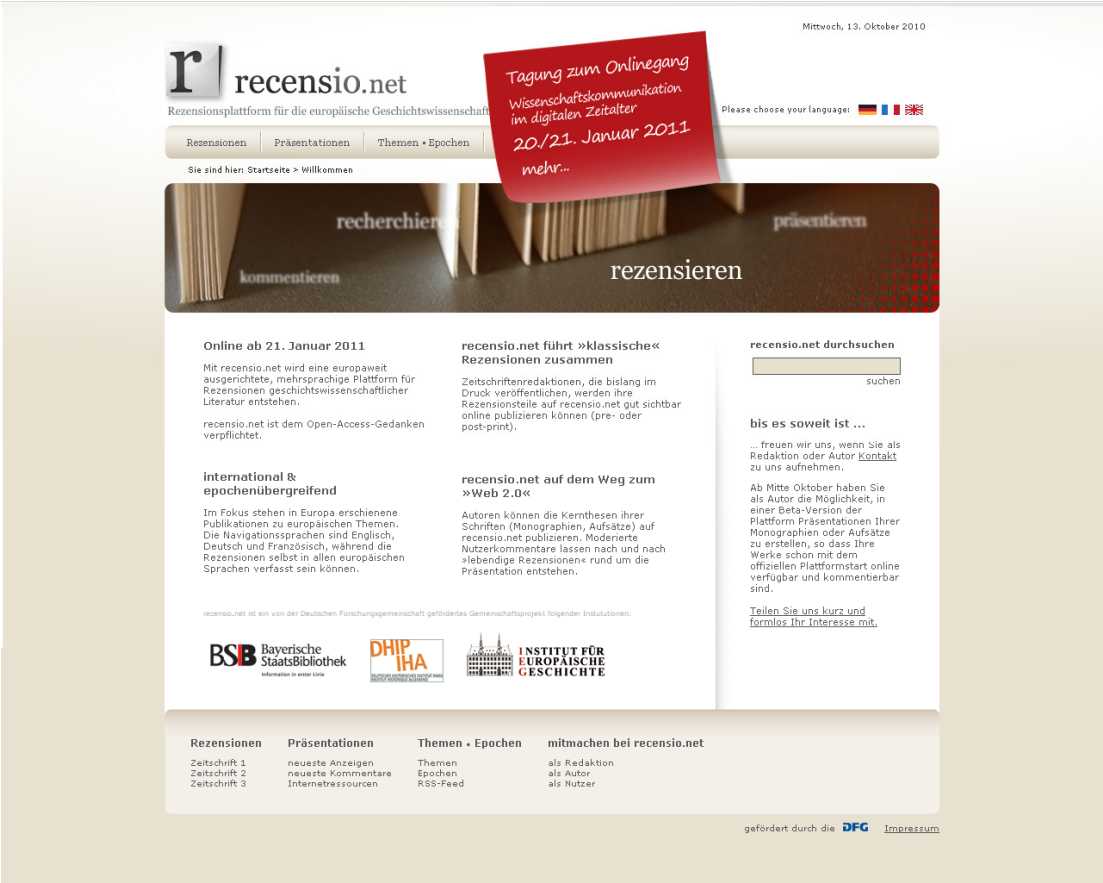
senschaften ,wie in den projektorientierten Strukturen der Forschungsförderung begründet ist). Diese Entwicklung soll damit berücksichtigt werden. Zusätzlich können Internetressourcen für die historische Forschung (Quellensammlungen, Forschernetzwerke, Bibliographien und vieles mehr) auf *recensio.net* präsentiert werden. Durch die zeitlich unbegrenzte Kommentarmöglichkeit wird der permanenten Veränderung, der das Medium „Internetressource“ unterworfen ist, Rechnung getragen.

Prinzipiell ist diese zweite Säule des Plattformkonzepts auf die aktive Beteiligung von Fachwissenschaftlern angewiesen. *recensio.net* ist sich bewusst darüber, dass die Web 2.0-Funktionalitäten der Plattform sicher eine gewisse Zeit des Anlaufens brauchen werden, zumal gerade unter traditionellen Fachvertretern die Befürchtung nicht ausgeräumt ist, dass diese Form der Kommunikation untrennbar mit einem inhaltlichen Qualitätsverfall einhergeht. Prinzipiell lebt der Gedanke des Web 2.0 von der Vorstellung einer weitgehenden Selbstregulierung von Auswüchsen, die den kommunizierten Inhalt oder die Umgangsformen betreffen. Insbesondere während der Startphase – in der die Kommentarfrequenz wachsen und diesen Mechanismus sicherstellen muss – wird die *recensio.net*-Redaktion dennoch nicht nur eingehende Präsentationen, sondern auch Kommentare prüfen, bevor diese online gehen.

Ein Instrument zum „Anschub“ der Diskussionskultur wird die Kontaktierung jener Wissenschaftler sein, mit denen sich ein Autor beschäftigt hat, der seine Schrift auf der Plattform präsentiert. Diese gibt er im Rahmen des Formulars zur pointierten Formulierung seiner Kernthesen als „Bezugsautoren“ an, so dass sie von der *recensio.net*-Redaktion kontaktiert und auf die Kommentarmöglichkeit aufmerksam gemacht werden können. Der Autor wird automatisch über den Eingang eines Kommentars zu seiner Präsentation benachrichtigt und kann so unmittelbar reagieren (worüber wiederum der Kommentierende informiert werden wird).

Projektstruktur

recensio.net ist ein von der DFG gefördertes Gemeinschaftsprojekt der Bayerischen Staatsbibliothek (BSB) München, des Deutschen Historischen Instituts Paris (DHIP) und des Instituts für Europäische Geschichte (IEG) Mainz unter der Leitung von Prof. Dr. Gudrun Gersmann. Inhaltlich stehen in Europa erschienene Publikationen zu europäischen Themen im Fokus. Sowohl „klassische“ Rezensionstexte als auch Präsentationen und Kommentare werden in allen europäischen Sprachen publiziert, während die Plattformnavigation selbst dreisprachig angeboten werden wird (Englisch, Deutsch, Französisch).



Interessierte Zeitschriften, Autoren und künftige Nutzer sind jederzeit eingeladen, mit *recensio.net* in Kontakt zu treten.

Fazit

Zum Zeitpunkt der Einreichung dieses Beitrags scheint die Hoffnung, mit der Gründung von recensio.net einen nachhaltigen Beitrag zur Internationalisierung und Flexibilisierung des „Schreibens über Schriften“ leisten zu können, berechtigt. Zahlreiche Redaktionen rezensionspublizierender Fachzeitschriften äußern Kooperationsinteresse. Wissenschaftler wiederum sehen mit Erleichterung einer Zusammenführung und Durchsuchbarmachung von Rezensionen im Open Access entgegen, die sicher nicht nur in den historischen Disziplinen ein Desiderat darstellt angesichts wachsender Publikationsmengen. Kontaktaufnahmen zu ähnlichen Initiativen anderer Fächer liegen sehr im Interesse der Redaktion von *recensio.net*.

RIHA Journal
Ein Beispiel für eine internationale Open Access-Zeitschrift für Kunstgeschichte

www.riha-journal.org

Ruth von dem Bussche-Hünnefeld

Talstraße 116
D-40217 Düsseldorf
rbussche@fotostoria.de

Regina Wenninger

Zentralinstitut für Kunstgeschichte
Katharina-von-Bora-Straße 10
D-80333 München
r.wenninger@zikg.eu

Zusammenfassung

Mit dem RIHA Journal, einer Open Access-Zeitschrift für Kunstgeschichte, ist im April 2010 ein internationales Gemeinschaftsprojekt von 28 Forschungsinstituten in 19 Ländern an den Start gegangen. Der Beitrag skizziert die Besonderheiten der Redaktionsstruktur, die technische Umsetzung sowie einige Herausforderungen des Projekts.

Journal of the International Association of Research Institutes in the History of Art
RIHA JOURNAL

Herausgeber & Redaktionsstruktur

Mit dem RIHA Journal, einer Open Access-Zeitschrift für Kunstgeschichte (www.riha-journal.org), ist im April 2010 ein ebenso innovatives wie ambitioniertes Publikationsprojekt an den Start gegangen: Die Zeitschrift wird gemeinschaftlich herausgegeben von den Mitgliedsinstituten von RIHA, der International Association of Research Institutes in the History of Art; sie ist damit ein Gemeinschaftsvorhaben von derzeit 28 verschiedenen kunsthistorischen, meist außeruniversitären Forschungseinrichtungen in 18 europäischen Ländern und den USA.[1]

[1] 1998 in Paris gegründet, gehören zu RIHA heute die folgenden 28 Mitgliedsinstitute (vgl. http://www.riha-institutes.org/members.htm): Amatller Institute of Hispanic Art, Barcelona; Institute of Art History of the Slovak Academy of Sciences, Bratislava; Royal Institute for Cultural Heritage (IRPA-KIK), Brüssel; Research Institute for Art History of the Hungarian Academy of Sciences, Budapest; „George Oprescu“ Institute for Art History, Buka-

Eine damit verbundene Besonderheit stellt auch die dezentrale Redaktionsstruktur dar: Anders als üblich gibt es nicht eine einzelne Redaktion, sondern jedes der 28 beteiligten Institute ist jeweils lokal für die redaktionellen Abläufe verantwortlich – vom Autorenkontakt über die Organisation des Review-Verfahrens bis zur Textredaktion.

Das Zentralinstitut für Kunstgeschichte (ZI) in München hat zusätzlich die Aufgabe einer Zentralredaktion übernommen, die u. a. die Endformatierung übernimmt, die Beiträge koordiniert und freischaltet, das gemeinsame Redaktionssystem betreut und zentraler Ansprechpartner für alle Beteiligten ist.

Regelmäßig stattfindende Workshops dienen der technischen Schulung der Redakteure sowie Erfahrungsaustausch und Diskussion. Die Treffen tragen zudem wesentlich dazu bei, die 28 äußerst unterschiedlichen Institutionen mit ihren unterschiedlichen Kapazitäten, unterschiedlichen Publikationskulturen und unterschiedlichen Erwartungen bezüglich der gemeinsamen Zeitschrift zu integrieren.

Gefördert wird das Projekt durch den Beauftragten der Bundesregierung für Kultur und Medien (BKM).

Grundzüge des RIHA Journals

Ziel des RIHA Journals ist es, aktuelle Forschungsergebnisse einem internationalen Fachpublikum schnell, dauerhaft und kostenfrei zugänglich zu machen und dabei die Vorzüge von Online-Publikationen einerseits mit höchsten Ansprüchen an wissenschaftliche Qualität andererseits zu verbinden.

Im Einzelnen heißt dies: Es werden ausschließlich aktuelle erstklassige Forschungsartikel veröffentlicht, die ein anonymes Begutachtungsverfahren („double blind peer review") erfolgreich durchlaufen haben. Die Namen der Gutachter werden ggf. zusammen mit dem Artikel ebenfalls veröffentlicht.

Entgegen der zunehmenden Spezialisierung von Fachzeitschriften will das RIHA Journal das Fach in seiner ganzen Breite und Vielfalt repräsentieren und steht dem gesamten Spektrum kunsthistorischer und kunstwissenschaftlicher Themen und Ansätze offen.

rest; Netherlands Institute for Art History (RKD), Den Haag; Visual Arts Research Institute Edinburgh (VARIE), Edinburgh; Kunsthistorisches Institut in Florenz – Max-Planck-Institut, Florenz; The Danish National Art Library, Kopenhagen; International Cultural Centre, Krakau; France Stele Institute of Art History, Ljubljana; Scientific Research Center of the Slovenian Academy of Sciences and Arts, Ljubljana; Courtauld Institute of Art, London; The Warburg Institute, London; The Getty Research Institute, Los Angeles; Zentralinstitut für Kunstgeschichte (ZI), München; Deutsches Forum für Kunstgeschichte – Centre allemand d'histoire de l'art, Paris; Institut national d'histoire de l'art (INHA), Paris; Institute of Art History, Prag; Bibliotheca Hertziana – Max-Planck-Institut für Kunstgeschichte, Rom; Istituto Nazionale di Archeologia e Storia dell'Arte, Rom; Nationalmuseum, Stockholm; Fondazione Giorgio Cini, Istituto di Storia dell'Arte, Venedig; Institute of Art of the Polish Academy of Sciences, Warschau; Center for Advanced Study in the Visual Arts (CASVA), Washington, D. C. ; Kommission für Kunstgeschichte an der Österreichischen Akademie der Wissenschaften, Wien; Clark Art Institute, Williamstown, MA; Institute of Art History, Zagreb; Schweizerisches Institut für Kunstwissenschaft (SIK-ISEA), Zürich.

Bevorzugte Publikationssprachen sind die fünf offiziellen CIHA[2]-Sprachen Deutsch, Englisch, Französisch, Italienisch und Spanisch. Weitere Sprachen sind nach Rücksprache möglich.

Neben Originalbeiträgen werden auch Übersetzungen von jüngeren Artikeln, die ursprünglich in einer Nicht-CIHA-Sprache erschienen sind, veröffentlicht.

Als Open Access-Zeitschrift stellt das RIHA Journal alle Artikel – versehen mit der Creative Commons-Lizenz CC-BY-NC-ND 3.0. – ohne Einschränkung kostenfrei zur Verfügung, sowohl als HTML-Seite als auch als PDF.

Jeder Artikel erhält neben der URL eine eigene URN und ist damit dauerhaft im Netz auffindbar und zitierfähig.

Zur Langzeitarchivierung werden die PDF/As der Artikel derzeit noch manuell an die Deutsche Nationalbibliothek in Frankfurt/Main geliefert. An der Integration einer OAI-Schnittstelle wird gearbeitet.

Als Online-Zeitschrift bietet das RIHA Journal seinen Autoren eine Reihe von Vorteilen: Die freie Zugänglichkeit begünstigt die höhere Sichtbarkeit der eigenen Forschung; die technischen Möglichkeiten des Mediums bieten zusätzliche Darstellungsoptionen (Verlinkungen, Einbettung von Video-Files etc.); zudem erlaubt es einen großzügigeren Umgang mit Artikellänge und Abbildungszahl: Selbst umfangreiche Beiträge oder auch Artikel mit Quellenanhang, die kaum Chancen hätten, in Print-Zeitschriften aufgenommen zu werden, können publiziert werden. Ein weiterer wesentlicher Vorteil ist die kurze Bearbeitungsdauer. Das RIHA Journal ist bestrebt, geeignete Manuskripte innerhalb von drei Monaten ab Einreichung zu veröffentlichen. Wissenschaftler können ihre Forschungsergebnisse auf diese Wiese erheblich rascher an die Fachöffentlichkeit bringen als mit herkömmlichen Print-Zeitschriften, bei denen sich der Redaktionsprozess unter Umständen über Jahre hinziehen kann. Dementsprechend erscheint das RIHA Journal nicht periodisch in Ausgaben, sondern laufend artikelweise. Über neue Beiträge können sich Leser per RSS Feed auf dem Laufenden halten. Zudem werden die Artikel über einschlägige Mailinglisten wie H-Arthist angekündigt.

Technische Umsetzung

Das RIHA Journal basiert auf dem Open Source Content Management System Plone, das zusammen mit dem Open Source Applicationserver ZOPE betrieben wird.[3] Es dient einerseits als Publikationsplattform, andererseits bildet es als virtuelle Arbeitsumgebung die Schnittstelle zwischen den Redaktionen und unterstützt den gesamten redaktionellen Workflow. Alle eingesetzten Komponenten, einschließlich der zusätzlichen, publikationsspezifischen Komponenten, sind Open Source.[4]

Verschiedene Gesichtspunkte waren ausschlaggebend für die Wahl von Plone. Generell lassen sich seine Grundfunktionalitäten gut an die Bedürfnisse wissenschaftlichen elektronischen Publizierens anpassen. Darüber hinaus stellt es ein relativ leicht bedienbares CMS zur

[2] CIHA = Comité International d'Histoire de l'Art (http://www.esteticas.unam.mx/CIHA/).
[3] Siehe http://plone.org/.
[4] Die öffentlichen Releases sind hier verfügbar: http://pypi.python.org/pypi/psj.site.

Verfügung und unterstützt die gängigen Office-Formate – beides entscheidende Aspekte angesichts der rund 30 am RIHA Journal beteiligten Redakteure, die in erster Linie Wissenschaftler sind, technisch ganz unterschiedliche Erfahrungen und Kompetenzen mitbringen und über die ganze Welt verteilt sind, was die technische Hilfestellung erschwert.

Zudem konnte mit der Entscheidung für Plone in weiten Teilen auf ältere Open Source-Entwicklungen zurückgegriffen werden, insbesondere auf Programmierungen, die für die Plattform perspectivia.net[5] realisiert worden waren.

Eine RIHA Journal-spezifische Anpassung stellt der Workflow dar, über den die Mitglieder in den verschiedenen Instituten weltweit verbunden sind und der auch das gesamt Review-Verfahren mit einschließt.

Ausgangsformate der Artikel bilden Office-Dokumente (DOC, DOCX, ODT); beim Hochladen werden diese mittels einer Transformationsengine in die beiden genannten Zielformate transformiert: HTML-Seite und PDF. Das bedeutet, dass die Artikel fertig formatiert sind und der Layout-Prozess abgeschlossen ist, bevor die Artikel in die Website geladen werden. Durch die Verwendung von Office-Dateien sind den gestalterischen Möglichkeiten zwar gewisse Grenzen gesetzt; umgekehrt haben sie, wie oben angedeutet, den Vorteil, dass die Redakteure mit vertrauten Programmen arbeiten können und sich nicht eigens in Layoutprogramme einarbeiten müssen.

Das RIHA Journal setzt als wesentliche Komponente das Produkt Plone Scholarly Journal (PSJ) ein, ein Zusatzprodukt für Plone, das die Grundelemente für das elektronische Publizieren liefert. Neben der erwähnten Transformationsengine sind dies insbesondere eigene Contenttypen, d.h. Dokumenttypen wie Artikel, Rezension, Buch, Monografie usw., die jeweils über anpassbare, frei konfigurierbare Metadatensets verfügen; das RIHA Journal verwendet die beiden Contenttypen „psj document“ und „psj review“.

Der angebundene Transformationsprozess wird mit OpenOffice.org realisiert. Dies hat den Vorzug, dass das Journal auch künftige Office Formate einsetzen kann und damit Zukunftssicherheit gewährleistet ist – es genügt, auf eine höhere Version von OpenOffice.org zu wechseln.

Herausforderungen

Neben der allgemeinen Schwierigkeit, eine neu gegründete Zeitschrift als konkurrenzfähiges Publikationsorgan zu etablieren, stellen die nach wie vor bestehenden Vorbehalte gegenüber Online-Publikationen, zumal Open Access-Publikationen, eine der größten Herausforderungen dar. Zwar tragen Online-Medien, ob in Form von Bilddatenbanken oder Textpublikationen, längst wesentlich zur kunsthistorischen Forschung bei, und unter den zahllosen Online-Zeitschriften zur Kunstgeschichte finden sich wissenschaftlich anspruchsvolle Zeitschriften renommierter Institutionen. Dennoch haben es Online-Publikationen zumindest innerhalb der Geisteswissenschaften weiterhin schwer, sich gegenüber etablierten Print-Organen zu behaupten. Zur individuellen Skepsis einzelner Wissenschaftler treten strukturelle Hindernisse in einzelnen Ländern. So finden etwa in Polen Online-Journale aus

[5] Siehe http://www.perspectivia.net.

formalen Gründen keinen Eingang in die Listen offiziell anerkannter akademischer Zeitschriften. Entsprechend gering ist der Anreiz für Wissenschaftler, in Online-Zeitschriften zu publizieren. Verstärkt wird dies durch den auch in den Geisteswissenschaften zunehmenden Druck, in prestigeträchtigen Journals zu publizieren, will man in der wissenschaftlichen Karriere vorankommen. Online-Zeitschriften schneiden dabei in der Regel nach wie vor schlecht ab.

Mit dem RIHA Journal trägt RIHA daher nicht nur der zunehmenden Bedeutung Rechnung, die Online-Publikationen für die kunsthistorische Forschung heute zukommt; mit seinen hohen Qualitätsstandards, strengen Auswahlverfahren sowie dem internationalen Renommee der beteiligten Institute will die Zeitschrift auch einen Beitrag leisten, jenen Vorbehalten und institutionellen Hürden entgegenzuwirken.

Ausblick

Technisch ist neben der Integration einer OAI-Schnittstelle (s. o.) geplant, die RIHA Journal-Artikel auch für die Nutzung an elektronischen Lesegeräten zu optimieren. Zudem sind intensivere Vernetzungen mit größeren Online-Plattformen in Vorbereitung.

Persönliche Publikationslisten im WWW – Webometrische Aspekte wissenschaftlicher Selbstdarstellung am Beispiel der Universität Bielefeld

Najko Jahn, Mathias Lösch und Wolfram Horstmann

Universität Bielefeld
Universitätsbibliothek
Universitätsstraße 25
D-33615 Bielefeld
{najko.jahn, mathias.loesch, wolfram.horstmann}@uni-bielefeld.de

Zusammenfassung

Persönliche Publikationslisten sind ein Bestandteil wissenschaftlicher Selbstdarstellung im Web. Der Beitrag stellt sich die Frage, wie sich wissenschaftliche Dokumente aus den persönlichen Webseiten heraus identifizieren und für empirische Untersuchungen im Rahmen von Linkanalysen nutzen lassen. Am Beispiel der Universität Bielefeld wurden 1.358 wissenschaftliche Volltexte eruiert und ihre Rezeption über zwei exemplarische Linkkontexte – Wikipedia und dem politischen Diskurs im Web – exploriert.

Einleitung

Das Web hat nicht nur die Art und Weise verändert, wie Wissenschaftler publizieren, sondern auch wie sie ihre Forschungsergebnisse und mithin sich selbst im Web darstellen. Die persönliche Publikationsliste ist für diese Entwicklung ein altbekanntes Beispiel, das disziplinübergreifend verbreitet ist. Eingebettet in den persönlichen CV einer Wissenschaftlerin oder eines Wissenschaftlers ist die persönliche Publikationsliste eine besonders kondensierte Darstellung wissenschaftlicher Aktivitäten. Neben der klassischen Funktion als Referenz auf die erschienenen Forschungsbeiträge nutzen Forschende zusätzlich die persönliche Publikationsliste, um auf Dokumente zu verlinken, sie verfügbar zu machen oder diese sogar als originären Erscheinungsort für die erste Fassung von zur Veröffentlichung bestimmten Beiträgen zu verwenden.

Eine Vielzahl an Services hat das Potential persönlicher Publikationslisten in den letzten Jahren aufgegriffen. Philpapers.org stellt ein prominentes Beispiel dar, wie aggregierte Volltexte, die von persönlichen Publikationslisten stammen, in ein kollaboratives Angebot innerhalb einer Fachgemeinschaft, hier die Philosophie, aufgenommen und von Fachkollegen erschlossen und diskutiert werden. Ein disziplinübergreifendes Beispiel ist researchGate.net. Der Social Network Service konzentriert sich neben kollaborativen Werkzeugen auf die wirkungsmächtige Außendarstellung eines Forschenden. Und auch Hochschulen greifen zunehmend das Potential der wissenschaftlichen Selbstdarstellung über die persönliche Publikationsliste auf (Horstmann & Jahn, 2010).

Ziel des Beitrages ist die beispielhafte Exploration der Selbstdarstellung wissenschaftlicher Aktivitäten im Web anhand von persönlichen Publikationslisten an der Universität Bielefeld. Insbesondere die Rezeption der Dokumente, die sich auf den persönlichen Publikationslisten befinden, ist von besonderem Interesse, da die freie Verfügbarkeit wissenschaftlicher Veröffentlichungen und ihre Vernetzung im Web ‚prima facie' Nachnutzungsszenarien über die Grenzen der persönlichen Fachgemeinschaft hinaus ermöglichen.

Die Exploration geschieht methodisch über eine Linkanalyse. Die Bibliotheks- und Informationswissenschaft erhebt Linkanalysen im Rahmen webometrischer Fragestellungen. Die Webometrie widmet sich beispielsweise den Webpräsenzen von akademischen Einrichtungen und Universitäten, um Linkmaße für die Außenwirkung von Forschenden zu validieren (Thelwall & Ruschenburg, 2006). Aufgrund der Komplexität der universitären Webpräsenzen und der Vielzahl an Linkkontexten nehmen wissenschaftliche Volltexte auf persönlichen Webseiten nur einen geringen Stellenwert in der webometrischen Debatte ein. Untersuchungen legen nahe, dass sich nur ein verschwindend geringer Anteil der Links auf universitäre Webpräsenzen überhaupt als ein intendierter Bezug auf die wissenschaftliche Erkenntnis interpretieren lässt (ebd.).

Um somit ein aussagekräftiges Bild der Rezeption wissenschaftlicher Volltexte auf persönlichen Publikationslisten zu erlangen, bedarf es des dezidierten Fokus auf persönliche Listen und die in ihnen enthaltenen wissenschaftlichen Dokumente. Und um Linkkontexte festzustellen, wird wiederum eine Gruppierung der Links benötigt. Somit stellen wir uns zunächst die Frage, wie sich wissenschaftliche Dokumente aus den persönlichen Webseiten heraus identifizieren lassen. Anschließend aggregieren wir die Links, die auf die Dokumente verweisen, und stellen exemplarisch zwei Linkkontexte der Selbstdarstellung wissenschaftlicher Veröffentlichungen auf persönlichen Webseiten vor: die Wikipedia und der politische Diskurs im Web.

Datengewinnung

Datengrundlage unserer Untersuchung ist ein Korpus aus wissenschaftlichen Publikationen, das wir von persönlichen Publikationslisten Bielefelder Wissenschaftler für eine frühere Studie zusammengestellt hatten (Jahn, Lösch & Horstmann, 2010). Ausgangspunkt war ein zuvor manuell erstelltes Verzeichnis der Publikationslisten von 1.214 Bielefelder Wissenschaftlern. Das Verzeichnis wurde insofern eingeschränkt, als wir nur Personen berücksichtigten, die zum Zeitpunkt der Erhebung promoviert waren. Das Korpus wurde dann erstellt, indem frei zugängliche PDF-Dokumente automatisiert von den persönlichen Publikationslisten der Wissenschaftler aggregiert und intellektuell als Publikationen klassifiziert wurden. Letzteres war nötig, weil nicht alle PDF-Dokumente, die auf solchen Webseiten veröffentlicht werden, notwendigerweise Publikationscharakter haben – man findet dort auch Lebensläufe, Seminarunterlagen, Präsentationen und andere, nicht wissenschaftlich relevante Dokumente. Die Beschränkung auf PDF-Dokumente wurde pragmatisch aus der Erfahrung der Autoren begründet, dass die meisten digitalen Volltexte als PDF veröffentlicht werden. Das Ergebnis dieser Prozedur ist ein Korpus der frei zugänglichen Dokumente der Bielefelder Forschenden. Zusätzlich zu den eigentlichen Dokumenten wurden auch deren URLs gespeichert. Diese bilden den Ausgangspunkt der vorliegenden Studie.

Um herauszufinden, welche Webseiten Hyperlinks auf die von uns als wissenschaftliche Publikationen verifizierten Dateien enthalten, verwendeten wir deren URLs als Eingabe für das API von Yahoo! Site Explorer (http://siteexplorer.search.yahoo.com/). Dabei handelt es sich um einen Service, der es erlaubt, Informationen über Webseiten aus dem Index der Suchmaschine Yahoo! auszulesen und insbesondere welche Seiten auf die Datei mit einer gegebenen URL zeigen (Backlinks). Wir verwendeten ein eigens entwickeltes Python-Skript, um die Backlinks für alle Dokumente in unserem Korpus zu aggregieren. Das Skript ordnet jedem Dokument seinen Autor und die entsprechenden Backlinks zu.

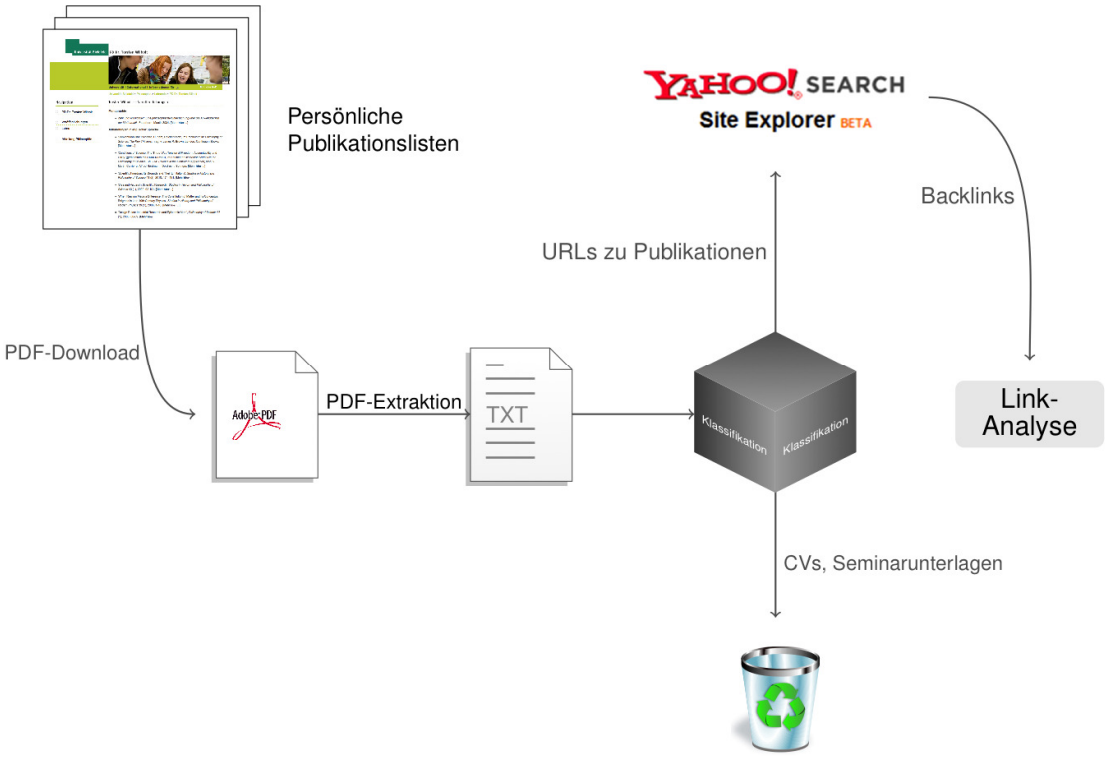


Abbildung 1: Aggregation von Backlinks auf wissenschaftliche Volltexte

Link-Netzwerke

Insgesamt konnten wir 1.358 wissenschaftliche Dokumente aggregieren, die sich auf 138 Wissenschaftler an der Universität Bielefeld verteilen. Auf die wissenschaftlichen Dokumente wiederum entfallen in der Summe 1.175 Backlinks, die über Yahoo! Site Explorer identifiziert wurden. Die weitere Exploration erfolgt mit Methoden der Sozialen Netzwerkanalyse (SNA). Die Visualisierungen wurden mithilfe der Software visone realisiert (Baur, 2008).

Zwei Aggregationsstufen standen hierbei im Vordergrund, um den Datenbestand stärker zu strukturieren. Zunächst lässt sich die persönliche Publikationsliste als Ganzes analog zu einer Forscher-Bibliographie auffassen. Die Backlinks können dann als ein Indikator für die thematische Ausrichtung eines Forschenden herangezogen werden. Des Weiteren ist es möglich, die Backlinks unter ihrer Domain zusammenzufassen. Damit lassen sich wiederum Domains, von denen die Links auf die wissenschaftlichen Volltexte stammen, als Indikator für die thematische Rezeption eines Dokuments heranziehen.

Es gehen daher in beide Untersuchungen jeweils zwei Mengen an Untersuchungsgegenständen ein, die bipartit angeordnet werden. Eine bipartite Anordnung bedeutet grob, dass Beziehungen im Netzwerk über Knoten einer Menge über die Verbindung zu einem Knoten der anderen Menge instantiiert werden. Das bipartite Arrangement wird auch für Zitationsstudien und die Untersuchung von Kollaborationsbeziehungen über Koautorschaften in der Bibliometrie verwendet (Havemann, 2009).

Wikipedia

Unser erstes Netzwerk, das Backlinks auf wissenschaftliche Volltexte als ein Indikator für die thematische Ausrichtung der Forschenden präsupponiert, basiert auf Backlinks, die aus der Wikipedia stammen. Die Wikipedia ist in der Vielzahl der Linkkontexte, die wir eruiert haben, aufgrund ihres enzyklopädischen Charakters und ihrer großen Popularität ein gutes Indiz für die themenbezogene Sichtbarkeit eines Forschenden.

Die Visualisierung des Netzwerks (Abbildung 2) offenbart die Eigenschaft der bipartiten Anordnung. Die grünen Knoten repräsentieren die persönliche Publikationsliste eines Forschenden, die violetten Knoten stehen für die Wikipedia-Artikel. Da die Wikipedia nach Sprachen unterteilt ist, fassen wir Backlinks aus Artikeln in mehreren Sprachen zusammen. So finden sich Werke eines Forschenden in Wikipedia-Beträgen zu „Xenoturbella“, eine Wurmart, die besonders für Genomforscher interessant ist. Publikationen dieses Autors werden in sechs Sprachversionen des Wikipedia-Artikels referenziert. Analog gehen wir vor, wenn ein Wikipedia-Beitrag mehrere Beiträge eines Forschenden zitiert.

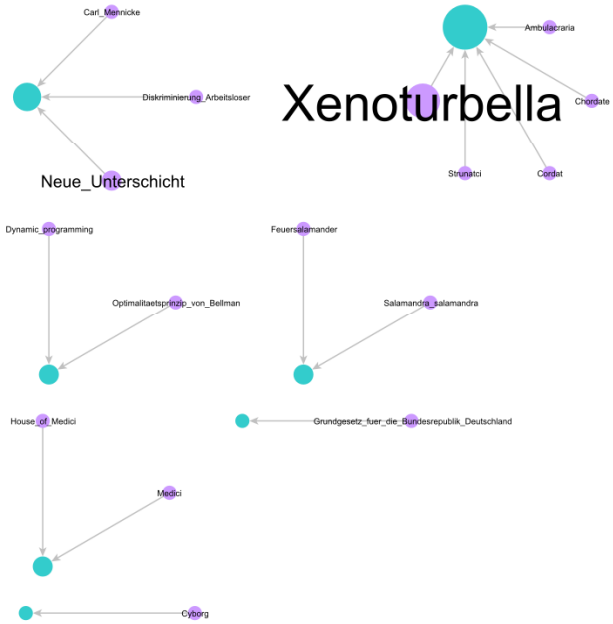


Abbildung 2: Linkkontext Wikipedia

Im Netzwerk ist die Anzahl der ein- bzw. ausgehenden Links eines Knoten über sein Größenverhältnis zu den anderen Knoten visualisiert. Das zugrundeliegende Maß ist in der SNA die Dichte, die die Anzahl der direkten Beziehungen eines Knotens berechnet.

Politischer Diskurs im Web

Das folgende Netzwerk ordnet Dokumentenlinks, die sich auf persönlichen Publikationslisten wiederfinden lassen, Domains zu, die ebenfalls auf das Dokument verweisen. Dabei muss ein Eintrag in einer persönlichen Publikationsliste nicht notwendigerweise selbstarchiviert sein, sondern kann auch auf ein bereits im Web veröffentlichtes Dokument verweisen.

Ein prominentes Beispiel in diesem Zusammenhang ist eine Unterrichtung für die Bundesregierung durch einen Bielefelder Juristen im Umfeld des Gesetzgebungsverfahrens des Telemediengesetzes. Parlamentarische Drucksachen werden von der Parlamentsdokumentation des Deutschen Bundestages fortlaufend erfasst, erschlossen und im DIP (Dokumentations- und Informationssystem für Parlamentarische Vorgänge) frei zur Verfügung gestellt.

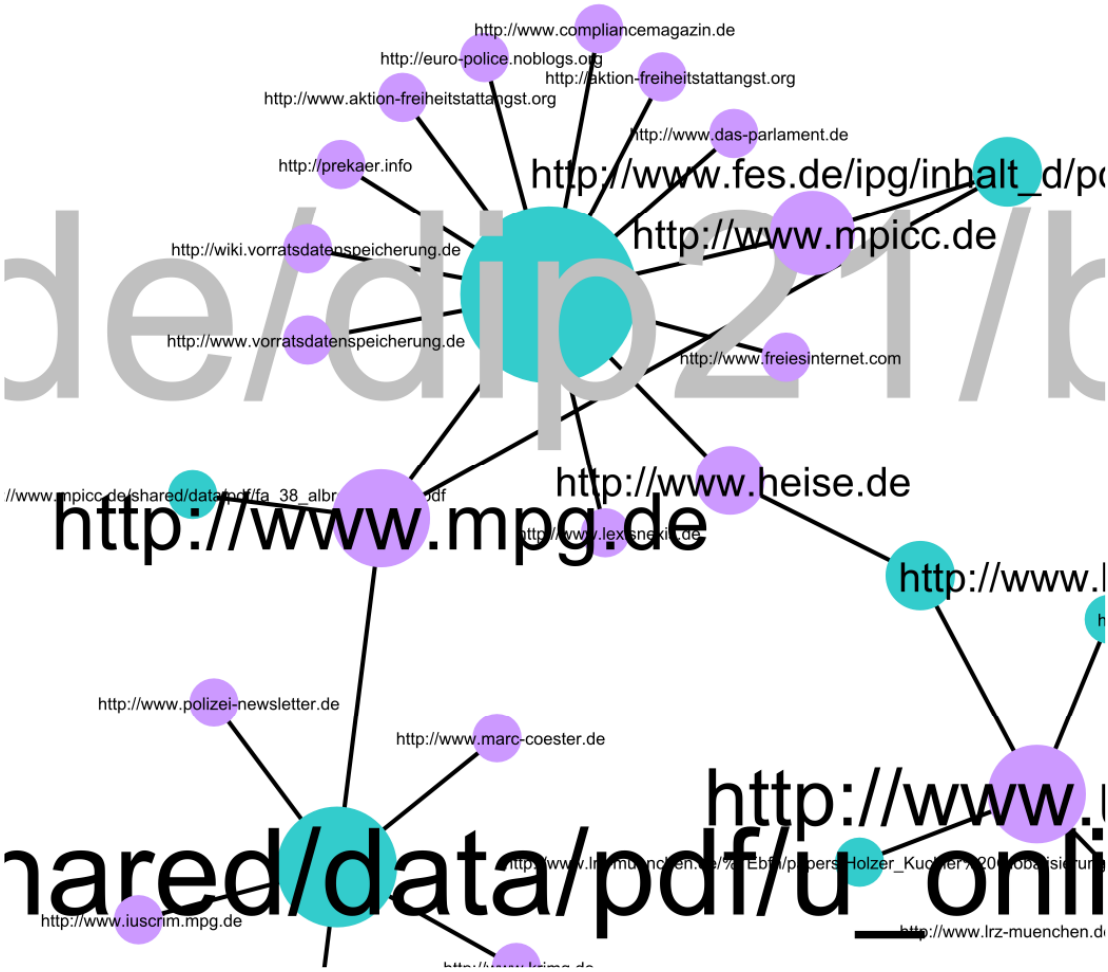


Abbildung 3: Linkkontext politischer Diskurs im Web

Die vorliegende Unterrichtung weist die formalen Merkmale einer wissenschaftlichen Arbeit auf, um den parlamentarischen Gesetzgebungsprozess zu informieren. Ihre Rezeption ist dabei besonders spannend, da sie sowohl von akademischen Einrichtungen im Umfeld der Rechts- und Politikwissenschaft als auch von Medien, in privaten Blogs und auf Seiten von Nichtregierungsorganisationen rezipiert wird.

Im Ausschnitt wird noch eine weitere Netzwerkeigenschaft deutlich: Domains, die auf die wissenschaftlichen Dokumente der persönlichen Seiten verweisen, koppeln diese Dokumente und bilden dadurch eine weitere Ebene der Darstellung wissenschaftlicher Veröffentlichungen an der Universität Bielefeld, die die Selbstdarstellung des einzelnen Forschenden ergänzt.

Fazit

Persönliche Publikationslisten im Web sind ein wichtiger Aspekt in der Selbstdarstellung wissenschaftlicher Aktivitäten. Verlinkte oder selbstarchivierte wissenschaftliche Volltextdokumente offenbaren eine Vielzahl an Untersuchungs- und Interpretationsmöglichkeiten. Die in der Arbeit explorierten Teilgraphen weisen Beziehungsgefüge auf, die sowohl eine thematische Einordnung wissenschaftlicher Autoren als auch die Bestimmung von Kontexten ihrer Rezeption unterstützen.

Insgesamt erscheinen Linkanalysen, die ihre Daten über verschiedene Aggregationsstufen gewinnen, ein vielversprechender methodischer Ansatz für das wachsende Feld der Digitalen Wissenschaft. Linkanalysen bieten Daten und Modelle für ihre interdisziplinären Fragestellungen und haben zugleich das Potential, die Entwicklung neuer Dienste durch Bibliotheken und Informationsdienstleister zu begleiten und anzuregen.

Literatur

Baur, Michael (2008). visone – Software for the Analysis and Visualization of Social Networks. Universität Karlsruhe. urn:nbn:de:swb:90-108975 .

Havemann, Frank (2009). Einführung in die Bibliometrie. Berlin: Ges. für Wiss.-Forschung c/o Inst. für Bibliotheks- und Informationswiss. urn:nbn:de:kobv:11-10097376 .

Horstmann, Wolfram und Najko Jahn (2010). Persönliche Publikationslisten als hochschulweiter Dienst – Eine Bestandsaufnahme. BIBLIOTHEK Forschung und Praxis 34, no. 2, 185-193. doi:10.1515/bfup.2010.032 .

Jahn, Najko, Mathias Lösch und Wolfram Horstmann (2010). Automatic Aggregation of Faculty Publications from Personal Web Pages. The Code4Lib Journal, September 21. http://journal.code4lib.org/articles/3765 .

Thelwall, Mike und Tina Ruschenburg (2006). Grundlagen und Forschungsfelder der Webometrie. Information – Wissenschaft und Praxis 57, no. 8, 401-406.