

In die Zukunft publizieren: Herausforderungen an das Publizieren und die Informationsversorgung in den Wissenschaften ; 11. Kongress der IuK-Initiative der Wissenschaftlichen Fachgesellschaften in Deutschland

Stempfhuber, Maximilian (Ed.)

Veröffentlichungsversion / Published Version
Konferenzband / conference proceedings

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Stempfhuber, M. (Hrsg.). (2006). *In die Zukunft publizieren: Herausforderungen an das Publizieren und die Informationsversorgung in den Wissenschaften ; 11. Kongress der IuK-Initiative der Wissenschaftlichen Fachgesellschaften in Deutschland* (Tagungsberichte / Informationszentrum Sozialwissenschaften, 11). Bonn: Informationszentrum Sozialwissenschaften. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-261151>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

In die Zukunft publizieren

Herausforderungen an das Publizieren und die Informationsversorgung in den Wissenschaften

**11. Kongress der luK-Initiative der
Wissenschaftlichen Fachgesellschaft in Deutschland**

Maximilian Stempfhuber (Hrsg.)

Tagungsberichte Band 11

In die Zukunft publizieren

Tagungsberichte

Herausgegeben vom Informationszentrum Sozialwissenschaften (IZ)
der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V. (ASI), Bonn.
Band 11

Das IZ ist Mitglied der Gesellschaft Sozialwissenschaftlicher
Infrastruktureinrichtungen e.V. (GESIS).
Die GESIS ist Mitglied der Leibniz-Gemeinschaft.

IuK - Initiative der Wissenschaftlichen Fachgesellschaften in Deutschland
Die einzelnen Artikel dieses Tagungsbandes sind als Open Access-Publikatio-
nen unter <http://www.iuk-initiative.org/> abrufbar.

Ausrichtung der Tagung unterstützt von:
Informationszentrum Sozialwissenschaften: <http://www.gesis.org/iz>

In die Zukunft publizieren
Herausforderungen an das Publizieren und die
Informationsversorgung in den Wissenschaften

10. Kongress der IuK-Initiative
der Wissenschaftlichen Fachgesellschaften in Deutschland

Maximilian Stempfhuber (Hrsg.)

Tagungsbericht Band 11

Informationszentrum Sozialwissenschaften, Bonn 2006

Bibliographische Information Die Deutsche Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliothek; detaillierte bibliographische Daten sind im Internet über www.ddb.de abrufbar.

ISBN 3-8206-0150-3

Herausgeber, Druck und Vertrieb:
Informationszentrum Sozialwissenschaften
Lennéstraße 30, 53113 Bonn
Tel.: 02 28 - 22 81 - 0
Printed in Germany

© 2006 Informationszentrum Sozialwissenschaften, Bonn. Alle Rechte vorbehalten. Insbesondere ist die Überführung in maschinenlesbare Form sowie das Speichern in Informationssystemen, auch auszugsweise, nur mit schriftlicher Einwilligung gestattet.

Inhalt

Vorwort	7
-------------------	---

Urheberrecht und Open Access

<i>Gerd Hansen</i> Urheberrecht für Wissenschaftler	9
--	---

<i>Rainer Kuhlen, Jochen Brüning</i> Potenziale von Creative Commons-Lizenzen für Open Innovation	21
--	----

<i>Alexis-Michel Mugabushaka</i> Open Access : die strategische Kluft?	29
---	----

<i>Christopher N. Carlson</i> Open Access oder Fair Use?	43
---	----

<i>Sandra Schaffert, Bernhard Schmidt</i> Open Access in der Bildungsforschung	55
---	----

Elektronisches Publizieren

<i>Patrick Danowski</i> Kollaboratives elektronisches Publizieren nach Wiki-Art	69
--	----

<i>Marco Jorio</i> In die Zukunft über die Vergangenheit publizieren – Das Historische Lexikon der Schweiz (HLS)	81
--	----

Informationssysteme und Suchmaschinen

<i>Doris Bambej</i> infoconnex – Angebotsstrategien für solche, die Nutzer werden wollen . . .	87
---	----

<i>Stefan Jakowatz</i> Neue Wege der Wissenspräsentation und –vermittlung in der Virtuellen Fachbibliothek Sozialwissenschaften (ViBSoz): Thematische Dokumentationen	99
--	----

<i>Gisela Minn, Anke S. Hoffmann, Yvonne Rommelfanger</i> Themenorientierte Fachinformationssysteme in den historischen Wissenschaften	113
--	-----

Christiane Baier, Peter Weiland

PsychSpider – Erfahrungen aus dem Betrieb einer spezialisierten
Suchmaschine 127

Ulrich Herb

Volltextserver für wissenschaftliche Dokumente aus der Psychologie:
PsyDok 145

Jochen Johannsen

Die Virtuelle Fachbibliothek Sportwissenschaft 157

Heike Schulte-Derne

Interdisziplinäre Aufsatzdatenbanken und ihre Verlinkung zum
(elektronischen) Volltext 171

Gregor Fels

eTeaching and eLearning in Chemistry. 185

Stephan Körnig

Generische Portale: Webbasierte Digital Library-Dienste für die
Wissenschaft 191

*Anne-Kathrin Walter, Philipp Mayr, Maximilian Stempfhuber,
Arne Ballay*

Crosskonkordanzen als Mittel zur Heterogenitätsbehandlung in
Informationssystemen 205

Wolfram Sperber

Suchmaschinen, Strukturierung und Metadaten 227

Philipp Mayr, Anne-Kathrin Walter

Google Scholar - wie tief gräbt diese Suchmaschine? 241

Forschungsprogramme und Evaluierung

Stefan Hornbostel

Forschungsrankings: Artefakte oder Sichtbarkeit der Forschungsarbeit? . . 263

Johannes Fournier

In die Zukunft publizieren. Positionen und Reflexionen. 279

Verzeichnis der Autorinnen und Autoren 295

Vorwort

Auch zur 11. Jahrestagung der Initiative Information und Kommunikation wissenschaftlicher Fachgesellschaften (IuK) erscheint zusätzlich zu den Einzelartikeln, die unter www.iuk-initiative.org als Open Access-Publikationen verfügbar sind, ein gedruckter Tagungsband. Etwa aus der Hälfte der referierten Beiträge der Tagung sind letztlich Beiträge zum Tagungsband hervorgegangen.

Die Jahrestagung 2005 stand unter dem Motto „In die Zukunft publizieren – Herausforderungen an das Publizieren und die Informationsversorgung in den Wissenschaften“. Der thematische Rahmen umspannte – angefangen beim Urheberrecht und Open Access – wichtige und innovative Aspekte des elektronischen Publizierens und der Informationsversorgung in den Wissenschaften, bis hin zur immer stärker an Bedeutung gewinnenden Evaluation wissenschaftlicher Leistung – und den dazu notwendigen Voraussetzungen im Publikationsprozess.

Der Tagungsband dokumentiert das Engagement der IuK-Initiative für die Verbesserung der Informationsversorgung in den Wissenschaften, dient als Plattform für die Kommunikation innovativer und richtungsweisender Ideen, und formuliert die Themen, zu denen innerhalb – und gerade auch zwischen – den deutschen Fachgesellschaften eine Standort- und Richtungsbestimmung notwendig erscheint.

Das positive Echo auf die Tagung von den Teilnehmerinnen und Teilnehmern sowie in den Medien hat die Relevanz der dort vorgetragenen Inhalte belegt. Insofern gilt unser Dank allen Vortragenden für die Gestaltung der Konferenz und den Autorinnen und Autoren für das Verfassen der Beiträge für den Tagungsband. Dank gilt auch dem Programmkomitee und den Mitarbeiterinnen und Mitarbeitern des IZ Sozialwissenschaften für die lokale Organisation und die Produktion des Tagungsbandes.

Maximilian Stempfhuber

IuK

Urheberrecht für Wissenschaftler

Risiken und Chancen der Urheberrechtsreform für das wissenschaftliche Publizieren

Gerd Hansen

Max-Planck-Institut für Geistiges Eigentum, Wettbewerbs- und Steuerrecht,
München

Zusammenfassung

Im Rahmen der laufenden Urheberrechtsreform steht der kontrovers diskutierte Vorschlag einer Anbietungspflicht für an Hochschulen beschäftigte Urheber im Raum. Wissenschaftler sollen danach verpflichtet werden, ein im Rahmen ihrer Lehr- und Forschungstätigkeit entstandenes Werk der Hochschule zur Veröffentlichung anzubieten. Im ersten Teil des Beitrags werden die insbesondere im Hinblick auf die Wissenschaftsfreiheit, Art. 5 Abs. 3 GG, bestehenden, erheblichen Bedenken gegen eine solche Anbietungspflicht ausgeleuchtet. Den Schwerpunkt des Beitrags bildet sodann der zweite Teil, in dem nach verfassungskonformen, alternativen urheberrechtlichen Regelungsansätzen für einen offeneren Zugang zu den Resultaten aus überwiegend öffentlich finanzierter Forschung und Lehre gesucht wird. Mit diesem Ziel werden verschiedene denkbare Instrumente wie eine neue Schrankenregelung (§ 52c UrhG *de lege ferenda*) durchgespielt. Der Beitrag mündet in einen konkreten urhebervertragsrechtlichen Vorschlag für einen neuen § 38 Abs. 1 Satz 3 UrhG, welcher dem Wissenschaftler das vertraglich nicht abdingbare Recht geben würde, seinen wissenschaftlichen (Zeitschriften-)Beitrag auch bei Einräumung ausschließlicher Nutzungsrechte nach Ablauf von grundsätzlich sechs Monaten seit Erstveröffentlichung z.B. auf seiner Website öffentlich zugänglich zu machen (Stichwort: „Self-Archiving“)¹.

1 Der Beitrag basiert auf einem im Rahmen der IuK-Jahrestagung am 09. Mai 2005 in Bonn gehaltenen Vortrag. Eine vertiefte juristische Auseinandersetzung zu den in diesem Beitrag angeschnittenen Fragen findet sich in dem Aufsatz des Verfassers, „Zugang zu wissenschaftlicher Information – alternative urheberrechtliche Ansätze“, GRUR Int. 2005, 378ff.

1 Einleitung

Die Reaktion der meisten Wissenschaftler auf Open Access lautet intuitiv: „Ja, klar, wie kann man gegen einen offenen Zugang zu Wissen sein?“ Die Zustimmung zu Open Access ist aus dem Bauch heraus groß. Nur schnell kommt die Frage: wie ist das mit den Rechten? Diese Unsicherheit lässt (zu) viele Wissenschaftler vor einer eigenen Open Access-Publikation zurückschrecken. Um hier Abhilfe zu schaffen, soll in diesem Beitrag nach urheberrechtlichen Lösungswegen gesucht werden.

2 Open Access

Wenn wir zwar potentiell immer mehr Informationen, aber immer weniger Informierte haben, droht die Informationsgesellschaft zur Farce zu werden. Die Ursachen für die paradoxe Situation, dass wir immer mehr wissen könnten, aber immer weniger wissen, sind vielfältiger Natur und vielfach beschrieben: die alltägliche Informationsflut, die auf uns hereinstürzt, immer kürzer werdende Aufmerksamkeitsspannen etc. Nicht zuletzt ist auch die sog. „Krise des wissenschaftlichen Publikationswesens“ wohl ein Grund für die suboptimale wissenschaftliche Informationsversorgung und -rezeption. Mit der „Krise des wissenschaftlichen Publikationswesens“ wird der Umstand umschrieben, dass sich die Preise für wissenschaftliche Fachveröffentlichungen in den letzten rund zehn Jahren explosiv erhöht haben. Sie haben sich offensichtlich derart alarmierend entwickelt, dass sich viele Bibliotheken die teuren Journals nicht mehr leisten können. Denn: die öffentlichen Kassen sind leer. Bemerkenswert ist dabei, dass diese Krise eine weltweite ist und sich ausnahmsweise nicht etwa Deutschland als der „kranke Mann Europas“ dafür verantwortlich machen lässt.

Das Absurde an dieser Situation ist, dass die öffentliche Hand weite Teile von Wissenschaft und Forschung an den Hochschulen finanziert. Profiteure des wissenschaftlichen Ausstoßes unserer Hochschulen sind aber nicht etwa die Steuerzahler, sondern primär die Wissenschaftsverlage. Diese Situation empfinden viele als unbefriedigend. Deshalb mehren sich die Stimmen, die sagen, die Resultate aus öffentlich finanzierter Forschung und Lehre sollten stärker öffentlich zugänglich sein². Das System scheint aus dem Gleichgewicht.

2 Siehe Asschenfeldt/Honekamp, ZRP 2004, 247; Stellungnahme des Aktionsbündnisses „Urheberrecht für Bildung und Wissenschaft“ zum RefE, S. 6, http://www.urheberrecht.org/topic/Korb-2/st/refentw/AB_Urheberrecht_BuW_261104.pdf; s.a. die Website der „Alliance for Taxpayer Access“, <http://www.arl.org/ata/>.

In den letzten ein, zwei Jahren hat sich der Widerstand formiert. Man spricht von der sog. Open Access-Bewegung³. Die Parolen lauten: „Open Access to Scientific Research!“ oder: „Public Information wants to be free!“ Es ist eine weltweite, gut vernetzte Bewegung, die rasant wächst. Sie verfolgt ihr Ziel im Wesentlichen mittels zweier Strategien: Open-Access-Journals und Self- bzw. Open-Access-Archiving. Erstere sollen sich über Publikationsgebühren finanzieren, um der Allgemeinheit einen kostenfreien Zugang zu ermöglichen. Self- bzw. Open-Access-Archiving bedeutet, dass Wissenschaftler selber oder durch Dritte ihre Beiträge online archivieren.

In Deutschland hat die sog. „Berliner Erklärung“ im Herbst 2003 bislang die größte Aufmerksamkeit für das Thema Open Access hervorgerufen und in der deutschen wissenschaftspolitischen Debatte verankert⁴. Eine große Zahl renommierter nationaler und internationaler Forschungseinrichtungen hat sich hier zusammengetan, um sich auf eine Strategie für einen offenen Zugang zu wissenschaftlicher Information zu verständigen. Zusammen mit den ähnlich gerichteten Erklärungen von Budapest und Bethesda besteht somit eine klare Definition dessen, was Open Access bedeutet (sog. BBB-Nenner)⁵. Kurz gesagt erlaubt ein Autor danach bei einer Publikation seines Beitrages gemäß der Open-Access-Prinzipien die unbegrenzte Vervielfältigung, Verbreitung, öffentliche Zugänglichmachung (§ 19a UrhG) und Bearbeitung seines Beitrages, um eine maximale Distribution seiner Ideen zu erreichen. Einige der Vorschläge, die unter dem Label „Open Access“ kursieren, sind insofern kein Open Access. Die nachfolgend abgehandelte Anbieterspflicht ist dafür ein Beispiel. Und auch der in diesem Beitrag vorgestellte urhebervertragsrechtliche Ansatz (§ 38 Abs. 1 Satz 3 UrhG) ist allenfalls „Open Access Light“, weil er nur die Option für eine Open-Access-Zweitveröffentlichung schafft und nicht dazu verpflichtet. Die zu beobachtende Popularisierung und Verwässerung des Begriffs „Open Access“ mag letztlich ein Erfolgssymptom sein, wie es jede Bewegung ereilt.

3 Anbieterspflicht für Urheber an Hochschulen

Auf Betreiben der Kultusministerkonferenz ist der Vorschlag für eine Anbieterspflicht für Urheber an Hochschulen in die (urheber-) rechtspolitische Diskussion eingebracht worden. Inspiriert durch § 42 Arbeitnehmererfindungsgesetz möchte man Autoren verpflichten, an Werken im Sinne von § 2 Abs. 1 Nr. 1 UrhG, die im Rahmen einer Lehr- und Forschungstätigkeit entstanden sind, der

3 Siehe die instruktive Einführung zum Thema „Open Access“ von Peter Suber unter: <http://www.earlham.edu/~peters/fos/overview.htm>.

4 Siehe http://www.mpg.de/pdf/openaccess/BerlinDeclaration_dt.pdf.

5 BOAI, <http://www.eifl.net/boai/boai.html>; Bethesda Statement on Open Access Publishing: <http://www.earlham.edu/~peters/fos/bethesda.htm>.

Universität ein einfaches Nutzungsrecht einzuräumen⁶. Ziel ist also eine obligatorische Zweitveröffentlichung in einem Hochschulverlag oder einem Online-Archiv der Hochschule.

3.1 Verfassungsrechtliche Bedenken

Dieses Ansinnen begegnet erheblichen verfassungsrechtlichen Bedenken. Durch eine solche Anbietungspflicht würde massiv in die mit der Wissenschaftsfreiheit (Art. 5 Abs. 3 GG) untrennbar verknüpfte Publikationsfreiheit eingegriffen. Wissenschaft ist ohne Mitteilung nicht denkbar. Der Wissenschaftler hat das Recht, über das Ob, das Wo und das Wie der Veröffentlichung seiner Forschungsergebnisse zu entscheiden. Diese Freiheiten würden erheblich eingeschränkt.

Durch eine Anbietungspflicht wäre zunächst einmal die negative Publikationsfreiheit betroffen, wenn der Urheber verpflichtet werden soll, jedes im Rahmen seiner Lehr- und Forschungstätigkeit entstandene Werk der Hochschule zur Veröffentlichung anzubieten.

Zwar besteht theoretisch die Möglichkeit einer verfassungskonformen Abschwächung, wenn man die Anbietungspflicht durch eine Andienungs- bzw. Anzeigepflicht für bereits zur Veröffentlichung vorgesehene Beiträge herunterstuft. Das Gegenargument hierzu drängt sich aber auf. Wie sollen die Universitäten einen solchen (zusätzlichen) bürokratischen Aufwand bewältigen? Die Erfahrungen mit § 42 Arbeitnehmererfindungsgesetz sind nicht unbedingt geeignet, diesbezüglich übergroße Hoffnungen zu hegen.

In vielen Disziplinen ist zudem die Publikation in einem international anerkannten „High-Quality-Journal“ maßgeblich. Als Beispiele seien hier nur „Nature“ oder „Science“ angeführt. Die Beschränkung auf eine Hochschulpublikation führt daher zu einer Behinderung des internationalen wissenschaftlichen Vorankommens gerade auch von Nachwuchswissenschaftlern und ist nicht im Interesse des Standorts Deutschland. Dies gilt auch dann, wenn man sich bei der gesetzlichen Ausgestaltung mit der Einräumung eines einfachen Nutzungsrechts begnügt. Durch ein Beschränken der Anbietungspflicht auf ein einfaches Nutzungsrecht bliebe dem Wissenschaftler nämlich nur eine Scheinfreiheit erhalten zu entscheiden, wo er seinen Artikel publiziert. In der Praxis kann er von dieser vermeintlichen Freiheit keinen Gebrauch machen, solange die Verlage von ihm eine exklusive Rechtseinräumung verlangen und ihm diese aber gerade wegen des obligatorischen einfachen Nutzungsrechts für die Universität verwehrt ist. Der Urheber kann nur solche Nutzungsrechte einräumen, die er auch noch hat. Es bleibt bei einem Eingriff in die Wissenschaftsfreiheit und einer internationalen Benachteiligung deutscher Wissenschaftler.

6 Siehe dazu Pflüger/ Ertmann, ZUM 2004, 436ff.

Neben der Einschränkung der Entscheidungsfreiheit über das Ob und das Wo einer Publikation stellt gleichermaßen das Wie ein Problem dar. So würde dem Autor zusätzlich die Freiheit entzogen, sich für eine Open-Access-Publikation, beispielsweise unter Verwendung einer Creative-Commons-Lizenz⁷, zu entscheiden.

3.2 Weitere Problemkreise bezüglich Anbietungspflicht

Neben den verfassungsrechtlichen Bedenken bestehen zahlreiche weitere einfachgesetzliche Probleme, die hier nur kurz angerissen werden sollen. Grundsätzliche Skepsis begegnet dabei insbesondere der Ansatz einer Verstaatlichung des Publikationswesens. Es stellt sich die Frage, ob die Aufgabe des wissenschaftlichen Publizierens bei den kommerziellen Anbietern nicht unter Umständen in professionelleren Händen liegt. Es ist anzunehmen, dass sich viele Open-Access-Enthusiasten an den Universitäten alsbald werden fragen müssen, ob sie den personellen, strukturellen und finanziellen Herausforderungen, die auf sie zukommen, gewachsen sind. Es besteht das Risiko, dass Open Access hier eher zu Kostensteigerungen denn zu -senkungen führt.

Ähnlich kritisch ist das Risiko von Publikationsverzögerungen zu bewerten. So ist bislang angedacht, eine zweimonatige Optionsfrist vorzusehen. Das hieße, dass die Universität zwei Monate Zeit bekäme, um sich zu überlegen, ob sie einen Beitrag zur Publikation annimmt oder nicht. Zu argumentieren, eine solche Option führe keineswegs zu Publikationsverzögerungen, schließlich könne der Autor seinen Beitrag problemfrei parallel zur Hochschule ganz konventionell einem Wissenschaftsverlag anbieten, erscheint realitätsfremd und kann nicht überzeugen. Man muss sich doch fragen, welcher Wissenschaftsverlag den Publikationsprozess mit Peer-Review etc. startet, solange eine mögliche Hochschulpublikation noch in der Schwebe ist. Auf den komplexen Konflikt zwischen Veröffentlichung und Patentanmeldung soll an dieser Stelle nicht weiter eingegangen werden⁸.

3.3 Zwischenergebnis Anbietungspflicht

Im Ergebnis bleibt festzuhalten, dass der Vorschlag einer Anbietungspflicht für Urheber an Hochschulen abzulehnen ist. Trotz aller Kritik ist aber hervorzuheben, dass dieser Vorstoß der KMK in seinem Grundanliegen grundsätzlich überaus begrüßenswert ist. Es scheint an der Zeit, darüber nachzudenken, welche neuen Wege im wissenschaftlichen Publikationswesen beschritten werden können: nicht gegen die Wissenschaftsverlage, sondern für die Wissenschaftler.

7 <http://creativecommons.org/>

8 Hier sei auf den eingangs bereits erwähnten Beitrag des Verfassers, GRUR Int. 2005, 378, 381, verwiesen.

4 Verfassungskonforme, alternative Regelungsansätze

4.1 Blick ins Ausland

Angenommen, man fragte den von der internationalen „Journal Crisis“ unbehelligten Mann auf der Straße: „Was glauben Sie, welches Land sorgt besser dafür, seinen Bürgern einen möglichst freien Zugang zu öffentlich finanzierten Ergebnissen aus Wissenschaft und Forschung zu ermöglichen? Die USA oder Deutschland? Welches von beiden Ländern ist stolz darauf, im sozialen Gemeinwohlinteresse dem Steuerzahler zu den mit Steuermitteln finanzierten Forschungsergebnissen offenen Zugang zu gewähren? Welches Land ist dagegen vor allem profitorientiert, privatnützig und begünstigt primär kommerzielle Interessen?“ – In Zeiten von Schröder und Bush ist leicht zu erahnen, wie die falsche Antwort des Mannes auf der Straße lauten wird. Diese vereinfachende Gegenüberstellung soll nicht die zugestandenermaßen natürlich auch bestehenden Vorzüge unserer Urheberrechtsschranken in Frage stellen noch die unbestrittenen Leistungen des deutschen Bibliothekssystems und anderer involvierter Gruppen diskreditieren. Es geht schlicht darum, wie in dieser Frage des Umgangs mit Steuermitteln für Wissenschaft und Forschung von den USA dazugelernt werden kann. Neben zahlreichen anderen internationalen Open-Access-Projekten – wie dem DAREnet-Projekt⁹ in den Niederlanden oder den Bemühungen des Wellcome Trust in Großbritannien¹⁰ – kommen v.a. der sog. NIH-Plan¹¹ und die Sabo-Bill¹² als Inspirationsquellen in Betracht. Beim NIH-Plan geht es im Kern darum, dass ausgehend von Empfehlungen des Haushaltsausschusses des US-Repräsentantenhauses solche wissenschaftlichen Beiträge, die im Rahmen einer mit staatlichen NIH-Geldern finanzierten Forschungstätigkeit entstanden sind, der Öffentlichkeit verstärkt zugänglich gemacht werden sollen, also Versionen der Beiträge auf einem kostenfrei zugänglichen Server – teilweise zeitlich verzögert – abgelegt werden sollen. Bei der sog. Sabo-Bill war überlegt worden, an solchen öffentlich finanzierten Forschungsergebnissen erst gar kein Urheberrecht entstehen zu lassen. Letzteres ist aber schon mit Art. 2 Abs. 1 der Revidierten Berner Übereinkunft (RBÜ) und dem im kontinentaleuropäischen Urheberrecht geltenden Schöpferprinzip unvereinbar.

9 Vgl. <http://www.darenet.nl/page/language.view/home>.

10 Siehe: http://www.wellcome.ac.uk/doc_WTX025197.html.

11 National Institutes of Health (NIH): <http://www.nih.gov/about/publicaccess/>.

12 Mehr Informationen zur sog. Sabo-Bill, offizieller Titel: „Public Access to Science Act“, (H.R. 2613), unter <http://www.biomedcentral.com/openaccess/archive/?page=features&issue=3>. Urheberrechtsschutz sollte danach ausgeschlossen werden “for any work produced pursuant to scientific research substantially funded by the Federal Government”.

Gleichwohl lassen sich diese beiden Projekte für die deutsche Diskussion fruchtbar machen. Konkret lassen sich ihnen zwei Prämissen entziehen: die Anknüpfung an die öffentliche Finanzierung und der pragmatische Kompromiss, nicht unmittelbar nach Open Access-Archiving zu rufen, sondern erst verzögert nach einer angemessenen Phase exklusiver Auswertung durch die Wissenschaftsverlage.

Versucht man diese Prämissen im deutschen Recht umzusetzen, ergibt sich theoretisch eine Vielzahl von Möglichkeiten. In Betracht kommen prinzipiell arbeitsvertragliche oder beamtenrechtliche Regelungen, öffentlich-rechtliche Anreizsysteme für Open-Access-Archiving beispielsweise in Gestalt einer konditionierten Fördermittelvergabe etc. etc. Letztlich wird nur eine Kombination all dieser Instrumente zum Erfolg führen können. Vorliegend soll ausschließlich der urheberrechtliche Lösungsansatz verfolgt werden.

4.2 Denkbare urheberrechtliche Regelungsansätze

Um eine gesetzlich garantierte Option für ein zeitlich verzögertes Open-Access-Archiving für überwiegend öffentlich-finanzierte Forschungsergebnisse im Urheberrecht zu verankern, kommen diverse Instrumente in Betracht. Denken lässt sich an eine Zwangslizenz, an eine neu einzufügende Schrankenregelung oder eine urhebervertragsrechtliche Regelung. Eine Lösung „One size fits all“ kann es dabei kaum geben, denn es lassen sich schwerlich alle Konstellationen des wissenschaftlichen Publizierens mit einer Regelung erfassen. Dafür sind allein die Amortisierungszeiträume in den einzelnen Disziplinen unterschiedlich. Im schnelllebigen STM-Bereich sind diese grundsätzlich kürzer, im Bereich der Geisteswissenschaften indessen länger. Es gilt daher u.a. zu differenzieren, an welchen Forschungsergebnissen ein übergeordnetes öffentliches Interesse besteht. Die Antwort lautet: an solchen, die im Rahmen einer überwiegend mit öffentlichen Mitteln finanzierten Lehr- und Forschungstätigkeit entstanden sind. Weiter scheint eine Begrenzung auf Periodika angeraten, denn Buchveröffentlichungen benötigen einen deutlich längeren Amortisierungszeitraum als aktuelle Zeitschriftenbeiträge.

Eine urheberrechtliche Zwangslizenz, die einen gesetzlichen Kontrahierungszwang zu angemessenen Bedingungen auf Einräumung eines einfachen Nutzungsrechts vorsähe, erscheint dabei wenig praktikabel. Nähere Aufmerksamkeit verdienen aber eine neue Schrankenregelung und eine urhebervertragsrechtliche Lösung.

4.3 § 52c UrhG als neue Schrankenregelung

Um den Zugang zu den Resultaten aus öffentlich finanzierter Forschung und Lehre zu verbessern, könnte man über folgende, neu einzufügende Schrankenregelung nachdenken:

§ 52c UrhG Öffentliche Zugänglichmachung öffentlich finanzierter Lehre und Forschung

- (1) Zulässig ist es, Werke i.S.v. § 2 Abs. 1 Nr. 1, die im Rahmen einer überwiegend mit öffentlichen Mitteln finanzierten Lehr- und Forschungstätigkeit entstanden sind und in Periodika erscheinen, sechs Monate nach ihrer Erstveröffentlichung zur Informationsteilhabe der Allgemeinheit öffentlich zugänglich zu machen, soweit dies zur Verfolgung nicht kommerzieller Zwecke gerechtfertigt ist. Für die Zugänglichmachung ist eine angemessene Vergütung zu zahlen. Der Anspruch kann nur durch eine Verwertungsgesellschaft geltend gemacht werden.
- (2) Der Urheber kann die öffentliche Zugänglichmachung verbieten, wenn das Werk seiner Überzeugung nicht mehr entspricht, ihm deshalb die Verwertung des Werkes nicht mehr zugemutet werden kann und er ein etwa bestehendes Nutzungsrecht aus diesem Grunde zurückgerufen hat (§ 42)¹³.

Bei der Ausgestaltung einer solchen Regelung ist neben der verfassungsrechtlichen Eigentumsgarantie (Art. 14 Abs. 1 GG)¹⁴ auch internationales Urheberrecht zu beachten. Namentlich ist der Richtlinie zum Urheberrecht in der Informationsgesellschaft aus dem Jahre 2001¹⁵ und dem sog. Dreistufentest¹⁶ Rechnung zu tragen. Während den Erfordernissen des Dreistufentests mit einem gewissen Argumentationsaufwand Genüge getan werden kann¹⁷, scheidet die wie oben angeregte Schrankenregelung an dem bislang geltenden europäischen Recht. Danach wäre die Schaffung einer solchen Schranke nur dann zulässig, wenn vorgesehen ist, dass der Privilegierte (ausschließlich) wissenschaftliche Forschungszwecke verfolgt (vgl. Art. 5 III RL 2001/29/EG bzw. deren Erwägungsgrund 32). Mit anderen Worten wäre eine Koppelung der Privilegierung an wissenschaftliche Forschungszwecke vonnöten. Man mag diesen vorgegebenen europarechtlichen Rahmen als zu eng kritisieren, zum gegenwärtigen Zeitpunkt besteht indes kein Spielraum für einen neuen § 52c UrhG. Dieses Resultat unterstreicht, wie sehr der abschließende, kasuistische Schrankenkatalog der Richtlinie die Mitgliedstaaten bei der Ausgestaltung sinnvoller Schrankenrege-

¹³ Vgl. § 46 Abs. 5 Satz 1 UrhG, dem die Regelung hier nachgebildet ist.

¹⁴ BVerfGE 31, 229, 240f.

¹⁵ Vgl. Art. 5 III a) RL 2001/29/EG vom 22.05.2001, ABl. L 167 vom 22.06.2001.

¹⁶ Vgl. Art. 5 Abs. 5 der RL 2001/29/EG, Art. 9 II RBÜ, Art. 13 TRIPs oder Art. 10 WCT und Art. 16 II WPPT.

¹⁷ Siehe der Verf., GRUR Int. 2005, 378, 385.

lungen für die Wissensgesellschaft behindert. Vor diesem Hintergrund ist zu überdenken, ob man den enumerativen Schrankenkatalog nicht doch mit dem Dreistufentest als flexiblem Korrektivinstrument öffnen sollte.

4.4 Urhebervertragsrechtliches Lösungsmodell – Formulierungsvorschlag für einen neuen § 38 Abs. 1 S. 3 UrhG

Nach gegenwärtiger Rechtslage werden in den allermeisten Fällen unreflektiert sämtliche Rechte - also auch die Online-Rechte (das sog. Recht der öffentlichen Zugänglichmachung, § 19a UrhG) den Wissenschaftsverlagen eingeräumt. Dies gibt den Verlagen eine übermächtige Verhandlungsposition bei der Preispolitik. Zum Schutze des Urhebers vor dem Verwerter bzw. vor sich selber erscheint es daher erforderlich, eine vertraglich nicht abdingbare zeitliche Befristung des Ausschließlichkeitscharakters bei Einräumung ausschließlicher Nutzungsrechte vorzunehmen. Dies bedeutete, dass die exklusiven Nutzungsrechte, die ein Autor einem Verlag eingeräumt hat, sich beispielsweise nach einem halben Jahr automatisch in ein einfaches Nutzungsrecht umwandeln. Dies hätte vereinfacht gesprochen zur Konsequenz, dass dem Urheber regelmäßig die gesetzlich garantierte Option zur Zweitveröffentlichung im Internet bliebe. In diesem Sinne könnte ein neu einzufügender Satz 3 in § 38 Abs. 1 UrhG lauten:

„An wissenschaftlichen Beiträgen, die im Rahmen einer überwiegend mit öffentlichen Mitteln finanzierten Lehr- und Forschungstätigkeit entstanden sind und in Periodika erscheinen, hat der Urheber auch bei Einräumung eines ausschließlichen Nutzungsrechts das Recht, den Beitrag nach Ablauf von sechs Monaten seit Erstveröffentlichung anderweitig öffentlich zugänglich zu machen, soweit dies zur Verfolgung nicht kommerzieller Zwecke gerechtfertigt ist.“

Zu klären verbleibt bei einer solchen Regelung allein die Berücksichtigung der bereichsspezifisch divergierenden Auswertungszeiträume (s.o.). Um hier angemessen zu differenzieren, kommen verschiedene Möglichkeiten in Betracht, die hier nur stichwortartig vorgestellt werden sollen¹⁸. Ein Ansatz wäre eine gesetzliche Ausnahmeregelung, bei der den Verleger die Beweislast träfe, nachzuweisen, dass für eine angemessene Amortisierung ein längerer Zeitraum als die grundsätzlich geltenden sechs Monate erforderlich ist. Weiter kommt eine staatlich kontrollierte Fristverlängerung mit dem Deutschen Patent- und Markenamt (DPMA) als Aufsichtsbehörde in Betracht. Schließlich sind kollektivvertragliche Rahmenvereinbarungen zu erwägen. In diesem Zusammenhang ist auch an eine Revision der zwischen dem Deutschen Hochschulverband und dem Börsenverein des Deutschen Buchhandels ausgehandelten Vertragsmuster zu denken.

18 Vertieft dazu der Verf., GRUR Int. 2005, 378, 387.

5 Zusammenfassung

Die vorgeschlagene, zwingende Regelung § 38 Abs. 1 Satz 3 UrhG würde über eine zeitliche Befristung der ausschließlichen Nutzungsrechtseinräumung eine gesetzlich garantierte Option für eine Open Access-Zweitveröffentlichung im Internet schaffen. Im Gegensatz zum Vorschlag einer Anbieterspflicht bliebe somit der Wissenschaftler in seiner Entscheidung frei. Aber nicht nur das: durch eine vertraglich nicht abdingbare Regelung in § 38 I S. 3 UrhG würde Klarheit geschaffen. Bislang sind viele juristische Laien damit überfordert zu klären, ob sie mit einer Open Access-Publikation irgendwelche Rechte verletzen. Eine transparente gesetzliche Regelung würde hier Abhilfe schaffen. Vor allem dieses Transparenzbedürfnis verdeutlicht die Notwendigkeit einer gesetzlichen Regelung.

Letztlich handelt es sich bei dem Vorschlag um einen Kompromiss zwischen Open Access-Enthusiasten und Verlegerinteressen. Durch bereichsspezifisch unterschiedlich lange exklusive Auswertungszeiträume für die Verlage wird dabei versucht, den berechtigten Amortisierungsinteressen der Verlage Rechnung zu tragen. Wer auf der anderen Seite befürchtet, eine solche zeitlich verzögerte, gesetzliche Open Access-Archiving-Option sei eine Verschlechterung zum Status quo, viele Verlage erlaubten schon jetzt unmittelbares Self-Archiving parallel zur Erstveröffentlichung, übersieht, dass diese Möglichkeit durch die gesetzliche Regelung unbenommen bleibt. Wer seine Beiträge ohne zeitliche Verzögerung ins Netz stellen möchte und in der Lage ist, seine Rechte dementsprechend wahrzunehmen, der kann dies auch weiterhin tun. Die breite Masse aber, die es als lästig begreift oder sich überfordert fühlt, sich mit diesen urheberrechtlichen Fragen gegenüber den Verlagen auseinanderzusetzen, bekäme eine klare Regelung an die Hand, was sie darf und was nicht.

Der Vorschlag kann dabei letztlich nur ein Mosaikstein sein. Nur eine isolierte gesetzliche Option für Open Access zu schaffen, reichte nicht aus. Deshalb ist es von eminenter Bedeutung, zusätzliche Anreizsysteme für Open Access zu schaffen und für Open Access zu werben. Als Schlagwort sei hier nur „konditionierte Fördermittelvergabe“ genannt. Wie gesehen, bietet die Urheberrechtsreform erhebliche Risiken, aber auch Chancen für das wissenschaftliche Publizieren; Chancen, die wir nutzen sollten.

Angaben zum Autor

Gerd Hansen
Max-Planck-Institut für Geistiges Eigentum,
Marshallplatz 1
80539 München

Gerd Hansen hat in Freiburg, Sevilla und Berlin (HU) Germanistik und Jura studiert. Seit 2004 promoviert er als Stipendiat am Max-Planck-Institut für Geistiges Eigentum in München bei Herrn Prof. Dr. Reto M. Hilty. Der Arbeitstitel der Dissertation lautet: „Der dreifache Schutzzweck im Urheberrecht – Zur Neujustierung von Urheber-, Nutzer- und Verwerterinteressen“. Nähere Informationen zum Verfasser finden sich unter www.gerd-hansen.net

Potenziale von Creative Commons-Lizenzen für Open Innovation

Rainer Kuhlen, Jochen Brüning

FB Informatik und Informationswissenschaft – Universität Konstanz

Zusammenfassung

Es wird der Zusammenhang von *Open Access* (OA) und *Creative Commons* (CC) im Interesse von *Open Innovation* herausgearbeitet. Auf die noch bestehende Zurückhaltung von Wissenschaftlern, sich auf OA und CC einzulassen, wird eingegangen. Demgegenüber werden die Vorteile und Erfolgsfaktoren von CC dargestellt, wobei besonders auf die digitale Signatur als notwendige Ergänzung zur Sicherung der Werksauthentizität eingegangen wird. Nach der Darstellung der CC-Praxis und den Perspektiven der Weiterentwicklung wird CC noch im Kontext der verschiedenen Ausprägungen von *Open Access gold road* und *Open Access green road* diskutiert, also zwischen Formen der (anarchischen) Direkt-/Eigenpublikation im Web, Publikationsservern, freien OA-Journalen und so genannten kommerziellen OA-Journalen. CC-Lizenzierung sollte sich in jedem Fall als fördernd für *Open Innovation* erweisen.

1 Umfeld

Kontrovers in Politik, Wirtschaft und Wissenschaft wird diskutiert, ob die Privatisierung und damit zusammenhängend die Kommerzialisierung von Wissen und Information in unserer Gesellschaftsordnung nicht nur nicht vermeidbar, sondern sogar Bedingung für weitere Innovationen sind, oder ob, im Gegenteil, erst die freizügige, öffentliche Nutzbarkeit von Wissen und Information ein gleichermaßen inventions- und innovationsfreundliches Klima, sozusagen ein Klima des *Open Innovation* erzeugt. Erkennbar ist, dass sich gegen die im Interesse der kommerziellen Nutzung kaum zu vermeidbaren Verknappungsstrategien vielfältiger Widerstand artikuliert und organisiert. Damit steigt auch die Chance, dass sich die skizzierte Kontroverse besser durch empirische Belege der Praxis von *Open Innovation* auflöst, sprich, dass Modelle von *Open Innovation* sich nicht nur als wissenschafts-, sondern auch als wirtschaftsfördernd erweisen.

In Deutschland ist Indiz für diesen Widerstand gegen die sich absolut setzende Verwertungs- und Verknappungspraxis die von den wichtigen Wissenschaftsorganisationen, sehr vielen Fachgesellschaften und noch mehr Einzelpersonen unterschriebene Göttinger Erklärung vom 5. Juli 2004 ([http:// www.urheberrechtsbuendnis.de/](http://www.urheberrechtsbuendnis.de/)), die sich im Urheberbündnis institutionalisiert hat und auf die politische Gestaltung des Urheberrechts im Interesse von Bildung und Wissenschaft Einfluss nimmt. *Open Access*-Journale (im Folgenden OA) und -Server sind längst keine Exoten mehr, sondern sind aus dem Alltag der Wissensproduktion, -verteilung und -nutzung nicht mehr wegzudenken, genauso wie die Praxis von freier und offener Software nicht mehr aus der Softwareproduktion und -nutzung wegzudenken ist. In diesem Zusammenhang – freie Nutzung der wissenschaftlichen Produktion zu sichern – ist auch die Lizenzierung der erstellten Werke durch *Creative Commons* (im Folgendem CC) zu sehen.

2 Noch Skepsis von Seiten der Wissenschaft

Weiterhin ist festzustellen, dass Wissen um alternative Publikationswege und die damit verbundenen Nutzungsmöglichkeiten des so öffentlich gemachten Wissens in den Hochschulen und Forschungseinrichtungen noch nicht in dem Maße verbreitet ist, wie es angesichts der zu Recht beklagten „Zeitschriftenkrise“ (die ja im Grunde eine Verlagskrise ist) wünschenswert wäre. Selbst wenn dieses Wissen vorhanden ist und darüber hinaus die technischen Voraussetzungen, z. B. in Form eines professionell betriebenen Online-Publikationsservers der Bibliothek, gegeben sind, ist die Bereitschaft der Wissenschaftler, auf dem OA-Weg zu publizieren, gering. Dafür sind verschiedene Gründe auszumachen:

- Mangelndes Vertrauen in die Langzeitverfügbarkeit und -lesbarkeit der elektronisch archivierten Dokumente.
- Die derzeit noch mangelhafte Zitierfestigkeit der OA-Quellen, für die bei digitalen Formaten neben anderen Gründen auch die fehlende Seitenangabe in bestimmten Online-Formaten angeführt wird. Dieses Manko, wie auch andere, ließe sich durch den konsequenten Einsatz verfügbarer Web-Standards (Textmarken, Xlink, URNs, DOIs ...) beseitigen.
- Schwieriger wiegt der mangelnde Reputationsgewinn, der bislang für Autoren mit OA-Publikationen verbunden zu sein scheint. Dahinter steht die vermeintlich fehlende Qualitätssicherung durch das *Peer Review*. Ein von den kommerziellen Verlagen gerne aufgegriffenes Argument, begründen sie doch gerade mit den hohen Kosten für das *Peer Review* die teilweise exorbitant hohen Preise ihrer Produkte. Dabei hat in Wirklichkeit OA nichts mit

fehlendem *Peer Review* zu tun. So gut wie alle OA-Journale bestehen auf *Review*-Verfahren, organisieren dies allerdings gänzlich selber.

- Unsicherheit hinsichtlich Authentizität, Integrität und Veröffentlichungsdatum. Die damit zusammenhängenden Fragen und die in unseren Augen angemessene Antwort darauf finden sich mit einem Vorschlag zur digitalen Signatur am Ende von Abschnitt 3.

Wir gehen im Folgenden in erster Linie auf die Herausforderung der Rechtesicherung durch CC ein, denn ein weiterer Grund für die Zurückhaltung der Autoren, im OA-Verfahren zu publizieren, liegt im mangelnden Verständnis des Urheberrechts und der daraus resultierenden Rechtsunsicherheit sowohl bei Autoren als auch bei Nutzern des öffentlich gemachten Wissens, verunsichert durch rigorose Autorenverträge, die einseitig alle Rechte für die Verlage, auch für eine nicht vorhersehbare Medienzukunft, reklamieren sowie die bislang politisch gewollte Stärkung der Verlage, die in der novellierten Fassung des Urheberrechts zum Ausdruck kommt. Wie man an dem MPI-/FIZ-Karlsruhe-Projekt eSciDoc sieht, gibt es allerdings durchaus auch politisch geförderte Gegentendenzen.

Die problematische Autorensituation und die Verunsicherung der Nutzer erkennend und dem entgegenwirkend, weil innovationshemmend und dem Wissenschaftsprinzip der freien Verfügbarkeit von Daten, Wissen und Information zuwiderlaufend, hat Lawrence Lessig, Rechtsprofessor an der Stanford University, 2001 das CC-Projekt ins Leben gerufen. Im Januar 2002 wurde eine erste Version der frei verfügbaren CC-Lizenz veröffentlicht. Diese Lizenz ermöglicht Kulturschaffenden und Wissenschaftlern, ihre Werke im Rahmen des geltenden Copyrights bzw. Urheberrechts in selbst bestimmter Weise der Öffentlichkeit zu von ihnen ausgewählten Bedingungen zur Verfügung zu stellen. Das CC-Projekt bezog grundlegende Ideen aus der von Richard Stallman initiierten *Open Source*-Initiative und der zugehörigen Software-fokussierten GPL-Lizenz. Derzeit existiert eine Vielzahl von OA-Lizenzen, die oft an Spezialgebiete und deren Besonderheiten und Anforderungen angepasst sind. Beschränkt man sich nur auf die inhaltsbezogenen (*content*: Text, Bild, Musik, ...) Lizenzen, stehen mehr als 30 zur Auswahl. Für die große Akzeptanz der CC-Lizenzen – es sind bereits mehr als 17 Millionen Werke damit ausgezeichnet – sind mehrere Erfolgsfaktoren auszumachen.

3 Creative Commons-Erfolgsfaktoren

- **Anpassung an länderspezifische Rechtssysteme.** Der ursprüngliche (*generic*) Lizenztext wurde mittlerweile an 19 Jurisdiktionen angepasst, u.a. auch an das deutsche Urheberrecht. Hieraus resultiert ein hohes Maß an Rechtssicherheit sowohl für die Autoren als auch für die (Nach-) Nutzer der lizenzier-

ten Werke. In mehr als 70 weiteren Ländern wird an der Anpassung gearbeitet. Diese Arbeit wird von einer eigenen CC-Organisationseinheit, *iCommons*, mit Sitz in Berlin koordiniert und vorangetrieben.

- **Verständlichkeit.** Die Lizenzen treten in drei gekoppelten Formaten auf. Zum einen gibt es eine *Commons Deed* genannte Ausführung, die dem juristischen Laien in verständlicher Formulierung die in der Lizenz enthaltenen bzw. geforderten Rechte und Pflichten erläutert.
- **Verbindlichkeit.** Zum zweiten gibt es die juristisch einwandfreie und damit auch bindende Formulierung der Lizenz, die, wie oben erläutert, an die nationalen Jurisdiktionen gebunden ist.
- **Maschinelle Interpretierbarkeit.** Drittens gibt es eine maschinenlesbare Fassung der Lizenz, die ein automatisches Suchen nach CC-lizenziertem Material ermöglicht. Im Hinblick auf verfeinerte Such- und Verarbeitungsprozesse (Stichwort: *Semantic Web*) besteht die Möglichkeit, diese Form mit standardkonformen (XML/RDF/Dublin Core) Metadaten anzureichern. Zur steigenden Akzeptanz der CC-Lizenzen trägt sicher die Tatsache bei, dass populäre Suchmaschinen mittlerweile die Option bieten, „nur nach CC-lizenziertem Material“ zu suchen.
- **Einfache Handhabung und Flexibilität.** Über ein intuitiv zu bedienendes Web-Interface werden die drei oben genannten Lizenz-Bausteine mit wenigen Mausklicks generiert. Die Auszeichnung der zu lizenzierenden Werke erfolgt dann nicht durch Einbindung des gesamten Lizenztextes in das Werk – dieser liegt wohlverwahrt auf einem Server in der Obhut von CC –, was bei einigen Formaten, z.B. MP3-kodierter Musik oder Bildmaterial, ohnehin schwierig bzw. unmöglich sein dürfte, sondern es wird stattdessen nur ein Link auf die Lizenz eingebunden. Strukturierte Datenformate wie HTML, XML, JPEG, MP3 und pdf – um nur einige zu nennen – sind darauf vorbereitet. Neben diesem Link werden so auch die maschinenlesbaren Metadaten in das Werk eingebunden.
- **Anpassung an ein weites Anwendungsspektrum.** Sind die *Generic*-Lizenz und ihre länderspezifischen Adaptionen inhaltsorientiert, aber ansonsten undifferenziert, so sind nun weitere Lizenzen spezieller Ausprägung für die unterschiedlichsten Anwendungsszenarien entwickelt worden. Es stehen für die „sample-, remix- and share“-Kultur der Musikszene darauf abgestimmte CC-Lizenzen zur Verfügung; Lizenzen für die Dritte Welt-Länder sind ebenso verfügbar wie eine für von Autorenkollektiven geschaffene Werke, z.B. Wikis.
- **Aktualität und Praxisnähe.** Aufbauend auf den Erfahrungen bei der Anwendung der Lizenzen und den sich in der Praxis ergebenden Fragen erfolgt die Weiterentwicklung der Lizenztexte. Die derzeitige aktuelle Version ist 2.5,

was schon erahnen lässt, dass demnächst mit einer Version 3.0 zu rechnen ist. Fragen zur Anwendung und sich daraus ergebende Änderungswünsche werden auf Mailing-Listen und im CC-Wiki (<http://iteamspace.creativecommons.org/wiki/>) diskutiert.

Unserer Einschätzung nach fehlt jedoch noch ein wichtiges Element in der CC-Praxis, das für die Sicherung der Werkauthentizität und damit für die Akzeptanz von Wissenschaftlern zentral ist: die digitale Signatur ist eine notwendige Ergänzung von CC.

Besonders im Umgang mit Online-verfügbaren OA-Materialien treten folgende Fragen auf: Wer ist der Autor des Werks? Ist das Dokument vollständig und unverändert? Ist es das Dokument, das referenziert wurde? Ist es die Version, die referenziert wurde? Wann wurde das Dokument veröffentlicht¹?

Wissenschaftler müssen sicher sein können, dass auch im elektronischen Umfeld Authentizität, Integrität und die Eindeutigkeit des Veröffentlichungsdatums ihrer eigenen sowie die der referierten Dokumente gewährleistet sind. Die von anderer Seite vorgeschlagene Verwendung von DRM-Systemen als Lösung ist unhandlich bis unpraktisch, läuft den erklärten OA-Zielen zuwider und ist mit den CC-Lizenzen unvereinbar. Die Lösung sollte in der digitalen Signatur der Dokumente gemäß XAdES- (ETSI) bzw. XMLDSIG- (W3C/IETF) Standard liegen und damit auf X.509v3-Zertifikaten basieren. Die Signatur stellt auf höchstem Niveau – nämlich Signaturgesetz konform – sicher, dass die Dokumente von dem in der Signatur genannten Autor (dem Zertifikatsinhaber) unterzeichnet und seitdem nicht verändert worden sind. Selbst die kleinste Veränderung am Dokument führt zu einer als ungültig gekennzeichneten Signatur. Die Signatur enthält darüber hinaus einen Zeitstempel, der den Veröffentlichungszeitpunkt darstellt. Da digitale Formate problemlos kopiert und verändert (in diesem Zusammenhang besser: angepasst und fortgeschrieben) werden können, ist über diesen in der Signatur enthaltenen Zeitstempel eine Versionskontrolle gegeben. Bei extrem hohen Anforderungen – beispielsweise einem Patentantrag – können Dokumente mit dem qualifizierten Zeitstempel eines dafür zertifizierten Dienstleisters versehen werden, der dann die „amtliche“ Zeit enthalten würde.

4 Open Access und Creative Commons

Vervielfältigung, Verbreitung und öffentliche Aufführungen der Werke sind grundsätzlich erlaubt – das ist die grundsätzlich gemeinsame Basis von OA und CC. Die gemäß Urheberrecht dem Autor zustehenden exklusiven Rechte des

1 Brünig, J. (2004) CC Acceptance in the Science Community.
http://www.inf-wiss.uni-konstanz.de/People/JB/berlin_proposal.pdf

Öffentlichmachens gibt dieser an die Öffentlichkeit zur freien Nutzung. Die Nutzer der Werke sind im Gegenzug dazu verpflichtet, bei Vervielfältigung, Verbreitung und öffentlichen Aufführungen den Namen des Autors anzugeben und die von ihm gewählten Lizenzbestimmungen unverändert beim Werk zu belassen. Urheber setzen so mehr auf reputative als auf monetäre Anerkennung. Wahlfreiheit hat der Autor bezüglich der Freigabe der kommerziellen Nutzung des Werkes: Diese kann ausgeschlossen werden, ebenso Veränderungen am Werk. Möchte der Autor Veränderungen zulassen, dann kann er verlangen, dass die Derivate nur unter den dem Original zugewiesenen Lizenzbestimmungen verbreitet werden dürfen. Ist also beim Original keine kommerzielle Nutzung erlaubt, so ist diese auch bei den Derivaten nicht zulässig. Mit dieser Regelung wird verhindert, dass Entwicklungen aus der *public domain* von gewinnorientiert agierenden Unternehmen „gekapert“ werden können. Eine wichtige Neuerung in der Version 2.5 betrifft die obligatorische Namensnennung, bei der jetzt auch auf ein Projekt und/oder eine Förderorganisation verwiesen werden kann und die Möglichkeit, bei Derivaten ausdrücklich die Nennung des Autors der Originalarbeit zu untersagen.

Wie weiter oben erwähnt, lassen sich über das intuitiv zu bedienende Web-Interface aus diesen vier Lizenz-Grundelementen – nämlich Namensnennung, kommerzielle Nutzung, keine Derivate und Weitergabe zu den gleichen Bedingungen – individuelle, den Wünschen der Autoren entsprechende Lizenzen generieren, die völlig im Rahmen des geltenden Urheberrechts bleiben. Dem strikten „all rights reserved“ wird so ein flexibles „some rights reserved“ entgegengesetzt. Die Autoren gewinnen ein Stück ihrer informationellen Selbstbestimmung² zurück.

Die generischen CC-Lizenzen sowie die Lizenzen für spezielle Anwendungsgebiete orientieren sich am Copyright/Urheberrecht, sind also geeignet, um inhaltliche Werke (*content*) damit auszuzeichnen. Mit dem neuen *Science Commons*-Projekt wendet sich CC einem verwandten, unzulänglich vom Urheberrecht abgedeckten Gebiet zu. Hier sollen der Ankündigung nach Lizenzen, Verfahren und Vereinbarungen für Wissenschaft und Forschung entwickelt und etabliert werden, die eher vor dem Hintergrund des Patentrechts zu sehen sind. Dieses Projekt wird von einer weiteren, eigenständig operierenden und in Boston MA angesiedelten CC-Organisationseinheit, *Science Commons*, durchgeführt. Im Rahmen seines *Open Access Law*-Programms wurde von *Science Commons* ein Modell und ein Vereinbarungstext entwickelt und der Öffentlichkeit zur Verfügung gestellt, nach dem juristische Fachjournale die in ihnen er-

2 Kuhlen, R. & Brüning, J. (2004) Creative Commons (CC) – für informationelle Selbstbestimmung, gegen den Trend des Urheberrechts/Copyright als Handelsrecht. Chancen für einen innovativen Drei-Stufen-Test? In: Information - Wissenschaft & Praxis (IWP / nfd) 8/2004. S. 449 - 454 - http://www.inf-wiss.uni-konstanz.de/People/RK/Publikationen2004/CC_fuer_IWP-rk+jb2291004_final.pdf

schienenen Artikel nach OA-Prinzipien archivieren und der Allgemeinheit zugänglich machen können. Die Allgemeinheit – speziell Forschung und Lehre – gewinnt durch den freien Zugang, die Autoren gewinnen durch die größere Verbreitung ihrer Arbeiten, und den Journalen steht das Recht zu, immer als Quelle der Erstveröffentlichung genannt zu werden.

Es bleibt abzuwarten, wie sich in der Zukunft OA und CC im Interesse von *Open Innovation* weiter entwickeln und ihr wechselseitiges Verhältnis bestimmen werden. Greift man die von Stevan Harnad et al. vorgeschlagene Unterscheidung von *gold road* und *green road to Open Access* auf³, so ist CC für Autoren, die *green road* bevorzugen, also ihre Werke original auf die eigene bzw. organisations-/institutionseigene Webseite stellen, unverzichtbar – gibt die CC-Lizenz ihnen doch einen guten Teil ihrer informationellen Selbstbestimmung zurück bzw. sichert diese. Im gleichen Zuge gibt die Lizenzierung den Nutzern der Werke Rechtssicherheit. Auch zu *green road* sind noch die OA-Ansätze zu rechnen, die ebenfalls auf die klassische Zeitschriftenpublikation verzichten, und die Werke über Publikationsserver, z.B. der Bibliotheken, oder über große Netzverbände wie in MathNet oder PhysNet zugänglich machen (wo also in verteilten Servern effizient gesucht wird). Hier könnte es durchaus zu Konflikten zwischen der bislang autorenbestimmten CC-Lizenzierung und der Publikationsphilosophie der zentralen oder verteilten Publikationsserver kommen. Noch stärker ist dies bei den nach *gold road* arbeitenden freien OA-Journalen zu erwarten, die sich auch als Initiativen aus der Wissenschaft ihre finanzielle Basis sichern müssen. Hier wird man genau hinsehen müssen, was noch autorenbestimmt möglich ist bzw. welche Interessenausgleiche gefunden werden können. Noch stärker wird das zum Problem bei kommerziellen OA-Journalen, die, obgleich sie freie Nutzung nach Zahlung der Autoren (wie bei den freien OA-Journalen) zusichern, eigentlich keine wirklichen OA-Produkte sind, da die Verwertungsrechte ansonsten gänzlich beim Verlag bleiben. Ob solche Publikationen mit CC-Lizenzen noch sinnvoll sind, ist mehr als fraglich.

Es wird nicht das eine OA geben, und es gibt nicht die eine CC-Lizenz. Wie auch immer, CC-Lizenzen, hoffentlich dann zusammen mit der digitalen Signatur, sichern Wissenschaftlern und Kulturschaffenden/Kreativen rechtsverbindlich ihre selbstbestimmten Rechte und gewährleisten Authentizität, Integrität, weltweiten Zugriff und Langzeitverfügbarkeit ihrer Werke. Von dieser Sicherheit profitieren nicht nur die Autoren, sondern gleichermaßen die Nutzer, letztendlich die Allgemeinheit von der so entstehenden innovationsfördernden Wissens- und Kulturallmende.

3 Harnad, S., Brody, T., Vallieres, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., Stamerjohanns, H. & Hilf, E. (2004) The Access/Impact Problem and the Green and Gold Roads to Open Access. <http://dx.doi.org/10.1016/j.serrev.2004.09.013> /Serials Review/ 30 (4) 2004

Angaben zu den Autoren

Jochen Brüning, Prof. Dr. Rainer Kuhlen
Universität Konstanz
FB Informatik und Informationswissenschaft
Universitätsstr. 10
78464 Konstanz

E-Mail: jochen.bruening@uni-konstanz.de, rainer.kuhlen@uni-konstanz.de

Als Ingenieur verfügt *Jochen Brüning* über einen eher technisch geprägten Hintergrund. Seit Februar 2004 engagiert er sich als Mitarbeiter am informationswissenschaftlichen Lehrstuhl von Prof. Dr. Rainer Kuhlen an der Universität Konstanz für den freien Zu- und Umgang zu und mit Wissen und Information. Sein Aufgabenschwerpunkt liegt bei der Einführung und der Anwendung der *Creative Commons*-Lizenzen und der digitalen Signatur von Dokumenten.

Open Access : die strategische Kluft?

Eine empirische Untersuchung des Rezeptions- und Publikationsverhaltens von Wissenschaftlern in Deutschland

Alexis-Michel Mugabushaka

Deutsche Forschungsgemeinschaft

Zusammenfassung

Die wissenschaftliche Kommunikation befindet sich gegenwärtig in einem tiefgreifenden Wandlungsprozess. Vor allem die durch das Internet dargebotenen Möglichkeiten zur Rezeption und Verbreitung wissenschaftlicher Ergebnisse haben die Debatte um die wirtschaftliche Bedeutung wissenschaftlicher Verlage einerseits und deren Rolle als Garant wissenschaftlicher Qualität andererseits mit neuen Akzenten erneut auf die Tagesordnung gesetzt.

Eine DFG-Studie, der eine Befragung von über 1.000 DFG-Antragstellern unterschiedlicher fachlicher Provenienz zugrunde liegt, untersucht Einstellungen zu und Erfahrungen mit dem Open Access-Publizieren (DFG, 2005). In diesem Aufsatz wird in erster Linie das Spannungsverhältnis thematisiert, zwischen einerseits dem Wunsch, wissenschaftliche Ergebnisse einer möglichst breiten Fachöffentlichkeit zugänglich zu machen, andererseits aber auch die als karriereförderlich geltende Reputation bekannter Publikationsorgane zu nutzen.

1 Einleitung

Die Kommunikation unter Wissenschaftlern hat in den letzten Jahren durch die Entwicklung und rasche Verbreitung der Informations- und Kommunikationstechnologie (IuK) tiefgreifende Änderungen erfahren. Das Internet bietet den Forschenden die Möglichkeit, mit Kollegen Kontakt aufzunehmen, die auf gleichem Gebiet arbeiten, um sich auf diesem Wege über die neuesten Entwicklungen auszutauschen oder Kooperationsprojekte einzugehen. E-Mail-Kommunikation, themenspezifische Diskussionsforen und Fachportale gehören heutzutage zum Repertoire jedes Wissenschaftlers. Auch Fachkonferenzen, die ausschließlich über das Internet abgehalten werden, sind keine Seltenheit mehr.

Die IuK bot auch wissenschaftlichen Zeitschriftenverlagen die Möglichkeit, ältere Jahrgänge von Fachzeitschriften zu digitalisieren und neueste Ausgaben als Druck- und elektronische Fassungen anzubieten. Dieses Angebot ging allerdings nicht mit einer Senkung der Gebühren einher, die die Wissenschaftler zu entrichten haben, um Zugang zu den Ergebnissen zu haben. Die erhoffte Entschärfung der „Zeitschriftenkrise“ blieb aus¹. Die Open Access-Bewegung, die seit Mitte der 90er Jahre den freien Zugang zu wissenschaftlicher Literatur fördert, erlebte mit der Zeit einen Aufschwung und konnte Widerhall in zahlreichen Initiativen und Erklärungen namhafter Forschungsorganisationen finden.

In der Bundesrepublik riefen die deutschen Forschungsorganisationen in der „*Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen*“ Forschungsorganisationen und Kulturinstitutionen im Oktober 2003 dazu auf, das Internet verstärkt zu nutzen, um Forschungsergebnisse und Ressourcen nach dem Prinzip des offenen Zugangs zur Verfügung zu stellen².

Bei den Überlegungen sowohl hinsichtlich der adäquaten Implementierung dieser Resolution als auch in Bezug auf eine nachhaltige Open Access-Förderpolitik wurde das Informationsdefizit hinsichtlich des Stellenwerts der Open Access-Publikationen in den verschiedenen Wissenschaftsdisziplinen in Deutschland sowie die Faktoren, die das Publikations- und Rezeptionsverhalten bestimmen, deutlich. Die im Sommer 2004 von der DFG in Auftrag gegebene Studie³ sollte diesen Fragen nachgehen und eine empirisch fundierte Basis für eine Open Access Förderpolitik liefern.

2 Methodisches

In die Studie wurden Wissenschaftler, die im Zeitraum 2002 bis 2004 eine Förderung seitens der DFG erhalten haben, einbezogen. Es wurde eine stratifizierte Stichprobe von 1.600 Personen gezogen, die zu je einem Viertel Wissenschaftler aus den Geistes- und Sozialwissenschaften, den Lebenswissenschaften, den Naturwissenschaften sowie den Ingenieurwissenschaften umfasste. Innerhalb dieser vier Wissenschaftsbereiche wurde wiederum zwischen „etablierten Wissenschaftlern“ und Nachwuchswissenschaftlern (i.e. Wissenschaftler, die im Rahmen eines der auf die Nachwuchsförderung zielenden Programme der DFG gefördert wurden) unterschieden. Entsprechend der Randverteilung in der Gesamtpopulation wurde die Quote auf 80:20 festgesetzt.

-
- 1 Mit der „Zeitschriftenkrise“ wird die paradoxe Situation beschrieben, in der Wissenschaftler die Forschungsergebnisse in Fachzeitschriften veröffentlichen, manchmal aus Kostengründen keinen Zugang zu den Forschungsergebnissen in denselben Zeitschriften haben.
 - 2 vgl. http://www.zim.mpg.de/openaccess-berlin/BerlinDeclaration_dt.pdf
 - 3 Mit der Durchführung der Umfrage wurde die Gesellschaft für Empirische Studien (GES), Kassel, beauftragt.

An der Erhebung, die im Oktober und November 2004 stattfand, konnten die Befragten mittels Druck- oder Online-Fragebogen teilnehmen. Insgesamt beteiligte sich 64 Prozent der Befragten an der Umfrage. Die realisierte Stichprobe ist im Hinblick auf die Verteilung der Wissenschaftsbereiche sowie auf den beruflichen Status der Befragten zum Zeitpunkt der Antragstellung repräsentativ.

Der eingesetzte Fragebogen⁴, der von einer Arbeitsgruppe der DFG-Geschäftsstelle unter Einbeziehung von Experten entwickelt wurde, ging auf zahlreiche Aspekte ein. In diesem Aufsatz beschränken wir uns auf vier Leitfragen:

- Welche Publikationsformen werden von Wissenschaftlern sowohl für die Rezeption als auch die Veröffentlichung der Forschungsergebnisse bevorzugt verwendet, und sind - auch im Zeitalter des elektronischen Publizierens - diesbezüglich Unterschiede innerhalb der einzelnen Wissenschaftsdisziplinen immer noch feststellbar?
- Wie häufig nutzen die Wissenschaftler Open Access Publikationsplattformen, um ihre Forschungsergebnisse zu veröffentlichen oder um sich über die neuesten Entwicklungen in ihrem Fach zu informieren?
- Wie bewerten die Wissenschaftler die in den Open Access-Debatten thematisierten Chancen und Risiken?
- Halten die Wissenschaftler Open Access-Publikationen für förderungswürdig, und falls ja, welche Maßnahmen erachten sie für sinnvoll?

3 Ergebnisse

3.1 Die Publikationskulturen in einzelnen wissenschaftlichen Disziplinen

Wissenschaftliches Wissen hebt sich von anderen Erfahrungsformen dadurch ab, dass die involvierten Akteure für sich beanspruchen, es im Rahmen von informellen Gemeinschaften zu produzieren. Diese „scientific communities“ oder „invisible colleges“ bestehen aus Wissenschaftlern, die sich über das Neueste in ihrem Fachgebiet austauschen, sich gegenseitig korrigieren oder ergänzen, aber auch bestätigen oder widerlegen. Kommunikationsprozesse zwischen den Beteiligten spielen eine zentrale Rolle. Man unterscheidet zwischen „formeller Kommunikation“, die über allgemein zugängliche Publikationen geführt wird und in der Regel über qualitätssichernde Mechanismen verfügt, und „informeller Kommunikation“, die überwiegend aus einem persönlichen Austausch besteht.

4 Der Fragebogen ist im Hauptbericht als Anhang wiedergegeben (DFG 2005). Der anonymisierte Datensatz wurde dem Zentralarchiv für empirische Sozialforschung zur Verfügung gestellt

Zwischen diesen beiden Polen sind andere Publikationstypen wie etwa „Graue Literatur“ anzusiedeln (Hornbostel, 1997).

Unterschiede innerhalb der einzelnen Wissenschaftsdisziplinen bezüglich der verwendeten Publikationstypen in der formellen Kommunikation sind gut erforscht (vgl. hierzu Small und Crane 1979). Es lassen sich in den letzten Jahren jedoch zwei Entwicklungen feststellen, die Anlass zur Vermutung liefern, dass sich in den Publikationskulturen eine gewisse Konvergenz eingestellt hat.

- Das durch das Internet gebotene Einsparpotenzial bei den Herstellungskosten für Druckerzeugnisse, so die Vermutung, würde die Bedeutung der einzelnen Publikationsmedien zugunsten der „Grauen Literatur“ verschieben und den Stellenwert der Monographien eher unterminieren (für eine Diskussion über die Kosten der wissenschaftlichen Literatur vgl. OECD 2005)
- Wissenschaftliche Veröffentlichungen dienen nicht nur dazu, Forschungsergebnisse zur wissenschaftlichen Diskussion zu stellen, sie werden auch zur Bewertung der Forschungsleistungen einzelner Wissenschaftler, Forschergruppen oder Forschungsinstitutionen herangezogen. Es wird vermutet, dass die Zunahme der Evaluationsaktivitäten, die den Publikationsdruck erhöht, Veränderungen im Publikationsverhalten der einzelnen Fachdisziplinen hervorrufen könnte. Demnach wurde der Trend zur „least publishable unit“ verstärkt, und bei der Veröffentlichung von Forschungsergebnissen wird Zeitschriftenaufsätze, die in die bibliometrischen Bewertungen einbezogen werden, gegenüber Monographien oder Bücherkapiteln der Vorzug gegeben.

Auf die Frage nach der Nutzungshäufigkeit der verschiedenen Publikationstypen zu Informationsgewinnung zeigen sich deutliche Unterschiede in den einzelnen Wissenschaftsbereichen (Tabelle 1).

In den Lebens- und Naturwissenschaften stellen wissenschaftliche Zeitschriften das dominante Rezeptionsmedium dar. Nahezu alle Befragten aus diesen Bereichen nutzen häufig Aufsätze in Zeitschriften, um sich über die neueste Entwicklung in ihrem Fach zu informieren. Aufsätze in Sammelbänden sind die zweitwichtigste Publikationsform; sie werden von rund 40 Prozent der Befragten aus diesen Bereichen häufig genutzt. In den Ingenieurwissenschaften werden neueste fachliche Entwicklungen ebenfalls primär über Aufsätze in Zeitschriften rezipiert (91 Prozent der Befragten), doch spielen hier Beiträge in Tagungsbänden, die von rund 85 Prozent der Befragten häufig genutzt werden, eine fast ebenso wichtige Rolle. Geistes- und Sozialwissenschaftler schließlich nutzen die breite Palette verschiedenartiger Publikationsformen. Neben Zeitschriftenaufsätze, die bei 94 Prozent der Befragten häufig Verwendung finden, kommt auch Beiträgen in Sammelbänden sowie Monographien eine herausragende Bedeutung zu: Gut drei Viertel der Befragten geben an, diese Medien bei der Informationsbeschaffung häufig zu nutzen.

Tabelle 1: Häufig genutzte Publikationsformen zur Beschaffung aktueller Informationen im Fachgebiet (Prozent)

	Geistes- u. Sozialwissen- schaften	Lebens- wissen- schaften	Natur- wissen- schaften	Ingenieur- wissen- schaften	Gesamt
Aufsätze in Zeitschriften	93,6	98,0	95,8	90,7	94,7
Beiträge in Sammelbänden	74,8	39,6	41,0	40,2	48,4
Monografien	70,9	20,5	33,1	36,1	39,6
Beiträge in Proceedings/ Tagungsbänden	44,6	26,0	31,9	83,9	44,9
Rezensionen	38,2	7,5	3,8	6,2	13,3
Graue Literatur	14,9	3,7	9,9	9,7	9,4
	100,0	100,0	100,0	100,0	100,0
Anzahl (n)	236	255	307	225	1.023

Basis: 1.028; fehlende Angaben: 5

Frage 1: Um sich in Ihrem Fachgebiet aktuell zu informieren, wie häufig benutzen Sie die folgenden Publikationsformen? (Antwortkategorie 1 = ‚sehr häufig‘ bis 5 = ‚sehr selten‘; Kategorie 1 und 2 zusammengefasst).

Die Präferenzen der Publikationstypen in der Rezeption entsprechen denen bei der Veröffentlichung eigener Forschungsarbeiten.

In den vier genannten Wissenschaftsbereichen haben nahezu alle Wissenschaftler in den letzten fünf Jahren mindestens einen Zeitschriftenaufsatz publiziert, andere Publikationsmedien wurden in unterschiedlichem Maße genutzt. Rund 92 Prozent der befragten Ingenieurwissenschaftler haben einen Beitrag in Proceedings publiziert; in den anderen Wissenschaftsbereichen trifft dies auf ca. zwei Drittel der Befragten (Geistes- und Sozialwissenschaften und Naturwissenschaften) oder weniger als die Hälfte (Lebenswissenschaften) der Befragten zu. Geistes- und Sozialwissenschaftler publizieren dagegen häufiger Beiträge in Sammelbänden (80 Prozent) und Monografien (60 Prozent).

Tabelle 2: Anzahl der bei wissenschaftlichen Verlagen auf konventionelle Weise publizierten Beiträge (Prozent)

	Geistes- u. Sozialwissen- schaften	Lebens- wissen- schaften	Natur- wissen- schaften	Ingenieur- wissen- schaften	Gesamt
Zeitschriftenaufsätze					
nicht publiziert	4,5	5,0	2,2	4,0	3,8
publiziert	95,5	95,0	97,8	96,0	96,2
Gesamt	100,0	100,0	100,0	100,0	100,0
Anzahl	201	219	268	201	889,0
Aufsätze in Proceedings/Tagungsbänden					
nicht publiziert	32,8	53,4	34,0	8,5	32,7
publiziert	67,2	46,6	66,0	91,5	67,3
Gesamt	100,0	100,0	100,0	100,0	100,0
Anzahl	201	219	268	201	889,0
Beiträge in Sammelbänden					
nicht publiziert	20,4	63,0	66,0	50,2	51,4
publiziert	79,6	37,0	34,0	49,8	48,6
Gesamt	100,0	100,0	100,0	100,0	100,0
Anzahl	201	219	268	201	889
Monografien					
nicht publiziert	40,3	79,0	85,1	71,1	70,3
publiziert	59,7	21,0	14,9	28,9	29,7
Gesamt	100,0	100,0	100,0	100,0	100,0
Anzahl	201	219	268	201	889

Basis: 1.028; fehlende Angaben: 139

Frage 16: In den letzten fünf Jahren: Wie viele Ihrer Arbeiten haben Sie in diesem Zeitraum auf konventionelle Weise publiziert - d.h. über Verlage in gedruckter oder digitaler Form entgeltpflichtig dem Leser angeboten?

Nach Rezeptions- und Publikationsverhalten lassen sich somit – trotz der unbestrittenen Bedeutung von Zeitschriften für die Wissenschaftskommunikation aller Fächer – Unterschiede zwischen Natur- und Lebenswissenschaften; Geistes- und Sozialwissenschaften sowie Ingenieurwissenschaften feststellen.

Betrachtet man außerdem die wichtigsten Kriterien bei der Auswahl von Zeitschriften für die Publikation eigener Arbeiten sowie die internationale Orientierung der Publikationsaktivitäten (vgl. DFG 2005, S. 25 ff.), zeigen sich bedeutende Unterschiede zwischen den einzelnen Wissenschaftsbereichen, die auf drei „Kulturen des wissenschaftlichen Kommunizierens“ hinweisen: eine Zeitschriftenkultur der Natur- und Lebenswissenschaftler, die eher an Procee-

dings orientierte Kultur der Ingenieurwissenschaftler und die auf Monographien ausgerichtete Kultur der Geistes- und Sozialwissenschaftler.

Alles in allem zeigt sich, dass die Publikationskultur sehr vom jeweiligen Wissenschaftsbereich geprägt ist. Unterschiede im Publikationsverhalten werden also bislang weder durch das Aufkommen des Internet und die damit verbundenen Möglichkeiten des elektronischen Publizierens, noch durch den gestiegenen Kostendruck oder den mit Evaluationen in Verbindung stehenden Publikationsdruck nivelliert; die fachspezifische Publikationskultur ist vielmehr sehr stabil geblieben. Die Persistenz unterschiedlicher Fachkulturen in den Natur- und Lebenswissenschaften, in den Ingenieurwissenschaften und in den Geisteswissenschaften erfordert daher eine fächerspezifische Betrachtungsperspektive hinsichtlich des Stellenwerts der neueren Publikationsformen unter Open Access.

3.2 Kenntnisse der Open Access-Plattformen und -Nutzung

Das in der Literatur uneinheitlich definierte Konzept von „Open Access“ wurde in dieser Studie in seinen drei wesentlichen Ausprägungen des entgeltfreien Zugangs zu wissenschaftlichem Wissen operationalisiert: Open Access-Zeitschriften, elektronische Postprints und elektronische Preprints (vgl. Abbildung 1).

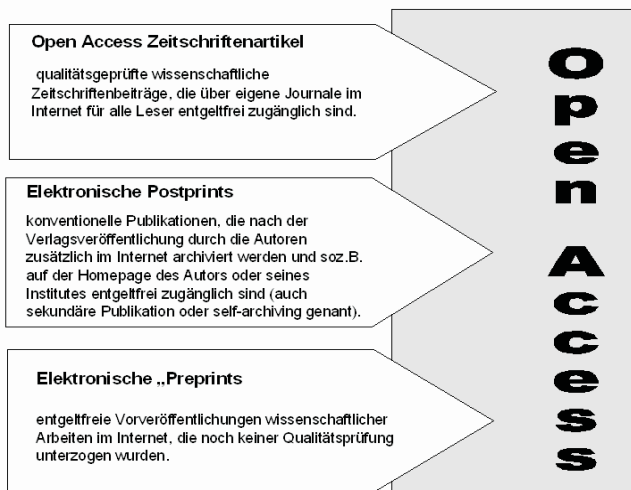


Abb. 1: Open Access – Begriffserläuterung

Die Wissenschaftler wurden gefragt, ob sie Open Access-Zeitschriften und Elektronische Preprints in ihrem Fach kennen und wie oft sie in den letzten Jahren in der einen oder anderen der oben genannten Plattformen publiziert haben (vgl. Tabelle 3).

Tabelle 3: Kenntnis und Nutzung von Open Access-Publikationsplattformen (in Prozent)

	Geistes- u. Sozialwissen- schaften	Lebens- wissen- schaften	Natur- wissen- schaften	Ingenieur- wissen- schaften	Gesamt
Open Access Zeitschriften					
hat mindestens einen Aufsatz in OA Zeitschrift publiziert	9,2	7,1	13,4	10,4	10,2
hat keinen Aufsatz in OA-Zeitschrift publiziert	28,4	38,8	27,5	14,5	27,5
kenne überhaupt keine OA- Zeitschrift in meinem Fach	62,4	54,2	59,1	75,1	62,2
Gesamt	229	240	298	221	988
Elektronische „Preprints“					
hat mindestens eine Publikation in Preprint-Server publiziert	8,0	12,6	34,5	10,7	17,6
hat keinen Aufsatz in Preprint-Server publiziert	5,5	12,6	13,9	9,6	10,6
kenne überhaupt keine Preprint-Server in meinem Fach	86,5	74,9	51,6	79,7	71,7
Gesamt	200	207	252	187	846
Elektronische „Postprints“					
hat mindestens einen Zeitschriftenaufsatz als elektronischen Postprint publiziert	18,4	24,7	31,3	30,8	26,7
hat keinen Zeitschriftenaufsatz als elektronischen Postprint publiziert	81,6	75,3	68,7	69,2	73,3
Gesamt	201	219	268	201	889

Frage 11: Kennen Sie Open Access Zeitschriften in Ihrem Fachgebiet?

Frage 14: In den letzten fünf Jahren – wie viele Aufsätze haben Sie in diesem Zeitraum veröffentlicht? Sind darunter auch welche, die in Open Access Zeitschriften erschienen sind?

Frage 18: In einigen Fächern ist es möglich, auch Vorabdrucke wissenschaftlicher Arbeiten für den entgeltfreien Zugriff in einem elektronischen Archiv im Internet bereitzustellen („Preprints“). Kennen Sie solche elektronischen Archive für Ihr Fach?

Frage 16: In den letzten fünf Jahren: Wie viele Ihrer Arbeiten haben Sie in diesem Zeitraum auf konventionelle Weise publiziert – d.h. über Verlage in gedruckter oder digitaler Form entgeltpflichtig dem Leser angeboten –, und wie viele davon haben Sie (oder Ihr Verlag) nochmals für den entgeltfreien Zugriff im Internet bereitgestellt?

Die Kenntnis von Open Access-Zeitschriften in ihrem Fach geben insgesamt knapp vier von zehn Befragten an. Hier bilden die Ingenieurwissenschaften die Ausnahme, denn nur jeder Vierte aus diesem Wissenschaftsbereich kennt Open Access-Zeitschriften. Preprint Server kennt jeder zweite befragte Naturwissenschaftler, gut jeder vierte Geistes- und Sozialwissenschaftler sowie jeder fünfte Ingenieurwissenschaftler. Indes sind Preprint-Server nur jedem zehnten Lebenswissenschaftler bekannt.

Noch geringer als der Bekanntheitsgrad der verschiedenen Publikationsformen ist die Zahl der eigenen Open Access-Veröffentlichungen. Nur jeder zehnte Befragte hat bereits mindestens einen Aufsatz in Open Access-Zeitschriften publiziert. Knapp jeder Siebte publiziert Vorabergebnisse seiner Forschung als „Preprints“. Hier schwankt der Anteil in den einzelnen Wissenschaftsbereichen: 35 Prozent der Naturwissenschaftler, aber nur 8 Prozent der Geistes- und Sozialwissenschaftler machen von dieser Möglichkeit Gebrauch. Häufiger jedoch wird die Möglichkeit einer „sekundären Publikation“ verfolgt. Jeder vierte Befragte hat mindestens einen Zeitschriftenaufsatz, der zunächst in einer „konventionellen Zeitschrift“ erschienen ist, der breiten Öffentlichkeit durch eine im Internet entgeltfrei zugängliche Archivkopie verfügbar gemacht.

Wissenschaftler, die Open Access-Zeitschriften in ihrem Fach kennen, wurden des Weiteren gebeten anzugeben, wie oft sie auf diese Zeitschriften zurückgreifen, um sich zu informieren. Die meisten Nutzer der Open Access-Zeitschriften zur Informationsgewinnung findet man in den Naturwissenschaften (jeder Vierte). In den Geistes- und Sozialwissenschaften sowie in den Lebenswissenschaften nutzt nur etwa jeder Fünfte Open Access-Zeitschriften „häufig“. In den Ingenieurwissenschaften, in denen drei Viertel der Befragten Kenntnis von OA-Zeitschriften haben, gibt immerhin knapp jeder Vierte, der Open Access-Zeitschriften kennt, an, diese häufig zu nutzen.

Festzuhalten bleibt, dass die befragten Wissenschaftler bislang relativ wenige Arbeiten im Open Access publiziert haben. Das liegt sicherlich zum großen Teil daran, dass ihnen die teils vielfältigen Möglichkeiten, Publikationen entgeltfrei zugänglich zu machen, nicht hinreichend bekannt sind. Andere Gründe für die Vorbehalte gegenüber diesen Publikationsformen lassen sich aus den Antworten auf die Frage, warum sie bislang nicht in Open Access-Zeitschriften publiziert haben, ableiten. Die meisten Befragten führen neben der Unkenntnis entsprechender Zeitschriften den als zu gering erachteten Verbreitungsgrad und deren in Fachkreisen oft als gering wahrgenommenes Renommee an. Die folgenden Zitate befragter Wissenschaftler belegen dies exemplarisch:

„Ich möchte von Fachleuten wahrgenommen werden. Diese lesen meines Wissens auch keine Open Access Journale.“ (Mann, Professor, analytische Chemie)

„Die OA-Zeitschriften in meinem Fachgebiet erreichen meine Zielgruppe (noch) nicht. Deren Renommee ist (noch) nicht (groß) genug.“ (Mann, Professor, Mathematik)

Insbesondere Wissenschaftler in frühen Stadien ihrer beruflichen Karriere stehen Open Access-Zeitschriften daher vielfach skeptisch gegenüber.

„Da ich noch am Beginn einer wissenschaftlichen Laufbahn stehe, ist die breite Wahrnehmung meiner Arbeit von immenser Bedeutung. Diese jedoch scheint mir momentan bei den gängigen Open Access-Zeitschriften noch nicht gegeben.“ (Mann, wiss. Nachwuchs, Molekulare Chemie)

3.3 Einstellungen zu Open Access

Den Befragten wurde eine Reihe von Aussagen, die in Debatten über Open Access die Chancen und Risiken des Open Access-Prinzips im Allgemeinen und der verschiedenen Open Access-Plattformen insbesondere thematisieren, vorgelegt. Sie sollten dann in Bezug auf diese Behauptungen ihre Zustimmung oder Ablehnung angeben. Gut zwei Drittel der Befragten sehen in Open Access einen Beitrag zur Verbesserung des Zugangs zu wissenschaftlichen Erkenntnissen; etwa der gleiche Anteil glaubt, dass sich die Publikationslandschaft durch Open Access nachhaltig verändern wird (Tabelle 4). Allerdings äußert die Mehrheit der Befragten Skepsis in Bezug auf die wissenschaftliche Qualität von Open Access-Veröffentlichungen (60 Prozent) und bezweifelt, dass diesen Publikationen die gebührende Anerkennung in Fachkreisen zuteil wird. Etwa zwei Drittel aller Befragten befürchtet, dass entgeltfrei zugängliche Veröffentlichungen bei der Beurteilung individueller wissenschaftlicher Leistungen wie auch bei der Begutachtung von Anträgen auf Fördermittel nur unzureichende Beachtung finden. Etwa drei Viertel sind der Auffassung, dass Publikationen im Open Access seltener zitiert werden als konventionelle Veröffentlichungen; 70 Prozent glauben, dass Open Access-Beiträge seltener in bibliographischen Systemen nachgewiesen werden als konventionelle Publikationen.

Tabelle 4: Zustimmung zu verschiedenen Aussagen zum Thema Open Access (in Prozent)

	Geistes- u. Sozialwissen- schaften	Lebens- wissen- schaften	Natur- wissen- schaften	Ingenieur- wissen- schaften	Gesamt
Erleichterung des Zugangs zu wissenschaftlichen Arbeiten für Wissenschaftler in Entwicklungsländern	87,1	90,3	86,7	85,0	87,3
Beitrag zur Verbesserung des Zugangs zu wissenschaftlichen Erkenntnissen	68,7	74,4	66,5	65,9	68,8
Beitrag zur nachhaltigen Veränderung des wissenschaftlichen Publikationswesens	56,7	61,1	66,8	58,6	61,4
Open Access ist als Medium des Publizierens noch zu wenig bekannt	85,0	82,7	79,8	81,3	82,0
Zitation ist seltener als bei konventionellen Veröffentlichungen	80,1	74,8	67,0	71,4	73,1
Bibliografischer Nachweis ist seltener als bei konventionellen Arbeiten	78,5	65,5	66,5	67,9	69,8
Unzureichende Berücksichtigung bei der Beurteilung individueller wissenschaftlicher Leistung	67,7	70,9	60,3	62,4	65,2
Unzureichende Berücksichtigung bei Anträgen auf Fördermittel	59,5	68,2	47,8	56,0	57,7
Langfristige Verfügbarkeit kann nicht garantiert werden	66,9	56,5	55,3	59,3	59,3
Autoren sollten durch Lizenzen in ihren Urheberrechten besser geschützt werden	67,4	60,0	56,3	49,2	58,2
Qualitätssicherung ist genauso sichergestellt wie bei konventionellen Veröffentlichungen	30,1	52,7	38,9	40,3	40,2
Insgesamt (n)	203	222	272	193	890

Basis: 1.028; fehlende Angaben: 138

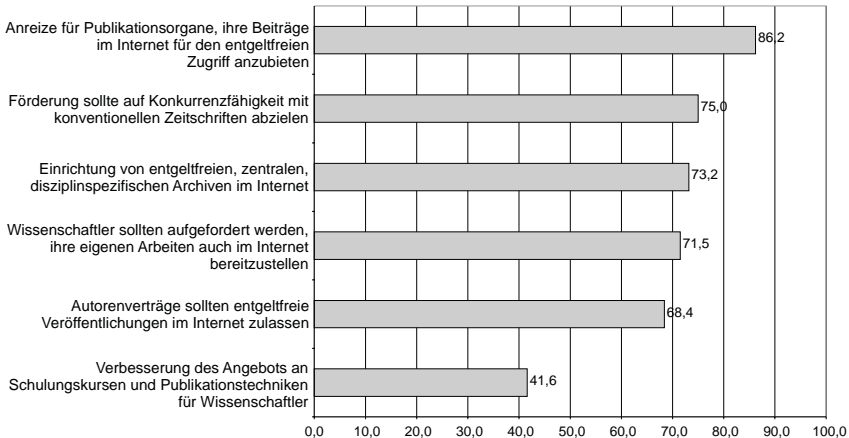
Frage 20: Hier haben wir eine Reihe von Aussagen zum Thema Open Access zusammengestellt, die in der gegenwärtigen Diskussion zum Thema wiederholt genannt werden. Welche halten Sie für zutreffend, welche nicht? (Antwortkategorie 1 = ‚stimmt genau‘ bis 5 = ‚stimmt überhaupt nicht‘; Kategorie 1 und 2 zusammengefasst).

Der starken Befürwortung von Open Access im Hinblick auf die Verbesserung des Zugangs zu wissenschaftlichen Arbeiten stehen also Bedenken gegenüber

dem geringen Bekanntheitsgrad und dem Stellenwert von Open Access-Publikationen in Fachkreisen gegenüber.

3.4 Open Access-Publikationen: Fördermaßnahmen und Vorschläge

Hinsichtlich konkreter Vorschläge zur Förderung von Open Access besteht über die Fächergrenzen hinweg relative Einigkeit darüber, dass vor allem etablierten Publikationsorganen Anreize geboten werden sollten, dort publizierte Beiträge auch im Internet entgeltfrei bereitzustellen (vgl. Abbildung 2). 86 Prozent der Befragten, die eine Förderung von Open Access für sinnvoll halten, sprechen sich für diesen Vorschlag aus. Im gleichen Zuge erhält auch der Vorschlag, Wissenschaftler zur Publizierung ihrer eigenen Arbeiten im Internet aufzufordern (72 Prozent), große Zustimmung. Drei Viertel der Befragten plädieren in diesem Zusammenhang für die Einrichtung von zentralen, disziplinspezifischen Archiven im Internet, in denen Autoren ihre Beiträge bereitstellen können. 68 Prozent wünschen sich, dass Autorenverträge die zusätzliche Veröffentlichung der Arbeiten im Internet zulassen sollten.



Frage 20: Hier haben wir eine Reihe von Aussagen zum Thema Open Access zusammengestellt, die in der gegenwärtigen Diskussion zum Thema wiederholt genannt werden. Welche halten Sie für zutreffend, welche nicht? (Skala 1 = ‚stimmt genau‘ bis 5 = ‚stimmt überhaupt nicht‘ sowie ‚kann ich nicht beurteilen‘; hier Skala 1 und 2 zusammengefasst)

Vor dem Hintergrund des geringeren Bekanntheitsgrades von Open Access-Plattformen sprechen sich die Befragten vor allem deutlich für die Förde-

rung von Diskussionen um Open Access in Hochschulen und Forschungseinrichtungen, sowie für die Verbesserung der Beratung und Information über die Möglichkeiten des Open Access-Publizierens aus. Auch die Förderung von Open Access-Zeitschriften wird von der Mehrheit der Befragten als wichtig erachtet: 75 Prozent sprechen sich dafür aus, diese Zeitschriften so zu stärken, dass sie mit konventionellen Zeitschriften mithalten könnten.

4 Zusammenfassende Betrachtungen

Ein Hauptanliegen der Studie bestand darin, auf der Grundlage der Entwicklung des Rezeptions- und Publikationsverhaltens in einzelnen Wissenschaftsdisziplinen das bisher herrschende Informationsdefizit bezüglich der Verbreitung und Nutzung von Open Access-Publikationen zu beheben. Die Ergebnisse zeigen, dass Veränderungen der Publikationstechnologie (Internet) die Fachkulturen im Publikationsverhalten noch wenig beeinflusst haben. Nach wie vor lassen sich drei Fachkulturen deutlich unterscheiden: Von den auf Zeitschriften fixierten und international orientierten Natur- und Lebenswissenschaften unterscheiden sich zum einen die Ingenieurwissenschaften, bei denen auch den „Proceedings“ eine wichtige Bedeutung zukommt. Zum anderen setzen aber auch die Geistes- und Sozialwissenschaften, die die breite Palette verschiedenartiger Publikationsmedien nutzen und stärker der Fachöffentlichkeit im eigenen Land zugewandt sind, andere Schwerpunkte.

Auf Seiten der Wissenschaftler lässt sich eine ambivalente Haltung gegenüber Open Access feststellen: auf der einen Seite sehen sie in Open Access eine Möglichkeit, den Informationszugang nachhaltig zu verbessern und plädieren für die Förderung von Open Access; auf der anderen Seite nutzen sie die Open Access-Plattformen selten, wenn es um die Veröffentlichung eigener Arbeiten geht. Die Gründe dieses Missverhältnisses zwischen Beurteilung und tatsächlicher Nutzung sind in den zum Ausdruck gebrachten Vorbehalten gegenüber dieser Publikationsform zu erkennen. Zwei Drittel der Befragten sind der Meinung, dass Open Access-Veröffentlichungen sowohl bei der Beurteilung individueller wissenschaftlicher Leistungen als auch bei der Begutachtung von Anträgen auf Fördermittel unzureichend anerkannt werden. Eine Mehrheit bezweifelt die Qualitätsstandards von Open Access-Publikationen. Es handelt sich um eine „strategische Kluft“ zwischen Zuspruch und Skepsis.

Überlegungen hinsichtlich einer adäquaten Förderpolitik⁵ müssen den geäußerten Vorbehalten Rechnung tragen und beachten, dass wissenschaftliche Ar-

5 Wege zum Wissen. Aktionsfelder zur Förderung des Open Access durch die DFG. Stellungnahme zur Studie „Publikationsstrategien im Wandel. Ergebnisse einer Untersuchung zum Publikationsverhalten DFG-geförderter Wissenschaftler unter besonderer Berücksichtigung von Open Access“ (www.dfg.de/lis/)

beiten nur in Organen veröffentlicht werden, denen die jeweilige Fachöffentlichkeit größtes Ansehen zuspricht.

Die sekundäre Publikation bereits anderweitig veröffentlichter Beiträge, die von immer mehr kommerziellen Verlagen unterstützt wird⁶, bietet hierfür eine optimale Lösung. Die Einbeziehung von Open Access-Zeitschriften in bibliometrischen Bewertungen dauert an (Lawrence, 2001; McVeigh, 2004; Antelman, 2005) und es werden die Zeitschriften aufgezeigt, die von der Fachöffentlichkeit wahr- und ernst genommen werden.

5 Literaturverzeichnis

Antelman, K. (2005). Do open-access articles have a greater research impact?

In: *College & Research Libraries*, 65 (1), p. 372-282.

DFG (2005). Publikationsstrategien im Wandel?

(www.dfg.de/zahlen_und_fakten/)

Hornbostel, Stefan (1997). *Wissenschaftsindikatoren. Bewertungen in der Wissenschaft*. Opladen: Westerdeutscher Verlag

Lawrence, S. (2001). Online or Invisible? In: *Nature*, 411 (6837), p. 521.

McVeigh, M. E. (2004). *Open Access Journals in the ISI Citation Databases: Analysis of Impact Factors and Citation Patterns. A citation study from Thomson Scientific*. Thomson Corporation <http://www.isinet.com/media/presentrep/essayspdf/openaccesscitations2.pdf>

OECD (2005). *Digital broadband content: scientific publishing*.

<http://www.oecd.org/dataoecd/42/12/35393145.pdf>

Angaben zum Autor

Dr. Alexis-Michel Mugabushaka
Deutsche Forschungsgemeinschaft
Kennedyallee 40
D - 53175 Bonn

Bereich Informationsmanagement
Referent für Evaluation und Statistik

6 Die sog. Romeo-Liste listet Verlage auf, die sekundären Publikationen zustimmen: www.sherpa.ac.uk/romeo.php

Open Access oder Fair Use?

Ein Vergleich nach Kosten-/Nutzen-Aspekten

Christopher N. Carlson
IWF Wissen und Medien gGmbH

Zusammenfassung

Der Terminus „Open Access“ ist erst seit wenigen Jahren gebräuchlich. Prima vista mag es erscheinen, dass die Fair-Use-Doktrin aus dem US-amerikanischen Urheberrecht das ältere Paradigma ist. Dennoch hat Open Access vom Denkanatz her schon eine gewisse Ahnenreihe. Hier eine Zeittafel wichtiger Wegbereiterereignisse von Open Access:

- 1971 Projekt Gutenberg
- 1979 Einführung USENET
- 1983 Arpanet stellt auf das TCP/IP-Protokoll um
- 1991 Einführung Gopher
- 1999 Open Archives Initiative (OAI)
- 2001 Budapester Open Access Initiative (BOAI)
- 2002 Einführung Creative Commons
- 2003 Bethesda Statement & Berliner Erklärung

Nach der Budapester Open Access Initiative (BOAI) aus dem Jahr 2001 hat die Berliner Erklärung über Open Access vom Herbst 2003 der einschlägigen Diskussion neue Impulse gegeben. Die Erklärung unterscheidet sich ganz grundlegend in der Herangehensweise etwa von der der USA, die die sog. Fair-Use-Doktrin im Urheberrecht gesetzlich verankert haben, während die Berliner Erklärung von einem freiwilligen System mit Internet-basierten Publikationsdepos und Anreizen für Deponenten ausgeht. Ziel beider Ansätze ist es, sowohl wissenschaftliche Arbeitsergebnisse als auch allgemeines Kulturerbe möglichst breit zugänglich zu machen.

1 Open Access: Definition

Für die Zwecke dieses Beitrags soll die Berliner Erklärung maßgeblich sein, zumal sie sich ausdrücklich auf die Aussagen der BOAI und des Bethesda Statements beruft. Demnach gilt folgende zweiteilige Definition:

- „1. Die Autoren und Rechteinhaber solcher Veröffentlichungen erteilen allen Benutzern das freie, unwiderrufliche und weltweite Zugangsrecht und die Erlaubnis, die Veröffentlichung für jeden verantwortlichen Zweck zu kopieren, zu benutzen, zu verteilen, zu übertragen und abzubilden unter der Bedingung der korrekten Nennung der Urheberschaft (wie bisher werden die Mechanismen der korrekten Berücksichtigung der Urheberschaft und der verantwortlichen Nutzung durch die Regeln der wissenschaftlichen Gemeinschaft zur Geltung gebracht) sowie das Recht, eine beschränkte Anzahl gedruckter Kopien für den persönlichen Gebrauch zu machen.
2. Eine vollständige Fassung der Veröffentlichung samt aller zugehörigen Begleitmaterialien wird zusammen mit einer Kopie der oben erwähnten Erlaubnis in einem geeigneten elektronischen Format auf mindestens einem online zugänglichen Archivserver mit geeigneten technischen Standards (wie die von Open Archive) hinterlegt und damit veröffentlicht. Der Archivserver muss betrieben werden von einer wissenschaftlichen Institution oder Gesellschaft, einer öffentlichen Institution oder einer anderen etablierten Organisation, die das „Prinzip des offenen Zugangs“, uneingeschränkte Verbreitung, Interoperabilität und Langzeitarchivierung zu verwirklichen sucht.“ (Berliner Erklärung 2003)

Der erste Teil der Definition zielt darauf, mit einer umfassenden Rechtseinräumung ein ungewöhnlich hohes Maß an Rechtssicherheit für die Nutzer/innen von Open Access-Publikationen zu erreichen. Der Hinweis auf das Fortbestehen der sog. Urheberpersönlichkeitsrechte ist ein Zugeständnis an das deutsche Urheberrechtsgesetz und nicht wirklich unbedingt konstitutiv für das Open Access-Paradigma.

Mit dem zweiten Teil der Definition wird die Dokumentenlieferung sichergestellt. Die Bestimmungen über die institutionelle Verankerung sollen zugleich die nachhaltige Konservierung und langfristige Verfügbarkeit der Dokumente gewährleisten.

1.1 Open Access und Veröffentlichungsreichweite

Die Menge an wissenschaftlicher Literatur ist so groß, dass Wissenschaftler/innen nicht alle relevanten Informationen finden und verwenden können. Dieses Phänomen, das als Informationsüberflutung oder bisweilen als data smog be-

zeichnet worden ist, führt nun einmal auch dazu, dass noch so werthaltige wissenschaftliche Arbeiten leicht im allgemeinen Rauschen untergehen.

Erste Zitierhäufigkeitsanalysen haben gezeigt, dass freie Online-Zugänglichkeit die Veröffentlichungsreichweite verbessern kann. Jedoch müssen auch verbesserte Recherchemöglichkeiten hinzukommen, um die Reichweite deutlich steigern zu können (Lawrence 2001). Hier kann der Open Access-Gedanke gut zum Tragen kommen: Die begrenzte Anzahl an OA-Archiven leitet Veröffentlichungen in eine begrenzte Anzahl von Kanälen. D.h., die Menge der in Frage kommenden URLs ist überschaubar - und auch bekannt. Diese können leichter „überwacht“ werden z.B. durch Blogs, Spinnen, Suchbots oder sog. Intelligente Agenten.

1.2 Open Access und Peer Review

Wie bisher werden die Mechanismen der korrekten Berücksichtigung der Urheberschaft und der verantwortlichen Nutzung durch die herkömmlichen Regeln der wissenschaftlichen Gemeinschaft zur Geltung gebracht. Open Access ist mit Peer Review vereinbar; das Internet unterstützt sogar innovative Formen von Peer Review wie z.B. das sog. open peer commentary (McKiernan 2004). Open peer commentary bedeutet, dass ein jedermann zugängliches Online-Diskussionsforum bzw. Weblog zur Verfügung steht, um Neuerscheinungen zu diskutieren. Es wird zwar kontrovers diskutiert, ob die Beiträge in einem solchen Forum unbedingt sehr werthaltig sein würden, aber es handelt sich ja um einen freien Markt von Ideen und Fachdiskursen. Jeder mag selbst entscheiden, wieviel Gewicht er den Kommentaren beimessen möchte. Außerdem kann open peer commentary durchaus zeitgleich mit der herkömmlichen Peer Review stattfinden.

Implizite Peer Review ist über Zugriffsstatistiken, Zitierhäufigkeiten und Verlinkungsstatistiken möglich. Dies ist schon relativ komfortabel mit den üblichen Webstatistikprogrammen möglich, die ohnehin standardmäßig in den meisten Einrichtungen arbeiten. Durch die Kanalisation der Open Access-Publikationen auf nur wenige URLs werden Zitieren und Verlinken erheblich erleichtert.

1.3 Open Access und die Kostenfrage

Open Access bedeutet immer die kostenlose Nutzung. Dem gegenüber steht die Forderung - z.B. in der Berliner Erklärung - nach einem Autorenhonorar. Es ist folglich eine Kosten-Reallokation notwendig: Von „Leserin zahlt“ zu „Autorin zahlt“, wobei es realistischerweise wohl eher der Arbeitgeber der Autorin sein wird, der an ihrer statt zahlt. Eine Reihe von wissenschaftlichen Verlagen haben sich inzwischen - hierzu durch die Open Access-Diskussion angeregt - zu der

Frage geäußert, wieviel es nun einmal kostet, um einen Beitrag in einer wissenschaftlichen Zeitschrift zu veröffentlichen. Es sind dabei Beträge zwischen US\$ 1500 und US\$ 100.000 genannt worden, wobei das obere Ende der Skala möglicherweise dazu dienen soll, durch einen Präemptivschlag die Diskussion im Keim zu ersticken.

Das „Open Choice“-Programm des Springer Verlags gewährt uns die Möglichkeit, in der Kostenfrage eine erste Hausnummer zu ermitteln. In diesem Programm zahlen Autor/innen US\$ 3000, um ihre Veröffentlichungen mit begrenzten Nutzungsrechten für Nutzer/innen kostenlos nutzbar zu machen (Springer Online 2005). Die starken Nutzungseinschränkungen, die Springer bei diesem Geschäftsmodell vorgesehen hat, bedeuten zwangsläufig, dass das Programm noch sehr weit vom Open Access entfernt ist. Auch kann getrost verneint werden, dass Springer das institutionelle Kriterium der Berliner Erklärung erfüllt, dass es „das Prinzip des offenen Zugangs ... zu verwirklichen sucht“.

Eine z.Z. weitgehend ungeklärte Frage dürfte sein, ob Einrichtungen der öffentlichen Hand ihre Ausgaben für die Beschaffung wissenschaftlicher Literatur etatmäßig in Kosten für die Bereitstellung wissenschaftlicher Publikationen werden umschichten können. Selbst wenn es sich im Endergebnis um ein reines Nullsummenspiel handelt, wird man hier bei den oft sehr traditionsbewußten Finanzverwaltungen voraussichtlich ein beachtliches Maß an Überzeugungsarbeit leisten müssen.

1.4 Open Access und Gemeinnützigkeit

Open Access-Publizieren ist gemeinwohlorientierter als kommerzielles Publizieren, da große Teile der wissenschaftlichen Gemeinschaft staatlich finanziert werden. Dieser Gesichtspunkt wird zusätzlich verstärkt, weil zur Zeit die staatlichen Bildungs- und Forschungssetats bestenfalls real stagnieren und mancherorts sogar schrumpfen. Schon 1999 hat die Kommission des DBI für Erwerbung und Bestandsentwicklung einen Offenen Brief an ausgewählte akademische Zeitschriftenverlage verschickt mit der Aufforderung, zumindest die explosionsartig wachsenden Kosten für wissenschaftliche Literatur etwas näher an die allgemeine Teuerungsrate zu bringen (DBI 1999). Die meisten der darauf antwortenden Verlage haben geltend gemacht, dass sie auf die Preisentwicklung keinen Einfluss hätten und haben stattdessen politische Lösungen angemahnt.

Die weitere Verbreitung (Reichweite) und die kostenlose Verfügbarkeit sind eine Art Gemeinschaftsdienst - auch und besonders gegenüber der Dritten Welt.

Ob dies genügen wird, um dem Open Access-Gedanken zum Durchbruch zu verhelfen? Die Berliner Erklärung führt eine relativ allgemein gehaltene Palette von möglichen Anreizen an. Forscher und Stipendiaten werden aufgefordert, ihre Veröffentlichungen nach dem Open Access-Prinzip zu machen, Einrichtungen sollen dafür eintreten, dass Open Access-Veröffentlichungen besonders bei

Berufungen und Beförderungen berücksichtigt werden usw. Das sind fraglos sehr unterstützenswerte Maßnahmen. Es ist jedoch zweifelhaft, ob dies - zusammen mit einem gewissen Rest-Idealismus der wissenschaftlichen Gemeinschaft - tatsächlich ausreichen wird, um Open Access entscheidend zu fördern. Wahrscheinlich bedarf es eines etwas handfesteren Anreizsystems, um die Bereitschaft der Publizierenden wirklich zu beflügeln.

2 Fair Use: Definition

Das US-Urheberrechtsgesetz (i.d.F. von 1976)¹ erlaubt die „Fair Use“ von urheberrechtlich geschützten Materialien: Für die Erschaffung neuer Werke, für Bildungszwecke und zur persönlichen Nutzung. Es bestimmt, dass für Zwecke wie Kritiken, Kommentare, Nachrichtenberichte, Unterricht und Lehre (einschließlich Mehrfachkopien) sowie für Wissenschaft und Forschung eine Urheberrechtsverletzung nicht vorliegt.

2.1 Fair Use: allgemeine Prinzipien

- Fair Use schränkt das Verwertungsmonopol des Urheberrechtinhabers ein.
- Fair Use ist nicht per se auf bestimmte Trägerformate beschränkt.
- Fair Use basiert auf einer gesetzlichen Regelung.
- Gemeinfreies Material in einem urheberrechtlich geschützten Werk bleibt frei.
- Man kann ein Werk immer ohne Zustimmung benutzen.
- Jedoch kann urheberrechtlich geschütztes intellektuelles Eigentum nur mit Zustimmung oder im Rahmen der Fair Use verwendet werden.

4 Faktoren helfen bei der Unterscheidung, ob eine Verwendung „fair“ ist:

- Der Verwendungszweck, z.B. kommerziell oder nicht-kommerziell.
- Das Wesen des Werks (schöpferisch/kompilativ/derivativ).
- Die Menge, die verwendet wird relativ zur Größe des Gesamtwerks.
- Die Auswirkungen der Verwendung auf das Marktpotential des Werkes.

Insbesondere Letzteres ist sehr wichtig, weil das Urheberrecht primär die kommerziellen Verwertungsrechte von Autoren schützen will. Fair Use findet folglich dort ihre Grenze, wo das Marktpotential erheblich beeinträchtigt wird.

1 Copyright Law of the United States. URL: <http://www.copyright.gov/title17/>

Im deutschen Urheberrechtsgesetz² findet Fair Use eine sinngemäße Entsprechung in den sog. ‘Schrankenbestimmungen’ (UrhG § 44a - § 63a). Jedoch muss man vermerken, dass die hier eingeräumten Nutzungsfreiheiten deutlich weniger weitreichend sind als im US-Urheberrecht. Die Rechtsposition der wissenschaftlichen Gemeinschaft hat sich bei der UrhG-Novelle des Jahres 2003 sogar deutlich verschlechtert; weitere Verschlechterungen sind durch den sog. „2. Korb“, der derzeit als Referentenentwurf beraten wird, zu erwarten.

Die drei Hauptformen der Fair Use sind:

- Schöpferische Nutzung durch Autoren/innen, die für eigene Werke aus Werken anderer Autoren/innen kopieren
- Persönliche Nutzung durch Individuen, die aus Werken zum Zweck der Unterhaltung oder der Information kopieren
- Nutzung für Bildungszwecke durch Lehrer/innen, Hochschullehrer/innen, Student/innen oder Schüler/innen

2.2 Fair Use und IKT

Es gibt unterschiedliche Einschätzungen, ob die Fair Use-Doktrin auf digitale und Online-Medien anwendbar ist. Obwohl das US-Urheberrechtsgesetz in der Fassung aus dem Jahr 1976 keinen Hinweis auf eine Einschränkung in bezug auf Formate oder Datenträger enthält, mag es schon plausibel argumentierbar sein, dass der Gesetzgeber einen primär auf Papier basierenden Informationsmarkt vor Augen hatte. Lineare Medien sind sachimmanent hinsichtlich ihrer möglichen Verbreitung begrenzt. Fair Use stammt aber eben aus der Ära der linearen Medien. Der Gesetzgeber hat womöglich nicht vorhergesehen, was Fair Use unter der Prämisse von leicht zu kopierenden digitalen und leicht zu downloadenden Online-Medien bedeuten würde. Der „digitale Unterschied“ könnte das Vermarktungsmonopol unterlaufen, das das Urheberrecht den Rechtsinhabern einräumen will. Dies könnte wiederum nach vorherrschender Rechtsprechung dazu führen, dass eine Anwendung nach Fair Use nicht mehr als zulässig anerkannt wird.

US-Gerichtsurteile aus jüngster Zeit schränken dann auch in diesem Sinne die Anwendung von Fair Use bei häufigen Nutzungstypen wie Fotokopien für Forschungszwecke, Hochschulunterricht, Selbststudium und selbst das Zitieren historischer Dokumente ein.

Die National Information Infrastructure (NII) Task Force³ in den USA hat Vorschläge in Bezug auf die „information superhighway“ gemacht, die die

2 Gesetz über Urheberrecht und verwandte Schutzrechte.

URL: <http://bundesrecht.juris.de/bundesrecht/urhg/>

3 Vgl. hierzu URL: <http://www.ibiblio.org/nii/NII-Task-Force.html>

Rechte der Urheberrechtsinhaber stärken, ohne entsprechende Fair Use-Bestimmungen vorzusehen. Viele dieser Empfehlungen sind dann in das Digital Millennium Copyright Act (DMCA)⁴ von 1998 eingeflossen. Besonders gravierend sind hier die neuesten Bestimmungen über DRM/ARM (digitales Rechtemanagement, automatisches Rechtemanagement). Sie schränken das gesetzlich verbriefte Recht auf Fair Use empfindlich ein. Eine solche Bestimmung hat auch über die UrhG-Novelle von 2003 ihren Weg in Gestalt des § 95a ins deutsche Urheberrecht gefunden: „Wirksame technische Maßnahmen zum Schutz eines nach diesem Gesetz geschützten Werkes oder eines anderen nach diesem Gesetz geschützten Schutzgegenstandes dürfen ohne Zustimmung des Rechtsinhabers nicht umgangen werden, soweit dem Handelnden bekannt ist oder den Umständen nach bekannt sein muss, dass die Umgehung erfolgt, um den Zugang zu einem solchen Werk oder Schutzgegenstand oder deren Nutzung zu ermöglichen.“ Mit anderen Worten, man kann durchaus ein Nutzungsrecht nach einer der o.g. Schrankenbestimmungen haben, das man aufgrund von § 95a dann auch nicht ausüben darf.

2.3 Fair Use und Marktfaktoren

Wie auch beim Patentrecht, geht das Urheberrecht von der Prämisse aus, dass der juristische Schutz intellektuellen Eigentums nötig ist, um Innovation zu ermöglichen. Dies mag in einigen Fällen so sein, absolut kann man dies nicht gelten lassen. Zum einen gibt es bekannte Fälle aus dem Wirtschaftsleben, wo das Gegenteil der Fall war: Fernsehen und Radio, ja sogar das Kaufvideogeschäft sind im Wesentlichen durch weitreichende Ausnahmebestimmungen zum Urheberrecht groß geworden. Sicherlich ist daher die Frage berechtigt, ob eine Güterabwägung zwischen dem Recht der Allgemeinheit auf leichte Zugänglichkeit zu wissenschaftlichen Veröffentlichungen und dem Recht der Urheberrechtsinhaberin (die die Wissenschaftlerin selbst sein kann, aber nicht sein muss) auf eine ökonomisch verstandene Verwertung ihrer Arbeit tatsächlich im Gesamtergebnis automatisch zugunsten des Letzteren führen muss?

Haben aber Wissenschaftsverlage ökonomische Nachteile durch Fair Use? Dies ist im Sinne des geltenden Rechts eine Schlüsselfrage der Fair Use: Wird das Vermarktungspotential des geschützten Werks im wesentlichen Umfang beeinträchtigt? Es würde den Rahmen dieser Arbeit sprengen, diese Frage umfassend behandeln zu wollen. Viel hängt davon ab, ob es gelingen kann, ein tragfähiges „Author Pays“-Modell zu entwickeln.

4 The Digital Millennium Copyright Act. Pub. L. No. 105-304, 112 Stat. 2860 (Oct. 28, 1998).
1. Copyright Office Summary, December 1998.
URL: www.copyright.gov/legislation/dmca.pdf

2.4 Wem gehören Urheberrechte?

Individualrechte an intellektuellem Eigentum werden in Deutschland sehr unterschiedlich wahrgenommen und behandelt. Die meisten Arbeitsverträge sehen vor, dass Erfindungen dem Arbeitgeber der Erfinderin gehören, aber eine Professorin besitzt das volle Urheberrecht an einer wissenschaftlichen Publikation, die sie geschrieben hat. Natürlich sollen die Urheberrechte an wissenschaftlichen Publikationen nicht qua Vertrag dem Arbeitgeber anheimfallen, aber die je unterschiedliche Behandlung von den beiden Typen von intellektuellem Eigentum sollte uns Anlass sein, darüber nachzudenken, wieso es sich so verhält.

Sowohl mit dem Patentrecht als auch mit dem Urheberrecht will der Gesetzgeber den Erschaffern intellektuellen Kapitals ein ökonomistisch zu verstehendes Verwertungsmonopol geben. Von wenigen - meist fächerbedingten - Ausnahmen abgesehen, sind weitaus die meisten wissenschaftlichen Publikationen kaum oder gar nicht einnahmeträchtig. Schon ein vergleichsweise kleiner Eingriff in das Urheberrecht würde genügen, um es Wissenschaftlern zu ermöglichen, zu beurkunden, dass sie mit ihrer Publikationstätigkeit keine kommerziellen Ziele verfolgen und somit auch bereit sind, ihre Werke für gemeinfrei zu erklären. Die rapide Ausbreitung des Open Access-Gedankens innerhalb der wissenschaftlichen Gemeinschaft belegt, dass viele Forscherinnen und Forscher hierzu bereit wären.

Ein wichtiges Gremium der OECD hat unlängst die OECD insgesamt dazu aufgerufen, einen Richtlinienentwurf für Open Access bei primären Forschungsdaten, die mit öffentlichen Mitteln gefördert wurden, zu entwickeln (OECD 2004). Auch wenn die erste Resonanz auf diesen Vorstoß vonseiten einiger Forschungseinrichtungen aus verschiedensten Gründen nicht sehr positiv gewesen ist, so ist der Grundgedanke dennoch bestechend: Was an Forschungsergebnissen entstanden ist, indem die Arbeit von der Allgemeinheit finanziert wurde, sollte möglicherweise tatsächlich nicht immer und unter allen Umständen proprietär gehalten werden dürfen.

In diesem Kontext seien die wichtigsten Ausprägungen der sog. Copyleft-Bewegung erwähnt.

Copyleft ist die Bezeichnung eines Schutzverfahrens, das verwendet werden kann, um sicherzustellen, dass freie Software (Freeware, Open Source) beziehungsweise freie Inhalte tatsächlich auch frei bleiben. Hierzu werden die Bestimmungen des Urheberrechts verwendet, um das proprietäre Monopol des Urhebers ins Gegenteil zu verkehren. Daher auch der Name - ein Wortspiel auf die Umfunktionierung des Urheberrechts (engl. = copyright).

Das Copyleft erzwingt die Freiheit von Weiterbearbeitungen und Fortentwicklungen eines gemeinfreien Ur-Werkes, um dadurch die proprietäre Vereinbarung freier Werke zu verhindern:

Bearbeitet ein Urheber (erlaubterweise) das Werk eines anderen, so erhält er nach geltender Rechtsprechung ein Mitspracherecht, darüber zu entscheiden, wie die Bearbeitung verwendet werden darf. War das ursprüngliche Werk noch für jeden frei kopierbar, verteilbar, veränderbar usw., so übertragen sich diese Freiheiten nicht automatisch auf die Bearbeitung - ganz im Gegenteil. Dies versuchen Copyleft-Lizenzen umzukehren. Da umgekehrt auch der ursprüngliche Autor ein Mitspracherecht an der Verwendung der Bearbeitung hat, erlaubt er nur dann die Weitergabe der Bearbeitungen, wenn sie zu den gleichen umfangreichen Rechten an jedermann lizenziert werden. Das Copyleft will somit verhindern, dass freie Software zum Ausgangsmaterial künstlich knapper proprietärer Software wird.

Das Copyleft-Verfahren wurde zunächst von der Free Software Foundation⁵ für die GNU-Lizenzen (GPL, LGPL und GFDL) verwendet. Inzwischen wird das Verfahren jedoch auch von anderen Organisationen emuliert. Dies betrifft beispielsweise diverse Musiklizenzen wie die OAL der Electronic Frontier Foundation⁶ sowie Versuche, eine für alle Werkstypen anwendbare Copyleft-Lizenz zu schaffen, zum Beispiel die Share-Alike-Lizenz des Projekts Creative Commons⁷.

Könnte dies womöglich der erste Schritt zu einem neuen Verständnis von Eigentum an Urheberrechten sein?

5 The Free Software Foundation. URL: <http://www.fsf.org/>

6 Electronic Frontier Foundation. URL: <http://www.eff.org/>

7 Creative Commons. URL: <http://creativecommons.org/>

2.5 Vergleich von Open Access und Fair Use

Tabelle 1

	Open Access		Fair Use	
	(+)	(-)	(+)	(-)
Verfügbarkeit	Selbstarchivierte OA-Veröffentlichungen sind schnell zu finden u. sofort zu nutzen	System ist freiwillig; die Auswahl an Publikationen könnte klein sein	Wissenschaftler/-innen kennen die traditionellen Methoden der Literaturbeschaffung sehr gut	Nutzer/innen müssen ihre Literatur selbst finden
Reichweite	Verbesserte Reichweite durch gezielte Recherche in den OA-Archiven		Verbesserte Reichweite durch gesetzlich verbriefte zustimmungsfreie Nutzung	
Qualitätssicherung	OA ist mit Peer Review vereinbar; innovative Peer Review-Methoden sind möglich		Traditionelle Peer Review funktioniert gut und wird nicht durch Fair Use beeinträchtigt	
Kosten	Nutzer/innen haben kostenlosen Zugang zu Literatur	Entweder „author pays“ - was Wissenschaftler/innen benachteiligt, die ihre Forschungsergebnisse mit der Gemeinschaft teilen, oder institutionelle Förderung für OA-Archive	„User pays“ ist marktorientierter: Angebot und Nachfrage sind im Gleichgewicht	„User pays“-Modell begrenzt den Zugang zu aktuellen Forschungsergebnissen für ökonomisch schwache Personen und Einrichtungen
Rechtssicherheit	Autor/innen räumen umfassende Nutzungsrechte als Teil der OA-Selbstarchivierung ein			Aushöhlung von Fair Use durch Gerichtsurteile und neue Gesetze hat die Rechtsunsicherheit vergrößert

Durch die o.a. Argumentenbilanz wird ersichtlich, dass Vor- und Nachteile beider Herangehensweisen sich in etwa die Waage halten.

Daher ist die Frage legitim ...

3 Können die beiden Ansätze miteinander kombiniert werden?

Open Access und Fair Use haben unterschiedliche Herkünfte und Befürworter: Open Access kommt aus der wissenschaftlichen Gemeinschaft, Fair Use kommt letztlich aus der Politik.

Es ist sehr unwahrscheinlich, dass der Staat eine zusätzliche Infrastruktur von Online-Archiven finanzieren wird. Für den Gesetzgeber ist Fair Use die kostengünstigere Variante.

Falls Open Access durch die wissenschaftliche Gemeinschaft im Wege der Selbstorganisation etabliert und durch ein „Autor zahlt“- oder „Arbeitgeber zahlt“-Modell finanziert werden kann, so kann der Gesetzgeber immer noch geeignete Fair Use-Bestimmungen in den nationalen Urheberrechtsgesetzen verankern, um Nutzer/innen mehr Rechtssicherheit bei der Nutzung von Werken zu geben, die nicht unter dem Open Access-Paradigma zur Verfügung gestellt werden. Eigentlich müßte das sogar passieren, denn viele Bestimmungen des deutschen bzw. europäischen Urheberrechts, wie derzeit kodifiziert, sind für die rechtliche Unterstützung von Open Access ungeeignet, z.T. sogar eher abträglich. Das Urheberrecht sollte schon Open Access-konform gestaltet sein.

4 Literaturangaben

Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen (2003)

http://www.zim.mpg.de/openaccess-berlin/BerlinDeclaration_dt.pdf

Copyright Law of the United States (1976 ff.). [Online]

<http://www.copyright.gov/title17/>

Creative Commons. [Online] URL: <http://creativecommons.org/>

DBI-Kommission (1999): Offener Brief. In: BIBLIOTHEKSDIENST 3. Jg. (1999), H. 5 S. 804 ff. [Online]

http://bibliotheksdienst.zlb.de/1999/1999_05_Erwerbung02.pdf

Digital Millenium Copyright Act, The (1998). Pub. L. No. 105-304, 112 Stat. 2860 (Oct. 28, 1998). 1. Copyright Office Summary. [Online]

www.copyright.gov/legislation/dmca.pdf

Electronic Frontier Foundation. URL: <http://www.eff.org/>

Free Software Foundation, The. [Online] URL: <http://www.fsf.org/>

Gesetz über Urheberrecht und verwandte Schutzrechte (2003).

[Online] URL: <http://bundesrecht.juris.de/bundesrecht/urhg/>

- Lawrence, S. (2001): Free online availability substantially increases a paper's impact. In: *Nature*, Volume 411, Number 6837, p. 521, 2001 [Online]
<http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>
- McKiernan, G. (2004): Peer Review in the Internet Age: Five (5) Easy Pieces. In: *Against the Grain* 16, no. 3 (June 2004) [Online]
www.public.iastate.edu/~gerrymck/DraftFive.doc
- NII Task Force (ca. 1995): [Online]
<http://www.ibiblio.org/nii/NII-Task-Force.html>
- OECD Committee for Scientific and Technological Policy at Ministerial Level (2004):
- Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29-30 January 2004 - Final Communique [Online]
http://www.oecd.org/document/0,2340,en_2649_34487_25998799_1_1_1_1,00.html
- Springer Online: <http://www.springeronline.com/sgw/cda/frontpage/0,11855,1-40359-12-115393-0,00.html>

Angaben zum Autor

Dr. Christopher N. Carlson
IWF Wissen und Medien gGmbH
Nonnenstieg 72
37075 Göttingen

1975-1980 M.A. in Literaturwissenschaften und Geschichte, 1984 Promotion in Literaturwissenschaften. 1987 Jahreslehrgang am Lehrinstitut für Dokumentation in Frankfurt/M. mit Abschluss als staatlich anerkannter Wissenschaftlicher Dokumentar. Forschungsschwerpunkte in Internet-Retrieval, intellektuelles Kapital, Open Access und Standardisierung in der Informationswirtschaft

Open Access in der Bildungsforschung

Die Einführung einer neuen Online-Zeitschrift

Sandra Schaffert, Bernhard Schmidt

Bildungsinstitut Martin Schön (BIMS), Traunstein

Lehrstuhl für Allgemeine Pädagogik und Bildungsforschung, LMU München

Zusammenfassung

Aktuell formiert sich eine Bewegung, die den freien und kostenlosen Zugang (Open Access) zu Texten wissenschaftlicher Zeitschriften fordert. Zunächst dokumentiert dieser Beitrag das aktuelle Angebot kostenlos zugänglicher deutschsprachiger Online-Literatur zur Bildungsforschung.

Im Bereich der Bildungsforschung lassen sich Online-Zeitschriften primär nach ihren unterschiedlichen Erscheinungsweisen systematisieren: als retropektive Digitalisierung, als Parallelausgabe zur Printversion oder als reine Online-Zeitschrift. Unter den vielen Beispielen befindet sich mit der „Erziehungswissenschaftlichen Revue“ auch eine reine Online-Zeitschrift eines wissenschaftlichen Verlags (Julius Klinkhardt). In Relation zum gesamten Publikationsaufkommen betrachtet ist „Open Access“ jedoch im Bereich der deutschsprachigen Sozialwissenschaften noch weit davon entfernt, Alltag wissenschaftlichen Publizierens zu sein. Befragungen von Bildungsforschern zeigen, dass diese maximal einmal im Jahr online publizieren und keine Veränderungen ihres Publikationsverhaltens erwarten.

Aus Sicht der Herausgeber der neuen Online-Zeitschrift „bildungsforschung“ werden anschließend die Herausforderungen und konzeptionellen Überlegungen vorgestellt, die mit einer solchen Neueinführung verknüpft sind: Die „bildungsforschung“ unterscheidet sich als reine Online-Zeitschrift, die kostenlos im World Wide Web zugänglich ist und weder eine Institution noch einen Verlag als Herausgeber hat, deutlich vom Angebot der gedruckten Fachzeitschriften, die in den Fachbibliotheken der Universitäten oder Forschungseinrichtungen zugänglich sind. (<http://bildungsforschung.org>)

1 Kostenlose Online-Literatur aus dem Bereich der Bildungsforschung

Im Frühjahr 2005 erschien die erste reguläre Ausgabe der wissenschaftlichen Online-Zeitschrift „bildungsforschung“ (<http://bildungsforschung.org>). Auch im herkömmlichen Printbereich gibt es keine Zeitschrift mit einer ähnlichen interdisziplinären Ausrichtung. Neben dem Themenbereich hat sich die Zeitschrift auch dem freien Zugang („Open Access“) verschrieben: Alle Beiträge sind kostenlos und frei im Internet zugänglich.

In diesem Beitrag wird ein Überblick über Open Access im Bereich der Bildungsforschung gegeben, die Idee und Konzeption der Online-Zeitschrift *bildungsforschung* vorgestellt und deren Entwicklung beschrieben.

Der Deutschen Forschungsgemeinschaft zufolge bedeutet Open Access einen für „Nutzer entgeltfreien Zugriff auf und die Möglichkeit umfassender Verwertung von qualitätsgeprüften wissenschaftlichen Publikationen im Internet.“¹ Nach Einschätzung von Mruck, Gradmann & Mey (2004) handelt es sich bei der Open Access-Bewegung nicht „nur um Überzeugungen vereinzelter, besonders internetaffiner Akteure [...], sondern um eine Bewegung von enormer Reichweite“ (Absatz 3); eine „wissenschaftliche Revolution formiert sich“ (vgl. Überschrift 2).

1.1 Typen von kostenlosen Online-Zeitschriften und Beispiele

Auch vor der *bildungsforschung*, und noch bevor die Bezeichnung des Open Access oder eine diesbezügliche wissenschaftliche Bewegung geläufig war, gab es schon Zeitschriften, die im Internet kostenlos zugänglich erschienen. Im Folgenden wird ein kurzer Einblick in das aktuelle Angebot der kostenlos zugänglichen Online-Literatur aus dem Gebiet der Bildungsforschung gegeben. Eine ausführlichere Dokumentation findet sich bei Schaffert (2004).

Drei Typen von kostenlosen Online-Zeitschriften lassen sich unterscheiden (s.a. Schaffert 2004): Online-Zeitschriften als Parallelausgabe von Printzeitschriften, retrospektiv digitalisierte Zeitschriften und reine Online-Zeitschriften.

Der derzeit wohl größte Anteil von wissenschaftlichen Online-Zeitschriften erscheint als Parallelausgabe von Printzeitschriften. Häufig sind dies Publikationen von Forschungseinrichtungen, beispielsweise sind die „Zeitschrift für Erwachsenenbildung“ des Deutschen Instituts für Erwachsenenbildung oder auch die „Kurzberichte des Instituts für Arbeitsmarkt- und Berufsforschung“ im Internet nachzulesen.

1 Aus der Ausschreibung „Umfrage zum Publikationsverhalten von Wissenschaftlern mit besonderer Berücksichtigung von Open Access“ der DFG, Vergabe-Nr.: LIS010/2004“ vom 2. August 2004.

Eine beachtliche Zahl der Online-Zeitschriften sind retrospektiv digitalisiert: So sind durch das Projekt „Retrospektive Digitalisierung von Bibliotheksbeständen“ der Universität Bielefeld deutschsprachige Zeitschriften von 1750 bis 1815, beispielsweise die „Neue Thalia“ von Friedrich von Schiller, im Internet zugänglich. Die Bibliothek für Bildungsgeschichtliche Forschung des Deutschen Instituts für Internationale Pädagogische Forschung (DIPF) offeriert diverse historische Zeitschriften zur Bildungsforschung aus dem Zeitraum von 1740 bis 1949 auf ihrer Website. Beispielsweise lassen sich dort die Ausgabe der Zeitschriften „Allgemeine Zeitung für Deutschlands Volksschullehrer“ (1817 - 1823); „Pädagogisches Museum“ (1778 - 1780), „Rossel's Wochenblatt für Elementarlehrer“ (1828 - 1833) und die „Zeitschrift für Geschichte der Erziehung und des Unterrichts“ (1911 - 1938) lesen. Retrospektiv digitalisierte Zeitschriften sind jedoch nicht nur für Bildungshistoriker interessant: Die Digitale Volltextbibliothek der Universität Innsbruck umfasst die Beiträge der älteren Jahrgänge von vier Zeitschriften zur Behindertenpädagogik (z.B. „impulse“).

Der dritte Typus von Online-Zeitschriften sind die „reinen“ Online-Zeitschriften. Die im Folgenden ausgewählten verfügen über ein Begutachtungsverfahren:

Die Erziehungswissenschaftliche Revue² des Verlags Julius Klinkhardt ist eine Rezensionszeitschrift für alle Teilbereiche der Erziehungswissenschaft und das einzige uns bekannte Angebot eines wissenschaftlichen Verlags. Eine Zeitschrift der Berufs- und Wirtschaftspädagogik, „bwp@“, wird seit 2002 von Mitgliedern der Universität Hamburg herausgegeben³. Auch die mehrsprachige Zeitschrift „Forum Qualitative Sozialforschung“⁴ beschäftigt sich mit Themen der Bildungsforschung und wird seit 1999 von Katja Mruck herausgegeben. Die Online-Zeitschrift für Theorie und Praxis der Medienbildung „Medienpädagogik“⁵ erscheint seit 2000 unter der Leitung der Kommission Medienpädagogik der Deutschen Gesellschaft für Erziehungswissenschaft und anderen. Zu den ältesten reinen Online-Zeitschriften aus dem Gebiet der deutschsprachigen Bildungsforschung ist wohl die „Zeitschrift für interkulturellen Fremdsprachenunterricht“⁶ (Deutsch als Fremdsprache) zu zählen, die bereits seit 1996 erscheint und von Britta Hufeisen und Manfred Prokop herausgegeben wird.

Schließlich noch ein Beispiel einer institutseigenen reinen Online-Zeitschrift mit den Mitarbeitern als Autoren: „Trends in Bildung International“ (Ti Bi) des Deutschen Instituts für internationale Pädagogische Forschung (DIPF) erscheint seit 2001.

2 <http://www.klinkhardt.de/EWR/>

3 <http://www.bwpat.de>

4 <http://www.qualitative-research.net/fqs/>

5 <http://www.medienpaed.com>

6 <http://www.tu-chemnitz.de/docs/ejournal/ejournal.html>

Soweit einige Beispiele von deutschsprachigen Online-Zeitschriften aus dem Gebiet der Bildungsforschung. Zusammenfassend können wir feststellen, dass die Online-Zeitschriften die Möglichkeiten des Internets kaum nutzen (z. B. die Einbindung von Hyperlinks oder farbigen Bildern). Wirft man einen Blick auf die Herausgeberschaften der reinen Online-Zeitschriften, fällt auf, dass zwar häufig formal als Herausgeber Institutionen agieren, die Tätigkeit an sich jedoch auf das Engagement von Einzelnen zurückzuführen ist. So ist es nicht weiter verwunderlich, dass das Angebot einer großen Dynamik unterliegt: Internet-Adressen ändern sich, Zeitschriften werden neu gegründet oder auch nach wenigen Ausgaben wieder eingestellt.

1.2 Disziplinen im Vergleich

Zahlen oder Studien zu kostenlos zugänglichen Online-Zeitschriften oder zum Publikationsverhalten von Bildungsforschern mit Open Access liegen kaum vor. Das Angebot und deren Nutzung mit anderen Disziplinen zu vergleichen, gelingt daher nur annähernd.

Der Statistik der Elektronischen Zeitschriftenbibliothek der Universität Regensburg zufolge liegt der Bereich „Pädagogik“ im Mittelfeld, was das Angebot an lizenzpflichtigen und kostenlosen Online-Zeitschriften insgesamt und auch das Angebot an kostenlosen Online-Zeitschriften betrifft. Zur Illustration zeigt Abbildung 1 ausgewählte Disziplinen und deren verzeichnete Online-Zeitschriften (Stand Mai 2005).

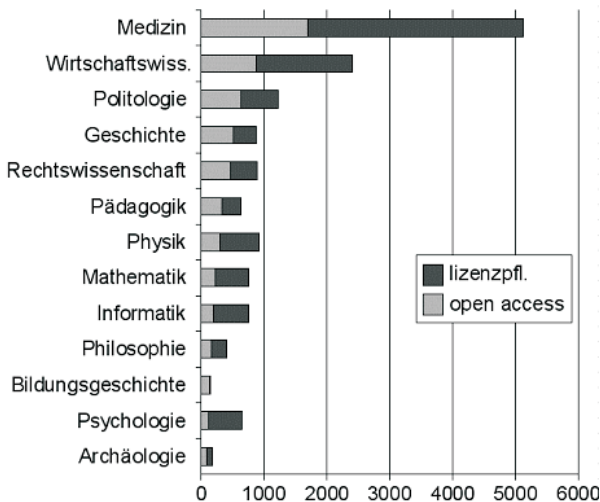


Abb. 1: Verzeichnis elektronischer Zeitschriften der Universität Regensburg, Stand 4.5.05

Aktuell gibt es noch wenig Untersuchungen zum Publikationsverhalten von Bildungsforschern. Das Prinzip und der Begriff des Open Access ist noch relativ jung und unbekannt, wie es beispielsweise Erich Weichselgartner (2005) in seiner Befragung von Psychologen darlegt. Die gerade veröffentlichten Ergebnisse einer DFG-Studie bestätigen zum Teil einen mangelnden Bekanntheitsgrad fachrelevanter Online-Zeitschriften bei den 1.600 befragten Wissenschaftlern aus Geistes-, Sozial-, Lebens-, Natur- und Ingenieurwissenschaften. Als wesentlichste Barriere für die Publikation in Open-Access-Zeitschriften (nicht einmal jeder zehnte Sozial- und Geisteswissenschaftler macht hiervon Gebrauch) erwies sich die Skepsis gegenüber deren wissenschaftlichem Renommee und die Befürchtung, dass frei zugängliche Online-Publikationen in der Fach-Öffentlichkeit weniger Beachtung fänden, als in Print-Publikationen (vgl. Deutsche Forschungsgemeinschaft 2005a).

An einer für ein so junges Thema relativ „alten“ Studie, einer Befragung von Miller-Kipp und Neuenhausen von 2003, nahmen 77 Erziehungswissenschaftler teil. Die Befragung ergab, dass sie maximal einmal im Jahr online publizieren, in den herkömmlichen Printmedien von einmal monatlich bis einmal jährlich. Miller-Kipp und Neuenhausen kommen zu dem Schluss: „Eine Veränderung des Publikationsverhaltens ist nicht in Sicht. Alle diesbezüglichen Aussagen werden von den Befragten mehr oder weniger verneint“ (S. 7).

2 Konzeption der „bildungsforschung“

Die oben angedeuteten fachkulturellen Differenzen in der Nutzung elektronischer Publikationsorgane verlieren an Bedeutung, wenn ein interdisziplinär zu bearbeitender Forschungsbereich zum Gegenstand einer Online-Zeitschrift wird. Die Bildungsforschung ist ein solcher Bereich (vgl. Tippelt 2002). Zwar widmen sich sozialwissenschaftliche Disziplinen – allen voran die Pädagogik und die pädagogische Psychologie – traditionell vielen Themen der Bildungsforschung, jedoch impliziert Bildungsforschung auch philosophische, soziologische, juristische, historische und politologische Fragestellungen und muss auch die aktuellen neurobiologischen (vgl. Tippelt/Schmidt 2005) und betriebswirtschaftlichen Diskussionen (vgl. Clar/Doré/Mohr 1997) aufgreifen. Es liegt also nahe – deutlicher gesagt: es ist unumgänglich –, bei Neueinführung einer (Online-)Zeitschrift zur Bildungsforschung diese interdisziplinäre Verhaftung des Themas zu berücksichtigen. Die Online-Zeitschrift *bildungsforschung* hat es sich explizit zur Aufgabe gemacht, auch den Austausch zwischen den an diesem Thema beteiligten Fächern zu unterstützen und die jeweiligen Schwerpunktthemen aus der Perspektive verschiedener Forschungstraditionen zu beleuchten (vgl. Schaffert/Schmidt 2004).

Ein verstärkter interdisziplinärer Austausch wird von verschiedenen nationalen und internationalen Organisationen ebenso gefordert und gefördert (vgl. Manathunga 2003), wie eine stärkere Förderung des wissenschaftlichen Nachwuchses (vgl. Enders 2005). Dazu gehört nach Auffassung der Herausgeber der *bildungsforschung* auch, jungen Wissenschaftlern ein Forum für die Publikation ihrer Forschungsergebnisse zu bieten. Dies bedeutet nicht etwa eine Reduzierung der methodischen und inhaltlichen Anforderungen an deren Beiträge, aber ein Peer-Review-Verfahren, das die Forschungsbiographie der Verfasser nicht berücksichtigt und sich rein auf die wissenschaftlichen Standards des jeweiligen Artikels bezieht.

Damit ist bereits ein weiteres zentrales Element der Online-Zeitschrift *bildungsforschung* angesprochen. Zur Sicherung der Qualität der Artikel nach den Standards wissenschaftlicher Publikationen wird jeder eingereichte Beitrag in einem Peer Reviewing von mindestens zwei externen Gutachtern evaluiert. Dieses Verfahren entspricht den Forderungen der Open-Access-Bewegung (vgl. Mruck/Gradmann/Mey 2004) und knüpft an die Vorgehensweise etablierter internationaler Online- und Print-Zeitschriften an.

Die bisher genannten konzeptionellen Grundgedanken der *bildungsforschung* sind sowohl für traditionell gedruckte Zeitschriften wie für Online-Journals gleichermaßen relevant. Die Entscheidung für eine Online-Zeitschrift ging im Wesentlichen auf die Ideen der bereits vorgestellten Open Access-Bewegung zurück, denn eine explizit der Bildungsforschung gewidmete Zeitschrift existiert – wie bereits gezeigt wurde – auch unter den gedruckten Publikationsorganen noch nicht. Neben den eher wissenschaftspolitischen Argumenten des Open Access (vgl. ebd.) gibt es jedoch weitere wesentliche Kriterien, die Internet-Publikationen von gedruckten Zeitschriften unterscheiden. Ein schneller Zugriff auf relevante Artikel gehört ebenso zu den Vorteilen des Mediums Internet, wie die zeitnahe Publikation von Forschungsergebnissen und der Wegfall von – bei Print-Publikationen unvermeidbaren – Einschränkungen hinsichtlich dem Umfang von Einzelbeiträgen und der Einbindung von Grafiken bzw. Bildern. Darüber hinaus ermöglichen Online-Publikationen auch die Integration von Audio- und Video-Files sowie die direkte Verlinkung mit anderen wissenschaftlich relevanten Inhalten im Netz und ermöglichen so die Integration eines an sich linear aufgebauten Textes in die netzwerkartige Metastruktur des World Wide Webs. Lernpsychologischen Studien, die in sich linear aufgebauten Texten einen lernförderlicheren Charakter zusprechen (vgl. Gerdes 1997), kann damit Rechnung getragen werden, ohne gleichzeitig auf die Vorteile der Einbettung von Wissensbeständen in komplexe Informationsnetzwerke zu verzichten (vgl. Winter 1998; Schmidt 2004, S. 76 ff.). Durch den Wegfall eines Verlags entstehen für die Herausgeber und Autoren allerdings zusätzliche, teilweise auch medienspezifische Aufgaben, auf die auch die Deutsche Forschungsgemeinschaft in einem Positionspapier aufmerksam macht (vgl. Deutsche For-

schungsgemeinschaft 2005b). Die Aufgaben des Lektorats wurden im Fall der *bildungsforschung* zunächst von den Herausgebern mitübernommen, ebenso wie das Einstellen der Beiträge in das ausgewählte Content Management System Plone⁷. Darüber hinaus erwies sich auch die Betreuung der Gutachter und Autoren als wichtige und zeitintensive Aufgabe für die Herausgeber, zumal auch für viele Autoren die medienpezifisch veränderten Richtlinien zur Manuskriptgestaltung noch ungewohnt waren.

The screenshot displays the website interface for *bildungsforschung.org*. At the top, there is a navigation bar with links: [artseite](#), [aktuelle ausgabe](#), [archiv](#), [editorial](#), [newsletter](#), [suche](#), [vorschau](#), and [bibliograph](#). Below this, a breadcrumb trail reads: [sie sind hier: startseite](#) » [archiv](#) » [2005 jahrgang 2 ausgabe 1](#) » [zum themenschwerpunkt "bildungsforschung: disziplinäre, ethische und empirische beispiele"](#).

The main content area is divided into several sections:

- navigation**: A sidebar menu with options like Editorial, Aktuelle Ausgabe, Vorschau, Archiv, and a tree view of the archive (2004 Jahrgang 1, Ausgabe 1; 2005 Jahrgang 2, Ausgabe 1).
- aktuelle artikel**: A list of recent articles, including "interdisziplinäre Online-Zeitschrift" (2004-10-02) and "Bibliographie" (2005-03-20).
- inhaltsverzeichnis**: A table of contents for the current issue, listing "1. Zum Inhalt", "2. Ausblick", "Autoren", "Literatur", and "Zitation".
- Article View**: The main content area shows the title "Zum Themenschwerpunkt 'Bildungsforschung: disziplinäre, methodische und empirische Beispiele'" by Sandra Schaffert and Bernhard Schmidt. The abstract states: "In diesem Beitrag wird ein kurzer Überblick über die Artikel der zweiten Ausgabe der 'bildungsforschung' mit dem Themenschwerpunkt 'Bildungsforschung: disziplinäre, methodische und empirische Beispiele' gegeben." Below this is the section "1. Zum Inhalt" with a sub-paragraph: "In der ersten regulären Ausgabe der bildungsforschung erscheinen sechs Beiträge. Trotz der Vielfalt ihrer inhaltlichen, theoretischen und methodischen Zugänge können sie das Schwerpunktthema 'Bildungsforschung: disziplinäre, methodische und empirische Beispiele' abdecken."

Abb. 2: Screenshot von bildungsforschung.org

Diese und andere Herausforderungen prägten den Weg zur Einführung einer neuen Online-Zeitschrift, wie er im Folgenden kurz nachgezeichnet werden soll.

⁷ <http://plone.org>

3 Entwicklungsschritte

Der folgende Abriss verschiedener Stadien und Schlüsselprobleme der Entstehung einer neuen Online-Zeitschrift sind nur Ausschnitte aus einem komplexen Projekt. Die sich daraus ableitenden Erfahrungen und Perspektiven sollen interessierten Wissenschaftlern und Publizisten, die selbst – ähnlich wie die Verfasser Anfang 2004 – nach digitalen Publikationsmöglichkeiten suchen, eine kleine Hilfestellung bieten. Vorab ist zu sagen, dass der Erfolg des beschriebenen Projekts zum jetzigen Zeitpunkt kaum beurteilt werden kann und sich erst an einer festen Etablierung in der Wissenschaftslandschaft im positiven Fall bzw. an dem allmählichen Verschwinden in der Bedeutungslosigkeit der Online-Zeitschrift im negativen Fall ablesen lassen wird.

Wertvolle Hinweise für die Gestaltung einer neuen Online-Zeitschrift *bildungsforschung* konnten aus bestehenden Formaten entnommen werden. Von den bereits kurz vorgestellten bisher vorliegenden Online-Zeitschriften aus dem sozialwissenschaftlichen Bereich ist insbesondere das Forum Qualitative Sozialforschung (fqs)⁸ hervorzuheben. Die internationale und interdisziplinäre Ausrichtung des fqs hatte hier ebenso Vorbildcharakter, wie die Verknüpfung der Beiträge mit weiteren Informationsangeboten (Newsletter, Linklisten, Tagungstermine). Auch wenn nicht alle der genannten Elemente übernommen wurden und stattdessen weitere Kommunikationstools ergänzt wurden, so war es in erster Linie der dahinter liegende Grundgedanke einer stärkeren Vernetzung von Herausgebern, Autoren und Lesern im Sinne einer Learning Community (vgl. Wilson/Ryder 1996) sowie ein fundiertes und transparentes Reviewing-Verfahren, die auch für die *bildungsforschung* fruchtbar gemacht wurden.

Als wesentliche Hilfe erwiesen sich bereits in der Konzeptionsphase zahlreiche Rücksprachen mit Bildungsforschern im In- und Ausland und Herausgebern von Online-Zeitschriften. Diese Gespräche verwiesen immer wieder auf einen deutlichen Bedarf eines interdisziplinären und online verfügbaren Publikationsorgans für Themen der Bildungsforschung. Diese – zwar sicherlich wenig repräsentative – Exploration der Bedarfslage signalisierte zumindest eine breite Aufgeschlossenheit gegenüber der Nutzung moderner Kommunikationswege zur Publikation wissenschaftlicher Befunde und zur eigenen Informationsrecherche, wie es auch bisherige empirische Studien nahe legen. So zeigt eine BMBF-Studie, dass Hochschuldozenten fachübergreifend die Bedeutung moderner Informations- und Kommunikationstechnologien zur Verbreitung wissenschaftlicher Ergebnisse durchaus hoch einschätzen, im Kontext des universitären Lehrbetriebs allerdings die Kompetenz der Studierenden zur gezielten Informationsrecherche im Internet noch für zu gering erachten (vgl. Klatt u.a. 2001).

8 <http://www.qualitative-forschung.de/publishing/fqs/index.php>

Viel kollegialer Zuspruch und die ermutigenden Forschungsbefunde bestärkten die Herausgeber in ihrem Vorhaben, über dessen technische Umsetzung nun nachgedacht werden musste. Als Open Source-Lösung bot sich hier das Content Management System (CMS) *Plone* an. Diese Plattform bietet neben der Möglichkeit zur systematisierten Publikation wissenschaftlicher Artikel und zur Einbindung multimedialer Elemente auch ein umfassendes Benutzerverwaltungssystem sowie zahlreiche Optionen zur Integration weiterer Inhaltsbereiche. Wichtig war für eine Online-Zeitschrift, v.a. im Hinblick auf das Reviewing Verfahren, einzelne Artikel während des Entstehungs- oder Begutachtungsprozesses nur einem begrenzten und klar definierten Benutzerkreis zugänglich zu machen. Dennoch erwies sich die umfassende Unterstützung durch einen Informatiker für die individuelle Anpassung des CMS als unabdingbar. Eine intensive Nutzung der Möglichkeiten, die durch das Medium Internet zur Verfügung stehen und eine permanente Betreuung des zugrundeliegenden CMS erfordert eine hohe informationstechnische Kompetenz, was bei der Planung eines Online-Publikations-Projekts bedacht werden muss (vgl. hierzu auch das Positionspapier der Deutschen Forschungsgemeinschaft 2005b, S. 12).

Nachdem im Falle der *bildungsforschung* diese technischen Rahmenbedingungen schnell zur Zufriedenheit der Herausgeber gestaltet werden konnten, galt es, Autoren für eine – praktisch noch nicht existente – Zeitschrift zu gewinnen. Hierbei bewährte sich neben einer ersten Vorab-Ausgabe im Herbst 2004, in der die Herausgeber ihr Vorhaben und das Konzept der Online-Zeitschrift skizzierten (vgl. Schaffert/Schmidt 2004), v.a. die gezielte persönliche Ansprache von Bildungsforschern. Wobei bewusst eine Mischung von Beiträgen namhafter und etablierter Wissenschaftler und Artikeln des wissenschaftlichen Nachwuchses angestrebt wurde. Aufgrund des großen Interesses vieler der angesprochenen Autoren konnte im März 2005 die erste reguläre Ausgabe der *bildungsforschung* mit dem Schwerpunkt „*Bildungsforschung: disziplinäre, methodische und empirische Beispiele*“ im Netz publiziert werden (vgl. Schaffert/Schmidt 2005), weitere Ausgaben sollen halbjährlich folgen. Im Herbst 2005 erschien das von Cathleen Grunert herausgegebene Themenheft „*Bildungsbiographien und Bildungsverläufe*“, zum Frühjahr 2006 plant Elisabeth Zwick die Herausgabe eines Themenhefts zu „*Bildung und Krieg*“.

Eine wesentliche Voraussetzung für die breite Wahrnehmung der Beiträge in der *bildungsforschung* und damit auch für das weitere Bestehen der Online-Zeitschrift ist deren Erreichbarkeit und Bekanntheitsgrad (vgl. Willinsky/Wolfson 2001). Zentral ist u. E. die Präsenz in einschlägigen Fachdatenbanken und die Auffindbarkeit durch Suchmaschinen, deren prominentestes Beispiel sicherlich Google ist. Der genaue Page Rank-Algorithmus von Google ist nur in Grundzügen bekannt. Bekannt ist beispielsweise, dass, wer in den „Trefferlisten“ der derzeit populärsten Suchmaschine möglichst weit vorne aufgelistet werden will, dies unter anderem durch die URL der Website, die Gestaltung ih-

rer Metadaten und die Häufigkeit der Verlinkung auf die Website beeinflussen kann (mehr hierzu bei Fries u.a. 2001). Gleichzeitig ist eine Präsenz in den einschlägigen wissenschaftlichen Datenbanken wie dem Directory of Open Access Journals, dem Deutschen Bildungsserver oder dem Regensburger Verbundkatalog auch bei gezielter Internet-Recherche wichtig. Als Ergebnis dieser manchmal mühseligen Arbeit der Eintragung in Verzeichnisse und Suchmaschinen befindet sich die bildungsforschung nun beispielsweise kontinuierlich an zweiter Stelle der Google-"Treffer" bei der Suche nach dem Begriff „Bildungsforschung“ (an erster Stelle liegt wohl uneinholbar das Max-Planck-Institut für Bildungsforschung) und wir können stetig steigende Zugriffszahlen auf die Website sowie kontinuierliche Bestellungen des regelmäßig erscheinenden Newsletters verzeichnen.

Wenn die Beiträge der bildungsforschung gefunden werden sollen, müssen sie auch in den traditionellen Systemen und Datenbanken zu finden sein. Für die bildungsforschung wurde daher einerseits eine ISSN (International Standard Serial Number) beantragt, die vielerorts erst die Aufnahme in die Bibliothekssysteme ermöglicht (ISSN 1860-8213). Zudem werden die Beiträge in der Datenbank „Fachinformation Bildung“ verzeichnet und sind auch in der Datenbank des Deutschen Bildungsservers zu recherchieren.

Im Interesse der Autor/innen und Leser/innen ist ein weitgehender, störungsfreier Zugang zu den einzelnen Artikeln notwendig. Dies beinhaltet auch eine stets gleichlautende Adresszeile eines Artikels, damit dieser auch zitierfähig ist sowie eine regelmäßige und verlässliche Datensicherung. Zudem werden die Beiträge der bildungsforschung auch auf dem Archivserver der Deutschen Bibliothek archiviert und sind daher dauerhaft verfügbar.

Trotz der wachsenden Zahl reiner Online-Zeitschriften mussten wir technische und einige administrative Aufgaben nicht nur erfüllen, sondern auch manchmal erst „entdecken“, da Erfahrungsberichte oder gar ein Manual für potentielle Herausgeber von Online-Zeitschriften nicht auffindbar waren. Zukünftig sollten Wissenschaftler, die eine elektronische Zeitschrift herausgeben wollen, auch die weitere Entwicklung und das Angebot von German Academic Publishers (GAP) beobachten: GAP will künftig allen, die neue Wege des elektronischen Publizierens im Wissenschaftsbereich gehen wollen, eine organisatorische und technische Infrastruktur im Verbund bieten.

4 Herausforderungen

Zu den zentralen Erfolgskriterien einer wissenschaftlichen Zeitschrift gehören deren Bekanntheitsgrad in der Scientific Community und die Bedeutung, die dem Publikationsorgan von Wissenschaftlern beigemessen wird. Diese Wertschätzung drückt sich nicht nur quantifizierend in Umfragen oder in dem –

meist über den Social Science Citation Index⁹ erhobenen und nicht unumstrittenen (vgl. Fröhlich 2003) – Impact-Faktor aus, sondern spiegelt sich im Interesse potenzieller Autoren und Mitherausgeber. Während bei gedruckten Zeitschriften die Auflagenhöhe und bei kostenpflichtigen Online-Zeitschriften die Zahl der registrierten Nutzer das kommerziell entscheidendste und unbestechlichste Maß für die Relevanz einer Fachzeitschrift ist, können diese Messgrößen im Falle von Open Access Journals durch Zugriffsstatistiken zumindest teilweise kompensiert werden. Allerdings lässt sich auf Basis der so genannten Logfile-Protokolle nur schwerlich ein zufälliges „Ansurfen“ der Online-Zeitschrift von einem interessierten Lesen einzelner Beiträge differenzieren. Nichtsdestotrotz kann so neben den Rückmeldungen von Autoren und Lesern auch ein empirisch härteres Maß für den Erfolg und Bekanntheitsgrad einer Online-Zeitschrift herangezogen werden.

Die genannten statistischen Daten lassen zunächst keine Rückschlüsse auf die Zusammensetzung der Leserschaft zu. Die *bildungsforschung* will dezidiert auch Zielgruppen jenseits der Universitäten und Forschungseinrichtungen ansprechen. Zwar gehören Wissenschaftler und Studierende sicherlich zu den wichtigsten Adressaten der neuen Online-Zeitschrift, ebenso sollen aber auch Praktiker aus dem Bildungssektor von dem interdisziplinären Austausch profitieren und diesen mitgestalten. Die *bildungsforschung* fordert daher insbesondere diesen Interessentenkreis auf, sich mit Praxisberichten oder Evaluationsstudien an der Fachdiskussion zu beteiligen.

Die damit anvisierte Erweiterung des Nutzerkreises erfordert zusätzliche Öffentlichkeitsarbeit sowie – bei erfolgreicher Umsetzung – erhöhten Aufwand hinsichtlich Organisation, Begutachtung und Lektorat. Um diese Aufgaben dauerhaft professionell wahrnehmen zu können, bedarf es mittelfristig einer Erweiterung des Teams, das aktuell aus drei ehrenamtlichen Herausgebern und einem ebenfalls ehrenamtlichen technischen Betreuer besteht. Der Einbezug von externen Herausgebern für einzelne Thementile – wie er bereits für die nächsten Ausgaben vorgesehen ist – kann hier als erster Schritt gesehen werden, darüber hinaus ist aber eine Einbindung drittmittelfinanzierter studentischer Hilfskräfte anzustreben. Entsprechende „Fördertöpfe“ existieren, wenn auch bislang nur sehr begrenzt und unter der Voraussetzung einer institutionellen Anbindung.

Insgesamt scheint das Projekt *bildungsforschung* bislang auf einem guten Weg zu sein, seine dauerhafte Existenz wird sich aber erst mit seiner Etablierung in der wissenschaftlichen Publikationslandschaft und langfristig sichergestellten personellen Ressourcen beweisen.

9 <http://isi15.isiknowledge.com/portal.cgi/>

Literatur

- Clar, Günter; Doré, Julia; Mohr, Hans (Hrsg.) (1997): *Humankapital und Wissen. Grundlagen einer nachhaltigen Entwicklung*. Berlin.
- Deutsche Forschungsgemeinschaft (Hrsg.) (2005a): *Publikationsstrategien im Wandel? Ergebnisse einer Umfrage zum Publikations- und Rezeptionsverhalten unter besonderer Berücksichtigung von Open Access*. Verfügbar unter:
http://www.dfg.de/dfg_im_profil/zahlen_und_fakten/statistisches_berichts_wesen/open_access/download/oa_ber_dt.pdf.
- Deutsche Forschungsgemeinschaft (2005b): *DFG-Positionspapier: Elektronisches Publizieren*. Verfügbar unter:
http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/pos_papier_elektron_publizieren_0504.pdf.
- Endres, Jürgen (2005): *Wissenschaftlicher Nachwuchs in Europa*. In: Teichler, Ulrich & Tippelt, Rudolf (Hrsg.): *Hochschullandschaft im Wandel*. 50. Beiheft der Zeitschrift für Pädagogik, S. 158-169.
- Fries, Rüdiger; Kinstler, Till; Schweibenz, Werner; Strobel, Johannes; Weiland, Peter (2001): *Was indexieren Suchmaschinen? Eine Untersuchung zu Indexierungsmechanismen von Suchmaschinen im World Wide Web*. In: B.I.T. online 1/2001, 49-56. Verfügbar unter:
<http://www.b-i-t-online.de/archiv/2001-01/nach2.htm>.
- Fröhlich, Gerhard (2003): *Gegen-Evaluation: Der Impact-Faktor auf dem Prüfstand der Wissenschaftsforschung*. In BUKO INFO 1/4, 61-65. Verfügbar unter: www.oeaw.ac.at/ita/ebene5/Froehlich.pdf.
- Gerdes, Heike (1997): *Lernen mit Hypertext*. Lengerich: Pabst.
- Klatt, Rüdiger; Gavriilidis, Konstantin; Kleinsimlinghaus, Kirsten; Feldmann, Marena u.a. (2001): *Kurzfassung der Endergebnisse des Projektes „Nutzung elektronischer wissenschaftlicher Informationen in der Hochschulausbildung“ (Eine Studie im Auftrag des Bundesministeriums für Bildung und Forschung)*. Verfügbar unter:
http://www.ibw.uni-hamburg.de/bwpat/papers/zu_2/kurzfas_SteFi.pdf.
- Manathunga, Catherine (2003): *Developing interdisciplinary research graduates: educational opportunities and dilemmas*. Paper für die NZARE/AARE Conference 2003. Verfügbar unter:
<http://www.aare.edu.au/03pap/man03133.pdf>.
- Miller-Kipp, Gisela & Neuenhausen, Benedikta (2003): *Erziehungswissenschaft und Internet. Ergebnisse einer Online-Erhebung*. In *Medienpädagogik. Online-Zeitschrift für Theorie und Praxis der Medienbildung*. Ausgabe

- 5.11.2003, Verfügbar unter:
<http://www.medienpaed.com/03-1/miller03-1.pdf>.
- Mruck, Katja; Gradmann, Stefan; Mey, Günter (2004): Open Access: Wissenschaft als Öffentliches Gut. In: Forum Qualitative Sozialforschung / Forum: Qualitative Social Research 5 (2), Art. 14. Verfügbar unter:
<http://www.qualitative-research.net/fqs-texte/2-04/2-04mrucketal-d.htm>.
- Schaffert, Sandra (2004): Kostenlose Online-Literatur der Bildungsforschung. p@psych e-zine, 9. Verfügbar unter:
<http://paedpsych.jku.at/ezine/2004/schaffert04/>.
- Schaffert, Sandra; Schmidt, Bernhard (2004): Inhalt und Konzeption der „bildungsforschung“. In: bildungsforschung 1/1. Abrufbar unter
www.bildungsforschung.org/Archiv/2004-01/einfuehrung.
- Schaffert, Sandra; Schmidt, Bernhard (2005): Zum Themenschwerpunkt „Bildungsforschung: disziplinäre, methodische und empirische Beispiele“. In: bildungsforschung 2/1, URL:
<http://www.bildungsforschung.org/Archiv/2005-01/zumthema/>.
- Schmidt, B. (2004): Virtuelle Lernarrangements für Studienanfänger. Didaktische Gestaltung und Evaluation des Online-Lehrbuchs Jugendforschung. Dissertationsschrift. München: Utz.
- Tippelt, Rudolf (2002): Einleitung des Herausgebers. In: ders. (Hrsg.), Handbuch Bildungsforschung. Opladen: Leske + Budrich, S. 9-20.
- Tippelt, Rudolf & Schmidt, Bernhard (2005): Was wissen wir über Lernen im Unterricht? In: Pädagogik 3/05, 6-11.
- Weichselgartner, Erich (2005): Perspektiven für Information und Kommunikation in der Psychologie: Vortrag gehalten im Rahmen der IuK-Tagung 2005 in Bonn. Verfügbar unter
www.iuk-initiative.org/pdf/sess1_weichselgartner.pdf.
- Winter, Alexander (1998): Arbeiten an und mit Hypertexten. In: Unterrichtswissenschaft 1998/1, 32-50.
- Willinsky, John; Wolfson, Larry (2001): The indexing of scholarly journals: a tipping point for publishing reform? In: The journal of electronic publishing 7/2. Verfügbar unter: <http://www.press.umich.edu/jep/07-02/willinsky.html>.
- Wilson, Brent; Ryder, Martin (1996): Dynamic Learning Communities: An Alternative to Designed Instructional Systems. Verfügbar unter:
<http://carbon.cudenver.edu/~mryder/dlc.html>.

Angaben zu den Autoren

Dr. Bernhard Schmidt
Lehrstuhl für Allgemeine Pädagogik und Bildungsforschung
Ludwig-Maximilians-Universität
Leopoldstr. 13
80802 München
b.schmidt@lmu.de

Studium der Pädagogik mit Abschluss Magister 1996-2000, Promotion zum Dr. Phil. 2004 an der LMU München, Wissenschaftlicher Assistent am Lehrstuhl von Prof. Tippelt seit 2004, Mitherausgeber der Online-Zeitschrift *Bildungsforschung*, Arbeitsschwerpunkte: E-Learning, Erwachsenenbildung, Informelles Lernen, Jugendbildung

Sandra Schaffert
Bildungsinstitut Martin Schön (BIMS)
Bahnhofplatz 4
83278 Traunstein

Jahrgang 1976, Studium der Pädagogik, Psychologie und Informatik an der Ludwig-Maximilians-Universität München mit dem Abschluss M.A., Tätigkeiten unter anderem als wissenschaftliche Mitarbeiterin im Deutschen Jugendinstitut e.V. München und im Bildungsinstitut Martin Schön (BIMS) in Traunstein (www.bims.com und <http://sandra.schaffert.ws>)

Kollaboratives elektronisches Publizieren nach Wiki-Art

Patrick Danowski

Zentral- und Landesbibliothek Berlin

Zusammenfassung

Das Internet ermöglicht der Wissenschaft eine neue Dimension der Kollaboration. Das Bundesministerium für Wissenschaft und Forschung fördert in dem Programm „e-Science“ die Entwicklung neuer Technologien für diese Zusammenarbeit¹. Die Wikipedia hat erfolgreich einen völlig neuen Ansatz des kollaborativen Arbeitens etabliert. Die Frage, die sich stellt, ist, ob sich dieser sehr erfolgreiche Ansatz auch im Rahmen des wissenschaftlichen Schreibens und Publizierens nutzen lässt. Mit dem Modell des kollaborativen elektronischen Publizierens soll eine Diskussionsgrundlage für eine solche neue Form der Zusammenarbeit und des Umdenkens in der Wissenschaft geschaffen werden.

1 Herausforderungen für das wissenschaftliche Publizieren

Die Wikipedia² hat es sehr erfolgreich vor gemacht. Gemeinsam wurde eine Enzyklopädie geschaffen, die immer öfter mit dem Brockhaus verglichen wird, auch wenn es noch Unterschiede gibt, was z.B. die Verlässlichkeit der Informationen angeht. (siehe auch Danowski, Voss 2004) Nicht zuletzt wurde das Konzept der Wikipedia von dem renommierten Grimme Institut bei der Verleihung des Grimme Online Awards von den Junioren als „herausragendes Beispiel kollaborativer Nutzung des Internet“ bezeichnet (Grimme 2005).

Da stellt sich die Frage, ob sich dieses erfolgreiche Verfahren auch für das wissenschaftliche Publizieren nutzen lässt, da Mehrautorenschaften beim wissenschaftlichen Publizieren zunehmen (Umstätter 2004) und gleichzeitig eine Steigerung bei den internationalen Kooperationen zu verzeichnen ist.

Diese Entwicklungen spiegeln sich auch im „DFG-Positionspapier zum elektronischen Publizieren“ (DFG 2005) wider. Die DFG spricht von einer „Verschiebung der Grenzen zwischen traditionellem Publizieren und informeller

1 <http://www.e-science-forum.de>

2 <http://de.wikipedia.org>

Kommunikation“ (DFG 2005, S. 4). Dokumente sollen in der Erstellungsphase und nach Fertigstellung *ergänzbar* sein durch Kommunikationskomponenten. Eine neue Publikationsform der „living documents“ (mit ständiger Aktualisierung) ist dabei, zu entstehen.

Dokumente in Wikis (z.B. in der Wikipedia) haben schon diese Eigenschaften. Kein anderes Nachschlagewerk enthält so schnell und umfassend Informationen über aktuelle Ereignisse wie die Wikipedia (beispielsweise über den neuen Papst und die Flutkatastrophe). Außerdem ist jederzeit nachvollziehbar, wer welche Änderungen vorgenommen hat.

2 Zusammenarbeiten beim Erstellen wissenschaftlicher Texte

Zunächst möchte ich theoretisch betrachten, wie Wissenschaftler zusammenarbeiten. Es existieren zwar viele Publikationen zum wissenschaftlichen Arbeiten, jedoch wird in diesen nur wenig dazu gesagt, wie man gemeinsam mit anderen eine wissenschaftliche Publikation verfassen kann. Aus diesem Grund habe ich zwei theoretische Modelle entwickelt, wie eine solche Kooperation ablaufen könnte.

2.1 Lokale Gruppe

Das erste Modell stellt dar, wie Wissenschaftler, die gemeinsam an einem Institut sind, zusammen einen Text verfassen könnten. Dieses Modell möchte ich als „Lokale Gruppe“ bezeichnen.

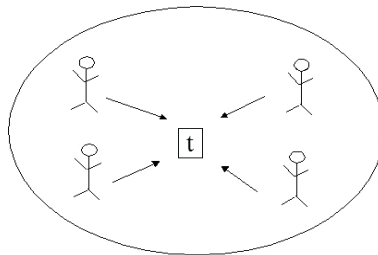


Abb. 1: Lokale Gruppe

Die Wissenschaftler können ohne große Probleme an einem Dokument, das sich in einer Datei befindet, zusammenarbeiten. Dies bedeutet zwar rein praktisch, dass nur einer zur Zeit an diesem Dokument arbeiten kann; da aber die Kommu-

nikation sehr direkt erfolgen kann, können sich die Wissenschaftler im Konfliktfall (also wenn zwei oder mehrere gleichzeitig das Dokument bearbeiten möchten) in einem Büro treffen und direkt gemeinsam arbeiten. Nachteil dieser Methode ist, dass sie ortsgebunden ist und Teams oft weltweit verteilt sind. Gleichzeitig könnte es passieren, dass diese Methode langsamer ist, da ein wirkliches gleichzeitiges Arbeiten nicht stattfinden kann, da Terminabsprachen unter den Mitgliedern zu einem gemeinsamen Treffen notwendig sind.

2.2 Die Puzzle Methode

Das zweite Modell möchte ich als die „Puzzle-Methode“ bezeichnen. Hier teilen die Wissenschaftler das Thema zunächst untereinander auf - hierzu erstellen sie beispielsweise zunächst eine Gliederung, die dann aufgeteilt wird. Jeder schreibt seinen Teil des Textes. Nachdem jeder seinen Teil fertig gestellt hat, werden die Teile zusammengeführt und wie ein Puzzle zusammengesetzt. Eventuell wird der gesamte Text noch einmal bearbeitet, um eine Interaktion, wie zum Beispiel Verweise zwischen den verschiedenen Teilen, herzustellen.

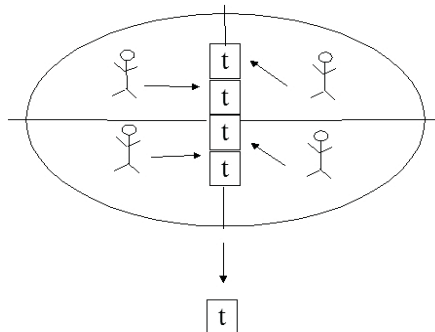


Abb. 2: Die Puzzle-Methode

Hier liegt auch einer der Nachteile dieser Methode. Gleichzeitiges Arbeiten verursacht hier eine starke Abgrenzung zwischen den einzelnen Teilen und somit findet keine große Interaktion zwischen den einzelnen Teilen statt, da der Inhalt der anderen Teile den anderen Autoren nicht bekannt ist. Außerdem eignet sich diese Methode nur für eine begrenzte Anzahl von Autoren (abhängig wie stark sich ein Text aufteilen lässt). Bei Themen, bei denen ein Aufteilen nur schwer möglich ist, eignet sich diese Methode überhaupt nicht.

3 Exkurs: Evolution der Wissenschaftlichen Kommunikation

Als kleinen Exkurs schauen wir uns an, inwieweit sich die wissenschaftliche Kommunikation verändert hat. Diese funktionierte zunächst hauptsächlich über Briefe, die von einem Wissenschaftler an einen anderen Wissenschaftler geschrieben wurden. Erst im 18. Jh. kamen wissenschaftliche Journale hinzu, die anfangs aus veröffentlichten Briefen bestanden, so dass sich die Kommunikation von einer eins-zu-eins- zu einer eins-zu-vielen-Kommunikation verändert hatte. Die Anzahl der Leser war noch stark abhängig von der Auflage oder dem Vorhandensein in einer bestimmten Bibliothek, so dass die Anzahl der Leser mit Steigen der Auflage zunahm, jedoch Beschränkungen unterlag. Gleiches gilt für die Anzahl der Autoren, die zwar im Laufe der Zeit wuchs, jedoch ebenfalls durch die Arbeitsmethode gewisse Beschränkungen aufwies.

Weitere radikale Änderungen brachte das Aufkommen von elektronischen Texten insbesondere im Zusammenhang mit dem Entstehen des Internets. Durch dieses war es möglich, den Text von überall, wo ein Computer an das Internet angeschlossen war, zu betrachten, so dass eine wesentliche Beschränkung für die Leser verschwand. Dies erhöhte die potenzielle Leserschaft und Zugänglichkeit eines Textes um ein Vielfaches, obwohl dadurch natürlich auch neue Schranken und Probleme aufgebaut wurden.³ Die Grenze in Bezug auf die begrenzte Anzahl von Autoren blieb jedoch bestehen, da es keine wesentlichen Änderungen im Erstellungsprozess der Texte gab, diese wurde durch die neuen Kommunikationswege „nur“ beschleunigt.

Eine Revolution in diesem Bereich brachte das Wiki, das im Jahre 1995 von Ward Cunningham erfunden wurde. In einer E-Mail an einen gewissen Steve schrieb er: „Steve - ich habe eine neue Datenbank auf meinem Web-Server installiert und bitte Dich, mal einen Blick darauf zu werfen. Es ist ein Web von Menschen, Projekten und Mustern, auf das man über ein cgi-bin-Skript zugreifen kann. Es bietet die Möglichkeit, ohne HTML-Kenntnisse mit Formularen Text zu editieren. Es wäre schön, wenn Du mitmachen oder wenigstens Deinen Namen in der Liste der RecentVisitors eintragen könntest. Die URL ist <http://c2.com/cgi-bin/wiki> - danke schön und beste Grüße.“⁴

Wikis machten es möglich, dass jeder Leser dem Text etwas hinzufügen, etwas ändern oder etwas löschen kann. Somit ist es möglich, dass ein einziger Text deutlich mehr Autoren als bisher haben kann, da jeder Leser gleichzeitig potenzieller Autor ist (eine Zusammenfassung des Ganzen nochmals in Tabelle 1).

3 Dies wird beispielsweise unter der Fragestellung des „digital and divides“ diskutiert. Diese Problematik wird unter anderen in (Kuhlen 2004) angesprochen.

4 <http://c2.com/wiki/mail-history.txt>. Übersetzung von Erik Möller

Tabelle 1: Übersicht der verschiedenen Publikationsarten (n steht für eine beschränkte endliche Zahl, x und y sind beliebig steigerbar)

	Autoren	Vermittler (Beispiel)	Leser	Besonderheiten
Brief	1	-	1	Direkt
Journal / Buch	n	Verlag	n	
elektronischer Text	n	Internet	x	
Wiki	y	Internet	x	Leser = möglicher Autor

4 Das kollaborative Publizieren

Nun komme ich zu der Theorie des kollaborativen Schreibens. Kollaborativ steht hier im Sinne von gemeinsam, nicht im Sinne des Kollaborateurs (siehe auch Kuhlen 2004). Gekennzeichnet ist es von drei Faktoren:

1. Eine Gruppe schreibt an *einem* Text
2. Somit existiert nur eine aktuelle Version des Textes (Versionskontrolle)
3. Gemeinsames Arbeiten von mehreren Personen an einem Text zur selben Zeit bzw. sehr zeitnah

Aus der Unterscheidung des letzten Punktes kann man zwei unterschiedliche Arten ableiten analog zu den Unterscheidungen der Kommunikationsarten (verwendet z.B. in (Nentwich 2003, S. 56)):

A) Synchrone Kollaboration - Zusammenarbeiten in Echtzeit

Zur Zeit sind Tools, die ein gemeinsames Arbeiten in Echtzeit unterstützen, erst am Entstehen, obwohl bereits im Jahre 1968 von Douglas Engelbart in einer Demo⁵ eine solche Funktionalität gezeigt wurde. Durch das Internet wurde diese Methode „wiederentdeckt“ und z.B. durch das Tool Moonedit⁶ realisiert.

B) Asynchrone Kollaboration – zeitnahes/zeitversetztes Zusammenarbeiten

Dies geht z.B. mit dem oben beschriebenen Wiki. Das bekannteste Projekt, dass dieses Tool erfolgreich einsetzt, ist die Wikipedia. Hier werden verschiedene Artikel einer Enzyklopädie von vielen Autoren gemeinsam erstellt. Wie dies geschieht, soll im Folgenden etwas genauer betrachtet werden.

5 Die Demo befindet sich vollständig im Netz:

<http://sloan.stanford.edu/mousesite/1968Demo.html>

6 <http://moonedit.com>

4.1 Schreiben in der Wikipedia

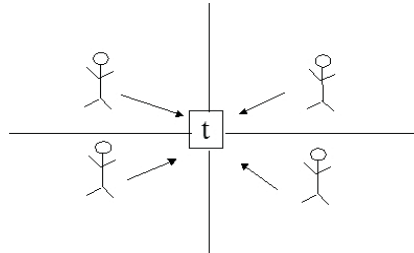


Abb. 4: Schreiben in der Wikipedia

Das Schreiben in der Wikipedia basiert auf dem „Prinzip der maximalen Offenheit“ (Erik Möller in: Schlieker, Lehmann 2005, S. 259). Dies bedeutet, dass jeder, der eine Seite betrachtet, diese auch ändern und somit durch sein Wissen Inhalte ergänzen kann und so in der Lage ist, einen Artikel zu verbessern. Somit kann jeder, der dies möchte, zum Autor werden. Aus diesem Grund hat Abbildung 3 auch keine Grenze nach außen. Die immer aktuelle Fassung des Textes liegt auf einem Server, auf den jeder zugreifen kann.

Nur sehr selten kommt es zu Bearbeitungskonflikten, bei denen zwei Autoren den selben Abschnitt eines Textes ändern möchten. Eine Endfassung gibt es nicht, sondern nur eine Fassung, die den aktuellen Stand wiedergibt. Dies ist besonders schwierig beim Zitieren⁷.

4.2 Kollaboratives elektronisches Publizieren

Nun versuche ich, drei wichtige Elemente zusammenzubringen. Da sind der Ansatz des kollaborativen Arbeitens, die Anforderungen des wissenschaftlichen Publizierens und der Trend des elektronischen Publizierens und somit das zur Verfügung stellen der Publikationen in elektronischer Form.

Die Kernfrage ist dabei, wie das Modell der Wikipedia abgewandelt werden kann, um es für wissenschaftliche Autoren und deren Publikationen attraktiver zu machen. Nun dazu möchte ich zwei Elemente der ersten beiden Modelle übernehmen. Zum einen soll die Offenheit ein wenig eingeschränkt werden, aber nur soweit, wie nötig. Zum anderen soll erreicht werden, dass eine endgültige Version der Dokumente existiert.

⁷ Die neue Version der MediaWiki Software (Version 1.5), mit dem die Wikipedia arbeitet, unterstützt nun auch persistente URL's für eine bestimmte Version, die das Zitieren eines bestimmten Standes eines Artikels zu einem bestimmten Zeitpunkt deutlich vereinfacht.

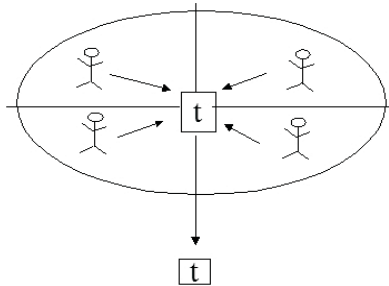


Abb. 5: Modell des kollaborativen elektronischen Publizierens

An genau diesem Problem arbeitet momentan auch die Wikipedia Community, an einer so genannten stabilen Version der Wikipedia oder auch Wikipedia 1.0 wie Jimmy Wales (Jüngling 2005) es genannt hat.

Die entscheidende Frage ist, wie man die Gruppe zum einen begrenzt, zum anderen aber möglichst offen hält, so dass auch neue Ideen von außerhalb der Gruppe einfließen können. Dazu schaut man sich an, wie Wissenschaftler im Allgemeinen ihre Mitautoren auswählen. Sie suchen sich Kollegen, denen sie fachlich vertrauen. Dies bedeutet, sie wählen Kollegen aus, die sie für kompetent halten. Genauer formuliert, sie schreiben mit Kollegen zusammen, die sie auf dem Gebiet, in dem der Text geschrieben werden soll, für kompetent halten. Eventuell würden sie sogar (dies wäre zu untersuchen) mit einem Kollegen zusammenarbeiten, der ihnen von einem anderen Kollegen, den sie selbst für kompetent halten, empfohlen wird. Sollte dies so sein, dann könnte man das ganze, ähnlich wie das „Web of Trust“ in der Kryptographie, modellieren. Analog könnte ein „Web of Competence“ gebildet werden, das jedoch einen wesentlichen Unterschied aufweisen müsste. Bei „Web of Trust“⁸ lautet die Schlussfolgerung:

- A vertraut B
- B vertraut C
- -> A kann auch C vertrauen.

Dies reichte bei einem „Web of Competence“ noch nicht aus, hier kommt noch ein Faktor hinzu:

- A hält B für kompetent in Bezug auf elektronisches Publizieren
- B hält C für kompetent in Bezug auf Urheberrecht
- B hält D für kompetent in Bezug auf elektronisches Publizieren

8 siehe auch den Wikipedia-Artikel zu Web of Trust:
http://de.wikipedia.org/wiki/Web_of_trust

Nur wenn die Bezugsgröße gleich ist lassen sich Verbindungen ziehen:

- -> A kann auch D für kompetent in Bezug auf elektronisches Publizieren halten
- aber: A kann nicht C für kompetent in Bezug auf das elektronische Publizieren halten.

Die grundlegenden Fragen, die sich nun stellen, sind: Wie beginnt man so ein Netzwerk? Und wie definiert man Kompetenz? Eine Möglichkeit wäre, das Ganze in einem „social tagging“-Verfahren (Dudeck, Voß 2005) zu realisieren, wobei jeder neue Tags (wobei jedes Tag den Namen einer Forschungsgruppe darstellt) generieren kann, bereits vergebene Tags jedoch nur von Personen, die diese schon erhalten haben, weiter vererbt werden können. Eine Reihe von Basistags, wie beispielsweise die institutionalisierten Fachrichtungen, können durch Nachweis eines entsprechenden Abschlusses erworben werden. So dass es nach einer gewissen Anlaufzeit aufgrund der Beiträge eines Autors und seinen erhaltenen Basistags möglich ist, sich ein Meinungsbild zu bilden und auch nur virtuell bekannte Autoren in eine virtuelle Forschungsgruppe aufzunehmen.

4.3 Kollaboratives Arbeiten mit Sozialer Software

Im Moment gibt es noch kein Tool, das den ganzen Prozess des Schreibens, von der Recherche über die Literaturverwaltung bis zum geschriebenen Text, unterstützt. Jedoch gibt es schon erste Möglichkeiten des kollaborativen Arbeitens im Bereich der Recherche und dem Erstellen von Literaturlisten.

Das Tool del.icio.us⁹ ermöglicht das gemeinsame Sammeln und Verwalten von Bookmarks. Verschiedene Personen können Links sammeln und diese mit frei wählbaren Stichwörtern so genannten Tags versehen. Zusätzlich ist es möglich, sich anzeigen zu lassen welche anderen Links mit den selben Tags von anderen versehen wurden. Des Weiteren kann man sehen, wer diese Seite noch gespeichert hat, um dann die Linklisten von anderen anzuschauen, die eventuell gleiche Interessen haben und so an weitere für einen selbst interessante Seiten zu kommen. Außerdem ist es möglich, alle neuen Links, die unter einem bestimmten Tag oder die von einer Person gespeichert worden sind, als RSS Feed¹⁰ zu abonnieren.

Vom Ansatz das selbe System, nur für wissenschaftliche Artikel und Bücher, verwendet [citeUlike](http://citeulike.org)¹¹. Die Suche kann in verschiedenen Datenbanken wie beispielsweise PubMed oder Amazon erfolgen. Hier ist es auch möglich, Gruppen zu beliebigen Themen zu bilden.

9 <http://del.icio.us>

10 Erläuterung zu RSS Feeds unter: <http://de.wikipedia.org/wiki/Rss>

11 <http://citeulike.org>

4.4 Hindernisse und Widerstände

Elektronisches Publizieren hat sich in der Wissenschaft noch nicht überall etabliert¹² und wird in einigen Fachbereichen noch sehr kritisch betrachtet. Da ist es selbstverständlich, dass ein neuer Weg des Publizierens, der weitere Änderungen mit sich bringt, auf Kritik und Widerstände stoßen wird. Doch wo könnten mögliche Ursachen für die Ablehnung liegen? Drei mögliche Ursachen sind:

1. Das westliche Geniebild: (Kuhlen 2004)
Historisch gewachsen in der westlichen Welt ist das Bild, dass am besten eine Person, ein Genie, für eine bestimmte Leistung in der Wissenschaft verantwortlich ist und somit auch dieser Person der Ruhm und die Ehre zugerechnet wird. Dies wirkt sich unmittelbar auf den zweiten Punkt aus.
2. Die Ausrichtung der Wissenschaft an (Print-) Publikationen
Beispielsweise bei der Leistungsbewertung eines einzelnen Wissenschaftlers in einem Berufungsverfahren. Dieser Punkt stellt schon für das elektronische Publizieren ein Hindernis dar. Für das kollaborative Publizieren ist das Problem noch gravierender, da das bisherige System davon ausgeht, dass eine Publikation nur von einigen wenigen geschrieben wird. Ein Entstehen, wie es das Modell des kollaborativen elektronischen Publizierens vorsieht, würde ein radikales Umdenken erfordern, wie es auch das Bundesministerium für Wissenschaft und Forschung in dem Programm zum e-Science¹³ fordert: „Erforderlich sind allerdings nicht nur technische und neue informationsmethodische Lösungen, sondern auch neue, kooperative Organisationsformen und eine neue Einstellung zum kooperativen Arbeiten, die in der Wissenschaft bislang noch wenig entwickelt sind und die über entsprechende Maßnahmen der Vertrauensbildung unterstützt werden müssen.“
3. Die Angst vor Plagiaten
Dieser weitere mögliche Einwand lässt sich ebenfalls aus dem ersten Problem ableiten. Denn die eigenen Texte und somit auch die eigene relevante Leistung für den Bewertungsprozess kann eventuell mehr korrekt der eigenen Person zugeordnet werden, wenn jemand anders den eigenen Ansatz fortführt und eventuell schneller zu der entsprechenden Lösung gelangt. Dies ist jedoch in diesem Modell nur bedingt richtig, denn da genau erfasst wird, wer wann was geschrieben, hat lässt sich viel leichter nachvollziehen, wer zuerst die bahnbrechende Idee gehabt hat.

12 siehe dazu auch verschiedene Beiträge in diesem Band

13 <http://www.bmbf.de/foerderungen/3179.php>

Sicher gibt es weitere Faktoren, die gegen dieses Modell zu sprechen scheinen, doch dies war vor ca. 20 Jahren bei den ersten elektronischen Publikationen nicht anders.

5 Ausblick

„Within a few years time, the scholarly publication system will be fundamentally different from what we came to know“ (Nentwich 2003, S. 366).

Der Erfolg der Wikipedia mit dem Modell des kollaborativen Arbeitens wird in einer immer breiteren Öffentlichkeit wahrgenommen und so überlegen auch verschiedene Wissenschaftler (z.B. Lambert, Heller¹⁴; Gerry McKiernan¹⁵), wie man diese Methode für das wissenschaftliche Schreiben nutzen kann. Hier zu behaupten, dass sich bereits ein erstes Umdenken in der Wissenschaft zeigt, wäre übertrieben, jedoch ist schon interessant, dass sich eine Idee völlig unabhängig voneinander entwickelt. In der Wikipedia Community wird unter dem Stichwort Wikiversity¹⁶ unter anderem über das kollaborative Erstellen von E-Learning-Inhalten diskutiert, aber auch das kollaborative Forsuchen wird unter diesem Label diskutiert.

Insbesondere da das Peer Review immer weiter unter Kritik gerät, ist über alternative Wege des Publizierens nachzudenken. Das Umdenken in Richtung elektronisches Publizieren und Open Access bildet hier einen ersten Schritt; eventuell stellen kollaboratives elektronisches Publizieren und freie Lizenzen einen weiteren dar. Momentan fehlt es noch an praktischen Implementierungen. Zu ermitteln, z.B. im Rahmen einer Studie, wäre, inwieweit Zusammenarbeiten in der Wissenschaft wirklich funktioniert und was weitere Rahmenbedingungen wären.

Literatur

Deutsche Forschungsgemeinschaft (DFG) (Hrsg.): Positionspapier zum elektronischen Publizieren

http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/pos_papier_elektron_publizieren_0504.pdf

14 Vortrag zu dem Thema „Wikis for scientific publishing“ auf der Wikimania 05:

<http://en.wikibooks.org/w/index.php?title=Wikimania05/LH1>

15 Blog zum Thema „Disruptive Scholarship“: <http://disruptivescholarship.blogspot.com/>

16 <http://www.wikiversity.de>

- Dudeck, Jochen; Voß, Jakob 2005: *Kooperation als wichtigster Bestandteil des Konzepts/Weblogs, Wikis & Co.: Social Software in Bibliotheken*. In: Buch und Bibliothek 3/2005
- Presseerklärung des Grimme Instituts 2005:
http://grimme-institut.de/event/goa/downloads/pressemitteilungen/pdf/PMGewinner_des_GOA_2005.pdf
- Jüngling, Thomas 2005: Die Erweiterung des globalen Wissens. In: Welt am Sonntag, 23.01.05
- Kuhlen, Rainer 2004: Kollaboratives Schreiben. In: C. Bieber; C. Leggewie (Hrsg.): *Interaktivität ein transdisziplinärer Schlüsselbegriff*. Frankfurt: Campus-Verlag. S. 216-239 online verfügbar:
http://www.inf-wiss.uni-konstanz.de/People/RK/Publikationen/2004/20040706_autoren_kollaborateure.pdf
- Nentwich, Michael 2003: *Cyberscience*. Wien: Austrian Academy of Science Press
- Schlieker, Christan; Lehmann, Kai; Möller, Erik 2005: Verknüpft, Verknüpfter, Wikis. In: Lehmann, Kai; Schetsche, Michael (Hrsg.): *Die Google-Gesellschaft. Vom digitalen Wandel des Wissens*. Bielefeld: transcript 2005.
- Umstätter, Walther 2004: Szientometrische Verfahren. In: Kuhlen, Rainer; Thomas Seeger; Dietmar Strauch (Hrsg.): *Grundlagen der praktischen Information und Dokumentation*. 5. Auflage. München : K. G. Saur S. 237-244.
- Voß, Jakob; Danowski, Patrick 2004: *Bibliothek, Information und Dokumentation in der Wikipedia*. In: *Information: Wissenschaft und Praxis* 8/2004 online verfügbar unter: <http://eprints.rclis.org/archive/00002566/>

Weiterführende Literatur

- Forschungsweblog zum kollaborativen elektronischen Publizieren:
<http://koelpu.twoday.net>
- Linkliste: <http://del.icio.us/tag/koelpu>
- Literaturliste: <http://de.citeulike.org/tag/koelpu>

Angaben zum Autor

Patrick Danowski
Zentral- und Landesbibliothek Berlin
Breite Str. 30-36
Berlin

Patrick Danowski ist Diplominformatiker und zurzeit Bibliotheksreferendar an der Zentral- und Landesbibliothek Berlin. Er hat an der Technischen Universität Berlin Informatik mit dem Schwerpunkt Informatik und Gesellschaft studiert. Seine fachlichen Schwerpunkte sind elektronisches Publizieren, Open Access und Open Content, Digitalisate und Digitalisierung, RFID in Bibliotheken sowie Wikis und die Wikipedia. Seit August 2005 ist er Mitglied im Vorstand von Wikimedia Deutschland e.V.

In die Zukunft über die Vergangenheit publizieren – Das Historische Lexikon der Schweiz (HLS)

Marco Jorio

Chefredaktor Historisches Lexikon der Schweiz

Zusammenfassung

Das 1988 von der Schweizerischen Akademie der Sozial- und Geisteswissenschaften ins Leben gerufene Historische Lexikon der Schweiz (HLS) ist das zur Zeit größte geistes- und sozialwissenschaftliche Projekt der Schweiz. Es arbeitet als historische Fachencyklopädie in rund 36.000 Artikeln die gesamte Schweizer Geschichte von der Ur- und Frühgeschichte bis zur Gegenwart auf. Basierend auf dem heutigen Forschungsstand (und zum Teil auf Grund von eigenen Forschungen) werden rund 26.000 Biographien, 4.000 Ortsartikel, 4.000 Sachartikel und 2.000 Familienartikel verfasst. Der Gesamtumfang des von einer Stiftung herausgegebenen und von der Eidgenossenschaft finanzierten Lexikons umfasst rund 60 Mio. Zeichen. Das Lexikon erscheint gleichzeitig in den drei Landessprachen deutsch, französisch und italienisch, d.h. alle Artikel erscheinen in allen drei Sprachen mit gleichem Inhalt. Damit ist das HLS unseres Wissens das weltweit größte und einzige mehrsprachige wissenschaftliche Lexikon.

Das HLS erscheint gleichzeitig in einer gedruckten und in einer elektronischen Version. Von den 12-13 gedruckten Bänden je Sprachausgabe sind bis heute die ersten vier Bände auf deutsch, französisch und italienisch (d.h. total 12 Bücher zu je 850 Seiten) publiziert. Im Herbst 2006 wird der 5. Band erscheinen. Daneben betreibt die Zentralredaktion in Bern eine elektronische Publikation, das sog. e-HLS, das der Scientific Community und der interessierten Öffentlichkeit die fertig redigierten Artikel im Web unentgeltlich zur Verfügung stellt (www.hls.ch). Zur Zeit sind rund 48.000 Artikel in den drei Sprachen abrufbar, was ungefähr dem Volumen von 14 gedruckten Bänden entspricht. Im Gegensatz zur gedruckten Ausgabe ist das e-HLS (noch) nicht illustriert. Der Verbund eines kommerziellen Printmediums (Verlag Schwabe AG, Basel) und einer frei zugänglichen Web-Publikation war und ist nicht ganz unproblematisch. Im Moment planen wir die Weiterführung des Lexikons in Form von Datenbanken sowie den Umbau des Unternehmens in ein historisches Informationszentrum für die Zeit nach Erscheinen des letzten Bandes.

1 Die Schweiz und ihre historischen Enzyklopädien

Ungefähr in jedem Saeculum machen sich die Schweizer Historikerinnen und Historiker auf, um das akkumulierte Wissen über die Vergangenheit des Landes in einer umfangreichen Enzyklopädie oder in einem Lexikon zusammenzufassen [1]. Neben der Aufgabe der Wissensvermittlung hatten und haben diese Publikationen auch immer wieder die Funktion der nationalen Selbstvergewisserung. Die Schweiz, die weder durch eine gemeinsame Sprache, noch durch eine gleiche Religion oder Konfession, noch durch eine geographische Einheit zusammen gehalten wird, stützt sich seit jeher auf die gemeinsame Geschichte und gleiche politische und gesellschaftliche Werte. Die historischen Lexika der Schweiz hatten und haben somit immer auch eine politische und eine identitätsstiftende Dimension.

Das erste derartige Werk gab der Zürcher Ratsherr und Bankier Johann Jakob Leu von 1747 bis 1765 in 20 Bänden unter dem Titel „Allgemeines Helvetisches, Eydgenössisches oder Schweitzerisches Lexicon“ heraus. Es erschien nur in deutscher Sprache. Von 1921 bis 1934 publizierte dann der Württemberger Immigrant Victor Attinger in Neuchâtel in 7 Bänden das „Historisch-Biographische Lexikon der Schweiz“ (HBL), das parallel auch auf Französisch unter dem Titel „Dictionnaire Historique et Biographique de la Suisse“ (DHBS) erschien.

Im Vorfeld zur 700-Jahrfeier von 1991 wurde in den frühen 1980er Jahren der Ruf nach einer neuen umfassenden Enzyklopädie zur Schweizer Geschichte laut, die vor allem auch die wissenschaftlichen Erkenntnisse der letzten Jahrzehnte, namentlich der neuen historiographischen Methoden und Fragestellungen wie etwa die Sozialgeschichte, Wirtschaftsgeschichte, Mentalitätsgeschichte lexikographisch aufbereitet einem breiten Publikum zur Verfügung stellen soll. Die Schweizerische Akademie der Geistes- und Sozialwissenschaften (SAGW) und die Schweizerische Gesellschaft für Geschichte (SGG) arbeiteten ein Projekt aus, das die Bundesbehörden 1987 genehmigten und seither gestützt auf das Forschungsgesetz mit rund 5 Mio. Fr. oder 3.3 Mio. Euro jährlich finanzieren. Es ist zur Zeit das größte geisteswissenschaftliche Einzelprojekt der Schweiz. Trägerschaft ist eine 1988 eigens dafür errichtete Stiftung [2].

2 Konzept und Publikation

Das Historische Lexikon der Schweiz arbeitet als historische Fachencyklopädie in rund 36.000 Artikeln – die Stichwortliste in allen drei Sprachen steht – die gesamte Schweizer Geschichte von der Ur- und Frühgeschichte bis zur Gegenwart auf. Vorgesehen ist die Erarbeitung von rund 26.000 Biographien von Schweizerinnen und Schweizern, die im In- und Ausland gewirkt haben, aber auch von

Ausländern, die in der Schweiz gelebt haben (z. B. Lenin) oder direkt auf die Schweiz eingewirkt haben (z. B. Adolf Hitler). Dann sollen 4.000 Ortsartikel (Kantone, Gemeinden, Seen, Flüsse, Länder, archäologische Fundorte etc.), 4.000 Sachartikel zu allen Bereichen und 2.000 Familienartikel publiziert werden. Das Werk soll in drei identischen Sprachausgaben auf deutsch, auf französisch unter dem Namen *Dictionnaire Historique de la Suisse* (DHS) und (zum erstenmal) auch auf italienisch unter dem Titel *Dizionario Storico della Svizzera* (DSS) zu je ungefähr 12-13 Bänden, also total ungefähr 36-39 Bänden erscheinen. Für die kleine rätoromanische Gemeinschaft ist eine einbändige Teilausgabe, das *Lexikon istoric retic* (LIR), in Arbeit. Zudem kooperieren wir eng mit dem neuen Historischen Lexikon des Fürstentums Liechtenstein (HLFL) zusammen. Seit 1990 arbeitet die 30-köpfige Redaktion in Bern, mit Filialen in der italienischen und rätoromanischen Schweiz, zusammen mit rund 2.000 Autorinnen und Autoren, 100 wissenschaftlichen Beratern und gegen 100 Übersetzern.

Sie haben sicher bemerkt: Die Angaben zum Umfang des HLS erfolgten in Bänden. Tatsächlich war im Konzept von 1987 vorgesehen, in erster Priorität auf ein gedrucktes Werk zuzusteuern. Allerdings war ebenfalls vorgesehen, in einer späteren Phase eine elektronische Ausgabe anzubieten. Deshalb wurde von Anfang an das HLS elektronisch erarbeitet. 1990 entschieden wir, dass dabei die ISO-Norm SGML angewendet werden soll. Wir wussten damals nicht, dass diese Norm die „Mutter“ von HTML und später XML werden sollte, womit sich dieser Entscheid als goldrichtig erwies.

Gerade, als die Produktion der Artikel so richtig begann, erwischte uns die „elektronische Revolution“: zuerst kam die CD-ROM, dann das Internet. Um 1995 begann eine epische Diskussion über elektronische versus gedruckte Publikation: Verzicht auf das Buch („Verrat an der abendländischen Buchkultur“: schrien die einen), Verbot ein e-HLS zu entwickeln („Rückschritt ins Mittelalter“: meinten die anderen). Schließlich entschied sich der Stiftungsrat 1997 nach langen Abklärungen für eine elektronische Publikation im Internet, die zeitlich vorgezogen wurde und für eine später folgende Druckversion. Die Publikation als CD-ROM wurde bereits damals als überholt beurteilt und den drei Verlegern überlassen. 1998 ging das e-HLS als (meines Wissens) weltweit erstes mehrsprachiges Lexikon mit 8.000 Artikeln in allen drei Sprachen ans Netz. Im Sinne eines *Service public* war und ist der Zugang zu den fertig bearbeiteten Artikeln seither gratis, aber aus verlegerischen Gründen ohne Bebilderung. Heute sind von den vorgesehenen 108.000 Artikel (3 mal 36.000 Artikel) rund 48.000 Artikel unentgeltlich zugänglich. Neben dem öffentlichen e-HLS betreiben wir für die Mitarbeiter noch eine passwort-geschützte Datenbank mit knapp 10.000 Originalartikeln, die erst in der Autorenversion vorhanden sind und von der Redaktion noch nicht bearbeitet sind.

Seit 2002 erscheinen nun auch die gedruckten Bände. Bisher sind vier Bände in jeder Sprache, also 12 Bücher, erschienen. Am 14. Oktober 2005 haben wir

den 4. Band, d. h. 3 weitere Bücher, der Öffentlichkeit übergeben. Es ist vorgesehen, sofern die finanziellen Mittel das erlauben, jedes Jahr einen Band (3 Bücher) zu produzieren und das e-HLS sukzessive zu füllen. 2012/13 soll so das gedruckte HLS vorliegen.

3 Das HLS zwischen Gutenberg und electronic publishing

Mit dem Medienverbund e-HLS und Bücher ist das HLS neue Wege gegangen. Warum diese Lösung? Warum überhaupt noch ein Buch? Nach langen Diskussionen national und international (Umberto Eco hat da schon Mitte der 1990er Jahre interessante Impulse gegeben) halten wir die beiden Datenträger nicht für Konkurrenten, sondern für Ergänzungen. Jedes Medium hat seine ihm spezifischen Stärken und Schwächen.

Die Stärken des Buches (und damit die Schwächen der e-Publikation) sehen wir in folgenden Punkten:

- Langlebigkeit und Nachhaltigkeit: Das gedruckte Buch wird voraussichtlich noch in mehreren Jahrhunderten zur Verfügung stehen – irgendwo. Die hohen Investitionen werden über Jahrzehnte, ja Jahrhunderte gesichert und amortisiert. Das Buch hat damit eine Archivfunktion. Die elektronische Publikation dagegen ist eigentlich für die heute lebende Generation nutzbar. Die Nachhaltigkeit ist (noch) nicht gewährleistet – technisch, personell, institutionell (denken wir nur an die verlassenen, aufgegebenen Sites!). Eine Datenbank muss dauernd unterhalten werden, ein Buch nicht! Oder provokativ formuliert: Datenbanken sind Wegwerfprodukte oder Datenspeicher auf Zeit, Bücher für die Ewigkeit.
- Das Buch hat den Vorteil, dass das Wissen einer Generation fixiert wird. Es hat damit auch die Funktion der Wissenssicherung. Die elektronische Publikation ist dauernd im Fluss.
- Für eine ganze Reihe von Abfragen ist das Buch der elektronischen Publikation überlegen: so z. B. für kurze Abfragen oder die Lektüre von langen Artikeln. Das Buch braucht kein Aufstarten, kein Öffnen, kein Suchen.
- Das Buch braucht keine technischen Hilfsmittel, die irgendwann einmal nicht mehr zur Verfügung stehen und ist damit jederzeit durch jedermann durch den angeborenen Sehnsinn zu benützen.
- Ein großer Teil der Informationsvermittlung in den Geisteswissenschaften läuft für einen großen Teil der Kundschaft heute immer noch über das Buch.
- Das Buch ist ein greifbares Kulturprodukt unserer abendländischen Zivilisation; das e-HLS ist virtuell und wird z. T. selbst von Internauten nicht als vollwertige Publikation wahrgenommen.

Die elektronische Publikation ihrerseits hat aber ihre unbestreitbaren Vorteile:

- Man kann anders recherchieren (z. B. Volltext-Suche) und sie mit anderen Datenbanken verlinken.
- Die strenge alphabetische Ordnung kann aufgehoben werden und über Suchkriterien die systematische wieder hergestellt werden, was dem vernetzten Zugang zur Geschichte besser entspricht.
- Sie kann dauernd dem neuen Forschungsstand angepasst werden.
- Sie ist wesentlich billiger.
- Sie hat keine Platzrestriktionen. Beim Buchdruck werden die Zeilen gezählt. Die redaktionellen Kürzungen sind zeitintensiv und kostspielig.

Wenn wir nun das gedruckte HLS betrachten, erfüllt es alle Anforderungen an das Buch (gutes, säurefreies Papier, haltbarer Einband) und nutzt dessen Möglichkeiten. Das heutige e-HLS hat aber noch Defizite: Die Volltextsuche ist noch ungenügend, die Abkürzungen und das Zeilenzählen sind eigentlich überflüssig, Links auf Datenbanken außerhalb des e-HLS fehlen noch. Das heutige e-HLS ist die elektronische Version eines gedruckten Werkes und muss noch aus- und umgebaut werden. Zurzeit läuft ein Projekt, das bis Ende 2005 das e-HLS modernisieren und verbessern wird.

Die Stiftung HLS hat zwei Stiftungszwecke: die Herausgabe des Lexikons und die Weiterführung in Form von Datenbanken. Tatsächlich sind wir auf dem Weg dazu. Ab Herbst 2005 werden wir die bereits gedruckten Artikel im e-HLS korrigieren und aktualisieren. Rund um das e-HLS haben sich eine Reihe von Datenbanken gruppiert, die z. T. öffentlich, z. T. noch nicht öffentlich sind (Familiennamenbuch, Glossare, Historikerdatenbank, das elektronische LIR). Zur Zeit sind wir daran abzuklären, ob und wie wir das seit Jahrzehnten geforderte Informations- und Dokumentationszentrum zur Schweizer Geschichte realisieren können.

4 Literaturangaben

- [1] Jorio, M. (2004): Zur Geschichte der Enzyklopädien und Lexiken in der Schweiz seit dem 17. Jahrhundert. In: Stamm, Th.; Weber, W. E.J. (Hrsg.): *Colloquia Augustana*, Band 18. Wissenssicherung, Wissensordnung und Wissensverarbeitung. Das europäische Modell der Enzyklopädien. 2004, Berlin: Akademie Verlag, S. 105 – 117.
- [2] *Historisches Lexikon der Schweiz*. Band 1. Stiftung Historisches Lexikon der Schweiz (Hrsg.). 2002, Basel: Schwabe-Verlag, S. V-XXX.

Angaben zum Autor

Dr. Marco Jorio
Chefredaktor Historisches Lexikon der Schweiz (HLS)
Historisches Lexikon der Schweiz (HLS)
Hirschengraben 11, Postfach 6576
3001 Bern, Schweiz

Studium der Geschichte und französischen Literatur in Freiburg (Schweiz) und Poitiers (Frankreich), Dissertation über den Untergang des Fürstbistums Basel (1792 – 1815), seit 1988 Chefredaktor des Historischen Lexikons der Schweiz (HLS) in Bern, Publikationen in den Bereichen Bundesstaatsgeschichte, Militärgeschichte, Religionsgeschichte.

infoconnex – Angebotsstrategien für solche, die Nutzer werden wollen

Doris Bambey

IZ Bildung des Deutschen Instituts für Internationale Pädagogische Forschung

Zusammenfassung

infoconnex als Zusammenschluss der Fachinformationsangebote der drei Disziplinen Pädagogik, Sozialwissenschaften und Psychologie verfolgt einen Ansatz, der neben der klassischen Klientel der universitär gebundenen Lehre und Forschung auch die außeruniversitären – oft im berufspraktischen Umfeld befindlichen – Interessenten ansprechen und an Fachinformation heranführen soll.

Dieser Anspruch wird über Angebotsstrategien verfolgt, die zielgruppenbezogene und vereinheitlichte Nutzungskonditionen für die drei Fächer bieten und ist eingebunden in ein umfassendes Konzept der nahtlosen Versorgung mit Volltexten im Sinne der appropriate copy, der Präsentation von Interdisziplinarität unter einer Oberfläche sowie einem flexiblen bedarfsgerechten Preis- und Geschäftsmodell, das für Datenbanken und Volltexte gleichermaßen gilt.

1 Bibliotheken und Fachinformationsanbieter als Partner

Auf der Basis einer gemeinsamen, strategisch ausgerichteten Förderung von DFG und BMBF¹ sind die federführenden Sondersammelgebietsbibliotheken und Fachinformationsanbieter der drei Fachgebiete arbeitsteilig bzw. mit verteilten Schwerpunkten an dem Vorhaben beteiligt:

- IZ Bildung des DIPF, Frankfurt a.M.
- IZ Sozialwissenschaften, Bonn

¹ Ein analoges Modell der gemeinsamen Förderung liegt auch vascoda zu Grunde, um so die Akteure - die Virtuellen Fachbibliotheken und die Fachinformationsanbieter - in eine konsistente und möglichst breit ausgerichtete Strategie einzubinden. Das Vorgehen basiert auf den in der Kooperationsvereinbarung zur Förderung von Informationsverbänden und Virtuellen Fachbibliotheken durch das Bundesministerium für Bildung und Forschung (BMBF) und die Deutsche Forschungsgemeinschaft (DFG), 2001, formulierten Grundsätzen,

http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/veroeffentlichungen/#2

- Zentrum für Psychologische Information und Dokumentation (ZPID), Trier
- Universitätsbibliothek Erlangen-Nürnberg (SSG Erziehungswissenschaft)
- Universitäts- und Stadtbibliothek Köln (SSG Sozialwissenschaften)
- Stadt- und Universitätsbibliothek Saarbrücken (SSG Psychologie)

Den Bibliotheken kommt die Aufgabe zu, unter der Federführung der Universitätsbibliothek Erlangen-Nürnberg Verlagsverhandlungen zur Lizenzierung von pay-per-view-Angeboten zu erzielen. Dazu wurden entsprechende Musterverträge als Grundlage entwickelt. Die Informationsanbieter wiederum sind schwerpunktmäßig zuständig für die Entwicklung einer interdisziplinären Recherche, die Generierung von Mehrwertdiensten und Verfügbarkeitsrecherchen sowie eines nutzerfreundlichen, tragfähigen Geschäftsmodells.

Als Auftragnehmer für die Speicherung von über infoconnex lizenzierten Volltexten sowie für die Bereiche Kundenverwaltung und e-payment fungiert die Gesellschaft für Betriebswirtschaftliche Information (GBI).

2 Umfassende Verfügbarkeitservices für eine nahtlose Literaturversorgung

Die nahtlose Versorgung der Nutzer mit Volltexten im Sinne der appropriate copy ist eine wesentliche Zielsetzung des Vorhabens. Hierbei gilt zu beachten, dass im geistes- und sozialwissenschaftlichen Verlagsbereich die Produktion und das Angebot kostenpflichtiger elektronischer Zeitschriften durch die überwiegend mittelständisch geprägte Verlagslandschaft weitaus schwächer ausgeprägt ist als im STM-Sektor. Insbesondere pay-per-use-Angebote sind im sozial- und geisteswissenschaftlichen Bereich bis dato eher nur in Ansätzen vorhanden, so dass die Gruppe der institutionell ungebundenen Einzelnutzer - etwa Journalisten, Therapeuten, Lehrer -, die über keine Campuslizenzen abgedeckt ist, neben dem klassischen Abonnement kein nennenswertes Angebot auf der Ebene der Einzelartikel vorfindet.

Mit Blick auf dieses Desiderat wurden unter der Regie der Universitätsbibliothek Erlangen-Nürnberg direkte Verlagsverhandlungen für alle drei Fächer durchgeführt, um hierüber neue elektronische Volltexte im pay-per-view-Verfahren zu gewinnen. Neben den institutionell ungebundenen Einzelnutzern kann dieses flexibel nutzbare Angebotssegment durchaus auch für Universitätsbibliotheken von Interesse sein. Denn der Kauf von Einzelartikeln kann in den Fällen, in denen keine Campuslizenz vorliegt, weil die Zeitschrift ggf. der Spitzenversorgung zugerechnet wird und auf Grund des Preisniveaus keine Lizenz vorhanden ist, helfen, eine Lücke in der lokalen Literaturversorgung der Universitätsangehörigen zu schließen.

Auf der Basis von aus den Fächern erstellten Verlags- und Zeitschriftenlisten wurden Verhandlungen mit 172 deutschen/deutschsprachigen Verlagen (355 Zeitschriften) und 48 ausländischen Verlagen (277 Zeitschriften) durchgeführt. Im Ergebnis konnten Verträge für 72 Zeitschriften, d.h. über ca. 10 000 Aufsätze, geschlossen werden. Die Erfahrungen mit den Verlagen zeugen davon, dass zumindest im sozial- und geisteswissenschaftlichen Bereich immer noch eine eher ambivalente Haltung gegenüber dem elektronischen Publikationsmarkt und neuen Vertriebsformen vorherrscht. Starke Umstrukturierungen der Verlagslandschaft (z.B. Bildung von Verlagsgruppen, Fusionierungen z.B.), Planungen zur Novellierung des Urheberrechtsgesetzes und die Klage gegen SUBITO treten hinzu und erzeugen eine z.T. eher abwartende Haltung.

Das heißt, dass das pay-per-view-Segment im Gesamtspektrum der Verfügbarkeit von Volltexten im sozial- und geisteswissenschaftlichen Bereich schon jetzt durchaus seinen Platz hat, aber ausgehend vom derzeitigen Volumen und dem Renommee der über pay-per-view angebotenen Zeitschriften (Peer-reviewed-Zeitschriften) eher von einer mittel- und langfristigen Entwicklung ausgegangen werden muss.

Eine Hauptprämisse und ein entscheidender Qualitätsindikator für Fachinformationsservices ist die möglichst lückenlose Versorgung von Wissenschaft und Forschung mit Volltexten unter Einschluss aller vorhandener Verfügbarkeitswege. Das heißt u.a., dass „die Weiterentwicklung der vorhandenen Informationssysteme [...] auf eine konsequente Koordination der überregionalen und lokalen Informationsangebote hinarbeiten“ muss.²

Bei infoconnex wird diese Zusammenführung der verschiedenen überregionalen und lokalen, kostenfreien und kostenpflichtigen Volltext- bzw. Bestelldienste über das in der Bibliothekswelt weit verbreitete Context-Linking-System SFX von der Firma ExLibris realisiert (Abbildung 1). Für den Nutzer bedeutet dies, dass ausgehend von einem recherchierten Nachweis über OpenURL eine Verfügbarkeitsrecherche durchgeführt wird, die ermittelt, welche Volltextserver, Verlage, bibliothekarischen Bestellservices usw. die gewünschte Publikation/den Volltext anbieten.

2 Strategisches Positionspapier »Information vernetzen – Wissen aktivieren«, BMBF, September 2002. Online verfügbar unter: www.dl-forum.de/dateien/Strategisches_Positionspapier.pdf. [Stand 21.01.2005]

The screenshot shows the 'infoconnex' search interface. On the left is a sidebar with navigation links: 'infoconnex-Suche', 'Wir über uns', 'About us', 'Pädagogik', 'Sozialwissenschaften', 'Psychologie', 'Konditionen', 'Registrierung', 'Intern', 'Kontakt', and 'Impressum'. The main area is titled 'Psychologie' and shows a search form with the query 'Überall=testverfahren'. Below the search bar, there are tabs for 'Pädagogik (300)', 'Sozialwissenschaften (90)', and 'Psychologie (1100)'. The 'Psychologie (1100)' tab is active, displaying a list of 8 search results under the heading 'Treffer: 1100'. Each result includes a checkbox, a year, a title, a database source (PSYINDEX), and an SFX availability icon.

	Jahr	Titel	Datenbank	Verfügbarkeit
<input type="checkbox"/>	2005	Nachwuchs für die Zukunft - Auswahl von Lehrlingen[...]	PSYINDEX	SFX
<input type="checkbox"/>	2005	Das Prokrustesbett der deutschen Lehrpläne für die Entwicklung von lehrplanvaliden Testverfahren - Zur Konstruktion der Deutschen Mathematiktests für fünfte und sechste Klassen (DEMAT 5+ und DEMAT 6+)[...]	PSYINDEX	SFX
<input type="checkbox"/>	2005	Analytische Psychotherapie mit Kindern und Jugendlichen - Eine Studie zur Behandlungswirksamkeit[...]	PSYINDEX	SFX
<input type="checkbox"/>	2004	Von der Beurteilung zur Notengebung : Ein Lehrbuch, Diagnostik und Benotung in der Aus-, Fort- und Weiterbildung[...]	PSYINDEX	SFX
<input type="checkbox"/>	2004	Grundlagen und Praxis des Hirnleistungsmanagements : Ein Leitfaden für Ärzte und Psychologen[...]	PSYINDEX	SFX
<input type="checkbox"/>	2004	Das Verständnis metaphorischer Sprache bei Grundschulkindern[...]	PSYINDEX	SFX
<input type="checkbox"/>	2004	Diagnostik von Konzentration und Aufmerksamkeitt[...]	PSYINDEX	SFX
<input type="checkbox"/>	2004	Lehrbuch der Psychotherapie für die Ausbildung zurzum Psychologischen Psychotherapeutin und für die ärztliche Weiterbildung : Band 1:	PSYINDEX	SFX

Abb. 1: SFX als Bestandteil der Trefferliste

SFX bietet eine Plattform, die an die eigenen Erfordernisse als Fachinformationsanbieter adaptierbar ist und sowohl in der Auswahl der Contentlieferanten als auch im Hinblick auf die Anordnung und Beschreibung der Dienste innerhalb des SFX-Menüs Erweiterungen erlaubt. Da infoconnex anders als eine Universitätsbibliothek, dem gängigen Einsatzort von SFX, ein offener, überregionaler Dienst ist, wurde eine spezielle Lizenzvereinbarung mit der Firma erwirkt, die dies berücksichtigt.

Folgende synchrone (direkt downloadbare) und asynchrone (mit Zeitverzug verbundene) Dienste werden derzeit unter SFX abgesehen bzw. angeboten:

- Elektronische Zeitschriftenbibliothek (EZB)
- Direkt-Links zu kostenfreien elektronischen Publikationen
- elektronische Bestellverfahren (SUBITO, JASON, Infotrieve, British Library)
- lokale und gedruckte Zeitschriftenbestände über die ZDB
- Aggregatoren wie CrossRef, EBSCO, Verlage, GBI (pay-per-view-Verkauf von Aufsätzen, die über infoconnex lizenziert sind)

Wenn man sich speziell den pay-per-view-Bereich anschaut, wird deutlich, dass die Verfügbarkeitsrate gerade dieser Angebotsform sehr stark zwischen den einzelnen Fächern differiert:

Während die Verfügbarkeitsrate von pay-per-view-Volltexten im Fach Psychologie ca. 70% beträgt, erreicht sie in der Pädagogik lediglich ca. 11-12%. Der Abdeckungsgrad in den Sozialwissenschaften wird zwischen beiden Werten, mit vermutlich stärkerer Nähe zu den Daten für die Pädagogik, anzusetzen sein (derzeit noch keine Überprüfung vorliegend).

Gerade die Psychologie mit ihrer starken Verortung im medizinisch-naturwissenschaftlichen Kontext profitiert von der früher und systematischer vorangegangenen Anwendung avancierter Publikations- und Vertriebswege in ihrem Bereich. Auch ist die aufgrund der exorbitanten Kostensteigerungen der verlegerischen Fachzeitschriften im STM-Bereich erzwungene Ausrichtung auf alternative Publikationswege (Open-Access-Ansatz) dort stärker ausgeprägt als in den Sozialwissenschaften und der Pädagogik.

Insgesamt beträgt die Verfügbarkeitsrate jedoch, wenn man die derzeit eingebundenen Services berücksichtigt - d.h. elektronische Bestelldienste, direktes Download bis hin zur klassischen Fernleihe –, ca. 98%.

Ein wichtiger Gesichtspunkt im Sinne der Nutzerfreundlichkeit ist die Beschreibung des jeweiligen Beschaffungsweges hinsichtlich Zeit- und Preisverhalten. Das heißt, dass der Nutzer angesichts der Vielfalt der angebotenen Dienste darauf angewiesen ist, Unterstützung bei der Auswahl des passenden Dienstes zu erhalten. Nicht immer ist davon auszugehen, dass es der kostengünstigste Dienst ist, der als optimal erachtet wird, ebenso kann das Kriterium der unmittelbaren Verfügbarkeit, d.h. die Schnelligkeit, ausschlaggebend sein und zur bewussten Wahl eines teureren Dienstes führen.

3 Interdisziplinarität effektivieren durch Überwindung von Heterogenität

infoconnex bietet über die renommierten Datenbanken SOLIS, PSYINDEX und FIS Bildung Literaturdatenbank einen direkten Weg von der Recherche in qualitativ hochwertigen Metadaten zu den elektronischen Volltexten. Diese drei Datenbanken mit einem Gesamtbestand von zusammen ca. einer Million Metadaten sind die anerkannten einschlägigen Anlaufstellen für die fachliche Recherche von Informationssuchenden aus Wissenschaft, Forschung und Praxis. Da die Datenbanken ihre Terminologien zur Beschreibung der Dokumente in dem jeweils fachspezifischen Kontext entwickelt haben, existiert eine verbleibende Heterogenität, die bei datenbankübergreifenden Recherchen zu Informationsverlusten führen kann.

Wird ein entsprechender Sachverhalt in der FIS Bildung Literaturdatenbank beispielsweise mit dem Term „Abneigung“ beschrieben, findet sich in der Datenbank PSYINDEX hierzu der Term „Aversion“. Ein Verfahren, dieses Problem aufzufangen, ist der Einsatz von Cross-Konkordanzen. Das heißt, dass bei datenbankübergreifenden Suchen ein automatischer Transfer des gewählten Suchbegriffs in die jeweilige Terminologie des Nachbarfaches stattfindet.

Dieses Verfahren lässt sich auch in den Bereich der Internationalisierung des Angebotes nutzbringend übertragen. In der Pädagogik beispielsweise wird analog zur Suche in deutschsprachigen Datenbanken die Methode des Termtransfers auf der Basis von bilingualen Cross-Konkordanzen auch bei der Anbindung internationaler Datenbanken angewandt. Der Mehrwert für den Nutzer besteht darin, dass bei einer sprachübergreifenden Suche – etwa der gleichzeitigen Suche in der FIS Bildung Literaturdatenbank und in British Education Index (BEI) - automatisch das korrekte, d.h. fach- bzw. datenbankspezifische englische Äquivalent zur Suche verwendet wird. D.h. ein Recherchierender, der mit dem Begriff „Arbeitsmarktstatistik“ in die FIS Bildung Datenbank einsteigt, muss nicht mehr durch trial and error mühsam ermitteln, ob der Begriff in seiner englischen Übersetzung in ERIC nun „employment statistics“ oder „job market statistics“ oder „labo(u)r market statistics“ heißt. Die Zuordnung des datenbankspezifischen englischen Äquivalents einschließlich seiner Synonyme erfolgt automatisch mit Hilfe der hinterlegten Äquivalenzlisten.

4 Harmonisiertes Geschäftsmodell für Datenbanken und Volltexte

Um insbesondere den Gelegenheitskunden viele Anreize und wenig Barrieren zu bieten, wird ein weit ausgedehnter Bereich der Datenbanken kostenfrei angeboten: Die gesamten Recherchefunktionalitäten, die Suchergebnisse in Form von Kurz-Titellisten und bis zu drei Vollanzeigen der Nachweise sind lizenzfrei nutzbar.

Der kostenpflichtige Bereich (Premiumbereich) des Angebotes umfasst dahingegen die Vollanzeige der Nachweise plus des Verfügbarkeitservice via SFX und den Druck und Download der Volltexte.

Das Geschäftsmodell sieht vor, dass zwei Kundengruppen gleichermaßen angesprochen werden. Mit Blick auf die institutionellen Kunden werden die bereits jeweils existierenden Preismodelle der drei Anbieter – DIPF, IZ Sozialwissenschaften und ZPID – weitergeführt. Eine neue einheitliche Preisgestaltung für institutionelle Kunden ist auf Grund vorhandener Lizenzverträge, großer Unterschiede zwischen den drei Anbietern und externer Verpflichtungen (Verträge mit Distributoren) nicht angeraten.

Im Bereich der Individualkunden konnte demgegenüber ein neues, gemeinsames Preismodell entwickelt werden. Es wurden Tageslizenzen und Jahreslizenzen eingeführt, die bewusst auf einem niedrigen Preisniveau angesiedelt sind. Auf einen Modus, der die Abrechnung von einzelnen Nachweisen (wie z.B. bei GBI) vorsieht, wurde dahingegen verzichtet, da dies – wie weiter unten ausgeführt – weder der Zahlungsakzeptanz der Kunden entsprochen hätte noch im Hinblick auf den damit verbundenen administrativen Aufwand zu rechtfertigen gewesen wäre.

In Bezug auf die Preisgestaltung bei den im Projekt akquirierten Volltexten (Einzelaufsätze der Fachzeitschriften) konnten bei den deutschsprachigen Verlagen überwiegend moderate Preise pro Aufsatz erwirkt werden. Auf der Basis einer zuvor erfolgten Marktuntersuchung hatte sich eine Spanne von Euro 3,- bis 5,- pro Einzelaufsatz als realistisch erwiesen. Diese Ergebnisse der Studie wurden in Form von Empfehlungen in die Verhandlungen eingebracht und überwiegend positiv aufgegriffen.

5 Bedarfs- und Marktuntersuchung des Recherche- und Volltextdienstes infoconnex

Um eine planvolle und nutzeradäquate Umsetzung der zuvor aufgeführten Einzelmaßnahmen mit Blick auf die Spezifika der Sozial- und Geisteswissenschaften zu gewährleisten, wurde zu Beginn des Vorhabens eine umfassende Markt- und Bedarfsuntersuchung durchgeführt.³

Folgende Fragen standen hierbei u.a. im Mittelpunkt des Interesses:

- Welches sind die jetzigen und zukünftigen Nutzergruppen, welche Personengruppen umfassen überwiegend Nichtnutzer?
- Was für eine Bedarfslage und welche Verwendungszwecke können festgestellt werden bzw. welche Informationen sind für welche Nutzergruppe interessant?
- Wie teuer darf die Leistung sein bzw. welches Preis-Leistungsverhältnis wird als akzeptabel angesehen?

Bezüglich der Informationsbedarfe ergab sich, dass Volltexte bzw. Primärdokumente für alle Nutzergruppen von prioritärer Bedeutung sind (Abbildung 2).

3 Studie: Entwicklung eines Marketingkonzeptes für den Aufbau eines Volltextdienstes im IV-BSB. Durchgeführt durch IMAC (Information & Managementconsulting), 2002. Im Auftrag des Informationsverbundes BSP (jetzt infoconnex).

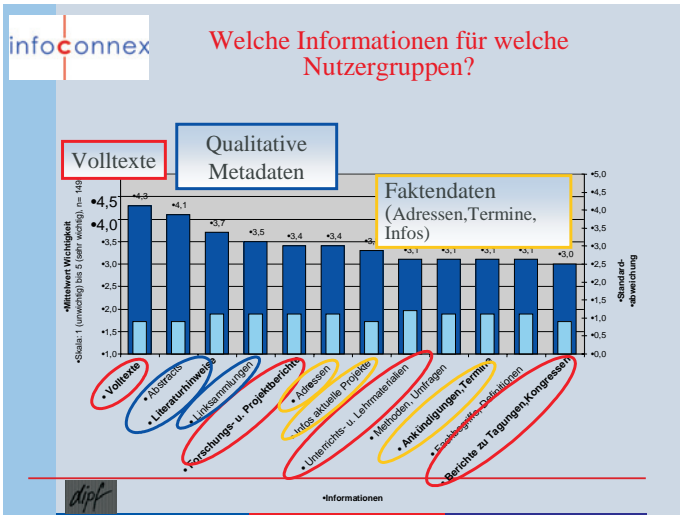


Abb. 2: Welche Informationen für welche Nutzergruppen

Es wurde allerdings festgestellt, dass Metadaten nicht für alle Zielgruppen gleichermaßen interessant sind; ihre Bedeutung nimmt mit zunehmendem Praxisbezug der Klientel ab (Abbildung 3).

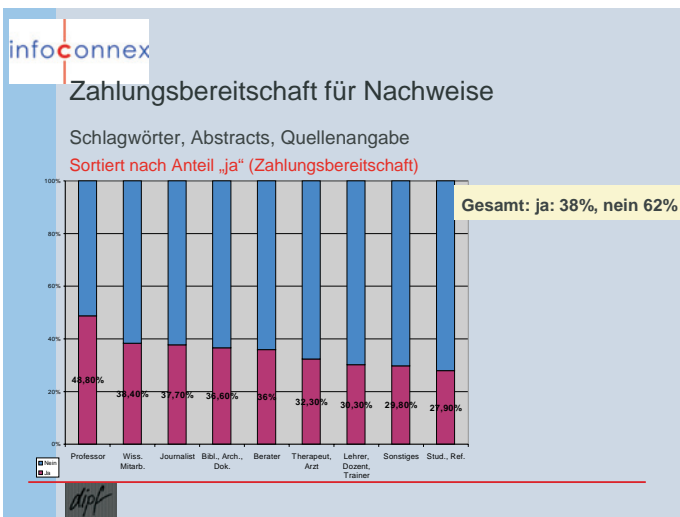


Abb. 3: Zahlungsbereitschaft für Nachweise

Bezüglich der Zahlungsbereitschaft für Metadaten wurde ermittelt, dass am ehesten Professoren und wissenschaftliche Mitarbeiter für Schlagworte, Abstracts, Quellenangaben bezahlen würden. Die ermittelten 38% der Antworten, die eine diesbezügliche Zahlungsbereitschaft bekundeten, kommen fast ausschließlich aus diesem Bereich.

Es lässt sich ein Zusammenhang annehmen zwischen dem wissenschaftlichen Tätigkeitsumfeld und der Einsicht in die Notwendigkeit und Eigenwertigkeit von systematischem, qualitativ gestütztem Wissenstransfer durch Sekundärinformationen. In eher praktischen Tätigkeitsfeldern – Journalisten, Berater, Lehrer, Therapeuten etc. – ist die Zahlungsbereitschaft für Metadaten bzw. die Datenbanknutzung dahingegen sehr gering ausgeprägt. Dieses Kundensegment – also die nicht universitätsgebundene Klientel – sollte von daher möglichst situativ nutzbare und flexible Zugangskonditionen erhalten, um hier nicht mit zusätzlichen Hürden konfrontiert zu sein.

Die Bereitschaft, für Volltexte – also die Primärinformationen - zu zahlen, ist mit 64,4% positiven Antworten weitaus stärker ausgeprägt (Abbildung 4). Was das Preis-Leistungsverhältnis angeht, so würde die Mehrzahl der Befragten für als nutzbringend erachtete Informationen zwischen Euro 3,- bis Euro 5,- zahlen. 68% der positiven Antworten bevorzugen eine Abrechnung je Einzelabruf (pay-per-view) und keine Pauschale.

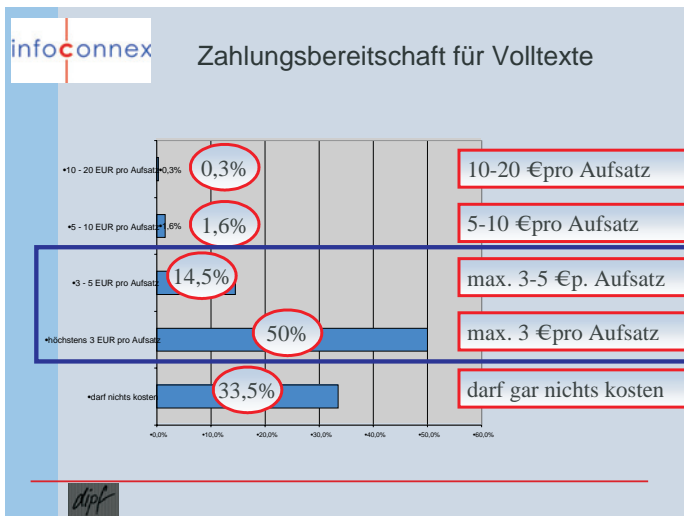


Abb. 4: Zahlungsbereitschaft für Volltexte

6 Angebotsstrategien für solche, die Nutzer werden wollen..., es aber vielleicht nicht wissen!

Ein wesentlicher Teil der akademisch ausgebildeten Klientel verliert mit dem Übergang in berufspraktische Zusammenhänge und Tätigkeitsfelder weitgehend den Anschluss an fachliche Informationsdienste, auch wenn die Forderung nach bzw. der Bedarf an kontinuierlicher Weiterqualifizierung und Wissensaneignung im beruflichen Alltag allerorten evident ist. Die Ergebnisse der im Projekt durchgeführten Untersuchung ergeben, dass der Bekanntheitsgrad der drei nachgefragten Fachdatenbanken auch im Bereich von Wissenschaft und Forschung mit einem Bekanntheitsgrad von 45,6 % (Pädagogik), 52,9% (Sozialwissenschaften und 52,9% (Psychologie) noch sehr gering ausgeprägt ist (Abbildung 5).

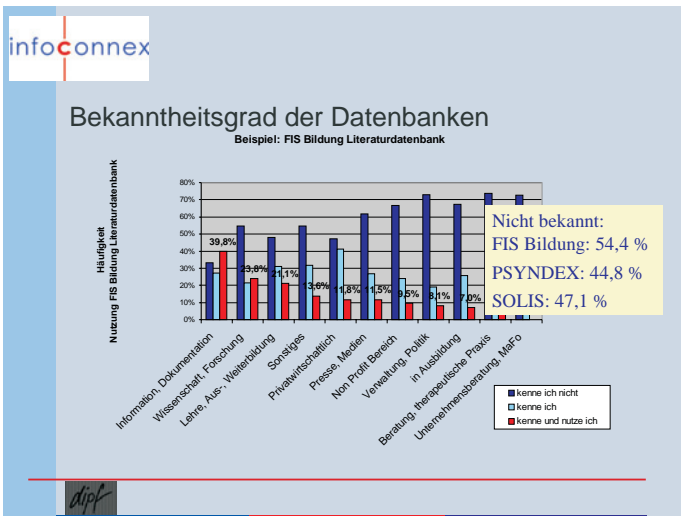


Abb. 5: Bekanntheitsgrad der Datenbanken

Ein Blick auf die Ergebnisse der im Jahr 2001 im Auftrag des BMBF durchgeführte SteFi-Studie⁴, wonach 61% der Studierenden und 66% der Lehrenden häufig allgemeine Suchmaschinen benutzen, aber nur 5% der Studierenden und 15% der Lehrenden fachspezifische Portale, stützten diese Beobachtung.

4 Studieren mit elektronischen Fachinformationen (SteFi), 2001, <http://www.stefi.de/>

Auch in den Universitäten gilt es nach wie vor, auf allen Ebenen – Studentenschaft, Mittelbau, Professuren - ein hohes, noch nicht aktives Potenzial von Endnutzern an fachliche Recherche- und Volltextdienste heranzuführen.

Ein Ansatz, der in der Pädagogik verfolgt wurde und verschiedentlich auch in anderen Fächern bzw. von anderen fachlichen Dienstleistern erprobt wird, zielt darauf, Google als Zubringer von Interessenten und potentiellen Nutzern zu instrumentalisieren. Dazu wurde die FIS Bildung Literaturdatenbank für Google geöffnet, so dass in der Trefferliste von Google Literaturnachweise mit direkten Links zur FIS Bildung Literaturdatenbank aufgeführt werden (Abbildung 6).



Abb. 6: Suchmaschinen als Zubringer von Nutzern

Die Folge war, dass im Beobachtungszeitraum (Nov. 2004 bis März 2005) eine 100-fach höhere Nutzungsfrequenz des kostenfreien Segments der Datenbank bezogen auf Sessions⁵ zu verzeichnen war. Nutzer, die über Google zur Datenbank gelangen, führen im kostenfreien Bereich durchschnittlich drei Transaktionen durch. Im kostenpflichtigen Bereich ergibt sich im Beobachtungszeitraum eine Steigerung der Sessions um 32%. Da im Beobachtungszeitraum nicht

5 Einer Session (Anwendersitzung) wird in diesem Fall eine Zeiteinheit von 65 Minuten (als empirisch sinnvolle zeitliche Taktung auf der Basis der Nutzungen des Deutschen Bildungsservers und FIS Bildung ermittelt) zugrundegelegt, das bedeutet, dass in dieser Zeiteinheit von einer IP-Adresse aus auf die Datenbank zugegriffen wird.

signifikant viele neuen Lizenzkäufe registriert wurden, handelt es sich bei diesen 32% hauptsächlich um schon über ihre Universität zugelassene Kunden. Das heißt, dass ein bisher nicht aktives Potenzial an Endnutzern, die sich hinter dem institutionellen Lizenznehmer Universitätsbibliothek verbergen, auf diesem Weg aktiviert werden könnte. Gleichzeitig gilt es, die bisher in noch größerer Entfernung zur Fachinformation sich bewegenden Internetnutzer im beruflichen Umfeld an qualifizierte fachliche Dienste heranzuführen und durch ein qualitätsgeprüftes, inhaltlich breites, einfach bedienbares und durch pay-per-use-Komponenten situativ nutzbares Serviceangebot gleichfalls an das Angebot zu binden.

Angaben zur Autorin

Doris Bambey

Deutsches Institut für Internationale Pädagogische Forschung (DIPF)

Schloßstr. 29

60468 Frankfurt/M.

Dipl. Soziologin, Wissenschaftliche Dokumentarin

Koordinatorin der Projekte *infoconnex*-Pädagogik, Fachportal Pädagogik sowie Barrierefreiheit von *vascoda*, Mitarbeit in der Koordinationsstelle Fachinformationssystem Bildung.

Neue Wege der Wissenspräsentation und –vermittlung in der Virtuellen Fachbibliothek Sozialwissenschaften (ViBSoz): Thematische Dokumentationen

Stefan Jakowatz

Informationszentrum Sozialwissenschaften Bonn (IZ)

Zusammenfassung

Die Virtuelle Fachbibliothek Sozialwissenschaften (www.vibsoz.de) bietet einen fachspezifischen Zugriff auf wissenschaftsrelevante Informationen und Dokumente über das Internet. Das Informationszentrum Sozialwissenschaften (IZ) betreibt dabei die inhaltliche Ausgestaltung in Kooperation mit drei fachlich kompetenten und erfahrenen Bibliotheks- und Serviceeinrichtungen - Bibliothek der Friedrich-Ebert-Stiftung (FES), Universitäts- und Stadtbibliothek Köln (USB) und Wissenschaftszentrum Berlin für Sozialforschung (WZB).

Einen Schwerpunkt des inhaltlichen Angebots bilden Dokumentationen zu aktuellen gesellschaftswissenschaftlichen Themen (z.B. das deutsche Gesundheitswesen, Mobilitätsforschung, Reformen des Sozialstaats, Europa etc.). Thematische Dokumentationen bieten unterschiedlichen Zielgruppen einen neuen Weg der Wissenspräsentation und -vermittlung, indem verschiedene Informationstypen und -ressourcen aus verteilten Beständen (z.B. aus Literatur-, Forschungs- und Institutionendatenbanken, Bibliothekskataloge, Internet) zu einem integrativen Informationssystem nachhaltig zusammengeführt und darauf aufbauend dynamisch erweitert werden.

Auf Basis der Nachnutzung von DBClear werden verschiedene Informationstypen wie Literaturnachweise, Forschungsreferenzen, Volltexte, Materialien, Daten, Fakten, Linklisten etc. unter Einbeziehung von Fachleuten und Nutzern intellektuell nach inhaltlichen und formalen Gesichtspunkten aggregiert.

Völlig neuartig ist die Möglichkeit der kooperativen Arbeit an thematischen Dokumentationen, unabhängig vom institutionellen/räumlichen Standort sowie die Aufbereitung der Inhalte durch einführende redaktionelle Textbeiträge und Kurzbeschreibungen.

Neuartig ist auch der interaktive und integrierte Browsing-Zugang für den Nutzer, der es ihm erlaubt, gleichzeitig sowohl in der Angebotsstruktur als auch in den Inhalten nach variabel gefilterten Gesichtspunkten navigieren zu können.

Darüber hinaus steht dem Nutzer ein Suchformular zur Verfügung, mit dem er entsprechend seinem Informationsbedürfnis nach spezifischen Angeboten innerhalb der jeweiligen thematischen Dokumentationen recherchieren kann.

1 Einleitung

Das von der DFG geförderte Projekt „Inhaltliche Ausgestaltung der Virtuellen Fachbibliothek Sozialwissenschaften“ ist ein Gemeinschaftsprojekt des Informationszentrums Sozialwissenschaften Bonn (IZ), der Bibliothek der Friedrich-Ebert-Stiftung (FES), der Universitäts- und Stadtbibliothek Köln (USB) sowie der Bibliothek des Wissenschaftszentrums Berlin für Sozialforschung (WZB) (Projektzeitraum 1.10.2003-30.9.2005). Zusammengeschlossen haben sich vier fachlich kompetente und erfahrene Bibliotheks- und Serviceeinrichtungen, um ein attraktives und qualitativ hochwertiges sowie nutzerorientiertes Informationsangebot über die sozialwissenschaftliche Community und ihre Aktivitäten aufzubauen. Der inhaltliche Aufbau der Virtuellen Fachbibliothek Sozialwissenschaften erfolgt ähnlich wie bei anderen Virtuellen Fachbibliotheken¹ und ergänzt somit die erste Projektphase der Entwicklung einer Metasuchmaschine. Der inhaltliche Ausbau der Virtuellen Fachbibliothek Sozialwissenschaften umfasst vier Segmente, aus denen der Aufbau von „thematischen Dokumentationen“ als neuer und im Umfeld der Virtuellen Fachbibliotheken in Deutschland einzigartiger Weg der Wissenspräsentation und -vermittlung herausragt.

2 Die Virtuelle Fachbibliothek Sozialwissenschaften

Der Aufbau der Virtuellen Fachbibliothek Sozialwissenschaften wurde von der Deutschen Forschungsgemeinschaft (DFG) im Kontext der Initiative zur „Weiterentwicklung der überregionalen Literaturversorgung“² gefördert (Projektzeitraum 1.5.1999-31.10.2002). Er ist somit in die von der DFG seit 1998 betriebene Erweiterung des bisherigen Systems der überregionalen dezentralen Literaturversorgung zu einzelnen Fachgebieten eingebunden. Ziel der Initiative ist es, durch Kooperation der Sondersammelgebietsbibliotheken (SSG) zusammen mit den zentralen Fachbibliotheken einen fachspezifischen Zugriff auf gedruckte und digitale Informationsressourcen ohne Medienbrüche über das Internet zu ermöglichen.³ Die DFG reagierte somit auf die allgemein zunehmende

1 Siehe Projekt- und Hintergrundinformationen zu den Inhalten der bestehenden Virtuellen Fachbibliotheken in Deutschland: <http://www.dl-forum.de/deutsch/projekte/projekte.php>

2 vgl. DFG 1998, Kapitel 2

3 vgl. DFG 1998, S. 12f.

Relevanz elektronischer Ressourcen in der wissenschaftlichen Informationsversorgung, die durch Veränderungen im Kommunikations- und Publikationsverhalten in Folge des informationstechnischen Entwicklungsschubes der letzten Jahre - vor allem bedingt durch die Ausbreitung des Internet – stetig zunahm.

Vorrangiges Ziel beim Aufbau der Virtuellen Fachbibliothek Sozialwissenschaften (ViBSoz) war es, Konsistenzbrüche, denen sich ein Wissenschaftler bei der Suche nach sozialwissenschaftlicher Literatur gegenübergestellt sieht, zu minimieren. Vor allem das Problem des inhaltlichen Zugriffs auf verteilte Dokumentenbestände, die bezüglich ihrer Erschließungsart (RSWK, Thesaurus Sozialwissenschaften, verschiedene Klassifikationen etc.) sowie in Bezug auf Inhalt und Struktur höchst unterschiedlich sind, sollte gelöst werden. Im Zentrum steht demnach die Vereinfachung und Beschleunigung des Informationsgewinnungsprozesses durch die „integrierte Bereitstellung sozialwissenschaftlicher Literaturinformationen aus verteilten, verschieden strukturierten und inhaltlich unterschiedlich erschlossenen Datenbeständen, die sich in miteinander nicht verbundenen, heterogenen Organisationsstrukturen und Zugänglichkeitskontexten befinden (Institutsbibliotheken, Sondersammelgebiete der Universitätsbibliotheken (SSG), wissenschaftliche Spezialbibliotheken, Referenzdatenbanken, digitale Volltexte)“⁴.

2.1 Informationstechnologischer und inhaltlicher Aufbau der ViBSoz

Der Aufbau der ViBSoz erfolgt in zwei projektgeförderten Phasen. In einer ersten Projektphase wurde vom Informationszentrum Sozialwissenschaften Bonn und der Technischen Universität Darmstadt⁵ eine Metasuchmaschine entwickelt, die eine integrierte Suche in den sozialwissenschaftlichen Kernbeständen deutschsprachiger Literatur ermöglicht. Der Nutzer dieser Suchmaschine wird in die Lage versetzt, mit nur einer Anfrage und dem ihm vertrauten Erschließungsinstrumentarium in verschiedenen Bibliotheken und Fachdatenbanken zu recherchieren⁶, ohne dass dieselbe Anfrage mehrfach unterschiedlich formuliert und in unterschiedlichen Recherchesystemen eingegeben werden muss. Mit

4 Kluck/Krause/Müller 2000, S. 4

5 In Kooperation mit dem Sondersammelgebiet Sozialwissenschaften der Universitäts- und Stadtbibliothek Köln, der Bibliothek der Friedrich-Ebert-Stiftung Bonn, des Zentrums für Interdisziplinäre Technikforschung der TU Darmstadt sowie dem Westdeutschen Verlag und dem Leske+Budrich Verlag.

6 Katalog der Universitäts- und Stadtbibliothek Köln; Darmstädter Virtueller Katalog; Datenbank SOLIS; Bibliothek der Friedrich-Ebert-Stiftung Bonn; Katalog des Wissenschaftszentrums Berlin für Sozialforschung; Clearinghouse-Datenbank des Informationszentrums Sozialwissenschaften Bonn

Hilfe von Transferkomponenten⁷ werden Anfragen an die Metasuchmaschine automatisch in die anderen beteiligten Sacherschließungssysteme (Thesauri und Klassifikationen) übersetzt. Zusätzlich zu den Transferkomponenten wurde eine einheitliche Benutzungsoberfläche und eine einheitliche Darstellung der Rechercheergebnisse implementiert.

Das seit Anfang 2002 im Einsatz stehende System der ViBSoz⁸, das die verschiedenen „klassischen“ Informationsquellen wie Bibliotheks-OPACs und Fachdatenbanken systemseitig zusammenführt, wird zusätzlich um eine Sammlung „neuer Medien“, d.h. durch intellektuell geprüfte WWW-Quellen des datenbankbasierten Clearinghouse Sozialwissenschaften (GESIS SocioGuide), erweitert. Das Informationszentrum Sozialwissenschaften baut seit 1995 den Fachinformationsführer SocioGuide auf und aktualisiert ihn kontinuierlich. Er enthält eine umfangreiche Sammlung ausgewählter Internet-Adressen und dient vorrangig der Orientierung über deutsche und internationale Internetangebote auf dem Gebiet der Sozialwissenschaften.⁹

2.2 Inhaltliche Ausgestaltung der Virtuellen Fachbibliothek Sozialwissenschaften

In dem Folgeprojekt „Inhaltliche Ausgestaltung der Virtuellen Fachbibliothek Sozialwissenschaften“ (InhalteViBSoz) wird seit Oktober 2003 ein umfassendes inhaltliches Informationsangebot für einen breiten Nutzerkreis aufgebaut. Der inhaltliche Aufbau umfasst dabei vier datenbankbasierte Segmente. Die Datenbanken für sozialwissenschaftliche Lehr- und Forschungseinrichtungen, Forschungsprojekte, Fachzeitschriften und der Fachinformationsführer SocioGuide existierten bereits vorher als singuläre Angebote im Internet und wurden auf Basis von DBClear reimplementiert und unter ViBSoz zusammengeführt:

Datenbank für sozialwissenschaftliche Lehr- und Forschungseinrichtungen

Die im Informationszentrum Sozialwissenschaften vorhandene Datenbank SOFO¹⁰ zu sozialwissenschaftlichen Forschungseinrichtungen in Deutschland wurde systematisch um nationale Lehrstühle und Lehrstuhlinhaber erweitert. Das Wissenschaftszentrum Berlin für Sozialforschung hat dabei das Angebot um Informationen zu wichtigen internationalen Forschungseinrichtungen und darin tätigen Schlüsselpersonen ergänzt.

7 Zum Thema ‚Hetreogenitätsbehandlung‘ und auch zur Entwicklung und zum Einsatz von Transferkomponenten siehe: Hellweg/Krause/Mandl/Marx/Müller/Mutschke/Strötgen 2001.

8 vgl. <http://www.vibsoz.de>

9 vgl. <http://www.gesis.org/SocioGuide/index.htm>; Rubrizierung nach GESIS SocioGuide Western World und GESIS SocioGuide Eastern Europe

10 vgl. <http://www.gesis.org/Information/SOFO/index.htm>

Datenbank für sozialwissenschaftliche Forschungsprojekte

Die Datenbank FORIS¹¹ des Informationszentrums Sozialwissenschaften, die eine umfangreiche Sammlung geplanter, aktueller und abgeschlossener Forschungsprojekte bietet, wurde im Projektverlauf vom Wissenschaftszentrum Berlin für Sozialforschung durch ein institutsinternes Meldeverfahren um Daten erweitert.

Datenbank für sozialwissenschaftliche Fachzeitschriften

Die Zeitschriftenkataloge bzw. -datenbanken des Informationszentrums Sozialwissenschaften¹², der Friedrich-Ebert-Stiftung und der Universität- und Stadtbibliothek Köln wurden für das gemeinsame Angebot in der Virtuellen Fachbibliothek abgeglichen, in den Strukturen harmonisiert und inhaltlich klassifiziert.

Thematische Dokumentationen

Im Zentrum der inhaltlichen Ausgestaltung der ViBSoz stehen die so genannten „thematischen Dokumentationen“. Sie bilden ein qualitativ hochwertiges Angebot durch die Verwendung institutseigener Informationsressourcen, unterstützt durch die professionellen Erfahrungen der vier beteiligten Bibliotheken bzw. Serviceeinrichtungen.

2.3 Technischer Rahmen

Technische Basis der Angebote in der ViBSoz ist die Software DBClear¹³. Das Softwaresystem entstand im Rahmen eines DFG-geförderten Projektes¹⁴ und wurde zur kooperativen Erstellung und Pflege von Fachinformationsführern zu Internetquellen (auch Clearinghouses oder Quality Controlled Subject Gateways genannt) vor allem für den Einsatz in den von der DFG geförderten Virtuellen Fachbibliotheken entwickelt. DBClear unterstützt systemseitig kooperative und automatische Arbeitsprozesse institutsübergreifend und/oder über mehrere Personen hinweg. Die Arbeitsprozesse können so gestaltet werden, dass in den einzelnen Arbeitsschritten unterschiedliche Kompetenzen genutzt werden. Darüber hinaus wird auch die Einbeziehung externen Spezialwissens ermöglicht (Workflow-Unterstützung). Um die Mitarbeiter beim arbeitsintensiven Aufbau und der Pflege einer solchen Quellensammlung von häufig wiederkehrenden Arbeiten zu entlasten, werden Arbeitsabläufe automatisiert (z.B. auto-

11 vgl. <http://www.gesis.org/Information/FORIS/index.htm>

12 vgl. <http://www.gesis.org/Information/Zeitschriften/index.htm>

13 Datenbankbasiertes Clearinghouse im Kontext digitaler Bibliotheken, vgl. Hellweg/Hermes/Stempfhuber/Enderle/Fischer 2002

<<http://www.gesis.org/Forschung/Informationstechnologie/Dateien/hellweg.pdf>>

14 vgl. <http://www.gesis.org/Forschung/Informationstechnologie/DBCclear.htm>

matischer Link-Checker). DBClear ist technisch so gestaltet, dass es flexibel an unterschiedlichste Anwendungsszenarien angepasst werden kann. So können das Metadatenschema für die Beschreibung der Quellen oder die Gestaltung der Benutzungsoberfläche mit geringem Aufwand den Anforderungen entsprechend verändert werden.

3 Thematische Dokumentationen

3.1 Definition und Ziel

Die thematischen Dokumentationen der ViBSoz sind inhaltlich und formal strukturierte Sammlungen von Informationsquellen zu gesellschaftspolitisch relevanten und aktuellen Themen, die sich an Fachwissenschaftler, Hochschulen, Schulen, an die Erwachsenenbildung sowie an die interessierte Öffentlichkeit richten. Sie eröffnen einen fachspezifischen Zugriff auf wissenschaftsrelevante Informationen und Dokumente. Maßgeblich für die Einbeziehung der Informationsquellen ist ausschließlich deren inhaltliche Relevanz für das betreffende Fachgebiet bzw. -thema, nicht der Produkttyp oder das Informationsmedium. Damit kann prinzipiell jede Art von Informationsquelle (z.B. Literaturnachweise, Forschungsreferenzen, Volltexte, Materialien, Daten, Fakten, Linklisten, Lehr- und Forschungseinrichtungen, Fachzeitschriften, Kongresse, Tagungen, Bilder, Filme, Tondokumente etc.) zur Verfügung gestellt werden. Dem Nutzer steht somit eine qualitätsgeprüfte, durch ein Metadatenschema beschriebene und i. d. R. kommentierte Sammlung von fachlich relevanten Quellen zur Verfügung, die das schwer zu überblickende Informationsangebot auf einen Fachauschnitt begrenzt.

Diese Form von Informationsversorgung nähert sich dem Informationsbedürfnis von Nutzern dadurch an, dass a) wissenschaftsrelevante Informationen - unabhängig von der Informationsart oder vom Informationstyp - integriert, aktuell und ortsunabhängig zugänglich gemacht und b) qualitativ hochwertig aggregierte Angebote ohne den so genannten „information overflow“, d.h. ohne irrelevante und randständige Informationen unter fachwissenschaftlicher Redaktion bereitgestellt werden. Darüber hinaus bieten thematische Dokumentationen c) im Gegensatz zu den traditionell singulären und zumeist fachspezifischen Informationsangeboten (z.B. Literaturdatenbanken, Institutionendatenbanken etc.) einen themenspezifischen bzw. interdisziplinären Zugang zur wissenschaftlichen Informationswelt.

3.2 Konzeption eines integrierten, interaktiven und kooperativen Informationsangebots

Thematische Dokumentationen sollen einen themenspezifischen, integrierten und interaktiven Zugriff auf fachrelevante Informationen und Dokumente über das Internet ermöglichen. Mit dieser Zielsetzung geht die Konzeption „über den klassischen Ansatz aller Virtuellen Fachbibliotheken, einen fachspezifischen Zugriff auf wissenschaftsrelevante Informationen und Dokumente zu bieten, hinaus“¹⁵. Sind für Informationssysteme (z.B. Literaturdatenbanken) der Informations- bzw. Medientyp für die Auswahl entscheidend, so ist bei den thematischen Dokumentationen allein die Relevanz einer Ressource - unabhängig von ihrer Beschaffenheit – für die Aufnahme in die Sammlung bestimmend¹⁶. Diese Art der Wissenspräsentation und -vermittlung ist eine konzeptionelle Weiterentwicklung bereits bestehender Informationsangebote - „sowiPlus“¹⁷ des Informationszentrums Sozialwissenschaften und „FES-Netzquelle: Geschichte und Politik“¹⁸ der Friedrich-Ebert-Stiftung Bonn - unter Nachnutzung der in den Projekten ViBSoz und DBClear geleisteten Instrumentenentwicklung.

Die Konzeption der thematischen Dokumentationen als elektronisches Publikationsmedium verbindet die Eigenschaften „klassischer“ Publikationsmedien wie die eines Buches oder einer Zeitschrift mit den Funktionalitäten eines modernen elektronischen Informationssystems. Adaptiert der inhaltliche Aufbau (Gliederung) eines Themas noch die übliche Strukturierung herkömmlicher wissenschaftlicher Printprodukte (z.B. das Inhaltsverzeichnis eines Buches oder einer Zeitschrift), so verwandeln interaktive Features (z.B. Browsingschemata, Links zum WWW, Filter, Suchmöglichkeiten etc.) die statischen Inhalte in dynamische, lebendige und jederzeit erweiterbare (aktualisierbare) Inhalte. Wichtig beim Aufbau von thematischen Dokumentationen ist die Nutzerorientierung¹⁹, vor deren Hintergrund sich vier grundlegende Anforderungen an eine innovative Form der Wissenspräsentation und -vermittlung ableiten lassen: Integration, Interaktivität, Kooperation und Mehrwert.

15 Otto 2005, S. 12

16 s.o.

17 Siehe <http://www.gesis.org/Information/SowiNet/index.htm>

18 Siehe <http://library.fes.de/library/netzquelle/>

19 Detaillierte Ausführungen zur Nutzerorientierung vgl. Otto 2005, S. 13f.

3.3 Integration

Thematische Dokumentationen sind inhaltlich und zugleich formal strukturierte Sammlungen von Informationsquellen. Die inhaltliche Strukturierung erfolgt unter hierarchischen Gesichtspunkten und ähnelt dementsprechend dem Inhaltsverzeichnis eines Buches. Dem gegenüber basiert die formale Strukturierung auf vordefinierten Typen von Informationsquellen. Dabei gibt es eine Vielzahl von Produkt- bzw. Medientypen, die als Objekttypen zu so genannten Objektgruppen aggregiert werden (Tabelle 1).

Tabelle 1

Objektgruppen	Objekttypen
Volltexte	Bücher, Zeitschriften, Aufsätze, Dissertationen. Graue Literatur
Literaturverweise	Bücher, Zeitschriften, Aufsätze, Dissertationen. Graue Literatur
Forschungsprojekte	Forschungsprojekte [laufend und abgeschlossen]
Datenbanken/Portale	OPACs, Fachdatenbanken, Portale/Fachinformationsführer
Veranstaltungen	Veranstaltungen
Organisationen	universitäre und außeruniversitäre Institute, Behörden, sonstige Organisationen
Materialien	Lehr- und Lernmaterialien, av-Medien, Statistiken
Kommunikation/Personen	Mailinglisten / Newsgroups / Personen

Quelle: Otto 2005, S. 12

Im Projektkontext erfolgt die Integration von Informationsquellen auf zwei unterschiedlichen Wegen. Während es bei den thematischen Dokumentationen des Informationszentrums Sozialwissenschaften, der Universitäts- und Stadtbibliothek Köln und des Wissenschaftszentrums Berlin für Sozialforschung eine differenzierte Rubrizierung und Erschließung einzelner Informationsquellen auf Einzelnachweisbasis gibt, beschränkt sich die Friedrich-Ebert-Stiftung auf die Verknüpfung kommentierter Linksammlungen zu Volltexten und Online-Materialien auf Basis statischer HTML-Seiten²⁰.

Die Integration heterogener Informationsquellen auf Einzelnachweisbasis lässt sich auf zwei Ebenen beschreiben. Auf inhaltlicher Ebene induziert die unterschiedliche Beschaffenheit von Informationsquellen z.B. Literaturnachweise, Forschungsreferenzen, Volltexte, Materialien, Daten, Fakten, Linklisten etc. keinen exkludierenden, sondern einen integrativen Selektionsprozess. Eine andere Art von Integration findet auf der Ebene der Präsentation der Inhalte von thematischen Dokumentationen statt. Hierbei handelt es sich um eine Verzah-

²⁰ Dies spiegelt sich auch in der optisch voneinander abgesetzten Darstellung der beiden Konzepte auf der Eingangsseite der Thematischen Dokumentationen wider, vgl. <http://vt-www.bonn.iz-soz.de/vibsoz/themen.htm>

nung von *integrativen* und *interaktiven* Elementen, die es dem Nutzer - ganz gegen die bisherigen Gepflogenheiten bei der Informationsbeschaffung - erlauben, die inhaltliche und formale Struktur gleichzeitig zu betrachten und jederzeit zwischen Inhaltsverzeichnis und Inhalten zu wechseln (siehe auch „Interaktivität“).

3.4 Interaktivität

Für die Präsentation der thematischen Dokumentationen wurde auf Basis von DBClear ein interaktiver Browsing-Zugang geschaffen, der es dem Nutzer erlaubt, gleichzeitig sowohl innerhalb der Angebotsstruktur als auch innerhalb der Inhalte nach variabel gefilterten Gesichtspunkten navigieren zu können (Abbildung 1).

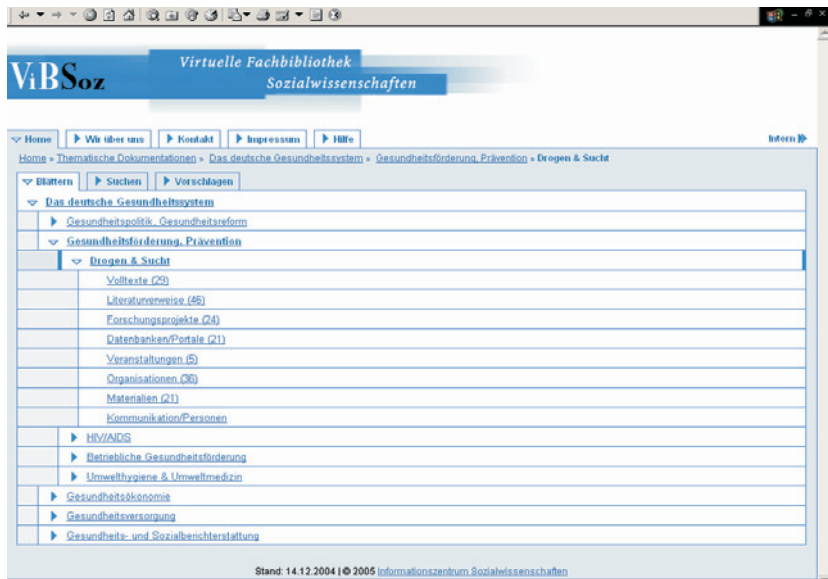


Abbildung 1

Sich innerhalb des Inhaltsverzeichnisses bzw. durch vielschichtige Hierarchien zu bewegen und gleichzeitig die Inhalte betrachten zu können, ist eine innovative Form der Wissenspräsentation und -vermittlung in Virtuellen Fachbibliotheken. Das traditionelle Vorgehen von Nutzern, sich erst ein Inhaltsverzeichnis anzusehen, bevor die Inhalte - genauer gesagt die Informationsressourcen - eingesehen werden, ist obsolet geworden. Das Browsen durch die Inhaltsglieder-

nung und der Zugriff direkt auf Inhalte ist somit zu einem iterativen und integrierten Prozess geworden, der die typischen Probleme von Nutzern bei der Navigation durch stark hierarchisierte Informationsangebote wesentlich minimiert.

Unterstützt wird die integrierte und interaktive Informationspräsentation durch ein neuartiges Grafikdesignkonzept. Dieses Konzept adaptiert den integrativen Gedanken und hebt die (allseits übliche) Links-Rechts-Darstellung von Navigation und Inhalt auf, indem die Navigationselemente und die Inhalte in einer Ansicht untereinander dargestellt werden (Abbildung 2).

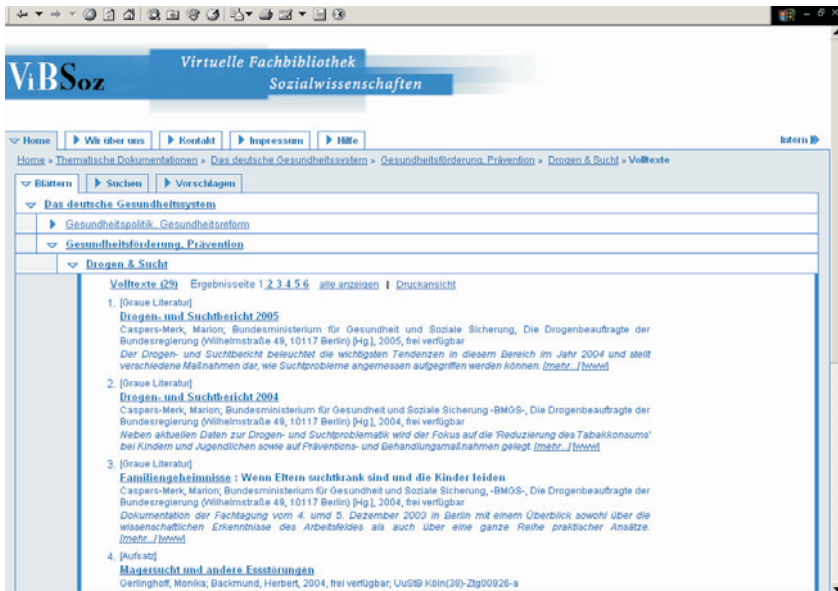


Abbildung 2

Des Weiteren eröffnet DBClear die interessante Möglichkeit, zusätzliche Elemente wie Spezialfilter, die den sichtbaren Inhalt nach formalen Kriterien strukturieren, einzubetten. Die beteiligten Projektpartner haben sich jedoch aus nutzerorientierten Gesichtspunkten dagegen entschieden, bei den bestehenden thematischen Dokumentationen neben dem Browsing über die Inhaltsstruktur auch ein Browsing über andere vordefinierte Kriterien wie z.B. Informationstyp, Nutzergruppe, Land, Sprache etc. anzubieten.

Ein interaktives und nutzerfreundliches Element ist die themenspezifische Suche in thematischen Dokumentationen auf Einzelnachweisbasis (Abbildung 3). Ein Suchformular mit verschiedenen Filterfunktionen (z.B. Objektgruppen)

ermöglicht es dem Nutzer, unabhängig von der Browsingarchitektur die Inhalte eines Themas wiederzufinden, die seinem speziellen Informationsbedürfnis entsprechen. Die Rechercheergebnisse werden nach Objektgruppen (s.o) sortiert und als Liste ausgegeben, so dass die Orientierung bei größeren Treffermengen erheblich erleichtert wird.

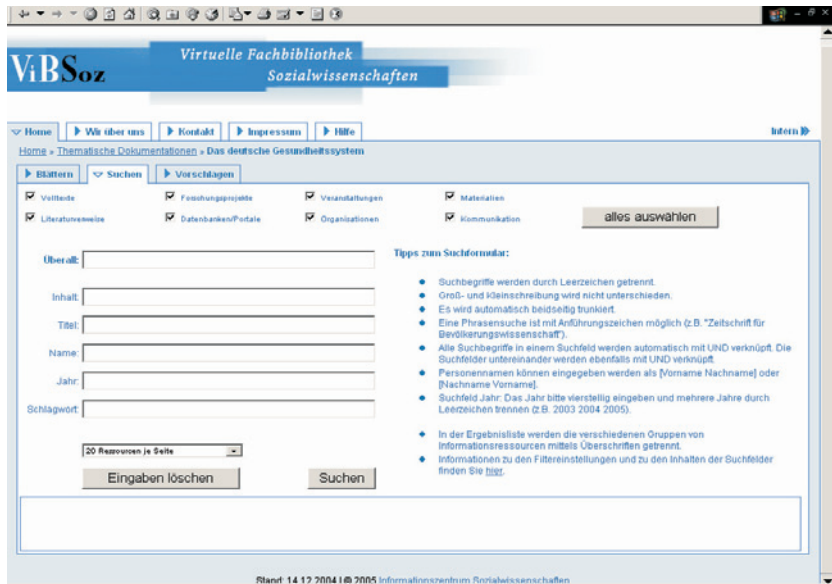


Abbildung 3

3.5 Kooperation und Mehrwert

Der Aufbau eines umfassenden Themenangebots in der ViBSoz ist dauerhaft angelegt und kann von einer Einrichtung allein nicht geleistet werden. Das Informationszentrum Sozialwissenschaften hat sich deshalb mit den o.a. Bibliotheks- und Serviceeinrichtungen zusammengeschlossen, um gemeinsam ein für unterschiedliche Nutzergruppen hochwertiges Angebot aufzubauen, das nach der Startfinanzierung durch die DFG weiterbetrieben werden soll. Kernelement der institutionellen Zusammenarbeit ist die Nachnutzung der Software DBClear. Das System ist sowohl in seinem Erscheinungsbild als auch in Art und Anzahl der verwendeten Kategorisierungs- und Beschreibungssysteme an die Bedürfnisse der jeweiligen Institution anpassbar²¹ und ermöglicht die kooperati-

21 vgl. <http://www.gesis.org/Forschung/Informationstechnologie/DBCclear.htm>

ve Arbeit an thematischen Dokumentationen unabhängig vom institutionellen und/oder räumlichen Standort. Durch die Möglichkeit der arbeitsteiligen Bearbeitung von Themen auf der Entwicklungsebene werden Mehrwerteffekte erzielt, die in diesem Umfang und in der vorhandenen Qualität ansonsten nicht erreichbar wären. Dem praktischen Themenaufbau vorausgegangen sind Absprachen sowohl über die intellektuelle Prüfung und Bewertung von Informationsquellen nach definierten Qualitätskriterien²² als auch über die inhaltliche und formale Erschließung dieser Informationsquellen durch standardisierte Metadaten²³. Ein zentraler webseitiger Zugang für alle Partner gestattet es, kooperativ, zeit- und ortsunabhängig an einem Thema, an Thementeilbereichen (z.B. Migration - Ausgewählte Themenbereiche) oder nur an einzelnen Informationstypen zu arbeiten. Gemeinsame redaktionelle Texte zum jeweiligen Themeneinstieg sowie die fachliche Kommentierung der einzelnen Informationsressourcen sind weitere Maßnahmen zur Steigerung des informationellen Mehrwerts für den Nutzer.

Die Anzahl der Kooperationspartner für bestehende und zukünftige Themen ist systemseitig nicht begrenzt, so dass beliebig viele Kooperationspartner aus dem Bibliotheks- und Informationswesen, aber auch aus dem Bereich der Wissenschaft in den Aufbau von thematischen Dokumentationen in der ViBSoz einbezogen werden können.

Kooperationen zwischen den projektbeteiligten Institutionen sind zwar notwendig, jedoch nicht ausreichend für den Aufbau von thematischen Dokumentationen. Der „Benutzer“ selbst ist sowohl bei der konzeptionellen Gestaltung der thematischen Dokumentationen aktiver Bestandteil²⁴ als auch beim weiteren Ausbau der Themen, z.B. durch die Möglichkeit zur Meldung von Veröffentlichungen, Projekten, Veranstaltungen etc. an redaktionell zuständige Mitarbeiter von ViBSoz über ein Kontakt- bzw. Vorschlagsformular.

4 ViBSoz - Thematische Dokumentationen

Die thematischen Dokumentationen greifen aktuelle gesellschaftspolitische Diskussionen auf, befassen sich mit neu etablierten Fachgebieten der Sozialforschung oder dokumentieren längerfristig benötigte bildungs- und sozialpoliti-

22 vgl. Bargheer 2002

23 vgl. die Empfehlungen zur Definition eines Metadaten-Core-Sets für die verteilte Suche über die Virtuellen Fachbibliotheken, vorgelegt vom META-LIB-Projekt Der Deutschen Bibliothek und der SUB Göttingen. Kurzfassung. Version 1.0, 2001-10-05.

24 Das Konzept und die ersten Entwürfe für inhaltliche Gliederungen des Themas Mobilitätsforschung wurden einer Gruppe von Fachwissenschaftlern des Wissenschaftszentrums Berlin für Sozialforschung vorgestellt und mit ihnen diskutiert, vgl. Otto 2005, S. 13f.

sche Informationen. Folgende Themen sind bereits abgeschlossen und werden online angeboten:

- Das deutsche Gesundheitssystem
- Ernährung in der öffentlichen Gesundheitspflege
- Gentechnologie am Menschen - soziale und ethische Implikationen
- Migration - Ausgewählte Themenbereiche
- Mobilitätsforschung
- Reform des Sozialstaats
- Europa / Europäische Union
- Frauenbewegung/ Genderpolitik
- Industrialisierung / Anfänge der Gewerkschaftsbewegung
- Europäische Gewerkschaftsbewegung
- Internationale Gewerkschaftsbewegung
- DDR-Geschichte
- Arbeitsmarkt und Beschäftigung

Weitere Themen befinden sich in Bearbeitung oder in der Konzeptionsphase.

Literaturverzeichnis

Bargheer, M. (2002): Qualitätskriterien und Evaluierungswege für wissenschaftliche Internetquellen.

<<http://webdoc.sub.gwdg.de/ebook/aw/2003/bargheer/v10.pdf>>

[letzter Zugriff am 31.07.2005]

DFG (Hrsg.) (1998): Weiterentwicklung der überregionalen Literaturversorgung. Memorandum

<http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/memo.pdf> [letzter Zugriff am 31.07.2005]

Die Deutsche Bibliothek; Niedersächsische Staats- und Universitätsbibliothek Göttingen (Hrsg) (2001): Empfehlungen zur Definition eines Metadaten-Core-Sets für die verteilte Suche über die Virtuellen Fachbibliotheken, Kurzfassung. Frankfurt.

<http://www2.sub.uni-goettingen.de/metacore/empfehlungen/core_set.pdf>

[letzter Zugriff am 31.07.2005]

Hellweg, H.; Hermes, B.; Stempfhuber, M.; Enderle, W.; Fischer, T. (2002): DBClear: A Generic System for Clearinghouses. In: Adamczak, W.; Nase, A. (Hrsg.): Gaining inside from research information. Kassel: Kassel Univ.

- Press, S. 131-139
 < <http://www.gesis.org/Forschung/Informationstechnologie/Dateien/hellweg.pdf>> [letzter Zugriff am 31.07.2005]
- Hellweg, H.; Krause, J.; Mandl, T.; Marx, J.; Müller, M. N.O.; Mutschke, P.; Strötgen, R. (2001): Treatment of Semantic Heterogeneity in Information Retrieval. Bonn (IZ-Arbeitsbericht Nr. 23)
http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab_23.pdf [letzter Zugriff am 31.07.2005]
- Kluck, M.; Krause, J.; Müller, M. N.O. (2000): Virtuelle Fachbibliothek Sozialwissenschaften: Projektskizze. Bonn (IZ-Arbeitsbericht Nr. 19)
 <http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab19.pdf> [letzter Zugriff am 31.07.2005]
- Otto, D. (2005): Inhaltliche Ausgestaltung der Virtuellen Fachbibliothek Sozialwissenschaften: Erfahrungsbericht aus dem kooperativen Aufbau eines nutzerorientierten Informationsangebots. In: Friedrich-Ebert-Stiftung (Hrsg.): Soziale Bewegungen online: Projekte der Friedrich-Ebert-Stiftung und ihrer Partnerorganisationen / 8. Inetbib-Tagung in Bonn vom 03. - 05. November 2004, Bd. 17. Bonn: Bibl. der Friedrich-Ebert-Stiftung, S. 9-16
 < <http://library.fes.de/pdf-files/bibliothek/02875.pdf>>
 [letzter Zugriff am 31.07.2005]

Angaben zum Autor

Stefan Jakowatz
 Informationszentrum Sozialwissenschaften
 Lennéstr. 30
 53113 Bonn

Studium der Soziologie, Politikwissenschaften, Psychologie sowie Wirtschaft- und Sozialgeschichte an der Phillips-Universität Marburg, seit 2000 wissenschaftlicher Mitarbeiter im Bereich Datenbankaufbau und -entwicklung (DuE) des IZ Sozialwissenschaften in Bonn, Projektmitarbeiter bei *infoconnex* und *Inhalte ViBSoz.*

Themenorientierte Fachinformationssysteme in den historischen Wissenschaften

Das Informationsnetzwerk zur Geschichte des Rhein-Maas-Raumes <RM.net>

Gisela Minn, Anke S. Hoffmann, Yvonne Rommelfanger
Universität Trier

Zusammenfassung

Das Projekt „Informationsnetzwerk zur Geschichte des Rhein-Maas-Raumes“ wird gemeinsam vom Fach Geschichte und dem Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier getragen.¹ Ziel des Projektes ist es, Modellstrukturen und prototypische EDV-Verfahren für die Schaffung eines themenorientierten Informationssystems in den Geisteswissenschaften aufzubauen und es in einem Institutionen übergreifenden Kooperationsverbund von nationalen sowie internationalen Forschungseinrichtungen sowie außeruniversitären Informationsanbietern zu einer zentralen Publikations- und Kommunikationsplattform weiterzuentwickeln.

Im Mittelpunkt der vierjährigen Pilotphase stehen die Zusammenführung und Aufbereitung heterogener Forschungsdaten unterschiedlicher fachwissenschaftlicher Herkunft und verschiedenen Dokumenttyps, ihre Klassifikation und Tiefenerschließung auf der Basis einheitlicher, normierter Metadatenysteme sowie ihre dauerhafte Bereitstellung in einem projekt unabhängigen, Fächer übergreifenden Informationsnetzwerk. Dabei wird die langfristige, Plattform unabhängige Sicherung und Nachnutzung der Forschungsdaten, die in interdisziplinären Projektverbänden oder Einzelvorhaben erhoben und in unterschiedlichen proprietären Dateiformaten gespeichert wurden, durch den konsequenten Einsatz internationaler Standards auf der Basis von SGML/XML gewährleistet.

1 An dieser Stelle möchten wir unserer ehemaligen Kollegin Sarah Neumann, M.A., jetzt Carl von Ossietzky-Universität Oldenburg, sowie unserem Kollegen Ansgar Schmitz für ihre Mitwirkung bei der Vorbereitung dieses Beitrages ganz herzlich danken.

1 Einleitung: Das <RM.net> - Genese, Projektkonzeption und Datenangebot

Bei dem „Informationsnetzwerk zur Geschichte des Rhein-Maas-Raumes“ (<RM.net>)² handelt es sich um ein fachbezogenes Netz, dessen Aufbau von der Deutschen Forschungsgemeinschaft seit 2001, zunächst im Rahmen des Programms „Informationsinfrastrukturen für netzbasierte Forschungskooperation und digitale Publikation“ und schließlich in dem neu geschaffenen Förderinstrument „Themenorientierte Informationsnetze“ ermöglicht wurde.³ Das kurz <RM.net> genannte Netzwerk stellt den Prototyp für ein themenorientiertes Fachinformationssystem in den historischen Wissenschaften dar, in dem Forschungsinformation unterschiedlicher institutioneller und fachwissenschaftlicher Herkunft gebündelt, nach einheitlichen Standards erschlossen und miteinander vernetzt wird. Ein solches Informationsangebot ist auf die Bedürfnisse kooperativ und international ausgerichteter Forschungsverbände ausgerichtet und hat zum Ziel, leistungsfähige, benutzerorientierte Strukturen und Werkzeuge zur Unterstützung der neuen wissenschaftlichen Kommunikations- und Publikationsplattformen zu schaffen sowie die technologischen Voraussetzungen für zeit- und ortsunabhängige Forschungskooperation zu verbessern (vgl. Eckelmann 1999, 493-501 / Eckelmann 2002, 174-179).

Der folgende Beitrag gewährt Einblick in die konkrete Projektpraxis, indem am Beispiel ausgewählter Informationskomponenten des Netzwerkes die notwendigen Arbeitsschritte zur Datenaufbereitung, inhaltlichen Tiefenerschließung und Vernetzung bis hin zur elektronischen Publikation vorgestellt werden. Weiterhin werden die Möglichkeiten des Zugangs und der Recherche, welche das Netz bietet, präsentiert. Um die Voraussetzungen und Rahmenbedingungen der Projektarbeit zu klären, wird zunächst auf Genese, Konzeption, Datenbestand und Ziele des Netzwerkes näher eingegangen.

2 Das Online-Angebot des <RM.net> ist unter der folgenden URL-Adresse zu finden: <http://www.RM.net.uni-trier.de> <11.07.05>.

3 Informationen zu den Förderprogrammen der DFG finden sich unter folgender URL-Adresse: http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/index.html <11.07.05>. Das neueste Positionspapier der DFG zum Thema „Elektronisches Publizieren“ ist im März 2005 erschienen und ist unter folgender URL-Adresse einzusehen: http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/pos_papier_elektron_publizieren_0504.pdf <11.07.05>.

1.1 Genese

Das an der Universität Trier angesiedelte Projekt geht zurück auf eine gemeinsame Initiative des Sonderforschungsbereichs 235 „Zwischen Rhein und Maas“⁴ und des Kompetenzzentrums für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften in den ausgehenden 1990er Jahren. In dem von der Deutschen Forschungsgemeinschaft geförderten interdisziplinären Forschungsverbund des SFB 235 wurden über einen Zeitraum von fünfzehn Jahren in mehr als dreißig Teilprojekten aus sechs Fächern Primärquellen und Forschungsdaten zur politisch-herrschaftlichen, wirtschaftlichen und kulturell-Entwicklung des Rhein-Maas-Raumes von der Spätantike bis in das 19. Jahrhundert zusammengetragen und ausgewertet.

Forschungsinformationen dieser Art, d.h. Forschungsinformationen, die in zeitlich befristeten Arbeitsvorhaben erhoben werden, sind in der Regel nach Projektabschluss für die weitere wissenschaftliche Forschung nicht mehr verfügbar. Nach wenigen Jahren können diese, da in proprietären Datenformaten auf Disketten oder CDs mit nur begrenzter Haltbarkeit gespeichert, nicht mehr gelesen werden. Um diesem Wissensverlust vorzubeugen und zugleich Datenbestände von hoher wissenschaftlicher und geprüfter Qualität für die Nachnutzung bereitzustellen, erarbeitete der SFB eine Konzeption für ein raumbezogenes Informationsnetzwerk, die auf dem konsequenten Einsatz internationaler Standards für die langfristige Datenhaltung und den orts- und zeitungebundenen Datenaustausch beruht.⁵

1.2 Datenbestand und Projektkonzeption

Umfangreiche SFB-Datenbestände aus unterschiedlichen Teilprojekten stehen inzwischen im <RM.net> elektronisch zur Verfügung. Weitere wichtige Anliegen waren, das Informationsnetzwerk für Daten externer Partner zu öffnen und die vom SFB begründete Zusammenarbeit im organisatorischen Rahmen eines Institutionen übergreifenden, offenen Forschungsnetzwerkes fortzuführen und auszubauen. Mittlerweile gehören dem Kooperationsprojekt zweiundzwanzig wissenschaftliche Einrichtungen an. Darüber hinaus steht das Projekt mit weiteren Institutionen aus dem Bibliotheks-, Archiv- und Verlagswesen in Kontakt. So sind derzeit zusammen mit den Datenbeständen der Kooperationspartner dreißig Informationseinheiten mit ca. 36.800 Einzeleinträgen als abgeschlosse-

4 Informationen zur Konzeption des Sonderforschungsbereichs 235 insgesamt und zu den Einzelprojekten sind über die Homepage des SFB 235 unter der folgenden URL-Adresse abrufbar: <http://www.uni-trier.de/infos/sfb235/sfb235.htm> <11.07.05>.

5 Zu Veröffentlichungen des Projektes vgl. Minn/Rapp 2003, 311-315 / Rapp 2003, 20-21 / Minn/Neumann/Rommelfanger 2003 (im Druck), elektronisch bereitgestellt unter der URL-Adresse: http://gepc189.uni-trier.de/RMnet/pdf/RMnet_hist2003.pdf <11.07.05>.

ne elektronische Publikationen im Netz zugänglich. Weitere Einheiten befinden sich für die Publikation in Vorbereitung.

Zudem besteht eine wichtige Aufgabe eines Informationsnetzwerkes darin, eine systematische Vorstrukturierung des Informationsangebots vorzunehmen, die den fach-spezifischen Anforderungen der Informationsbereitstellung und zudem den speziellen Interessen der potentiellen Nutzergemeinde entspricht. Das Informationsangebot des <RM.net> ist deshalb in vier Foren gegliedert:

Das Forschungsforum dient der Bereitstellung von Fachinformationen in Form von Bibliographien und Literaturlisten zu ausgewählten Themen der Geschichte des Rhein-Maas-Raumes.⁶ Im Themenforum werden orts- und regionenbezogene Kataloge und Lexika nach einheitlichen Erschließungsprinzipien kategorisiert und in einem übergreifenden elektronischen Verbund von Nachschlagewerken für das Rhein-Maas-Gebiet zusammengeführt. Im Kartenforum werden Einzelkarten und Kartenserien sowie ausgewählte Kartenwerke mit Bezügen zur rhein-maas-ländischen Geschichte online bereitgestellt. Im Quellenforum werden Primärquellen – Originalurkunden, Regesten und Rechtssatzungen – als digitale Editionen angeboten.

1.3 Zielsetzungen

Ziel des in der Pilotphase auf vier Jahre angelegten <RM.net>-Projektes ist es, ein fachbezogenes Informationsnetzwerk zur Geschichte des Rhein-Maas-Raumes aufzubauen und dabei Modelllösungen für ein raumbezogenes, themenorientiertes Informationssystem für die europäische Regionalgeschichte zu entwickeln. Zu diesen prototypischen Lösungen, die nach Projektabschluss zur Nachnutzung für vergleichbare Projektvorhaben zur Verfügung gestellt werden und die derzeit erarbeitet werden, gehören:

1. standardisierte Auszeichnungssysteme für ein breites Spektrum zentraler Textgattungen der Geschichtswissenschaft, welche im <RM.net> in fünf Dokumentklassen⁷ zusammengefasst werden, die den Datenaustausch zwischen Informationssystemen erleichtern,
2. EDV-Werkzeuge für Register zur indexbasierten, integrierten Recherche, wie z.B. hierarchisch strukturierte Raumindizes, mit denen man parallel

6 Das Angebot des Forschungsforums soll künftig um eine Sammlung von Links zu Forschungseinrichtungen, Institutionen, Museen, Zeitschriften etc. erweitert werden, die sich mit der Geschichte des Rhein-Maas-Raumes beschäftigen.

7 Die fünf Hauptgruppen der Dokumentklassifikation des <RM.net>: Dokumentgruppe I: Orts- und Regionen bezogene Nachschlagewerke, Datensammlungen / Dokumentgruppe II: Quelleneditionen unterschiedliche Gattungen / Dokumentgruppe III: Bibliographien / Dokumentgruppe IV: Kartenwerke, Einzelkarten, Kartenserien, kommentierte Atlaswerke / Dokumentgruppe V: Personen- und Institutionen bezogene Informationseinheiten.

über das Gesamtnetzwerk bzw. in ausgewählten Publikationen recherchieren kann,

3. standardisierte Benutzeroberflächen zur Darstellung elektronischer Publikationen gleichen Texttyps, z.B. für die elektronische Publikation von Urkunden oder Kartenwerken.

Diese Musterlösungen sollen dazu beitragen, den Aufbau weiterer Informationssysteme zu unterstützen. Welche konkreten Arbeitsschritte anfallen, um Online-Publikationen, wie sie das <RM.net> zusammenführt, zu realisieren, wird im Folgenden erläutert.

2 Arbeitsschritte: Von der WORD-Datei des Forschungsprojektes zum SGML-Dokument

Das <RM.net>-Projekt steht vor der Aufgabe, zumeist in proprietären Datenformaten vorliegende Forschungsinformation verschiedener fachlicher und institutioneller Herkunft und unterschiedlichen Dokumenttyps aufzubereiten, zu erschließen und miteinander in einem Informationsverbund zu vernetzen. Diese durch die Heterogenität des Datenmaterials geprägten Ausgangsbedingungen spiegeln sich im Ablauf der Projektarbeit wider. Die wissenschaftlichen Daten werden in ein SGML-/XML-konformes Format überführt, welches eine langfristige und Plattform unabhängige Archivierung sichert sowie einen Datentransfer zwischen unterschiedlichen Systemen ermöglicht. Auf die notwendigen Arbeitsschritte von der Dokumentanalyse bis zur Gestaltung der Benutzeroberfläche wird nun einzugehen sein (Abb. 1).

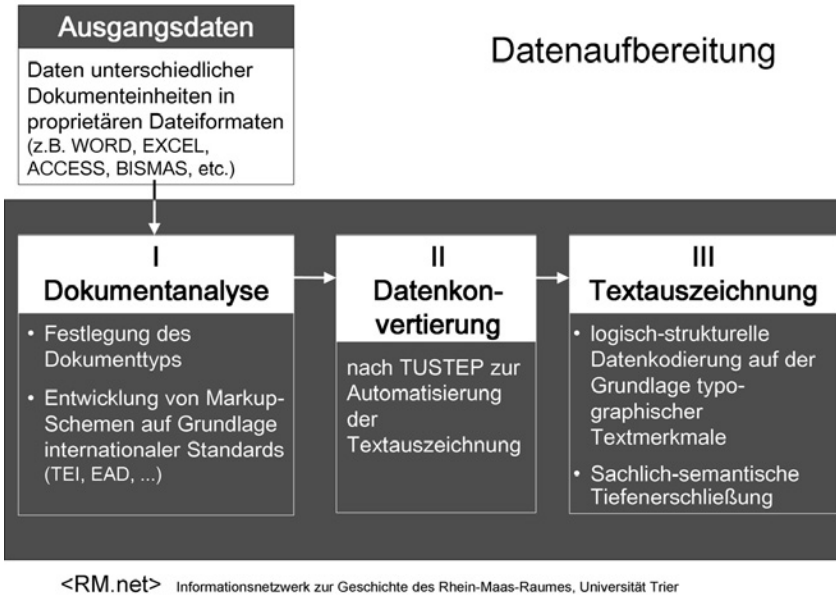


Abb. 1: Datenaufbereitung

2.1 Arbeitsschritte I - III: Von der Dokumentanalyse bis zur Tiefenerschließung der Inhalte

Die Projektarbeit beginnt mit der Dokumentanalyse. Sie ist Grundlage für die Festlegung des Dokumenttyps und die Entwicklung von Auszeichnungsschemata zur logisch-strukturellen und sachlich-semantischen Tiefenerschließung des jeweiligen Textes. Das im <RM.net>-Projekt aufbereitete Material erfordert Markup-Systeme für zentrale Dokumenttypen historischer Forschung. Für die Kodierung dieser Dokumenttypen sind auf der Grundlage einer eingehenden Analyse der inhaltlich-formalen Struktur und der Layout-Gestaltung der Dateneinheiten einheitliche Auszeichnungsrichtlinien festzulegen. Dabei wird auf die internationalen Vorgaben der Text Encoding Initiative (TEI)⁸ zurückgegriffen, die speziell für die Geisteswissenschaften Auszeichnungsregeln entwickelt hat. Die dort gefundenen Kodierungsschemata müssen allerdings in den meisten Fällen modifiziert und auf die Bedürfnisse der einzelnen Dokumenttypen des

8 Vgl. zu Informationen über TEI deren Homepage unter folgender URL-Adresse: <http://www.tei-c.org/<11.07.05>>.

<RM.net> angepasst werden, da für diese bisher kein SGML-konformes Markup vorliegt. Nachdem der Dokumenttyp bestimmt und die wesentlichen Strukturmerkmale des zu bearbeitenden Textes ermittelt worden sind, kann ein Markup-Schema festgelegt werden.

In einem zweiten Arbeitsschritt werden die in proprietären Software-Formaten vorliegenden Ausgangsdaten in vorstrukturierte TUSTEP-Dateien⁹ überführt. Mit diesem über TUSTEP-Programmroutinen gesteuerten Konvertierungsvorgang wird zugleich die Datenkodierung in Phase III eingeleitet.

Die Textauszeichnung verläuft in mehreren Stufen: Zunächst erfolgt eine weitgehend automatisierte Vorauszeichnung. Hierbei werden die in der Layout-Analyse ermittelten typographischen TUSTEP-Kennungen über spezielle Programmskripte in SGML-/XML-konforme Kodierungen überführt, die die logisch-formale Struktur des Textes abbilden. Nach Prüf- und Korrekturarbeiten beginnt die halbautomatisch-manuelle Feinauszeichnung mit sachlich-semantischem Markup, also die eigentliche Tiefenerschließung der Inhalte. Die konkreten Arbeitsgänge zur Datenaufbereitung werden am Beispiel des Dokumenttyps „Ortskatalog“ erläutert (Abb. 2).

Das Beispiel zeigt einen Eintrag aus einem klar strukturierten, nach Kategorien aufgebauten Katalog zur Geschichte der Leprosorien im Rheinland (aus: Uhrmacher 2000, 49), der insgesamt 207 Ortsartikel einheitlicher Struktur umfasst. Auf den jeweiligen Ortsnamen und seine Verwaltungszugehörigkeit folgen die nach den zuvor festgelegten Kriterien gegliederten, hier über einen Zahlenschlüssel kodierten Sachinformationen – der eigentliche Artikel-Inhalt – mit abschließenden Literaturhinweisen. Diese Einheit des Dokumenttyps I hat einen klaren und übersichtlichen Artikelaufbau, der sich in eine geographische, inhaltliche und bibliographische Komponente unterteilen und dementsprechend in das Markup für ortsbezogene Daten umsetzen lässt.

9 Das Textverarbeitungsprogramm TUSTEP ist ein Produkt des Zentrums für Datenverarbeitung der Universität Tübingen. Es ist ein Werkzeug zur wissenschaftlichen Bearbeitung von Textdaten, wie es insbesondere die Geisteswissenschaften benötigen, das nun seit mehr als 25 Jahren im Einsatz ist und ständig weiterentwickelt wird. Es enthält Bausteine für alle Arbeitsgänge beim Computer gestützten wissenschaftlichen Umgang mit Textdaten, von der Erfassung über Abfrage, Analyse, Sortierung, Speicherung und andere Arten der Verarbeitung von Textdaten bis zur Ausgabe in gedruckter oder elektronischer Form. Vgl. die URL-Adresse: <http://www.zdv.uni-tuebingen.de/tustep/> <11.07.05>.

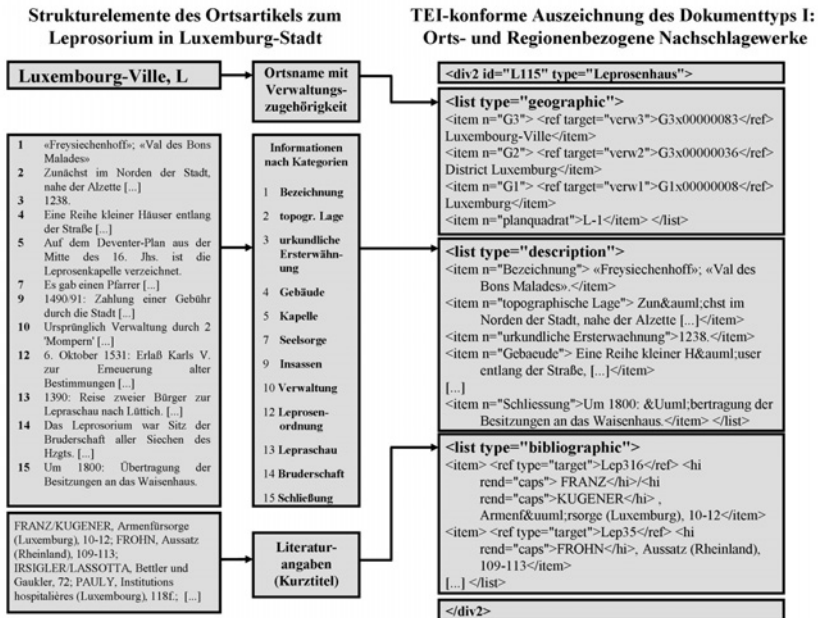


Abb. 2: Strukturelemente und Markup für den Dokumenttyp Ortskatalog

2.2 Arbeitsschritte IV - VI: Erstellen der Internet-Publikation

Die drei weiteren Arbeitsphasen können wie folgt zusammengefasst werden (Abb. 3): Mit der Anbringung inhaltlich-semantischen Markups in Phase III werden die Grundlagen für den Aufbau der graphischen Benutzerschnittstellen gelegt. Auf der Basis der Metadaten-Klassifikationen werden mit hierarchisch aufgebauten Raum-, Personen- und Sachthesauri komplexe Rechercheinstrumente erstellt. Sie bilden das Grundgerüst für die Verschlagwortung des weiteren Datenbestandes und werden als dynamisches Erschließungsinstrument ausgebaut, das mit der Integration jeder neuen Informationseinheit umfassender, differenzierter und – durch interne Vernetzung der Informationseinheiten – komplexer wird.

Mittels Tcl-/Tk-basierter Suchroutinen eröffnet sich so ein differenzierter Zugang zu den Informationseinheiten des Netzwerkes, die in einer MySQL-Datenbank abgelegt sind. Die auf diese Weise entstehenden Indices gehören zu den zentralen, dem Nutzer im <RM.net> zur Verfügung stehenden Zugangsmöglichkeiten in die Dateneinheiten.

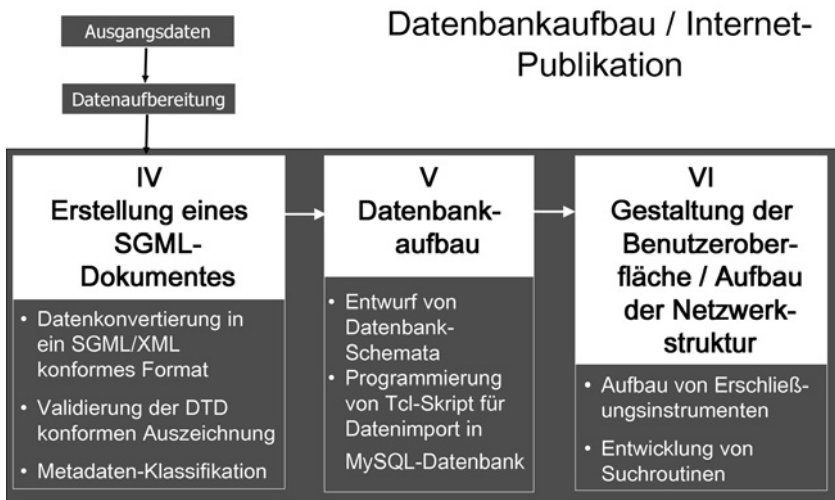


Abb. 3: Arbeitsschritte zum Erstellen der Internetpublikation

3 Das <RM.net> aus der Nutzerperspektive: Möglichkeiten des Zugangs und der Recherche

Neben komplexen Zugangs- und Recherchemöglichkeiten steht dem Nutzer als erster und auch einfachster Zugang in die digitale Publikation die Option offen, vergleichsweise wie in einem „Buch“ zu blättern und die einzelnen Dokumente in ihrer genuinen Abfolge kennen zu lernen. Durch das Anklicken der Illustration auf der Startseite der jeweiligen Edition gelangt man zum ersten Datensatz und kann von hier aus weiterblättern, sich in der Edition vor- und zurückbewegen.

3.1 Die Indices des <RM.net>

Das <RM.net> bietet für jede Einheit globale und Einheiten spezifische Zugangsmöglichkeiten an. Für das Informationsnetzwerk insgesamt wird ein Raum-, Personen- und Sachindex angeboten. Zum Aufbau der Indices werden die entsprechenden Informationselemente – Personennamen, Ortsnamen, Sachbezüge – mit eindeutigen Kennungen versehen und in gesonderten Thesaurusdateien gespeichert, über die die Erstellung der Netzwerk übergreifenden Suchinstrumente gesteuert wird.

Als Beispiel für einen solchen globalen Index seien der hierarchisch gegliederte Raumindex und der Personenindex herausgegriffen. Der Raumindex ist derzeit in drei Ebenen (Staat – Verwaltungseinheit – Ort) unterteilt und weist zurzeit etwa 3200 Orte in achtzehn Ländern aus. So werden unter „Deutschland“ die Bundesländer aufgeführt, zu denen im <RM.net> Informationen vorhanden sind. Wählt man ein Bundesland aus, erhält man eine nach Foren sortierte Auflistung aller Einheiten, die diesem Bundesland zugeordnet sind. Von dieser Trefferliste ausgehend ist der Zugang in die einzelnen Publikationen möglich. Ein weiteres zentrales Rechercheinstrument – insbesondere für die digitalen Quelleneditionen – ist der Personenindex. Hier sind bisher ca. 4660 Personen indiziert und mit den entsprechenden Einheiten vernetzt. Über dieses Rechercheinstrument kann sich der Nutzer zielgerichtet Urkunden ausgeben lassen, die eine Verbindung zu einer bestimmten historischen Persönlichkeit aufweisen.¹⁰

3.2 Dokumentgruppen und Einheiten spezifische Rechercheinstrumente

Neben den globalen Indices stehen für jede Dateneinheit darüber hinaus individuelle Suchanfragen zur Verfügung, zu deren Realisierung inhaltlich-strukturelle Merkmale herangezogen werden. Eine solche, gerade für die Dokumentgruppe „Quellenedition“ relevante Rechercheoption ist beispielsweise die Suche über das Ausstellungsdatum. Hiermit bietet sich dem Nutzer die Möglichkeit, gezielt Dokumente aus einem bestimmten Jahr zu recherchieren.¹¹ Weiterhin gibt es für die Bibliographien einen Autorenindex, der mittlerweile etwa 6.440 Autoren umfasst, sowie eine Volltextsuche. Beides ist global für alle Bibliographien als Gesamtindex verfügbar sowie eingeschränkt auf die jeweilige Einheit.¹²

10 Sucht man beispielsweise Dokumente zu Heinrich II. von Virneburg, Erzbischof von Köln, so werden im Quellenforum in der Edition „Mittelfränkische Urkunden, 1300 bis 1330“ fünfunddreißig Dokumente gefunden. Vgl. zum Suchergebnis die URL-Adresse: <http://gepc189.uni-trier.de/cgi-bin/RMnetIndex.tcl?hea=&nav=su&for=&cnt=supi&page=1&mode=persdetail&typ=&id=890<11.07.05>>.

11 Sucht man beispielsweise in der Edition „Volkssprachige Originalurkunden aus Metzger Archiven bis zum Jahr 1270“ nach Dokumenten aus dem Jahr 1250, werden 14 Regesten antizipiert. Vgl. zum Suchergebnis die URL-Adresse: <http://gepc189.uni-trier.de/cgi-bin/RMnetIndex.tcl?hea=qf&nav=su&for=qpitzreg&cnt=sus&mode=ausdat&such=1250<11.07.05>>.

12 Neben den hier vorgestellten Indices verfügt das <RM.net> weiterhin über einen Projekt- und einen Sachindex. Diese globalen Indices wie auch die Volltextsuche können über die Foren übergreifende Suchseite des Netzes unter folgender URL-Adresse angesteuert werden: <http://gepc189.uni-trier.de/cgi-bin/RMnetIndex.tcl?hea=&for=&nav=su&cnt=sus<11.07.05>>.

Neben diesen, für eine Dokumentgruppe spezifischen Recherchen bietet auch jede Einheit Ansatzpunkte für ganz individuelle Suchanfragen. Die Edition „Luxemburger Bevölkerungs- und Steuerlisten von 1805/06 und 1852“ hat z.B. eine Vielzahl solcher Zugänge aufzuweisen. So kann man sich hier, um nur eine für sozialhistorische Fragestellungen wichtige Option herauszugreifen, für das Jahr 1852 alle verwitweten Personen in Luxemburg auflisten lassen. Man erhält eine Liste von 733 Personen, die dieses Kriterium erfüllen.¹³

3.3 Vernetzungsstrukturen

Der Mehrwert elektronischen Publizierens liegt unter anderem in der Möglichkeit begründet, sehr komplexe Vernetzungsstrukturen aufzubauen. So können Verknüpfungen innerhalb der einzelnen Einheiten, z.B. Querverweise zwischen einzelnen Dokumenten oder zwischen Dokument und Karte, genauso realisiert werden wie Verlinkungen über die Forengrenzen hinaus. Eine gängige Form der Foren übergreifenden Netzwerkstruktur stellt die Vernetzung zwischen den Publikationen in Themen-, Karten- und Quellenforum mit den entsprechenden Bibliographien im Forschungsforum dar. So kann z.B. aus dem Kartenarchiv des SFB 235 auf den entsprechenden Eintrag in der kommentierten Bibliographie des SFB 235 zugegriffen werden.¹⁴

4 Zusammenfassung und Ausblick

Themenorientierte Informationsnetzwerke wie das <RM.net> bieten ein digitales Informationsangebot, das speziell auf die Bedürfnisse und Interessen von Forschergruppen abgestimmt ist. Derartige Informationsnetze dienen im Vergleich mit den Virtuellen Fachbibliotheken primär der Bereitstellung und Vernetzung forschungsrelevanter digitaler Fachinformation. So stellt das <RM.net> eine themenspezifische Informationsumgebung für die regionalhistorischen Wissenschaften dar. Mit den hierarchisch und alphabetisch strukturierten Thesauri und einer Vielzahl weiterer Optionen des Zugangs werden dem Nutzer elaborierte und fachspezifische Rechercheinstrumente zur präzisen, integrierten Informationsermittlung bereitgestellt.

¹³ Vgl. zum Suchergebnis die URL-Adresse:

<http://gepc189.uni-trier.de/cgi-bin/RmnetIndex.tcl?hea=tf&&nav=su&for=qfluxewst&cnt=sus&mode=famStat&id=verwitwet&cbx=1852<11.07.05>>.

¹⁴ Vgl. beispielsweise im elektronischen Kartenarchiv des SFB 235 die Karte „Die urbanen Zentren am Mittelrhein um 1300“ von Monika Escher und Frank G. Hirschmann. Unterhalb der Kartenansicht wird über den Verweis „publiziert in“ der Link in die kommentierte Bibliographie des SFB 235 angeboten: <http://gepc189.uni-trier.de/cgi-bin/RMnetIndex.tcl?hea=kf&nav=&for=karch&cnt=karch&id=B2S218<11.07.05>>.

Das Leistungsspektrum der Netzwerke ist künftig um eine weitere Komponente zu erweitern. Die Informations- und Publikationsplattform wird um eine netzbasierte Arbeitsumgebung zu ergänzen sein, die bereits unmittelbar im Forschungsprozess die Aufnahme von Forschungsdaten, ihre Erschließung und ihre tiefer gehende Analyse unterstützt. Erste Schritte in diese Richtung unternimmt das <RM.net> in Verbindung mit dem ebenfalls an der Universität Trier angesiedelten Sonderforschungsbereich 600 „Fremdheit und Armut“ bei der Konzeption und Programmierung eines Projekt übergreifenden EDV-Systems zur Organisation kollaborativen verteilten wissenschaftlichen Arbeitens in einem Verbund von 16 Teilprojekten.¹⁵

Voraussetzung für den Transfer der prototypisch am <RM.net> entwickelten Lösungen auf andere Forschungs- und Informationsverbünde ist die offene Konzeption des Netzwerkes. Der konsequente Einsatz internationaler SGML-/XML-basierter Standards ist dabei von besonderer Bedeutung, da eine wichtige und dringliche Aufgabe für die zukünftige Entwicklung der Informationsnetze darin bestehen wird, die technologischen Grundlagen für ihre dauerhafte Fortführung und ihre Integration in übergeordnete Informations-Infrastrukturen zu schaffen. In diesem Zusammenhang stellt auch die Vernetzung zwischen virtuellen Fachbibliotheken und themenorientierten Fachinformationssystemen ein wichtiges Aufgabenfeld dar.

5 Literaturangaben

DFG-Positionspapier (März 2005): Elektronisches Publizieren:

http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/pos_papier_elektron_publizieren_0504.pdf <11.07.05>.

Eckelmann, S. (1999): Neue Kommunikations- und Publikationsinfrastrukturen auf der Basis digitaler Medien. Schlußfolgerungen aus den Multimedia-Rundgesprächen der Deutschen Forschungsgemeinschaft. In: Zeitschrift für Bibliothekswesen und Bibliographie 46 (1999), S. 493-501.

Eckelmann, S. (2002): Die Archivförderung der DFG – Schwerpunkte und Perspektiven. In: Menne-Haritz, A. (Hrsg.): Online-Findbücher, Suchmaschinen und Portale. Beiträge des 6. Archivwissenschaftlichen Kolloquiums der Archivschule Marburg, Marburg (Veröffentlichungen der Archivschule Marburg 35), S. 173-181.

Minn, G.; Rapp, A. (2003): The Information and Reference Network for the History of the Rhine-Meuse-Area. An Area-Oriented Subject Information

¹⁵ Informationen zur Konzeption des Sonderforschungsbereichs 600 insgesamt und zu den Einzelprojekten sind über die Homepage des SFB 600 unter der folgenden URL-Adresse abrufbar: <http://www.sfb600.uni-trier.de/> <11.07.05>.

- System for the Humanities. In: Burch, T.; Fournier, J.; Gärtner, K.; Rapp, A. (Hrsg.): Standards und Methoden der Volltextdigitalisierung. Beiträge des Internationalen Kolloquiums an der Universität Trier, 8./9. Oktober 2001, Mainz (Akademie der Wissenschaften und der Literatur Mainz, Abhandlungen der Geistes- und sozialwissenschaftlichen Klasse 9), S. 311-315.
- Minn, G.; Neumann, S.; Rommelfanger, Y. (2003): Das Informationsnetzwerk zur Geschichte des Rhein-Maas-Raumes (RM.net). Ein raum- und themenorientiertes Fachinformationssystem für die europäische Regionalgeschichte. In: Hohls, R.; Nippel, W. (Hrsg.): .hist2003. Geschichte und neue Medien. Tagung des Kooperationsverbundes clio-online, Berlin, 9.–11.4.2003 (im Druck), elektronisch bereitgestellt:
http://gepc189.uni-trier.de/RMnet/pdf/RMnet_hist2003.pdf <11.07.05>.
- Uhrmacher, M. (2000): Leprosorien in Mittelalter und früher Neuzeit, Köln (Geschichtlicher Atlas der Rheinlande, 7. Lieferung, Karte und Beiheft VIII.5).
- Rapp, A. (2003): Taalkundige, diplomatische en historische Apecten van een gezamenlijk onderzoeksproject. In: Archievenblad 107 (2003), Nr. 3, S. 20-21.

Angaben zu den Autorinnen

Dr. Gisela Minn, Anke Susanne Hoffmann, Yvonne Rommelfanger, M.A.
<RM.net> - Informationsnetzwerk zur Geschichte des Rhein-Maas-Raumes
Universität Trier
Fachbereich III / Geschichte
54286 Trier

Dr. Gisela Minn

Wissenschaftliche Mitarbeiterin im Sonderforschungsbereich 600 „Fremdheit und Armut. Wandel von Inklusions- und Exklusionsformen von der Antike bis zur Gegenwart“ an der Universität Trier, Leiterin des Projektes „Informationsnetzwerk zur Geschichte des Rhein-Maas-Raumes“ in Verbindung mit Prof. Dr. Lutz Raphael und Dr. Thomas Burch, Universität Trier.

Anke Susanne Hoffmann

Wissenschaftliche Mitarbeiterin im Projekt „Informationsnetzwerk zur Geschichte des Rhein-Maas-Raumes“ und Doktorandin im Fach Germanistik / Neuere deutsche Literaturwissenschaft an der Universität Trier.

Yvonne Rommelfanger, M.A.

Wissenschaftliche Mitarbeiterin im Projekt „Informationsnetzwerk zur Geschichte des Rhein-Maas-Raumes“ und Doktorandin im Fach Neuere / Neueste Geschichte an der Universität Trier.

PsychSpider – Erfahrungen aus dem Betrieb einer spezialisierten Suchmaschine

Christiane Baier, Peter Weiland

Zentrum für psychologische Information und Dokumentation

Zusammenfassung

Spezialisierte Suchmaschinen werden vor dem Hintergrund der Unzufriedenheit mit den Ergebnissen universeller Suchmaschinen als Alternative für die Suche nach Fachinformationen im WWW angepriesen. In diesem Beitrag wird untersucht, inwieweit spezialisierte Suchmaschinen eine sinnvolle Ergänzung für die Recherche von Fachinformationen darstellen können und welche Maßnahmen zur Attraktivitätssteigerung geeignet sind.

Das ZPID betreibt bereits seit mehr als vier Jahren eine spezialisierte Suchmaschine für das Fach Psychologie. Der Index enthält Webseiten von mehr als 370 Servern u.a. von psychologischen Instituten und Lehrstühlen, die Angebote von Fachgruppen und Fachgesellschaften und elektronische Zeitschriften und Volltexte. Die kostenpflichtige Datenbank PSYINDEX kann ebenfalls durchsucht werden, wobei ein komfortabler Einstieg in die Web-Rechercheoberfläche der Datenbank gegeben ist.

Aufbauend auf den Erfahrungen beim Betrieb von PsychSpider wird beschrieben, welche technischen (und auch redaktionellen) Probleme beim Erstellen von Suchkollektionen für eine spezialisierte Suchmaschine bewältigt werden müssen. Hierzu zählen beispielsweise die Indexierung nicht-standardkonformer Datenformate oder auch das Spidern von Seiten, die von Content Management Systemen ausgeliefert werden. Zusätzlich wird das Nutzerverhalten mit der spezialisierten Suchmaschine PsychSpider über einen Zeitraum von 9 Monaten analysiert. Dabei wird auch ein Vergleich zur Nutzung universeller Suchmaschinen hergestellt. Untersucht werden hier vor allem die verwendeten Suchbegriffe und die Benutzung.

1 Defizite von Universalsuchmaschinen

Universalsuchmaschinen wie Google, MSN oder Yahoo ermöglichen einem Nutzer den problemlosen Zugriff auf eine Vielzahl von Informationen im Web. Der Index von Google umfasst ca. 8.000.000.000 Seiten, MSN gibt

5.000.000.000 als Indexgröße an¹. Dennoch bieten diese Suchmaschinen für verschiedene Anfragen nicht immer die besten Suchergebnisse. Gerade Nutzer im akademischen Bereich klagen über die mangelnde Abdeckung von „Deep Web“-Angeboten. Diese Angebote sind in der Regel durch Crawler nicht automatisch zu erschließen. Nach Angaben von Bergmann (2000) ist der Anteil der Informationen im „Deep Web“ um ein Vielfaches größer als im „Surface Web“. Darüber hinaus sind „Deep Web“-Informationen in der Regel höherwertiger (vgl. Bergmann 2000). Die Suchmaschinenbetreiber bemühen sich – sofern sich der Aufwand ihrer Meinung nach lohnt – um „Deep Web“-Angebote, dennoch ist der Anteil spezialisierter Angebote in den Indexen von Universalsuchmaschinen eher gering².

Studien von searchenginewatch.com zeigen, dass einige wenige Suchportale wie Google, Yahoo oder MSN (siehe Abbildung 1) von der Mehrheit der Internetnutzer für die Recherche benutzt werden. In jüngerer Zeit wächst daher die Besorgnis, dass die Bündelung des Informationszugangs in den Händen weniger Monopolisten zu einer Behinderung der Informationsversorgung führen kann, beispielsweise durch kommerzielle Interessen der Suchmaschinenanbieter, sich die Aufnahme in den Index und ein gutes Ranking bezahlen zu lassen oder auch politisch motivierte Zensurbestrebungen³.

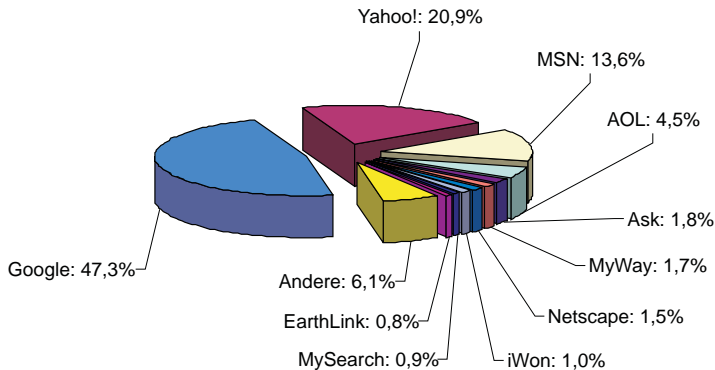


Abb. 1: Marktanteil Suchmaschinen in den USA - März 2005
(Quelle: Nielsen NetRatings Search Engine Ratings, 22.04.2005,
URL <http://searchenginewatch.com/reports/article.php/2156451> am
13.05.2005)

- 1 Vgl. Search Engine Size Wars V Erupts, online verfügbar unter <http://blog.searchenginewatch.com/blog/041111-084221> am 13.05.2005
- 2 Yahoo! gibt an, dass sein Index zu 99% aus „Surface Web“-Angeboten besteht.
- 3 Die Suchergebnisse von Yahoo und Google werden beispielsweise in der Volksrepublik China gefiltert, vgl. <http://www.heise.de/tp/r4/artikel/13/13729/1.html> (online verfügbar am 24.05.2005)

Ein weiteres Problem, mit dem Universalsuchmaschinen mehr und mehr zu kämpfen haben, ist die Qualität der indextierten Seiten. Crawler können zum einen die abgerufenen Seiten nicht auf Qualität prüfen, dies ist intellektuell aufgrund der schieren Menge an Seiten unmöglich, zum anderen versuchen Site-Betreiber durch verschiedene, nicht immer legale Maßnahmen ihre Seiten in den Rankings nach oben zu treiben. Hier spricht man auch von Suchmaschinen-Spam. Die Suchmaschinenbetreiber stehen hier in einem regelrechten Wettbewerb mit den Sitebetreibern.

Universalsuchmaschinen bieten mit ihren riesigen Indexen zwar Zugang zu einer unerschöpflichen Menge an Informationen, ihnen fehlt allerdings der Kontext der Fragestellung des jeweiligen Benutzers. Eine einfache Suche nach dem Begriff „Adler“ sowohl in Google als auch in PsychSpider⁴ zeigt, dass der an Psychologie interessierte Nutzer mit der einfachen Anfrage „Adler“ bei der spezialisierten Suchmaschine PsychSpider eher ans Ziel gelangt, nämlich an Informationen über den Individualpsychologen Alfred Adler (siehe Abbildung 2).

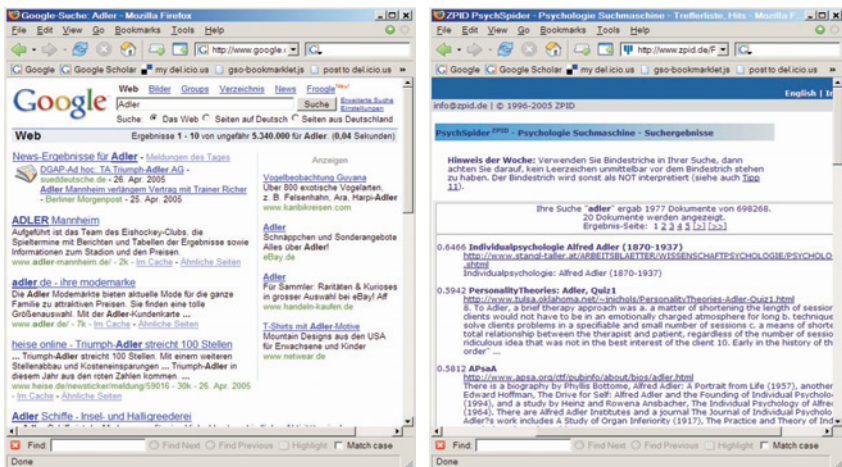


Abb. 2: Suche nach „Adler“ bei Google und bei PsychSpider

4 Spezialisierte Suchmaschine für die Psychologie des Zentrum für psychologische Information und Dokumentation (ZPID)

2 Spezialisierte Suchmaschinen

Spezialisierte Suchmaschinen oder Suchdienste bieten dem Nutzer im Gegensatz zu den allgemeinen Suchmaschinen wie Google, MSN oder Yahoo nur einen kleinen Ausschnitt der im WWW verfügbaren elektronischen Informationen an. Sie treffen bereits vor der Indexierung eine Vorselektion, da nur diejenigen Dokumente in den Index aufgenommen werden, die bestimmten Kriterien entsprechen. Es gibt mittlerweile zu fast jeder Lebenssituation oder jedem Thema spezialisierte Suchmaschinen, das Angebot ist nahezu unüberschaubar⁵.

2.1 PsychSpider als ein Beispiel für eine spezialisierte Suchmaschine

Das Zentrum für psychologische Information und Dokumentation betreibt seit 2001 PsychSpider als spezialisierte Suchmaschine für Web-Ressourcen in der Psychologie und angrenzender Bereiche. PsychSpider umfasst ca. 700.000 Dokumente von mehr als 370 Servern (Stand: Mai 2005). Neben den Kollektionen *Veranstaltungen, Fachgesellschaften und Fachgruppen, Bibliotheken, Fachinformationszentren, E-Journals, (Test-) Diagnostik, Institute* und *Web (allgemein)* sind die Datenbanken *PSYNDEX Lit & AV* und *PSYNDEX Tests* im Index enthalten, die Letztgenannten nur mit dem Titel des Dokuments in der Trefferansicht.

Jede URL, die PsychSpider indexiert, wird manuell auf Relevanz überprüft. Da über die eingegebenen Startpunkte (URLs) auch nicht-relevante Seiten in den Index gelangen können, wird dieser ebenfalls kontinuierlich manuell überprüft.

Die technische Basis von PsychSpider ist die Suchmaschine Verity K2 in der Version 5.

5 Vgl. <http://www.searchengineguide.com/> oder auch <http://www.completeplanet.com/>

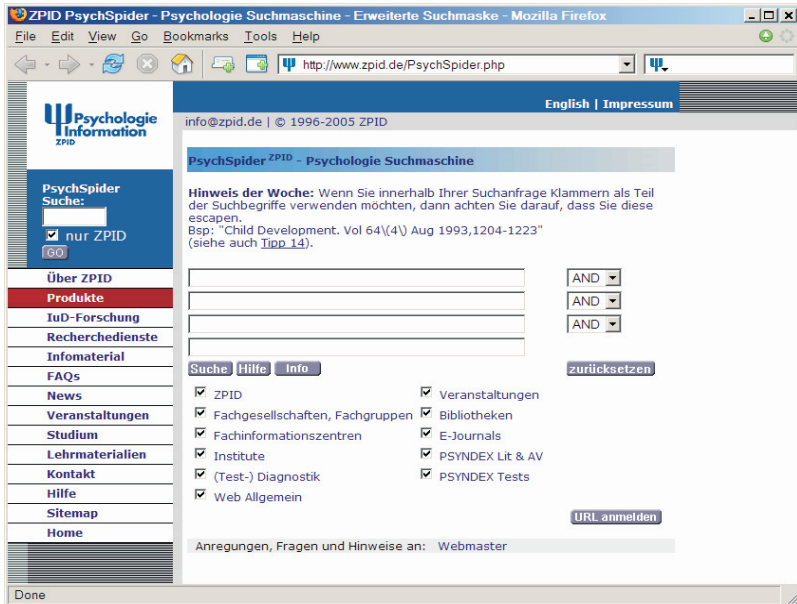


Abb. 3: Erweiterte Suchmaske von PsychSpider

2.2 Aufbau spezialisierter Suchmaschinen

Aus dem Hauptvorteil einer spezialisierten Suchmaschine – nämlich der Vorselektion relevanter Quellen – erwächst auch die Hauptschwierigkeit beim Aufbau der Kollektionen. Jede URL, die zum Index hinzugefügt wird, muss manuell auf Relevanz überprüft werden. Darüber hinaus muss sichergestellt werden, dass der Crawler tatsächlich nur relevante Seiten einer Site absucht. Eine Möglichkeit, dies zu erreichen, ist über Mustervergleich der URLs gegen zuvor festgelegte Werte.

Die zunehmende Verbreitung von Content Management Systemen erschwert die Filterung über die URL, da die jeweilige Organisationsstruktur nicht mehr in der Verzeichnisstruktur des Servers abgebildet ist. Hier kann erst während des Indexierens über die Einschließung bestimmter Regeln und Optionen eine Filterung vorgenommen werden. Im Betrieb führt dies zu einer erhöhten Last sowohl beim Suchmaschinenbetreiber als auch beim Informationsanbieter⁶.

6 Im Falle der psychologischen Fakultät einer großen deutschen Hochschule, die universitätsweit ein Content Management System einsetzt, bedeutet dies, dass alle Seiten der Universität heruntergeladen und indiziert werden müssten, um anschließend nur die psychologielevanten Seiten im Index zu belassen.

Neben der (manuellen) Erstellung eines spezialisierten Indexes, wie dies PsychSpider und Scirus von Elsevier tun, gibt es beim Aufbau spezialisierter Suchmaschinen auch die Möglichkeit des Setzens von Filtern auf einen allgemeinen Index. Diese Vorgehensweise wurde bei Google Scholar gewählt, d.h. Google Scholar bietet eine Untermenge des vollständigen Google Index, automatisch gefiltert nach relevanten Dokumenten für die Wissenschaft (Hamaker, 2005). Abbildung 4 illustriert die Vorgehensweisen beim Aufbau spezialisierter Suchmaschinen.

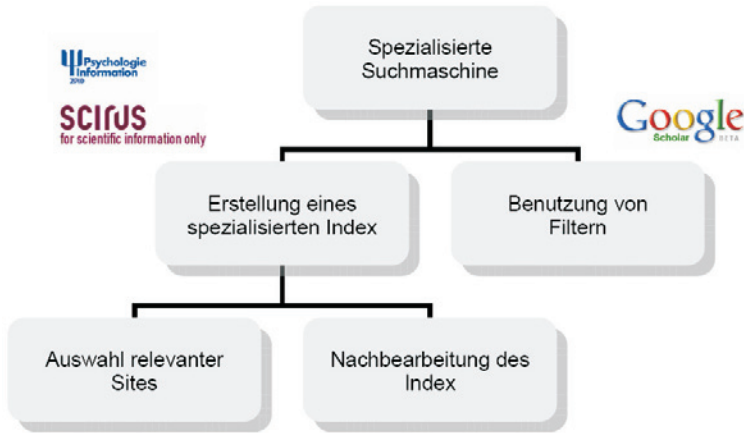


Abb. 4: Aufbau spezialisierter Suchmaschinen am Beispiel PsychSpider, Scirus und Google Scholar

2.3 Kriterien für die Kategorisierung spezialisierter Suchmaschinen

Die bekanntesten Beispiele für spezialisierte Suchmaschinen für die Wissenschaft sind Google Scholar und Scirus von Elsevier. Beide Suchdienste haben zum Ziel, eine möglichst vollständige Abdeckung aller Wissenschaftsdisziplinen anzubieten, wobei Scirus neben Fachartikeln auch einen redaktionell betreuten Index von sonstigen Webseiten enthält und erweiterte Filtermöglichkeiten – beispielsweise nach Fächern – bietet. Neben der Indexierung frei zugänglicher Quellen benutzen beide Crawler auch die Datenbanken der großen Verlage und Fachgesellschaften: Scirus kann sich hier des gesamten Elsevier-Spektrums bedienen, Google hat offensichtlich Abkommen u.a. mit der IEEE, der ACM, Macmillan und Wiley geschlossen (vgl. Jacso 2005).

Abbildung 5 vergleicht PsychSpider, Scirus und Google Scholar anhand der Kriterien *Neutralität*, *Themenspezifität*, *Qualität*, *Einbindung von „Deep Web“-Angeboten*, *Indexgröße*, *Genre* und *Zusatzangebote*.

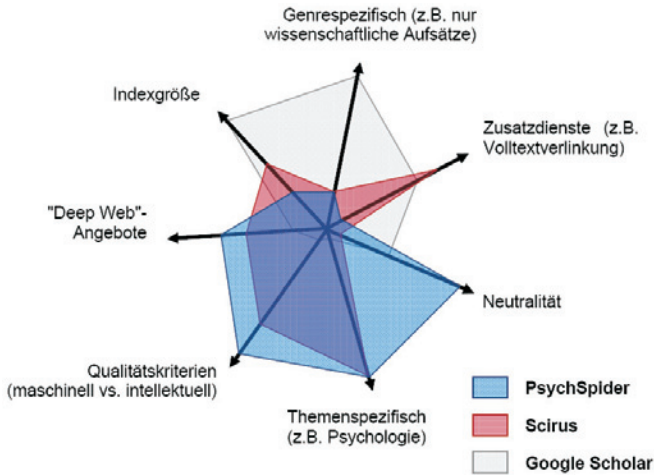


Abb. 5: Vergleich der Spezialsuchmaschinen PsychSpider, Scirus und Google Scholar anhand verschiedener Kriterien

3 Erfahrungen aus dem Betrieb

3.1 Einfluss der Seitengestaltung auf die Indexierung

Die Betreiber von Websites sind sich häufig nicht bewusst, welchen Einfluss die Gestaltung ihrer Seiten auf die Indexierbarkeit durch Crawler hat. Aus der Betreibersicht einer Suchmaschine fallen hier insbesondere folgende Punkte auf, die bei der Indexierung eines Angebotes zu beachten sind.

- **Verwendung von JavaScript in der Navigation**

Die meisten Crawler haben Schwierigkeiten beim Folgen von Links in Menüs, die durch JavaScript aufgebaut werden. Häufig kann daher nur die Startseite eines Angebotes indexiert werden, Folgeseiten können nicht in den Index aufgenommen werden.

Diese Aussage gilt auch für Navigationen, die über Flash oder Java Applets aufgebaut werden.

- **Session Cookies**

Viele Website-Betreiber verfolgen die Pfade ihrer Benutzer im Angebot über sogenannte Session Cookies. Dies führt in der Regel dazu, dass ein Crawler nicht mehr zuordnen kann, welche Seiten er bereits heruntergeladen hat.

- **Nichtvalide XML-Dokumente**

Das XML-Format wird immer häufiger zum Aufbau von Webseiten verwendet. Die meisten Web-Browser sind in der Regel sehr tolerant, wenn es um die Bildschirmdarstellung von „nicht-wohlgeformtem“ XML geht; auch bei Fehlern im Quellcode wird versucht, die Seite darzustellen. Verwendet die Suchmaschine einen XML-Parser, dann müssen die betreffenden XML-Dokumente „valide“ sein, d.h. syntaktisch korrekt, damit eine Indexierung durchgeführt werden kann. Eine häufige Problemquelle in diesem Zusammenhang sind nicht geschlossene XML-Tags. Dieses Problem kann durch die Verwendung des HTML-Parsers für XML-Dokumente – ggfs. unter Verlust von Metadaten - umgangen werden.

- **PDF-Dateien (Formate / Versionen)**

Das *Portable Document Format* von Adobe ist das bevorzugte Dateiformat für elektronische Artikel im Internet. Mit jeder neuen Version fügt Adobe neue Funktionalitäten hinzu, die Änderungen am Dateiformat mit sich ziehen. Diese Versionswechsel müssen in der Suchmaschinensoftware mitberücksichtigt werden⁷, damit eine Indexierung funktionieren kann. Darüber hinaus können die Sicherheitsoptionen von PDF verhindern, dass Dokumente von einer Suchmaschine indexiert werden können.

3.2 Nichtöffentliche Angebote

Im WWW existiert eine Vielzahl von Angeboten, die zwar einem Nutzer mit einem Webbrowser zugänglich sind, nicht aber dem Crawler von PsychSpider (und anderen Suchmaschinen). Die Anbieter dieser Informationsangebote möchten aus unterschiedlichen Gründen verhindern, dass auf Dokumente innerhalb ihrer Site von einer Suchmaschine aus zugegriffen wird bzw. dass Crawler ihre Site indexieren. Erreicht wird dies in der Regel über eine Datei robots.txt auf dem Webserver oder über die Verwendung des robots-Metatag. Hier kann der Site-Administrator angeben, ob ein Crawler auf ein Angebot zugreifen darf oder welche Teile eines Angebotes nicht zugänglich sein sollen. Die meisten Site-Betreiber reagieren zu Recht sehr empfindlich auf Missachtung dieser Direktiven und sperren den Zugriff für einzelne IP-Adressen.

⁷ Für die für den PsychSpider verwendete Verity-Software gab es erst einige Zeit nach Erscheinen des PDF-Format 1.5 einen Patch, der es ermöglicht hat, Dokumente in diesem Format korrekt zu indexieren.

In jüngerer Zeit stößt man auf immer mehr Angebote, die nur dem Google Spider (dieser identifiziert sich als Googlebot) den Zugriff erlauben; andere Suchmaschinen werden ausgesperrt. Der Grund für dieses Vorgehen ist in der Marktmacht von Google als dominierende Suchmaschine zu sehen: Da es vielen Site-Betreibern ausreicht, über Google gefunden zu werden, sperren sie andere Suchmaschinen aus, um den Traffic auf ihrer Site zu reduzieren.

4 Nutzerverhalten

Zur Untersuchung des Nutzerverhaltens wurde eine Logfileanalyse von PsychSpider durchgeführt. Die Daten wurden im Zeitraum vom 15. Juli 2004 bis 14. April 2005 erhoben und umfassen 51.403 Suchanfragen. Erfasst wurde neben Zeitstempeln, Nutzerkennung und der Suchanfrage auch, ob es sich um eine initiale Suchanfrage (Übergabe der Suchparameter per POST) oder um den gezielten Aufruf einer Ergebnisseite (Übergabe der Parameter per GET) handelte. Analysiert wurde die Navigation über die Ergebnisseiten anhand der 12.864 gezielten Aufrufe von Ergebnisfolgeseiten. Die 37.877 initialen Suchanfragen wurden hinsichtlich *Suchanfragen per Session*, *Terme pro Suchanfrage*, *Verwendung von Operatoren*, *Modifikationen der Suchanfrage* und *Suchanfragen- bzw. Termhäufigkeiten* näher betrachtet.

Um typisches Verhalten aufzeigen zu können, war ein Vergleich mit anderen Studien zu Suchmaschinenanfragen unumgänglich. Die Wahl fiel auf Studien, bei denen die Suchmaschinen ein ähnliches Suchverhalten hinsichtlich Unterstützung von Operatoren (AND, OR, NOT, +, -) und Aufbereitung einer Suchanfrage zeigen wie PsychSpider. Konkret sind dies die Studie zu AltaVista, welche ungefähr 1 Mrd. Suchanfragen umfasst, mit Erhebungszeitpunkt 1998 (Silverstein et al. 1999), drei Studien zu Excite und zwei zu FAST. Die älteste Studie zu Excite hat als Erhebungszeitpunkt 1997 und umfasst ungefähr 51.000 Anfragen (Jansen et al. 2000), im Weiteren als „excite 51K“ bezeichnet, die nachfolgenden Studien untersuchen jeweils über 1 Mio. Suchanfragen mit einem Erhebungszeitpunkt zum einen von 1997 (Spink et al. 2001), im Weiteren als „excite 1M“ bezeichnet, und zum anderen von 2001 (Spink et al. 2002), im Weiteren als „excite 2001“ gekennzeichnet. Die im Folgenden als „FAST“ bezeichnete Studie entstand 2001 und umfasst 451.551 Suchanfragen (Spink, Ozmutlu et al. 2002) und die als „AlltheWeb“ gekennzeichnete Studie analysiert fast 1 Mio. Suchanfragen, die in 2002 erhoben wurden (Jansen, Spink, 2005). Die zeitliche Differenz zwischen den Vergleichsdaten und unserem Material ist uns bewusst, aber aufgrund der in den Studien festgestellten geringen Veränderung des Nutzerverhaltens über die Jahre vernachlässigbar. Ebenso spielen nationale Unterschiede (Europa versus USA) eher im Hinblick auf die Suchterme ein

Rolle, die aber bei einer spezialisierten Suchmaschine wie PsychSpider einen gänzlich eigenen Charakter haben.

4.1 Navigation über Ergebnisseiten

12.864 Aufrufe bei PsychSpider, d.h. ungefähr 25% aller Suchanfragen sind gezielte Anforderungen von Ergebnisseiten. PsychSpider liefert pro Ergebnisseite 20 Treffer und eine Navigation über 5 Folgeseiten sowie Links zur letzten und ersten Trefferseite aus.

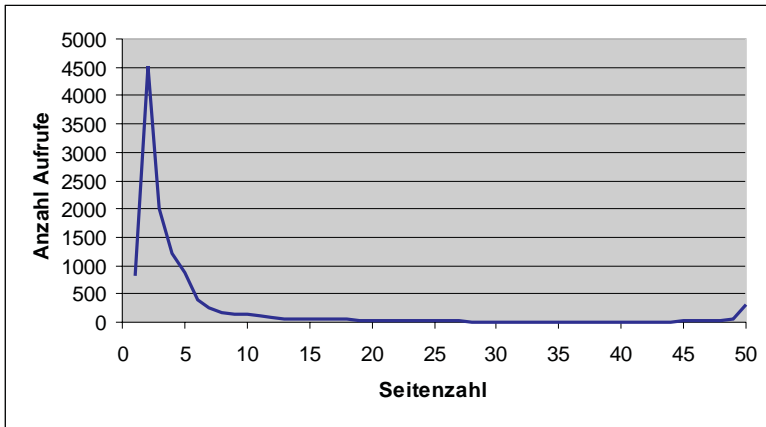


Abb. 6: Verteilung der expliziten Anforderungen von Ergebnisseiten bei PsychSpider

Damit zeigt sich eine ganz typische Verteilung (siehe Abbildung 6) mit einem Peak bei der 2. Seite und kontinuierlichem Abflachen zur 5. Seite hin. D.h. die ersten 100 Treffer werden vom Nutzer noch angesehen, weitere Ergebnisse eher nicht mehr. Der leichte Anstieg bei 50 erklärt sich als Sprung auf die letzte Seite, da PsychSpider nur maximal 50 Trefferseiten ausliefert. Der Wert bei 1 ist der gezielte Rücksprung auf die erste Ergebnisseite. Diese Verteilung ist absolut normal und auch die Vergleichsstudien berichten, dass insgesamt nur wenig Ergebnisseiten angesehen werden (Silverstein et al. 1999; Jansen et al. 2000; Spink et al. 2001; Spink, Ozmutlu et al. 2002; Jansen, Spink 2005).

4.2 Suchanfragen per Session

Suchanfragen per Session geben Anhaltspunkte über das Informationsbedürfnis des Nutzers. PsychSpider zeigt eine normale Verteilung (siehe Abbildung 7) mit über 25% der Sessions, die nur aus einer Suchanfrage bestehen. Eine deutliche Abnahme an Sessions, je mehr Suchanfragen pro Session gestellt werden, ist

aus der Kurve zu ersehen. Das Maximum lag bei 83 Suchanfragen innerhalb einer Session.

Beim Vergleich mit den genannten Studien (siehe Abbildung 8) sieht man, dass PsychSpider im Mittelfeld liegt. Im Durchschnitt wurden 2,14 Suchanfragen pro Session gestellt. AltaVista berichtet von 2, Excite von 2,3 über 2,5 bis 2,8 und FAST bzw. AlltheWeb von 2,9 bzw. 2,8 Suchanfragen per Session. Unterschiede in der Berechnung, was eine Session ist und welche Suchanfragen mit einbezogen werden, sind vernachlässigbar.

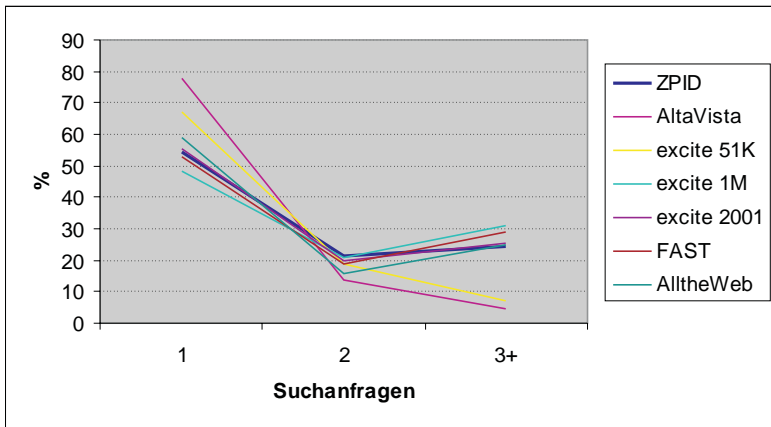


Abb. 8: Suchanfragen per Session – Vergleich mit anderen Studien

4.3 Terme pro Suchanfrage

Terme pro Suchanfrage liefern einen Hinweis auf die Komplexität der gestellten Suchanfragen. Bei PsychSpider überwiegen die Suchanfragen mit nur einem Term (siehe Abbildung 9). Die deutliche Abflachung der Kurve in Abhängigkeit von der Zunahme der Terme ist nicht ungewöhnlich. Im Durchschnitt wurden

1,89 Terme pro Suchanfrage verwendet. Dies ist gering im Vergleich zu anderen Studien, die 2,2 – 2,6 Terme pro Suchanfrage messen (siehe Abbildung 10).

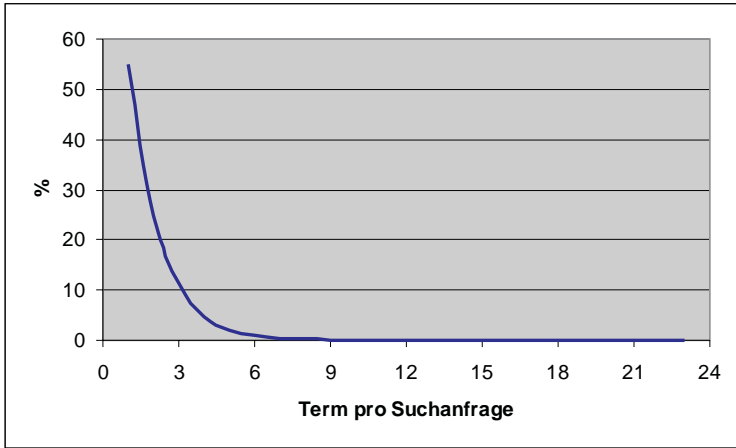


Abb. 9: Terme pro Suchanfrage – Verteilung am ZPID

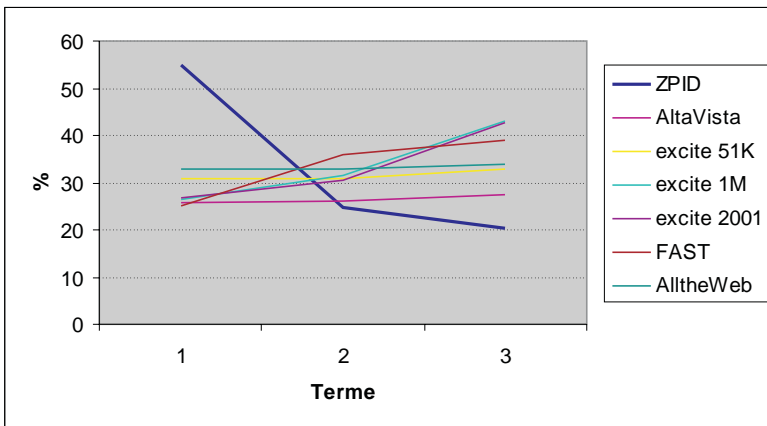


Abb. 10: Terme pro Suchanfrage – Vergleich mit anderen Studien

Eine mögliche Erklärung ist, dass ein weiterer Term zur Festlegung des Kontextes bei einer spezialisierten Suchmaschine nicht nötig ist und der Nutzer deswegen vermehrt mit eintermigen Suchanfragen auskommt. Unterschiede in der Berechnung, d.h. die Begriffe innerhalb der Phrase oder die Operatoren auch als

Terme gezählt, sind keine Ursache für dieses Phänomen. Dass das Mitzählen von Wörtern innerhalb von Phrasen oder Operatoren das Verhältnis Term/Suchanfrage nicht wirklich verändert und die Zählweise beim Studienvergleich damit irrelevant ist, zeigt sich anhand der Suchanfragenmenge, in der Phrasen oder Operatoren benutzt werden.

4.4 Verwendung von Operatoren

Phrasen kommen nur in knapp 2% der Anfragen vor (siehe Tabelle 1). Der hohe Anteil von 12,1% bei AND ist in der Art und Weise begründet, wie PsychSpiders Suchinterfaces gestaltet sind (siehe Abbildung 3). Die Haupteingabemaske besteht aus vier Texteingabefeldern, die defaultmäßig mit AND verknüpft werden. OR oder NOT müssen aus einem Pulldown-Menü ausgewählt werden. Bei AltaVista wird nur pauschal von einer geringen Verwendung Boole'scher Operatoren berichtet (Silverstein et al. 2001). In den Excite-Studien sind es 8,5% (Jansen et al. 2000), weniger als 5% (Spink et al. 2001) bzw. 10% (Spink et al. 2002), bei FAST nur 1% (Spink et al. 2002; Jansen, Spink 2005). Obwohl PsychSpider mit annähernd 13% Suchanfragen mit Boole'schen Operatoren über den Ergebnissen der anderen Studien liegt, darf man sich nicht täuschen lassen. Der Anteil ist insgesamt gering und der Gebrauch oft fehlerhaft.

Tab. 1: Verwendung von Operatoren

Operator	absolut	prozentual
Phrase	723	1,91
AND	4582	12,10
OR	271	0,72
NOT	61	0,16
+	392	1,03
-	79	0,21
Komma	751	1,98
Klammer	86	0,23
Wildcard *	286	0,76
Wildcard ?	16	0,04

4.5 Modifikation der Suchanfragen

Über 45% der Sessions bestehen aus mehr als einer Suchanfrage. Hier stellt sich die Frage, welcher Art die Modifikationen sind, die Suchanfragen innerhalb einer Session erfahren. 33,8% sind Begriffsmodifikationen, z.B. wird Altruismus

zu Altruismusskalen verändert, 25,4% sind völlige Neuformulierungen der Suchanfragen, z.B. wird Angst umformuliert zu Ärger. In 6,4% der Fälle erfolgt ein Operatorenwechsel, d.h. alle oder einer der zuvor verwendeten Operatoren wird durch einen anderen ersetzt. In 46,4% der Fälle werden Terme hinzugefügt oder entfernt, d.h. die Termanzahl ändert sich. Selbstverständlich gibt es auch Kombinationen dieser Modifikationen, z.B. Veränderung der Termanzahl und Begriffsmodifikationen. Betrachtet man nur die Änderung der Termanzahl, so folgt das Nutzerverhalten bei PsychSpider dem anderer Suchmaschinen. Eine wie in Abbildung 11 dargestellte Kurve findet sich zum Beispiel auch bei Excite (Spink et al. 2001).

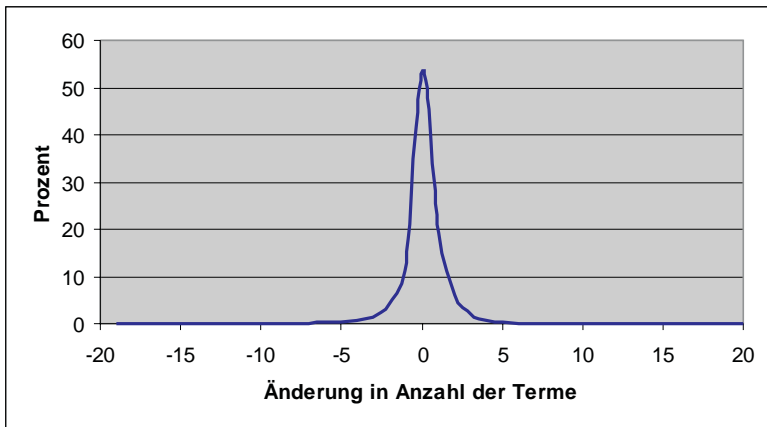


Abb. 11: Verteilung der Termanzahländerung

Wenn man lediglich die Suchanfragen analysiert, in denen keine Änderung der Termanzahl erfolgt, immerhin über 50%, dann ändern sich die Zahlen nur bei den Begriffsmodifikationen deutlich. Begriffsmodifikationen nehmen um 11% auf 44,8% zu. Neuformulierung nehmen um 4% zu, von 25,4% auf 29,4% und der Anteil an Operatorenwechseln nimmt um 3,3% auf 3,1% ab. Insgesamt erfolgen die Veränderungen eher in kleinen Schritten. Ein Verhalten, welches auch bei den anderen Studien beschrieben wurde (Jansen et al. 2000; Spink et al. 2001; Spink, Ozmutlu et al. 2002).

4.6 Suchanfragen- und Termhäufigkeiten

Um einen weitgehend unverfälschten Eindruck vom allgemeinen Informationsbedürfnis der PsychSpider-Nutzer für die Untersuchung der Suchanfragen bzw. Termhäufigkeiten zu erhalten, wurden aus den 37.877 initialen Suchanfra-

gen diejenigen Suchanfragen eliminiert, die am gleichen Tag mit gleicher IP gestellt wurden. Zusätzlich erfolgte eine Umwandlung in Kleinbuchstaben, da für PsychSpider Groß-/Kleinschreibung für die Suche irrelevant ist⁸. Unter den Top 100 der Suchanfragen dominieren eintermige deutschsprachige Anfragen. Nur vier Anfragen bestehen aus mehr als einem Wort. Auf Rang 29 ‘was ist psychologie’, auf Rang 77 ‘assessment center’, auf Rang 87 ‘emotionale intelligenz’ und auf Rang 94 ‘pädagogische psychologie’. Aus Platzgründen sind nur die ersten 10 Suchanfragen aufgeführt (siehe Tab. 2). Analysiert man nur die Terme, so ändert sich das Bild. Der Term ‘psychologie’ war vorher auf Rang 22, ‘test’ auf 17, ‘fragebogen’ auf Rang 58, ‘kinder’ auf 63, ‘soziale’ und ‘therapie’ noch nicht einmal unter den ersten Hundert. Diese Begriffe sind im Kontext einer Psychologiesuchmaschine wohl zu allgemein, weswegen dort vermehrt eine Kombination mit einem weiteren Term erfolgt. Auffällig ist die Dominanz der deutschsprachigen Begriffe. Nur 9% der Anfragen sind in englischer Sprache, 72% sind deutsch. Der Rest sind Anfragen, die sowohl im deutschen als auch englischen Sprachraum sinnvolle Abfragen darstellen. Dies sind u.a. Suchen nach Testabkürzungen, Personen, Zahlen oder Begriffen wie ‘borderline’, ‘test’, ‘depression’ etc. Durchweg alle Suchanfragen betreffen psychologische Aspekte. Die Dominanz von Begriffen aus dem sexuellen Bereich (Silverstein et al. 1999; Jansen et al. 2000), Entertainment und Freizeit (Silverstein et al. 1999; Jansen, Spink 2005), Personen und Computer (Spink, Ozmutlu et al. 2002; Jansen, Spink 2005) oder Commerce (Spink et al. 2002) wie bei Universalsuchmaschinen üblich, ist bei PsychSpider in keiner Weise gegeben. Die Spezialisierung von PsychSpider auf psychologischen Content wird von den Nutzern für die Suchanfragen auch wahrgenommen.

Tab. 2: Top 10 der Suchanfragen bzw. Terme

Rang	Suchanfrage	Term
1	litzcke	psychologie
2	stress	test
3	borderline	kinder
4	depression	fragebogen
5	angst	soziale
6	motivation	stress
7	burnout	depression
8	aggression	therapie
9	schizophrenie	motivation
10	mobbing	angst

8 Die Nutzer verwenden aber sehr wohl Groß-/Kleinschreibung.

5 Fazit

Spezialisierte Suchmaschinen wie PsychSpider stellen eine Ergänzung und keine Konkurrenz zu den großen Universalsuchmaschinen wie Google oder MSN dar. Durch die Einbindung von „Deep Web“-Angeboten, wie dies bei PsychSpider durch die Einbindung von *PSYNDEX Lit & AV* und *PSYNDEX Tests* geschieht, können sie ihren Nutzern einen echten Mehrwert anbieten. Des Weiteren bieten spezialisierte Suchmaschinen von vornherein einen Kontext für Suchanfragen an, so dass nur themenrelevante Treffer angezeigt werden.

Die Untersuchung des Nutzerverhaltens der PsychSpider-Benutzer hat kaum signifikante Unterschiede zu den vorliegenden Vergleichsstudien zum Nutzerverhalten bei Universalsuchmaschinen gezeigt. Auffällig war hier nur der vergleichsweise hohe Anteil von Ein-Wort-Suchbegriffen bei PsychSpider. Dieses Phänomen kann auf den Kontext, den PsychSpider für Suchanfragen zur Verfügung stellt, zurückgeführt werden. Ein Problem, das sowohl die aktuelle Studie als auch die Vergleichsstudien beschreiben, ist die fehlerhafte Verwendung boole'scher Operatoren durch die Suchmaschinennutzer.

Die Attraktivität von PsychSpider als sinnvolles Recherchetool für die Psychologie kann zum einen durch den weiteren Ausbau des Index erreicht werden, zum anderen aber auch durch eine Verbesserung der Rechercheoberfläche und andere Ranking-Algorithmen. Weitere Nutzerstudien müssen zeigen, inwieweit die Nutzer mit der Darstellung der Ergebnisse und den vorhandenen Recherchemöglichkeiten zufrieden sind.

6 Literatur

- Bergmann, M. (2000). The deep Web: Surfacing the hidden value. In: The Journal of Electronic Publishing, 7 [1], online verfügbar unter <http://www.press.umich.edu/jep/07-01/bergman.html> am 23.05.2005.
- Hamaker, Chuck and Brad Spry (2005). Key Issues: Google Scholar Search. Serials 18 [1]
- Jansen, Bernard; Spink, Amanda; Saracevic, Tefko (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. In: Information Processing and Management: an International Journal, 36 [2], S. 207-227.
- Jansen, Bernard J.; Spink, Amanda (2005). An analysis of web searching by European AlltheWeb.com users. In: Information Processing and Management 41 [2], S. 361-381.
- Jacso, P. (2005). Google Scholar: the pros and the cons. In: Online Information Review, 29 [2].

- Silverstein, C.; Marais, H; Henzinger, M.; Moricz, M. (1999). Analysis of a very large web search engine query log. In: ACM SIGIR Forum, 33 [1], S. 6-12.
- Spink, Amanda; Jansen, Bernard J.; Wolfram, Dietmar; Saracevic, Tefko (2002). From E-Sex to E-Commerce: Web Search Changes. In: Computer, 35 (3), S. 107-109.
- Spink, Amanda; Ozmutlu, Seda; Ozmutlu, Huseyin C.; Jansen, Bernard J. (2002). U.S. versus European web searching trends. In: ACM SIGIR Forum, 36 [2].
- Spink, Amanda; Wolfram, Dietmar; Jansen, Bernard; Saracevic, Tefko (2001). Searching the Web: the public and their queries. In: Journal of the American Society for Information Science and Technology, 52 [3], S. 226-234.

Angaben zu den Autoren

Christiane Baier, Peter Weiland

DV-Einsatz und Anwendung

Zentrum für psychologische Information und Dokumentation (ZPID)

Universitätsring 15, 54296 Trier

Postanschrift: 54286 Trier

E-Mail: info@zpid.de

Volltextserver für wissenschaftliche Dokumente aus der Psychologie: PsyDok

Ulrich Herb

Saarländische Universitäts- und Landesbibliothek (SULB)

Zusammenfassung

Im Gemeinschaftsprojekt „Digitale Psychologie-Information (DPI)“ setzten sich das Zentrum für Psychologische Information und Dokumentation (ZPID)¹ und die Sondersammelgebietsbibliothek für Psychologie, die Saarländische Universitäts- und Landesbibliothek (SULB)², das Ziel der Verbesserung des Informationsangebotes und -nachweises im Fach Psychologie bezogen auf Online-Informationen und digitale Publikationen. DPI wurde bis März 2005 von der Deutschen Forschungsgemeinschaft (DFG)³ gefördert.

Einer der zentralen Arbeitsschritte war die Einrichtung des (sowohl für Autoren als auch Nutzer kostenlos nutzbaren) Volltextservers PsyDok, <http://psydok.sulb.uni-saarland.de>, an der SULB und seine Einbindung in disziplinäre und interdisziplinäre Nachweissysteme.

1 Genese

1.1 SSG Bibliotheken und Virtuelle Fachbibliotheken

Seit 1966 beheimatet die Saarländische Universitäts- und Landesbibliothek (SULB) das Sondersammelgebiet (SSG) Psychologie. Zentrale Aufgabe ist die überregionale Versorgung der Wissenschaft mit Fachinformationen zur Psychologie. Diese Versorgungsaufgabe bezieht sich auf unterschiedliche Medientypen: Konventionelle Medien (Print), Mikroformen, Multimedia (Bücher + CD-Rom). 1998 verabschiedete die DFG ein Memorandum zur Weiterentwicklung der Literaturversorgung und zielte damit auf die Ausweitung des Systems der überregionalen Literaturversorgung unter Berücksichtigung der neuen Informationstechniken. Die DFG führte in der Folge ein Förderprogramm für Vir-

1 <http://www.zpid.de>

2 <http://www.sulb.uni-saarland.de>

3 <http://www.dfg.de>

tuelle Fachbibliotheken ein. 1999 ging die Virtuelle Fachbibliothek Psychologie online. Virtuelle Fachbibliotheken erfüllen ganz allgemein folgende Aufgaben:

- Nachweis fachwissenschaftlicher Informationsquellen im Internet in Zusammenarbeit mit Kooperationspartnern
- Sicherung der langfristigen Verfügbarkeit von elektronischen Fachinformationen
- Bündelung des Zugangs zu Angeboten
- Erschließung fachwissenschaftlicher Informationsquellen.

1.2 Ausbau und Integration der Virtuellen Fachbibliothek im Projekt DPI

Gerade die Flüchtigkeit wissenschaftlicher elektronischer Informationen im WWW stellt die Subject Gateways (resp. Fachinformationsführer, Linkkataloge) der Virtuellen Fachbibliotheken vor große Probleme. Daher wurde im Rahmen des Projekts DPI an der SULB der Volltextserver PsyDok eingerichtet. Generell diente DPI der Verbesserung der Versorgung mit elektronischen Fachinformationen in der Psychologie. Dieses Vorhaben bedurfte der intensiven Zusammenarbeit zwischen SULB und ZPID. Es umfasste folgende Arbeitspakete:

- Entwicklung eines Metadatenstandards für das Fach Psychologie
- Anpassung der Psychologie-Suchmaschine PsychSpider⁴ (ZPID)
- Integration von PsychLinker (ZPID) und Fachinformationsführer Psychologie FIPS (SULB)⁵
- Beteiligung der Fachcommunity an der Erschließung hochwertiger Internetressourcen im PsychLinker
- Einrichtung des Volltextservers PsyDok⁶ (SULB)

4 <http://www.psychspider.de>

5 <http://www.psychlinker.de>, der PsychLinker existierte als reines ZPID-Angebot schon seit langem. In DPI wurden PsychLinker und FIPS unter einer Oberfläche integriert und unter dem schon eingeführten Namen PsychLinker weitergeführt. Nach wie vor wird der PsychLinker beim ZPID gehostet.

6 <http://psydok.sulb.uni-saarland.de/>

2 Anspruch und Umsetzung

2.1 Ziel: Verbesserung der Informationsversorgung

Der Beitrag, den die SULB mit PsyDok zur Verbesserung der Informationsversorgung mit psychologierelevanten Informationen leistet, lässt sich wie folgt umreißen:

PsyDok bietet ein deutschlandweites, kostenloses, zentrales Veröffentlichungsangebot für wissenschaftliche Dokumente aus der Psychologie.⁷ Im Gegensatz zur Veröffentlichung auf lokalen Hochschulschriftenservern – die selbstverständlich parallel zur Veröffentlichung auf PsyDok möglich ist – erreichen PsyDok-Dokumente eine um ein Vielfaches höhere Sichtbarkeit. Diese beruht zum einen auf der zentralen Funktion von PsyDok innerhalb der Disziplin, zum anderen auf der vielfältigen Einbindung in disziplinäre Nachweissysteme. Darüber hinaus wird eine maximale Sichtbarkeit auch durch Nutzung interdisziplinärer Nachweissysteme erreicht.⁸ Dieser systematische Nachweis bietet vor allem gegenüber der immer noch weit verbreiteten Ablage wissenschaftlicher Dokumente auf den Webservern von Fachbereichen und Instituten einen immensen Vorteil, werden solche Texte doch eher zufällig und nach click-by-click-retrieval gefunden. Auch in anderer Hinsicht sind solche spontanen Veröffentlichungen durch Ablegen auf Fachbereichs-/Institutsservern kritisch zu sehen: Ihre Verwertbarkeit im wissenschaftlichen Diskurs ist äußerst begrenzt, da ihre Lebensspanne bzw. Verfügbarkeit meist kurz (zumindest aber unkalkulierbar) und ihre Unveränderbarkeit und Zitierbarkeit fraglich sind. Im Gegensatz dazu sichert die SULB den Autoren die dauerhafte Verfügbarkeit und Zitierfähigkeit für auf PsyDok veröffentlichte Dokumente zu.

PsyDok fungiert als überregionale fachlich ausgerichtete Sammlung elektronischer Dokumente. Der Server bietet eine zentrale Anlaufstelle für Veröffentlichung und Suche von elektronischen Dokumenten (vor allem graue Literatur, auch Postprints/Preprints). Es handelt sich bei PsyDok um ein Open Access-Angebot, das Fachwissenschaftlern die Möglichkeit zum Self-Archiving gibt. PsyDok ist kein originäres respektive genuines Veröffentlichungsangebot, das die Möglichkeit zum Self-Publishing gibt.⁹ Sowohl die Veröffentlichung von Dokumenten als auch die Nutzung der auf PsyDok veröffentlichten Dokumente ist kostenlos.

7 PsyDok wird allerdings auch im deutschsprachigen Ausland zur Veröffentlichung genutzt.

8 Vgl. Kapitel 2.9 Sichtbarkeit und Nachweis

9 Derzeit befasst sich die SULB konzeptionell mit der Option, auch ein originäres Open Access-Angebot mit Möglichkeit zum Self-Publishing bereitzustellen.

2.2 Technik

PsyDok basiert auf der Software OPUS¹⁰, die an der Universität Stuttgart entwickelt wurde und die an über 50 deutschen Hochschulen zum Betrieb eines institutional repositories genutzt wird. OPUS beruht auf einer MySQL-Datenbank, verwendet UNIX/LINUX als Betriebssystem und benötigt eine APACHE-/PHP-Umgebung. Bevorzugt werden auf PsyDok PDF-, PostScript- und HTML-Dateien veröffentlicht. Die SULB sichert für veröffentlichte Dokumente - wie bereits erwähnt - Langzeitverfügbarkeit und Zitierbarkeit. Die Zitierbarkeit wird über die Vergabe einer dauerhaften URL und durch Vergabe von Uniform Resource Names (URN)¹¹ garantiert. Ein URN ist - genau wie der weniger bekannte Persistent Uniform Resource Locator (PURL)¹² oder der aus dem Verlagswesen bekannte Digital Object Identifier (DOI)¹³, ein so genannter „Persistent Identifier“. Für Dokumente, die einen Persistent Identifier (in Form des URN, DOI oder PURL) besitzen, kann dauerhafter Zugriff und damit Referenzierbarkeit und Zitierbarkeit auch bei einer etwaigen Änderung der URL gewährleistet werden. Über einen Resolving-Mechanismus¹⁴ wird dem URN der korrekte URL zugeordnet.

Die Langzeitlesbarkeit und -nutzbarkeit wird ebenfalls zugesichert; hier ist die SULB Nachnutzer von Techniken, die auf übergeordneter Ebene, z.B. auf Verbundebene oder in Vorhaben wie nestor¹⁵ oder kopal¹⁶ entwickelt werden. Auf größere Zeitspannen bezogen ist die Langzeitlesbarkeit abhängig von Entwicklungen der Techniken zur Emulation und Migration.

2.3 Recht

Der Autor schließt mit der SULB einen Veröffentlichungsvertrag. Damit erhält die SULB unter anderem das nicht-ausschließliche Veröffentlichungs- und Vielfältigungsrecht, das Recht der öffentlichen Zugänglichmachung des Dokuments und das Recht der öffentlichen Zugänglichmachung der Metadaten (insbesondere des Abstract).

10 <http://elib.uni-stuttgart.de/>

11 <http://www.persistent-identifier.de/>

12 <http://www.purl.org/>

13 <http://www.doi.org/>

14 Ein Resolving-Mechanismus weist einer abstrakten Kennung (z.B. einer URN wie urn:nbn:de:bsz:291-psydok-818) ein konkretes Objekt zu (in diesem Fall den Bericht „Berufsfindung und Geschlecht: Wege in die Berufe Medizin und Psychologie“ von Susanne Stroux und Ernst-Hartmut Hoff).

15 <http://www.langzeitarchivierung.de/>

16 <http://kopal.langzeitarchivierung.de/>

2.4 Dokumentarten

Für die Veröffentlichung auf PsyDok kommt eine recht große Bandbreite an Dokumentarten in Frage:

- Magister-/Diplomarbeiten, Dissertationen, Habilitationen
- Aufsätze/Artikel als Postprints oder Preprints
- Bücher, Monographien
- Lehrmaterialien, Vorträge, Poster, Skalen
- Teile/Kapitel aus Büchern
- Journals/Komplette Ausgaben von Zeitschriftenheften
- Arbeits- und Forschungsberichte, Research Papers

2.5 Akquise von Dokumenten

Um Dokumente einzuwerben, wurden alle psychologischen Institute, Fachbereiche und Lehrstühle in Deutschland angeschrieben und auf PsyDok als Publikationsangebot hingewiesen. Selbstverständlich wurde PsyDok auch bei der Deutschen Gesellschaft für Psychologie (DGPs) publik gemacht und es wurde darum gebeten, Mitglieder auf dieses Publikationsinstrument aufmerksam zu machen. Nachdem die Mitgliederversammlung der Deutschen Gesellschaft für Psychologie (DGPs) sich im September 2004 mehrheitlich für den Beitritt zur Berliner Erklärung zum Open Access entschieden hat (vgl. Weichselgartner in diesem Band), bietet sich die Nutzung einer - in Form von PsyDok - schon existierenden Veröffentlichungsplattform an. Zusätzlich wurden Prüfungsämter und Fachschaften angeschrieben, um PsyDok gezielt als Veröffentlichungsmöglichkeit für Diplomarbeiten zu bewerben. Außerdem wurden, soweit möglich, alle Produzenten grauer Literatur auf PsyDok aufmerksam gemacht. Die SULB wollte auf diesem Weg den Bezug von grauer Literatur in Printform auf den Bezug in elektronischer Form durch die Veröffentlichung auf PsyDok umstellen. Schließlich wurden in der Projektphase von DPI DFG-geförderte Projekte aus dem Fach Psychologie durch ein mit der DFG abgestimmtes Anschreiben auf PsyDok hingewiesen und ausdrücklich darum gebeten, Projektberichte und andere projektbezogene Publikationen via PsyDok der Fachwelt zugänglich zu machen.

2.6 Veröffentlichung

Die Veröffentlichung erfolgt in der Regel durch den Autor bzw. dessen Institution über ein passwortgeschütztes Web-Formular. Dabei werden auch die Metadaten zum Dokument erstellt. Bei eingespielten Postprints oder bei Unsicherheiten auf Seiten der Autoren, ob sie ohne Rückfrage beim Verlag einen Post-

print auf PsyDok veröffentlichen können, wird auf die Angaben des Projekts SHERPA (Securing a Hybrid Environment for Research Preservation and Access) zurückgegriffen. SHERPA veröffentlicht eine Verlagsliste¹⁷, aus der die - meist erstaunlich liberale und Open Access-freundliche - Copyrightpolitik verschiedener Verlage ersichtlich ist. Um Missbrauch des Servers zu verhindern, sind eingespielte Dokumente nicht unmittelbar nach dem Hochspielen auf PsyDok im WWW sichtbar, sondern erst nach einer Überprüfung durch Mitarbeiter der SULB.

2.7 Support

Der Support durch die SULB umfasst unter anderem die Autorenunterstützung per Mail und Telefon sowie ggf. die Dokumentkonvertierung von proprietären in Standard-Dateiformate. Anbietern größerer Schriftsammlungen kann ein eigenes Portal eingerichtet werden, das nur die selbst beigesteuerten Dokumente anzeigt. Dieses Angebot wird von verschiedenen Herausgebern von Schriftenreihen in Anspruch genommen; es ermöglicht die - für einen hochschulübergreifenden Wissenschaftsserver wie PsyDok wichtige - Individualisierbarkeit der Darstellung. Solche Portale werden dem Layout der herausgebenden Einrichtung angepasst.¹⁸ Dokumente, die ein institutsinternes Peer Review durchlaufen haben, erhalten eine entsprechende Auszeichnung. Schließlich übernimmt die SULB in Ausnahmefällen das Einspielen der Dokumente.

2.8 Qualität

Da PsyDok kein originäres Publikationsangebot ist, sondern zumeist Dokumente aufnimmt, die in anderen Zusammenhängen erstellt wurden, kann seitens der SULB bei Entstehung der Dokumente kein Einfluss auf die Qualität genommen werden. Dennoch greifen - abgesehen von der Tatsache, dass die SULB sich das Recht vorbehält, nicht jedes vorgeschlagene Dokument auf PsyDok zu veröffentlichen - Mechanismen der Qualitätskontrolle:

- Bei Schriftenreihen kann von einer Kontrolle per Editorial Review ausgegangen werden.
- Bei Dissertationen und Habilitationen liegt in aller Regel ausreichende Qualität qua Dokumentart vor.
- Bei Diplom- und Magisterarbeiten erfolgt eine Veröffentlichung nur auf Empfehlung bzw. nach Rücksprache mit der Gutachterin/dem Gutachter.
- Bei Preprints kann eine überprüfbare Reputation des Autors aus bisherigen Veröffentlichungen abgeleitet werden.

17 <http://www.sherpa.ac.uk/romeo.php?all=yes>

18 Beispiele solcher Portale finden sich unter: <http://psydok.sulb.uni-saarland.de/portal.php>

- Bei Postprints darf davon ausgegangen werden, dass eine Begutachtung durch Verlage in die Wege geleitet wurde.
- Zusätzlich können Fachbereiche und Institute ein eigenes Peer Review durchführen, bevor Dokumente auf PsyDok veröffentlicht werden.

Im Bereich der technischen oder organisatorischen Qualitätsprüfung digitaler Publikationsangebote ist in Deutschland vor allem die Deutsche Initiative für Netzwerkinformation e.V. (DINI)¹⁹ zu nennen. DINI strebt die Etablierung von Gütesiegeln für Dokumentenserver an. Neben technischen und formalen Standardisierungsaspekten steht das Ziel im Vordergrund, elektronisches Publizieren durch Vergabe von Gütesiegeln noch seriöser zu machen. Die Kriterien für die erfolgreiche Zertifizierung berühren organisatorische Fragen, Fragen der Authentizität und Integrität der Daten, zugesagte Archivierungszeiträume, Funktionsumfang des Angebots, Export von Dokumenten in Archivsysteme und Fragen der inhaltlichen Erschließung sowie Fragen des Metadatenexports. Server, die den Anforderungen der DINI genügen, erhalten ein entsprechendes Siegel. PsyDok erfüllt alle diese Anforderungen und ist zertifizierter DINI-Server.

2.9 Sichtbarkeit und Nachweis

Bei der Diskussion der Vorteile der elektronischen Publikation gegenüber der Printpublikation wird neben der wesentlich höheren Zirkulationsgeschwindigkeit elektronischer Dokumente immer auch die erhöhte Sichtbarkeit dieser Dokumente genannt. Gerade hinsichtlich der Sichtbarkeit hebt sich PsyDok durch seine Einbindung in disziplinäre und interdisziplinäre Suchsysteme von anderen Volltextservern ab. Der Server eignet sich für die Einbindung in fachliche Angebote hervorragend, da PsyDok im Unterschied zu den meisten üblichen institutional repositories kein „Gemischtwarenladen“, sondern ein „Fachhändler“ ist.

PsyDok verfügt unter anderem über eine Schnittstelle zum kooperativ von ZPID und SULB betriebenen Subject Gateway PsychLinker²⁰ und zur Datenbank PSYNDEX²¹ (Literatur, Testverfahren, AV-Medien) des ZPID. Darüber hinaus werden Diplomarbeiten in der Diplomarbeitendatenbank²² des ZPID nachgewiesen. Sämtliche Dokumente werden durch die Psychologie-Suchmaschine PsychSpider²³ des ZPID (vgl. Baier/Weiland in diesem Band) indiziert und sind über diese recherchierbar. Als weiteres wichtiges fachspezifisches

19 <http://www.dini.de/>

20 <http://www.psychlinker.de>, die Übernahme der Links zu den Dokumenten in den Psych-Linker hängt von inhaltlichen Gewichtungen ab

21 <http://www.psyndex.de>, die Auswahl der Dokumente, die im PSYNDEX erscheinen, erfolgt durch ZPID-Mitarbeiter

22 <http://www.zpid.de/index.php?wahl=products&uwahl=frei&uwahl=thesefinfo>

23 <http://www.psychspider.de>

Nachweissystem ist der OPAC der SULB²⁴ als SSG-Bibliothek zu nennen, über den täglich zahlreiche überregionale Recherchen nach psychologischer Fachliteratur durchgeführt werden.

An interdisziplinären Nachweissystemen ist zum Beispiel der Katalog des Südwestdeutschen Bibliotheksverbundes²⁵ oder der Karlsruher Virtuelle Katalog (KVVK)²⁶ zu erwähnen. Neben diesen klassisch-bibliothekarischen Suchsystemen sind PsyDok-Dokumente aber auch über Suchsysteme, die gezielt Volltextserver und elektronische Dokumente berücksichtigen, zugänglich. Hier sind in erster Linie OASE (Open Access to Scientific Literature)²⁷, früher Karlsruher Virtueller Volltextkatalog (KVVK), und die Metasuche²⁸ über alle OPUS-basierten Volltextserver zu nennen. Als registrierter Dataprovider der Open Archives Initiative (OAI)²⁹ stellt PsyDok die Metadaten zu den Dokumenten über das OAI Metadata Harvesting Protokoll OAI-basierten Suchmaschinen zur Verfügung. Solche Suchmaschinen (oder „OAI Service Provider“) bieten die Möglichkeit, in einem wissenschaftlichen Dokumentenpool nach elektronischen, sofort verfügbaren und in aller Regel kostenlosen Dokumenten zu recherchieren. Das Metadata Harvesting Protokoll der OAI ermöglicht damit die wohl gelungenste Umsetzung der Vorstellung einer digitalen Bibliothek.³⁰ Andere interdisziplinäre Suchsysteme, in denen PsyDok-Dokumente nachgewiesen sind, sind z.B. Scirus³¹, die wissenschaftliche Suchmaschine von Elsevier, und das Forschungsportal des Bundesministeriums für Bildung und Forschung (BMBF)³². Dokumente, die im Subject Gateway PsychLinker oder in der Datenbank PSYINDEX verzeichnet sind, erscheinen auch im interdisziplinären Wissenschaftsportal vascoda³³, einem von der DFG und dem BMBF initiierten Gemeinschaftsprojekt deutscher Bibliotheken und Fachinformationseinrichtungen. Dokumente, die in der Datenbank PSYINDEX verzeichnet sind, erscheinen zusätzlich im Informationsverbund infoconnex³⁴, einer ebenfalls von DFG und BMBF geförderten Initiative zur Versorgung mit Volltexten in den Fächern Psychologie, Pädagogik, Sozialwissenschaften.

24 <http://opac.sulb.uni-saarland.de>

25 <http://swb.bsz-bw.de/>

26 <http://www.ubka.uni-karlsruhe.de/kvk.html>

27 <http://www.ubka.uni-karlsruhe.de/kvbk.html>

28 http://opus.uni-stuttgart.de/opus/gemeinsame_suche.php

29 <http://www.openarchives.org>

30 Als eine der besten OAI-basierten Suchmaschinen sei hier OAIster mit fast 6 Millionen Datensätzen genannt, <http://oaister.umdl.umich.edu/o/oaister/>

31 <http://www.scirus.com/srsapp/>

32 <http://www.forschungsportal.net/>

33 <http://www.vascoda.de/>

34 <http://www.infoconnex.de/>

Auch fachfremde Informationsanbieter integrieren PsyDok in ihr System, so werden einige Dokumente aus PsyDok auch in der „Collection of Computer Science Bibliographies“³⁵ und im CiteSeer³⁶ nachgewiesen.

Die SULB will die Liste der Suchmaschinen und Nachweissysteme, in denen PsyDok-Dokumente zu finden sind, kontinuierlich erweitern. So wurde PsyDok bei Google Scholar³⁷, dem Google-Ableger für wissenschaftliche Literatur, zur Indizierung gemeldet.³⁸ Weiterhin bemüht sich die SULB derzeit, PsyDok durch die Suchmaschine Social Science Search Engine³⁹ des Social Science Information Gateway (SOSIG)⁴⁰ indizieren zu lassen. Gleiches gilt für die Suchmaschine PsychCrawler der American Psychological Association (APA)⁴¹. Ein Nachweis in der neuen APA-Datenbank für graue Literatur, PsycExtra, wird erwoogen und nach Klärung rechtlicher Fragen entschieden.

Auf PsyDok archivierte Dokumente erzielen folglich einen um ein Vielfaches höheren Verbreitungsfaktor, als es durch klassische (Bibliotheks-) Kataloge, die üblichen WWW-Suchmaschinen oder einen lokalen, nicht fachlich ausgerichteten, wissenschaftlichen Volltextserver möglich ist. Ende 2004 wurden die Zugriffe auf PsyDok so zahlreich, dass PsyDok auf einen neuen Rechner portiert werden musste - auch das ein Beleg für die Verbreitung, die PsyDok-Dokumente finden.

2.10 Retrieval

PsyDok verfügt neben der Metadatensuche in definierten Feldern und Volltextsuche über verschiedene Browsingoptionen. Der Benutzer kann über Einrichtungen, Dokumentarten, die Dewey Decimal Classification (DDC) und Schriftenreihen bzw. Kongresse browsen. Nach Abschluss eines entsprechenden Vertrages mit der APA wird auch das Browsing über die in der Psychologie gebräuchliche Klassifikation, die PsycINFO Classification Categories and Codes der APA, möglich sein. Weitere Features sind ein RSS-Feed⁴², das über die neuesten Veröffentlichungen auf PsyDok informiert, und Plugins für die Searchbox des Browsers Firefox. Beide können über die PsyDok-Homepage abgerufen werden.

35 <http://iinwww.ira.uka.de/bibliography/>

36 CiteSeer ist eine Art digitale Bibliothek, die frei zugängliche, elektronische wissenschaftliche Volltexte zur Informatik und Informationswissenschaft nachweist. Derzeit wird auf weit über 700.000 Dokumente verwiesen. Der Clou: CiteSeer verwaltet dabei auch Zitationen und verlinkt auf referenzierte Literatur. <http://citeseer.ist.psu.edu/>

37 <http://scholar.google.com/>

38 Google Scholar befindet sich noch im Betastadium und indiziert PsyDok derzeit teilweise.

39 <http://www.sosig.ac.uk/harvester.html>

40 <http://www.sosig.ac.uk/>

41 <http://www.apa.org/>

42 <http://psydok.sulb.uni-saarland.de/rss.php>

3 Ausblick

Neben der Einbindung von PsyDok in weitere Nachweissysteme und der Implementierung der PsycINFO Classification Categories and Codes werden in nächster Zeit zahlreiche Erweiterungen und konzeptionelle Weiterentwicklungen umgesetzt werden. Einige dieser Erweiterungen und Entwicklungen entspringen der OPUS-Community, andere sind SULB-spezifisch und werden selbstredend - insofern möglich - auch innerhalb der OPUS-Community weitergegeben.

Hier ist zunächst die Einrichtung einer Schnittstelle zum Print-On-Demand-Anbieter ProPrint⁴³ und die Verankerung der Creative Commons⁴⁴ (vgl. auch Euler sowie Brüning/Kuhlen in diesem Band) in der Lizenzgestaltung zu nennen. Beide Module wurden an der Universitätsbibliothek Tübingen entwickelt und sind am dortigen Volltextserver TOBIAS-lib⁴⁵ schon im Einsatz (vgl. Roth-Steiner in diesem Band). Die SULB entwickelt ein Modul zur Mehrsprachigkeit für OPUS, von der PsyDok als Server für ein stark englischsprachig geprägtes Fach sehr profitieren dürfte. Um die Authentizität und Integrität der Dokumente überprüfen zu können, werden in naher Zukunft MD5-Prüfsummen generiert. Bereits in PsyDok verwirklicht ist die Anzeige einer anonymisierten Download-Statistik, die anzeigt, wie oft ein Dokument aufgerufen wurde, ohne dabei preiszugeben, wer den Server besucht hat oder gar ein spezielles Dokument gelesen hat. In der Diskussion um Open Access und elektronisches Publizieren allgemein werden von Gegnern dieser Bestrebungen regelmäßig zwei Argumente angeführt, die das elektronische Publizieren außerhalb der klassischen Verlagsangebote als gewissermaßen zweitklassig darstellen sollen:

- das Fehlen von Zitationsnachweisen
- das Fehlen verlässlicher, öffentlicher Statistiken über die Nutzung der elektronischen Dokumente

Es gibt bereits Modelle und technische Lösungen, die das erste Argument entkräften.⁴⁶ Diese können aber von Betreibern einzelner Server nicht im Alleingang sinnvoll umgesetzt werden. Anonyme Zugriffsstatistiken auf Volltextservern können aber das zweite Argument widerlegen. Eine solche Statistik belegt, dass frei zugängliche Informationen in aller Regel eine wesentlich höhere Leserate und eine schnellere Verbreitung erreichen, als es bei konventionellen Publikationen der Fall ist. Dabei darf aber nicht vergessen werden, dass hier eine

43 <http://www.proprint-service.de/>

44 <http://creativecommons.org/>

45 <http://www.uni-tuebingen.de/ub/elib/tobias.htm?http://w210.ub.uni-tuebingen.de/dbt/intro/>

46 Neben dem schon erwähnten CiteSeer sind vor allem Citebase,

<http://citebase.eprints.org/cgi-bin/search>, und OpCit, <http://opcit.eprints.org/>, zu nennen.

Vielzahl technischer Probleme zu beachten und zu lösen sein wird, bis unter Betreibern von Volltextservern eine Standardisierung erreicht ist, die wirklich verlässliche Zugriffsdaten ermöglicht. Ziel ist es, den Counter Standard⁴⁷ zu erreichen. Statistiken wie die, die PsyDok derzeit bereitstellt, sind davon noch relativ weit entfernt, können aber den Benutzern wertvolle Hinweise über Leseraten eines Dokumentes liefern und bieten zumindest einen Ausgangspunkt, um Protagonisten des elektronischen Publizierens im Allgemeinen und Open Access im Speziellen schlagkräftige Argumente zu liefern.

Angaben zum Autor

Ulrich Herb
Saarländische Universitäts- und Landesbibliothek
Postfach 15 11 41
66041 Saarbrücken

Dipl.-Soziologe, wissenschaftlicher Mitarbeiter an der Saarländischen Universitäts- und Landesbibliothek. Projektstätigkeiten: Bisherige Projektstätigkeiten (Virtuelle Fachbibliothek Psychologie, Digitale Psychologie Information, vas-coda) Aktuelle Tätigkeitsschwerpunkte: elektronisches Publizieren, alternative Publikationsmodelle, Open Access.
E-Mail: u.herb@sulb.uni-saarland.de

47 <http://www.projectcounter.org/>

Die Virtuelle Fachbibliothek Sportwissenschaft

Zum Stand des Gemeinschaftsprojekts

Jochen Johannsen

Zentralbibliothek der Sportwissenschaften der Deutschen Sporthochschule
Köln

Zusammenfassung

Vorgestellt wird das Konzept einer Virtuellen Fachbibliothek („ViFa“) Sportwissenschaft, wie es von einer Projektgruppe erarbeitet worden ist, die neben Bibliotheken und Institutionen der sportwissenschaftlichen Fachinformation mit der Deutschen Vereinigung für Sportwissenschaft (dvs) auch ein Mitglied der IuK-Initiative der wissenschaftlichen Fachgesellschaften umfasst. Ziel des Projekts ist die Etablierung eines zentralen Internetportals zur qualifizierten Fachinformation in Form eines „One Stop Portals“ für die Sportwissenschaft. Ermöglicht werden soll eine integrierte Metasuche in den von den Partnern erschlossenen hybriden Beständen, die durch Nachweis und Integration digitaler Volltexte und durch die Möglichkeit von Online-Fernleihe und elektronischer Dokumentlieferung ohne Medienbruch für die Benutzung bereitgestellt werden sollen. Der Aufsatz skizziert vor allem Ziel und Umfeld dieses der Deutschen Forschungsgemeinschaft im März 2005 zur Förderung vorgeschlagenen Projekts.

1 Ein Desiderat der Sportwissenschaft

In der jüngeren Vergangenheit ist verschiedentlich darüber geklagt worden, dass „die deutsche Sportwissenschaft als eines der ‚Schlusslichter‘ im deutschen Wissenschaftsbereich insgesamt“ anzusehen sei, sowohl was die Quantität als auch was die Qualität der digitalen Fachinformation im Internet angehe (Elpel 2000, 104). Anstatt „ein Bit voraus“ zu sein, „hinke“ die Sportwissenschaft auf diesem Gebiet hinter anderen Wissenschaftsdisziplinen her, so lautete schon vor einem Jahrzehnt der Vorwurf (Borkenhagen 1995, 4). Zweifellos hat sich seitdem vieles gebessert, schaut man etwa auf die wichtigsten in Deutschland produzierten sportwissenschaftlichen Datenbanken, die nunmehr online verfügbar sind. Dennoch ist nach wie vor eine recht starke Zersplitterung des sportwissenschaftlichen Informationsangebotes im Internet zu konstatieren und entspre-

chend von den Nutzern ein hoher Rechercheaufwand in Bibliothekskatalogen, Datenbanken und Webverzeichnissen zu betreiben. Der Vorschlag, dieses Problem durch den Aufbau eines einschlägigen Informationsportals für die Sportwissenschaft zu lösen, damit diese sich nicht mehr „hinter den Diensten anderer Disziplinen zu verstecken“ braucht (Borkenhagen 2001, 32), ist ebenso alt wie die zitierten Klagen – und soll nunmehr im Rahmen des im Folgenden vorgestellten Vorhabens endlich umgesetzt werden. Das Ziel der kooperativ entwickelten Virtuellen Fachbibliothek Sportwissenschaft ist es, eine zentrale Anlaufstelle für die sportwissenschaftliche Informations- und Literaturversorgung im Internet zu schaffen, die im Sinne eines „One Stop Shops“ oder „One Stop Portals“ ihren Nutzern die Möglichkeit bietet, in einem Anlauf die für sie wichtigsten Informationsmittel aus allen Bereichen der multidisziplinären Sportwissenschaft zu ermitteln und zu nutzen. Nachweis und Bereitstellung, Recherche und Zugriff sollen also möglichst nahtlos miteinander verknüpft werden, so dass die Nutzer ohne Medienbruch und durch wenige Klicks von einem Titelnachweis direkt zu einem elektronischen Volltext oder wenigstens zur Online-Bestellung eines Dokuments per Lieferdienst (*subito*) oder Fernleihe gelangen können. Dass es hierbei nicht um einen bibliothekarischen Selbstzweck, sondern um ein Desiderat der sportwissenschaftlichen Community geht, ist die leitende Überzeugung der an diesem Projekt beteiligten Institutionen und Personen.

Das erwähnte Desiderat ergibt sich nicht zuletzt aus den fachlichen Strukturen der Sportwissenschaft. Zu den Besonderheiten des Fachs gehört seine Multi- bzw. Transdisziplinarität: wie nur wenige Wissenschaften überschreitet es traditionelle Fakultätsgrenzen und weist gemeinsame Interessen, Themen und Fragestellungen sowohl mit den Geistes- und Kulturwissenschaften (z.B. Sportgeschichte und Sportethik) und den Sozial- und Wirtschaftswissenschaften (z.B. Sportsoziologie und Sportökonomie), als auch mit den Natur- und Biowissenschaften (z.B. Biomechanik und Sportmotorik), der Medizin (z.B. Sport- und Rehabilitationsmedizin) sowie den Ingenieurwissenschaften bzw. der Informatik (z.B. Sportgerätebau und Sportinformatik) auf. Zugleich stellt die Sportwissenschaft jedoch eine eigenständige Disziplin mit spezifischen Fragestellungen dar, die, etwa im Bereich der Trainings- und Bewegungslehre, gleichermaßen praktische wie theoretische Bezüge besitzt. Die im Fach angelegte Interdisziplinarität, die damit verbundene institutionelle Vielfalt und die Unterschiedlichkeit der Ansprüche an bibliothekarische und informationelle Dienstleistungen prägen die Ausgangslage des Projektes. Sportwissenschaftlerinnen und Sportwissenschaftler sind *einerseits* auf die Ergebnisse vieler Nachbar- bzw. Mutterwissenschaften angewiesen und tragen mit ihren Forschungen zu ihnen bei. Sie bewegen sich deshalb häufig in den Informationsumgebungen für sie relevanter Wissenschaften. Die Sportwissenschaft insgesamt kann deshalb wohl beanspruchen, ein gleichsam ‚natürlicher‘ Partner für das interdisziplinäre

Wissenschaftsportal *vascoda* (www.vascoda.de) zu sein¹. Dies wird nicht zuletzt dadurch unterstrichen, dass mit der Zentralbibliothek der Sportwissenschaften der Deutschen Sporthochschule Köln (ZBS) und dem Bundesinstitut für Sportwissenschaft (BISp) zwei der Träger der projektierten ViFa Sportwissenschaft zu den Partnerorganisationen von *vascoda* gehören, wobei das BISp bereits 2004 sein Datenbankangebot als erstes sportwissenschaftliches Modul SPORTIF in die *vascoda*-Metasuche integriert hat.

Über solche interdisziplinären Einbindungen hinaus besteht jedoch – wie eingangs hervorgehoben – *andererseits* ein dringender Bedarf an einem originär sportwissenschaftlichen Informationsportal als einem zentralen Vernetzungspunkt der Disziplin. Die im Kontext anderer Fächer entwickelte Einsicht, dass Fachportale und Virtuelle Fachbibliotheken nur fachlich einschlägige Nachweise, diese aber möglichst vollständig anbieten sollten (Görlitz 2005, 69), trifft im Falle der äußerst vielgestaltigen Sportwissenschaft in besonderem Maße zu. Die Erfüllung dieser Anforderung ist für die Akzeptanz und damit für die nachhaltige Nutzung der künftigen Virtuellen Fachbibliothek in der sportwissenschaftlichen Fachwelt von großer Bedeutung. Ein Andocken der Sportwissenschaft an das Informationsangebot einer einzelnen Mutterwissenschaft würde die Sichtbarkeit der Disziplin für Fachleute und Fachfremde reduzieren und die Nutzungsbereitschaft der Studierenden, Lehrenden und Forschenden in der Sportwissenschaft und im Leistungssport deutlich minimieren². Auch würden dort Teilbereiche betont, wo es um das Abdecken einer äußerst vielgestaltigen Disziplin und um die Befriedigung durchaus verschiedener Nutzerinteressen und Informationsbedürfnisse geht. Sinnvoller erscheint deshalb ein von der Nutzer-sicht ausgehendes Konzept einer eigenständigen ViFa Sportwissenschaft, welche die institutionelle und inhaltliche Vielgestaltigkeit des Faches abbildet und gleichzeitig enge Kooperationen mit fachlich ‚benachbarten‘ ViFas – wie etwa aus den Bereichen Pädagogik oder Medizin - eingeht. Ein solches Portal kann aufgrund der gesellschaftlichen Omnipräsenz und Attraktivität des Sports auch über die fachwissenschaftlichen Grenzen hinaus mit Aufmerksamkeit rechnen.

1 *vascoda* stellt zur Zeit hauptsächlich eine Metasuche über verschiedene Informationsbestände einzelner Fächer wie OPACs, Fachinformationsführer und Literaturdatenbanken dar. An der Gebrauchstauglichkeit dieser Metasuche ist zum Teil berechtigte Kritik geübt worden. Vgl. dazu die anderorts gemachten Anmerkungen des Verfassers (Johannsen 2005, 121f) sowie den Überblicksartikel von Tamara Pianos (2005).

2 Ein solches Andocken impliziert die 2004 von der Deutschen Forschungsgemeinschaft veröffentlichte Zielvorstellung zur Weiterentwicklung des Sondersammelgebietsplans, denn dort wird die Sportwissenschaft dem Themencluster „Bildung“ und dem Fachportal Pädagogik zugeordnet (DFG 2004).

2 Zum Umfeld der Virtuellen Fachbibliotheken

Virtuelle Fachbibliotheken streben danach, prinzipiell alle für eine wissenschaftliche Disziplin relevanten Informationsressourcen ungeachtet der Medienart unter einem Dach anzubieten und mit bibliothekarischen Serviceleistungen und informationellen Mehrwertdienstleistungen zu verknüpfen. In der Informationslandschaft der Bundesrepublik werden Virtuelle Fachbibliotheken mit jenen Fachinformationsangeboten identifiziert, die sich, seit Ende der 1990er Jahre gefördert durch die Deutsche Forschungsgemeinschaft (DFG), in dem interdisziplinären Dachportal *vascoda* zusammengefunden haben (Pianos 2005). Der Zugang zu dem Fächer- und Angebotsspektrum der 25 (Stand Frühjahr 2005) meist modular aufgebauten ViFas erschließt sich wohl am einfachsten über die Seiten dieses Portals. Die Leistungsfähigkeit der einzelnen Angebote ist – abhängig von den Erfordernissen der einzelnen Fächer, den Möglichkeiten der beteiligten Institutionen und der langsamen Entwicklung guter Praxis – durchaus heterogen. Dennoch haben sich, befördert auch durch die Präzisierung von Vorgaben durch die Förderinstitution (DFG 2004) und durch zusammenfassende Veröffentlichungen (Rösch 2004), mittlerweile einige Standards herausgebildet. Dabei ist vor allem die Integration des jeweiligen SSG-OPACs, eines Fachinformationsführers (*subject gateway*), wichtiger Datenbanken und eines Current-Contents-Dienstes sowie die Implementierung einer Metasuche über die einzelnen Module zu nennen. Der möglichst unmittelbare Zugang zu den Ressourcen über E-Volltexte, Dokumentlieferdienste oder Online-Fernleihe ist zu gewährleisten. Eine (im Idealfall) möglichst einheitliche, jedoch stets fachlich angemessene (und deshalb meist heterogene) sachliche Erschließung der Ressourcen zur Ermöglichung eines übergreifenden inhaltlichen Retrievals stellt einen wichtigen integrativen Kernaspekt dar (Krause 2004).

Für dieses hier nur grob zu umreißende Anforderungsprofil sind im Rahmen der mit Hilfe der DFG seit 1998 aufgebauten ViFas konzeptionelle, organisatorische und technische Lösungen oder Lösungsansätze entwickelt worden, die von jüngeren Projekten wie der ViFa Sportwissenschaft nachgenutzt werden können. Dennoch bringt der Aufbau einer ViFa nach wie vor einen erheblichen Ressourceneinsatz mit sich, der von den beteiligten Institutionen nicht ‚nebenher‘ oder mit ‚Bordmitteln‘ zu bewältigen ist. So ist der Aufwand an qualifizierter Arbeitszeit für die konzeptionelle Planung und für die Kommunikation mit Kooperationspartnern beträchtlich, ebenso wie die entstehenden Kosten, etwa für die Implementierung einer Metasuche über die einzelnen Module. Der Aufbau der ViFas in Deutschland ist deshalb angesichts stagnierender und sinkender Bibliotheksetats und trotz beträchtlicher Eigenleistungen der beteiligten Bibliotheken und Informationseinrichtungen von Beginn an nur durch die Personal- und Sachbeihilfen der DFG möglich gewesen, die im Jahr 2004 insgesamt 3.708.000 € für Projekte in dem zuständigen Förderprogramm „Elektroni-

sche Publikationen im wissenschaftlichen Literatur- und Informationsangebot“ bewilligt hat (DFG 2005, 127). Diese Fördermittel werden ausdrücklich als Anschubfinanzierung in der Aufbauphase von ViFas gewährt, wobei es das Ziel der DFG ist, das System der Sondersammelgebiete in ein Netzwerk virtueller Fachbibliotheken zu transformieren und damit die Konsequenz aus den veränderten Informationsanforderungen der Wissenschaft und aus den zunehmend hybriden Beständen der Bibliotheken und Informationseinrichtungen zu ziehen (DFG 2004). Ein vorrangiges Ziel der aufbauenden Institutionen muss also sein, ihr Angebot so zu strukturieren, dass dessen Fortbestand und Weiterentwicklung auch langfristig gesichert ist. Naturgemäß hatte dieser Aspekt eine besondere Bedeutung auch in den Vorplanungen des hier vorgestellten Projektes. Die Entscheidung, den Aufbau einer ViFa Sportwissenschaft in Angriff zu nehmen und einen diesbezüglichen Antrag auf Förderung bei der DFG zu stellen, wäre nicht erfolgt, wenn bei den Beteiligten nicht die Überzeugung bestehen würde, diese Nachhaltigkeit gewährleisten zu können.

3 Projektgemeinschaft und Kooperationspartner

Die Kooperation mehrerer Institutionen hat sich als sinnvolles Paradigma für den Aufbau wissenschaftlicher Portalsysteme erwiesen (Görlitz 2005). Wie auch bei anderen ViFas üblich, soll die ViFa Sportwissenschaft von mehreren Partnern aufgebaut und betrieben werden. Es handelt sich dabei um fünf Institutionen, die im März 2005 einen Gemeinschaftsantrag auf Förderung des Projektes bei der DFG eingereicht haben. Diese Institutionen – zwei Spezialbibliotheken, zwei Einrichtungen der wissenschaftlichen Fachinformation und der Spitzensportförderung sowie die zentrale wissenschaftliche Fachgesellschaft – sollen nun kurz vorgestellt werden.

3.1 Zentralbibliothek der Sportwissenschaften der Deutschen Sporthochschule Köln (ZBS)

Die ZBS (www.zbsport.de) ist mit einem Bestand von über 375.000 bibliographischen Einheiten und 1.600 Zeitschriften die weltweit größte Bibliothek des Sports und der Sportwissenschaft. Sie dient der Deutschen Sporthochschule Köln (DSHS) als Universitätsbibliothek und betreut als wissenschaftliche Spezialbibliothek im Rahmen des DFG-Systems der überregionalen Literaturversorgung durch Sammelschwerpunkte das Sondersammelgebiet 31 (SSG Sportwissenschaft). Der Sammelauftrag der ZBS ist dabei ebenso einfach wie komplex: jeder für die Sportwissenschaft relevante Titel sollte prinzipiell von ihr beschafft werden. In dieser Funktion im Zentrum der sportwissenschaftlichen Literatur- und Informationsversorgung in Deutschland hat die ZBS die Feder-

führung im Projektantrag inne und steht in besonderem Maße für die nachhaltige Weiterbetreuung des Projektes durch die Gewährleistung von Projektkoordination, technischer Betreuung und inhaltlicher Integration.

3.2 Bundesinstitut für Sportwissenschaft (BISp), Bonn

Das BISp (www.bisp.de) ist als dem Bundesinnenministerium (BMI) unterstellte Bundesanstalt für die sportwissenschaftliche Forschungsförderung und für den Transfer von Forschungsergebnissen in den Spitzensport zuständig. Im Rahmen seiner IuD-Aufgaben zur Vermeidung von Doppelforschung hat sich das BISp mit den von ihm erstellten sportwissenschaftlichen Datenbanken SPOLIT, SPOMEDIA und SPOFOR als eine Fachinformationseinrichtung profiliert, die international zu den „big five“ der Sportdokumentation gezählt wird (Ghent 2001, 173f)³. Neben seinen Datenbanken erarbeitet das BISp einen Fachinformationsführer zu sportwissenschaftlich relevanten Internetquellen, der ein weiteres wichtiges Element der künftigen ViFa darstellen wird (Störk 2003).

3.3 Institut für Angewandte Trainingswissenschaft Leipzig (IAT)

Das IAT (www.sport-iat.de) – formal ein von einem Trägerverein im Deutschen Sportbund (DSB) getragenes Institut, das vom BMI gefördert wird - zählt neben dem BISp zu den herausragenden Einrichtungen der sportwissenschaftlichen Fachinformation in Deutschland und betreibt ebenso wie dieses und in der Nachfolge des DDR-Forschungsinstituts für Körperkultur und Sport (FKS) sportwissenschaftliche Grundlagenforschung und praktische Sportförderung. Das IAT steuert mit der Internetdatenbank SPONET und mit der abgeschlossenen bibliographischen Datenbank SPOWIS wichtige Module einer sportwissenschaftlichen *virtual library* (Sandner & Regner 2001) bei und wird sich durch die Retrodigitalisierung rarer Quellen aus dem Bereich des DDR-Sports an der Versorgung der Wissenschaft mit elektronischen Volltexten beteiligen.

3.4 Deutsche Vereinigung für Sportwissenschaft (dvs), Hamburg

Mit der dvs (www.sportwissenschaft.de) ist die zentrale wissenschaftliche Fachgesellschaft der Disziplin und ein Mitglied der IuK-Initiative einer der Träger des Projektes. Diese Konstellation stellt in der Landschaft der Virtuellen Fachbibliotheken (leider) eher eine Ausnahme dar, ist aber, aus Sicht der Part-

3 Als die vier anderen Institutionen der „big five“ nennt Ghent (ebd.) das kanadische Sport Information Resource Centre (SIRC), das australische National Sport Information Centre (NSIC), die Amateur Athletic Foundation of Los Angeles (AAF) und das französische Institut National du Sport et de l'Education Physique (INSEP).

ner, ein gleichsam natürlicher Ausdruck der konsequenten Wissenschafts- und Nutzerorientierung des Projekts. Mit ihren ca. 900 in verschiedenen Sektionen und Kommissionen organisierten Mitgliedern aus allen Bereichen der sportwissenschaftlichen Community kann die dvs entscheidend dazu beitragen, das Angebot der ViFa Sportwissenschaft auf die Bedürfnisse der Sportwissenschaftlerinnen und Sportwissenschaftler hin zu evaluieren und zu gestalten. Zugleich bringt die dvs als Produzent einschlägiger Fachliteratur wichtige digitale Volltexte in das Angebot ein.

3.5 Bibliothek der Friedrich-Ebert-Stiftung (FES), Bonn

Die Bibliothek der Friedrich-Ebert-Stiftung (www.fes.de/library) zählt mit einem Bestand von ca. 730.000 Medieneinheiten zu den großen historischen und sozialwissenschaftlichen Spezialbibliotheken in der Bundesrepublik und wird von der DFG im Rahmen der überregionalen Literaturversorgung gefördert. Sie ist einer der Träger von *ViBSoz*, der Virtuellen Fachbibliothek Sozialwissenschaften, bringt aber ihre zum Teil unikalenen Bestände auch in eine Reihe regionalwissenschaftlicher ViFas wie *Cibera*, *MENALIB*, *NedGuide* und *ViFaOst* ein und ist darüber hinaus in der digitalen Archivierung aktiv (FES 2005). In den Fach-OPAC der ViFa Sportwissenschaft wird die Bibliothek vor allem ihren großen Bestand zur Arbeitersportbewegung einbringen, von dem ein Teil im Rahmen des Projekts digitalisiert und als Volltext online verfügbar gemacht werden soll. Darüber hinaus stellt sie Dokumente der bundesdeutschen Parteien zur Sportpolitik bereit.

3.6 Kooperations- und Technologiepartner

Die Entscheidung zum Aufbau einer eigenständigen ViFa Sportwissenschaft ist gekoppelt an die Kooperation mit verschiedenen weiteren Partnern – ein Konzept der informationellen Autarkie macht weder Sinn noch entspräche es dem oben angesprochenen trans- bzw. multidisziplinären Charakter der Sportwissenschaft. Außer der sportwissenschaftlichen Community als einem ständigen Gesprächs- und Kooperationspartner sind hier besonders *vascoda* und das Netzwerk der ViFas zu nennen, außerdem weitere Bibliotheken und Einrichtungen der wissenschaftlichen Fachinformation sowie bibliothekarische und informationswissenschaftliche Dienstleistungszentren.

Eine Reihe von Institutionen haben ihre Bereitschaft, als *Kooperationspartner* mit der Projektgemeinschaft der ViFa Sportwissenschaft zusammenzuarbeiten, mit einem *letter of intent* dokumentiert. Die Art dieser Zusammenarbeit unterscheidet sich naturgemäß von Fall zu Fall stark. Die Bandbreite reicht von der Beteiligung an der Evaluation der ViFa und der Mitarbeit in ihrem wissenschaft-

lichen Beirat über die Bereitstellung von Katalogdaten für den virtuellen Fach-OPAC bis hin zum Austausch von Datenbankinhalten:

- Deutscher Sportbund, Frankfurt (DSB)
- Sportwissenschaftliches Institut der Universität des Saarlandes, Saarbrücken (SWI)
- Arbeitsgemeinschaft für Sportpsychologie in der Bundesrepublik Deutschland (asp)
- Arbeitsgemeinschaft Sportwissenschaftlicher Bibliotheken (AGSB)
- Deutsche Zentralbibliothek für Medizin, Köln (ZB MED)
- Deutsches Institut für Internationale Pädagogische Forschung, Frankfurt (DIPF)
- Informations- und Dokumentationsstelle am Institut für Ernährungswissenschaft, Justus-Liebig-Universität Gießen
- Verbundzentrale des Gemeinsamen Bibliotheksverbundes GBV, Göttingen (VZG)

Die technologischen Herausforderungen, die das Projekt für die Trägergemeinschaft mit sich bringt, ließen es ihr als ratsam erscheinen, sich für die Bewältigung der Aufgaben der Expertise eines kompetenten und erfahrenen *Technologiepartners* zu bedienen. Es wurde deshalb entschieden, die ViFa Sportwissenschaft in enger Kooperation mit dem Hochschulbibliothekszenrum Nordrhein-Westfalen (HBZ) aufzubauen und sie von ihm hosten zu lassen. Da das HBZ (www.hbz-nrw.de) seit Anfang 2005 für den technischen Betrieb des Dachportals *vascoda* zuständig ist und die von ihm dabei verwendete Software *Information Portal Suite (IPS)* auch beim Betrieb der ViFa zum Einsatz kommen soll, sind von diesem Schritt, der eine Besonderheit der ViFa Sportwissenschaft darstellt, wichtige Synergieeffekte zu erwarten. Dies gilt etwa für Probleme wie die Barrierefreiheit (*accessibility*) des Angebots, für Profildienste und Personalisierungsfunktionen, für die Durchführung der Metasuche oder für die Integration der ViFa in *vascoda*.

4 Module der ViFa Sportwissenschaft

Wie bei den meisten der bereits realisierten ViFas orientiert sich das Konzept der ViFa Sportwissenschaft an einem *core set* von Modulen, die allgemein als unverzichtbar angesehen werden (DFG 2004, Rösch 2004) und die per Metasuche und inhaltliche Integration über sacherschließende Metadaten einer gemeinsamen Recherche zugänglich gemacht werden. Im Folgenden soll nur ein summarischer Überblick über das Angebotsspektrum der projektierten ViFa gegeben

werden, da eine detaillierte Vorstellung der einzelnen Module den vorgegebenen Rahmen sprengen würde.

4.1 Datenbankmodule und Metasuche

Den informationellen Kern der ViFa bilden die verschiedenen Datenbanken, die den Zugriff auf die sportwissenschaftlich relevanten Informationsressourcen ermöglichen, sowie die integrierte *Metasuche* über die verteilt nachgewiesenen Bestände. Die Metasuche wird, wie oben erläutert, durch die Kooperation der Projektgemeinschaft mit dem HBZ und durch den Einsatz der Portal-Software IPS realisiert. Diese Software ermöglicht, analog zu der ebenfalls mit IPS aufgebauten Digitalen Bibliothek NRW (www.digibib.net) des HBZ, eine Vielzahl von nutzerfreundlichen und hilfreichen Services. Hierbei sind besonders die Personalisierungsfunktionen und Profildienste zu nennen, die es den Nutzern ermöglichen, Suchabos oder individualisierte Datenbankprofile festzulegen. Der Kernbestand an Datenbanken der ViFa Sportwissenschaft soll in der Folge kurz skizziert werden.

4.1.1 Bibliothekskataloge

Ziel dieses Moduls ist der Aufbau eines virtuellen Fach-OPACs, der neben dem ZBS-OPAC als dem zentralen Katalog des SSG Sportwissenschaft sportbezogene Auszüge aus weiteren relevanten Katalogen enthält. In der Aufbauphase werden zunächst folgende Kataloge einbezogen, die Integration weiterer sportbezogener Sonderbestände wird im Projektverlauf geprüft werden:

- Onlinekatalog der ZBS als OPAC des SSG Sportwissenschaft
- Segment Sportmedizin des OPACs der Deutschen Zentralbibliothek für Medizin
- Segment Arbeitersport des OPACs der Bibliothek der Friedrich-Ebert-Stiftung

4.1.2 Fachinformationsführer

Ein qualitätskontrollierter *subject gateway* zu nach sportwissenschaftlichen Kriterien ausgewählten und erschlossenen Internetquellen wird aktuell vom BISp auf seiner Plattform SPORTIF (www.sport-if.de) aufgebaut und wird später in das Angebot der ViFa integriert werden. Dabei nutzt das BISp mit DB Clear eine vom Informationszentrum Sozialwissenschaften Bonn entwickelte Software, die auch in anderen ViFas (ViBSoz, ViFaArt) zum Einsatz kommt. Ermöglicht wird sowohl ein Suchzugang als auch ein systematischer Browsingzugang (Störk 2003).

4.1.3 Online Contents Sportwissenschaft

Die ZBS wird in Kooperation mit der GBV-Verbundzentrale eine Datenbank mit Nachweisen aktueller Zeitschriftenaufsätze (Current-Contents-Dienst) nach dem bewährten Muster der Online Contents der Sondersammelgebiete (OLC-SSG)⁴ aufbauen und in ihre Metasuche integrieren. Aus der Recherche in der entstehenden Datenbank OLC-SSG Sportwissenschaft können nahtlos die recherchierten Zeitschriftenartikel per Lieferdienst bestellt werden.

4.1.4 Fachdatenbanken

Die Fachdatenbanken von BISp und IAT stehen im Zentrum der sportwissenschaftlichen Fachinformation in Deutschland. Die Datenbanken beider Anbieter bieten durch Abstracts und inhaltsbeschreibende Metadaten wichtige Mehrwertdienste und decken weite Teile der deutschen und der internationalen sportwissenschaftlichen Forschung ab. Die *Fachdatenbanken des BISp* (Kloock 2003, Störk 2003) sind bereits in *vascoda* eingebunden und werden über die BISp-Plattform www.sport-if.de in das Angebot der ViFa integriert⁵. Im Einzelnen handelt es sich dabei um die Datenbanken:

- SPOLIT (ca. 147.000 bibliographische Nachweise),
- SPOFOR (Nachweis von ca. 5.200 Forschungsprojekten),
- SPOMEDIA (Nachweis von ca. 1.500 AV-Medien).

Auch die *Fachdatenbanken des IAT* sind bereits über die Internetseiten des IAT (www.sport-iat.de) online verfügbar und konzentrieren sich gemäß der Ausrichtung des Instituts vor allem auf den Bereich der Trainingswissenschaft (Sandner & Regner 2001). Es handelt sich dabei um die beiden folgenden Datenbanken:

- SPOWIS (ca. 120.000 inhaltlich erschlossene Titelnachweise aus den Jahren 1960-1995 einschließlich grauer Literatur und ehemals geheimer Forschungsberichte),
- SPONET (Nachweis von ca. 10.000 Internetressourcen, zumeist Volltexte aus dem Bereich der Trainingswissenschaft).

4.1.5 Weitere Datenbanken

Analog zu den Bibliothekskatalogen sollen auch die Fachdatenbanken der Projektpartner durch andere Fachdatenbanken bzw. durch sportbezogene Auszüge aus Datenbanken anderer Anbieter und durch Kooperationen mit anderen ViFas (wie dem Fachportal Pädagogik des DIPF) ergänzt werden. Im Projektverlauf sollen zu den im Folgenden aufgeführten Angeboten noch weitere hinzutreten:

4 Siehe dazu <http://www.gbv.de/du/info/ViFa.shtml>.

5 Darüber hinaus sind sie nach wie vor online unter www.bisp-datenbanken.de verfügbar.

- *FIS Bildung Literaturdatenbank*: Integration eines sportpädagogischen und sportdidaktischen Auszugs in Kooperation mit dem Deutschen Institut für Internationale Pädagogische Forschung, Frankfurt (DIPF),
- *Nutrisport*: Informationssystem zur Sporternährung der IuD-Stelle am Institut für Ernährungswissenschaft der Universität Gießen (www.nutrisport.de),
- *E-Learning-Content* und Volltextdatenbank aus der Kooperation von dvs und dem Sportwissenschaftlichen Institut der Universität des Saarlandes (SWI).

4.2 Verbesserung der elektronischen Volltextversorgung

Ein wichtiges Anliegen der Projektgemeinschaft ist es, die Versorgung der Sportwissenschaftlerinnen und Sportwissenschaftler mit relevanten E-Volltexten zu verbessern. Im Rahmen des Projekts sollen deshalb drei Digitalisierungsprojekte durchgeführt werden, wobei die gewonnenen Digitalisate über die oben angeführten Datenbanken recherchiert und aufgerufen werden können. Es handelt sich dabei um folgende Bestände:

- ca. 30 Bände der „Schriftenreihe“ der dvs zu verschiedenen Themenbereichen,
- die Zeitschrift „Theorie und Praxis des Leistungssports“, die 1963-1990 in der DDR erschien und als „geheime Dienstsache“ galt (IAT),
- ca. 450 Bücher und Broschüren aus der Zeit vor 1945 zur Geschichte des Arbeitersports (Bibliothek der Friedrich-Ebert-Stiftung).

4.3 Inhaltliche Integration durch Crosskonkordanzen

Von entscheidender Bedeutung für die Errichtung einer nutzerfreundlichen übergreifenden Suche in einer ViFa ist die Lösung des Problems bestehender Unterschiede der inhaltlichen Erschließung. Die ViFa Sportwissenschaft will eine Optimierung des Retrievals durch eine intelligente Transformation der Suchanfragen anhand von intellektuell erstellten Crosskonkordanzen erreichen (Krause 2004). Zu diesem Zweck sollen als Teil des Projektantrags semantische Entsprechungen zwischen der feingegliederten Fachsystematik der ZBS und den Deskriptoren des BISp und des IAT identifiziert werden, die dann als terminologische Grundlage einer intelligenten Metasuche dienen können. Ergänzt wird diese ‚sportinterne‘ Heterogenitätsbehandlung durch Crosskonkordanzen zwischen den SPOLIT-Deskriptoren und den Deskriptoren der ‚sportexternen‘ Datenbanken FIS-Bildung, SOLIS und Psyndex, die das BISp in Kooperation mit dem Informationszentrum Sozialwissenschaften in Angriff genommen hat.

4.4 Wissenschaftlicher Beirat und Evaluation

Der Gradmesser für den nachhaltigen Erfolg einer ViFa ist die Akzeptanz und Nutzung durch die wissenschaftliche Community. Es ist deshalb das Konzept der Projektgemeinschaft, die Wissenschaftlerinnen und Wissenschaftler in den Aufbau und die Weiterentwicklung der ViFa einzubinden. Ihre kritische Begleitung durch einen fachwissenschaftlichen Beirat soll sicherstellen, dass Vertreter der Zielgruppe einen festen Platz in den Strukturen der ViFa haben und von dort aus die Nutzerinteressen vertreten können. Der Beirat wird aus Einzelpersonen aus verschiedenen Bereichen der Disziplin und aus institutionellen Mitgliedern bestehen, wobei der Deutsche Sportbund (DSB) und die AG Sportwissenschaftlicher Bibliotheken (AGSB) ihre Mitarbeit bereits zugesagt haben. Eine ständige Qualitätskontrolle soll jedoch nicht allein durch den Beirat erfolgen. Für die Kommunizierung des Angebots in die Community und für die Durchführung von Evaluierungsverfahren ist die Beteiligung der wissenschaftlichen Fachgesellschaft dvs als ein besonderer Vorteil der ViFa Sportwissenschaft anzusehen. Darüber hinaus werden kooperierende Facharbeitsgemeinschaften wie die AGSB oder die AG für Sportpsychologie (asp) durch die Expertise ihrer Mitglieder zur Optimierung und Kontrolle des Angebotes beitragen und auch am kooperativen Aufbau einzelner Module (Fachinformationsführer) partizipieren können.

Mit diesem Punkt – der Einbindung der Wissenschaftlerinnen und Wissenschaftler in den Aufbau und die Weiterentwicklung der ViFa – ist dieser kurze Projektbericht wieder zu seinem Ausgangspunkt zurückgekehrt: das hier nur grob zu skizzierende „One Stop Portal“ stellt ein Desiderat der Sportwissenschaft dar und ist in engem Kontakt mit der *scientific community* als Angebot von Fachwissenschaftlern für Fachwissenschaftler aufzubauen. Einzelne Module der ViFa liegen bereits vor oder sind im Aufbau. Es bleibt zu hoffen, dass auch das Gesamtvorhaben in naher Zukunft verwirklicht werden kann. Die Projektgemeinschaft wird unter der URL www.vifasport.de über den Fortgang des Projekts berichten.

5 Literaturverzeichnis

- Borkenhagen, F. (1995): Immer ein Bit voraus... – Zur Einführung in das Schwerpunktthema. In: dvs-Informationen 10, H. 2, S. 4-6.
- Borkenhagen, F. (2001): <www.sportscience.de> – Gedanken zu einem künftigen Internet-Portal der deutschen Sportwissenschaft. In: dvs-Informationen 16, H. 1, S. 30-32.
- Deutsche Forschungsgemeinschaft (2004): Das DFG-System der überregionalen Sammelschwerpunkte im Wandel. Weitere Schritte zur Umsetzung des

- Memorandums zur Weiterentwicklung der überregionalen Literaturversorgung. In: Zeitschrift für Bibliothekswesen und Bibliographie 51, H. 5-6, S. 328-345.
- Deutsche Forschungsgemeinschaft (2005): Jahresbericht 2004: Aufgaben und Ergebnisse. Bonn: Deutsche Forschungsgemeinschaft.
- Elpel, K.-P. (2000): Sportwissenschaft und vernetzte digitale Fachinformation. Zustand, konzeptionelle Ansätze und Perspektiven computergestützter Verfahren für Forschung und Lehre im Hochschulbereich. Hamburg: Edition Akademie.
- Friedrich-Ebert-Stiftung (Hrsg.) (2005): Soziale Bewegungen online. 8. Inet-Bib-Tagung in Bonn vom 03.-05. November 2004. Projekt der Friedrich-Ebert-Stiftung und ihrer Partnerorganisationen. Bonn: Bibliothek der Friedrich-Ebert-Stiftung (Veröffentlichungen der Bibliothek der Friedrich-Ebert-Stiftung, 17).
- Ghent, G. (2001): The Development of Sports Documentation and Information Centres: Highlights from Recent History. In: International Association for Sports Information (Hrsg.): Sports Information in the Third Millenium. Proceedings of the 11th IASI World Congress, Lausanne, 25th, 26th and 27th April 2001. Lausanne: Olympic Museum and Studies Centre, S. 171-186.
- Görlitz, D. (2005): Kooperative Portalsysteme in der Wissenschaft. In: Wiemeyer, J. (Hrsg): Education, Research and New Media. Chances and Challenges for Science. 10. Tagung der IuK-Initiative der wissenschaftlichen Fachgesellschaften vom 11.-14. März 2004 in Darmstadt. Hamburg: Czwalina (Schriften der Deutschen Vereinigung für Sportwissenschaft, 146), S. 69-78.
- Johannsen, J. (2005): Tagungsbericht: InetBib 2004 in Bonn. In: Bibliotheksdienst 39, H. 1, S. 118-122.
- Kloock, W. (2003): BISp-Datenbanken liefern fast 150.000 Dokumente. SPOLIT komplett im Internet. In: BISp-Jahrbuch, S. 73-80.
- Krause, J. (2004): Konkretes zur These, die Standardisierung von der Heterogenität her zu denken. In: Zeitschrift für Bibliothekswesen und Bibliographie 51, H. 2, S. 76-89.
- Rösch, H. (2004): Virtuelle Fachbibliotheken - in Zukunft Fachportale? In: Information - Wissenschaft & Praxis 55, H. 2, S. 73-80.
- Sandner, H.; Regner, R. (2001): Development and Application of a Virtual Library in Sports Science. In: International Association for Sports Information (Hrsg.): Sports Information in the Third Millenium. Proceedings of the 11th IASI World Congress, Lausanne, 25th, 26th and 27th April 2001. Lausanne: Olympic Museum and Studies Centre, S. 191-201.

Störk, M. (2003): Die Weiterentwicklung von SPORTIF: Die Integration der Online-Datenbanken und des Fachinformationsführers Sport im Wissenschaftsportal vascoda. In: BISp-Jahrbuch, S. 81-92.

Pianos, T. (2005): Was macht vascoda? Vision und Wirklichkeit. In: Zeitschrift für Bibliothekswesen und Bibliographie 52, H. 2, S. 67-78.

Angaben zum Autor

Dr. Jochen Johannsen, MA (LIS)
Technische Universität Berlin
Universitätsbibliothek im VOLKSWAGEN-Haus
Fasanenstr. 88
10623 Berlin

Studium der Geschichte und Germanistik in Freiburg, Köln und Witten/Herdecke (Promotion 2004), postgraduales Fernstudium der Bibliothekswissenschaft (Humboldt-Universität Berlin), 2004/05 an der Zentralbibliothek der Sportwissenschaften der Deutschen Sporthochschule Köln Koordinator des Projekts ViFa Sportwissenschaft, seit Oktober 2005 tätig an der Universitätsbibliothek der TU Berlin

Interdisziplinäre Aufsatzdatenbanken und ihre Verlinkung zum (elektronischen) Volltext

Heike Schulte-Derne

Fachhochschulbibliothek Ludwigsburg

Zusammenfassung

Die stetige Zunahme sowohl von Zeitschriftentiteln als auch von wissenschaftlichen Publikationen in Aufsatzform erfordert neue Werkzeuge für die umfassende Suche nach Zeitschrifteninhalten. Mittlerweile existiert eine Vielzahl von frei zugänglichen Rechercheangeboten, die durch graphische Oberflächen leicht zu handhaben sind und umfangreiche Ergebnisse liefern. Problematisch ist meist jedoch die an eine Recherche anschließende Beschaffung der gewünschten Artikel. Der „klassische“ Weg, die erneute Recherche in lokalen Bibliothekskatalogen oder Verbänden mit anschließendem Kopieren oder Bestellen über Fernleihe, scheint heutzutage, da Arbeitsgänge zunehmend vereinfacht und elektronisiert werden, nicht mehr zeitgemäß. Ein Großteil der Benutzer wünscht sich so genannte „One-Stop-Shops“, die von der Recherche bis zum gewünschten Treffer im Volltextformat alles aus einer Hand anbieten und zeitraubende Zwischenschritte und Medienbrüche ersparen.

Welche fachübergreifenden Aufsatzdatenbanken enthalten weiterführende Angebote? Wie wird der Zugang zu elektronischen Volltexten geregelt? Welche Kosten entstehen? Welche Angebote machen die eigenen Bibliotheksbestände effektiv benutzbar? Wie funktionieren Pay-per-View-Verfahren für Einzelnutzer ohne Abonnement? Werden Verlinkungen benutzerbezogen und kontextsensitiv erzeugt oder ergeben sich Klicks ins Leere?

Der Beitrag betrachtet aus dem umfangreichen im Internet frei zugänglichen Angebot zwei fachübergreifende Aufsatzdatenbanken (IBZ Online, Springer-Link) in der Konzeption ihrer Inhalte, den Konditionen des Zugriffs auf die Suchmasken und Treffer sowie der Usability der Recherchemasken und Trefferanzeigen. Ein Fokus liegt auf Zusatzdiensten wie Speichermöglichkeiten von Suchanfragen und Titellisten sowie dem Anlegen von kontextsensitiven Einstellungen oder Benutzerprofilen für Current-Awareness-Dienste. Schließlich werden die bestehenden Verlinkungen der Datenbanken zu anderen Nachweisinstrumenten wie der EZB oder zu lokalen Verbänden, zu Pay-per-View- und Dokumentenliefersystemen oder – idealerweise – direkt zum gesuchten elektronischen Volltext dargestellt und erläutert.

1 Situation auf dem Zeitschriftenmarkt

In den letzten Jahren ist eine stetige Zunahme sowohl von Zeitschriftentiteln als auch von Publikationen in Aufsatzform zu erkennen. In vielen Bereichen, vor allem den Naturwissenschaften, werden neue Forschungsergebnisse hauptsächlich, teils sogar ausschließlich in Zeitschriften veröffentlicht, um eine schnellere Verbreitung der Inhalte zu erreichen. Für die Erstellung wissenschaftlicher Arbeiten wird somit auch die Recherche nach Zeitschriftenartikeln immer wichtiger. Viele Verlage haben diesen Bedarf an effektiven Suchwerkzeugen erkannt und bieten mittlerweile eine Vielzahl von Zeitschriftenaufsatzdatenbanken, oftmals frei über das Internet zugänglich, an. Durch zunehmend graphische Benutzeroberflächen sind die Datenbanken relativ leicht zu bedienen und garantieren aufgrund großer Mengen erschlossener und aufbereiteter Inhalte schnellen Sucherfolg.

Für die weitere Verwendung der gefundenen Treffer spielt die Verlinkung der Datenbank mit weiteren Nachweisinstrumenten, Dokumentenliefersystemen oder idealerweise den elektronischen Volltexten eine entscheidende Rolle. Ein Großteil der Benutzer wünscht sich so genannte „One-Stop-Shops“ (vgl. Träger, 2001, S. 179 und Swets Blackwell 2003, S. 182), bei denen sie von der Recherchemöglichkeit bis zum Treffer im Volltextformat alles aus einer Hand bekommen.

Die vorliegende Arbeit betrachtet zum einen die IBZ Online, die von ihrer Konzeption und ihrem Preis vorrangig eine Datenbank für Bibliotheken darstellt. Sie erschließt Zeitschrifteninhalte fach- und verlagsübergreifend und bietet kontextsensitive Verlinkungen. Dem gegenüber wird die Web-Plattform der Springer-Verlagsgruppe, SpringerLink, dargestellt, die nur Publikationen der eigenen Verlage erfasst, diese aber auch dem Einzelnutzer ohne Abonnements über Pay-per-View-Systeme zugänglich macht.

2 Internationale Bibliographie der geistes- und sozialwissenschaftlichen Zeitschriftenliteratur (IBZ)

2.1 Datenbankbeschreibung

Die internationale Bibliographie der geistes- und sozialwissenschaftlichen Zeitschriftenliteratur (IBZ) erscheint seit 1896. Seit 2001 zeichnet der K.G. Saur-Verlag in München verantwortlich für die Bibliographie. Sie erscheint neben der Print- und CD-ROM-Version seit 1995 auch online. Die Online-Version wird zunächst zusätzlich und kostenlos zur CD-ROM-Ausgabe, seit 1999 auch separat vertrieben. Für eine Lizenz (ein gleichzeitiger Zugriff) entstehen Kosten in

Höhe von 3.014 € pro Jahr. Im Zusammenhang mit dem Kauf der gedruckten oder der CD-ROM-Version kostet die Online-Lizenz zusätzlich 1.390 €. Der Zugriff auf die Datenbank ist generell zulassungspflichtig. Der Gemeinsame Bibliotheksverbund (GBV) fungiert für diese Aufsatzbibliographie als Host und für die teilnehmenden Verbundbibliotheken sowie andere lizenzierte Institutionen ist so der Zugang über die Oberfläche des GBV möglich. ReDI bietet für seine Teilnehmerbibliotheken ebenfalls seit Oktober 2003 Zugriff auf die beim GBV gehostete Online-Version an. Die Daten der IBZ sind darüber hinaus auch in JADE eingespielt und somit für lizenzierte Nutzer von JADE-Anbietern auch über diese Oberfläche recherchierbar. Generell ist ein einmonatiger kostenloser Testzugang für Institutionen möglich.

Derzeit sind etwa 2.500.000 Zeitschriftenaufsätze aus ca. 11.000 Zeitschriften nachgewiesen. Jährlich wächst die Datenbank um etwa 120.000 weitere Aufsatznachweise. Eine Liste der ausgewerteten Zeitschriften ist auf der Homepage des Saur-Verlags abrufbar

(http://www.saur.de/_download/quellenlisten/0000010139.pdf). Das Profil der IBZ bildet eine Art „interdisziplinärer Aufgeschlossenheit“ (Schnelling 2002, S. 378). Der Schwerpunkt liegt mit ca. 3.700 ausgewerteten Periodika aber weiterhin klar auf den Geistes- und Sozialwissenschaften sowie deren angrenzenden Fachgebieten. Für die Sozial- und Wirtschaftswissenschaften werden ca. 1.500, im naturwissenschaftlichen Bereich ca. 2.500 Periodika ausgewertet. Etwa 700 Titel betreffen die medizinischen Fachgebiete. In begrenztem Umfang werden auch populärwissenschaftliche Beiträge berücksichtigt. „Der Schwerpunkt der Bibliographie liegt auf der Auswertung europäischer Zeitschriften. Die Vielfalt der Sprachen, in denen die Zeitschriften publiziert werden, spiegelt die Entwicklung der multilingualen und multikulturellen Forschungslandschaft in Europa wider“ (Saur 2003, S. 2). Dadurch bedingt liegt die Mehrzahl der Artikel in Englisch (51%) und Deutsch (37%) vor. Insgesamt sind jedoch Publikationen in mehr als 40 Sprachen enthalten, darunter hauptsächlich französische, italienische, niederländische und spanische Titel. Die Beiträge werden von Wissenschaftsredakteuren durch Schlagworte in Englisch, Deutsch und teilweise Italienisch in Anlehnung an Personennamen- (PND) und Schlagwortnormdatei (SWD) erschlossen und aufbereitet.

Als Berichtszeitraum der Online-Version gelten die Jahre 1983 bis heute. Das Update der Online-Edition erfolgt, im Gegensatz zur halbjährlichen Aktualisierung der CD-ROM-Version, monatlich.

2.2 Recherchemasken

Die Gestaltung der Suchmasken und Trefferanzeigen erinnert an die Benutzeroberflächen der anderen über den GBV angebotenen Datenbanken, den Verbundkatalog des GBV, die Zeitschriftendatenbank oder die neue Oberfläche

des Südwestdeutschen Bibliotheksverbundes (SWB). Die Benutzeroberflächen können über ein Flaggensymbol in Deutsch und Englisch aufgerufen werden.

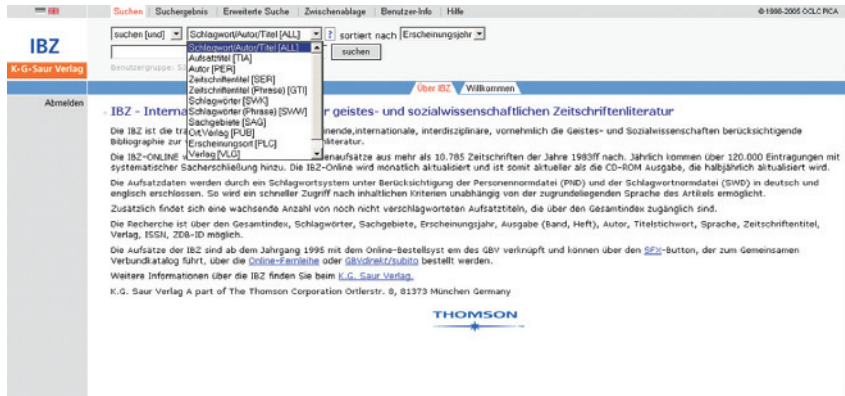


Abb. 1: Einfache Suche der IBZ Online

Auf der Startseite findet man zunächst die einfache Suche (vgl. Abb. 1), bei der man durch Auswahllisten bereits in allen Feldern recherchieren kann. Suchkriterien sind dabei Wörter im Freitext, Personennamen, Aufsatz- und/oder Zeitschriften-/Serientitel, Schlagwörter oder Sachgebiete, die Sprache, der Verlag und Erscheinungsort bzw. das Erscheinungsjahr sowie verschiedene Identnummern (ISSN, IBZ- oder Pica-Nummer). Die eingegebenen Suchbegriffe sind mittels Boolescher Operatoren aus einer Auswahlliste verknüpfbar. Zu jedem Suchbegriff kann mit dem Befehl *Blättern* der Thesaurus herangezogen werden, der englische und deutsche Begriffe enthält. Dabei sind sowohl formale Register für die Personennamen als auch sachliche Indizes für Sachgebiete, Schlagwörter etc. verfügbar. Über das Fragezeichen neben den Auswahlfeldern kann eine Anleitung zum effektiven Gebrauch der Suchfelder angesehen werden. Die Trefferliste lässt sich über Erscheinungsjahr bzw. Relevanz des Treffers bezüglich des Suchbegriffes sortieren. Relevanz bedeutet, dass diejenigen Treffer zuerst angezeigt werden, die der Suchanfrage inhaltlich am ehesten entsprechen. Dies wird festgestellt mithilfe der Ähnlichkeit des eingegebenen Suchbegriffs mit dem Begriff im Treffer sowie der Häufigkeit, mit der der Begriff im gefundenen Dokument auftaucht. Weitere Kriterien sind die Länge des Dokuments (kürzere werden bevorzugt), die so genannte reziproke Dokumentenhäufigkeit (seltener Suchbegriffe werden höher eingestuft) sowie die Übereinstimmung der Suchbegriffe mit den für das Dokument charakteristischen Begriffen. Über die Auswahlliste der Booleschen Operatoren und die Eingabe zusätzlicher Suchbegriffe kann das Suchergebnis anschließend weiter verfeinert werden.

Die erweiterte Suche ist analog zur einfachen Suche in den gleichen Feldern möglich, es sind jedoch vier Suchbegriffe in einem Formular kombinierbar. Sortiert werden kann auch hier nach Erscheinungsjahr und Relevanz. Zusätzlich kann die Suche hier auch auf ein bestimmtes Jahr bzw. einen Zeitraum eingeschränkt werden.

In der oberen Navigationsleiste (vgl. Abb. 1) ist bei allen Arbeitsschritten die *Hilfe* erreichbar. Angeboten werden ausführliche Hilfetexte zur einfachen und erweiterten Suche sowie speziell zu den *Such-Operatoren*, *Suchergebnissen* und den Bestellmöglichkeiten über *Fernleihe* oder *Lieferdienste*. Durch diese Aufteilung sowie die anschaulichen Beispiele kann man die Recherchemöglichkeit effektiv nutzen.

2.3 Output-Funktionen

The screenshot shows the IBZ-Online search results interface. At the top, there are navigation tabs: Suchen, Suchergebnis, Erweiterte Suche, Zeichensablage, Benutzer-Info, and Hilfe. Below this is a search bar with the query 'trebs' and a search button. The main content area displays the search results for 'Ihre Aktion suchen [und] (Schlagwort/Autor/Titel [ALL]) trebs'. The first result is an article with the following details:

- Aufsatztitel:** [gleichheit ohne grenzen? die kosmopolitische überforderung](#)
- Autor:** [trebs, angela](#)
- In:** [information.philosophie](#), 2004, j. 32, n. 5, p. 7-14.
- Sprache des Aufsatzes:** [deutsch](#)
- Suchgebiete:** [philosophie](#), [politik](#)
- Zeitschriftentitel:** [information.philosophie - literach : moser, claudia, verlag](#)
- ISSN:** [1434-5250](#) [ReDi-Bestandsabfrage](#)
- Fernleihe und mehr:** [GBV!](#) [SFX](#)

At the bottom of the result, there are navigation options: '5 von 359', 'Inhaltsverzeichnis', 'alle Aufsätze', 'Alle Hefte', and a 'gehe zu' field with '5 von 359' on the right.

Abb. 2: Treffer-Vollanzeige IBZ-Online

Im Anschluss an eine Recherche wird zunächst eine Kurzzliste mit allen gefunden Treffern angezeigt. Die bibliographischen Daten beschränken sich dabei auf den Titel und Verfasser des Aufsatzes sowie die Angabe des Zeitschriftentitels mit Erscheinungsvermerk. Die Treffer sind mit Symbolen versehen, die die jeweilige Materialart bzw. die Publikationsform (wie Aufsätze, Bücher, Briefe, Online Ressourcen, Zeitschriften/Serien) kennzeichnen. In der Mehrzahl handelt es sich hier um Aufsätze.

Über die Funktion *Titeldaten* bzw. durch das Anklicken eines Artikels kann man in die Vollanzeige der einzelnen Aufsätze wechseln (vgl. Abb. 2). Jeder Titelnachweis besteht in dieser ausführlichen Beschreibung aus Aufsatztitel und Autor sowie Schlagworten und Verweisungen in Deutsch und teilweise Englisch. Die Zeitschrift, in der der Artikel erschienen ist, wird mit Zeitschriftentitel, Erscheinungsjahr, Jahrgang, Nummer sowie Seitenzahl des Aufsatzes angegeben. Die Sprachen des Aufsatzes und der vorhandenen Abstracts werden

ebenso wie das Sachgebiet von Aufsatz und Zeitschrift vermerkt. Im Gegensatz zur CD-ROM-Edition fehlt hier der Verlag. Über das Feld *Zeitschriftentitel* ist der Aufsatz mit der Gesamtaufnahme der Publikation verknüpft. Die Registerkarte *Suchgeschichte* führt von der Trefferliste als auch von der Einzeltrefferanzeige zu den bislang in dieser Sitzung gestellten Suchanfragen. Über Markierungsmöglichkeiten lassen sich dabei mehrere der gestellten Anfragen mit *und* oder *oder* kombinieren und erneut abschicken.

Downloaden und Ausdrucken der Ergebnisse ist auf allen Ergebnisseiten möglich. Über den Button *Download* in der linken Sub-Navigationsleiste kann man die zum Abspeichern ausgewählten Treffer bestimmen und das Format der Speicherung (Vollanzeige, Kurzanzeige oder International Standard Book Description) festlegen. Der Download der Trefferliste erfolgt in Form einer E-Mail, die dem Benutzer zugeschickt wird. Für einen Ausdruck wird die Suchanfrage mit den entsprechenden Treffern in einem druckbaren Format (HTML) angezeigt. Einzelne Ergebnisse aus der Vollanzeige können in der *Zwischenablage* abgespeichert werden und zu einem späteren Zeitpunkt überarbeitet und gespeichert und/oder gedruckt werden.

2.4 Verlinkungen

Innerhalb einer ausführlichen Trefferanzeige sind sämtliche Bestandteile der bibliographischen Daten verknüpft (vgl. Abb. 2). Über die blau unterlegten Begriffe können neue Suchabfragen mit dem jeweiligen Begriff im Freitext, den Schlagwörtern bzw. Sachgebieten ausgelöst werden. Über den *Zeitschriftentitel* gelangt man zur Gesamtaufnahme der Zeitschrift mit Verlinkungen zu allen Heften und Aufsätzen der Publikation. Unterhalb der Vollanzeige einzelner Artikel gibt es Verknüpfungen mit dem Inhaltsverzeichnis des entsprechenden Zeitschriftenheftes, einem vollständigen Inhaltsverzeichnis der gesamten Zeitschrift sowie einer Auflistung der erschlossenen Hefte des Titels. Diese können dann wiederum auf ein Erscheinungsjahr bzw. einen Erscheinungszeitraum eingeschränkt werden.

Über einen SFX-Button kann eine Bestandsabfrage im GBV erfolgen, eine direkte Bestellung der gewünschten Artikel über Dokumentenlieferdienste (GBV*direkt*/Subito) oder Online-Fernleihe veranlasst werden oder in einer Web-Suchmaschine (u.a. Google, HotBot, MetaCrawler etc.) weitergesucht werden. Die Aufsatzdaten werden dabei größtenteils in Such- oder Bestellmasken übernommen. Für Nutzer, die direkt über ReDI auf die IBZ zugreifen, sind die Aufsätze über die ISSN mit der ReDI-Bestandsabfrage verknüpft. Dadurch kann der Bestand im SWB sowie die Verfügbarkeit eines elektronischen Volltextes für die Heimatbibliothek über das Nachweissystem der Elektronischen Zeitschriftenbibliothek (EZB) eingesehen werden.

Bislang sind über die IBZ nur Verlinkungen zu weiterführenden Nachweissystemen realisiert. Die Übernahme der Aufsatzdaten in die Bestell- und Nachweissysteme stellt jedoch eine komfortable Möglichkeit dar, die gefundenen Treffer weiter zu verarbeiten. Bei der ReDI-Bestandsabfrage gehen jedoch die aufsatzspezifischen Daten verloren und es wird nur noch nach der übergeordneten Publikation und deren Bestandsnachweisen gesucht. Laut Auskunft des Saur-Verlags sind jedoch in naher Zukunft Verlinkungen der Artikeldaten mit elektronischen Volltexten geplant. Wie dies aussehen soll, steht noch nicht fest, es werden noch verschiedene Modelle getestet. Grundsätzlich gibt es Überlegungen, bereits bestehende Volltextdatenbanken der gesamten Verlagsgruppe Thompson Corporation, der der Saur-Verlag angehört, dafür heranzuziehen.

3 SpringerLink

3.1 Datenbankbeschreibung

Urheber des Online-Angebots ist die Springer-Verlagsgruppe mit den beteiligten Verlagen Urban & Vogel, Steinkopff und Birkhäuser und – seit der Fusion 2004 – auch Kluwer Academic Publishers. Seit 1996 wird SpringerLink kontinuierlich aufgebaut. Die laufend aktualisierten Datenbestände sind über Webplattform <http://www.springerlink.de> und von vielen anderen Portalen her möglich. Es verlinken z.B. mehrere Abstract & Indexing Services sowie Zeitschriftenagenturen oder CrossRef auf das Online-Angebot. Die technische Unterstützung von Metapress seit Juni 2003 ermöglicht es, „die Millionen täglicher Zugriffe und Anfragen besser zu bewältigen und einen weltweit effizienten und ausfallfreien Service zu garantieren“ (SpringerLink 2003, S. 1148).

Das Angebot umfasst derzeit ca. 1.250 laufende Zeitschriften und 2500 Online-Bücher in elf Fachgebietenbibliotheken. Das Angebot der Zeitschriftenartikel wird durch elektronische Supplemente wie Simulationen, Video, Ton oder 3-D-Darstellungen ergänzt. Fachlichen Schwerpunkt bilden die Naturwissenschaften, Technik und Medizin – entsprechend dem Profil der beteiligten Verlage. Ca. 80% der Publikationen liegen in englischer Sprache vor. Der Berichtszeitraum für die elektronischen Volltexte ist 1997 bis heute, Abstracts liegen ungefähr seit 1994 vor. Durch das so genannte *Online Journals Archive* wird das Angebot momentan sukzessive um weitere 1,5 Mio. Artikel mit etwa 13 Mio. Seiten Inhalt von vor 1996 ergänzt. Zusätzlich sind so genannte *Online-First-Artikel* enthalten, die vor der gedruckten Publizierung hier recherchiert- und zitierbar veröffentlicht werden.

Der Zugang zu den Inhaltsverzeichnissen, Abstracts, Suchfunktionen und dem Alert-Service für Inhaltsverzeichnisse und Stichwörter ist generell kostenfrei. Zusatzfunktionen wie die Alert-Dienste oder das Zusammenstellen einer

Favoritenliste von Zeitschriftentiteln werden nur für registrierte Benutzer angeboten. Die Anmeldung ist für Einzelpersonen oder Institutionen möglich und in beiden Fällen kostenlos. Institutionelle Nutzer, wie z.B. Bibliotheken, erhalten übergangslosen Zugriff mittels Freischaltung von IP-Adressen. Der *Table of Contents-Alert* sendet dem Benutzer Inhaltsverzeichnisse und Abstracts der neuesten Ausgabe vorher definierter Publikationen via eMail zu. Über die *Keyword Alerts* können Stichwörter festgelegt werden, die für einen Benutzer relevant sind. Erscheint eine Publikation, in der einer dieser Begriffe im Titel oder einem anderen vorher festgelegten Teil der bibliographischen Daten, wie z.B. dem Abstract, Volltext oder Autorennamen auftaucht, so wird der Benutzer darüber informiert. Nach einer Registrierung können ebenfalls bereits bestehende Abonnements gedruckter Titel gemeldet werden. Auf diese Publikationen ist dann im Anschluss an die Recherche ein übergangsloser Volltextzugriff möglich, da der Zugriff auf die Online-Version im Abonnementpreis mit enthalten ist. Neben diesen Kombinationen sind aber auch reine Online-Abonnements erhältlich, die Preise sind vom jeweiligen Zeitschriftentitel abhängig. Für einige Publikationen werden darüber hinaus einzelne freie Ausgaben angeboten.

3.2 Recherchemasken

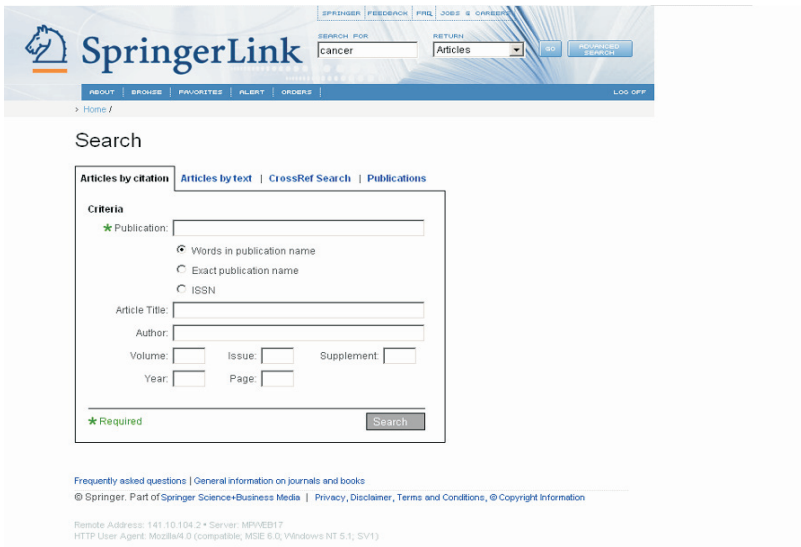


Abb. 3: Erweiterte Suche bei SpringerLink

Die Recherchemasken liegen jeweils nur in Englisch vor. Die einfache Suche befindet sich direkt auf der Startseite sowie auf jeder folgenden Seite im oberen Bereich (vgl. auch Abb. 3). Hier kann eine Stichwortrecherche in den Artikeln, Zeitschriftentiteln oder Verlagsnamen durchgeführt werden. Die erweiterte Suche (vgl. Abb. 3) wird zusätzlich in vier Registerkarten *Articles by citation*, *Articles by text*, *CrossRef Search* sowie *Publications* unterteilt. Erstgenannte ermöglicht die Suche nach Artikeln oder Autoren innerhalb einer bestimmten Veröffentlichung. Dabei sind die Suchkriterien Wörter aus dem Titel der Zeitschrift, der exakte Publikationsname oder die ISSN. Die Suche kann auf einzelne Jahrgänge bzw. Hefte einer Publikation eingeschränkt werden.

Über die *Articles by text*-Suche kann man Stichwörter im Titel oder Abstract eines Aufsatzes, in beidem kombiniert oder im Autorennamen suchen. Hier kann über eine Auswahlliste bestimmt werden, ob man mit der exakten Phrase, mit allen oder mit keinem der Wörter suchen will. Über die *Boolean Search* können Begriffe mit *and*, *or*, *not* oder *near* verknüpft werden. Es ist eine Einschränkung auf die Artikel möglich, für die Volltextzugriff besteht. Über die Funktion *Show Advanced Options* können weitere Spezifizierungsmöglichkeiten wie die Einschränkung auf ein Erscheinungsjahr oder eine Bandnummer sowie die Beschränkung der Suche innerhalb einzelner Publikationen definiert werden. Die Treffer können nach Erscheinungsdatum oder Relevanz sortiert werden. Im Anschluss an eine Suche kann die Trefferliste durch Eingabe weiterer Begriffe im Feld *Search Within Results* spezifiziert werden. Über die Registerkarte *Publications* kann man im Titel und/oder der Beschreibung der Veröffentlichungen suchen. Auch hier können die Suchbegriffe mittels Boolescher Operatoren verknüpft sowie die Suche auf die zugänglichen Volltexte beschränkt werden. Die *CrossRef-Suche* ermöglicht eine Quer-Recherche über Verweise zwischen verschiedenen Publikationen (vgl. dazu <http://www.crossref.org>).

Hilfen zur Recherche sind nur sehr eingeschränkt gegeben. Neben den Suchmasken bei der erweiterten Suche werden am rechten Seitenrand teilweise so genannte *Search Tips* angeboten, über die man Erläuterungen zum Gebrauch der Operatoren abrufen kann. Erklärungen zu den einzelnen Suchfeldern gibt es generell nicht. Auch die FAQs, die am oberen Seitenrand angeboten werden, bieten nur allgemeine Hinweise zu SpringerLink, zur Benutzung der Alert-Dienste sowie zu Problemen bei Abonnements. Die Suche bleibt hier außen vor. Für zusätzliche Unterstützung wird ein eMail-Kontakt angegeben, der bei eigenen Anfragen aber oftmals sehr verzögert oder mit Hinweisen auf die bestehenden Erläuterungen geantwortet hat.

3.3 Output-Funktionen

The screenshot shows the SpringerLink search results page. At the top, there is a navigation bar with the SpringerLink logo and search options. The search criteria are set to 'cancer' and 'Articles'. The results are displayed in a list format, showing the first three items. Each item includes a checkbox, a title, publication information, authors, publisher, issue, pages, and an excerpt. On the right side, there are navigation links for 'First page', 'Previous page', 'Next page', and 'Last page', along with a 'Search Within Results' section.

Search Results

Criteria: cancer 1 to 10 of first 1,000

Filter Selected Items OR Export Selected Citations: RIS Text
Select All Deselect All

1. **The association of cancers of the larynx with cancers of the lung**
Publication: European Archives of Oto-Rhino-Laryngology (Historical Archive)
Authors: A. Rinaldo, C. Marchioni, L. Faggionato, A. Ferlito, U. Saffiotti
Publisher: Springer-Verlag GmbH
Issue: Volume 253, Numbers 4-5
Pages: 256 - 259
Excerpt: We report 77 cases of associated primary cancers of the larynx and lung that were managed at the Department of Otolaryngology, University of Padua, between 1993 and 1994. To our knowledge, this is the largest series reported in the literature to date. This association ...

2. **Risk factors of female cancers in ragusa population (sicily) - I. Endometrium and cervix uteri cancers**
Publication: European Journal of Epidemiology (Historical Archive)
Authors: R. Cusimano, O. Dardanoni, L. Dardanoni, M. Rosa, G. Pavone, R. Tumino, L. Galà
Publisher: Springer Science+Business Media B.V., Formerly Kluwer Academic Publishers B.V.
Issue: Volume 5, Number 3
Pages: 363 - 371
Excerpt: A case-control study on breast, cervix and endometrium cancer cases registered in Ragusa between January 1, 1993 and June 30, 1995 has been conducted. Information on risk factors has been obtained by means of a structured questionnaire.

3. **Diminutive cancers of the colon and rectum: comparison between flat and polypoid cancers**
Publication: International Journal of Colorectal Disease
Authors: A. F. P. K. Laong, F. Seow-Choen, C. L. Tang
Publisher: Springer-Verlag GmbH

Search Within Results
 Search within these results...
 For:

Abb. 4: Trefferliste SpringerLink

Die Artikelnachweise werden mit Titel, Publikation in der der Aufsatz veröffentlicht wurde, Autoren, Verlag, Erscheinungsangaben, Relevanz des Treffers bezüglich des Suchbegriffes sowie einem kurzen Auszug aus dem Abstract angezeigt. Die eingegebenen Suchbegriffe sind dabei im Artikelnachweis grau unterlegt (vgl. Abb. 4). Durch Anklicken der Treffer gelangt man zur Einzeltrefferanzeige mit ausführlicheren Angaben. Dort findet man beispielsweise ein Cover der Zeitschrift, alle am Aufsatz beteiligten Personen mit ausführlichen Kommunikationsdaten sowie den Abstract und die vergebenen Keywords. Auf der rechten Seite findet man verschiedene Exportmöglichkeiten und den Hinweis zur Verfügbarkeit des Artikels. Ist der Aufsatz frei zugänglich bzw. liegt ein Abonnement vor, kann der Artikel direkt als PDF aufgerufen werden. Sind keine Abonnements bzw. Lizenzen hinterlegt, schließen sich die Verlinkungen zum einzelnen Erwerb des Artikels über das Pay-Per-View-System an. Darüber hinaus bestehen so genannte *Linking Options* (vgl. 3.4 Verlinkungen).

3.4 Verlinkungen

Im Anschluss an eine Recherche werden Artikel, die frei erhältlich sind bzw. vorher vom Benutzer subskribiert wurden, direkt mit dem Volltext als PDF verlinkt. Ist der Artikel nur kostenpflichtig erhältlich, so wird dies durch den Hinweis *Full Text Secured* angezeigt. Über *Add To Shopping Cart* und *Purchase Item Now* kann der Artikel zunächst im Einkaufswagen abgelegt und später bezogen oder aber direkt gekauft werden. Der Kauf erfolgt über eine gesicherte Verbindung und öffnet zunächst ein Formular für die Adressinformationen. Bei registrierten, eingeloggtten Nutzern werden diese Daten aus dem angelegten Benutzerprofil übernommen. Abschließend werden die Preise für den Artikel (i.d.R. 30 US\$ plus 16% Steuern, unabhängig von der Seitenanzahl!) aufgeführt und durch Eingabe einer gültigen Kreditkarte wird der Kauf abgeschlossen.

Die *Linking Options* ermöglichen eine direkte Verlinkung auf die Artikel über *MetaPress Direct Linking*, OpenURLs oder Digital Object Identifiers (DOIs). Leider bieten sich diese Links auch nicht registrierten/nicht lizenzierten Benutzern an, sodass man gegebenenfalls erneut auf der Vollanzeige mit dem Hinweis auf die Verfügbarkeit landet, also „ins Leere klickt“.

Ruft man einen Artikel im Volltext auf, weil man ein Abonnement für die Zeitschrift hat bzw. ein freies Heft angefordert hat, so wird dieser mit Publikationstitel, Band, Jahrgang sowie Seitenzahl und den vergebenen DOIs angezeigt.

4 Zusammenfassende Beurteilung

Die hier untersuchten Aufsatzdatenbanken bilden nur einen kleinen, beispielhaften und keineswegs repräsentativen Ausschnitt des Angebots, das derzeit besteht. Es existiert mittlerweile eine Vielzahl von Produzenten und noch mehr Datenbanken, die über unterschiedlichste Vermittler und Anbieter zugänglich gemacht werden. Generell kann man sagen, dass mit den Aufsatzdatenbanken umfangreiche und internationale Sammlungen erschlossen werden. Die Berichtszeiträume gehen teilweise sehr weit zurück, wodurch die Inhalte bezogen auf die erschlossenen Fachgebiete sehr umfassend dokumentiert werden. Bei den Verlagsdatenbanken beschränkt man sich eher auf die aktuellen Inhalte. „Die Berichtszeit geht kaum weiter zurück als bis 1996. Die Effizienz der hier genannten Datenbanken [Verlagsdatenbanken] liegt darin, dass sie auch die Abstracts erfassen, die sie auch (...) in den Suchergebnissen anzeigen“ (Hehl 2001, S. 104). Allerdings werden diese Angebote sukzessive auch um ältere Inhalte erweitert. Die meist kostenlose Einsicht in die Kurzfassungen der Artikel gewährleistet eine effektivere Recherche, zumal die Titel oft sehr wenig aussagen. Sehr gut geeignet sind VerlagsSuchsysteme insbesondere dann, wenn man

bei einem dieser Verlage (mehrere) laufende Abonnements hat, über die man dann nahtlos auf den elektronischen Volltext zugreifen kann. Leider werden diese Verknüpfungen aber oft nicht kontextsensitiv erzeugt, sodass der Volltext rein theoretisch zwar verfügbar ist, ob der Benutzer aber tatsächlich eine Zugriffsberechtigung hat, wird nicht geprüft.

Ein entscheidendes Kriterium ist die Aktualität der Aufsatzdatenbanken, die über das Internet zugänglich sind. Die schnelle, teilweise vor Erscheinen der Printausgabe vollzogene Bereitstellung der elektronischen Ausgaben bietet den Nutzern dieser Angebote einen schnelleren Zugriff auf aktuelle Forschungsliteratur. Daneben sollte allerdings nicht vernachlässigt werden, dass dies – im Gegensatz zur Behauptung der Anbieter – nicht immer der Fall ist. Oftmals ergeben sich Verzögerungen in der Veröffentlichung der Online-Artikel, die dem Nutzer meist verborgen bleiben (vgl. dazu Heinen 2003, S. 584ff).

Die Gestaltung der Suchmasken sowie die Präsentation der Ergebnisse sind größtenteils als gut und funktional zu beurteilen. Die Produzenten der Datenbanken bemühen sich, intuitiv zu bedienende und leicht verständliche Oberflächen zu schaffen. Hilfeseiten werden zum Teil allerdings etwas vernachlässigt und setzen eine gewisse Recherchekompetenz voraus. Dennoch sollten die Benutzer von Bibliotheken dabei nicht außer Acht gelassen werden, ein hoher Schulungsaufwand kann nicht das Ziel der Anbieter sein. Eine einheitliche Gestaltung der Oberflächen, wie sie bei den GBV-Datenbanken GVK, SWB etc. erfolgt ist, ist deshalb zu begrüßen. Die Zusatzdienste, die zu den meisten Datenbanken angeboten werden, sind zielgruppenorientiert. Viele Datenbanken erzielen durch personalisierte Dienste einen Mehrwert, der für Nutzer ein entscheidendes Kriterium im Hinblick auf die Auswahl der Aufsatzdatenbank sein kann.

Zu den Verlinkungen ist allgemein zu sagen, dass nur in wenigen Fällen wirklich gute und unkomplizierte Verlinkungen zu den elektronischen Volltexten bestehen. Ein direkter Zugriff auf die elektronischen Volltexte ist meist nur bei kommerziellen Angeboten wie den Verlagsdatenbanken und nur nach Registrierung laufender Abonnements möglich. In den meisten Fällen jedoch bleibt es bei dem Zugriff auf die bibliographischen Daten und Abstracts. Viele Anbieter ergänzen ihr Angebot deshalb um eine Verknüpfung mit einem (teilweise externen) Dokumentenlieferdienst, mit dem lokalen Verbund des Nutzers oder mit der Elektronischen Zeitschriftenbibliothek, um wenigstens eine weiterführende Recherche ohne Medienbruch zu ermöglichen. Dabei werden mittlerweile meist die Bestelldaten automatisch übernommen.

Lösungen, wie sie die IBZ mit der ReDI-Bestandsabfrage realisiert hat, sind richtungweisend: über integrierte Nachweisinformationen bzw. standortbezogene Bestandsabfragen kann der Nutzer die für ihn relevante Verfügbarkeit des Artikels prüfen, bevor er den ebenfalls integrierten Bestellservice (Lieferdienst oder Fernleihe) in Anspruch nimmt.

5 Literatur

Die vorliegende Arbeit basiert auf folgender Diplomarbeit:

Schulte-Derne, Heike: Interdisziplinäre Aufsatzdatenbanken und ihre Verlinkung zum elektronischen Volltext: Diplomarbeit im Studiengang Bibliotheks- und Medienmanagement (Diplom) an der Fachhochschule Stuttgart - Hochschule der Medien, 2003. – 124 S. - URL:
<http://digbib.iuk.hdm-stuttgart.de/epub/volltexte/2004/367/>

Verwendete Literatur:

Hehl, Hans (2001): Die elektronische Bibliothek - Literatur- und Informationsbeschaffung im Internet. - 2., überarb. und erw. Aufl. - 236 S. - München: Saur, 2001. - 3-598-11496-6.

Heinen, Ingrid; Bickar, Elmar (2003): Untersuchung zur Verfügbarkeit von Zeitschriftenaufsätzen oder ist online wirklich immer schneller als Papier? In: Bibliotheksdienst, 37.2003 [5], S. 584-593.

Saur 2003: Bibliographien zur geistes- und sozialwissenschaftlichen Literatur URL:
http://www.saur.de/service/05_promotion/sammel/Bibl_geistes_d_0703.pdf
f (Zugriff: 27.07.2005)

Schnelling, Heiner (2002): IBZ - Internationale Bibliographie der geistes- und sozialwissenschaftlichen Zeitschriftenliteratur. In: ABI-Technik, 22.2002 [4], S. 378.

SpringerLink 2002a: Springer LINK: Reference Linking. In: Bibliotheksdienst, 36.2002 [3], S. 354.

SpringerLink 2002b: Wissenschaftliches Recherchieren im Internet wird noch einfacher. In: ABI-Technik, 22.2002 [1], S. 71.

SpringerLink 2003: Springer-Verlag: Online-Plattform Springer-Link durch Kooperation mit Metapress strategisch gestärkt. - In: Bibliotheksdienst, 37.2003 [8/9], S. 1148. – und in: ABI-Technik, 23.2003 [2], S. 151-152.

Swets Blackwell (2003): Die Zukunft des Informationsmanagements. In: B.I.T. Online, 6.2003 [2], S. 182-183.

Tröger, Beate (2001): Elektronische Publikationen in der Spezialbibliothek. In: Spezialbibliotheken heute – Wettbewerb und Kooperation. – Jülich: Arbeitsgemeinschaft der Spezialbibliotheken, Sektion 5 im Deutschen Bibliotheksverband, 2001. S. 175 – 192.

Angaben zur Autorin

Dipl.-Bibl. (FH) Heike Schulte-Derne
Bibliothek der
Fachhochschule Ludwigsburg –
Hochschule für öffentliche Verwaltung und Finanzen
Reuteallee 36
71634 Ludwigsburg

10/2000 - 2/2004 Studium Bibliotheks- und Medienmanagement, Hochschule
der Medien, Stuttgart

seit 3/2004 stellvertr. Ltg. HVF Bibliothek Ludwigsburg, Arbeitsschwerpunkte:
Katalogisierung, Auskunft, Datenbankinformation

eTeaching and eLearning in Chemistry

Gregor Fels

University of Paderborn, Faculty of Science, Department of Chemistry

Summary

The pilot project „Networked Chemistry Studies“ (VS-C) sponsored by the ministry for science and education in Germany presents multimedia learning content from all topics in Chemistry and related subjects. Teachers and students use the interactive, chemistry related e-learning platform to build courses and to explore the huge XML based content library. All kinds of visualizations invite the student to conduct computer based experiments. The option to choose between an explorative way of learning and the traditional, book oriented studies increases the motivation of the students.

Results and Discussion

„Vernetztes Studium - Chemie“ (Networked Chemistry Studies) [1] – VS-C – has been one of the leading eLearning projects of the German Federal Ministry of Education and Research [2] during the years 1999 and 2004. In this project, academic participants from Germany, the United Kingdom and Switzerland together with Germany’s Chemistry Information Centre (FIZ CHEMIE Berlin) [3] (see Fig. 1) have developed a novel internet-based modular system for teaching and learning chemistry on the level of „Bachelor in Chemistry“. The project is a milestone in terms of the adjustment of chemical education to the demands

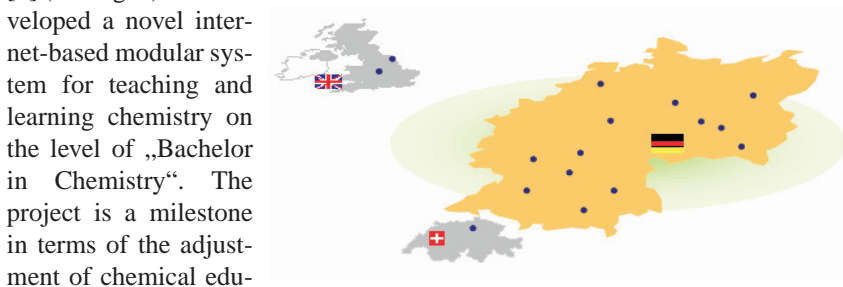


Fig. 1: Location of VS-C participants

of the twentieth century and to the available techniques for multimedia assisted, interactive presentation of chemical knowledge. In addition, the material made available by the

project serves as a German contribution to the “European Higher Education Area” [4,5] and will be used for the harmonization of European chemical education.

The interactive teaching- and learning-platform is an assembly of multimedia knowledge modules from the various chemistry subjects, and it includes such elementary subjects of other sciences as mathematics, physics, toxicology etc. as much as they are part of the chemistry Bachelor program. Throughout the VS-C cross-linking is a big issue (Fig. 2).

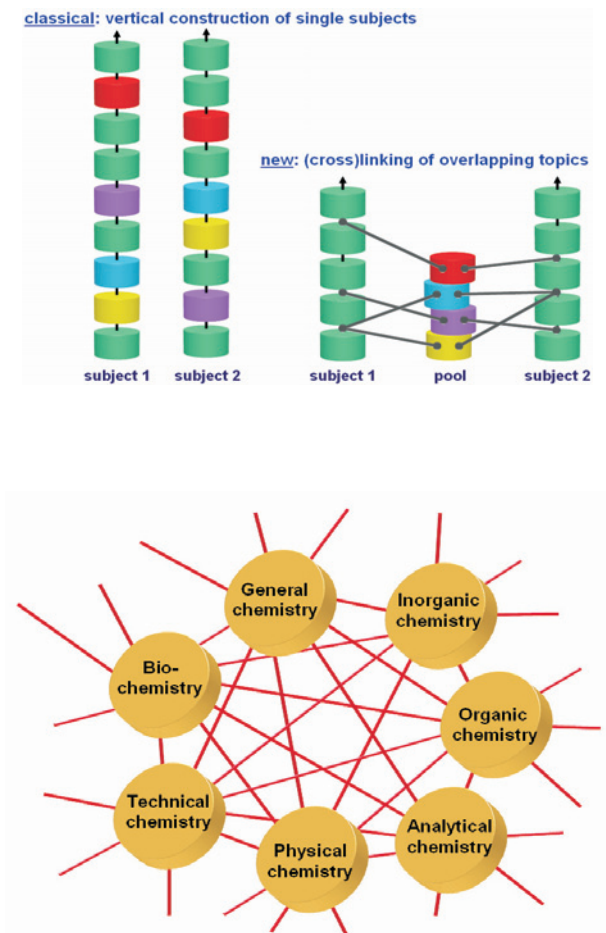


Fig. 2: Crosslinking of subjects with the chemistry curricula (left) and between various chemistry related subjects

While in the more traditional way of teaching chemistry, the various chemical subjects are taught independently from each other, the VS-C project makes profit from the numerous cross-links between them. As an example, the subject “kinetics” is covered say in Organic Chemistry as much as it is necessary to understand the particular issue, for instance a kinetic versus a thermodynamic reaction pathway. However, this material is additionally linked to the more fundamental treatment of kinetics in Physical Chemistry as well as for instance to the biochemical issues of enzyme reaction.

Technically and didactically the chemical knowledge is arranged in “valid learning units” (VLU) [6] which cover a specific learning subject on 5-10 web-pages that can be worked through within 20-40 min depending on the student’s basic knowledge of the particular topic. A VLU starts with a summary page that explains the aim of the VLU and defines the requirements for a successful study, and it gives hints for follow up material. Several of these VLUs are then combined to a trajectory which covers a broader topic.

As an example (Fig. 3), teaching the subject of alkenes in Organic Chemistry will include a general introduction into this family of hydrocarbons, as well as a description of their preparation, reaction and industrial use. These latter topics will be separate VLUs, while the VLU on introductory aspects, could perhaps cover the nomenclature, the MO- and VM-model, the physical properties, or the stereochemistry, which are all described on separate web-pages. Together with the VLUs on application, preparation and reaction of alkenes this will make up a sample trajectory on the alkenes.

It should be noted at this point that so far VS-C has been a German project with all the objects (XML-pages, graphics, flash, video etc.) being presented in German language. It is intended, however, to translate the complete VS-C material to English, a process particularly facilitated by the employed XML-technique [7] throughout the project.

As described in Fig. 4, a VLU is prepared by linking the appropriate objects from the huge database of about 100,000 objects put together during the 5 years of the project. This proceeds via a drag and drop technique that easily allows one to set up, change and rearrange a given VLU according to specific needs. In the same way, a trajectory is built from a series of VLUs. VLUs and trajectories, therefore, can be arranged to fit the specific requirements of the class and can be submitted to the students as the reading for the next lecture or a complete course.



Fig. 3: Example of a trajectory

While the instructor is using the multimedia teaching aids from the web-pages, the student receives the corresponding pages as part of a textbook with all the multimedia tools embedded in the relevant text.

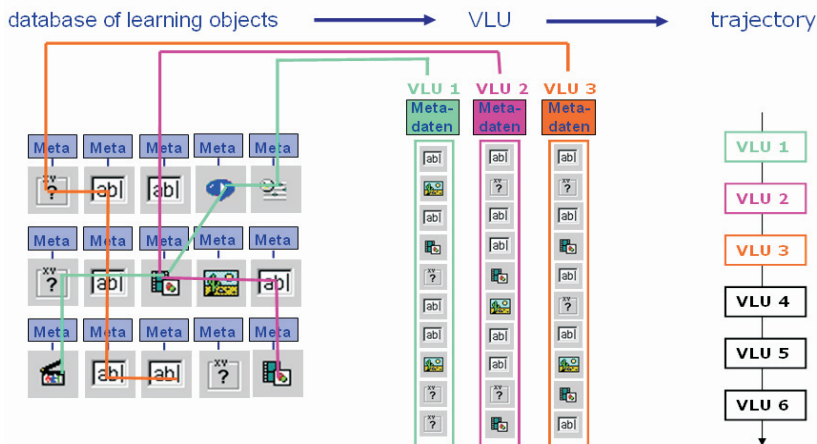


Fig. 4: Exemplary composition of VLU and trajectory

Although in general a user could even change the text and choose the embedded media within a given web-page, the average VS-C user will start from given web-pages in order to set up a VLU or trajectory. On this basis one can search the database for relevant content or one can start from given VLUs or trajectories which are provided by the VS-C authors. Again, all this is done by a simple drag and drop technique. The web-pages provided in the VS-C database are already linked to other VS-C web-pages or VLUs whenever it is appropriate or necessary (Fig. 5).

These links can be to specific other sites of the same chemistry subject, i.e. from an alkene page to various other parts of the Organic Chemistry modules within the VS-C (light blue fields in the above graphic), or to specific topics of other subjects like Theoretical, Macromolecular or Technical Chemistry (red fields), and – as a more general link – to a glossary. This glossary is an open system that is generated automatically via keywords from the pool of available XML-pages of the total VS-C content. It directs the user to either a relevant VLU or to a specific page of a VLU that deals with the particular topic in question (Fig. 6). The underlying technique is based on cross-linking of keywords, which allows one to link not only to already prepared web-pages or VLUs but also to newly prepared material whenever it enters the VS-C database at the FIZ CHEMIE Berlin server [1].

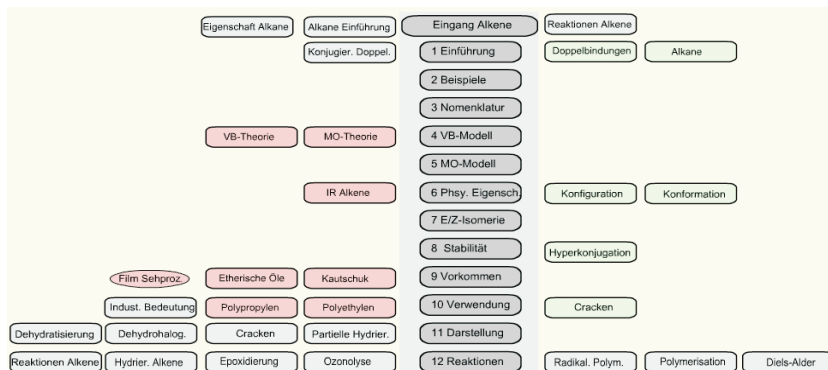


Fig. 5: Exemplary crosslinking of an alkene VLU

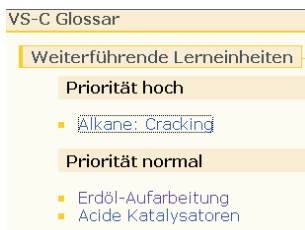


Fig. 6: Glossary example

Results and Discussion

In summary, the VS-C is not a virtual study of chemistry but rather it represents a supplement – not a replacement – for lecturing and learning chemistry. The project is offered as an add on to the conventional text books, representing the material in multimedia knowledge modules, the valid learning units (VLUs), that facilitate teaching and learning the science of chemistry on the level of Bachelor in chemistry.

For further details the reader is pointed to the freely available web-pages of the VS-C project which are hosted at <http://www.vs-c.de>. At this website one can access the presently available content (“Lernmaterial”), and in addition a database of videos with organic chemical experiments as well as some other related material.

References

- [1] <http://www.vs-c.de>
- [2] <http://www.bmbf.de>
- [3] <http://www.fiz-chemie.de>
- [4] http://www.bmbf.de/pub/communique_bologna-berlin_2003.pdf
- [5] http://www.gdch.de/vas/tagungen/archiv/issues_paper.pdf
- [6] <http://www.bibb.de/de/limpact12295.htm>
- [7] <http://www.w3.org/XML/>

Angaben zum Autor

Prof. Dr. Gregor Fels
Universität Paderborn
Fakultät für Naturwissenschaften, Department Chemie
Warburgerstr. 100
33098 Paderborn

1977 Promotion an der Universität Münster (Organische Chemie); 1978-79 Postdoc-Aufenthalt an der University of California, Berkeley; 1979-86 Max-Planck-Institut Dortmund (Neurobiochemie); 1986-2003 WasagChemie Synthesen GmbH (1986-90 Entwicklung und Qualitätskontrolle, 1990-2003 Produktionsleiter, Prokurist), seit 1993 Professor für Organische Chemie an der Universität Paderborn (interzelluläre Kommunikation, Rüstungsaltslasten, Computerchemie).

Generische Portale: Webbasierte Digital Library-Dienste für die Wissenschaft

Stephan Körnig

Technische Universität Darmstadt

Zusammenfassung

Das vom Bundesministerium für Bildung und Forschung geförderte Projekt „Generische Portale“ (Leitung Prof. Dr. Rudi Schmiede, TU Darmstadt) dient der Erarbeitung eines offenen und entwicklungsfähigen Modells für Digital Library-Dienste. Diese sind nach heutigem Stand der Technik als „Web Services“ zu implementieren. Dies ist allerdings ein sehr allgemeiner Begriff, der unterschiedliche Architekturen abdeckt und einen weiten Spielraum für die mögliche Implementierung von Komponenten zulässt. Ein wichtiger Teil des Projekts bestand darin, die unterschiedlichen Technologien im Hinblick auf die Bedürfnisse des Austausches von wissenschaftlicher Fachinformation zu bewerten. Der konkrete Implementierungsvorschlag zielt darauf ab, das HTTP-basierte REST-Modell und die XML/SOAP-gestützten Spezifikationen des W3C zu integrieren. Es wird das konkrete pipelinebasierte Komponenten-Framework vorgestellt, das diese unterschiedlichen Formen des Zugriffs unterstützt. Die Realisierung wird beispielhaft auf der Basis der Net-Umgebung von Microsoft realisiert - die bei der Berücksichtigung der erarbeiteten Richtlinien durchaus den Aufbau einer „Offenen Plattform“ gestattet.

1 Motivation

Im Bereich elektronischer Fachinformation besteht eine große Heterogenität bezüglich der verwendeten Formate (Daten/Metadaten) und Protokolle. Ebenso unterscheiden sich die Anbieter von Inhalten, was die Organisationsform oder ihre Interessen betrifft, beträchtlich. Das Spektrum reicht vom einzelnen Wissenschaftler bis hin zu kommerziellen und organisatorisch „schwergewichtigen“ Anbietern von Fachinformation. Diese verfügen über unterschiedliche Kapazitäten bzgl. der Erschließung von Fachinformation oder der Bereitstellung standardkonformer Technologien.

Es sollte keinesfalls vernachlässigt werden, dass dem einzelnen Wissenschaftler oder dem an Fachinformation Interessierten zunehmend eine bedeu-

tende Rolle in diesem Bereich zukommt. „Distributed Authoring“ und „Self-Publishing“ Modelle sind in bestimmten Bereichen der Fachinformation mittlerweile unverzichtbar. Das Know-how, das einzelne Wissenschaftler bei ihrem Umgang mit Fachinformation entwickeln (Recherchestrategien, Selektionskriterien, Typologien usw.), sollte zudem wieder zurückfließen und die Gestaltung von Informationsangeboten mitbestimmen können.

Als ein Hemmschuh der aktuellen Entwicklung hat sich sowohl die fehlende Standardisierung von DL-Diensten (einschließlich der zugehörigen Meta-Information) erwiesen, als auch das Fehlen von Werkzeugen zur Unterstützung der Anbieter von Inhalten bei der Bereitstellung von Diensten.

Zugleich hat sich bestätigt, dass der Beschreibungssprache XML zunehmend die Bedeutung des Standardformats zukommt. Dieses Format wird sowohl von kommerziellen Anbietern als auch von unterschiedlichen Communities unterstützt. Die XML-basierte Verarbeitung von Fachinformation ist aus unserer Sicht als einzige zukunftsfähig und bietet zugleich (aufgrund der auf dieses Format bezogenen weiteren Standards des W3C-Konsortiums) hervorragende Chancen zur Etablierung von Standards bzgl. der Bereitstellung von Fachinformationen bzw. entsprechenden Diensten.

Wir schlagen eine Strategie vor, die sowohl die Spezifikation von Standards als auch die Entwicklung von konkreten Tools zur standardkonformen Bereitstellung von Diensten unterstützt. Wir sind davon überzeugt, dass eine solche Strategie sogar unabhängig von der künftigen technischen Entwicklung nachhaltige Verbesserungen erzielen wird, da durch sie Aspekte der Erschließung und Strukturierung von Fachinformation gefördert und eine Institutionen übergreifende Spezifikation der Anforderungsprofile von Diensten angestoßen wird.

Ein erster Schritt bestand in der Spezifikation von DL-Diensten auf der Ebene von Diensttypen und Typen von Informationsobjekten. Die aktuelle Arbeit besteht in der Entwicklung einer konkreten Applikation zur Bereitstellung eines solchen Typsystems für DL-Dienste: ein komponentenorientiertes Framework.

Bei der Umsetzung dieser Projektziele wurden die folgenden Prinzipien berücksichtigt:

- Generizität des Ansatzes, und in eins damit die
- Erweiterbarkeit im Hinblick auf fachspezifische Bedürfnisse
- Prinzip des „Selbstenthaltenseins“ (self-containment), also der Bereitstellung von Informationen über den Aufbau von Elementen der Spezifikation im Element selbst
- Nutzung der etablierten Standards
- Unterstützung des anerkannten REST-Modells, das Schnittstellen weitgehend auf dem HTTP-Protokoll basieren lässt und eine „menschenslesbare“

Strukturierung sowohl des Protokolls als auch des „Informationsraumes“ einschließt

- Unterstützung der vom W3C-Konsortium auf den Weg gebrachten Initiativen zum Semantic Web und der Web Services-Aktivität.

2 Web Services

DL-Dienste, wie sie hier verstanden werden, sind verteilte und webgestützte modulare Einheiten, die für den Aufbau fachspezifischer Angebote oder von Portalen genutzt werden können. Dieser Begriff ist erst einmal technisch gefasst - er bezeichnet eine Komponente im Sinne eines Softwaremoduls. Als Grundlage für die Kommunikation oder das Interagieren der Module untereinander dient das HTTP-Protokoll (auf dessen Grundlage mittlerweile fast alle in diesem Bereich verbreiteten Standards implementiert werden, wie OAI-PMH oder SOAP). Inhaltlich wird ein DL-Dienst als ein „Kontrakt“ aufgefasst, der sowohl Zusicherungen über das enthält, was der Dienst leistet, als auch Anforderungen spezifiziert, die für die Inanspruchnahme des Dienstes gelten (formale Korrektheit der Anfrage, Berechtigung zur Nutzung usw.). Beide Aspekte zusammen charakterisieren das, was sich in den letzten Jahren als Grundprinzipien komponentenbasierter Softwareentwicklung für das Web etabliert hat. Sie implizieren damit nur ein Minimum an Technologieabhängigkeit. Die Technologie, die festgeschrieben wird, ist die Plattform, die ohnehin schon anbieterübergreifend genutzt wird: das Internet als World Wide Web.¹

Ein wichtiger Grund für die komponentenbasierte Entwicklung von Web-Applikationen besteht in dem Aspekt der Wiederverwendbarkeit. Mit der modularen Softwareentwicklung bezieht sich dieser Aspekt auf die Softwareentwicklung „im Hause“. Applikationen werden aus Modulen zusammengesetzt, die im Prinzip auch für andere Applikationen benutzt werden können. Im Bereich der Portalentwicklung wird dieser Aspekt der Wiederverwendbarkeit zunehmend auch auf eine echte verteilte Nutzung von Modulen erweitert. Erfolgreiche Sites im Internet basieren bekanntlich auf dem „One-Stop Shop“-Prinzip. Dieses besagt, dass ein potentieller Kunde alle Vorgänge, die ihn dazu bewegen, eine Site zu besuchen (oder die sich im Verlaufe seiner Transaktionen ergeben), auch dort abwickeln kann. Für die Betreiber von Angeboten ergibt sich – neben dem Kostenaufwand – eine hohe Anforderung in Bezug auf die Kompetenzen, die nötig sind, all diese Vorgänge zu unterstützen. Internetportale, die webba-

¹ Diese Begriffe werden nahezu synonym benutzt. Daher sollte der Unterschied auch hier einmal festgehalten werden. Das Internet spezifiziert ein (relativ ausfallsicheres) WLAN, während das „Web“ als verteilte Struktur, als Medium für den Hypertext zu charakterisieren ist.

sierte Komponenten nutzen, können eine entsprechende Vielfalt an Diensten anbieten, ohne diese selbst entwickeln zu müssen. Dienste werden dann als „verteilte“ Module von verschiedenen Anbietern zur Verfügung gestellt, die sich auf deren Entwicklung und Unterstützung spezialisiert haben. Für diese sog. business-to-business-Applikationen (B2B) gibt es entsprechende Frameworks, die diese Modularisierung unterstützen und Richtlinien für die Erstellung, Nutzung und Verfügbarmachung von Komponenten bereitstellen.

Im Bereich der wissenschaftlichen Fachinformation wäre eine solche Herangehensweise ebenso vielversprechend. Gerade in diesem Bereich sind die Kompetenzen der einzelnen Anbieter sehr unterschiedlich und nur die größeren Fachinformationszentren oder Verlage könnten alles aus einer Hand liefern. Zudem: im wissenschaftlichen Umfeld gibt es mitunter kleine Communities, die zu ihrer Unterstützung gleichwohl auf angepasste (domänenspezifische) und hochwertige Werkzeuge zur Nutzung von Fachinformation angewiesen sind. Mit der höheren Modularisierung wächst die Chance, dass die Nutzersicht die Entwicklung von Werkzeugen stärker prägt, als dies bisher geschehen ist. Es gibt hier mögliche Akteure - wie Bibliotheken, Institute, kleinere Fachgesellschaften oder Softwareschmieden -, die ein hochspezialisiertes Fachwissen besitzen und schon bei geringerer Unterstützung Komponenten für DL-Dienste bereitstellen könnten - die aber kaum in der Lage wären, Komplettlösungen zu entwickeln oder zu warten. Es ist daher ersichtlich, dass die Verfügbarkeit und die Akzeptanz von Standards, welche die konkrete Gestaltung der komponentenbasierten Entwicklung von DL-Diensten betreffen, bedeutende Impulse für die Entwicklung in diesem Bereich geben könnte. Insbesondere dann, wenn Frameworks zur Verfügung stehen, die Teile dieser Aufgabe automatisieren könnten.

2.1 SOA - Service Oriented Architecture

Unter dem Begriff „Web Services“ werden gegenwärtig unterschiedliche Konzepte gefasst. Diesen Konzepten ist allein dies gemeinsam, dass Dienste im Internet angeboten werden und dass sie üblicherweise über Schnittstellen verfügen, die (auch) mit einem Web Browser genutzt werden können. Es werden hier unter DL-Diensten nur solche Web-Dienste verstanden, die dem ausgeführten Aspekt der Modularisierung Rechnung tragen und daher geeignet sind, den Aufbau von fachwissenschaftlichen Angeboten auf der Basis einer echt verteilten Struktur Rechnung zu tragen.

Ein engerer Begriff von „Web Services“ bezieht sich auf eine aktuelle Anstrengung zur Standardisierung solcher Dienste durch das W3C. Dieses Konsortium, das von den „Großen“ der IT-Branche unterstützt wird, entwickelt auf der Basis von XML eine Reihe von Standards, die Beschreibungssprachen, Protokolle und Anforderungen an modulare und sich selbst beschreibende Web-Dienste festlegen. Diese Initiative versucht damit, eine sog. „Service Oriented

Architecture“ (SOA) umzusetzen, die das bisher Gesagte über Web-Dienste als Komponenten festschreibt. Auch wenn diese Festschreibungen nicht wesentlich über bisher bekannte Designprinzipien für Komponenten hinausgehen, so ist doch bemerkenswert, dass nun eine anerkannte und Unternehmen übergreifende Standardisierung für diesen Bereich initiiert wurde.

SOA legt die folgenden Punkte fest:

- Dienste sind als Komponenten zu implementieren
- der Zugang zu ihnen muss unabhängig von der technischen Plattform über Web-Schnittstellen möglich sein
- der Zugriff auf sie erfolgt „kontrakt-basiert“, d.h. die Beschreibung der Schnittstelle kann analog zu einem Vertrag verstanden werden, in dem vereinbart wird, welche Aufgaben der Aufrufer zu erledigen hat und welche Funktionen der Dienst erfüllt.
- Dienste müssen dynamisch lokalisiert werden können
- Dienste müssen dem Prinzip des „Selbst-Enthaltenseins“ gehorchen, d.h. sie sind für ihren internen Zustand selbst verantwortlich
- die Kommunikation mit dem Dienst erfolgt über den „Austausch von Botschaften“- es dürfen keine Zugriffe auf interne Zustände oder Datenstrukturen direkt erfolgen

Die dieser Architektur entsprechenden „Web Services“ des W3C erfüllen damit die Forderung nach einer Komponenten-Entwicklung von „Diensten“ als Module, die unabhängig voneinander genutzt werden können. Auf ihrer Grundlage ist es möglich, die eigenen Web-Angebote mit Hilfe von Diensten zu unterstützen, die Dritte entwickelt haben und fortlaufend pflegen bzw. weiterentwickeln. Die sog. UDDI-Registries – einem Standard, der außerhalb der Aktivitäten des W3C entwickelt wurde – gestatten es zudem, Dienste zu publizieren, sie auf einem offenen „Markt“ anzubieten. Eine statische Verknüpfung von Diensten widerspricht dem SOA-Modell. Die Standardisierung der Schnittstellenbeschreibung verspricht, dass Dienste ohne Eingriffe durch Entwickler ausgetauscht (dynamisch „gebunden“) werden können und mit den Registries gibt es entsprechende Anlaufstellen für alle möglichen Nachfrager von Diensten.

2.2 Semantic Services - Probleme der „Standardisierung des Wissens“

Die Web Services des W3C bieten damit anscheinend eine hervorragende Grundlage für die Bereitstellung und den Austausch wissenschaftlich relevanter Informationen. Es ist schon jetzt absehbar, dass diese Web Services auch in diesem Bereich zu einem unverzichtbaren Baustein für den Aufbau webbasierter

Angebote werden. Dies gilt insbesondere für die Bereiche, in denen für ganz konkrete Aufgabenstellungen bestimmte Dienste zu integrieren sind. In dem Anwendungsbereich, den die IuK-Initiative voranbringen will, stellen sich allerdings weitere Anforderungen, bei denen genau zu prüfen ist, ob sie durch diese Technologie erfüllt werden können. Von diesen Forderungen möchte ich zwei herausheben: Leichtgewichtigkeit und Integrierbarkeit unterschiedlicher Semantiken.

Der erste Aspekt basiert auf der Vielfalt der für Fachinformation relevanten „Anbieter“. Will man kleinere Einrichtungen oder gar den einzelnen Wissenschaftler erreichen, dann muss die Einstiegsschwelle zur Nutzung einer Technologie möglichst niedrig sein. Personalkapazitäten für umfangreiche Anpassungs- oder Wartungsarbeiten können hier nicht aufgebracht werden. Der zweite Aspekt betrifft die Eigentümlichkeit wissenschaftlichen Arbeitens. Die Kernkonzepte in den unterschiedlichen Wissenschaften definieren sich über die Verfahren ihrer jeweiligen Gegenstandskonstitution. Unterschiedliche Disziplinen verfügen über unterschiedliche Konzeptualisierungen ihres Gegenstandsreichs, die mitunter nicht aufeinander abbildbar sind. Gleichwohl sind für die jeweilige Disziplin die Informationen, die aus Quellen einer anderen Disziplin stammen, dadurch nicht irrelevant. Es besteht an vielen Nahtstellen das dringende Bedürfnis nach einer fächerübergreifenden Informationsauswertung.

Die „Web Services“ beschreiben ein generisches Modell, das für nahezu alle Zwecke im Bereich webbasierter Dienste eingesetzt werden kann. Um in einem heterogenen Umfeld eingesetzt werden zu können, bedarf es daher spezialisierter Frameworks, die eine konkrete Typisierung vornehmen. Diese Entwicklung ist aus dem Bereich der XML-Formate ja schon seit längerem bekannt. Mit XML lassen sich beliebige Formate, Strukturbeschreibungen, Datenmodelle usw. erzeugen, die zudem explizite Informationen über ihren Aufbau enthalten (sofern die bei der Auszeichnung benutzten Tags mit sinnvollen Bezeichnern definiert werden). Damit sich ein entsprechendes XML-Format auch durchsetzt, muss das Schema, nach dem es aufgebaut ist, von den verschiedenen Nutzern akzeptiert und auf dieselbe Weise genutzt werden - die konkrete Form der Nutzung des XML-Standards muss wiederum konfektioniert oder standardisiert werden.

Bezogen auf Web Services für DL-Dienste ergibt sich ein analoges Problem. Spezielle Typen von Diensten, von „Objekten“, die über diese Dienste bearbeitet werden und von Operationen, die für den Anwendungsbereich typisch sind, müssen mit einem enormen Aufwand an Spezifikationsarbeit „aufgebaut“ werden. Im wohl erfolgreichsten Anwendungsbereich von Web Services, dem B2B-Bereich (s.o.), sind aus diesem Grunde schon umfangreiche Frameworks entstanden. Diese Frameworks erfüllen die Aufgabe, unternehmensspezifische IT-Konzepte zu „normalisieren“. Eine Automation des Zusammenspiels von Web Services unterschiedlicher Anbieter ist nur auf der Grundlage von „geteil-

ten“ Konzepten möglich. Diese Einigung muss weitreichend sein, sie betrifft sowohl Datenmodelle als auch funktionale Zusicherungen für alle Beteiligten, die auf eine gemeinsame Art und Weise „expliziert“ werden müssen. Diese aufwendige Anpassung findet immerhin (im Unterschied zu früheren Lösungen) nur einmal statt: im Übergang von den unternehmensspezifischen Lösungen hin zum Framework. In früheren Szenarien war zwischen allen beteiligten Partnern jeweils eine entsprechende Anpassung nötig. Eine automatisierte Anpassung, wie sie etwa im Rahmen der „Semantic Web“-Spezifikationen möglich gemacht werden sollte, hat sich dagegen als eine Chimäre erwiesen. Im Bereich der Fachinformation potenziert sich dieses Problem – hier gibt es nur wenige Konzepte, die fachübergreifend geteilt werden können. Diese betreffen grundlegende Typen von Diensten, typische Objekte, sofern man diese auf medienspezifische Dokumenttypen reduziert, und die auf diese Typen bezogenen elementaren Schnittstellen. Ein „generisches Framework“ ist aber nur dann leistungsfähig, wenn es - über den Umgang mit diesen grundlegenden (generischen) Typen hinaus - an Aufgaben angepasst werden kann, die für einen bestimmten Anwendungsbereich typisch sind. Diese Anpassung müsste anwendungsübergreifend mit den Mitteln des Frameworks selbst durchgeführt werden können. Das Framework müsste eine Semantik bereithalten, die beschreibt, wie die unterschiedliche Semantik der Schnittstellen für einen bestimmten Anwendungsbereich genutzt werden kann. Im Bereich der „Web Services“ Standards hat sich daher eine rege Entwicklung nachfolgender Standards ergeben, die etwa das Verhalten von Diensten und ihren Schnittstellen in der Zeit beschreiben – es soll die zugesicherte „Choreographie“ oder „Orchestrierung“ von Diensten austauschbar und damit das Zusammenspiel von Diensten maschinenlesbar ausgegeben werden können. Dies ist, solange dies keine *Maschinenverständlichkeit* impliziert, zwar machbar, in diesem Fall aber auch nicht zielführend, da unterschiedliche Choreographien ohne ein Verständnis der zugrunde liegenden Semantik nicht angepasst werden können.

Aus diesem Grunde muss die Möglichkeit eines einheitlichen Frameworks für wissenschaftliche Fachinformation in Frage gestellt werden, das ohne Eingriffe von Entwicklern das Zusammenspiel von DL-Diensten automatisiert. Gleichwohl wird der Aufbau von fachspezifischen Angeboten von diesem Standard enorm profitieren können. Die Nutzung allgemein verfügbarer Web Services wird vereinfacht, die durchgängige Nutzung von XML gestattet es, Angebote aus heterogenen Quellen in ihrer unterschiedlichen Struktur auszuwerten und nicht zuletzt steht ein standardisierter Mechanismus zur Verfügung, eigene Dienstangebote und seine Schnittstellen zu beschreiben. Eine solche Nutzung der W3C Web Services muss dann aber nicht eigens durch ein Framework beschrieben und organisiert werden - die Tools der Hersteller können genutzt werden, um entsprechende Schnittstellen zu erzeugen. Das eigentliche Framework kann dagegen auf eine technisch weniger anspruchsvolle Basis gestellt werden,

um grundlegende Funktionalitäten mittels einfacher HTTP-Aufrufe performant bereitzustellen.

2.3 Web Services mit einfachen Mitteln: REST

In seiner Dissertation mit dem Titel „Architectural Styles and the Design of Network-based Software Architectures“ (2000) beschreibt Roy T. Fielding ein Modell für die Nutzung und Bereitstellung von Angeboten und Diensten im Web. Dieses Modell verhält sich in vielerlei Hinsicht konträr zur Web Services Initiative des W3C. Fielding schlägt eine Vorgehensweise vor, die darauf verzichtet, eine „Middleware“ zu etablieren. Anstatt (wie diese Web Services es tun) eine zusätzliche Protokollschicht zur Unterstützung der Interaktion zwischen Diensten einzuführen, beschreibt er eine Nutzung des HTTP-Protokolls, die er mit dem Kürzel REST versieht, dem „Representational State Transfer“.

Das Kernproblem der Nutzung von HTTP für Dienste besteht darin, dass dieses Protokoll selbst „zustandslos“ ist, d.h. es baut keine Verbindung auf, deren Eigenschaften (wie z.B. Informationen über den Nutzer, bisher übertragene Daten usw.) abgefragt oder verändert werden können. Die Verbindung besteht einzig in der Übertragung einzelner Kommandos und es ist Sache der beiden an der Kommunikation beteiligten Seiten entsprechende Zustände zu verwalten. Diese Verwaltung sollte Fielding zufolge durch die Übertragung einer „Repräsentation des Zustands“ bei jedem Kommandoaufruf unterstützt werden. Bei dieser Repräsentation einer „resource“ handelt es sich um einen URI, für dessen Benutzbarkeit der Webdienst zu sorgen hat. Fielding schlägt also vor, allgemein anerkannte (und leichtgewichtige) Standards zu benutzen: URIs dienen der „Adressierung“ von Ressourcen, HTTP-Kommandos dienen der Kommunikation und HTML wird zur Verlinkung (URIs für weitere Schnittstellen, etwa zum Rendern oder Löschen einer Information) benutzt. Dieses einfache Modell wird durch naheliegende Maßnahmen wie Caching oder der Erzeugung von clientseitigem Code flankiert.

Ein Beispiel für diesen Stil ist das OAI Protokoll zum Harvesten von Metadaten. Mittels GET und POST-Requests werden Anfragen und Ergebnisdokumente in XML direkt über das HTTP-Protokoll verschickt. Da die Aufgabe des Protokolls sich darauf beschränkt, einen sequentiellen Datenstrom abzurufen (in der Regel bibliographische Einträge für Dokumente), ist die Implementierung der zugehörigen Zustandsverwaltung recht einfach. OAI-Server und -Clients können also in der Tat mit sehr geringem Aufwand entwickelt werden.

Aber auch bei komplexeren Aufgaben, wie z.B. der Nutzung einer webgestützten XML-Datenbank, werden Schnittstellen im REST-Stil angeboten und oftmals anderen Schnittstellen vorgezogen. Diese Schnittstellen nutzen einfache HTTP-Kommandos, die von gängigen Skriptsprachen erzeugt, verschickt, empfangen und ausgewertet werden können. Damit ist der Programmierauf-

wand niedrig, die Lokalisierung von Programmfehlern ist einfach und der Transport der Informationen erfolgt ohne den zusätzlichen Verlust, den das XML-basierte Kommunikationsprotokoll (XML/SOAP) der „Web Services“ mit sich bringt.

3 Generisches Framework für DL-Dienste

Auch wenn es sich bei REST um einen Entwicklungsstil handelt, der sich zu den „Web Services“ konträr verhält, so widerspricht er in keiner Weise den Prinzipien von SOA. Die Semantik von Diensten wird hier aber strikt vom Protokoll getrennt und als ein „implizit geteiltes Wissen“ der an der Kommunikation beteiligten Seiten angesehen. In unserem Szenario besteht aber zur Zeit keine Aussicht, mit den „Web Services“-Spezifikationen eine Beschreibung zu entwickeln, die eine automatische Abbildung der Konzeptualisierungen aufeinander garantiert. (Der neu geprägte Begriff von den „Semantic Services“ ist deshalb auch untauglich, eine neue Qualität von Web Services zu bezeichnen.) Die komponentenbasierte Entwicklung von DL-Diensten nach den SOA-Prinzipien sollte sich daher auf die Aspekte konzentrieren, die „verallgemeinerbar“ sind. Die Anpassbarkeit an fachspezifische Bedürfnisse, die „Generizität“ des Frameworks muss gegeben sein, ohne den möglichen Nutzern ein schwergewichtiges Modell aufzubürden. Einfache Aufgaben müssen naheliegend umgesetzt werden können, während auch die Umsetzung eines komplexeren Workflows durch die Eigenschaften des Frameworks unterstützt werden muss.

Diese Anforderungen werden durch zwei Merkmale berücksichtigt: einerseits durch ein Komponentenmodell, das gegenüber anderen Ansätzen, wie z.B. Avalon oder Spring, vereinfacht und auf die Bedürfnisse von DL-Diensten zugeschnitten wurde. Zum anderen wird eine Pipeline basierte „Verarbeitungslogik“ favorisiert. Im Unterschied zu universellen Workflow-Beschreibungssprachen gestattet die Konfiguration von Pipelines keine beliebigen Verarbeitungslogiken. Die Leistungsfähigkeit von pipelinebasierten Konzepten wurde aber schon von Systemen wie Cocoon2 (einem Framework zur Bereitstellung dynamischer Webinhalte) aufgezeigt. Der Verlust an Modellierungsmöglichkeiten wird durch einen Zugewinn an Verständlichkeit und Wartbarkeit aufgewogen.

3.1 Inversion of Control: Komponenten-Frameworks

Komponenten-Frameworks sehen die sog. „Inversion of Control“ vor. Dieser Begriff verdeutlicht die Entwicklungsprinzipien komponentenbasierter Software. Entwickler bestimmen selbst nicht über den Einsatz ihrer Komponenten. Diese müssen so entworfen sein, dass sie in unterschiedlichen Kontexten genutzt werden können. Die Kontrolle über die Nutzung der von ihnen zur Verfü-

gung gestellten Funktionalität obliegt dabei dem Framework. Komponenten-Frameworks wie Avalon definieren Schnittstellen, um den „Lebenszyklus“ von Komponenten bestimmen zu können: z.B. ihre Erzeugung, Initialisierung, Reinitialisierung oder ihre Freigabe. Diese Schnittstellen müssen von den Komponenten zur Verfügung gestellt werden - insbesondere muss der mit ihnen verbundene „Kontrakt“ eingehalten werden. Darüber hinaus werden weitere Rollen und Verhaltensweisen der Komponenten und die Schnittstellen, über die sie zu realisieren sind, in XML-Konfigurationsdateien festgehalten. Die (typischer Weise in Java) erstellten Komponenten können auf diese Weise in einen komplexen Workflow eingebunden werden. Avalon, Spring und weitere Frameworks gestatten auf diese Weise den Aufbau von Application Servern, ohne dass die Komponentenentwickler mit den unterschiedlichen Schichten der Gesamtarchitektur in Berührung kommen müssen. Wiederkehrende Aufgaben, wie die Übergabe von Konfigurationsparametern an die Komponente, die in Konfigurationsdateien festgehalten sind, werden vom Framework übernommen.

Mit der Komplexität des Frameworks steigt allerdings auch die Komplexität dessen, was als „Kontrakt“ den Entwicklern der Komponenten bekannt sein muss. Wenn eine einfache Aufgabenstellung vorliegt, dann werden die Zugewinne so gering ausfallen, dass sich eine Nutzung des Frameworks nicht mehr lohnt. Entwicklungsziel unseres Frameworks war es, die notwendigen Schnittstellen möglichst einfach zu halten. Dies wurde durch zwei Maßnahmen erreicht. Einmal wurde, wie oben gesagt, eine Engführung der Strukturierungsmöglichkeiten vorgenommen, indem ein einfaches Pipelinemodell für die Kopplung der Komponenten vorgesehen wurde. Zweitens wurde mit dem „Proxy“-Konzept dafür gesorgt, dass Funktionalitäten, die von der Komponente nicht auf eine besondere Weise zur Verfügung gestellt werden müssen, auf eine standardisierte Weise vom Proxy erfüllt werden können. Ein Proxy in diesem Sinne ist also ein Stellvertreter der Komponente, der durch das Framework angesprochen wird und aufgrund der Konfiguration der Komponente entscheidet, welche Funktionalität von ihm erfüllt wird und welche durch die Komponente selbst erbracht werden muss.

3.2 Pipelinebasierter Workflow

Die Verwendung von Pipelines ist ein schon seit Jahrzehnten bewährtes Mittel, um Komponenten miteinander zu verknüpfen. Im Betriebssystem UNIX wurden sie benutzt, um Prozesse miteinander auf eine einfache Weise hintereinander zu schalten. Diese „Pipes“ verknüpfen die Standard Ein- und Ausgabekanäle der Prozesse miteinander. Sie sind wenig effizient, da die Verarbeitung in den Prozessen voneinander abhängt - auf die Eingaben eines anderen Prozesses muss (synchronisiert) gewartet werden. Aktuelle Frameworks verwenden daher ein ereignisbasiertes Modell, bei dem - z.B. beim Parsen eines Dokuments - Da-

ten zu einer in sich sinnvollen Einheit gebündelt und zur Behandlung an die nächste Komponente in der Pipeline weitergereicht werden. Beispielsweise werden in Cocoon2 SAX-Events beim Parsen von XML-Dokumenten erzeugt. Die Behandlung von SAX-Events erlaubt die Bearbeitung umfangreicher Dokumente bei geringem Speicherbedarf und hohem Durchsatz - fast alle XML Application Frameworks (XAPs) basieren daher auf diesem Modell. Ein typisches Problem in diesem Zusammenhang besteht darin, dass es Arbeitsvorgänge gibt, die nicht durch ein sequentielles Abarbeiten von Ereignissen behandelt werden können. XSLT-Transformationen, bei denen XML-Dokumente für unterschiedliche Zwecke gerendert werden, setzen die Verfügbarkeit des gesamten Dokuments voraus. Aus diesem Grund sind in unserem Ansatz Pipelines konfigurierbar im Hinblick auf die Objekte, die zwischen den Komponenten ausgetauscht werden.

3.3 Implementierung des Frameworks

Die Implementierung erfolgt in der Programmiersprache C# unter der Microsoft.Net-Umgebung. C# wurde gewählt, da diese Programmiersprache einerseits Java sehr ähnlich ist - die mittlerweile sehr gut bekannten objektorientierten Konzepte dieser Programmiersprache können eingesetzt und dabei sowohl auf die Robustheit als auch auf die Performanz vertraut werden. Zum anderen sind die sprachspezifischen Erweiterungen (die mittlerweile auch in Java - wenn auch unvollständig - übernommen worden) für die Entwicklung von Frameworks sehr nützlich. C# gestattet erstens eine bessere Kontrolle der Speichersemantik als Java, da es neben dem Datentyp, der über Objekterzeugung entsteht, auch einen Typ vorsieht, der per Wert übergeben wird. Damit wird die Skalierbarkeit von Softwaresystemen gewährleistet, die „massenhaft“ Instanzen eines Typs erzeugen und wieder freigeben müssen. In C# kann dies ohne die teuren Aspekte der Objektkonstruktion und -destruktion geschehen. Zweitens werden die Generischen Datentypen in der (schon im Betastadium verfügbaren) Version 2.0 von C# umfassend unterstützt. Damit kann besonders elegant wieder verwendbarer Code geschrieben werden, der typischer und effizient für unterschiedliche Datentypen eingesetzt werden kann. Die generischen Datentypen in Java unterstützen nicht die integrierten elementaren Datentypen (wie z.B. für Integer). Besonders wichtig ist aber drittens die Verfügbarkeit von sog. „Attributen“. Hierbei handelt es sich um Metadaten, die allen Elementen des Codes hinzugefügt werden können. Die Schnittstellen eines Moduls können auf diese Weise markiert und vom Framework später eingebunden werden. Dieser Mechanismus wird von Microsoft in Teilen der Klassenbibliothek und bei weiteren Tools intensiv benutzt. In unserem Prototypen besteht der Einsatz darin, dass die Einbindung von benutzerdefinierten Komponentenschnittstellen über Attribute stattfindet. Diese Schnittstellen müssen im Code nur mit entsprechenden Attri-

buten markiert werden - das Framework wertet diese Informationen aus und generiert die in der Konfigurationsdatei definierten Pipelines aus den von Benutzern bereitgestellten Modulen. Im Unterschied zu anderen Komponentenframeworks ist es also nicht nötig, die Schnittstellen der von Benutzern bereitgestellten Komponenten mit vorgegebenen Bezeichnern zu versehen - jede Schnittstelle, die den benötigten Kontrakt erfüllt, kann markiert und somit dem Framework bekannt gemacht werden. Zusammen mit dem o.g. Proxy-Konzept - es müssen nur die Schnittstellen bereitgestellt werden, die in die Verarbeitung des Datenstroms einer Pipeline eingreifen - ergibt sich damit ein überschaubarer Aufwand für die Entwickler von Komponenten.

(Weitere Implementierungsdetails können auf unserer Website www.generische-portale.de eingesehen werden.)

3.4 Microsoft.Net und Offenheit

Ein wichtiger Punkt, den es bei der Wahl der Implementierung zu berücksichtigen galt, war die Frage nach der Offenheit einer Plattform. Die von Microsoft selbst propagierte Form der Offenheit - die beeindruckende und massive Unterstützung des Web Services-Standards - erwies sich als eher unzulänglich. Die Web Service Description Language (WSDL), mit der die Schnittstellen von Web Services beschrieben werden, sind XML-Dateien. Für die Beschreibung der Typen, die an den Schnittstellen übergeben werden, ist ein XML-Schema zu definieren. Im Fall von Microsofts Schemas werden hierbei allerdings nicht elementare Typen benutzt - vielmehr wird das objektorientierte Typsystem von .Net unterstützt. Diese Lösung ist elegant, solange sich auf beiden Seiten der Kommunikation .Net-Applikationen befinden. Anpassungen können im Prinzip entwickelt werden - die Komplexität des benutzten Typsystems stellt hierbei allerdings eine Hürde dar, die kleinere Anbieter nicht „überspringen“ werden. Die Web Services Tools, die Microsoft bereitstellt, sind nur dann im Sinne von Offenheit gut einsetzbar, wenn bewusst auf die Übergabe komplexerer Typen an den Schnittstellen vermieden wird. Verzichtet man also auf die „nahtlose Integration“ von Webdiensten in die .Net-Welt, dann können durchaus Webdienste implementiert werden, die plattformunabhängig genutzt werden können. Hinzu kommt, dass mit dem Mono-Projekt eine .Net kompatible Applikationsschicht unter Linux bereitgestellt wird.

4 Zusammenfassung

Eine Initiative zur Standardisierung von DL-Diensten hat mit weit größeren Problemen zu kämpfen, als man dies angesichts der aktuellen Entwicklungen bei den „Web Services“ vermuten wird. Die Vision des „maschinenlesbaren“

Webs wird sich auch in naher Zukunft nicht in den Bereichen umsetzen lassen, die auf eine fachspezifische Modellierung von Diensten und Angeboten angewiesen ist. Die fächerübergreifende Integration findet hier meist nur unter Inkaufnahme des Verlustes von inhaltsbezogenen Informationen statt, da Mechanismen für eine automatisierte Anpassung nur selten zur Verfügung stehen. Der Ansatz, mit einem Framework DL-Dienste auf eine gemeinsame Grundlage zu stellen, stößt auf dasselbe Problem. Die geschichtete Architektur der W3C-Spezifikationen rund um die Web Services gestattet es zwar, das automatisierte und plattformunabhängige Binden von Diensten auf der Grundlage vorhandener Metadaten durchzuführen - es beinhaltet damit aber keineswegs ein „semantisches Framework“. Dieses müsste durch Vereinbarungen, die aus den jeweiligen „communities of practice“ kommen, geschaffen werden.

Gegenwärtig besitzt das REST-Modell gegenüber der Initiative des W3C aufgrund seiner „Leichtgewichtigkeit“ klare Vorteile. Es verzichtet auf eine geschichtete Protokollstruktur und setzt auf überall vorhandene und leicht beherrschbare Technologien auf. Durch die Ergänzung mit Diensten, die implizite Modellbildungen explizit machen, könnte sich dieser Entwicklungsstil durchaus Vorteile gegenüber den „Web Services“ erarbeiten. Diese Ergänzung könnte in einem „Explain-Service“, wie ihn Z39.50 vorsieht, oder in Beschreibungen nach Maßgabe von „typekits“ bestehen - einer Auflistung von Tools oder Diensten, die zur Behandlung von XML-Fragmenten geeignet sind (siehe <http://typekit.org>).

Solche Ergänzungen können nun nicht durch ein isoliertes Projekt geschaffen und etabliert werden. Diese Aufgabe kann nur durch eine Clearingstelle gelöst werden, in der die verschiedenen Beteiligten durch konkrete Spezifikationsarbeit vertreten sind. Die IuK-Initiative wäre ein Umfeld, aus dem heraus eine solche Clearingstelle entstehen kann.

Angaben zum Autor

Dr. Stephan Körnig
TU Darmstadt
Institut für Philosophie
Schloß
64283 Darmstadt

Studium der Philosophie, Ethnologie und Religionswissenschaft in Marburg, Tübingen und Basel. Promotion in der Philosophie an der Universität Basel. IT-Spezialist in den Bereichen objektorientierte Analyse, Design und Entwicklung. Mitarbeit in unterschiedlichen Projekten im DL-Bereich - zuletzt in „Generische Portale“. Lehrbeauftragter an der TU Darmstadt (Philosophie)

Crosskonkordanzen als Mittel zur Heterogenitätsbehandlung in Informationssystemen

*Anne-Kathrin Walter, Philipp Mayr, Maximilian Stempfhuber,
Arne Ballay*

Informationszentrum Sozialwissenschaften, Bonn

Zusammenfassung

Die nutzerfreundliche Zusammenführung von Informationsangeboten erfordert nicht nur die Kopplung der Einzelangebote auf technischer Ebene – erst die semantische Integration der in der Regel unterschiedlich strukturierten und erschlossenen Information ermöglicht auf der Nutzerseite einen einheitlichen und effizienten Zugriff. Als Ergänzung zur Standardisierung von Metadaten und Inhaltserschließungsverfahren, die sich vor allem beim Neuaufbau von Informationssammlungen anbietet, spielen Crosskonkordanzen eine zentrale Rolle bei der Vernetzung von kontrollierten Erschließungsvokabularen unterschiedlicher Datenbanken oder Bibliothekskataloge. Erst deren Transferleistung – die Abbildung von Termen eines Thesaurus auf die Terme eines anderen Thesaurus durch semantische Relationen – ermöglicht es Nutzern, mit Begriffen eines Vokabulars gleichzeitig in mehreren Datenbanken zu recherchieren, ohne die Suchanfrage mehrfach und datenbankspezifisch zu formulieren. Am Beispiel des Informationsverbunds infoconnex¹ wird das Konzept der Crosskonkordanzen erläutert und anhand von konkreten Beispielen gezeigt, welche Möglichkeiten sie bei der integrierten Recherche in Fachdatenbanken aus Pädagogik, Sozialwissenschaften und Psychologie bieten.

1 Einleitung

Ein Blick auf die aktuelle Informationslandschaft zeigt eine Diversifizierung in viele dezentrale Informationsangebote, die online und zumeist unverbunden zur Verfügung stehen. Mit zunehmender Vernetzung und interdisziplinärer Ausrichtung von Forschungsaktivitäten sehen sich Nutzer immer häufiger in der Situation, auf mehrere Informationsangebote zugreifen zu müssen, um ihr Infor-

¹ Das Projekt infoconnex wurde gefördert vom Bundesministerium für Bildung und Forschung (BMBF) und von der Deutschen Forschungsgemeinschaft (DFG).

mationsbedürfnis zu befriedigen. Die alleinige Nutzung z. B. des Online-Katalogs (OPAC) der lokalen Bibliothek führt genauso wenig zu einem befriedigenden Ergebnis wie die Beschränkung auf eine einzelne Fachdatenbank. Mit jedem Wechsel zwischen Informationsangeboten sieht sich der Nutzer aber vor das Problem gestellt, die eben bereits für den OPAC formulierte Suche für eine oder mehrere Fachdatenbanken neu zu formulieren.

Nicht nur die unterschiedlichen Benutzungsoberflächen bedingen eine erneute Einarbeitung in jedes der Recherchesysteme, auch die Schlagwortsuche, die Suche über das zur inhaltlichen Beschreibung von z. B. Publikationen verwendete, kontrollierte Vokabular, erfordert die wiederholte Auswahl von Suchbegriffen, die das Informationsbedürfnis des Nutzers möglichst gut beschreiben. Aufgrund der unterschiedlichen Größe, Spezifität und Fachbezogenheit der Vokabulare (z. B. Schlagwortnormdatei der Deutschen Bibliothek als großes, allgemeines Vokabular im Gegensatz zu einem relativ kleinen, aber fachspezifischen Thesaurus) können die für eine bestimmte Datenbank gewählten Suchbegriffe nur in seltenen Fällen direkt zur Suche in einer anderen Datenbank verwendet werden. Neben für den Nutzer relativ leicht zu behebenden Unterschieden in der Ansetzung der Thesaurusbegriffe (z. B. Ansetzung „Kind“ im Singular vs. Ansetzung „Kinder“ im Plural) erfordern vor allem die Verwendung unterschiedlicher Begriffe für das gleiche semantische Konzept (z. B. „Kommunikationsgeschichte“ in der Pädagogik vs. „Mediengeschichte“ in den Sozialwissenschaften) oder die Kombination mehrerer Begriffe (z. B. „Medien + Geschichte“ in der Psychologie) eine intensive Einarbeitung in den Sprachgebrauch der Fachgebiete oder Datenbanken. Nur wenige Nutzer dürften bereit und in der Lage sein, diesen aufwändigen Transfer bei der täglichen Informationsbeschaffung zu leisten – zumindest legen neuere Untersuchungen dies nahe [s. Boekhorst et al. 2003, Poll 2004]. Was aber ist zu tun, um den Nutzern den bequemen Zugriff auf die benötigten Informationen zu ermöglichen?

Eine Strategie, die bereits erste positive Resultate zeigt, ist die Zusammenführung bislang unverbundener Angebote. So bündeln die von der Deutschen Forschungsgemeinschaft (DFG) geförderten Virtuellen Fachbibliotheken und die vom Bundesministerium für Bildung und Forschung finanzierten Informationsverbände bereits relevante Informationen der betreffenden Fächer. Darüber hinaus führt das fachübergreifende Wissenschaftsportal *vascoda*² beide Förderlinien zusammen und schafft einen zentralen Zugang zu wissenschaftlicher Information über Fächergrenzen hinweg. Für den Nutzer ist dadurch mehr – bislang verstreute – Information an einer Stelle zugänglich und der Aufwand beim Wechsel zwischen Informationsangeboten sinkt deutlich. Auf der qualitativen Ebene wurde das Potential der Integration aber bislang noch wenig genutzt. Zwar bieten viele Portale eine integrierte Suche (Metasuche) an, häufig werden

2 <http://www.vascoda.de>

die Suchbegriffe des Nutzers aber ohne weitere Behandlung an die angeschlossenen Datenbanken und OPCAs weitergeleitet – mit allen negativen Konsequenzen für die Größe (Recall) und Qualität (Precision) des Ergebnisses.

Die qualitative Verbesserung der integrierten Suche in Datenbanken unterschiedlicher Fächer hat sich in diesem Umfeld erstmalig der Informationsverbund Pädagogik – Sozialwissenschaften – Psychologie (infoconnex³) zum Ziel gesetzt. Dabei wurde versucht, der Heterogenität zu begegnen, die sich aus der Zusammenführung verschiedener Informationsangebote auf mehreren Ebenen ergibt (s. Stempfhuber 2002 und Strötgen 2004):

- Technische Heterogenität aufgrund der Zusammenführung unterschiedlicher Datenbanksysteme.
- Strukturelle Heterogenität aufgrund unterschiedlicher Modellierung der in den Datenbanken beschriebenen Entitäten (z. B. Literatur, Internetquellen, Institutionen, Lehrmaterialien) durch jeweils eigene Attributmengen (Metadatenschemata).
- Semantische Heterogenität aufgrund unterschiedlicher Inhaltserschließungssysteme, durch die die zulässigen Attributausprägungen definiert werden (z. B. Thesauri, Klassifikationen, Dokumentarten, Personenrollen, Länderlisten und andere Normdaten).

Der Fokus von infoconnex lag auf der Behandlung von semantischer Heterogenität, die durch die Verwendung jeweils eigener Attributausprägungen für semantisch äquivalente Attribute entsteht. Im Kontext von Literaturinformation betrifft dies häufig die Attribute „Schlagwort“ und „Klassifikation“, die zur Beschreibung der wesentlichen Inhalte von Publikationen verwendet werden und zwischen OPACs und Fachdatenbanken sehr starke Unterschiede aufweisen. So ist jede der Fachdatenbanken in infoconnex mit einem eigenen Thesaurus erschlossen, der sich hinsichtlich Struktur, Scope und Spezialisierungsgrad wesentlich von den anderen unterscheidet. Um dem Benutzer die Möglichkeit zu geben, mit einem beliebigen – in der Regel dem ihm am besten bekannten – Thesaurus eine Anfrage zu formulieren, die dann auf die drei unterschiedlich erschlossenen Fachdatenbanken angewendet wird, wurden die Thesauri bilateral durch so genannte Crosskonkordanzen miteinander verknüpft. Die Menge der zwischen den Begriffen unterschiedlicher Thesauri definierten Beziehungen (Relationen) erlaubt es im Rechercheprozess dann, die Anfrage des Nutzers vom Vokabular des Ausgangsthesaurus automatisch auf andere Vokabulare abzubilden und mit den für die Zieldatenbank geeigneten Begriffen zu suchen.

Im Folgenden werden zunächst mögliche Verfahren zur semantischen Heterogenitätsbehandlung vorgestellt und dann vertieft auf Crosskonkordanzen eingegangen. Am Beispiel der fachübergreifenden Recherche im Informationsver-

3 <http://www.infoconnex.de>

bund infoconnex wird dann die Anwendung der Crosskonkordanzen in der Praxis erläutert und ihre Wirksamkeit zur Vokabularerweiterung belegt.

2 Semantische Heterogenitätsbehandlung

Ein zentrales Problem des Information Retrieval ist die Vagheit der Abbildung zwischen der Anfrage eines Benutzers und den Indexierungsbegriffen, die den Inhalt – häufig nur die wahrscheinlich für Nutzer relevanten Aspekte – der im Informationssystem gespeicherten Dokumente beschreiben (siehe Abbildung 1: V1). Konkret bedeutet dies, dass ein Benutzer in seiner Anfrage für den gleichen semantischen Aspekt andere Begriffe verwendet, als – im Falle intellektueller Erschließung – die für die Inhaltserschließung verantwortliche Person vergeben hat. Auch bei der automatischen Indexierung tritt dieses Problem auf, allerdings hier zwischen dem Wortschatz des Informationssuchenden und den Autoren der Dokumente. Traditionell wird die Vagheit zwischen Anfrage- und Dokumentenebene durch Verfahren zur Termerweiterung behandelt, was „manuell“ durch den Nutzer geschehen kann (z. B. durch Verwendung eines Thesaurus mit Synonym- oder Ober-/Unterbegriffsbeziehungen) oder in ähnlicher Weise auch durch das Informationssystem (teil-)automatisch geschehen kann (z. B. durch Vorschläge für zusätzliche und alternative Suchbegriffe).

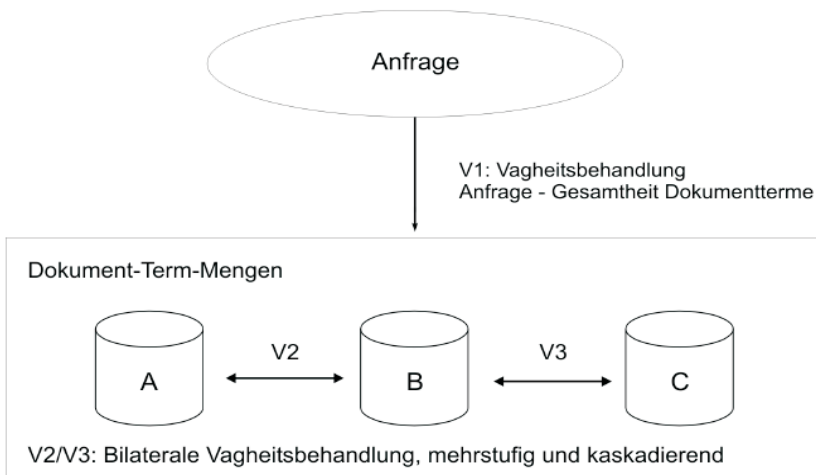


Abb. 1: Vagheitsbehandlung (aus Krause 2003)

Handelt es sich bei den zu durchsuchenden Datenbeständen um homogen erschlossene Datenbanken, sind diese Verfahren ausreichend. Anders bei heterogenen Dokumentenbeständen: Durch die unterschiedlichen Thesauri entsteht Vagheit bereits auf der inhaltlichen Beschreibungsebene der Dokumente. Stellt man die Datenbanken einander gegenüber und vergleicht, mit welchen Begriffen gleiche Dokumente oder solche mit ähnlichem Inhalt erschlossen sind, so stellt man fest, dass für gleiche semantische Konzepte je nach Erschließungsvokabular unterschiedliche Indexierungsbegriffe verwendet wurden, die sich in ihrer semantischen Präzision unterscheiden können. Die Konsequenz ist, dass die Anfrage des Nutzers nicht nur auf ein Indexierungsvokabular, sondern parallel auf mehrere abgebildet werden muss.

In der Praxis würde dies bedeuten, dass zwischen dem Vokabular, das der Nutzer zur Formulierung der Anfrage benutzt (z. B. den ihm gut bekannten Thesaurus einer Fachdatenbank) und jedem der Vokabulare der anderen Datenbanken eine Abbildungsvorschrift (z. B. eine Konkordanz) vorliegen muss. Dies ist häufig aber nicht der Fall. Als Ausweichstrategie kann die Abbildung der Nutzeranfrage bilateral und kaskadierend zwischen den beteiligten Vokabularen erfolgen, so dass die Anfrage zunächst auf ein (zentrales) Vokabular abgebildet und von dort aus schrittweise auf die anderen propagiert wird. Die Abbildung 1 verdeutlicht dies. Die Anfrage des Nutzers wird hier zunächst auf das Vokabular der Datenbank B abgebildet und erst von dort aus auf die Vokabulare der Datenbank A (Transformation V2) und Datenbank B (Transformation V3). Dabei ist konzeptionell auch eine Kombination von unterschiedlichen Verfahren bei den einzelnen Schritten zulässig.

Zur Behandlung der Vagheit bzw. der Heterogenität der Dokumentenbestände gibt es verschiedene Verfahren. Die zwei wesentlichen werden im Folgenden kurz vorgestellt.

- Intellektuell erstellte Crosskonkordanzen zwischen Klassifikationen und Thesauri

Crosskonkordanzen sind gerichtete, relevanzbewertete Relationen zwischen Termen zweier Thesauri oder Klassifikationen. Sie ermöglichen eine „Übersetzung“ von Anfragetermen eines Erschließungssystems in ein anderes. Die Erstellung der Relationen erfolgt intellektuell, d.h. ein Terminologie-Experte vergleicht verschiedene Begriffssysteme und setzt die Terme zueinander in Beziehung. Die Verknüpfung erfolgt dabei nach einem Verfahren, das im Projekt CARMEN AP12⁴ definiert wurde und folgende Arten von Relationen zwischen Termen unterschiedlicher Vokabulare festlegt:

- Eine Äquivalenzrelation wird für identische, synonyme oder quasi-synonyme Terme verwendet.

4 siehe <http://www.bibliothek.uni-regensburg.de/projects/carmen12/index.html> de

- Eine Oberbegriffsrelation wird von einem semantisch engeren zu einem semantisch weiteren Term gesetzt.
- Eine Unterbegriffsrelation wird von einem semantisch weiteren zu einem semantisch engeren Begriff gesetzt.
- Die Ähnlichkeitsrelation wird für verwandte Begriffe gesetzt.

Das Verfahren ist mit einem nicht zu vernachlässigenden intellektuellen Aufwand verbunden und betrachtet die Terme der Vokabulare unabhängig von ihrem konkreten Gebrauch bei der Indexierung von Dokumenten.

- **Quantitativ-statistische Ansätze**

Bei diesen Verfahren werden die semantischen Relationen mit Hilfe von statistischen Methoden automatisch erzeugt. Grundlage hierfür sind Parallelkorpora, d.h. Bestände, deren Dokumente gleichzeitig nach zwei Sacherschließungssystemen indexiert sind. In einer „Kookkurrenzanalyse“ werden im Parallelkorpus die gemeinsam vorkommenden Terme aus jedem Vokabular auf Dokumentenebene ausgewertet und daraus Wahrscheinlichkeiten für ein gemeinsames Vorkommen abgeleitet. Es werden also solche Gruppen von Termen beider Thesauri ermittelt, die besonders häufig bei der Indexierung von Dokumenten zusammen auftreten.

Die Schwierigkeit bei diesem Verfahren ist, einen geeigneten Parallelkorpus bereitzustellen. In Strötgen 2004, Binder et al. 2002 und Hellweg et al. 2001 werden Möglichkeiten zur Simulation von Parallelkorpora beschrieben. Anders als bei den Crosskonkordanzen kann bei statistischen Verfahren nicht nach unterschiedlichen Relationstypen differenziert werden, die erzeugten Relationen spiegeln jedoch den Gebrauch beider Vokabulare bei der Indexierung konkreter Dokumente wider.

3 Einsatz von Crosskonkordanzen am Beispiel von infoconnex

3.1 Crosskonkordanzen im Informationsverbund infoconnex

infoconnex ist ein interdisziplinärer Informationsverbund für die Fächer Pädagogik, Sozialwissenschaften und Psychologie, der die integrierte Recherche in den zentralen Fachdatenbanken der drei Disziplinen ermöglicht und auch in das deutsche Wissenschaftsportal *vascoda*⁵ integriert ist. Im Rahmen von infoconnex kamen erstmalig die in Kapitel 2 erläuterten Crosskonkordanzen im Echtbetrieb zum Einsatz.

5 siehe <http://www.vascoda.de/>

Das Projekt infoconnex wurde zwischen Juni 2001 und Juli 2005 durch das BMBF und die DFG gefördert und als Gemeinschaftsprojekt vom Informationszentrum Sozialwissenschaften (IZ), dem Deutschen Institut für Internationale Pädagogische Forschung (DIPF) und dem Zentrum für Psychologische Information und Dokumentation (ZPID) sowie den entsprechenden Sondersammelgebietsbibliotheken (Universitätsbibliothek Erlangen-Nürnberg, SSG Bildungsforschung; Stadt- und Universitätsbibliothek Köln, SSG Sozialwissenschaften; Saarländische Universitäts- und Landesbibliothek, SSG Psychologie) ins Leben gerufen. Ziel war es, ein qualitativ hochwertiges, durch Verfahren der Heterogenitätsbehandlung gestütztes und interdisziplinär angelegtes Fachinformationsportal zu entwickeln, dessen Literaturnachweisdatenbanken auch mit Volltexten kommerzieller Verlage verknüpft werden sollten:

- Fachinformationssystem (FIS) Bildung (Pädagogik),
- PSYINDEX (Psychologie),
- SOLIS (Sozialwissenschaften).

Der Gesamtbestand umfasst ca. eine Million Nachweise. Neben der fachspezifischen Suche in den einzelnen Fachdatenbanken ermöglicht infoconnex zusätzlich eine fachübergreifende Suche (siehe Abb. 2).

The screenshot shows the 'infoconnex' search portal. At the top, there are three tabs: 'Pädagogik' (green), 'Sozialwissenschaften' (red), and 'Psychologie' (blue). Below the tabs is a search bar containing the query 'Schlagwort=ökonomische theorie' -> (ökonomie und theorien)[28]. The search results are displayed in a table with columns for 'Jahr', 'Titel', 'Datenbank', and 'Verfügbarkeit'. The results are filtered by 'Pädagogik (121)', 'SOLIS (1657)', and 'Psychologie (28)'. The table shows 6 results, all from the PSYINDEX database.

	Jahr	Titel	Datenbank	Verfügbarkeit
<input type="checkbox"/>	1983	Marketing ohne verhaltenswissenschaftliche Fundierung? [...]	PSYINDEX	6 5 FX
<input type="checkbox"/>	1979	Theorie und Erfahrung : Beitrage zur Grundlagenproblematik der Sozialwissenschaften [...]	PSYINDEX	6 5 FX
<input type="checkbox"/>	1983	Wirtschaftskriminalitaet als Prozess kollektiver Selbstschaedigung? : Ein Beispiel und einige theoretische Ueberlegungen [...]	PSYINDEX	6 5 FX
<input type="checkbox"/>	1982	Ergebnisse der experimentellen Wirtschaftsforschung zum Optimierungsverhalten [...]	PSYINDEX	6 5 FX
<input type="checkbox"/>	1978	Nicht-intendierte soziale Folgen individueller Handlungen [...]	PSYINDEX	6 5 FX

Abb. 2: Fachübergreifende Schlagwortsuche mit Einsatz von Crosskonkordanzen in infoconnex

Da jede der drei Datenbanken mit einem eigenen Thesaurus erschlossen ist, hätte eine Schlagwortsuche ohne den Einsatz von Verfahren zur Heterogenitätsbehandlung zur Folge, dass nur mit Suchbegriffen, die in allen drei Thesauri vor-

handen sind, auch in allen drei Datenbanken Treffer gefunden würden. Alle semantischen Konzepte, die in den drei Thesauri zwar vorkommen, dort aber sprachlich unterschiedlich ausgedrückt sind, könnten bei der Suche ohne zusätzliche intelligente Verfahren nicht entsprechend nachgewiesen werden.

Um dem zu begegnen, werden in infoconnex Crosskonkordanzen eingesetzt. Sie ermöglichen die Transformation der vom Nutzer eingegebenen Schlagwörter, die in der Regel aus dem ihm am besten bekannten Thesaurus stammen werden, in entsprechende Schlagwörter der anderen Thesauri. Damit wird es möglich, an jede der drei Datenbanken spezifische Anfragen zu stellen, ohne dass der Nutzer diesen intellektuellen Transfer selbst leisten muss. Die Abbildung 3 veranschaulicht das prinzipielle Vorgehen. Die an alle drei Datenbanken gerichtete Anfrage des Nutzers (Suchoption „infoconnex“) wird in einem ersten Schritt auf die im System vorhandenen Thesauri abgebildet. Es wird also überprüft, in welchen Thesauri die eingegebenen Schlagwörter enthalten sind. In einem zweiten Schritt wird ermittelt, zu welchen anderen Thesauri Crosskonkordanzrelationen existieren. Diese werden dann dazu benutzt, um für die Suche in der entsprechenden Datenbank die Originalbegriffe aus der Anfrage durch die über die Crosskonkordanzen transformierten Begriffe zu ersetzen und eine datenbankspezifische Anfrage zu erzeugen. Die Crosskonkordanzen liegen in infoconnex grundsätzlich bidirektional vor, d. h. die Thesauri sind immer paarweise und in beide Richtungen miteinander verbunden, so dass Transfers in jede Richtung stattfinden können. Bei einer größeren Zahl von Thesauri ist diese Art der Verknüpfung aufgrund der Vielzahl an Kombinationen möglicherweise nicht mehr zu leisten. Daher wird in Kapitel 4.2 ein alternatives Verfahren mittels indirekter Transfers dargestellt.

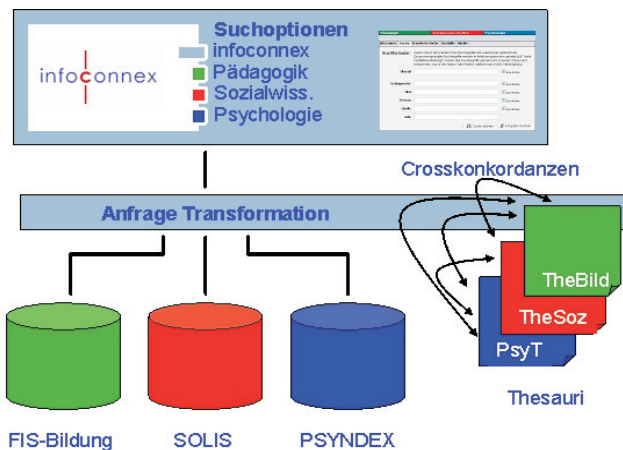


Abb. 3: Schematische Darstellung der Anfragetransformation in infoconnex

The image displays three sequential screenshots of a search interface, illustrating the transformation of a user query across three different databases. Each screenshot shows a search form with a header bar containing three colored tabs: Pädagogik (green), Sozialwissenschaften (red), and Psychologie (blue). The search form includes a search bar, a 'Suchformular > Trefferliste' header, and an 'Anfrage' section. Below the search bar, a table shows the results for each database: Pädagogik (121), SOLIS (1657), and Psychologie (28). The 'Anfrage ändern' button is visible in each screenshot.

Screenshot 1 (Top): The search bar contains the query 'Schlagwort=WIRTSCHAFTSTHEORIE[1 21]'. The results table shows: Pädagogik (121), SOLIS (1657), Psychologie (28).

Screenshot 2 (Middle): The search bar contains the query 'Schlagwort = ökonomische Theorie (1657)'. The results table shows: Pädagogik (121), SOLIS (1657), Psychologie (28).

Screenshot 3 (Bottom): The search bar contains the query 'Schlagwort='ökonomische theorie'-> (ökonomie und theorien)[28]'. The results table shows: Pädagogik (121), SOLIS (1657), Psychologie (28).

Abb. 4: Beispiel einer Anfragetransformation zum Schlagwort „ökonomische Theorie“

Die Abbildung 4 zeigt ein Beispiel dieser Anfragetransformation durch Crosskonkordanzen. Übergreifend über alle drei Datenbanken wird eine Schlagwortsuche nach „ökonomische Theorie“ durchgeführt. Im Thesaurus Sozialwissenschaften ist dieser Begriff ein gültiger Deskriptor. Die Anfrage „Schlagwort = ökonomische Theorie“ kann direkt an die Datenbank SOLIS weitergeleitet werden, wobei 1657 Dokumente gefunden werden. Im Thesaurus Psychologie hingegen ist dieser Begriff kein gültiger Deskriptor. Allerdings verweisen Crosskonkordanzrelationen auf äquivalente Terme in diesem Thesaurus, die stattdessen verwendet werden können. Die Anfrage des Nutzers wird für die Datenbank PSYINDEX zur Schlagwortkombination „Ökonomie + Theorien“ transformiert und es werden 28 Treffer gefunden. Auch im Thesaurus Pädagogik findet sich der Deskriptor „ökonomische Theorie“ nicht, eine Transformation zu „Schlagwort = Wirtschaftstheorie“ ergibt jedoch 121 Treffer in FIS Bildung.

Derzeit werden in infoconnex lediglich Äquivalenzrelationen genutzt, da sie in jedem Fall automatisch – also ohne Nutzerinteraktion – verwendet werden können. Die Verwendung weiterer Crosskonkordanz-Relationen (Oberbegriffe, Unterbegriffe, Ähnlichkeiten) soll dem Nutzer später über entsprechende Elemente auf der Benutzungsoberfläche schrittweise ermöglicht werden.

3.2 Weitere Beispiele für die Effekte von Crosskonkordanzen

Im Folgenden werden nun erste empirische Ergebnisse präsentiert, die zeigen, welchen Einfluss die Ausweitung des Suchvokabulars durch Crosskonkordanzen bei der datenbankübergreifenden Suche hat.

Beim Einsatz von Crosskonkordanzen stellt sich zunächst die Frage nach der Wirksamkeit dieser Rechercheunterstützung auf das Ergebnis. Wie sinnvoll der Einsatz von Crosskonkordanzen für die Literaturrecherche ist und wie Crosskonkordanzen aufgebaut werden müssen, um einen maximalen Effekt zu erzielen, kann letztendlich nur durch eine qualitative Bewertung der Ergebnisse durch Nutzer beantwortet werden, auf deren Basis dann detaillierte Messungen des Recall und der Precision erfolgen können. Diese Untersuchungen sind im Projekt „Kompetenzzentrum Modellbildung und Heterogenitätsbehandlung“⁶ geplant, das vom BMBF im Rahmen der „Kompetenznetzwerke Neue Dienste, Standardisierung, Metadaten“ gefördert wird.

Als vorbereitender Schritt für diese Untersuchungen wurde die Erhöhung der Trefferzahlen untersucht, die durch Äquivalenzrelationen (siehe Kap. 2) zwischen den drei kontrollierten Vokabularen von infoconnex erzielt werden können. Die in den Tabellen 1 bis 3 dargestellten Ergebnisse sind repräsentativ für Termtransformationen, für die in mindestens zwei der infoconnex-Datenbanken Treffer generiert werden können. Dabei werden nur solche Termtransformationen betrachtet, bei denen die Zielterme nicht zeichengleich mit dem Ausgangsterm sind, also eine „intelligente“ Transferleistung erbracht wurde (kursiv gekennzeichnet). In Tabelle 1 (Zeile 3) finden sich zwei dieser Termtransformationen. Der Deskriptor „berufliche Weiterbildung“ aus dem sozialwissenschaftlichen Thesaurus wird in Berufsbildung (Thesaurus Bildung) und „Weiterbildung im Beruf“ (Thesaurus Psychologie) transformiert.

Die Relevanz der zusätzlichen Treffer in den folgenden Darstellungen wurde nicht bewertet. Auf die Überlappung der drei kontrollierten Vokabulare und die Zusammensetzung der einzelnen in infoconnex eingesetzten Crosskonkordanzen wurde bereits an anderer Stelle hingewiesen (Mayr, Stempfhuber, Walter 2005).

Die folgenden Beispiele zeigen – ausgehend von Deskriptoren des Thesaurus Sozialwissenschaften TheSoz – die Relationen zu den Deskriptoren bzw. Deskriptorkombinationen in den beiden Vokabularen Thesaurus Bildung TheBild (CK) und Psyndex Terms PsyT (CK) sowie die Anzahl der Treffer dieser Konkordanzen in den drei Datenbanken SOLIS, FIS Bildung und Psyndex. Die Deskriptoren und Treffer, die ohne intellektuelle Konkordanzherstellung für die integrierte Recherche der drei Datenbanken nicht zur Verfügung gestanden hätten, sind in den folgenden Tabellen kursiv gekennzeichnet. Zeichengleiche Deskrip-

6 <http://www.gesis.org/Forschung/Informationstechnologie/komohe.htm>

toren spielen eine Sonderrolle (siehe unten) und werden nicht gesondert hervorgehoben.

Tabelle 1: Auswahl häufig zur Inhalterschließung verwendeter TheSoz-Deskriptoren mit Termtransformationen in die Vokabulare TheBild und PsyT. Zusätzlich ist die Trefferanzahl angegeben, die mit dem Deskriptor bzw. den Transformationen in der jeweiligen Datenbank erreicht werden kann (Stand Anfang 2005).

TheSoz	SOLIS	TheBild (CK)	FIS Bildung	PsyT (CK)	Psyindex
Arbeitsloser	1.995	Arbeitsloser	1.432	Arbeitslosigkeit + Arbeiter	12
berufliche Weiterbildung	4.869	Berufsbildung	3.494	Weiterbildung im Beruf	529
Entwicklungsland	18.278	Entwicklungsland	2.756	Entwicklungsländer	111
Erfahrung	1.875	Erfahrung	1.727	Erfahrungen (Erlebnisse)	463
Gewerkschaftspolitik	2.193	Gewerkschaftspolitik	50	Gewerkschaften + Politik	2
Informationstechnologie	1.778	Informationstechnik	3.358	Information + Technologie	108
Kulturpolitik	1.599	Kulturpolitik	1.441	Kultur (Anthropologie) + Politik	12
Modellentwicklung	832	Modellentwicklung	231	Modelle + Entwicklung	154
ökonomische Theorie	1.613	Wirtschaftstheorie	91	Ökonomie + Theorien	28
politische Entwicklung	1.867	politische Entwicklung	18	Politik + Entwicklung	122
politische Kommunikation	1.383	Politische Kommunikation	15	Politik + Kommunikation	42
Prävention	1.624	Präventivmaßnahme	2.526	Prävention	3.355
soziale Umwelt	1.044	Soziales Milieu	189	Soziale Umwelt	548
Technologiepolitik	1224	Technologiepolitik	113	Technologie + Politik	11

Tabelle 1 zeigt Deskriptoren des TheSoz (linke Spalte), die vergleichsweise häufig zur Indexierung von Dokumenten in SOLIS verwendet werden (siehe 2. Spalte, Treffer in SOLIS). Genuin sozial- bzw. politikwissenschaftliche Begriffe wie politische Entwicklung, politische Kommunikation, ökonomische Theorie oder Gewerkschaftspolitik liefern insbesondere in SOLIS deutlich mehr als 1.000 Dokumente. Dass auch in den beiden Datenbanken der anderen Fächer (FIS Bildung und Psyindex) durch die Crosskonkordanzen zusätzliche Doku-

mente gefunden werden, verdeutlicht ihren Mehrwert für interdisziplinäre Fragestellung. Auffällig ist, dass aufgrund der inhaltlichen Nähe des Thesaurus Sozialwissenschaften und des Thesaurus Bildung relativ viele Deskriptoren zeichengleich vorliegen⁷. Dies trifft z. B. auf folgende Deskriptoren zu: Arbeitsloser, Entwicklungsland, Erfahrung, Kulturpolitik usw. Die Überlappung zweier Thesauri durch zeichengleiche Terme, die sehr stark von der Größe und Spezialisierung der Vokabulare abhängt, kann beispielsweise als ein Indikator gewertet werden, ob es sinnvoll ist, eine Crosskonkordanz für diese zwei Vokabulare zu erstellen bzw. wie hoch der intellektuell zu leistende Aufwand sein wird.

Weiterhin fällt auf, dass die Überführung der TheSoz-Deskriptoren in das Vokabular des PsyT (CK) terminologisch schwieriger ist und folglich viel häufiger Deskriptorkombinationen verwendet werden müssen, um die Semantik der TheSoz-Deskriptoren auszudrücken. Beispiele hierfür sind Arbeitslosigkeit + Arbeiter, Gewerkschaft + Politik oder Modell + Entwicklung. Die Analyse der Crosskonkordanzen TheSoz→PsyT und TheSoz→TheBild in infoconnex bestätigt diesen Eindruck sehr deutlich.

Tabelle 2 zeigt als Kontrast ausgewählte Deskriptoren des TheSoz, die vergleichsweise selten zur Indexierung von Dokumenten in SOLIS verwendet werden. Beispielsweise liefert die Schlagwortsuche nach den Deskriptoren Biologieunterricht, Ethikunterricht oder Pharmakologie in SOLIS kaum Treffer, während in FIS Bildung sehr viele und in Psyndex zumindest einige zusätzliche Treffer gefunden werden. Dies zeigt, dass die in infoconnex erstellten Crosskonkordanzen über das gesamte Spektrum des Ausgangsthesaurus hinweg zu einer Erweiterung des Dokumentenraums auf zwei verwandte Fächer (Pädagogik und Psychologie) führen.

Die beiden Deskriptoren Biologieunterricht und Ethikunterricht müssen im Fall des PsyT durch Kombination von zwei Deskriptoren (Ethik + Lehren bzw. Naturwissenschaftlicher Unterricht + Biologie) ausgedrückt werden. Das Beispiel des TheSoz-Deskriptors schulpsychologischer Dienst zeigt, dass die Kombination zweier Deskriptoren nicht immer positive Effekte für eine Konkordanz bringt. Die Kombination der beiden PsyT-Deskriptoren Schulpsychologie + Soziale Dienste bringt im Fall Psyndex keine Treffer, was beim Scope dieser Datenbank überrascht. Recherchiert man lediglich mit dem Psyndex-Deskriptor Schulpsychologie, liefert Psyndex 280 Treffer. Folglich werden durch die Kombination mit dem Deskriptor Soziale Dienste relevante Dokumente ausgeschlos-

⁷ Dies hat mit der Entstehungsgeschichte des Thesaurus Bildung zu tun, der sich aus den Vokabularen mehrerer Einzelthesauri zusammensetzt und dadurch zwangsläufig auch andere Fachgebiete überlappt. Der Thesaurus befindet sich zur Zeit in Überarbeitung und wird in seiner neuen Version diesen Grad der Überlappung mit dem Thesaurus Sozialwissenschaften wohl nicht mehr aufweisen.

sen⁸. Dieser negative Effekt intellektuell erstellter Crosskonkordanzen, insbesondere bei Termkombinationen, kann letztlich nur ausgeschlossen werden, wenn Relationen neben der semantischen Prüfung zusätzlich über den Recall in der Zieldatenbank geprüft werden. Diese Prüfung ließe sich relativ leicht automatisieren und könnte in die laufende Crosskonkordanzerstellung eingliedert werden.

Tabelle 2: Auswahl selten verwendeter TheSoz-Deskriptoren mit Termtransformationen in die Vokabulare TheBild und PsyT, sowie die Trefferanzahl der Deskriptoren bzw. Deskriptor-Kombinationen in den Datenbanken SOLIS, FIS Bildung & Psyndex (Trefferzahlen sind vom Stand Anfang 2005).

TheSoz	SOLIS	TheBild (CK)	FIS PsyT (CK) Bildung	Psyndex
Berufsgewerkschaft	2	Gewerkschaft	2175 Berufe + Gewerkschaften	2
Biologieunterricht	7	Biologieunterricht	4815 Naturwissenschaftlicher Unterricht + Biologie	10
Ethikunterricht	5	Ethikunterricht	818 Ethik + Lehren	9
Facharbeiterberuf	71	Facharbeiter	853 Facharbeiter + Berufe	2
frauenspezifische Verfolgung	16	Frau + Verfolgung	3 Verfolgung + Frauen	2
Islamismus	64	Islamischer Fundamentalismus	4 Islam + Fundamentalismus	3
moralisches Urteil	23	Moralisches Urteil	12 Moral + Beurteilung	65
nichteheliches Kind	104	Uneheliches Kind	4 Uneheliche Kinder	20
Pharmakologie	8	Pharmazie	80 Pharmakologie	52
Piaget, J.	147	Piaget, Jean	230 Piaget (Jean)	208
Psychometrie	36	Psychometrie	74 Psychometrie	438
schulpsychologischer Dienst	13	Schulpsychologische Beratung	306 Schulpsychologie + Soziale Dienste	0
Softwareergonomie	14	Softwareergonomie	10 Software + Ergonomie	194
Täter-Opfer- Beziehung	189	Taeter-Opfer- Beziehung	14 Täter + Kriminalitätsoffer	77
Umweltethik	49	Oekologische Ethik	2 Umwelt + Ethik	18

Dass auch das Problem der Eigennamen – und insbesondere der mangelnden Standardisierung dieser – durch Crosskonkordanzen positiv beeinflusst werden kann, wird am Beispiel des TheSoz-Deskriptors Piaget, J. deutlich. Durch die zwischengeschaltete Crosskonkordanz können – obwohl alle drei Thesauri den

8 Die Suche nach dem Schlagwort Schulpsychologie kombiniert mit der Freitextsuche Dienst bringt in Psyndex 43 relevante Treffer.

Namen unterschiedlich ansetzen – in allen drei Datenbanken Publikationen des Entwicklungspsychologen Jean Piaget gefunden werden.

Die weiteren, in Tabelle 3 aufgezeigten Beispiele sind dem Systematikbereich 3.4 Medizin und Gesundheitswesen des Thesaurus Sozialwissenschaften entnommen.

Tabelle 3: TheSoz-Deskriptoren und Termtransformationen des Systematikbereichs 3.4 „Medizin und Gesundheitswesen“ (Trefferzahlen: Stand August 2005)

TheSoz	SOLIS	TheBild (CK)	FIS Bildung	PsyT (CK)	Psyn- dex
Krankheit	6450	Krankheit	1506	Erkrankungen	2998
öffentliches Gesundheitswesen	239	Gesundheitswesen	628	Gesundheits- versorgungssystem	1835
Arzt-Patient- Beziehung	485	-	-	Ärzte + Patienten	234
Ethnomedizin	72	-	-	Volksmedizin	204
körperliche Entwicklung	75	Koerperliche Entwicklung	160	Physische Entwicklung	109
gesetzliche Krankenversicherung	491	Krankenversicherung	144	Krankenversicherung von Arbeitnehmern	15
Knochenkrankheit	-	Knochenkrankheit	2	Osteoporose	13
Impfung	7	Impfung	33	Immunisierung	12
Kinderlähmung	5	Kinderlaehmung	7	Poliomyelitis	7
geriatrische Rehabilitation	1	-	-	Geriatrische Patienten + Rehabilitation	123

Anhand einiger Deskriptoren (z. B. Ethnomedizin, Knochenkrankheit, Kinderlähmung und geriatrische Rehabilitation), die in den Sozialwissenschaften sicherlich ausgesprochene Spezialgebiete repräsentieren und für die z. T. im Thesaurus Bildung keine Entsprechungen gefunden werden konnten, wird sehr gut deutlich, warum die Crosskonkordanzen in Verbindung mit der interdisziplinären Suche eine entscheidende Verbesserung für den Nutzer bewirken. Für die oben genannten Deskriptoren, die in der Datenbank SOLIS kaum Treffer generieren, bringt gerade die Crosskonkordanz zur Psychologie eine große Erhöhung der Trefferzahlen, die einem sozialwissenschaftlichen Nutzer als intelligente Mehrwertleistung automatisch angeboten werden kann.

3.3 Zusammenfassung

- Für den Recherchierenden bedeutet der Einsatz von Crosskonkordanzen eine deutliche Vergrößerung des Suchraums, den er von jedem einzelnen der verwendeten Thesauri aus nun erreichen kann.
- Erste Untersuchungen in infoconnex zeigten, dass sich durch Crosskonkordanzen eine signifikante Erweiterung des Suchvokabulars ergibt (im Schnitt um ca. 50 % gegenüber dem Ausgangsvokabular).
- Besonders im interdisziplinären Bereich konnte belegt werden, dass Crosskonkordanzen einen informationellen Mehrwert bieten, da sie Nischen eines Fachgebiets mit potentiell zentraleren Bereichen eines anderen Fachgebiets verbinden können.
- Die vorliegenden, rein quantitativen Ergebnisse lassen zwar noch keine präzisen Aussagen zur Relevanz der durch Crosskonkordanzen gelieferten, zusätzlichen Treffer zu, die erreichte Vergrößerung der Treffermenge stellt aber eine notwendige Voraussetzung für die angestrebte Verbesserung des Ergebnisses dar.

4 Weitere Einsatzmöglichkeiten von Crosskonkordanzen

Aufgrund der positiven Erfahrungen mit dem Einsatz von Crosskonkordanzen im Rahmen von infoconnex soll in den künftigen Ausbaustufen des Portals das Potential dieser semantischen Technologie in größerem Umfang genutzt werden. Dies betrifft vor allem den Einsatz aller in den Crosskonkordanzen enthaltenen Relationen sowie die Implementierung von indirekten Transformationen in Fällen, in denen ein Transfer zwischen zwei Thesauri nicht auf direktem Weg möglich ist.

4.1 Einsatz der zusätzlichen Relationstypen

Die bislang in infoconnex eingesetzten Äquivalenzrelationen stellen nur einen kleinen Teil der insgesamt in den Crosskonkordanzen enthaltenen Beziehungen zwischen Deskriptoren zweier Thesauri dar. Sie unterscheiden sich von den anderen Relationstypen (Oberbegriff, Unterbegriff, Ähnlichkeit) im Grad der Übereinstimmung der semantischen Konzepte, die verbinden.

Während bei Äquivalenzrelationen davon ausgegangen werden kann, dass die Präzision der Termtransformation so hoch ist, dass sie ohne Rückfrage automatisch für den Nutzer durchgeführt werden kann, ist dies bei den anderen Relationen nicht zwangsläufig der Fall. Besonders bei Oberbegriffsrelation wie zwischen Hochadel und Adel verringert sich die Präzision und es ist zu erwar-

ten, dass deutlich mehr und vor allem weiter vom Informationsbedürfnis des Nutzers entfernte Treffer nachgewiesen werden. Das Gegenteil ist bei Unterbegriffsrelation wie zwischen Kampagne und Werbekampagne zu erwarten, bei der Ähnlichkeitsbeziehung zwischen Konfliktverhalten und Konfliktfähigkeit ist die Auswirkung auf das Ergebnis – und vor allem die Bewertung durch den Nutzer – weitgehend offen. Weitere Untersuchungen werden hier erst zeigen müssen, ob bestimmte Relationen ebenfalls automatisch, also ohne explizite Aktivierung durch den Nutzer, ausgeführt werden können, ohne das Ergebnis zu verschlechtern. Eine wesentliche Rolle werden in jedem Fall die Menge der zusätzlichen Treffer, der Einsatz dieser Relationen für sich alleine oder in Kombination mit anderen und die Fachgebiete spielen (mehrere Fachgebiete im Gegensatz zum Einsatz zwischen Thesauri eines Fachgebietes). Daneben entscheidet auch die Form der Bereitstellung dieser Möglichkeiten für den Nutzer über ihre Akzeptanz. In Stempfhuber 2003 werden Gestaltungsideen vorgestellt, wie Crosskonkordanzen unterschiedlichen Nutzergruppen in jeweils spezifischer Form und interaktiv zur Rechercheformulierung angeboten werden können.

4.2 Indirekte Termtransformation

Werden mehrere Datenbanken integriert durchsucht, wird die Anfrage im Idealfall durch eine Crosskonkordanz direkt in alle beteiligten Erschließungssysteme übersetzt (direkter Transfer). Allerdings ist es nicht immer möglich und sinnvoll, sämtliche Erschließungssysteme untereinander zu verknüpfen, da die Erstellung von Crosskonkordanzen – je nach Größe der Vokabulare – mit nicht zu vernachlässigendem Aufwand verbunden ist. Dies gilt nicht nur für fachlich weit entfernte Thesauri und Klassifikationen wie Physik und Psychologie, auch innerhalb eines Faches kann der Aufwand zu groß sein. Beispielsweise besteht in den Sozialwissenschaften der Wunsch, die Datenbank SOLIS des IZ gemeinsam mit der Datenbank SoLit des Deutschen Zentralinstituts für soziale Fragen (DZI) und dem Bibliothekskatalog des Sondersammelgebiets (SSG) Sozialwissenschaften an der Universitäts- und Staatsbibliothek (USB) Köln zu durchsuchen. Letzterer ist mit der Schlagwortnormdatei (SWD) der Deutschen Bibliothek und SoLit mit dem Thesaurus des DZI verschlagwortet. Der Aufwand, jedes der drei Vokabulare paarweise miteinander zu verbinden, ist bereits relativ groß; zieht man in Betracht, dass dies nur ein kleiner Teil der in den Sozialwissenschaften relevanten Datenbanken ist, so wird schnell deutlich, dass eine paarweise bidirektionale Verknüpfung von Vokabularen nicht skaliert, mit vernünftigem Aufwand also nicht zu leisten ist.

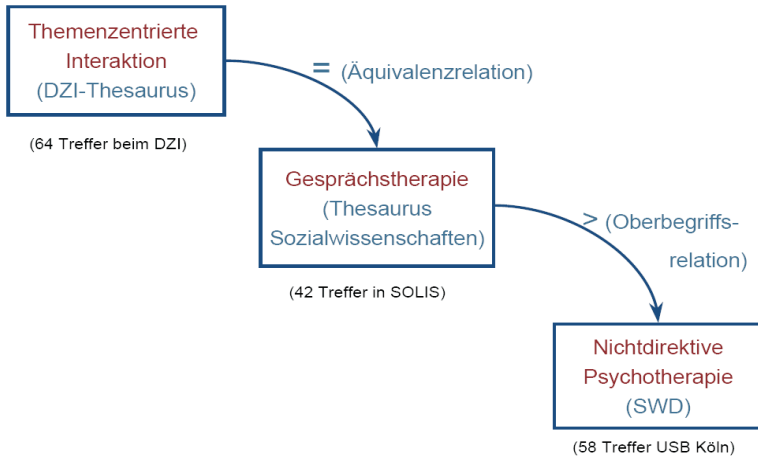


Abbildung 5: Indirekte Termtransformation (Beispiel)

Gelöst werden kann dieses Problem durch einen indirekten Termtransfer, also indem bei der Termtransformation ein „Umweg“ – oder Mehrschrittverfahren – gewählt wird. Dabei wird versucht, durch sequentielle Anwendung mehrerer Crosskonkordanzen einen Transfer zwischen zwei Erschließungssystemen zu erreichen, die nicht direkt miteinander verbunden sind. Abbildung 5 zeigt ein Beispiel, in dem ein Nutzer gleichzeitig in allen drei Datenbanken suchen möchte und dazu das Schlagwort themenzentrierte Interaktion aus dem DZI-Thesaurus verwendet. Im ersten Schritt – zur Suche in SOLIS – wird eine direkte Transformation (Gesprächstherapie) zum Thesaurus Sozialwissenschaften durchgeführt, woraus 42 Treffer resultieren. Für den zweiten Schritt – der Suche im Bibliothekskatalog – kann nun keine direkte Transformation vom DZI-Thesaurus aus durchgeführt werden, da eine derartige Crosskonkordanz nicht existiert. Stattdessen wird ein Zweischnittverfahren angewendet, in dem zunächst vom DZI-Thesaurus zum IZ-Thesaurus und erst von dort aus dann zur SWD transformiert wird (nichtdirektive Psychotherapie). Zur Sicherstellung einer möglichst hohen Präzision werden – wenn vorhanden – Äquivalenzrelationen verwendet. Da aber mit zunehmender Zahl an beteiligten Crosskonkordanzen dies nicht immer sichergestellt werden kann, dem Nutzer aber ein Ergebnis geboten werden soll, muss auch auf andere Relationen zurückgegriffen werden. Die theoretische Fundierung und praktische Erprobung von Mehrschrittverfahren stellt daher auch einen zukünftigen Forschungsschwerpunkt am IZ dar.

5 Ausblick

Crosskonkordanzen zwischen Erschließungsvokabularen, so haben erste Untersuchungen gezeigt, stellen ein probates Mittel dar, den Nutzer bei der gleichzeitigen Recherche in unterschiedlich erschlossenen Datenbanken zu unterstützen. Die Crosskonkordanzen leisten dabei automatisch den Transfer der Suchbegriffe aus der Anfrage des Nutzers in die unterschiedlichen Erschließungsvokabulare, mit denen die Datenbanken jeweils erschlossen sind. Der Nutzer wird in vielen Fällen vom entsprechenden kognitiven Aufwand befreit und kann sich intensiver der Bewertung der Rechercheergebnisse widmen. Entsprechende Benutzungsschnittstellen erlauben es dem fortgeschrittenen Nutzer, eine Vorschau auf das Ergebnis der Transformationen zu erhalten und diese sehr detailliert zu beeinflussen.

Der hier beschriebene Einsatz von Crosskonkordanzen bezog sich ausschließlich auf Textinformationen, also Literaturnachweisdatenbanken und Bibliothekskataloge. Doch auch im Bereich von Fakteninformationen, zum Beispiel bei statistischen Zeitreihen aus der Marktforschung, spielen sie eine Rolle, da hier ebenfalls eine Vielzahl von Datenbanken gemeinsam durchsucht werden müssen, um die benötigten Informationen zusammenzutragen (s. Stempfhuber et al. 2003). Da die statistischen Ämter weltweit in der Regel länderspezifische Nomenklaturen (also kontrollierte Vokabulare für Fakteninformation) zur Erschließung ihrer Statistiken verwenden, müssen gerade im internationalen Bereich häufig mehrere Nomenklatorsysteme verwendet werden, um die gewünschte Information zu erhalten.

Eine konzeptionelle Erweiterung erfährt der Einsatz von Crosskonkordanzen bei der gleichzeitigen Recherche von Text- und Faktendaten (s. Stempfhuber 2005). Hier gilt es nicht nur die Sprache der amtlichen Statistik auf die Sprache in Fachpublikationen abzubilden. Auch die grundsätzlich verschiedene Art der Inhaltserschließung von Texten und Fakten bereitet in manchen Fällen Probleme. Dies wird besonders dann deutlich, wenn nicht Aggregatdaten, sondern Primärdaten (z. B. sozialwissenschaftliche Umfragen) betrachtet werden. Die Komplexität der semantischen Konstrukte, die zur Inhaltserschließung von Primärdaten benötigt werden, um einem Nutzer nicht nur das Auffinden, sondern auch die korrekte Interpretation der Daten zu ermöglichen, geht weit über den Aufwand in der normalen dokumentarischen Praxis hinaus und lässt typische Thesauri oder Nomenklaturen nicht als geeignet erscheinen. Zukünftige Forschungsarbeiten werden daher verstärkt auch Ontologien mit einbeziehen, wenn es darum geht, Text- und Faktendaten zu dokumentieren und integriert recherchierbar zu machen (siehe Krause&Stempfhuber 2005).

Literatur

- Boekhorst, P.; Kayß, M.; Poll, R. (2003): Nutzungsanalyse des Systems der überregionalen Literatur- und Informationsversorgung. Teil 1: Informationsverhalten und Informationsbedarf der Wissenschaft. Universitäts- und Landesbibliothek Münster und infas Institut für angewandte Sozialwissenschaft GmbH.
- Binder, G. et al.: Heterogenitätsbehandlung bei textueller Information verschiedener Datentypen und Inhalterschließungsverfahren. IZ-Arbeitsbericht Nr. 24, 2002. S. 64 ff.,
http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab_24.pdf
- Hellweg, Heiko; Krause, Jürgen; Mandl, Thomas; Marx, Jutta; Müller, Matthias N.O.; Mutschke, Peter; Strötgen, Robert (2001): Treatment of Semantic Heterogeneity in Information Retrieval. Bonn: IZ Sozialwissenschaften. 47 S. (IZ-Arbeitsbericht; Nr. 23) URL:
http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab_23.pdf
- Krause, Jürgen (2003): Standardisierung von der Heterogenität her denken - Zum Entwicklungsstand bilateraler Transferkomponenten für digitale Fachbibliotheken. IZ-Arbeitsbericht Nr. 28, 2003. S.14 ff.,
http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab_28.pdf
- Krause, Jürgen (2004): Konkretes zur These, die Standardisierung von der Heterogenität her zu denken. In: *ZfBB: Zeitschrift für Bibliothekswesen und Bibliographie* 51, Nr. 2, S. 76 - 89
- Krause, Jürgen; Stempfhuber, Maximilian (2005): Nutzerseitige Integration sozialwissenschaftlicher Text- und Dateninformationen aus verteilten Quellen, (01. Juli). - 6. Wissenschaftliche Tagung „Datenfusion und Datenintegration“, Statistisches Bundesamt Wiesbaden, 30. Juni - 01. Juli 2005
- Mayr, Philipp; Stempfhuber, Maximilian; Walter, Anne-Kathrin (2005): Auf dem Weg zum wissenschaftlichen Fachportal – Modellbildung und Integration heterogener Informationssammlungen. In: Schmidt, Ralph (Hrsg.): 27. DGI-Online-Tagung. Frankfurt am Main: DGI.
- Poll, Roswitha (2004): Nutzungsanalyse des Systems der überregionalen Literatur- und Informationsversorgung Teil 1: Informationsverhalten und Informationsbedarf der Wissenschaft. In: *ZfBB* 51(2004) S. 59 – 75.
- Schott, Hannelore; Schroeder, Albert (2004): Crosskonkordanzen von Thesauri und Klassifikationen. S. 41 - 49. In: Budin, Gerhard; Ohly, H. Peter (Hrsg.):

- Wissensorganisation in kooperativen Lern- und Arbeitsumgebungen. Würzburg: Ergon. (Fortschritte der Wissensorganisation 8).
- Stempfhuber, Maximilian (2003): Objektorientierte Dynamische Benutzungsoberflächen – ODIN. Behandlung semantischer und struktureller Heterogenität in Informationssystemen mit den Mitteln der Softwareergonomie. Bonn: Informationszentrum Sozialwissenschaften. (Forschungsberichte; Band 6).
- Stempfhuber, Maximilian (2005): Text-Fakten-Integration in Informationssystemen. S. 263 - 274. Knowledge eXtended: die Kooperation von Wissenschaftlern, Bibliothekaren und IT-Spezialisten; 3. Konferenz der Zentralbibliothek; 2. - 4. November 2005; Vorträge und Poster. Jülich: Forschungszentrum, Zentralbibliothek. (Schriften des Forschungszentrums Jülich: Reihe Bibliothek/Library; 14).
- Stempfhuber, Maximilian; Hellweg, Heiko; Schaefer, André (2002): ELVIRA: User Friendly Retrieval of Heterogenous Data in Market Research. S. 299 - 304. In: Callaos, Nagib; Hernandez-Encinas, Luis; Yetim, Fahri (Hrsg.): SCI 2002: The 6th World Multiconference on Systemics, Cybernetics and Informatics; July 14 - 18, 2002, Orlando, USA; Proceedings, Vol. I: Information Systems Development I. Orlando: TPA Publ.
- Strötgen, Robert (2004): ASEMOS. Weiterentwicklung der Behandlung semantischer Heterogenität. In: Bekavac, Bernard; Herget, Josef; Rittberger, Mark (Hrsg.): Information zwischen Kultur und Marktwissenschaft; Proceedings des 9. Internationalen Symposiums für Informationswissenschaft (ISI 2004). Konstanz: UVK. (Schriften zur Informationswissenschaft; Bd. 42), S. 269 - 281.

Angaben zu den Autoren

Informationszentrum Sozialwissenschaften (IZ)
 Abt. Forschung und Entwicklung
 Lennéstraße 30
 53113 Bonn
 walter@bonn.iz-soz.de

Anne-Kathrin Walter studierte Informatik an der Universität Bremen. Seit September 2004 arbeitet sie im Projekt „Modellbildung und Heterogenitätsbehandlung“ am Informationszentrum Sozialwissenschaften in Bonn. Ihre Forschungsinteressen liegen im Information Retrieval, u.a. im Bereich der semantischen Heterogenitätsbehandlung.

Philipp Mayr studierte Bibliothekswissenschaft, Informatik und Soziologie an der Humboldt-Universität zu Berlin. Zu seinen Forschungsinteressen gehören

Information Retrieval im Bereich der Internet-Suchmaschinen sowie Metriken des Internet (Webometrie). Seit November 2004 arbeitet er im Projekt „Modellbildung und Heterogenitätsbehandlung“. Philipp Mayr ist wissenschaftlicher Mitarbeiter am Informationszentrum Sozialwissenschaften in Bonn.

Maximilian Stempfhuber studierte Informationswissenschaft und Wirtschaftsinformatik an der Universität Regensburg und promovierte in Informatik an der Universität Koblenz-Landau. Er ist stellvertretender Direktor des Informationszentrum Sozialwissenschaften in Bonn und leitet dort die Abteilung informationswissenschaftliche Forschung und Entwicklung. An der Universität Koblenz-Landau führt er im Rahmen von Lehraufträgen Veranstaltungen in der Informatik und Computervisualistik durch. Seine Forschungsschwerpunkte sind Softwareergonomie, Visualisierung und Information Retrieval.

Arne Ballay studierte Geografie, Schwerpunkt Geo-Informatik and der Rheinischen Friedrich-Wilhelms Universität Bonn. Seit November 2001 arbeitet er als wissenschaftlicher Mitarbeiter am Informationszentrum Sozialwissenschaften unter anderem im Projekt „Informationsverbund Pädagogik - Sozialwissenschaften - Psychologie (infoconnex)“. Seine Arbeits- und Forschungsschwerpunkte liegen in der Entwicklung internetbasierte Informationssysteme, vor allem in der Nutzung geo-kodierter Informationen bei der Recherche und Ergebnisvisualisierung.

Suchmaschinen, Strukturierung und Metadaten

Wolfram Sperber

Konrad-Zuse-Zentrum für Informationstechnik Berlin

Zusammenfassung

Suchmaschinen sind für die meisten Nutzer die erste Adresse im Internet auf der Suche nach interessanten Web-Seiten. Suchmaschinen sind prägend für das Nutzerverhalten im Web. Selbst in die Umgangssprache haben die Suchmaschinen Einzug gehalten: „Googeln“ etwa ist zum Synonym für Suche im Web geworden.

Neben den großen (universellen) Suchmaschinen gibt es eine Vielzahl weiterer Suchmaschinen / Portale, die sich auf ein bestimmtes Thema konzentrieren. Für einen effizienten Betrieb solcher Suchmaschinen ist die Entwicklung neuer Techniken und Methoden erforderlich.

Der Beitrag diskutiert die Anforderungen an spezialisierte Suchmaschinen und den erreichten Entwicklungsstand, insbesondere für den wissenschaftlichen Bereich.

1 Suchmaschinen

Nicht nur an der Börse sorgen die Internet-Suchmaschinen für Furore. Die universellen Suchmaschinen, allen voran Google, sind für viele Anwender der Schlüssel für die Nutzung des Internet. Sie sind einfach und intuitiv in der Handhabung und bieten eine Suche über das komplette Web, zumindest den sichtbaren Teil desselben, an. Die Nutzer müssen in der Regel nur ihre Suchbegriffe eintippen.

Dafür erhalten sie in Sekundenbruchteilen eine Trefferliste der gefundenen Web-Seiten. Bei mehr als 8.000.000.000 Objekten, die etwa Google nach eigenen Angaben derzeit nachweist, eine wichtige Voraussetzung für die Akzeptanz der Suchmaschine.

Und die Suchmaschinen leisten noch mehr:

- Die Trefferlisten sind sortiert, die relevanten Treffer sind auch tatsächlich häufig unter den ersten Treffern.

- Die Nutzung der Suchmaschinen ist kostenfrei (zumindest für die Anfrage). Trotzdem lässt sich mit Suchmaschinen viel Geld verdienen, die Werbung ist etwa für Google eine der Haupteinnahmequellen.
- Die (universellen) Suchmaschinen sind nicht nur bequem für die Anfragenden, sondern auch für die Ersteller der Informationen. Die (universellen) Suchmaschinen nehmen das Web so, wie es ist. Sie stellen keine Voraussetzungen an die Struktur der Dokumente, verlangen weder spezifische Formate noch eine zusätzliche inhaltliche Erschließung etc. Einzige Bedingung ist, dass die Dokumente im Netz gefunden werden können (das ist in der Regel dann gegeben, wenn mindestens ein Link auf ein Dokument verweist).

Ist damit das Problem der Suche im Internet ein für allemal gelöst? Wie die Fragestellung schon vermuten lässt, sprechen diverse Gründe dagegen.

Im Allgemeinen

Suchmaschinen sammeln die Information aus dem Internet zusammen. Dazu haben die Suchmaschinen so genannte Seed-Listen, die URLs enthalten. Die Links, die auf diesen Seed-Listen vorhanden sind, werden ausgelesen und als Startpunkt für das Auffinden weiterer Ressourcen benutzt. Informationen, die zwar im Netz vorhanden, aber nicht verlinkt sind, können mit diesem Vorgehen nicht gefunden werden, sondern müssen zusätzlich in die Suchmaschinen eingebunden werden.

Die Informationen müssen aber dann in irgendeiner Weise auch suchbar gemacht werden. Dazu ist es nötig, die gesammelten Informationen auszuwerten. Das beginnt mit der Volltextindexierung einer HTML-Seite und ist beliebig erweiterbar, z.B. Kontextanalysen, Klassifizierung, bis hin zu logischen Schlussfolgerungen.

Ziel der universellen Suchmaschinen des Web ist es, dem Nutzer eine Suche über möglichst viele Ressourcen des Web zu bieten. Diese Art von Suchmaschinen ist nicht selektiv, der Inhalt der Informationen ist zunächst von untergeordneter Wichtigkeit. Im Gegenteil, universelle Suchmaschinen wollen ja gerade das Web so komplett wie möglich nachweisen. Abbildung 1 zeigt die ersten Treffer in Google.de für den Suchbegriff „Funktion“. Der Nutzer, der ausschließlich an Informationen eines einzigen Themengebietes interessiert ist, sieht sich häufig mit weiteren Informationen außerhalb seines konkreten Informationsbedürfnisses konfrontiert.

Es gibt allerdings Entwicklungen, die dem Nutzer eine an seinen Interessen orientierte Auswahl der Suchergebnisse ermöglichen. Beispielsweise bieten etwa die erweiterten Suchmöglichkeiten von Google und Yahoo an, Suchergebnisse in Abhängigkeit von der Sprache oder vom Domainnamen zu filtern.

The image shows a screenshot of a Google search page for the term "Funktion". At the top, the Google logo is visible on the left, and navigation links for "Web", "Bilder", "Groups", "Verzeichnis", "News", "Frage", and "Mehr" are on the right. Below the logo is a search bar with the text "Suche: Das Web Seiten auf Deutsch Seiten aus Deutschland". The search results are displayed in a list format, with each entry including a title, a brief description, and a URL. The first result is "Funktion-Online" with a description about a math assistant for Windows. Other results include "Funktion - Wikipedia", "Hash-Funktion - Wikipedia", "Funktionen 1 - Mathematische Hintergründe", "funktion.de - Internet dienstleistungen", "Frich Mendelssohn: Dynamik und Funktion", and "Funktion Ono". On the right side of the page, there is an "Anzeigen" (Ads) section with several advertisements, including one for "Die große Homöopathie" and another for "Linux Email & Cal-Syncer".

Abb. 1: Google-Abfrage nach „Funktion“

Einen höheren Stellenwert als der Inhalt hat die Vernetzung der Information. Google analysiert die Links, die von einer Web-Seite ausgehen bzw. auf die Web-Seite zeigen und schließt daraus auf die Bedeutung der Web-Seite. Je mehr Links auf eine Web-Seite zeigen und je wichtiger die Ausgangsseiten der Links sind, desto bedeutender ist die Seite, d.h. der so genannte Page Rank der Seite. Der Page Rank lässt sich mathematisch als Gleichungssystem modellieren, das iterativ gelöst werden kann, siehe (Brin, Page 1998).

Universelle Suchmaschinen müssen mit den verschiedensten Formaten fertig werden, in denen die Informationen vorliegen können. Die Auswertung verschiedener Formate erfordert spezielle Werkzeuge zur Aufbereitung. Ob und wie strukturierte Dokumente, etwa XML kodierte Dokumente, von den universellen Suchmaschinen ausgewertet werden, ist nicht bekannt. Die spezifische Auswertung strukturierter Dokumente, die bis zur Auswertung von MathML kodierten Formeln reichen kann, erfordert die Entwicklung von Software, die genau diese spezifischen Auswertungen vornehmen kann.

In diesem Zusammenhang ist auch zu erwähnen, dass die universellen Suchmaschinen bewusst auf die Auswertung von Metadaten verzichten. Ihre Begründung: Metadaten können missbraucht werden, etwa für Spam Informationen, die einzig und allein das Ziel haben, das Ranking einer Seite zu verbessern. Tatsächlich werden Web-Seiten und Metadaten auch immer wieder manipuliert.

Damit entfallen natürlich die Möglichkeiten, die sich aus der zusätzlichen Auswertung der Metadaten ergeben. Gerade für den wissenschaftlichen Bereich ist aber die Gefahr missbräuchlicher Nutzung relativ gering und die Bedeutung

der Metadaten für eine zusätzliche semantische inhaltliche Erschließung wichtig.

Die universellen Suchmaschinen, allen voran Google, beherrschen derzeit den Suchmaschinenmarkt.

Zudem indexieren die Suchmaschinen die Informationen nicht nur, sondern sie archivieren die gefundenen Informationen auch. Google strebt an, zum führenden Informationsprovider im Web zu werden: „Google’s mission is to organize the world’s information and make it universally accessible and useful“, siehe (Google – Mission Statement). Eine Monopolisierung des Zugangs zu den Informationen liegt aber sicher nicht im Interesse der Internet-Nutzer.

Im Speziellen

Die universellen Suchmaschinen geben zwar bei allgemeinen Anfragen eine riesige Anzahl von Treffern an, die Anzeige ist jedoch auf 1.000 Treffer beschränkt. Um alle Treffer zu sehen, müssen die Anfragen eingeschränkt werden, etwa durch zusätzliche Suchbegriffe oder durch Beschränkung der Suche auf einzelne Web Sites. Spezielle Anfragen (die Anzahl der Treffer ist kleiner als 1.000) zeigen, dass die Trefferlisten der Suchmaschinen nicht alle Dokumente enthalten, die im Web tatsächlich vorhanden sind.

Die prinzipielle Arbeitsweise der Suchmaschinen ist kanonisch:

- Die Informationen des Web werden von Web Crawlern gesammelt.
- Die gesammelten Informationen werden ausgewertet (Indexierung, Klassifizierung, Ranking etc.).
- Die Informationen sind über ein User Interface suchbar.

Bemerkung: Die Arbeitsgänge können parallel ablaufen, beispielsweise können das Sammeln und die Auswerten zusammen ablaufen.

Die genauen Methoden und Arbeitsweisen, insbesondere die Auswertung der Suchergebnisse und deren Ranking, sind Betriebsgeheimnis der Suchmaschinenbetreiber und werden ständig modifiziert.

Bemerkung: Die Positionierung einer Web-Seite auf vorderen Plätzen bei den universellen Suchmaschinen bedeutet Sichtbarkeit im Web, was heute auch wichtig für den kommerziellen Erfolg eines Unternehmens ist. Es haben sich inzwischen viele Firmen etabliert, die ihr Geld mit Suchmaschinenoptimierung verdienen, d.h. vorhandene Web Seiten so aufbereiten, dass die Web Seiten bei Anfragen an vorderer Stelle der Trefferlisten erscheinen. Die Suchmaschinenbetreiber sind so gezwungen, ihre Rankingmethoden geheim zu halten und häufig zu ändern, um Missbrauch vorzubeugen. Das heißt aber auch, dass Methoden und Arbeitsweisen der universellen Suchmaschinen nicht transparent sind.

Neben den universellen Suchmaschinen gibt es eine Vielzahl von Suchmaschinen/Portalen, die sich auf spezielle Inhalte oder Themengebiete spezialisiert haben. Seit Mitte der 90-iger Jahre beschäftigen sich verschiedene Entwicklergruppen mit dem Thema spezialisierte Suchmaschinen und „Topical Web Crawlers“, siehe etwa Pant, Srinivasan, Menczer 2004 oder McCallum, Nigam, Renne, Seymore 1999.

Die prinzipielle Arbeitsweise spezialisierter Suchmaschinen unterscheidet sich nicht von der oben geschilderten. Der wesentliche Unterschied besteht darin, dass die spezialisierten Suchmaschinen „nur“ die für sie relevanten Webseiten berücksichtigen wollen und eben nicht das ganze Web.

Die beste Möglichkeit besteht darin, schon während des Crawlings der Webseiten (allgemeiner der Web-Objekte) zu entscheiden, ob eine Webseite gesammelt werden soll oder nicht. Es muss also während des Crawlings über die Relevanz der Seite entschieden werden, mit anderen Worten: es muss der Inhalt der Seite evaluiert werden, also festgestellt werden, ob die Seite zu einem Themengebiet passt oder nicht.

Dieser Prozess wird als Klassifizierung bezeichnet. Die Klassifizierung kann sich auf verschiedene Dinge beziehen: etwa die Klassifizierung von Texten entsprechend vorgegebener Inhalte oder die Klassifizierung von Informationstypen wie Publikationen, Software, Homepages von Personen und Institutionen etc.

Die Klassifizierung kann sowohl manuell als auch automatisch erfolgen:

- Die Seite hat manuell erstellte Metadaten, die den Inhalt der Webseite charakterisieren, etwa Schlüsselwörter oder Klassifizierungen der Seite. Ob man den Metadaten vertraut, kann vom Provider der Webseite abhängig gemacht werden. Beispielsweise können die Metadaten akademischer Institutionen als vertrauenswürdig eingestuft werden. Wie die Metadaten kodiert werden, ist letztlich unwichtig. Natürlich vereinfacht die Anwendung von Standards für die Metadaten die Auswertung der Seiten wesentlich. Viele Webseiten, auch im wissenschaftlichen Bereich, haben allerdings keine manuell erstellten Metadaten.
- Die Information muss aus den Webseiten, z.B. HTML-Seiten, automatisch extrahiert werden. Dazu können lexikalische Methoden, Struktur- und Kontextanalysen der Seite verwendet werden.

Die Analyse des Inhalts einer Seite ist also immer mit zusätzlichem Aufwand verbunden. Die maschinelle Behandlung kann auch zu neuen Metadaten über die Webseite selbst führen, etwa zur Generierung von Schlüsselwörtern oder der Zuweisung von Klassifikationen zu der Seite.

Die Vorteile spezialisierter Suchmaschinen liegen auf der Hand: Dieser Typ von Suchmaschinen sammelt nur einen speziellen Ausschnitt der Informationen, die im Web vorhanden sind. Spezialisierte Suchmaschinen wenden sich an bestimmte Nutzergruppen, die spezifische Informationsbedürfnisse haben. Spezialisierte Suchmaschinen können die Informationen in spezieller Form aufbereiten und qualifizierte Such-/Navigationsmöglichkeiten bieten, die auf den Themenbereich der Suchmaschine zugeschnitten sind. Die Ergebnislisten etwa können entsprechend ihrer Klassifikation geordnet oder nach speziellen Rankingkriterien sortiert werden.

Der Nachteil spezialisierter Suchmaschinen: Der Aufwand steigt. Ein Nutzer muss i.A. verschiedene Suchmaschinen und Suchstrategien benutzen, wenn er Informationen aus verschiedenen Themengebieten sucht.

Bemerkungen:

- Die Vorteile spezialisierter Suchmaschinen sind identisch mit den oben geschilderten Vorteilen der Nutzung von Metadaten. Das ist nicht zufällig: Spezialisierte Suchmaschinen und Metadaten hängen eng miteinander zusammen. Die Evaluierung ist mit einer Analyse der Web-Seiten verbunden. Die Ergebnisse der Analyse machen Aussagen über die Web-Seite, sind also Metadaten.
- Spezialisierte Suchmaschinen müssen einen Mehrwert gegenüber den universellen Suchmaschinen bieten. Eine Volltextsuche, eingeschränkt auf ein spezielles Gebiet allein, ist nicht ausreichend.
- Die Analyse der Web-Seiten durch die spezialisierten Suchmaschinen ist die Grundlage für den Mehrwert der spezialisierten gegenüber den universellen Suchmaschinen.
- Eine Kooperation/Kombination zwischen den universellen Suchmaschinen und den spezialisierten Suchmaschinen ist sinnvoll. Die universellen Suchmaschinen können etwa auf die spezialisierten Suchmaschinen verweisen oder spezielle Suchmaschinen abfragen und deren Ergebnisse in ihre Resultate integrieren.

2 Brauchen die Wissenschaften eigene Suchmaschinen?

Eigentlich eine überflüssige Frage. Gerade in den Wissenschaften ist es wichtig, gezielt und präzise nach den vorhandenen Informationen suchen zu können. Die universellen Suchmaschinen, zumindest in der heutigen Form, genügen diesen Ansprüchen nicht.

In den Wissenschaften wurden auch schon vor dem Web spezialisierte Suchmaschinen verwendet.

Als Beispiel sei hier die Mathematik aufgeführt: Die Datenbanken Zentralblatt Math (Zentralblatt Math) und Math Reviews (Math Reviews) haben sich in den 80-er Jahren aus gedruckten Referateorganen entwickelt, die eine spezialisierte Suche über mathematische Publikationen bieten. Die Datenbanken erfassen mathematik-relevante Publikationen nahezu vollständig, und erstellen dazu bibliographische Metadaten, u.A. Autor, Titel, Quelle, Klassifikation, Schlüsselwörter etc.

Hier wird Klassifikation im Sinn einer inhaltlichen Einordnung einer Publikation in ein mathematisches Spezialgebiet verstanden. Für die Klassifizierung mathematischer Publikationen steht die „Mathematical Subject Classification“, MSC (MSC 2000), zur Verfügung.

Zusätzlich enthalten die Einträge der Datenbanken Besprechungen der Arbeiten, die von Fachkollegen erstellt worden sind.

Die Datenbanken ermöglichen den Nutzern einen schnellen Überblick über die auf ihren Forschungsgebieten neu erschienenen Arbeiten.

Bis in die 90er Jahre war die Nutzung der Datenbanken wenig intuitiv (setzte u.a. die Kenntnisse spezieller Retrieval-Sprachen voraus). Einfache graphische Nutzeroberflächen wurden erst mit dem Aufkommen des Web in den 90er Jahren entwickelt.

Neben den klassischen Formen für die Publikation wissenschaftlicher Ergebnisse wie Journalartikel und Bücher haben sich in den letzten Jahren neue Informationstypen etabliert, die für das wissenschaftliche Arbeiten wichtig sind, z.B. Software und Simulationen, Visualisierungen, Messdaten etc.

Für Software wurde mit dem „Guide to Available Mathematical Software“, GAMS (GAMS), ein eigenes Klassifikationssystem entwickelt, das einen effizienten Zugang speziell zu mathematischer Software unterstützen soll.

Die oben beschriebene Situation ist nicht nur für die Mathematik typisch, sondern trifft auch auf andere Wissenschaften zu:

- Die Anzahl der Informationen hat zugenommen. Es sind neue Typen für das Publizieren wissenschaftlicher Informationen entstanden.
- Die Informationen in den Wissenschaften werden durch unterschiedliche Merkmale/Klassifikationen erschlossen und beschrieben. Die Klassifikationssysteme können sich überlappen. Die Merkmale/Klassifikationen umfassen sowohl formale als auch inhaltliche Aspekte.
- Suchmaschinen/Portale in den Wissenschaften sollen dem Nutzer die Suche nach verschiedenen Merkmalen/Klassifikationen gestatten und die Ergebnisse entsprechend den speziellen Anforderungen der Community aufbereiten und präsentieren.

Ein weiteres wesentliches Argument, das für den Aufbau und den Betrieb fachspezifischer Suchmaschinen spricht, sind spezielle Formate, die in den Wissenschaften verwendet werden. In der Mathematik und den Naturwissenschaften stellen insbesondere mathematische Formeln eine Schwierigkeit für die Präsentation und Auswertung dar.

Heute wird TeX als de facto-Standard für das Publizieren mathematischer Arbeiten eingesetzt. TeX ist eine Markup-Sprache mit dem Schwerpunkt auf mathematischen Inhalten. Die TeX-Dateien werden für den Druck nach PS oder PDF konvertiert. PS oder PDF formatierte Dokumente lassen sich allerdings nicht mehr nach mathematischen Termen oder Formeln absuchen.

Eine Lösung ist die Extensible Markup Language, XML (XML). XML bietet einen Rahmen zur Entwicklung fachspezifischer Sprachen. Die Mathematik hat etwa mit der Mathematics Markup Language, MathML (MathML), OpenMath (OpenMath) und Open Mathematical Documents, OMDoc (OMDoc), spezifische XML-Sprachen unterschiedlicher Mächtigkeit entwickelt. In so kodierten Dokumenten stecken zusätzliche semantische Informationen, die ein besseres Retrieval ermöglichen.

Eine fachspezifische Auswertung XML formatierter Dokumente wäre zwar für die universellen Suchmaschinen prinzipiell möglich, angesichts des geringen Anteils solcher spezifisch formatierter Informationen am gesamten Volumen des Web ist der Aufwand für die Entwicklung und den Einsatz spezieller Methoden zur Auswertung solcher Dokumente für die Betreiber universeller Suchmaschinen schlicht zu hoch.

Bleibt die Frage offen, wer fachspezifische Suchmaschinen/Portale entwickeln und betreiben soll und kann?

2.1 Fachspezifische Suchmaschinen sind eine Entwicklungsaufgabe

Wie oben angedeutet, sind Entwicklung und Betrieb fachspezifischer Suchmaschinen/Portale aufwändig. Der Aufwand und die Methoden hängen vom Thema ab. Eine Suchmaschine zu einem inhaltlich eng umrissenen Thema, das aus einer „überschaubaren“ Menge an Objekten besteht, die zudem auf wenige Domains beschränkt sind, lässt sich durchaus händisch verwalten und betreiben. Bei anderen Themengebieten, die inhaltlich breiter sind und die Web-Seiten sich auf viele unterschiedliche Domains verteilen, wie z.B. der Mathematik, müssen möglichst viele Komponenten fachspezifischer Suchmaschinen automatisch betrieben werden. Nur so lässt sich der Aufwand für den Betrieb der Suchmaschine in Grenzen halten.

Der Aufbau fachspezifischer Suchmaschinen für das Web fängt beim Crawling an. Die Web-Seiten müssen - wie oben schon beschrieben - die für das Themengebiet relevante Information finden.

Die Arbeit von G. Pant, P. Srinivasan und F. Menczer (Pant, Srinivasan, Menczer 2004) gibt einen Überblick über die Thematik des themenspezifischen Crawlings und stellt verschiedene Algorithmen für „intelligentes“ Crawling vor. Diese Verfahren wurden von den Autoren auch schon in der Praxis eingesetzt, etwa um ein Portal für biomedizinische Web-Seiten (Srinivasan, Mitchell, Bodenreider, Pant, Menczer 2002) aufzubauen.

Fünf Verfahren werden in (Pant, Srinivasan, Menczer 2004) vorgeschlagen, um eine Seite zu analysieren:

1. Naive Best-First Crawler

Der Naive Best-First Crawler basiert auf dem Vergleich der Wörter verschiedener Dokumente. Eine Web Seite ist ein Vektor von Worten, die entsprechend ihrer Häufigkeit gewichtet sind. Der Abstand zweier Dokumente wird über das Skalarprodukt im R^n gemessen, wobei ein neues Dokument mit dem „besten“ Dokument verglichen wird.

2. SharkSearch

SharkSearch misst die Ähnlichkeit zweier Dokumente anhand der Links. Der Text im Anker des Links, der Text des Links und seiner Umgebung (Link-Kontext) und der Wert der Dokumente, die den Link auf das Dokument enthalten, werden als Maß für die Relevanz einer Seite für das Themengebiet genommen. Die Textanalyse erfolgt wie im Fall des Naive Best-First Crawlers.

3. Focused Crawler

Der Focused Crawler setzt auf die Klassifizierung der Seiten entsprechend vorgegebener Klassifikationssysteme (Taxonomien) auf. In der Arbeit wird vorgeschlagen, dafür die Bayes Methode zu nutzen, siehe unten.

4. Context Focused Crawler

Der Context Focused Crawler startet mit einer Link Analyse. Dazu wird der Kontext Graph untersucht, der entsteht, wenn die ausgehenden Links einer Web-Seite analysiert werden. Man misst den Linkabstand zwischen den Web-Seiten. Web-Seiten, die denselben Linkabstand vom Ausgangspunkt haben, werden zusammengefasst. Diese Mengen werden dann automatisch klassifiziert, etwa wiederum mit Bayes.

5. InfoSpider

Mit InfoSpidern bezeichnen Pant, Srinivasan, Menczer eine sich dynamisch steuernde Menge von Web Agenten, die das Web auf relevante Seiten untersucht. Für die Untersuchung werden Schlüsselwortlisten und neuronale Netze genutzt.

Generell sind die Methoden und Verfahren für fachspezifische Suchmaschinen in der Entwicklung, insbesondere betrifft das die Methoden für die automati-

sche Klassifizierung der Web Seiten, die auch für den Aufbau prototypischer Suchmaschinen für die Wissenschaften/Themenbereiche eingesetzt worden sind.

Die Methoden für die automatische Klassifizierung von Web-Seiten hängen vom Typ der Seite und dem Inhalt der Seite ab. Am intensivsten sind die Methoden zur Klassifikation von Texten untersucht. Lexikalische/linguistische Methoden analysieren das Wortmaterial in den Texten. Daraus ergeben sich Dokumentvektoren, die die in den Dokumenten auftauchenden Terme und deren Häufigkeiten enthalten. Üblicherweise werden zuvor noch Stoppwörter aus der Liste aller Wörter entfernt und die Wörter auf Wortstämme reduziert. Es gibt verschiedene Möglichkeiten und Verfahren, um die Ähnlichkeit (den Abstand) zwischen den Dokumentvektoren zu messen, z.B. mit dem bekannten Skalarprodukt zwischen Vektoren im \mathbb{R}^n . Klassifikationsprobleme werden damit auf mathematische Modelle zurückgeführt.

Automatische Klassifikation ist ein Problem des automatischen Lernens. Als Startpunkt wird immer eine Menge von Dokumenten verwendet, deren Klassifikation bekannt ist: die Trainingsmenge. Aus diesen Dokumenten werden dann Term-Häufigkeitsvektoren abgeleitet, die eine Klasse beschreiben. Neue Dokumente werden mit den Trainingsmengen der verschiedenen Klassen verglichen.

Es gibt eine Vielzahl von Methoden zur automatischen Klassifizierung, etwa

- Nearest Neighbors (Berechnung des Abstandes zwischen Web-Seiten mittels des Skalarproduktes im \mathbb{R}^n),
- Naive Bayes (Berechnung der Wahrscheinlichkeit, dass ein Dokument zu einer Klasse gehört, dazu wird das Konzept der bedingten Wahrscheinlichkeit eingesetzt),
- Support Vector Machines (Berechnung von Hyperebenen, die die Dokumente einer Klasse von den übrigen separieren).

Erste eigene Versuche mit verschiedenen Methoden haben gezeigt, dass die Verfahren für die automatische Klassifizierung sehr stark von den konkreten Klassifizierungsaufgaben abhängen. Viele Fragen sind noch offen, siehe auch (Bekermann, McCallum, Huang 2005), etwa

- die Güte und Zuverlässigkeit der Klassifizierungsmethoden, auch in Abhängigkeit von der üblichen hierarchischen Struktur von Klassifikationssystemen (Taxonomien).
- die Auswahl der Parameter, etwa die Größe und der Einfluss der Trainingsmengen auf den Klassifizierungsprozess. Die Trainingsmengen müssen „hinreichend“ groß sein, d.h. das relevante Wortmaterial der Klassen enthalten.

Ein Teilschritt zur automatischen Klassifizierung kann die Erzeugung von Schlüsselwörtern sein, die dann ebenfalls semantische Aussagen über eine Webseite (Dokument) machen.

Die zur Zeit größte fachspezifische Suchmaschine, die ausschließlich auf die Wissenschaften fixiert ist, ist nach eigener Darstellung Scirus (Scirus). Diese Suchmaschine wird vom Elsevier Verlag betrieben.

Scirus ist eine Suchmaschine für wissenschaftliche Publikationen (u.a. die Zeitschriftenartikel, die im Elsevier Verlag erschienen sind, aber auch aus community-basierten Systemen wie E-Print ArXiv (ArXiv) und versucht darüber hinaus, weitere wissenschaftsrelevante Web-Seiten in die Suchmaschine einzubeziehen.

Scirus bietet verschiedene Suchmöglichkeiten, siehe Abb. 2, für die wissenschaftliche Information:

- nach Informationstypen
- nach Quellen
- nach Wissenschaftsgebieten

SCIRUS
for scientific information only

About Us Newsroom Advisory Board Submit Web Site Search Tips Contact Us

Advanced Search [Basic Search](#) [Search Preferences](#)

 The complete document All of the words

AND The complete document All of the words

Published between	1920	and	2006
Information types	<input checked="" type="checkbox"/> All <input type="checkbox"/> Abstracts <input type="checkbox"/> Articles <input type="checkbox"/> Books <input type="checkbox"/> Company homepages <input type="checkbox"/> Conferences <input type="checkbox"/> Patents <input type="checkbox"/> Preprints <input type="checkbox"/> Scientist homepages		
File formats	<input checked="" type="checkbox"/> All <input type="checkbox"/> HTML <input type="checkbox"/> PDF		
Content sources	<input checked="" type="checkbox"/> All Journal sources <input type="checkbox"/> BioMed Central <input type="checkbox"/> Crystallography Journals Online <input type="checkbox"/> MEDLINE/PubMed <input type="checkbox"/> Project Euclid <input type="checkbox"/> ScienceDirect <input type="checkbox"/> Scitation <input type="checkbox"/> Society for Industrial & App. Mathematics <input type="checkbox"/> PubMed Central <input checked="" type="checkbox"/> All Web sources <input type="checkbox"/> E-Print ArXiv <input type="checkbox"/> CogPrints <input type="checkbox"/> NaGA <input type="checkbox"/> Patent Offices <input type="checkbox"/> RePEc <input type="checkbox"/> OJS <input type="checkbox"/> Umeo, Toronto, T-Spain <input type="checkbox"/> MIT OpenCourseWare <input type="checkbox"/> NDLTD <input type="checkbox"/> Catech <input type="checkbox"/> Digital Archives <input type="checkbox"/> Other		
Subject areas	<input checked="" type="checkbox"/> All <input type="checkbox"/> Life Sciences		

Abb. 2: Ausschnitt aus der erweiterten Suche in Scirus

Scirus setzt u.a. auch automatische Verfahren für die Klassifizierung der Informationen ein (sowohl linguistische Verfahren als auch Methoden zur Strukturanalysen z.B. für die Informationstypen). Zudem werden aus den Dokumenten Schlüsselwörter ermittelt, die für eine Verfeinerung der Suchanfrage benutzt werden können.

Obwohl Scirus mehr als ein Prototyp ist, ist die Qualität der Klassifikationen nicht immer überzeugend.

Aus dem hier Gesagten wird klar, dass die Entwicklung fachspezifischer Suchmaschinen eine anspruchsvolle Aufgabe ist - auch in wissenschaftlicher Hinsicht. Die wissenschaftliche Community ist gefordert, sich in der Entwicklung geeigneter fachspezifischer Suchmaschinen zu engagieren.

2.2 Betrieb fachspezifischer Suchmaschinen

Ob die wissenschaftliche Community den permanenten Betrieb von fachspezifischen Suchmaschinen übernimmt, muss diskutiert werden.

Pros

- Die heutigen universellen Suchmaschinen wie Google haben Archive aufgebaut, in denen die Dokumente gespeichert werden. Damit sind die Suchmaschinen in der Lage, den Zugang zu Informationen zu kontrollieren. Die wissenschaftliche Community gerät so in ein Abhängigkeitsverhältnis von den Betreibern der universellen Suchmaschinen.
- Die wissenschaftliche Community hat andererseits keinen Einfluss auf die Geschäftsmodelle der Suchmaschinenbetreiber.
- Suchmaschinen unter eigener Kontrolle können leichter an neue Erfordernisse und Entwicklungen angepasst werden.

Cons

- Der mit der Pflege und dem Betrieb verbundene Aufwand ist mit den derzeitigen Strukturen durch die wissenschaftliche Community nicht zu leisten. Eventuell könnte das von wissenschaftlichen Infrastruktureinrichtungen wie Bibliotheken oder in Deutschland den Fachinformationszentren übernommen werden.
- Ein offenes Problem ist auch die Zusammenarbeit von universellen und fachspezifischen Suchmaschinen. Der Nutzer wird fachspezifische Suchmaschinen nur akzeptieren, wenn für ihn die Suche bessere und vollständigere Ergebnisse liefert und diese möglichst automatisch in die Ergebnisse der universellen Suchmaschinen integriert werden. Im Moment sind allerdings für die universellen Suchmaschinenbetreiber die fachspezifischen Suchmaschinenbetreiber als Kooperationspartner noch nicht interessant.

- Die wissenschaftliche Community verfügt derzeit nicht über die Expertise und technische Kompetenz und Ausstattung, größere Suchmaschinen stabil und zuverlässig zu betreiben.

3 Fazit

Fachspezifische Suchmaschinen können ein wichtiges Werkzeug für die Suche nach wissenschaftlichen Informationen im Web sein, indem sie wissenschaftliche Informationen einfacher und effizienter zugänglich machen. Der Aufbau leistungsfähiger fachspezifischer Suchmaschinen ist jedoch aufwändig und erfordert noch viel Entwicklungsarbeit speziell bei der automatischen Aufbereitung und Verarbeitung der Inhalte der Informationen. Die wissenschaftliche Community sollte sich in die Entwicklung fachspezifischer Suchmaschinen einbringen.

Literaturangaben

Brin, Sergey; Page, Lawrence (1998): The Anatomy of a Large-Scale Hypertextual Web Search Engine,

<http://www-db.stanford.edu/~backrub/google.html>

Google - Mission Statement (2005):

<http://www.google.com/corporate/index.html>

Pant, Gautam; Srinivasan, Padmini, and Menczer, Filippo (2004): Crawling the Web, S. 153 – 177 in Levenc, Mark; Poulouvassilis, Alexandra: Web Dynamics, Springer 2004 ISBN 3-540-40676-X

McCallum, Andrew; Nigam, Kamal; Rennie, Jason; Seymore, Kristie (1999): A Machine Learning Approach to Building Domain-Specific Search Engines,

http://www.ri.cmu.edu/pub_files/pub1/mccallum_andrew_1999_2/mccallum_andrew_1999_2.pdf

Zentralblatt Math: <http://www.emis.de/ZMATH/>

Math Reviews: <http://www.ams.org/mr-database>

MSC 2000: <http://www.google.com/corporate/index.html>

GAMS: <http://gams.nist.gov/>

XML: <http://www.w3.org/XML/>

MathML: <http://www.w3.org/Math/>

OpenMath: <http://www.w3.org/Math/>

OMDoc: <http://www.openmath.org/cocoon/openmath/index.html>

Srinivasan, Padmini; Mitchell, Joyce; Bodenreider, Olivier; Pant, Gautam; Menczer, Filippo (2002): Web Crawling Agents for Retrieving Biomedical Information, <http://dollar.biz.uiowa.edu/~pant/Papers/bixmas.pdf>

Bekkermann, Ron; McCallum, Andrew, Huang, Gary (2005): Automatic Categorization of Email into Folders: Benchmark Experiments on Enron and SRI Corpora, <http://www.cs.umass.edu/~mccallum/papers/foldering-tr05.pdf>

Scirus: <http://www.scirus.com>

ArXiv: <http://www.arxiv.org/>

Angaben zum Autor

Dr. Wolfram Sperber

Konrad-Zuse-Zentrum für Informationstechnik Berlin

Takustr. 7

14195 Berlin

Studium der Mathematik, im Anschluss Promotion, Arbeit als wissenschaftlicher Mitarbeiter, seit 1992 in verschiedenen Projekten zum Aufbau elektronischer Informations- und Kommunikationssysteme in den Wissenschaften tätig, u.A. in der Math-Net Initiative und im Math&Industry Projekt.

Gegenwärtige Arbeitsschwerpunkte: Suchmaschinen, automatische Klassifikation.

Google Scholar - wie tief gräbt diese Suchmaschine?

Philipp Mayr, Anne-Kathrin Walter

Informationszentrum Sozialwissenschaften, Bonn

Zusammenfassung

Der Beitrag widmet sich dem neuen Google-Suchdienst Google Scholar. Die Suchmaschine, die ausschließlich wissenschaftliche Dokumente durchsuchen soll, wird mit ihren wichtigsten Funktionen beschrieben und anschließend einem empirischen Test unterzogen. Die durchgeführte Studie basiert auf drei Zeitschriftenlisten (STM-Zeitschriften von Thomson Scientific, Open Access-Zeitschriften des Verzeichnisses DOAJ und in der Fachdatenbank SOLIS ausgewerteten sozialwissenschaftliche Zeitschriften), deren Abdeckung durch Google Scholar überprüft wurde. Die Studie zeigt Defizite in der Abdeckung und Aktualität des Google Scholar Index. Weiterhin macht die Studie deutlich, wer die wichtigsten Datenlieferanten für den neuen Suchdienst sind und welche wissenschaftlichen Informationsquellen im Index repräsentiert sind. Die Pluspunkte von Google Scholar liegen in seiner Einfachheit, seiner Suchgeschwindigkeit und letztendlich seiner Kostenfreiheit. Die Recherche in Fachdatenbanken kann Google Scholar trotz sichtbarer Potenziale (z. B. Zitationsanalyse) aber heute aufgrund mangelnder fachlicher Abdeckung und Transparenz nicht ersetzen.

1 Einleitung

Der Start des neuen Google-Dienstes Google Scholar¹ hat kurz nach seiner Veröffentlichung im November 2004² wie gewohnt ein großes Medienecho nach sich gezogen. Sowohl in der allgemeinen Presse (Markoff 2004, Terdiman 2004, Frankfurter Allgemeine Zeitung vom 22.11.2004) als auch unter Wissenschaftlern, Fachverlagen und Wissenschaftsgesellschaften hat Google Scholar insbesondere wegen der Nähe zu den viel diskutierten Themen Open Access und Invisible Web für großes Aufsehen gesorgt (Asbrand 2004, Banks 2004, Kennedy & Price 2004, Payne 2004, Sullivan 2004). Das Besondere an Google Scholar ist neben der zugrunde liegenden Technologie sicherlich seine Bemü-

1 Siehe <http://scholar.google.com/>

2 Siehe <http://googleblog.blogspot.com/2004/10/scholarly-pursuits.html>

hung zur Beschränkung auf wissenschaftliche Information. Google Scholar gibt dazu Folgendes auf seinen Seiten an:

„Google Scholar enables you to search specifically for scholarly literature, including peer-reviewed papers, theses, books, preprints, abstracts and technical reports from all broad areas of research. Use Google Scholar to find articles from a wide variety of academic publishers, professional societies, preprint repositories and universities, as well as scholarly articles available across the web.“³

Damit positioniert Google seinen neuen Service in einem strategisch sehr interessanten und wachsenden Bereich.

Zur Ausgangssituation: Weltweit arbeiten immer mehr Wissenschaftler, die in einer steigenden Zahl von Fachzeitschriften immer häufiger publizieren. *Umstätter* spricht in diesem Zusammenhang von der Wissenschaftsgesellschaft (Umstätter 1998, 303ff.). Als Folge dieser Wissenschaftsgesellschaft lässt sich seit Jahrzehnten ein rasant ansteigender wissenschaftlicher Output, mit einhergehender Beschleunigung der wissenschaftlichen Kommunikation, messen. Dieser Prozess wird durch den gerade stattfindenden Paradigmenwechsel im Publikationswesen (von der Print- zur Online-Publikation) zusätzlich beschleunigt und verkompliziert. Für die Wissenschaftler bedeutet das zunehmende Schwierigkeiten bei der Auswahl und Beschaffung relevanter Literaturquellen für ihre Arbeit. *Krause* beschreibt z.B. typische Probleme und Lösungsansätze der polyzentrischen Informationsversorgung (Krause 2003). Wissenschaftler sind demnach heute gezwungen, bzgl. der inhaltlichen Erschließung hochgradig heterogene Dokumentenräume (verteilte Digitale Bibliotheken) zu durchsuchen. Häufig stehen ihnen für ihre umfassenden Recherchen kaum semantisch integrierte Angebote zur Verfügung.

Allem Anschein nach will Google diese wissenschaftlich relevanten Dokumentenräume mit seinem neuen Suchdienst Google Scholar automatisch erschließen. Da Google über die Reichweite, Aktualität und Abdeckung von Google Scholar keine Informationen bereithält, soll mit dieser empirischen Studie untersucht werden, wie tief Google Scholar sich in das wissenschaftliche Web vorgearbeitet hat. Wir haben dazu den Umfang des Services anhand der Abdeckung unterschiedlicher Zeitschriftenlisten gemessen. Weiterhin wurde untersucht, welche Typen von Nachweisen und welche Webserver sich in den analysierten Trefferdaten befinden.

Der Beitrag beschreibt zunächst die Funktionsweise und Besonderheiten von Google Scholar. Im zweiten Teil gehen wir auf die Ergebnisse der Google Scholar-Studie (April 2005) ein und fassen unsere Beobachtungen zu diesem neuen Service knapp zusammen.

3 Siehe <http://scholar.google.com/scholar/about.html>

2 Google Scholar

Das Pilotprojekt CrossRef Search⁴ kann als Test und Vorläufer von Google Scholar angesehen werden. Google hat bei CrossRef Search Volltext-Bestände einer größeren Zahl von Fachverlagen (z.B. Blackwell, Nature Publishing Group, Springer-Verlag usw.) und Wissenschaftsgesellschaften (z.B. ACM, IEEE, usw.) indexiert und über eine typische Google-Oberfläche bereitgestellt. Die CrossRef-Suche wird bei den einzelnen CrossRef-Partnern nach wie vor angeboten⁵.

Ähnlich vom Ansatz, aber viel breiter und unspezifischer im Scope ist die „wissenschaftliche Suchmaschine“ Scirus⁶, die laut eigenen Angaben 200 Millionen „science-specific Web pages“ durchsucht. Unter diesen Webseiten befinden sich viele frei zugängliche Dokumente auf universitären Webservern, auf denen z.B. auch Studenten ihre Dokumente ablegen, die aber nicht unbedingt wissenschaftlichen Ansprüchen genügen. Für eine Recherche nach wissenschaftlich geprüfter Information (z.B. durch das Peer Review) ist diese Tatsache oft ein Ausschlusskriterium für ein Suchsystem.

Wie sich am Pilotprojekt CrossRef Search ablesen lässt, hat Google Scholar über die Kooperation mit wissenschaftlichen Verlagen einen anderen Ansatz gewählt. Was ist interessant am Google Scholar-Ansatz? An vorderster Stelle ist sicherlich die bereits erwähnte Beschränkung auf nachweislich wissenschaftliche Dokumente zu nennen, die bislang von keiner Internetsuchmaschine konsequent umgesetzt werden konnte. Google Scholar selbst ist zunächst ein kostenfreier Service. Allerdings befinden sich viele Inhalte, die über Google Scholar nachgewiesen werden, auf Verlagsservern, auf denen der Volltext-Abruf kostenpflichtig wird. Die Abstracts der Dokumente werden dem Recherchierenden aber mindestens angezeigt. Der Google-Ansatz beinhaltet weiterhin Dokumente aus dem stetig wachsenden Open Access und Self Archiving-Bereich (Swan & Brown 2005).

Für den Nutzer sind neben dem direkten Volltextzugang aber unter Umständen die von Google implementierten Analysen und darauf aufbauend das Dokumentenranking interessant. Google Scholar's Relevanzranking basiert laut eigenen Angaben auf unterschiedlichen Kriterien (siehe Zitat unten). Insbesondere die automatische Zitationsextraktion und -analyse, auch Autonomous Citation Indexing (ACI) genannt (Lawrence, Giles & Bollock 1999), kann für den Nutzer Hilfen bei der Informationssuche und -beschaffung bringen. Hochzitierte Arbeiten werden nach diesem Verfahren oben in die Ergebnisliste gerankt und sind für Recherchierende damit gut sichtbar. Das automatische Verfahren ACI

4 Siehe <http://www.crossref.org/crossrefsearch.html>

5 Siehe z.B. die CrossRef Suche bei Nature
http://www.nature.com/search/search_crossref.html

6 Siehe <http://www.scirus.com>

setzt allerdings voraus, dass die Literaturangaben der analysierten Dokumente zur Verfügung stehen, was bei den Volltexten per se gegeben ist. Google Scholar kann damit über die Referenzen analysierter Dokumente hinaus auch Literaturquellen nachweisen, die nicht auf den indexierten Webservern liegen.

Weiterhin ist an Google Scholar interessant, dass diese Suchmaschine interdisziplinär konzipiert ist. Im Gegensatz zu Spezialsuchmaschinen wie z.B. dem CiteSeer-System⁷, das ausschließlich wissenschaftliche Informatikliteratur indexiert, wäre mit dem Google Scholar-Ansatz eine umfassende Wissenschaftsuchmaschine für alle Disziplinen denkbar.

Nachfolgend werden die wichtigsten Features von Google Scholar knapp dargestellt.

- *Erweiterte Suche*: die erweiterte Suche von Google Scholar bietet neben der Suche im Titel eines Dokuments die Möglichkeit, nach Autorennamen, einem Zeitschriftentitel und dem Publikationsjahr eines Artikels oder Buches zu recherchieren. Diese Attribute stellen für wissenschaftliche Fachrecherchen nur ein Minimalset an Suchkriterien dar (vgl. Abfragemöglichkeiten von Literatur- und Fachdatenbanken); für ein automatisches System bereitet die zuverlässige Extraktion dieser Daten aus z. T. un- oder teilstrukturierten Dokumenten jedoch große Schwierigkeiten (vgl. Lawrence, Giles & Bollacker 1999).
- *Volltextzugang*: im Gegensatz zu den klassischen Nachweis- bzw. Referenzdatenbanken, die in den bibliografischen Angaben einschließlich Abstract und Schlagwörtern suchen, basiert die Google Scholar-Suche auf einem Volltextindex. D.h., dass der Nutzer mit kleineren technischen Einschränkungen (Price 2004)⁸ und allen Vor- und Nachteilen dieses Recherchetyps direkt in den Volltexten der Dokumente recherchiert und idealerweise sofort auf den Volltext zugreifen kann.
- *Relevanzranking*: Google gibt dazu an: „Just as with Google Web Search, Google Scholar orders your search results by how relevant they are to your query, so the most useful references should appear at the top of the page. This relevance ranking takes into account the full text of each article as well as the article’s author, the publication in which the article appeared and how often it has been cited in scholarly literature. Google Scholar also automatically analyzes and extracts citations and presents them as separate results, even if the documents they refer to are not online. This means your search results may include citations of older works and seminal articles that appear only in books or other offline publications.“⁹

7 Siehe <http://citeseer.ist.psu.edu/>

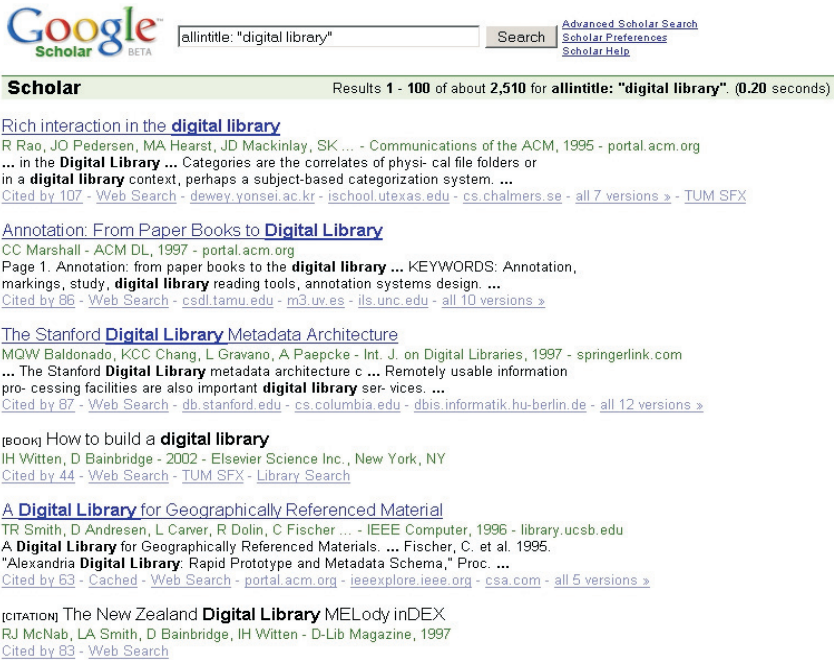
8 Siehe zu den Einschränkungen <http://blog.searchenginewatch.com/blog/041201-105511>

9 Siehe <http://scholar.google.com/scholar/about.html>

- *Web Search*: Die Verknüpfung zum Google-Gesamtindex bietet insbesondere dann eine Hilfestellung, wenn die Dokumente nicht direkt über die Google Scholar-Trefferliste verfügbar sind und über die Standard-Websuche die Anfrage auf das „gesamte“ Web ausgeweitet wird.
- *Institutional Access*: Das Pilotprojekt „Institutional Access“ bietet hauptsächlich für institutionelle Benutzer (z.B. Studenten und Hochschulmitarbeiter) Mehrwerte, da Google die elektronischen Bestandsnachweise der Bibliotheken über Linkresolver wie SFX nutzt.
- *Weitere Features*: Google Scholar bietet weitere interessante Features wie z. B. die Funktion *Library Search*, die eine Anfrage an den OCLC WorldCat¹⁰ weiterleitet und lokale Bibliotheken ausgibt, die z.B. ein gewünschtes Buch nachweisen. Zusätzlich werden alternative Fundstellen eines Dokuments im Web ausgewiesen (siehe Abb. 1 *versions*).

Abbildung 1 zeigt eine typische Google Scholar-Trefferliste. Auf die einzelnen Bestandteile eines Treffers wird zu einem späteren Zeitpunkt noch intensiver eingegangen (siehe Kapitel 3.1). Vorab soll nur darauf hingewiesen werden, dass sich die Treffer, die Google Scholar liefert, bzgl. der Verfügbarkeit unterscheiden. So sind zwei Treffer in Abbildung 1 (siehe Kennzeichnung BOOK und CITATION) nicht über einen Hyperlink erreichbar, sondern wurden lediglich aus indextierten Dokumenten extrahiert.

10 Siehe <http://www.oclc.org/worldcat/>



The screenshot shows a Google Scholar search interface. At the top left is the Google Scholar logo. A search bar contains the text "allintitle: \"digital library\"". To the right of the search bar are buttons for "Search", "Advanced Scholar Search", "Scholar Preferences", and "Scholar Help". Below the search bar, a green bar indicates "Scholar" and "Results 1 - 100 of about 2,510 for allintitle: \"digital library\". (0.20 seconds)".

The first search result is titled "Rich interaction in the [digital library](#)". The author information is "R Rao, JO Pedersen, MA Hearst, JD Mackinlay, SK... - Communications of the ACM, 1995 - portal.acm.org". The abstract snippet reads: "... in the **Digital Library** ... Categories are the correlates of physical file folders or in a **digital library** context, perhaps a subject-based categorization system. ...". It is cited by 107 and includes links for "Web Search", "dewey.yonsei.ac.kr", "ischool.utexas.edu", "cs.chalmers.se", "all 7 versions", and "TUM SFX".

The second result is titled "Annotation: From Paper Books to [Digital Library](#)". The author is "CC Marshall - ACM DL, 1997 - portal.acm.org". The snippet says: "Page 1. Annotation: from paper books to the **digital library** ... KEYWORDS: Annotation, markings, study, **digital library** reading tools, annotation systems design. ...". It is cited by 66 and includes links for "Web Search", "csdl.tamu.edu", "m3.uv.es", "ils.unc.edu", and "all 10 versions".

The third result is titled "The Stanford [Digital Library](#) Metadata Architecture". The authors are "MQW Baldonado, KCC Chang, L Gravano, A Paepcke - Int. J. on Digital Libraries, 1997 - springerlink.com". The snippet reads: "... The Stanford **Digital Library** metadata architecture c ... Remotely usable information processing facilities are also important **digital library** services. ...". It is cited by 87 and includes links for "Web Search", "db.stanford.edu", "cs.columbia.edu", "dbis.informatik.hu-berlin.de", and "all 12 versions".

The fourth result is titled "[BOOK] How to build a [digital library](#)". The author is "IH Witten, D Bainbridge - 2002 - Elsevier Science Inc., New York, NY". It is cited by 44 and includes links for "Web Search", "TUM SFX", and "Library Search".

The fifth result is titled "A [Digital Library](#) for Geographically Referenced Material". The authors are "TR Smith, D Andresen, L Carver, R Dolin, C Fischer ... - IEEE Computer, 1996 - library.ucsb.edu". The snippet says: "... The Stanford **Digital Library** metadata architecture c ... Remotely usable information processing facilities are also important **digital library** services. ...". It is cited by 63 and includes links for "Cached", "Web Search", "portal.acm.org", "ieeexplore.ieee.org", "csa.com", and "all 5 versions".

The sixth result is titled "[CITATION] The New Zealand [Digital Library](#) MELody inDEX". The authors are "RJ McNab, LA Smith, D Bainbridge, IH Witten - D-Lib Magazine, 1997". It is cited by 63 and includes a link for "Web Search".

Abb. 1: Typische Google Scholar-Ergebnisliste. Gesucht wurde die Phrase „digital library“ im Titel.

3 Google Scholar-Studie

Wie an anderer Stelle bereits mehrfach kritisiert (Lewandowski 2004, Jacso 2004, Jacso 2005a, Jacso 2005b), ist über die eigentliche Größe und Abdeckung von Google Scholar sehr wenig bekannt. Auch Fragen, wie häufig und ob der Index der Suchmaschine aktualisiert wird, können aus öffentlich zugänglichen Informationsquellen praktisch nicht beantwortet werden. Aus diesem Grund wollen wir mit der Studie ein etwas genaueres Bild der aktuellen Situation zeichnen.

Die folgende Untersuchung konzentriert sich auf die Fragestellung: Wie tief gräbt Google Scholar? Die Studie soll Aussagen zu folgenden Fragen ermöglichen:

- Wie groß ist die Abdeckung unterschiedlicher wissenschaftlicher Zeitschriften in Google Scholar? Die Studie testet über die Abfrage von unterschiedlichen Zeitschriftenlisten, ob Google die Zeitschriften indexiert hat und Arti-

kel aus diesen Zeitschriften nachweisen kann. Die Zeitschriftenlisten kommen aus drei sehr unterschiedlichen Bereichen Science, Technology, Medicine (STM), Open Access und Sozialwissenschaften, und ermöglichen Rückschlüsse zu den thematischen Schwerpunkten des aktuellen Google Scholar-Angebots.

- Welche Dokument- bzw. Treffertypen sind in Google Scholar enthalten? Die analysierten Trefferdaten geben Hinweise über die Zusammensetzung der Ergebnisse aus den Treffertypen Volltextlink, Nachweislink und Zitationsnachweis.
- Von welchen Anbietern kommen die meisten Dokumente? Die Studie soll deutlich machen, wer die größten Datenlieferanten für den neuen Suchdienst sind und welche wissenschaftlichen Informationsquellen aktuell im Index unterrepräsentiert sind. Die Verteilung der Webserver bzw. Anbieter ist interessant, weil sich daraus schließen lässt, ob Google Scholar eher kostenpflichtige Dokumente oder frei zugängliche erschließt.
- Wird der Google Scholar-Index regelmäßig aktualisiert?

3.1 Methodik

Im Zeitraum April/Mai 2005 wurden drei Zeitschriftenlisten abgefragt und die zurück gelieferten Daten analysiert. Zeitschriften stellen in den meisten Fachdisziplinen die wichtigsten Publikationsorgane und Orte der wissenschaftlichen Fachdiskussion dar. Zudem sind sie gut prozessierbar und man erhält trotz einer relativ geringen Anfragemenge eine repräsentative und auswertbare Menge an Treffern.

Da wir nicht alle existierenden Zeitschriften abfragen konnten, haben wir folgende öffentlich zugängliche Zeitschriftenlisten als Grundlage der Untersuchung gewählt:

1. Zeitschriftenliste von *Thomson Scientific (ISI)*¹¹. Bei dieser Liste handelt es sich vorrangig um internationale Science Technology Medicine Journals (STM). Für die Untersuchung konnten 10.645 Zeitschriftentitel berücksichtigt werden.
2. Frei zugängliche elektronische Zeitschriften des *Directory of Open Access Journals (DOAJ)*¹². Diese Liste umfasste insgesamt 1.415 internationale Open Access Journals aus allen Wissenschaftsbereichen.

11 Masterliste des ISI siehe <http://www.isinet.com/cgi-bin/jrnlst/jlresults.cgi?PC=MASTER>

12 DOAJ siehe <http://www.doaj.org/>

3. *Zeitschriften der Datenbank SOLIS (IZ)*¹³. Diese Liste umfasst insgesamt 317 hauptsächlich deutschsprachige Zeitschriften aus unterschiedlichen Fachgebieten der Soziologie und angrenzenden Bereichen.

Die drei Zeitschriftenlisten decken jeweils einen sehr unterschiedlichen Bereich ab und können daher inhaltlich und vom Umfang her nicht direkt miteinander verglichen werden. Sie sollen vielmehr Aufschluss darüber geben, welche wissenschaftlichen Disziplinen von Google Scholar in welcher Form und Tiefe nachgewiesen werden. Erwähnt werden soll, dass die drei untersuchten Zeitschriftenlisten lediglich einen kleinen Teil der erscheinenden Zeitschriften widerspiegeln. Die Elektronische Zeitschriftenbibliothek in Regensburg¹⁴ weist beispielsweise über 22.800 Zeitschriftentitel nach, davon sind mehr als 2.650 reine Online-Zeitschriften.¹⁵ Harnad et al. gehen von etwa 24.000 peer reviewed Journals (Harnad et al. 2004) aus. Andere Schätzungen gehen sogar von über 100.000 periodisch erscheinenden Publikationen aus (Ewert & Umstätter 1997).

Die Untersuchung gliedert sich in folgende Schritte:

Schritt 1: Abfrage der Zeitschriftentitel der drei Zeitschriftenlisten: Um die Abdeckung von Google Scholar zu ermitteln, wurden die oben genannten Zeitschriftenlisten Ende April 2005 abgefragt. Die erweiterte Suche bietet hierfür das Suchfeld „Return articles published in“. Die Untersuchung beschränkte sich auf die ersten 100 Treffer pro Zeitschrift.

Schritt 2: Speicherung der Google Scholar-Ergebnisseiten: Es wurden für jeden abgefragten Zeitschriftentitel maximal 100 Treffer (Records) zur weiteren Bearbeitung lokal abgespeichert.

Schritt 3: Extraktion der Daten aus den Ergebnisseiten: Datenbasis der Untersuchung waren die einzelnen Records der Ergebnisseiten. Um die Vorgehensweise bei der Analyse zu verdeutlichen, wird im Folgenden kurz der Aufbau typischer Google Scholar-Treffer beschrieben.

13 Siehe Liste der Datenbank SOLIS (Sozialwissenschaftliches Literaturinformationssystem) <http://www.gesis.org/Information/Zeitschriften/index.htm>

14 Siehe <http://www.bibliothek.uni-regensburg.de/ezeit/>

15 Siehe <http://rzblx1.uni-regensburg.de/ezeit/about.phtml>

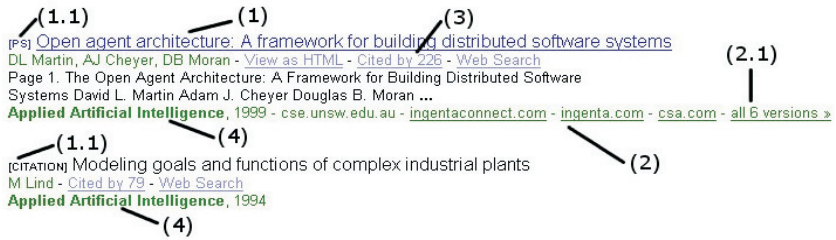


Abb. 2: Zwei typische Records einer Google Scholar-Trefferliste

(1) Titel des Nachweises und Dokumenttyp

Neben der Relevanz eines Nachweises interessiert einen Nutzer vor allem auch die Verfügbarkeit. Im besten Fall wird er direkt zum Volltext weitergeleitet, im ungünstigsten Fall bekommt er nur die Zitation angezeigt mit der Möglichkeit zur Suche in Google Web Search. Die erste Zeile eines Records bestimmt die Art des Nachweises. Dabei werden bestimmte Dokumenttypen durch eine Kennzeichnung in eckigen Klammern vor dem eigentlichen Titel des Nachweises kenntlich gemacht.

- Direkter Link zum Volltext im Postscript- oder PDF-Format: Handelt es sich bei einem Record um den Nachweis eines Volltexts im Postscript- oder PDF-Format, wird „PDF“, bzw. „PS“ in eckigen Klammern dem Treffer vorangestellt (1.1 in Abb. 2). Bei Links zu PDF-Dateien trifft dies nicht immer zu, daher wurde in diesem Fall auch die Endung des Links berücksichtigt.
- Normale Nachweise: Die meisten Treffer sind Links, die zunächst zum bibliografischen Nachweis des Dokuments führen. Dieser Nachweis sollte laut Google Scholar mindestens einen Abstract enthalten.
- Zitationen: Viele Zeitschriftenartikel führt Google Scholar nur als Zitation auf. Diese Treffer sind dadurch gekennzeichnet, dass dem Treffer „CITATION“ vorangestellt ist (1.1 in Abb. 2) und dass der Name des Treffers nicht mit einem Link unterlegt ist.
- Bücher: Google Scholar weist auch Bücher nach, die durch das Kürzel „BOOK“ gekennzeichnet sind. Da in dieser Untersuchung nur die Nachweise von Zeitschriften interessieren, werden sie nicht weiter beachtet.

(2) Domains

Neben dem Hauptlink, mit dem der Titel unterlegt ist (siehe Abb 2., (1)), werden Links zu weiteren Servern gegeben, die den Artikel vorhalten. Angezeigt wird dabei nicht die gesamte Adresse, sondern nur die Domains. Auch diese wurden

ausgewertet, um ein Ranking der Server zu erstellen. Gibt es mehrere Quellen, erreicht man diese durch Anklicken des Links „all x versions“ (siehe Abb. 2, (2.1)). Diese Links wurden für die Auswertung allerdings nicht verfolgt.

(3) Zitationszahlen

Google Scholar baut das Ranking der Nachweise unter anderem auf der Anzahl der Zitationen eines Artikels auf. Diese werden ebenfalls angezeigt (siehe Abb. 2, (3)), wurden aber für die Untersuchung nicht weiter ausgewertet.

(4) Zeitschriftentitel

Da Google Scholar nur eingeschränkt Phrasensuche unterstützt, werden auch Zeitschriften durchsucht, die die Suchterme nicht zwingend als Phrase beinhalten. Daher wurden die Records bei der Auswertung einzeln überprüft und nur als Treffer gezählt, wenn der genaue Titel (siehe Abb. 2 (4)) gefunden wurde.

Schritt 4: Analyse und Aggregation der extrahierten Daten: Schwierigkeiten bei der Analyse traten bei der Überprüfung der Zeitschriftentitel auf. Beispielsweise gibt Google Scholar bei der Suche nach Artikeln der Zeitschrift „Applied Intelligence“ auch Treffer der Zeitschrift „Applied Artificial Intelligence“ aus. Ein weiteres Problem stellt die uneinheitliche Darstellung der Titel dar, die auf die automatische Zitationsextraktion zurückzuführen ist: zum Beispiel werden Artikel der Zeitschrift „Analyse und Kritik“ auch unter dem Titel „Analyse and Kritik“ oder „Analyse & Kritik“ aufgeführt. Die aus den Trefferlisten extrahierten Daten wurden über einfache Auszählungen aggregiert. Zunächst haben wir die Zeitschriften ausgezählt, deren Titel eindeutig erkannt oder nicht erkannt wurden (siehe Tab. 1). Die Treffer, die eindeutig einer Zeitschrift zugeordnet werden konnten, wurden den vier unterschiedlichen Dokumenttypen zugewiesen und ausgezählt (siehe Abb. 3). Für jeden Treffer, der einer Zeitschrift zugeordnet werden konnte, wurden anschließend alle Domains (Webserver) extrahiert und die Häufigkeit der einzelnen Webserver pro Zeitschriftenlisten bestimmt.

3.2 Ergebnisse

Die Ergebnisse der Untersuchung teilen sich in zwei Bereiche. Zum einen Ergebnisse, die sich ausschließlich auf die analysierten Trefferlisten beziehen, zum anderen Ergebnisse, die auf stichprobenartigen Tests basieren und keine repräsentativen Aussagen zulassen.

1. Identifikation der Zeitschriften

Als erstes haben wir geprüft, wie viele Zeitschriftentitel der jeweiligen Listen sich in den Trefferdaten von Google Scholar identifizieren lassen. Als „gefundene Titel“ werden nur die Zeitschriftentitel gewertet, die eindeutig in den zurück gelieferten Daten identifiziert werden konnten. Alle nicht eindeutig identifizierten Titel, wie die oben (Schritt 4) genannten Beispiele, werden als nicht gefundene Titel gewertet. Titel, die keine Treffer in Google Scholar generieren, sind ebenfalls in der Spalte „nicht gefundene Titel“ enthalten.

Tab. 1: Identifikation der Zeitschriftentitel in den Google Scholar-Daten

Liste	Titel	gefundene Titel	nicht gefundene Titel
IZ (SOLIS)	317	228 (72%)	89 (28%)
DOAJ	1.415	1.078 (76%)	337 (24%)
ISI	10.645	8.931 (84%)	1.714 (16%)

Tabelle 1 zeigt, dass der Großteil der angefragten Zeitschriftentitel der drei Listen (IZ, DOAJ, ISI) in den zurück gelieferten Google Scholar-Daten identifiziert werden kann (siehe Spalte „gefundene Titel“) und damit Artikel der jeweiligen Zeitschriften nachgewiesen werden können. Die genaue Anzahl der Artikel einer Zeitschrift wurde nicht bestimmt, da uns max. 100 Treffer pro Zeitschrift zur Analyse zur Verfügung standen. Von den 317 Zeitschriften der IZ-Zeitschriftenliste (SOLIS) können beispielsweise 228 Titel (ca. 72 % der Liste) eindeutig identifiziert werden („gefundene Titel“). Bei 89 Zeitschriftentiteln (ca. 28 % der Liste) lässt sich der Titelstring der Zeitschrift nicht eindeutig identifizieren („nicht gefundene Titel“) oder es werden keine Treffer geliefert. Dies trifft auf 20 Zeitschriften bzw. etwa 6 % der Zeitschriften der IZ-Liste zu. Auffällig sind die relativ hohen Werte (zwischen 72 % und 84 %) der gefundenen Zeitschriftentitel für alle drei Listen. Überraschenderweise werden 337 der frei zugänglichen Open Access-Zeitschriften (24 % der DOAJ-Liste) in Google Scholar nicht gefunden. Die hauptsächlich englischsprachigen STM-Journals der ISI-Liste haben prozentual mit 84 % die beste Abdeckung.

2. Verteilung der Dokumenttypen

Als nächstes haben wir die von Google Scholar zurück gelieferten Daten bzgl. der Zugehörigkeit zu einem Dokumenttyp analysiert. Insgesamt wurden über 601.000 Google Scholar-Treffer analysiert. Die Google Scholar-Treffer lassen sich in vier Typen einordnen (siehe Beschreibung zu Link, Citation, PDF, PS in Kapitel 3.1). Die Verteilung der Dokumenttypen (siehe Abb. 3) steht in engem Zusammenhang mit den in 1. aufgeführten Ergebnissen. Der hohe Anteil der gefundenen Zeitschriften spiegelt sich in einem sehr hohen Anteil des Dokument-

typs *Citation* wider. Dieser Typ, der von Google als „offline-Nachweis“ bezeichnet wird, kann nicht als klassischer Literaturnachweis beschrieben werden, da er lediglich auf aus anderen Dokumenten extrahierten Referenzen basiert und nur minimale bibliografische Informationen bietet. *Citation* macht in den analysierten Daten über alle drei Listen mit 44 % den größten Anteil aus. Der Dokumenttyp *Link*, ein umfangreicherer Literaturnachweis mit Abstract, macht einen Anteil von 43 % aus. Die Nachweise mit direktem Zugriff auf den Volltext im Format PDF oder PS sind mit 12 % (PDF) bzw. 1 % (PS) deutlich seltener vertreten.

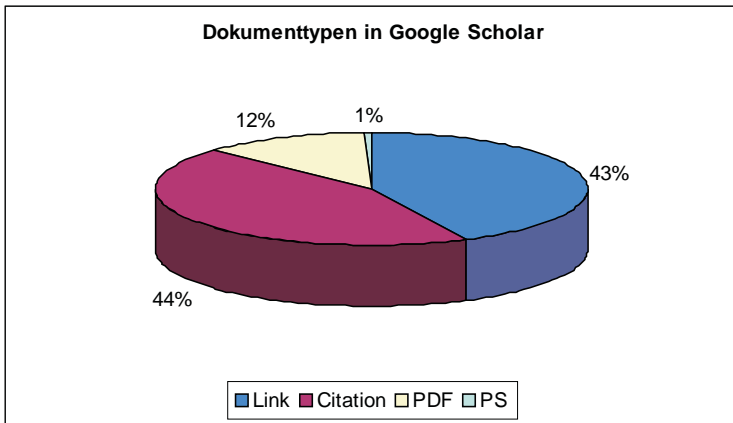


Abb. 3: Verteilung der Dokumenttypen in analysierten Google Scholar-Trefferlisten

In Tabelle 2 werden die Werte der Dokumenttypen der Trefferanalyse für alle drei Zeitschriftenlisten aufgeführt.

Tab. 2: Verteilung der Dokumenttypen über die drei abgefragten Listen

Liste	Link %	Citation %	PDF %	PS %
IZ (SOLIS)	1,32	92,95	5,73	0,00
DOAJ	37,72	39,94	21,46	0,88
ISI	43,88	43,70	11,91	0,51

Auffällig ist, dass die Zeitschriften der Datenbank SOLIS zum überwiegenden Teil Zitationsangaben (siehe 92,95 % beim Dokumenttyp *Citation*) generieren. Der Grund dafür ist, dass Google Scholar diese meist deutschsprachigen Aufsätze auf den indexierten Webservern nicht direkt nachweisen kann und folglich nur die Referenzen indexierter Dokumente ausgibt. Der Anteil der Zitations-

nachweise bei den beiden internationalen Zeitschriftenlisten (DOAJ und ISI) ist zwar deutlich niedriger, aber dennoch relativ hoch. Rund 40 % der Open Access-Artikel (DOAJ) können nicht als Volltext oder Link ausgegeben werden. Die STM-Zeitschriften der ISI-Liste liefern den höchsten Anteil an Link-Nachweisen (rund 44 %).

3. Verteilung der Webserver

Verweist ein Treffer auf einen Nachweis, gibt Google neben dem Hauptlink (siehe Abb. 2) noch weitere Links an, unter denen das Dokument zu finden ist. Hierbei interessiert die Verteilung dieser Webserver pro Zeitschriftenliste. Tabelle 3 zeigt die 25 häufigsten Server, die Zeitschriften der ISI-Liste nachweisen. Die Spalte „Beschreibung der Anbieter“ trifft eine Aussage über die Art der Server. „Verlag“ sind kommerzielle Verlagsserver, bei denen der Volltextabruf kostenpflichtig ist. „Digitale Bibliothek“ steht für Server, die kostenfreie Nachweise bieten, die aber nicht in jedem Fall den Volltext direkt liefern können. Unter Umständen treffen bei einem Server beide Beschreibungen zu, wie bei portal.acm.org. „OA Volltext“ bezeichnet Open Access-Server, die frei zugängliche Volltexte liefern.

Tab. 3: Verteilung der 25 häufigsten Webserver (ISI-Liste)

Webserver	Name des Anbieters	Beschreibung des Anbieters	Häufigkeit
ncbi.nlm.nih.gov	National Center for Biotechnology Information	Digitale Bibliothek	150.616
ingenta.com	Ingenta	Verlag	68.925
csa.com	CSA	Verlag	54.652
ingentaconnect.com	Ingenta	Verlag	52.051
springerlink.com	Springer-Verlag	Verlag	21.114
doi.wiley.com	Wiley Publishers	Verlag	19.280
kluweronline.com	Kluwer	Verlag	18.196
adsabs.harvard.edu	NASA Astrophysics Data System	Digitale Bibliothek	16.381
portal.acm.org	Association for Computing Machinery	Verlag, Digitale Bibliothek	15.280
blackwell-synergy.com	Blackwell Publishing	Verlag	14.216
dx.doi.org	Digital Object Identifier System	Linkresolver ¹⁶	13.697

¹⁶ Das Digital Object Identifier System identifiziert Objekte (Artikel, Bücher usw.) über ihre eindeutige ID, die DOI und leitet die Nutzer zu den Verlagen, die die Dokumente nachweisen. Es übernimmt somit die Aufgabe eines Linkresolvers.

Fortsetzung Tabelle 3

taylorandfrancis. metapress.com	Taylor & Francis Group	Verlag	13.221
ideas.repec.org	RePEc Economics database	Digitale Bibliothek	7.681
ieeexplore.ieee.org	IEEE	Verlag, Digitale Bibliothek	6.405
journals.cambridge.org	Cambridge University Press	Verlag	5.379
nature.com	Nature Publishing Group	Verlag	4.680
content.karger.com	Karger Medical and Scientific Publishers	Verlag	4.219
muse.jhu.edu	Muse Scholarly journals online	Digitale Bibliothek	3.944
link.aip.org	American Institute of Physics	Digitale Bibliothek	3.602
pubmedcentral.nih.gov	National Institutes of Health	OA Volltext	3.377
extenza-eps.com	Extenza e-Publishing Services	Verlag	3.303
papers.ssrn.com	Social Science Electronic Publishing	Digitale Bibliothek	3.271
iop.org	Institute of Physics	Digitale Bibliothek	2.259
arxiv.org	e-Print archive	OA Volltext	2.076
leonline.com	Lawrence Erlbaum Associates	Verlag	1.838

Auffällig ist die Häufung der Verlage am Anfang der Liste, die auf die Kooperation von Google Scholar mit Verlagen (CrossRef Partner) zurückzuführen ist. Im Anhang finden sich die 25 häufigsten Webserver der beiden Zeitschriftenlisten DOAJ und IZ SOLIS.

Die weiteren Ergebnisse 4. - 6. beziehen sich auf sehr einfache Tests, die im Vorfeld und während der Untersuchung durchgeführt wurden.

4. *Ungefähre Größe von Google Scholar*

Zur Größe von Google Scholar lassen sich eigentlich nur sehr vage Schätzungen abgeben. Google selbst macht, wie bereits erwähnt, keine Aussagen zur Größe des Index sowie der Zeitschriften- und Webserver-Abdeckung ihres Services. Daher haben wir einzelne Jahrgänge über das Datumfeld der erweiterten Suche abgefragt. Einschränkend muss dazu gesagt werden, dass die auf unsere Anfragen hin von Google Scholar ausgegebenen Daten z. T. sehr widersprüchliche Ergebnisse liefern. Auf die unterschiedliche Abfrage des Zeitraums 1995 - 2000 gibt Google Scholar folgende verwirrende Ergebnisse zurück:

Tab. 4: Abfrage des Zeitraums 1995 - 2000 in Google Scholar

Anfrage	Zeitraum	ungefähre Treffer
1	1995 - 2000	887.000
2	1995 - 1996	526.000
3	1997 - 1998	572.000
4	1999 - 2000	555.000

Die Anfrage 1 (siehe Tab. 4) nach dem gesamten Zeitraum 1995 - 2000 ergibt einen deutlich anderen Wert als die Summe der Anfragen 2 - 4 nach den Dokumentzahlen der einzelnen Zwei-Jahresabschnitte (1995 - 1996; 1997 - 1998; 1999 - 2000). Demzufolge ist die folgende Abbildung (Abb. 4) mit großer Vorsicht zu betrachten. Abbildung 4 visualisiert die Entwicklung der Dokumentzahlen (Hits) des Zeitraums 1950 - 2004. Die Daten wurden für jedes Jahr einzeln abgefragt. Die Kurve zeigt deutlich sichtbar ein exponentielles Wachstum für das Publikationsaufkommen in diesem Zeitraum. Dieser Verlauf entspricht im Verhältnis dem real messbaren Verlauf, der insbesondere durch *Derek de Sol-la Price*¹⁷ untersucht wurde. Die ca. 8.000.000 aufsummierten Treffer der einzelnen Jahresabfragen (1950 - 2004) geben daher eher einen groben Richtwert als eine genaue Messung der Größe des aktuellen Google Scholar Index.

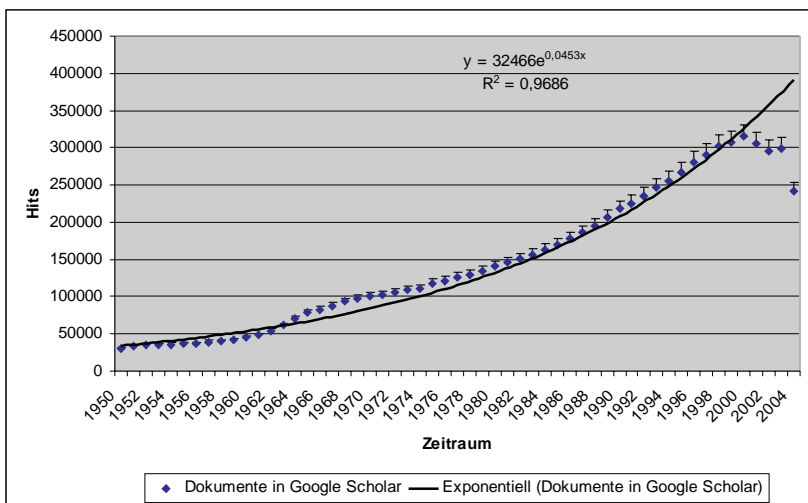


Abb. 4: Ungefähre Anzahl der Dokumente in Google Scholar (Zeitraum 1950 - 2004)

17 Siehe http://de.wikipedia.org/wiki/Derek_de_Solla_Price

5. Abdeckung und Aktualität von Google Scholar

Wie bei Google Web Search kann auch bei Google Scholar die Suchanfrage durch das Schlüsselwort *site* auf eine Domain beschränkt werden. Auf diese Weise erhält man die Anzahl der Artikel, die Google Scholar auf diesem Webserver indiziert hat (vgl. Jacso 2004). Vergleicht man dieses Ergebnis mit den Angaben der Betreiber der Sites (siehe Tab. 5, Spalte “Trefferangaben auf den Webservern“), erhält man einen groben Überblick über die Abdeckung einzelner Server. In Tabelle 5 finden sich acht ausgewählte Webserver. Ingenta, Springerlink, Wiley, Blackwell gehören in die Gruppe der kommerziellen Verlage. Die Angebote der IEEE und ACM sind kommerzielle Angebote von Fachgesellschaften. Das ArXiv und der Astrophysik-Server in Harvard sind nichtkommerzielle frei zugängliche Angebote. Führt man die gleiche Abfrage zeitversetzt durch, kann anhand der Änderung der Ergebnisse eine Aussage über die Aktualität gemacht werden. Für die Server, die keine Angaben zur Anzahl ihrer Nachweise oder Dokumente machen, kann keine Aussage zur Abdeckung getroffen werden. Es kann aber davon ausgegangen werden, dass der Verlagsserver des Blackwell Verlagsservers deutlich mehr Dokumente nachweist, als die von Google Scholar indizierten 71.500 Artikel. Das gleiche trifft für die Digitale Bibliothek der ACM zu. Für alle anderen Server gilt, dass Google Scholar bis jetzt nur einen Teil der Dokumentenbestände abdeckt. Die Abweichungen sind z.T. erheblich und lassen sich schwer erklären. Wir vermuten, dass Google Scholar für den Start nur einen Teil der Angebote indiziert hat. Bei einer Wiederholung der Abfragen Mitte Juli 2005 konnte keine Aktualisierung der Dokumentzahlen dieser Server festgestellt werden. Dieses Ergebnis verdeutlicht den Beta-Status des Services und lässt darauf schließen, dass Google Scholar momentan den Index nicht laufend aktualisiert.

Tab. 5: Abdeckung einzelner in Google Scholar erfasster Webserver

Ausgewählte Webserver	Trefferangaben in Google Scholar	Trefferangaben auf den Webservern
site:adsabs.harvard.edu	303.000	4.200.000
site:ieeexplore.ieee.org	193.000	1.100.000
site:springerlink.com	146.000	2.200.000
site:doi.wiley.com	111.000	4.500.000
site:ingentaconnect.com	108.000	18.000.000
site:portal.acm.org	94.700	?
site:blackwell-synergy.com	71.500	?
site:arxiv.org	56.400	330.000?

6. Vergleich SOLIS und Google Scholar

Den Abschluss unserer Untersuchung bildet ein Vergleich von Google Scholar mit der sozialwissenschaftlichen Fachdatenbank SOLIS anhand von zwei Beispielanfragen, die einen Eindruck von der Abdeckung einer renommierten sozialwissenschaftlichen Fachzeitschrift und der Zahl der Nachweise für einen sehr selektiven Fachbegriff geben sollen.

- Beispiel 1: Gesucht werden Artikel der „Koelner Zeitschrift fuer Soziologie und Sozialpsychologie“, einer renommierten deutschsprachigen Fachzeitschrift aus dem Bereich der Sozialwissenschaften. Die Recherche in SOLIS ergibt 2.756 sozialwissenschaftlich relevante Nachweise, die intellektuell inhaltlich erschlossen sind und aussagekräftige Abstracts vorweisen. Google Scholar liefert 753 Treffer, die jedoch überwiegend unerschlossene Zitationen darstellen. Der Service gibt lediglich minimale bibliografische Angaben (Titel, Autor, Zeitschrift, Jahr und Zitationswert) eines Artikels an. Von einer hochwertigen intellektuellen Erschließung sowie einer umfassenden Abdeckung der Artikel der Zeitschrift kann nicht gesprochen werden.
- Beispiel 2: Gesucht wurde bei dieser Anfrage mit dem Schlagwort „Anarchosyndikalismus“¹⁸, einem sozialwissenschaftlichen Fachbegriff, der für eine spezifisch sozialwissenschaftliche Fragestellung steht. In SOLIS findet man 37 Nachweise. Alle SOLIS-Treffer sind sozialwissenschaftlich relevant und weisen fachwissenschaftlich begutachtete und publizierte Aufsatznachweise bzw. Monographien nach. Google Scholar hingegen liefert 5 Nachweise, die sich aus 3 nichtwissenschaftlichen Ressourcen und 2 Zitationen zusammensetzen.

4 Fazit¹⁹

Wie der herkömmliche Suchdienst Google Web Search bietet auch Google Scholar eine sehr schnelle Suche und einfach zu bedienende Benutzungsoberfläche. Pluspunkte sind weiterhin, dass die Recherche kostenfrei ist und dass interdisziplinär in Volltextbeständen gesucht werden kann. Der Ansatz von Goo-

18 Dieser Deskriptor ist dem Thesaurus Sozialwissenschaften entnommen. Anarcho-Syndikalismus ist laut Schmidt, Manfred G.: Wörterbuch zur Politik, Stuttgart: Kröner 1995 eine „Bezeichnung für eine Allianz von Anarchismus und Syndikalismus, eine vor allem in romanischen Ländern verbreitete Spielart des Anarchismus, die insb. die Abschaffung staatlicher und klassengebundener Herrschaft und die Übernahme der Produktionsmittel durch Arbeiter-Assoziationen, insb. Gewerkschaften zum Ziel hat, ...“. Zur weiteren Information siehe auch <http://de.wikipedia.org/wiki/Anarchosyndikalismus>

19 Wir sind uns bewusst, dass wir die Aussagen, die wir hier getroffen haben, u. U. beim nächsten Update von Google Scholar revidieren müssen. Alle Ergebnisse der Studie sind eine Momentaufnahme und basieren auf Stichproben (100 Treffer pro Anfrage).

gle Scholar bietet für Literatursuchende einige Potenziale, wie z.B. die automatische Zitationsanalyse (siehe dazu Belew 2005) und das darauf aufbauende Ranking sowie in vielen Fällen den direkten Volltextzugriff.

Die Studie zeigt, dass sich zwar ein Großteil der Zeitschriften der drei abgefragten Listen in Google Scholar finden lassen. Genauer betrachtet, wird dieses Ergebnis jedoch durch den hohen Anteil an extrahierten Referenzen relativiert (siehe Abb. 3, 44 % Zitationen). Überwiegend vertreten sind die internationalen STM-Journals der ISI-Liste. Die Analyse der Webserver zeigt, dass ein Großteil der analysierten Treffer von Verlagen gestellt wird. Vermutlich wurden vorrangig die Fachangebote der CrossRef-Partner sowie von weiteren kommerziellen Fachverlagen *teilweise* indexiert (siehe Tab. 3). Der deutschsprachige Anteil an Fachzeitschriften, getestet anhand der sozialwissenschaftlich ausgerichteten IZ-Liste, ist aller Wahrscheinlichkeit nach sehr gering.

Unsere Ergebnisse zeigen, dass umfangreiche frei zugängliche Bestände, insbesondere aus dem Open Access-Bereich bislang wenig berücksichtigt werden. Unverständlich ist, dass Artikel, die sich auf frei im Internet verfügbaren Webservern befinden, häufig von Google Scholar nicht nachgewiesen werden, obwohl sie meistens über eine klassische Google-Suche zu finden sind. Obwohl angekündigt wird, „scholarly articles across the web“ anzubieten, ist der Anteil der nachgewiesenen Artikel aus Open Access-Zeitschriften bzw. der Volltexte (Eprints, Preprints) vergleichsweise gering.

Unsere Tests zeigen weiterhin, dass Google Scholar *keine hochaktuellen Daten* präsentieren kann. Der Google Scholar-Index scheint auf einem „alten“ Crawl zu basieren (wahrscheinlich Anfang 2005). Index-Updates konnten jedenfalls im Untersuchungszeitraum April, Mai, Juli 2005 nicht festgestellt werden. Die Erfahrungen von *Peter Jacso* zur Abdeckung (Jacso 2005) können wir über die Abfrage der Zeitschriftenlisten empirisch bestätigen. Allerdings muss dem Service zugute gehalten werden, dass er sich in einem Beta-Stadium befindet. Diese Tatsache erklärt aber weitere Defizite wie Dubletten in den Daten, fehlerhafte Trefferergebnisse und z. T. nichtwissenschaftliche Quellen nicht gänzlich.

Im Vergleich zu Fachdatenbanken bietet Google Scholar z. Z. nicht die Transparenz und Vollständigkeit, die viele Nutzer von einem wissenschaftlichen Informationsangebot erwarten werden. Als Ergänzung der Recherche in Fachdatenbanken - v. a. durch die Abdeckung einer Reihe von Open Access-Zeitschriften - kann Google Scholar aber durchaus nützlich sein.

Literatur

- Asbrand, Deborah (2004): *Wie das Wissen in das Internet kommt*. Heise Verlag.
URL: <http://www.heise.de/tr/aktuell/meldung/print/54249>
- Banks, Marcus A. (2005): *The excitement of Google Scholar, the worry of Google Print*. URL: <http://www.bio-diglib.com/content/2/1/2>
- Belew, Richard K. (2005): *Scientific impact quantity and quality: Analysis of two sources of bibliographic data*. URL: <http://arxiv.org/abs/cs.IR/0504036>
- Ewert, Gisela; Umstätter, Walther (1997): *Lehrbuch der Bibliotheksverwaltung*. Stuttgart: Hiersemann. ISBN 3-7772-9730-5
- Google (2004): *About Google Scholar*.
URL: <http://scholar.google.com/scholar/about.html>
- Harnad, Stevan; et al. (2004): *The green and the gold roads to Open Access*. In: *Nature*. URL: <http://www.nature.com/nature/focus/accessdebate/21.html>
- Jasco, Peter (2004): *Google Scholar Beta*. Thomson Gale.
URL: <http://www.galegroup.com/servlet/HTMLFileServlet?imprint=9999®ion=7&fileName=/reference/archive/200412/googlescholar.html>
- Jasco, Peter (2005a): *Google Scholar Beta (Redux)*. Thomson Gale.
URL: <http://www.gale.com/servlet/HTMLFileServlet?imprint=9999®ion=7&fileName=/reference/archive/200506/google.html>
- Jasco, Peter (2005b): *Google Scholar: the pros and the cons*. In: *Online Information Review* 29, Nr. 2
- Kennedy, Shirly; Price, Gary (2004): *„Google Scholar“ is Born*. ResourceShelf.com.
URL: <http://www.resourceshelf.com/2004/11/wow-its-google-scholar.html>
- Krause, Jürgen (2003): *Standardization, heterogeneity and the quality of content analysis: a key conflict of digital libraries and its solution*, (06. August). - IFLA 2003, World Library and Information Congress: 69th IFLA General Conference and Council.- International Federation of Library Associations and Institutions, Social Science Libraries Section, Berlin, 01. - 09. August. URL: http://www.ifla.org/IV/ifla69/papers/085e_trans-Krause.pdf
- Lawrence, Steve; Giles, C. Lee; Bollacker, Kurt (1999): *Digital Libraries and Autonomous Citation Indexing*. In: *IEEE Computer* 32, Nr. 6, S. 67-71.
URL: <http://citeseer.ist.psu.edu/aci-computer/aci-computer99.html>
- Lewandowski, Dirk (2004): *Spezielsuche für wissenschaftliche Informationen*. In: *Password*, S. 24.
URL: http://www.durchdenken.de/lewandowski/doc/suchmaschinen-news_dez2004.pdf

- Markoff, John: Google Plans New Service For Scientists And Scholars. In: New York Times vom 18.11.2004
- Payne, Doug (2004): Google Scholar welcomed.
URL: <http://www.biomedcentral.com/news/20041123/01/>
- Price, Gary (2004): Google Scholar Documentation and Large PDF Files. SearchEngineWatch.com.
URL: <http://blog.searchenginewatch.com/blog/041201-105511>
- Suchdienst für Wissenschaftler gestartet. In: Frankfurter Allgemeine Zeitung vom 22.11. 2004. URL:
<http://www.faz.net/s/Rub21DD40806F8345FAA42A456821D3EDFF/Doc~EE9B89791578A43DC94D3EFC756117E72~ATpl~Ecommon~Scontent.html>
- Sullivan, Danny (2004): Google Scholar Offers Access To Academic Information. Searchenginewatch.com.
URL: <http://searchenginewatch.com/searchday/article.php/3437471>
- Swan, Alma; Brown, Sheridan (2005): Open access self-archiving: An author study, 2005. URL: <http://cogprints.org/4385/>
- Terdiman, Daniel (2004): A Tool for Scholars Who Like to Dig Deep, In: New York Times vom 25.11.2004.
- Umstätter, Walther (1998): Die Rolle der Digitalen Bibliothek in der modernen Wissenschaft. S. 297-316. In: Fuchs-Kittowski, Klaus; Laitko, Hubert; Parthey, Heinrich; Umstätter, Walther (Hrsg.): Wissenschaft und Digitale Bibliothek. Berlin: Gesellschaft für Wissenschaftsforschung.
URL: http://www.wissenschaftsforschung.de/JB98_297-316.pdf

Angaben zu den Autoren

Philipp Mayr, Anne-Kathrin Walter
Informationszentrum Sozialwissenschaften (IZ)
Abt. Forschung und Entwicklung
Lennéstraße 30
53113 Bonn
mayr@bonn.iz-soz.de, walter@bonn.iz-soz.de

Philipp Mayr studierte Bibliothekswissenschaft, Informatik und Soziologie an der Humboldt-Universität zu Berlin. Zu seinen Forschungsinteressen gehören Information Retrieval im Bereich der Internet-Suchmaschinen sowie Metriken des Internet (Webometrie). Seit November 2004 arbeitet er im Projekt „Modellbildung und Heterogenitätsbehandlung“. Philipp Mayr ist wissenschaftlicher Mitarbeiter am Informationszentrum Sozialwissenschaften in Bonn.

Anne-Kathrin Walter studierte Informatik an der Universität Bremen. Seit September 2004 arbeitet sie im Projekt „Modellbildung und Heterogenitätsbehandlung“ am Informationszentrum Sozialwissenschaften in Bonn. Ihre Forschungsinteressen liegen im Information Retrieval, u.a. im Bereich der semantischen Heterogenitätsbehandlung.

Anhang

Anhang 1: Verteilung der 25 häufigsten Webserver (DOAJ-Liste)

Webserver	Häufigkeit
ncbi.nlm.nih.gov	10289
dx.doi.org	4582
pubmedcentral.nih.gov	4424
citebase.eprints.org	2495
bmc.ub.uni-potsdam.de	2282
biomedcentral.com	2264
scielo.br	2256
csa.com	1368
ajol.info	1129
emis.ams.org	940
bioline.org.br	865
adsabs.harvard.edu	854
gdrs-intranet.ath.cx	495
tspace.library.utoronto.ca	484
medind.nic.in	368
portal.acm.org	344
ideas.repec.org	291
fizika.hfd.hr	287
emis.de	267
copernicus.org	266
ingentaconnect.com	239
arxiv.org	235
scielosp.org	230
hindawi.co.uk	229
bioline.uts.utoronto.ca	223

Anhang 2: Verteilung der 25 häufigsten Webserver (IZ-Liste)

Webserver	Häufigkeit
ideas.repec.org	167
springerlink.com	107
papers.ssrn.com	91
qualitative-research.net	72
eiop.or.at	67
ncbi.nlm.nih.gov	59
netec.mcc.ac.uk	54
demographic-research.org	42
ingentaconnect.com	34
hsr-trans.zhsf.uni-koeln.de	33
webdoc.sub.gwdg.de	28
webdoc.gwdg.de	27
wu-wien.ac.at	26
diw.de	21
muse.jhu.edu	21
cesifo.de	20
olymp.wu-wien.ac.at	18
sofi-goettingen.de	17
gwdu05.gwdg.de	16
thieme-connect.com	13
wwwuser.gwdg.de	11
repec.iza.org	8
gespraechsforschung-ozs.de	8
wifak.uni-wuerzburg.de	7
uni-bielefeld.de	7

Forschungsrankings: Artefakte oder Sichtbarkeit der Forschungsarbeit?

Stefan Hornbostel

IFQ Institut für Forschungsinformation und Qualitätssicherung, Bonn
Institut für Sozialwissenschaften der Humboldt-Universität zu Berlin

Zusammenfassung

Forschungsrankings gehören inzwischen zum Alltag. Die diesen Rankings zugrunde liegenden Indikatoren, ihre Konstruktion und Angemessenheit werden im Beitrag diskutiert und die Beziehungen zwischen einzelnen Indikatoren und verschiedenen Rankings exemplarisch dargestellt.

1 Einleitung

Was noch vor 10 Jahren von vielen Beobachtern als modisches Strohfeuer abgetan wurde, ist heute fester Bestandteil der Wissenschaftslandschaft geworden: Die Evaluation von Forschung und Lehre. Nicht nur die inzwischen in fast allen Bundesländern in der einen oder anderen Form institutionalisierte Evaluation der Lehre und zunehmend auch der Forschung, sondern ebenso die regelmäßig erscheinenden Rankings zu Forschungs- und Lehrleistungen und last not least die zunehmende Verbreitung von kennzahlgestützten Systemen leistungsorientierter Mittelvergabe deuten darauf hin, dass es sich nicht um eine Modewelle, sondern um eine dauerhafte Veränderung handelt. Eine Veränderung, die zwar bisher nur in geringem Umfang zu einer unmittelbaren Umverteilung finanzieller Ressourcen geführt hat (anders als etwa in Großbritannien), aber zunehmend mittelbare Auswirkungen, sei es in Gestalt der Studienortentscheidung von Studienanfängern, sei es in Form ministerialer Strukturentwicklungsplanungen oder hochschulinterner Profilbildungen zeitigt.

Dabei war zu Beginn dieser Entwicklung in den 1980er Jahren die Forschung das bevorzugte Untersuchungsobjekt, das zunächst aus eher akademischen Interessen mit bibliometrischen Instrumenten vermessen wurde; in den 1990ern rückte jedoch die Evaluation der Lehre deutlich in den Vordergrund und seit einigen Jahren lässt sich beobachten, dass sich das öffentliche Interesse neuerlich auf die Forschung verlagert (vgl. van Raan 2004).

Dieses neue Interesse ist weniger durch aufwändige Entwicklungsarbeit an bibliometrischen Verfahren gekennzeichnet, als vielmehr durch die Anwendung derartiger Verfahren, meist kombiniert mit anderen (nicht-bibliometrischen) Forschungsindikatoren und gekennzeichnet durch eine breite öffentliche Kommunikation der Ergebnisse in Form von Ranglisten: Das CHE zusammen mit der ZEIT und der FOCUS veröffentlichen regelmäßig nationale Forschungsrankings, die DFG informiert ebenfalls regelmäßig mit einem „Förder-Ranking“, der Wissenschaftsrat arbeitet an einer Pilotstudie für ein bundesweites Forschungsrating, die Europäische Kommission legte jüngst den Versuch eines europäischen Forschungsrankings in ausgewählten Fächern vor, das CEST in der Schweiz ermittelt die Mitglieder der internationalen „Champions League“, die Jiao Tong University in Shanghai vermarktet ein „Academic Ranking of World Universities“, und das sind bei weitem nicht alle kursierenden Ranglisten zu nationaler oder internationaler Forschungsperformanz. Zumindest diejenigen Hochschulen, die auf derartigen Ranglisten respektabel platziert sind, nutzen verständlicher Weise die jeweiligen Rankings für ihre Marketingzwecke. Ebenso verständlich ist unter diesen Bedingungen allerdings die Frage danach, ob derartige Rankings Forschungsleistungen tatsächlich angemessen abbilden oder vielmehr künstliche Differenzen produzieren, mit möglicherweise fatalen Fehlsteuerungseffekten in der Folge (vgl. Weingart 2005).

2 Sichtbare und unsichtbare Leistungen?

Es sind allerdings nicht nur die Institutionen des Forschungssystems (Institute, Fachbereiche, Hochschulen), die mit der dauerhaften Bewertung der Forschungsperformanz unter Anforderungsdruck geraten. Auch auf der disziplinären Ebene provozieren die Forschungsrankings eine Auseinandersetzung um die Kriterien angemessener Bewertung. Derartige Auseinandersetzungen waren bisher den methodologischen Debatten vorbehalten und endeten meist in einer mehr oder weniger friedlichen Koexistenz von Paradigmen und Schulen. Über Evaluationen und Rankings kehrt nun die Frage nach einer „disziplinären Gestalt“ gewissermaßen durch die Hintertür zurück. Die Messung von Effektivität – verstanden als Grad der Zielerreichung – setzt voraus, dass disziplinäre Ziele und Standards vorhanden sind, aus denen sich Kriterien der Bewertung ableiten lassen. Effizienzbeurteilungen – also Zweck-Mittel-Relationen – erfordern neben der Erfassung eines Outputs auch eine Inputmessung. Diese Anforderung auf der disziplinären wie auf der institutionellen Ebene, Zielvorstellungen so zu explizieren, dass sie operationalisiert werden können, stößt bei den verschiedenen Disziplinen auf ganz unterschiedliche Voraussetzungen.

Auf der einen Seite stehen die Naturwissenschaften mit einem hohen Internationalisierungsgrad, einer vergleichsweise gut dokumentierten Forschungskom-

munikation, die sich überwiegend auf Fachzeitschriftenbeiträge stützt, einer weitgehend projektförmigen, unter Einsatz von Drittmitteln organisierten Forschung und einer Nachwuchsförderung, die relativ eng in die Forschungsaktivitäten eingebunden ist. Auf der anderen Seite stehen die Geisteswissenschaften, für die diese Bedingungen nicht oder zumindest weitaus weniger zutreffen. Dazwischen lassen sich jene Sozial- und Gesellschaftswissenschaften positionieren, die starke empirische Orientierungen mit geisteswissenschaftlichem Reflexionswissen verbinden. Die Fähigkeit, ein konsistentes Selbstbild zu entwickeln, ist damit bereits recht unterschiedlich ausgeprägt und wird insbesondere in den relativ jungen Disziplinen meist durch interne Ausdifferenzierungsprozesse zusätzlich belastet. Unter solchen Bedingungen entsteht sehr schnell ein öffentlicher Rechtfertigungsdruck, der sich entlang der in anderen Disziplinen eingesetzten Kriterien wissenschaftlicher Leistung entfaltet. Es wird dann schwer zu unterscheiden, ob das angelegte Kriterium ungeeignet ist oder aber durchaus zutreffend ein Problem des Faches beschrieben wird, das im Fach selbst unzureichend thematisiert ist.

Beispiele dafür, wie unter solchen Bedingungen „Artefakte“ entstehen können, lassen sich sicher viele anführen. Hier sei pars pro toto auf einen jüngeren Versuch hingewiesen, „die Forschungsleistungen der Soziologie an zehn deutschen Universitäten“ mit Hilfe bibliometrischer Methoden zu erfassen (Litzenberger/Sternberg 2005, S. 174). Die Autoren der Studie, die durchaus die Beschränkungen ihrer Methode sehen, benutzen den Social Sciences Citation Index (SSCI), um die Forschungsleistungen zu vergleichen. Nun ist dies angesichts einer sprachlich stark national orientierten Disziplin und einer extrem geringen Repräsentanz deutscher Soziologiezeitschriften im SSCI ohnehin ein problematisches Vorgehen. Seine eigentliche Brisanz gewinnt es aber weniger aus nationalen Besonderheiten, sondern aus den disziplinären Publikationsgepflogenheiten. In den Sozialwissenschaften spielt – auch international – die Monographie und der Beitrag in Sammelbänden eine sehr große Rolle, wie die kleine Analyse in Abbildung 1 zeigt.

Im SSCI werden jedoch ausschließlich Zeitschriftenbeiträge ausgewertet. Publikations- und Zitationsanalysen auf dieser Basis machen also einen großen Teil der Forschungsarbeit unsichtbar. Welche Effekte dies auf Leistungsvergleiche hat, ist schwer zu sagen, aber es spricht wenig für die Vermutung, dass es „zu keiner systematischen Verzerrung“ kommt (Litzenberger/Sternberg 2005, S. 187). Die ausgeschlossene Literatur gehört jedenfalls sicherlich nicht zum irrelevanten Teil des Forschungsoutputs, sonst wäre sie kaum in den führenden Zeitschriften zitiert worden¹.

1 Die Diversität der Publikationsmedien ist – anders als in den Naturwissenschaften – typisch für geistes- und sozialwissenschaftliche Fächer. Eine ausschließliche Konzentration auf Zeitschriftenartikel blendet daher wichtige Teile der Fachkommunikation aus. Vgl. hierzu

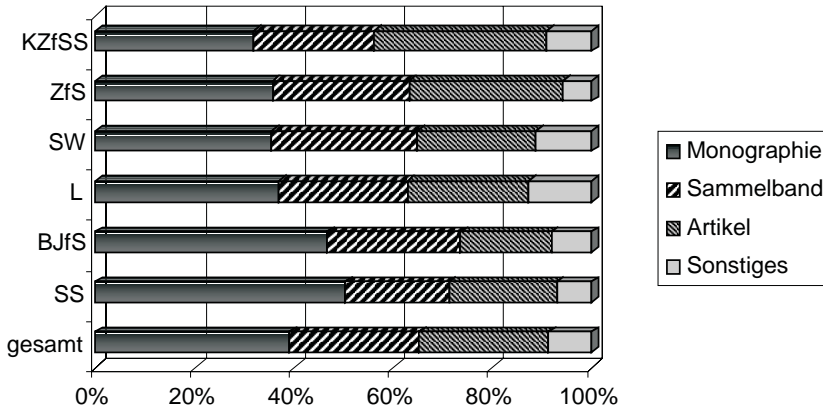


Abb. 1: Wissenschaftliche Quellen, die in sechs deutschen soziologischen Fachzeitschriften des Jahrganges 2002 zitiert wurden – Angaben in Prozent aller Referenzen in der jeweiligen Zeitschrift, Quelle: eigene Erhebungen

3 Artefakte? Natürlich!

Rankings beruhen auf quantitativen Indikatoren und solche Indikatoren sind im ursprünglichen Wortsinne selbstverständlich etwas „künstlich Hergestelltes“. In der sozialwissenschaftlichen Methodenlehre ist mit dem Artefakt allerdings weniger das künstliche oder gar künstlerische der Herstellung gemeint, sondern vielmehr eine Art „Kunstfehler“ im Forschungsprozess, wie am obigen Beispiel dargestellt. Methodisch nicht reflektierte Randbedingungen suggerieren ein Ergebnis, das „prima facie“ als wissenschaftlich solide erscheint und dennoch unzureichend oder gar „falsch“ ist. Artefakte in diesem Sinne sind keine Fälschungen oder offenkundigen Fehler, sie offenbaren sich erst im kritischen Methodendiskurs.

Bei der Konstruktion von Indikatoren wird, in der Regel von einer theoretischen Modellannahme ausgehend, versucht, einen Sachverhalt in seinen wesentlichen Aspekten in einem numerischen Relativ abzubilden. Durch Vergleich

Fortsetzung Fußnote 1

die in der jüngsten Überblicksstudie von Hicks (2004) herausgehobene Bedeutung von vier Kategorien sozialwissenschaftlicher Fachliteratur: Von hoher Relevanz sind neben Veröffentlichungen in internationalen (SSCI) und nationalen Journalen vor allem Monographien und Sammelbände, aber auch die so genannte Transferliteratur (non-scholarly literature), die sich nicht an das direkte Fachpublikum, sondern an Praktiker richtet (vgl. auch Jeanin/Devillard 2005).

erlauben Indikatoren dann auch Bewertungen im Sinne einer mehr oder weniger großen Annäherung an eine Zielvorgabe. Indikatoren sind daher immer begleitet von Objekt- und Instrumenttheorien, um deren Gültigkeit gestritten wird. Sie ähneln insofern durchaus dem wissenschaftlichen Prozedere der Erkenntnisgewinnung, als sie die Randbedingungen ihrer Aussagefähigkeit thematisieren, Vorwissen voraussetzen und als hypothesengeleitete Konstrukte immer wieder an das empirische Wissen angepasst werden müssen.

In diesem Sinne sind Indikatoren hoch artifizuell. Sie konstruieren nämlich gewissermaßen die Welt, die sie abbilden, selbst. Das ist kein Defekt – wie häufig vermutet wird –, sondern eine basale Eigenschaft von Indikatoren. Auffällig wird diese Eigenschaft allerdings nur dort, wo Indikatoren neu entwickelt und implementiert werden. Sind sie einmal zur akzeptierten Routine geworden, wird diese Konstruktionseigenschaft in der Regel übersehen, vergessen oder einfach nicht mehr thematisiert.

Wir haben uns z.B. alle daran gewöhnt, dass wir mit einer Reihe von Indikatoren regelmäßig über den Pulsschlag der Wirtschaft und die Entwicklung gesellschaftlicher Wohlfahrt informiert werden. Sobald das Bruttosozialprodukt steigt oder sinkt, werden Reaktionen ausgelöst. Dabei wird in aller Regel nicht mehr thematisiert, was mit dem Indikator eigentlich gemessen wird, obwohl jeder weiß, dass das Bruttosozialprodukt als Indikator höchst problematisch ist, denn natürlich kann man durch eine stark umweltzerstörende Produktion und die anschließende Reparatur der Schäden das BSP kräftig steigern. Nur: wohlfahrtssteigernd ist dieser Vorgang sicherlich nicht.

Grundsätzlich tritt dieses Problem auch dann auf, wenn Forschungsleistungen nicht aufgrund quantitativer Indikatoren bewertet werden, sondern Beurteilungen durch ein Peer Review-Verfahren erzeugt werden. Auch beim Peer Review entstehen „Artefakte“ im o.g. methodischen Sinne. Der Vorteil des Peer Reviews liegt zweifellos darin, dass Gutachter sehr flexibel die Kontextbedingungen berücksichtigen können, der Preis dafür liegt in einem hohen Grad an subjektiver Färbung des Urteils.

Der wesentliche Unterschied zwischen beiden Verfahren liegt darin, dass dem Gutachtervotum ein expliziter Beurteilungsprozess zugrunde liegt, während Indikatoren massenhaft im Forschungsprozess implizite Urteile (Herausgeberentscheidungen, Mittelbewilligungen etc.) verarbeiten, die zu anderen Zwecken als den mit den Indikatoren verfolgten getroffen wurden.

Weder im Peer Review-Verfahren, noch in einer indikatorenbasierten Bewertung sind es vorrangig wissenschaftsexterne Kriterien, die eine potentielle Produktion von „Artefakten“ vorantreiben, sondern es ist eher eine verstärkte Inanspruchnahme wissenschaftsinterner Kriterien. Die Operationalisierung eines Konzepts von „guter Forschung“ ist es, die die Debatte um die „Artifizialität“ von Forschungsindikatoren immer wieder aufleben lässt.

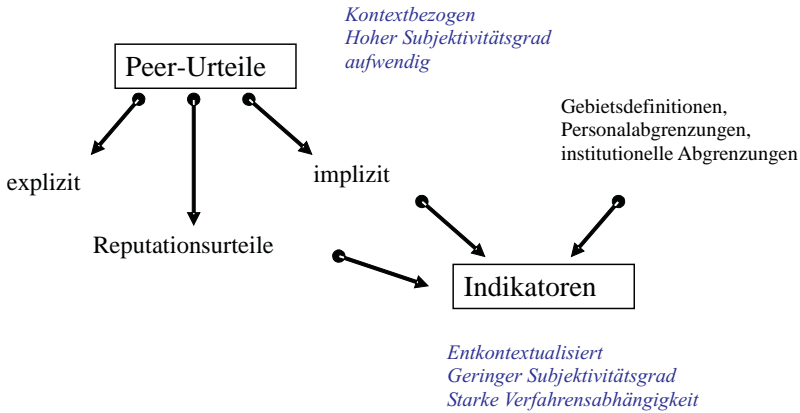


Abb. 2: Bewertungen von Forschungsleistungen

4 Indikatoren

Da die Indikatoren die aus derartigen Operationalisierungsversuchen entstehen, im Großen und Ganzen bekannt sind, beschränke ich mich hier auf einen Überblick (vgl. Hornbostel 1997, 2004, van Raan 2004).

Die erste Gruppe von Indikatoren, bemüht sich darum, Peer-Urteile, die im Förderungs-, Publikations- oder Reputationssystem routinemäßig anfallen, zu quantifizieren. Es kann sich dabei um die Begutachtung von Drittmittelanträgen (Drittmittelindikator), die Manuskriptbegutachtung (Publikationsindikator), Preisverleihungen (z.B. Nobelpreis), Stipendien oder Reputationsurteile (auf Basis von Expertenbefragungen) handeln (vgl. DFG-Förder-Ranking 2003, Sommer 2005). All diese Indikatoren haben spezifische Vor- und Nachteile. Wissenschaftspreise sind meist zu selten oder zu fachspezifisch, um sie für eine Indikatorenbildung zu nutzen, Reputationsurteile weisen nicht nur ein erhebliches time-lag auf, sondern sind auch durch die Quantität des Forschungsoutputs beeinflusst, Drittmittelindikatoren setzen voraus, dass Forschung üblicherweise projektförmig und drittmittelfinanziert organisiert wird, Publikationsindikatoren sagen zunächst etwas über Aktivität, aber nur eingeschränkt etwas über die Qualität der einzelnen Beiträge aus.

Die zweite Gruppe von Indikatoren versucht nicht allein die Peer-Urteile, sondern das Verhalten großer Teile der scientific community im wissenschaftlichen Kommunikationsprozess als Basis zu nutzen. Typischerweise werden in dieser Gruppe nicht nur Publikationen auf der Basis geeigneter Literaturdatenbanken erfasst, sondern auch die Reaktionen der community in Gestalt von Zitationen, die Produktionsstrukturen (Kooperationen, Koautorschaften, Netzwerk-

analysen) sowie die Herausbildung thematischer Vernetzungen (Kozitationsanalysen, inhaltsanalytische Untersuchungen). In dieser Indikatorengruppe sind sehr ausgefeilte Analysetechniken entwickelt worden, gleichwohl ist auf der Theorieebene der Status des Zitats nach wie vor umstritten, denn der symbolischen Handlung des Zitierens können sehr unterschiedlich Signifikate zugeordnet werden, so dass die Interpretation von Zitationsindizes – abgesehen von ihrer Gebundenheit an (sub)disziplinäre Gepflogenheiten – noch immer Fragen aufwirft (vgl. Moed 2005).

Eine vollständige Auflistung aller Publikationen ist äußerst schwer zu realisieren, i. d. R. auch nicht erwünscht, weil meist nur ein kleiner Ausschnitt der gesamten Publikationsmenge für relevant gehalten wird (Bradford's Law). Die Abgrenzung „relevanter“ Literatur beginnt mit der Unterscheidung forschungsrelevanter Literatur von anderer (populärwissenschaftlicher, rein dokumentarischer etc.) und endet bei einer relativ scharfen Selektion der Literatur nach Relevanz, Resonanz, Anschlussfähigkeit o.ä.

Für die letztere Variante steht insbesondere der Science Citation Index (SCI) und dessen Derivate, die bereits bei der Auswahl auszuwertender Zeitschriften das Kriterium einer „Mindestsichtbarkeit“ in Form von Zitaten berücksichtigen. Diese fachübergreifende Literaturdatenbank ist also nicht durch Vollständigkeit, sondern (zumindest vom Anspruch her) durch die Auswahl besonders relevanter Nachweise ausgezeichnet. Sie bietet als (bisher²) einzige Datenbank darüber hinaus auch die Möglichkeit, die Zitationshäufigkeit der erfassten Publikationen zu ermitteln. Im Gegensatz dazu erfassen (nationale) Fachdatenbanken ein sehr breites Spektrum (d.h. nicht nur Zeitschriftenartikel, sondern auch Monographien und Sammelbände) der fachlich einschlägigen Forschungsliteratur, meist sogar auch graue Literatur. Diese Datenbanken orientieren sich weitaus stärker an der Vollständigkeit der gesammelten Nachweise, verfügen aber meist weder über klare Auswahlkriterien noch über eine Relevanzeinschätzung der erhobenen Literatur. Schließlich dienen auch fachlich ausgerichtete europäische oder internationale Datenbanken (die allerdings ebenfalls keine Zitate nachweisen) als Informationsquelle.

Besonders relevant wird in Zukunft die Frage werden, wie sich die wachsende Zahl der Open Access-Publikationen angemessen erfassen lässt (vgl. DFG 2005). Insgesamt bietet die Technik des elektronischen Publizierens ein sehr großes Potential für die automatisierte Analyse und Identifikation „relevanter“ Literatur. Allerdings ist bisher das Feld zwischen online-Zeitschriften mit seriösem Review-System, Dokumentenservern und mehr oder weniger privaten

2 Neuerdings bietet auch die Datenbank SCOPUS die Möglichkeit, fachübergreifend nach Zitationen zu suchen. Wie gut sich diese Datenbank für bibliometrische Zwecke eignet, ist bisher aber noch unklar.

websites sehr weitgespannt und die Analysetechniken noch in den Anfängen (vgl. Thelwall/Vaughan/Björneborn 2005).

Eine dritte Gruppe von Indikatoren setzt nicht an der Urteilskompetenz der scientific community an, sondern an den Reaktionen externer Auditorien. Dieser Schritt über die Wissenschaftlergemeinschaft hinaus ist überall dort von Bedeutung, wo Anwendungsbezüge eine wichtige Rolle spielen. Die zugehörigen Indikatoren greifen entweder auf „attestierter Neugigkeit“ (Patentindikatoren), die wirtschaftliche Verwertung (Lizenzentnahmen), die Entwicklung von Kooperationen mit externen Partnern oder auch auf die soziale, kulturelle oder politische Wirkung in einer größeren Öffentlichkeit zurück. Auch hier gilt, dass sowohl Konstruktion wie Interpretation der Indikatoren erheblichen Sachverstand voraussetzen, denn etwa aus Patentanmeldungen lässt sich keineswegs umstandslos auf die Innovationsintensität schließen (vgl. McAleer/Slottje 2005).

Eine letzte – in Rankings häufig benutzte – Gruppe von Indikatoren greift auf die Schnittstelle von Lehre und Forschung, nämlich die Nachwuchsausbildung, zurück. Diese Indikatoren lassen sich nur sehr eingeschränkt als Forschungsindikatoren verstehen. Ob sie mit anderen Indikatoren harmonisieren, hängt wesentlich davon ab, ob die Nachwuchsausbildung eng in das Forschungsprogramm einer Einrichtung integriert ist (wie es häufig in den Naturwissenschaften der Fall ist), oder ob Promotionen weitgehend unabhängig von institutionellen Forschungsprofilen bearbeitet werden (wie häufig in den Geistes- und Sozialwissenschaften der Fall). Davon abgesehen, ist allein die Zahl von Promotionen und Habilitationen wenig aussagefähig, wenn sie nicht um eine Qualitäts- oder Relevanzgewichtung ergänzt wird. Die Tabellen 1 und 2 zeigen sehr deutlich, dass in der Chemie die Zahl der Promotionen pro Professor deutlich positiv mit den per capita-Werten der eingeworbenen Drittmittel und der Publikationen korrelieren und auch zur mittleren Zahl der Zitate pro Artikel eine positive Korrelation besteht. Für die Soziologie hingegen zeigt sich nicht nur kein signifikanter Effekt, sondern sogar eine leicht negative Korrelation.

Die Übersicht in den beiden Tabellen deutet bereits an, dass Forschungsrankings deutlich an Gehalt und Aussagekraft gewinnen, wenn sie sich nicht nur auf einen Indikator stützen, sondern möglichst verschiedene Indikatoren zur Beleuchtung des Konstrukts „Forschungsqualität“ heranziehen. Das ist nicht unbedingt gleichbedeutend mit Multidimensionalität, denn ob sich auf den verschiedenen Indikatoren tatsächlich unterschiedliche Dimensionen forschungsbezogener Aktivitäten abbilden, scheint von Fach zu Fach sehr unterschiedlich zu sein.

Tab. 1: Chemie - Korrelationen zwischen Indikatoren (CHE)

Chemie - Korrelationen zwischen Indikatoren (CHE)

	Drittm. pro Jahr	DFG- Mittel pro Jahr	Publ. pro Jahr	Zitate pro Jahr	Prom. pro Jahr	Drittm. je Wis- ensch.	Publ. pro Prof.	Zitate pro Papier	Prom. pro Prof.
Reputation	,664**	,532**	,443**	,440**	,635**	,213	,353**	,243*	,447**
Drittmittel pro Jahr		,761**	,572**	,586**	,676**	,640**	,485**	,378**	,500**
DFG-Mittel pro Jahr			,594**	,655**	,568**	,472**	,488**	,455**	,356**
Publikationen pro Jahr				,920**	,716**	,341**	,823**	,411**	,529**
Zitate/Jahr					,721**	,401**	,791**	,694**	,566**
Promotionen pro Jahr						,387**	,508**	,431**	,882**
Drittmittel je Wissensch.							,248*	,351**	,263*
Publikationen pro Professor								,471**	,453**
Zitate/paper									,385**

* Die Korrelation ist auf dem Niveau von 0,05 (1-seitig) signifikant.

** Die Korrelation ist auf dem Niveau von 0,01 (1-seitig) signifikant.

Quelle: CHE Fachbereichserhebung 2002

Tab. 2: Soziologie - Korrelationen zwischen Indikatoren (CHE)

Soziologie - Korrelationen zwischen Indikatoren (CHE)

	Drittm. insgesamt pro Jahr	DFG- Mittel pro Jahr	Publ. pro Jahr	Prom. pro Jahr	Drittm. je Wissensch. schaftler	Publ. je Prof.	Prom. je Prof.
Reputation	,597**	,676**	,501**	,457**	,319*	,357**	,102
Drittmittel ins- gesamt pro Jahr		,740**	,664**	,561**	,534**	,237	,039
DFG-Mittel pro Jahr			,531**	,580**	,327*	,131	,171
Publikationen pro Jahr				,635**	-,003	,615**	,028
Promotionen pro Jahr					-,158	,015	,616**
Drittmittel je Wissenschaftler						,149	-,181
Publikationen je Professor							-,146

* Die Korrelation ist auf dem Niveau von 0,05 (1-seitig) signifikant.

** Die Korrelation ist auf dem Niveau von 0,01 (1-seitig) signifikant.

Quelle: CHE Fachbereichserhebung 2001

5 Fettnäpfchen und Fallstricke

Anders als bei der Evaluation der Lehre werden für die Forschungsevaluationen und darauf basierenden Rankings sehr unterschiedliche Abgrenzungen der Untersuchungseinheiten verwandt. Während in der Lehre ein Studiengang und die ihn verantwortenden Organisationseinheiten maßgeblich sind, mäandert Forschung zwischen organisatorischen und fachlichen Grenzen. Die Wahl einer angemessenen Abgrenzung ist daher weniger eine Frage von richtig und falsch, als vielmehr eine Frage des Erkenntnisinteresses oder der intendierten Anwendung.

Grundsätzlich lassen sich zwei Fragestellungen unterscheiden, die sich auf die Operationalisierung auswirken: Gilt das Interesse vor allen Dingen der Beurteilung organisatorischer Einheiten, dann bildet in der Regel das Personal, das diesen Einheiten zugeordnet ist, die Ausgangsbasis für alle weiteren Analysen. Organisatorische Einheiten wie Lehrstühle, Arbeitsgruppen, Institute, Fachbereiche etc. sind historisch gewachsene Einheiten, die von Hochschule zu Hochschule mehr oder weniger stark in der fachlichen Komposition und Ausrichtung variieren. Insbesondere wenn sich neue Forschungsfelder ausdifferenzieren, kann die Institutionalisierung dieser Gebiete höchst unterschiedlich erfolgen (z.B. kann molekulare Biologie innerhalb der Biologie, der Chemie oder der Medizin beheimatet sein). Umgekehrt beherbergen Fachbereiche u.U. Personal, das von seiner fachlichen Ausrichtung (und entsprechend in der Forschungs- und Publikationsaktivität) auf „fremde“ Forschungsfelder orientiert ist (z.B. Didaktiker in einem naturwissenschaftlichen Fachbereich, die für die Lehrerausbildung zuständig sind). Schließlich werden Kooperationsbeziehungen eingegangen, die sowohl die Organisationsgrenzen als auch die nationalen Grenzen überschreiten.

Der „institutionelle Ansatz“ hat den wesentlichen Vorteil, dass tatsächlich Leistungen einer organisatorisch abgegrenzten Einheit untersucht werden und insofern die Ergebnisse einer Evaluation auch handlungsfähigen Einheiten zugerechnet werden können. Das bedeutet auch, dass sich die „Produktivität“ von organisatorischen Einheiten sehr genau berechnen lässt, da das zugrunde gelegte Personal bekannt ist. Die Nachteile liegen darin, dass u.U. „fachfernes“ Personal in fachbezogenen Datenbanken nicht identifiziert werden kann, dass die Erfassung des akademischen Mittelbaus aufgrund hoher Fluktuation sehr schwierig ist, die Zuordnung von Publikationen teilweise aufwändige Prüfungen der Identität des Autors verlangt (Homonyme) und unter Umständen nur Teile eines Forschungsgebietes erfasst werden (wenn sich ein Forschungsgebiet auf sehr unterschiedliche Institute verteilt). In Deutschland folgt das CHE-Ranking diesem Ansatz.

Der zweite Ansatz geht von einer fachlichen Klassifikation der Drittmittel bzw. der Publikationen aus. Für bibliometrische Analysen bedeutet dies tech-

nisch, dass fachliche Klassifikationen eines Datenbankbetreibers übernommen werden. Meist handelt es sich um die „Subfield-Klassifikation“ des Science Citation Index. Dabei werden nicht Autoren – auch nicht einzelne Zeitschriftenartikel – klassifiziert, sondern ganze Zeitschriften einem Fachgebiet zugeordnet. Das kann konkret bedeuten, dass ein Autor mal als Chemiker, mal als Biologe gezählt wird, je nachdem, wo seine Artikel erschienen sind. Besondere Probleme bereiten dabei die angesehenen multidisziplinären Zeitschriften wie *Science* oder *Nature*, die sich fachlich nicht zuordnen lassen.

Die institutionelle Zuordnung erfolgt dann aufgrund der Adressangabe, die der Autor in seinem Artikel als Korrespondenzadresse angegeben hat. Diese Adressen werfen in der Praxis eine ganze Reihe von Problemen auf, da die Angaben weder vollständig noch standardisiert sind. Das Magazin FOCUS folgt mit seinem Ranking einer Mischung aus Fachklassifikation und Zuordnung nach institutionellen Adressen.

Im Fall des DFG-Rankings dient die interne Zuordnung der Drittmittelanträge zu fachlich kompetenten Fachausschüssen als Klassifikationsgrundlage. Diese Zuordnung ist eine andere als die Klassifikation im Science Citation Index, beide wiederum sind anders als die Klassifikation des Statistischen Bundesamtes. Das DFG-Ranking versucht, durch ausgefeilte Konkordanzen Kompatibilität zwischen diesen abweichenden Klassifikationen herzustellen. Letztlich bleibt aber das jedem Bibliothekar in Gestalt des Thesaurus bekannte Problem, dass Taxonomien nur mehr oder weniger gut auf den (zudem veränderlichen) Gegenstandsbereich passen.

Wie immer man vorgeht, die Abgrenzung der Untersuchungseinheit – mit dem Ziel vergleichbare Einheiten zu definieren – hat immer etwas „Synthetisches“. Problematisch wird dies insbesondere, wenn es um „Effizienz- oder Produktivitätsanalysen“ geht, denn „Sichtbarkeit“ in der Forschung hängt nicht nur von Qualität, sondern auch von einem Größenfaktor ab (je höher der absolute Output an Forschungsergebnissen, desto sichtbarer). Bei der Berechnung von größenunabhängigen pro-Kopf-Angaben wird es sehr wichtig, dem identifizierten „Output“ den zugrunde liegenden personellen „Input“ exakt zuzuordnen. In den CHE-Analysen werden für die bibliometrischen Analysen Namenslisten der Professoren benutzt. In diesem Fall ist also der „Input“ klar definiert und abgegrenzt. Nachteil dieser Strategie ist aber, dass der akademische Mittelbau nicht erfasst wird. Der Mittelbau wird lediglich bei der Kennzahl Drittmittel pro Wissenschaftler berücksichtigt (vgl. Tab. 1 und 2). Auch dies hat allerdings seine Tücken, da die Relation von Professoren und Mittelbau (ohne Drittmittelstellen) sich in Deutschland in den einzelnen Bundesländern sehr unterschiedlich entwickelt hat.

Berücksichtigt man darüber hinaus, dass für die Konstruktion von Indikatoren eine Reihe weiterer Entscheidungen notwendig sind, wie etwa die Zurechnung bei Mehrfachautorschaften, der Umgang mit Selbstziten, die Frage, ob

man nach der „work-done-at“-Methode nur die Publikationen mit der korrekten institutionellen Adresse berücksichtigt oder nach der „current-potential“-Methode alle Publikationen eines Autors, ob man Schwellenwerte oder Gewichtungen benutzt, prozentuale Indizes, Rohwerte oder Abweichungen von Erwartungswerten, dann wird deutlich, dass „Forschungsleistung“ oder „Forschungsqualität“ eine Konstruktion ist, die nicht einfach auf einer Abbildung beruht, sondern auf einem „Herstellungsprozess“. Je klarer innerhalb der jeweiligen scientific communities der Forschungsbegriff und die zugehörigen Qualitätsmerkmale umrissen sind und konsensual getragen werden, desto eher decken sich hinreichend differenzierte Indikatorkonstruktionen mit den Wahrnehmungen der betroffenen Forscher (vgl. van Raan 2005).

Auch die Übereinstimmung unterschiedlicher Forschungsrankings wird je nach Indikator durch die Datengrundlage und die Konstruktion des Indikators mehr oder weniger befriedigend ausfallen. Vergleicht man beispielsweise die Ergebnisse von FOCUS und CHE-Ranking im Fach Psychologie, dann zeigen sich für die Drittmittelwerbungen – trotz unterschiedlicher Quellen – mit einer Korrelation von 0,71 noch erhebliche Übereinstimmungen. Für die Zitate je Artikel gilt dies jedoch nicht mehr (Pearson's Corr. 0,29), wie Abbildung 3 zeigt. Neben echten Fehlern (nicht berücksichtigte Publikationen und Zitationen) in beiden Untersuchungen sind es abweichende Konstruktionsentscheidungen, unterschiedliche Personengruppen und nicht ganz deckungsgleiche Untersuchungszeiträume, die zu diesen Abweichungen führen, obwohl die gleiche Datenbasis (SCI) genutzt wurde.

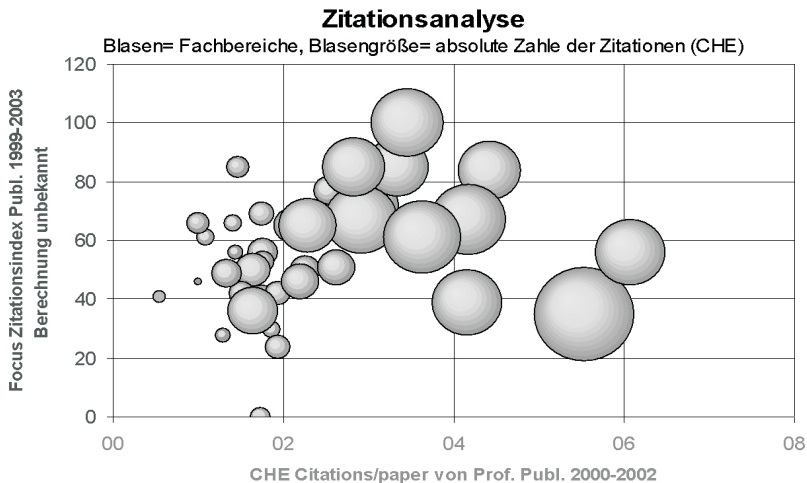


Abb. 3: Zitationsanalyse Fach Psychologie

6 Resümee

Forschungsrankings und -ratings sind einflussreiche Steuerungsinstrumente, wenn sie sorgfältig eingesetzt werden. Umgekehrt haben sie durchaus das Potential fataler Fehlsteuerungen, wenn sie Anreize produzieren, die nicht mit den disziplinären Formen der Wissensakkumulation kompatibel sind. Forschungsindikatoren sind keine Instrumente, die umstandslos in Steuerungsprozessen eingesetzt werden können, sie ähneln eher Diagnoseinstrumenten. Inwieweit sie Forschungsleistungen sichtbar machen können, hängt einerseits an der Solidität ihrer Konstruktion, andererseits daran, wie weit in den jeweiligen Disziplinen Konsens über Qualitätsstandards der Forschung herrscht. Das Grundkonzept von Wissenschaftsindikatoren basiert gerade auf der Überlegung, dass Wissenschaft nur von Wissenschaftlern bewertet werden kann. Solche Bewertungen fallen explizit als Urteile über Forschung oder Forschungsvorhaben an, implizit durch den Umgang mit den Erkenntnisansprüchen anderer. Urteile über die Güte, Nutzbarkeit oder Relevanz von Forschung entstehen also in der Kommunikation über Forschung. Die Urteile, die sich in diesem sozialen Raum bilden, sind nicht mit einer methodologischen Würdigung der Forschungsbefunde zu verwechseln. Die Güte von Forschung in einem methodologischen Sinne besteht nur bedingt aus Akzeptanz und Nutzung. Forschungsindikatoren sind daher Konstruktionen, die bestenfalls das wiedergeben, was Wissenschaftler für gut, wichtig oder nützlich halten.

Die Entwicklung geeigneter Leistungsindikatoren ist ein iterativer Prozess. Erst die Anwendung, empirische Überprüfung und Kritik der Indikatoren führt zu verfeinerten Messinstrumenten. Ein solcher Prozess bedarf allerdings der beständigen Stimulierung und Wissensakkumulation.

Literaturangaben

Deutsche Forschungsgemeinschaft (DFG) (2005): Publikationsstrategien im Wandel? Ergebnisse einer Umfrage zum Publikations- und Rezeptionsverhalten unter besonderer Berücksichtigung von Open Access. Weinheim: Wiley-VCH Verlag.

http://www.dfg.de/dfg_im_profil/zahlen_und_fakten/statistisches_berichtswesen/open_access/download/oa_ber_dt.pdf [Stand: 15.10.05]

Deutsche Forschungsgemeinschaft (DFG) (2003): Förder-Ranking 2003. Institutionen - Regionen – Netzwerke. DFG-Bewilligungen und weitere Basisdaten öffentlich geförderter Forschung. Bonn.

http://www.dfg.de/ranking/download/dfg_foerder_ranking_2003.pdf [Stand: 15.10.05]

- Hicks, D. (2004): The Four Literatures of Social Science. In: Moed, H. F.; Glänzel, W.; Schmoch, U. (Hrsg.): Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems. Dordrecht: Kluwer Academic. S. 473-496.
- Hornbostel, S. (2004): Leistungsparameter und Ratings in der Forschung. In: Evaluation – ein Bestandteil des Qualitätsmanagements. HRK/Beiträge zur Hochschulpolitik 9/2004, S.173-187.
- Hornbostel, S. (1997): Wissenschaftsindikatoren. Bewertungen in der Wissenschaft. Opladen: Westdeutscher Verlag.
- Jeannin, P.; Devillard, J. (2005): Implementing relevant disciplinary evaluations in the social sciences. National vs International interactions in scientific communities. In: Scientometrics, Vol. 63, Nr. 1, S. 121 – 144.
- Litzenberger, T.; Sternberg, R. (2005): Die Forschungsleistung der Soziologie an zehn deutschen Universitäten. In: Soziologie, Jg. 34, Heft 2, S. 74-190.
- McAlee, M.; Slottje, D. (2005): A new measure of innovation: The patent success ratio. In: Scientometrics, Vol. 63, Nr. 3, S. 421 – 429.
- Sommer, S. (2005): Bibliometric analysis and private research funding. In: Scientometrics, Vol. 62, Nr. 1, S. 165 – 171.
- Moed, H. F. (2005): Citation Analysis in Research Evaluation. Dordrecht: Springer.
- Thelwall, M.; Vaughan, L.; Björneborn, L. 2005: Webometrics. In: Annual Review of Information Science and Technology 39, S. 81-135.
- Raan A. F.J., van (2005): Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. In: Scientometrics, Vol. 62, Nr. 1, S. 133-143.
- Raan A. F.J., van (2004). Measuring Science. In: Moed, H. F.; Glänzel, W.; Schmoch, U. (Hrsg.): Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems. Dordrecht: Kluwer Academic, S. 19-50.
- Weingart, P. (2005): Impact of bibliometrics upon the science system: Inadvertent consequences? In: Scientometrics, Vol. 62, Nr. 1, S. 117-131.

Internetquellen

- CEST. „Champions League“: <http://www.cest.ch>
- Shanghai Jiao Tong University. „Academic Ranking of World Universities“: <http://ed.sjtu.edu.cn/ranking.htm>
- Wissenschaftsrat. Forschungsrating: http://www.wissenschaftsrat.de/pilot_start.htm

Angaben zum Autor

Prof. Dr. Stefan Hornbostel
Humboldt-Universität zu Berlin
Philosophische Fakultät III
Institut für Sozialwissenschaften
Unter den Linden 6
10099 Berlin

In die Zukunft publizieren. Positionen und Reflexionen

Johannes Fournier

Deutsche Forschungsgemeinschaft

Zusammenfassung

Die Bedeutung entgeltfrei zugänglicher elektronischer Publikationen und die Entwicklung von einer eher formalisierten Kultur des Publizierens hin zu einer stärker informell geprägten Kommunikation bestimmen den wissenschaftlichen Diskurs zunehmend. Der Beitrag geht von der These aus, dass die zur Akzeptanz jeglicher Publikation nötige Qualitätssicherung wesentliche Voraussetzung für wichtige Funktionen jeder Veröffentlichung ist, nämlich für das Einwerben von Renommee und gesteigerter Sichtbarkeit der eigenen Forschung. Diese Funktionen sind zugleich die Prämissen für die künftige Verbreitung von Open Access, für das umfassende Einbeziehen informeller Diskurse und für die aufwändige Publikation wissenschaftlicher Primärdaten.

1 Einleitung

Die gegenwärtigen Entwicklungen im Bereich des wissenschaftlichen Publizierens sind so vielfältig, die Diskussionen finden auf so vielen verschiedenen Ebenen statt und gelten so vielen unterschiedlichen Themen, dass es wahrlich nicht leicht fällt, Positionen zu formulieren, die dem aktuellen Sachstand gerecht werden. Hinzu kommt, dass wir im Grunde nicht einmal von einer „Zukunft *des* wissenschaftlichen Publizierens“ sprechen können, wie es der ursprünglich vorgesehene Titel der Tagung, deren Beiträge der vorliegende Band versammelt, vorsah. Wir können uns nämlich immer nur auf die Zukunft des wissenschaftlichen Publizierens in gänzlich *verschiedenen* wissenschaftlichen Disziplinen beziehen, in denen die jeweils zu beobachtenden Tendenzen oft unterschiedlich sind und den Entwicklungen in anderen Wissenschaftsbereichen gegebenenfalls zuwiderlaufen können. Lediglich eins lässt sich mit großer Gewissheit behaupten: Über das Thema „In die Zukunft publizieren“ zu schreiben, bedeutet heute unabweislich, sich mit der Bedeutung, Funktion und Akzeptanz vornehmlich des elektronischen Publizierens zu befassen, das heute auch aus den so genannten Buchwissenschaften nicht mehr wegzudenken ist.

Um der Fülle der mit dem Thema „In die Zukunft publizieren“ verbundenen Diskussionen halbwegs gerecht zu werden, sollen die folgenden Ausführungen im Großen und Ganzen vier zentralen Themenbereichen zugeordnet werden, die das Gerüst dieses Beitrags bilden. Ich will mich zunächst kurz mit der Akzeptanz des elektronischen Publizierens befassen, zweitens über die zunehmende Bedeutung von Open Access schreiben, drittens die Entwicklung von einer eher formell geprägten Publikationskultur hin zu einer Kultur wissenschaftlicher Kommunikation kommentieren und schließlich wenige Aspekte der Publikation wissenschaftlicher Primärdaten beleuchten. Für manche der im Folgenden formulierten Positionen sei auf das vom Unterausschuss für elektronisches Publizieren der DFG erarbeitete und im März diesen Jahres vom Ausschuss für wissenschaftliche Bibliotheken und Informationssysteme verabschiedete Positionspapier zum elektronischen Publizieren verwiesen;¹ doch bleibt festzuhalten, dass der folgende Beitrag im Wesentlichen persönliche Einschätzungen referiert.

2 Akzeptanz durch Qualitätssicherung

Manche Diskussionen über das elektronische Publizieren sind erst dann adäquat zu verstehen, wenn man sich bewusst macht, dass jede Publikation mindestens zwei Funktionen erfüllt. Erstens nämlich ist es vornehmste Aufgabe von Wissenschaftlerinnen und Wissenschaftlern, die Ergebnisse ihrer Forschung durch Publikationen zugänglich zu machen. Zweitens aber, und dies gilt für alle Fächer gleichermaßen, sollen Publikationen die Sichtbarkeit auf die Person der Autoren lenken, um ihnen Renommee zu verschaffen und so Karrierechancen zu erhöhen. Diese zweite Funktion ist ganz wesentlich und darf nie vergessen werden, wenn über die Zukunft des Publizierens gesprochen wird. Renommee wird ein Autor nur einwerben, wenn seiner Publikation eine dem höchsten Niveau seiner Wissenschaft entsprechende Qualität bescheinigt wird; also entscheidet die Qualitätssicherung über die Akzeptanz elektronischer Publikationen auf Seiten der Leser *und* der Autoren.

Die Qualitätskriterien, die zur Akzeptanz elektronischer Zeitschriften beitragen, können je nach Wissenschaftsbereich unterschiedlich sein; auch deshalb heißt es immer wieder, dass unterschiedliche Publikationskulturen fächerspezifische Publikationsplattformen erfordern. Auf einem DFG-Rundgespräch zum Publizieren in den Geistes- und Sozialwissenschaften, das Mitte Dezember

1 DFG-Positionspapier: Elektronisches Publizieren. Empfehlungen des Unterausschusses für elektronische Publikationen und des Ausschusses für wissenschaftliche Bibliotheken und Informationssysteme. März 2005. URL: http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/p_os_papier_elektron_publizieren_0504.pdf

2004 in Berlin stattfand, betonten die anwesenden Fachkollegiaten und Vertreter wissenschaftlicher Fachgesellschaften in diesem Zusammenhang vor allem den Aspekt der langfristigen Verfügbarkeit elektronischer Publikationen; dass diese noch nicht hinreichend gesichert sei,² erkläre eine gewisse Zurückhaltung dem neuen Medium gegenüber. Da in den Kulturwissenschaften oftmals Texte aus weit zurückliegenden Zeiten herangezogen werden müssen, ist deren Betonung der Langfristverfügbarkeit als wesentliches Akzeptanzmerkmal evident. Allerdings wird die langfristige Verfügbarkeit auch für die Natur-, Lebens- und Ingenieurwissenschaften immer wichtiger werden, zumal hinsichtlich der Notwendigkeit einer langfristigen Speicherung etwa von Datenbanken oder multimedialer Inhalte.

Vor allem in den Lebenswissenschaften spielt der Impact Faktor nach wie vor eine herausragende Rolle als Qualitätsmerkmal, da Mediziner und Biologen einen engen Zusammenhang zwischen einem rigorosen Peer Review und einem hohen Impact-Faktor annehmen. Damit hängt die Akzeptanz einer elektronischen Publikation derzeit oftmals von dem ihr zugeschriebenen Impact-Faktor ab. Zugleich ist zu beobachten, dass manche Lebenswissenschaftler, offenbar bedingt durch dessen Missbrauch, eine zunehmend kritische Haltung zum Impact-Faktor einnehmen. Folgerichtig werden in diesem Bereich auch neue Ansätze zur Qualitätsbewertung entwickelt, die nicht ganzen Zeitschriften als solchen, sondern allein der Bewertung einzelner Aufsätze gelten, z.B. indem die auch sonst bei elektronischen Zeitschriften zu beobachtende Tendenz zum sog. *unbundling* gewissermaßen auf die Qualitätsbewertung übertragen wird.

Unter „Faculty of One Thousand“ (www.facultyof1000.com, vgl. Abb. 1) werden einzelne lebenswissenschaftliche Aufsätze durch die angesehensten Wissenschaftler der jeweiligen Subdisziplinen bewertet und mit deren Bewertungen in einem sog. „literature awareness tool“ bereitgestellt. Die Gutachter – wenn man sie überhaupt so nennen kann, denn sie bewerten Aufsätze, die sie erst nach ihrer Publikation in gänzlich unterschiedlichen Fachzeitschriften lesen und für die „Faculty of One Thousand“ evaluieren – bewerten die zwei bis vier ihrer Meinung nach wichtigsten und interessantesten Aufsätze, die sie monatlich lesen; auf der Basis der Evaluierung durch Spitzenvertreter der Fächer wird dann in einem bestimmten mathematischen Verfahren der sog. „Faculty 1000 Factor“ ermittelt, mit dem jeder Subskribent des Angebots informiert wird, ob ein bestimmtes Paper als „recommended“, „must read“ oder „exceptional“ gilt. Der Ansatz, die Qualität eines Beitrags somit nicht nach der Zeitschrift zu beurteilen, in der er erschienen ist, sondern die Bewertung allein aufgrund des Beitrags selbst vorzunehmen, scheint mir durchweg überzeugend. Zwar funktio-

2 Einen aktuellen Überblick über die Dauerhaftigkeit unterschiedlicher Speichermedien bietet Stefan Rohde-Enslin: Nicht von Dauer. Kleiner Ratgeber für die Bewahrung digitaler Daten in Museen. Berlin 2004 (Materialien aus dem Institut für Museumskunde – Sonderheft 2; nestor ratgeber 1).

niert dieser Ansatz nur aufgrund des hohen Engagements der bewertenden Wissenschaftler, doch dieses Engagement ist gerade im Publikationswesen ohnehin vorhanden und wird hier auf eine Weise kanalisiert, die der *community* in hohem Maße zugute kommt.

Abb. 1: Faculty of 1000

Ein derartiges System einer durch Fachleute vorgenommenen ex post-Evaluierung, das eine bessere Orientierung im Literaturschunzel ermöglicht, scheint mir persönlich tragfähiger als andere Formen des Open Peer Review,³ nach denen z.B. alle im Zuge der Begutachtung angemerkten Kommentare – oftmals nach interner und externer Begutachtung gestaffelt – begleitend zu dem dann publizierten Text gelesen werden können (vgl. Abb. 2). Meine Skepsis diesem Ansatz gegenüber gründet sich auf der zugegeben sehr persönlichen Einschätzung, dass die im Netz nachzulesenden Gutachterkommentare teilweise wenig zum Verständnis und zur Beurteilung der wissenschaftlichen Qualität der Publikation beitragen; da sie zugleich mit der Publikation erscheinen, wird jeder

3 Das Begutachtungsverfahren der in Abb. 2 gezeigten Zeitschrift beschreibt in aller Kürze Michael Nentwich: Die Zukunft der wissenschaftlichen Zeitschrift – Beispiele für innovative E-Journals. Online-Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare (2002), No. 74, S. 18-20. URL: http://www.uibk.ac.at/sci-org/voeb/om/om74_nentwich.pdf.

Wissenschaftler sie lesen wollen und ist somit zu einer Lektüre quasi ‚genötigt‘, die seinem Erkenntnisinteresse in manchen Fällen wenig dienlich sein mag. Zugleich soll nicht verkannt werden, dass eine offene Begutachtung von Artikeln durchaus zur Qualitätssteigerung einer wissenschaftlichen Zeitschrift beitragen kann.⁴

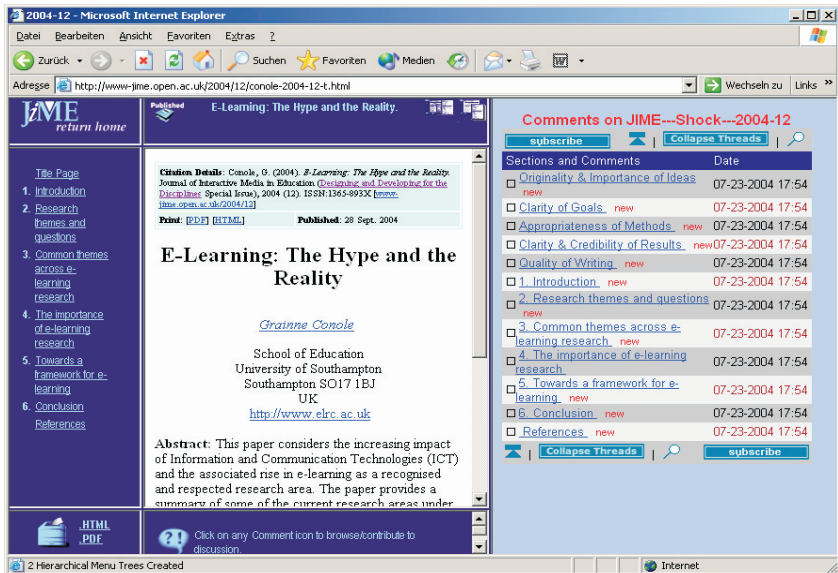


Abb. 2: Journal of Interactive Media in Education

3 Open Access

Im Sommer 2004 hatte die DFG eine Umfrage zum Publikationsverhalten DFG-geförderter Wissenschaftler in Auftrag gegeben, insbesondere um sich ein Bild über die Relevanz entgeltfrei zugänglicher Publikationen für die Kommunikation in unterschiedlichen wissenschaftlichen Disziplinen zu verschaffen. Da der Beitrag von Mike-Alexis Mugabushaka im vorliegenden Band Konzeption und Ergebnisse der Umfrage ausführlich referiert,⁵ soll an dieser Stelle auf eine eingehendere Schilderung der eigentlichen Umfrage verzichtet werden. Doch müssen einige der den Open Access-Abschnitten der Befragung entnom-

4 Vgl. Ulrich Pöschl: Interactive Journal Concept for Improved Scientific Publishing and Quality Assurance. In: Learned Publishing 17 (2004), S. 105-113.

5 Die Ergebnisse der Umfrage sind außerdem unter www.dfg.de/lis (s. unter „Publikationen“) entgeltfrei zugänglich.

mene Resultate hier dargelegt werden, die geeignet sind, wichtige Beiträge für eine in die Zukunft gerichtete Publikationsstrategie zu liefern.

3.1 Breite Zustimmung zur Beförderung von Open Access

Die Umfrage hat zwar gezeigt, dass Wissenschaftler derzeit noch sehr wenig im Open Access publizieren. Von den 1.028 Befragten, deren Antworten ausgewertet werden konnten, haben lediglich 141 angegeben, mindestens einmal in einer Open Access-Zeitschrift publiziert zu haben. Darüber hinaus konnten 341 Wissenschaftler darauf verweisen, dass sie bereits mindestens einen Aufsatz, der konventionell publiziert wurde, sekundär für den entgeltfreien Zugriff im Internet zur Verfügung gestellt haben. Zugleich sprach sich eine deutliche Mehrheit aller befragten Wissenschaftler – erwartungsgemäß mit leichten fachspezifischen Unterschieden – dafür aus, dass die DFG als Organisation Open Access als Publikationsform stärker befördern solle (vgl. Abb. 3). Dabei ist der Nachwuchs der Geistes- und Sozialwissenschaftler offenbar etwas zurückhaltender als der anderer Disziplinen, seine Publikationen im Open Access vorzulegen. Das dürfte teils daran liegen, dass es in den Geisteswissenschaften noch relativ wenige wirklich renommierte Open Access-Publikationsorgane gibt. Zugleich ist zu bedenken, dass gerade in den Geisteswissenschaften der Habilitation bzw. dem sog. „Zweiten Buch“ nach wie vor eine bedeutende Rolle zukommt; als vornehmlich auf Zeitschriftenbeiträge ausgerichtete Publikationskultur eignet sich Open Access für diese Aufgabe in geringerem Maße.

Wie bereits eingangs dargelegt, sollen Publikationen ihren Autoren Renommee verschaffen. Nun ist es inzwischen mehrfach erwiesen, dass frei zugängliche Online-Publikationen die Sichtbarkeit der eigenen Forschung und damit die Voraussetzung für das Einwerben von Renommee deutlich erhöhen. Aus diesem Grund dürfte es in der Tat vor allem eine Frage der Zeit sein, bis Open Access stärker und umfassender praktiziert wird. Auch dass es derzeit Bestrebungen verschiedener Wissenschaftsorganisationen gibt, die klaren Empfehlungen der sog. Berlin 3-Konferenz⁶ in eine regelrechte Publikationspolitik umzusetzen, wird den unterschiedlichen Open Access Publikationsmodellen zugute kommen. Diesen Empfehlungen zufolge sollen Wissenschaftsorganisationen ihren Angehörigen nämlich empfehlen oder gar vorschreiben, erstens eine Archivkopie sämtlicher Publikationen in mindestens einem frei zugänglichen Server zu deponieren und zweitens nach Möglichkeit in Open Access-Zeitschriften zu publizieren. Im Rahmen der Jahresversammlung hat sich der Hauptausschuss der DFG Anfang Juli 2005 dafür ausgesprochen, dass Projektnehmer künftig mit den Bewilligungsschreibern aufgefordert werden sollen, ihre Zeit-

6 Berlin 3 Open Access: Progress in Implementing the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Feb 28th – Mar 1st, 2005, University of Southampton, UK. URL: <http://www.eprints.org/berlin3/>.

schriftenbeiträge durch eine sekundäre Publikation auf institutionellen oder disziplinspezifischen Servern oder durch originäre Veröffentlichungen in Open Access-Zeitschriften entgeltfrei verfügbar zu machen. Auf europäischer Ebene ist vor allem auf die entsprechenden Politiken des Wellcome Trust, der Association of Swedish Higher Education (SUHF) sowie der Research Councils UK (RCUK) zu verweisen.⁷

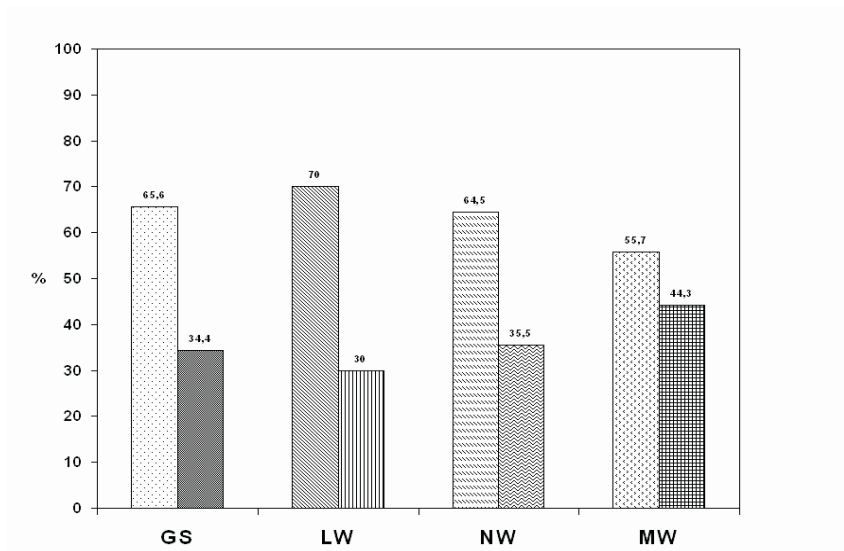


Abb. 3: Sollte die DFG Open Access verstärkt fördern?
 ‚Ja‘-Stimmen je links, ‚Nein‘-Stimmen je rechts. GS = Geistes- und Sozialwissenschaften; LW = Lebenswissenschaften; NW = Naturwissenschaften; MW = Materialwissenschaften.

3.2 Institutional Repositories

Wenn mich mein Eindruck nicht vollends täuscht, wird Open Access in Zukunft vor allem durch das Einstellen bereits anderweitig veröffentlichter Publikationen in institutionelle und disziplinspezifische Server unterstützt werden, während die originäre Publikation in Open Access-Zeitschriften demgegenüber wohl langsamer zunehmen wird, vor allem, da diese als in aller Regel neu aufge-

⁷ Vgl. die Dokumente unter http://www.wellcome.ac.uk/doc_WTD002766.html (Wellcome Trust), <http://www.suhf.se/BinaryLoader.aspx?ObjectID=108&PropertyName=FileList&PropertyValueIndex=0&CollID=File> (SUHF) und <http://www.rcuk.ac.uk/press/20050628openaccess.asp> (RCUK).

legte Zeitschriften gewisse Zeit brauchen, bis sie hohe Impact-Faktoren erreichen und sich somit als attraktive Publikationsorgane anbieten.

Die Umfrage zum Publikationsverhalten DFG-geförderter Wissenschaftler hat ergeben, dass Wissenschaftler disziplinspezifische Server als sinnvollste Variante für die sekundäre Archivierung ihrer bereits anderweitig veröffentlichten Beiträge ansehen. In Deutschland gibt es zwar ein flächendeckendes Netz von Hochschulschriftenservern,⁸ doch die über diese Server verfügbaren Inhalte sind offenbar, auch in qualitativer Hinsicht, für Wissenschaftler derzeit nicht attraktiv genug.⁹ Daher muss die Frage erlaubt sein, ob die Pflege institutioneller Server eine wirklich attraktive Dienstleistung für die Forschung darstellt? Könnte es, wenn man vornehmlich die Interessen der einzelnen wissenschaftlichen Disziplinen, nicht aber die der jeweiligen Institutionen, die an der eigenen Sichtbarkeit mehr interessiert sein dürften als an der Attraktivität angebotener Inhalte für ein bestimmtes Fach, für Hochschulen insgesamt nicht lohnender sein, sich über die Grenzen der jeweiligen Universitäten hinweg zusammenzutun, um, den jeweiligen Profilen entsprechend, gemeinsam Server mit je disziplinspezifischen Angeboten aufzubauen? Die mit dem Etablieren solcher Server verbundene Intention, eine gute Visitenkarte für die hervorragende Forschung der jeweiligen Einrichtungen und Institutionen abzugeben, lässt sich vielleicht am besten erreichen, wenn Universitäten und Hochschulen vornehmlich Sekundärpublikationen genau der Fächer einstellen, für die sie jeweils eine Spitzenstellung einnehmen.¹⁰ Solches Vorgehen käme den Bedürfnissen der Fachwissenschaftler, für deren Literaturbeschaffung das von einer bestimmten Internet-Adresse begonnene Browsing vielfach eine bedeutendere Rolle spielt als die Recherche über Suchmaschinen,¹¹ offenbar weit entgegen.

Die hier nur kurz angedeutete Diskussion zeigt vor allem, dass Infrastruktureinrichtungen künftig noch enger mit Wissenschaftlern als ihren „Kunden“ zu-

8 Vgl. die Liste der unter <http://www.dini.de/dini/wisspub/dokuserver.php> [Stand: 01.07.05] angeführten Dokumentenserver.

9 Erst kürzlich hat sich die Universität Bielefeld dazu entschieden, neben dem allgemeinen Hochschulschriftenserver einen eigenen Server für qualitativ hochwertige Forschungsleistungen anzubieten, um die Attraktivität der über ein solches Repositorium angebotenen Inhalte deutlich zu erhöhen.

10 Ein gelungenes Beispiel für das Hervorgehen disziplinspezifischer Angebote über institutionelle Repositorien findet sich im Kontext des niederländischen DARE-Projekts: Auf der Basis der über Repositorien einzelner Hochschulen verfügbaren Informationen wurde „Connecting Africa“ (<http://www.connecting-africa.net/About.aspx>) als disziplinspezifisches Portal aufgebaut; weitere Informationen vgl. auch unter <http://www.darenet.nl/en/page/language.view/diensten.connectingafrica>.

11 Darauf hat z.B. Carol Tenopir in ihrem Vortrag „The Role of Publishing in the Research Process“ hingewiesen, den sie im Rahmen des JISC International Colloquium „International solutions for the dissemination of research: Considering international solutions for purchasing electronic journals and disseminating research outputs“ am 21. Juni 2005 in London gehalten hat.

sammenarbeiten müssen, um ein Auseinanderklaffen zweier getrennter Publikationswelten von vornherein zu vermeiden. Wenn auf der einen Seite einzelne Wissenschaftler fachspezifische Publikationsangebote aufbauen und dabei auf in Informationseinrichtungen bereits vorhandene Kompetenzen verzichten, wenn auf der anderen Seite eben diese Informationseinrichtungen fachlich nicht weiter spezifizierte Publikationsplattformen einrichten, die in ihrer mangelnden fachlichen Spezifikation den Ansprüchen einer wissenschaftlichen Kommunikationsplattform eben nicht genügen, besteht die drohende Gefahr einer Zerteilung der Informationsversorgung. Hier gilt es, die gemeinsame Arbeit von Wissenschaftlern und Informationsspezialisten mit einer jeweils klar geregelten Aufgabenteilung verstärkt in den Blick zu nehmen und an konkreten Publikationsvorhaben kooperativ umzusetzen.

3.3 Rechtsfragen

Damit entgeltfrei zugängliche Sekundärpublikationen als Angebot an die Wissenschaft attraktiv bleiben, müssen sie rechtlich abgesichert werden. Nach Ausweis der von SHERPA gesammelten und über die sog. ROMEO-Liste verfügbaren Daten gestattet zwar die überwiegende Mehrheit wissenschaftlicher Verlage die zusätzliche Veröffentlichung der Beiträge, die in den jeweiligen Verlagszeitschriften erschienen sind, wenngleich die Konditionen für Sekundärpublikationen von Verlag zu Verlag variieren.¹² Da diese Konditionen letzten Endes auf dem *goodwill* der Verlage basieren, ist jeder Wissenschaftler gut beraten, das Einstellen seiner Veröffentlichungen in Repositorien bzw. die Möglichkeit dazu über entsprechende Formulierungen im Verlagsvertrag abzusichern,¹³ indem er oder sie nicht alle Verwertungsrechte an einen Verleger abtritt, sondern sich einem Verlag gegenüber ein nicht ausschließliches Verwertungsrecht für eben diesen mit Open Access verbundenen Zweck vorbehält. Für die wenigen Zeitschriften, deren Herstellung mit einem Zuschuss der DFG unterstützt wird, wird dies den Projektnehmern seit Februar 2005 mit unseren Bewilligungsschreiben verbindlich vorgeschrieben.

Der gerade beschriebene Rechtsvorbehalt ist umso wichtiger, wenn man die Entwicklungen im Bereich des Urhebergesetzes verfolgt. Denn mit den Neuregelungen, die im sog. „Zweiten Korb“ eingeführt werden sollen, steht zu befürchten, dass die Informationsversorgung im Bereich von Lehre und For-

12 Diese Liste ist verfügbar unter <http://www.sherpa.ac.uk/romeo.php> (Stand: 01.07.05).

13 Vgl. dazu auch Ulrike Specht: Rechtliche Aspekte der Werkverwertung im Rahmen von Open Access. In: Theke aktuell. Für Heidelberger Bibliotheken von Heidelberger Bibliotheken 12/2 (2005), S. 70-78. – Für elektronische Dissertationen bietet das Projekt „Dissertationen Online“ einen Mustervertrag zwischen Autor und Verlag an, der Verlagen hinsichtlich der Online-Fassung ein lediglich einfaches Nutzungsrecht einräumt. URL: http://www.dissonline.de/convert/index_recht.htm

schung, Bildung und Wissenschaft massiv beeinträchtigt wird.¹⁴ Mit einer rechtlichen Absicherung von Sekundärpublikationen und dem Verfügbarmachen ihrer Beiträge im Open Access können Wissenschaftler selbst entscheidend dazu beitragen, dass der Austausch von Forschungsinformationen möglich bleibt und nicht an zunehmend schwierigeren kommerziellen und juristischen Rahmenbedingungen scheitert.

Hinsichtlich der mit Open Access verbundenen Rechtsfragen sollte schließlich auch bedacht werden, wie und mit welchem Nachdruck die verschiedenen Elemente der Definition von Open Access verfolgt werden. Nach der mit der „Berliner Erklärung“ gegebenen Definition beinhaltet Open Access auch eine Abtretung gewisser Rechte. Denn die Autoren hinterlegen die Ergebnisse ihrer Arbeiten „in einem geeigneten elektronischen Format auf mindestens einem online zugänglichen Archivserver mit geeigneten technischen Standards“ *zusammen* mit der Erlaubnis, die Veröffentlichung unter korrekter Nennung des Urhebers für jeden verantwortlichen Zweck weiter verwenden zu dürfen.¹⁵ Die durch diese Bestimmung gedeckte Möglichkeit zur Herstellung sog. abgeleiteter Werke dürfte vielen Wissenschaftlern entschieden zu weit gehen. Daher ist künftig darauf zu drängen, dass z.B. über den Gebrauch von Creative Commons-Lizenzen klar definiert wird, inwieweit eine über den reinen entgeltfreien Zugang hinausgehende Rechteinräumung stattfinden darf, um auch so einer möglichen Verunsicherung der Autoren entgegen zu treten. Nach meiner Auffassung ist der entgeltfreie Zugang zu Texten und Daten das entscheidende und auch das wichtigste Merkmal des Open Access; kurzfristig kann Open Access vermutlich nur in diesem Sinne auch durchgesetzt werden. Deshalb sollte m.E. die in einem weiten Verständnis von Open Access inbegriffene, umfassende Rechtsabtretung wenigstens derzeit nicht *forciert* verfolgt werden.

3.4 Verlage und Open Access

Auch wenn Open Access in manchen Diskussionen nach wie vor als Kampfmittel der Bibliotheken gegen die Macht vornehmlich international agierender Verlagshäuser beschrieben wird, dürfte Open Access kaum die Lösung der Zeitschriftenkrise bringen, schon gar nicht kurzfristig. Die anhaltende Debatte um Modalitäten einer möglichen und notwendigen Finanzierung von Open Access,

14 Zu den Änderungen, die die Informationsversorgung im Wissenschaftsbereich deutlich erschweren könnten, vgl.: Urheberrecht für Bildung und Wissenschaft. Anforderungen an das Zweite Gesetz zur Regelung des Urheberrechts in der Informationsgesellschaft. Hrsg. von Ulrich Sieber und Thomas Hoeren. (Hochschulrektorenkonferenz; Beiträge zur Hochschulpolitik 2/2005). Bonn 2005. – Der Gang der Diskussion ist gut dokumentiert unter www.urheberrechtsbuendnis.de.

15 Zitat nach der deutschen Fassung der „Berlin Declaration on Open Access in the Sciences and Humanities“; s. unter <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>.

die uns weiter begleiten wird, zeigt sehr deutlich, dass die Entwicklung valider Geschäftsmodelle nach wie vor ein Desiderat ist, das nur über praktische Umsetzungen verschiedenen Lösungen zugeführt werden kann. Zwar mag Open Access à la longue Auswirkungen auf die Verlagslandschaft haben. Doch zeigen nicht nur Experimente wie etwa Springers „Open Choice“-Modell, dass auch kommerzielle Verlage neuen Geschäftsmodellen durchaus aufgeschlossen gegenüberstehen. Verleger haben auch begriffen, dass sie von Open Access profitieren können, wie z.B. eine am 20. April über liblicense-1 versendete E-Mail zeigt. In dieser E-Mail berichtet Ann Okerson, dass ein Verlag die Inhalte eines Open Access Repository „abgreift“, diese Inhalte mit diversen Mehrwertdiensten aufbereitet und anschließend kommerziell vermarktet.¹⁶ So bezahlt der Kunde nicht für die reine Information, sondern für den zusätzlichen, vom Verleger bereitgestellten Service. Warum nicht, wenn es sich lohnt, diesen Service in Anspruch zu nehmen, solange die Information selbst frei verfügbar ist? Auch der Umstand, dass Google Scholar künftig, sobald mehrere Versionen eines Dokumentes im Netz verfügbar und nachgewiesen sind, stets die Version des jeweiligen Verlages als erste anführen wird, deutet darauf hin, dass traditionelle Vertriebs- und Verlagsstrukturen bestehen bleiben werden, ohne Open Access an sich Abbruch zu tun. Diese Koexistenz scheint mir auch richtig zu sein.

4 Wissenschaftliche Kommunikation

In jeder Diskussion zur Zukunft des wissenschaftlichen Publizierens ist heute zu bedenken, dass das Publizieren lediglich einen Teilausschnitt der wissenschaftlichen Kommunikation darstellt. Wissenschaftliche Kommunikation ist weit umfassender und oft weniger formal als das wissenschaftliche Publizieren an sich, für das, je nach Fächerkultur, mehr oder weniger strenge, eingespielte Regeln greifen. In den letzten Jahren ist zu beobachten, dass die informelle Kommunikation, die über neue Möglichkeiten der Medien bestens unterstützt wird, zunehmende Bedeutung erlangt. Publikation und Kommunikation werden sich daher in vielfältigen Formen vermischen; auf lange Sicht werden die Grenzen zwischen formeller und informeller Kommunikation vielleicht verschwinden, worin Chancen liegen – etwa in Bezug auf die Aktualität wissenschaftlicher Erkenntnisse oder die Lebendigkeit ihrer Vermittlung in der Diskussion –, aber auch Gefahren – etwa hinsichtlich der Befürchtung, nie einen aktuellen, gültigen Erkenntnisstand definieren zu können.

16 S. unter <http://www.library.yale.edu/~llicense/ListArchives/0504/msg02152.html>.

4.1 Zunehmende Bedeutung der informellen Kommunikation

Die zunehmende Bedeutung der informellen Kommunikation äußert sich nicht nur in enthusiastischen Ankündigungen, wie sie z.B. am 19. April 2005 von Gerry McKiernan voller Emphase über liblicense-l versendet wurde:

„Indeed there is Life After Self-Archiving and Institutional Repositories, and Social Software such as Blogs and, in particular, Wikis, Will/Can/Could/Might create a MajorMindShift !!! [TrustMeOnThisOne]”¹⁷

Dass diese Emphase nicht nur ein frommes Credo ist, sondern die derzeit ablaufenden Entwicklungen des wissenschaftlichen Diskurses zutreffend charakterisiert, wird dadurch bestätigt, dass in der Umfrage zum Publikationsverhalten DFG-geförderter Wissenschaftler einige der Befragten die Frage nach den aus ihrer Sicht wichtigsten Open Access-Zeitschriften für ihr Fach mit der Nennung von Diskussionslisten wie H-Soz-u-Kult oder Virtuellen Fachbibliotheken wie *historicum.net* beantwortet haben, die nicht in eigentlichem Sinne Zeitschriften, sondern Foren einer nicht immer streng formalisierten Kommunikation sind. Diese Antworten zeigen deutlich, dass der informelle Austausch über die Datennetze von zunehmend größerer Relevanz ist.

Wenn außer Frage steht, dass neue Formen der wissenschaftlichen Kommunikation über Maillisten, Newsgroups, Weblogs oder Wikis das traditionelle Publizieren ergänzen werden, ist aber weiter zu fragen, inwieweit Beiträge über eher informelle Kommunikationskanäle eine wesentliche Funktion der klassischen Publikation erfüllen, nämlich das Einwerben von Renommee für die Person des Autors. Wie oben dargelegt, ist diese Funktion in den meisten Fächern eng an die Bewertung der Qualität wissenschaftlicher Beiträge gebunden. Doch das klassische Peer Review kann im Bereich der informellen Kommunikation naturgemäß nicht richtig greifen. Andererseits eröffnet gerade die informelle Kommunikation die Chance, dass öffentliche Beiträge möglichst breit rezipiert werden. In dieser Hinsicht könnte man vielleicht die These wagen, dass für Beiträge etwa in Wikis eine Art Open Peer Review greift, das sich zusehends positiv auf die Qualität entsprechender Beiträge auswirkt, so dass Diskussionen auf hohem Anspruchsniveau im offenen Kommunikationskanal die Regel werden. Es bleibt allerdings abzuwarten, wie schnell sich solche Kommunikationsformen in unterschiedlichen Disziplinen etablieren werden.

4.2 Trends, Tendenzen und Gefahren?

Formen offener Kommunikation basieren auf neuen Technologien; insbesondere Web-Services spielen eine bedeutende Rolle, um die Informationen, die Gegenstand der wissenschaftlichen Diskussion sind, rasch zu verbreiten oder zu erhalten. In einer zunehmend schnelllebigen Welt sind solche Services unver-

17 <http://www.library.yale.edu/~llicense/ListArchives/0504/msg01884.html>.

zichtbar, um die auch in der Wissenschaft notwendige Orientierung zu unterstützen. Zugleich möchte ich kritisch hinterfragen, ob und inwieweit nicht auch die Gefahr besteht, dass Informationen, wenn sie stets in kleinen und kleinsten Häppchen serviert werden, wegführen von einem Gesamtbild, das wir zum Verständnis von Forschungsfragen und letzten Endes für unser Verständnis der Welt benötigen. Besteht z.B. hinsichtlich der zunehmend häufiger eingesetzten RSS-Feeds nicht die Gefahr einer „Zerstückelung der Welt“, die nur noch in Ausschnitten wahrgenommen, diskutiert und erforscht wird? Und besteht nicht auch die Gefahr, dass Neuigkeitendienste, die Information betont kurz vermitteln müssen, gerade die *mainstream*-Forschung verstärken, weil die Zeit – und oft auch die Geduld – der Empfänger vielfältiger Nachrichten nicht ausreicht, auch die Grundlagen weiter zu ermitteln, die Wissenschaftler benötigen, um aktuelle Diskussionen adäquat einzuschätzen? Und nehmen Forscher sich die Zeit, auch die abseits des *mainstream* liegenden Gebiete zu erforschen, für die es keine geeigneten Neuigkeitendienste und keine personalisierten *services* gibt? Auch wenn die offene Kommunikation und der Einsatz der Technologien, die sie ermöglichen, in dem Maße, in dem die Informationsflut wächst, unverzichtbar ist, muss der sinnvolle Umgang mit neuen Technologien vielfach erst eingeübt werden, da es nicht allein darum gehen kann, stets den letzten, angeblich neuesten Meldungen hinterher zu rennen und darüber alles andere, dessen Erforschung durchaus die Mühe lohnt, aus dem Blick zu verlieren.

5 Wissenschaftliche Primärdaten

Wissenschaft und Forschung basieren auf Quellen diverser Herkunft. Dies können Texte, Bilder, Töne, Filme und Daten gänzlich verschiedener Art sein. Gerade in Zeiten knapper Ressourcen ist es nicht mehr zu verantworten, dass solche Quellen, die ja in aller Regel nicht nur von einem Einzelnen oder einzelnen Teams ausgewertet werden (können), sondern von einer ganzen Heerschaar weltweit arbeitender und disziplinar vernetzter Wissenschaftler, nicht so publiziert werden, dass deren sinnvolle und ökonomisch verantwortbare Nachnutzung möglich ist. Die Publikation wissenschaftlicher Primärdaten wird daher zunehmende Bedeutung erlangen; doch um die wichtige Publikation von Primärdaten zu befördern, müssen Wege erprobt und gefunden werden, auch diesen Zweig des Publikationsbaumes über Akzeptanz und Renommee zu steuern.

5.1 Bedeutung korrekter Zitation

Die Publikation von Primärdaten ist mit erheblichem Aufwand verbunden, wenn die einmal erhobenen Daten so dokumentiert und präsentiert werden sollen, dass ihre sinnvolle Nachnutzung auch durch Dritte möglich sein soll. Die-

sen Aufwand werden Wissenschaftler, wenn sie nicht geradezu altruistisch veranlagt sind, vor allem dann auf sich nehmen, wenn sie mit publizierten Primärdaten genau das erlangen, was sie generell von ihren Veröffentlichungen erhoffen: Sichtbarkeit ihrer Forschung und Renommee für sich selbst. Das setzt die Entwicklung einer Kultur voraus, in der gewissermaßen das geistige Eigentum an Primärdatenpublikationen entsprechend attribuiert wird, z.B. indem Mechanismen auch für die korrekte Zitation von Datensätzen aus spezifischen Informationssystemen entwickelt und umgesetzt werden. Das ist meines Erachtens ein unverzichtbarer Anreiz, um die Publikation von Primärdaten in zunehmendem Maße für Wissenschaftler interessant werden zu lassen. Es muss sich, schlicht und ergreifend gesagt, für Wissenschaftler lohnen, Zeit, Geld und Energie in die oft mühsame Aufbereitung ihrer Daten zur Nachnutzung durch Dritte zu investieren.

5.2 Vernetzung

Wenn sich dies schließlich lohnt, können auch die vielfältigen Möglichkeiten der Vernetzung von Publikationen mit Primärmaterialien unterschiedlicher Art besser umgesetzt werden, als es meiner Einschätzung nach bislang geschieht. Zwar wird immer wieder darauf hingewiesen, dass das Internet sich in hervorragender Weise eigne, multimediale Elemente einzubinden und so die empirische Begründung für publizierte Aussagen gleich mitzuliefern, aber dieser Weg ist bei weitem noch nicht in dem wünschenswerten Maße ausgeschöpft. Dabei soll es nicht um eine muntere, oft zweckfreie Illustration der Thesen gehen, sondern um deren punktgenaue Begründung durch Nachweis der Quellen- und Datengrundlage. Dass das gerade jetzt erarbeitete, künftige 7. Rahmenprogramm der Forschungsförderung der Europäischen Union eine eigene Priorität auf die Förderung wissenschaftlicher Infrastrukturen legen wird, ist wohl ein deutliches Zeichen dafür, dass die ungeheure Bedeutung eines gemeinsamen, übernationalen Zugriffs auf wissenschaftliche Quelldaten unterschiedlichster Art erkannt wurde und dringender Umsetzung bedarf.

6 Zukunft des wissenschaftlichen Publizierens

Resümierend, doch keineswegs resignierend lässt sich konstatieren, dass es kaum möglich ist, die Zukunft des wissenschaftlichen Publizierens, die durch viele teils voneinander abhängige Determinanten bestimmt wird, heute halbwegs zutreffend vorauszusagen. Doch eines ist gewiss: Wie sich das wissenschaftliche Publizieren entwickelt, hängt weniger davon ab, welche Publikationspolitik Forschungsorganisationen oder Förderorganisationen mit mehr oder weniger großem Nachdruck verfolgen. Publikationspolicies verschiedens-

ter Art sind gewiss sinnvoll, oftmals notwendig und leisten einen wichtigen Beitrag zur Entwicklung des wissenschaftlichen Publizierens. Doch die Zukunft des wissenschaftlichen Publizierens in den einzelnen Fächern hängt letzten Endes allein davon ab, wie Forscher selbst publizieren. Mit Open Access für die Publikation von Primärdaten und Forschungsergebnissen haben Wissenschaftler wirksame Werkzeuge, nicht nur die Aufmerksamkeit der jeweiligen Fachöffentlichkeit in hohem Maße auf ihre Arbeiten zu lenken, sondern die Forschung zugleich durch den leichten und unmittelbaren Zugang zu ihren Publikationen entscheidend zu befördern.

Angaben zum Autor

Dr. Johannes Fournier
Deutsche Forschungsgemeinschaft
Kennedyallee 40
D-53175 Bonn

Studium der Germanistik, Geschichte und Philosophie. Promotion zum Dr. phil. (Altgermanistik) an der Universität Trier 1997. „Kompetenzzentrum für elektronisches Publizieren in den Geisteswissenschaften“ an der Universität Trier 1997-2003. Seit August 2003 Referent in der Gruppe „Wissenschaftliche Literaturversorgungs- und Informationssysteme“ der DFG und dort seit Mai 2005 Programmdirektor für das Förderprogramm „Elektronisches Publizieren“.

Verzeichnis der Autorinnen und Autoren

Christiane Baier

DV-Einsatz und Anwendung, Zentrum für psychologische Information und Dokumentation (ZPID), Universitätsring 15, 54296 Trier, Postanschrift: 54286 Trier

Arne Ballay

Informationszentrum Sozialwissenschaften (IZ), Abt. Forschung und Entwicklung, Lennéstraße 30, 53113 Bonn
E-Mail: ballay@bonn.iz-soz.de

Doris Bambey

Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstr. 29, 60468 Frankfurt/M.

Jochen Brüning

Universität Konstanz, FB Informatik und Informationswissenschaft, Universitätsstr. 10, 78464 Konstanz
E-Mail: jochen.bruening@uni-konstanz.de

Dr. Christopher N. Carlson

IWF Wissen und Medien gGmbH, Nonnenstieg 72, 37075 Göttingen

Patrick Danowski

Zentral- und Landesbibliothek Berlin, Breite Str. 30-36, Berlin

Prof. Dr. Gregor Fels

Universität Paderborn, Fakultät für Naturwissenschaften, Department Chemie, Warburgerstr. 100, 33098 Paderborn

Dr. Johannes Fournier

Deutsche Forschungsgemeinschaft, Kennedyallee 40, 53175 Bonn

Gerd Hansen

Max-Planck-Institut für Geistiges Eigentum, Marstallplatz 1, 80539 München

Ulrich Herb

Saarländische Universitäts- und Landesbibliothek, Postfach 15 11 4, 66041 Saarbrücken

Anke Susanne Hoffmann

<RM.net> - Informationsnetzwerk zur Geschichte des Rhein-Maas-Raumes, Universität Trier, Fachbereich III / Geschichte, 54286 Trier

Prof. Dr. Stefan Hornbostel

Humboldt-Universität zu Berlin, Philosophische Fakultät III, Institut für Sozialwissenschaften, Unter den Linden 6, 10099 Berlin

Stefan Jakowatz

Informationszentrum Sozialwissenschaften, Lennéstr. 30, 53113 Bonn

Dr. Jochen Johannsen, MA (LIS)

Technische Universität Berlin, Universitätsbibliothek im VOLKSWAGEN-Haus, Fasanenstr. 88, 10623 Berlin

Dr. Marco Jorio

Chefredaktor Historisches Lexikon der Schweiz (HLS), Historisches Lexikon der Schweiz (HLS), Hirschengraben 11, Postfach 6576, 3001 Bern, Schweiz

Dr. Stephan Körnig

TU Darmstadt, Institut für Philosophie, Schloß, 64283 Darmstadt

Prof. Dr. Rainer Kuhlen

Universität Konstanz, FB Informatik und Informationswissenschaft, Universitätsstr. 10, 78464 Konstanz
E-Mail: rainer.kuhlen@uni-konstanz.de

Philipp Mayr

Informationszentrum Sozialwissenschaften (IZ), Abt. Forschung und Entwicklung, Lennéstraße 30, 53113 Bonn
E-Mail: mayr@bonn.iz-soz.de

Dr. Gisela Minn

<RM.net> - Informationsnetzwerk zur Geschichte des Rhein-Maas-Raumes, Universität Trier, Fachbereich III / Geschichte, 54286 Trier

Dr. Alexis-Michel Mugabushaka

Deutsche Forschungsgemeinschaft, Kennedyallee 40, 53175 Bonn

Yvonne Rommelfanger, M.A.

<RM.net> - Informationsnetzwerk zur Geschichte des Rhein-Maas-Raumes, Universität Trier, Fachbereich III / Geschichte, 54286 Trier

Sandra Schaffert

Bildungsinstitut Martin Schön (BIMS), Bahnhofplatz 4
83278 Traunstein

Dr. Bernhard Schmidt

Lehrstuhl für Allgemeine Pädagogik und Bildungsforschung,
Ludwig-Maximilians-Universität, Leopoldstr. 13, 80802 München
E-Mail: b.schmidt@lmu.de

Dipl.-Bibl. (FH) Heike Schulte-Derne

Bibliothek der Fachhochschule Ludwigsburg – Hochschule für öffentliche
Verwaltung und Finanzen, Reuteallee 36, 71634 Ludwigsburg

Dr. Wolfram Sperber

Konrad-Zuse-Zentrum für Informationstechnik Berlin, Takustr. 7, 14195
Berlin

Dr. Maximilian Stempfhuber

Informationszentrum Sozialwissenschaften (IZ), Abt. Forschung und
Entwicklung, Lennéstraße 30, 53113 Bonn
E-Mail: stempfhuber@bonn.iz-soz.de

Anne-Kathrin Walter

Informationszentrum Sozialwissenschaften (IZ), Abt. Forschung und
Entwicklung, Lennéstraße 30, 53113 Bonn
E-Mail: walter@bonn.iz-soz.de

Peter Weiland

DV-Einsatz und Anwendung, Zentrum für psychologische Information und
Dokumentation (ZPID), Universitätsring 15, 54296 Trier, Postanschrift:
54286 Trier

Der Tagungsband zur 11. Jahrestagung der Initiative Information und Kommunikation wissenschaftlicher Fachgesellschaften (IuK) dokumentiert ihr Engagement für die Verbesserung der Informationsversorgung in den Wissenschaften, dient als Plattform für die Kommunikation innovativer und richtungsweisender Ideen und formuliert die Themen, zu denen innerhalb – und gerade auch zwischen – den deutschen Fachgesellschaften eine Standort- und Richtungsbestimmung notwendig erscheint.

Der Tagungsband ergänzt die Informationen zur Jahrestagung, die unter <http://www.iuk-initiative.org/> verfügbar sind.



InformationsZentrum
Sozialwissenschaften

der Arbeitsgemeinschaft
Sozialwissenschaftlicher Institute e.V.

Lennéstraße 30 • D-53113 Bonn
Telefon 02 28 / 22 81 - 0
Telefax 02 28 / 22 81 - 120

GESIS

Das IZ ist Mitglied der
Gesellschaft Sozialwissenschaftlicher
Infrastruktureinrichtungen e.V.

Die GESIS ist Mitglied der
Leibniz-Gemeinschaft.

ISBN 3-8206-0150-3
EUR 25,-