

gegen ein anderes auszutauschen. Stattdessen sollte man beide Regelwerke vereinfachen, die Datenformate angleichen und die Normdateien (authority files) zusammenfassen.

DIE VERFASSER

Annette Rath-Beckmann, Staats- und Universitätsbibliothek, Bibliotheksstr., 28359 Bremen, rathb@uni-bremen.de

Dr. Thomas Elsmann, Staats- und Universitätsbibliothek Bremen, Bibliotheksstr., 28359 Bremen, elsmann@uni-bremen.de

OAI

»CREATING A EUROPEAN FORUM FOR OPEN ARCHIVES ACTIVITIES« BERICHT DES WORKSHOPS VOM 13./14. MAI IN PISA

Offene Archive in Europa

Die Idee der »offenen Archive« erfreut sich zunehmenden Interesses nicht nur in der Welt der so genannten Preprint-Server, sondern auch in wissenschaftlichen, kulturellen und Bildungsbereichen. Insbesondere der Ansatz der Open Archives Initiative (OAI)¹ stellt eine alternative kostengünstige Technologie zur Verfügung, die es ermöglicht, auch bisher schwer zugängliche Ressourcen über das Internet zu erschließen und damit einem breiteren Publikum zur Verfügung zu stellen.

Die Veranstalter:

Das Europäische Projekt »Open Archives Forum«²

Das Projekt »Open Archives Forum«³ wird von der Europäischen Union im Rahmen des IST-Programms⁴ (Information Societies Technologies) gefördert. Im Unterschied zu anderen Projekten wird dieses Mal nicht vordergründig Software entwickelt, sondern ein bestehender Gedanke, der der Öffnung digitaler Ressourcen und Archive, vor allem durch die OAI, in der Verbreitung der Ideen und Technologien im europäischen Raum unterstützt.

Die Partner in diesem Projekt sind zum einen das United Kingdom Office of Library Networking (UKOLN)⁵ an der University of Bath, Großbritannien, das Istituto di Elaborazione della Informazione, Na-

tional Research Council (IEI-CNR), Italien⁶, und das Rechenzentrum der Humboldt-Universität zu Berlin⁷. UKOLN ist ein Institut, welches im Bereich der Bibliotheken Technologien entwickelt und Forschungsprojekte im Bereich Metadaten und digitale Bibliothek durchführt. Ansprechpartner sind dort Rachel Heery und Leona Carpenter. IEI-CNR ist ein Informatik-Forschungsinstitut, welches seit mehreren Jahren aktiv an der Entwicklung neuer Technologien im Bereich digitale Bibliotheken und Information-Retrieval arbeitet. Die Projektverantwortliche hier ist Donatella Castelli. An der Humboldt-Universität fungiert Susanne Dobratz als Ansprechpartner. Das Projekt ist auf eine Dauer von zwei Jahren angelegt.

Oft treten bei der Anwendung technischer Spezifikationen, so z.B. des OAI-Protokolls auch Probleme auf, welche die Standardisierungsorganisationen, wie die OAI selbst, nicht regulieren können. Hier ist es die Aufgabe von Fach- und Forschungsgemeinschaften digitaler Archive, eigene Metadatensätze oder auch eigene Anpassungen an das Protokoll für ihren spezifischen Bereich zu definieren und umzusetzen. So ist z.B. häufig auch die Frage zu stellen, welche Kategorieneinteilung für ein Archiv angeboten werden sollte. Auch gibt es Bereiche, in denen die Idee der weltweit über eine einfache Schnittstelle vernetzten digitalen Archive noch nicht Einzug gehalten hat; Bereiche, in denen spezielle Anforderungen an Archive und Metadaten gestellt werden, wie z. B. im Umfeld der Museumsinformation oder von Digitalisierungsprojekten.

In genau diese Lücke stößt nun das europäische Projekt »Open Archives Forum«. Nachdem die eigentliche Entwicklung der OAI-Spezifikation seit der Herausgabe der Version 2 des OAI-Protokolls⁸ seit Mai 2002 von der OAI selbst als abgeschlossen angesehen wird, ist es notwendig, unterstützende Maßnahmen zu organisieren und durchzuführen, um die Idee der vernetzten Archive und der so genannten »low-barrier-interoperability« weiter in die Bibliotheken, Archive, Museen und wissenschaftlichen Gemeinschaften hineinzutragen.⁹

Derartige Maßnahmen sind zum einen die Organisation von thematischen, auf spezielle Nutzergruppen abgestimmte Workshops und Studien, zum anderen aber auch der Aufbau eines WWW-Forums, in dem Informationen zu Protokollen, Metadatenschemata, Archiven, Software und Werkzeugen zur Installation verteilter Archive erfasst, ausgewertet und zur Verfügung gestellt werden, ein Forum, das zu virtuellen Diskussionen und Partnerschaften ermutigt.

Weitere Informationen zu dem Projekt erhalten Sie über die WWW-Seite des Projektes <http://>

IST-Programm

Problematik spezieller Anforderungen

Aufbau eines WWW-Forums

www.oaforum.org bzw. aus bereits existierenden Veröffentlichungen.^{10 11 12}

Ziele des Workshops in Pisa

Der Workshop »Creating a European Forum for Open Archives Activities«, der am 13. und 14. Mai 2002 am Istituto di Elaborazione della Informazione / Consiglio Nazionale delle Ricerche in Pisa stattfand, hatte zum Ziel, den Startschuss für ein europäisches Forum zum Thema »Offene Archive« im Allgemeinen und zur Bündelung europäischer Aktivitäten innerhalb der OAI im Speziellen zu geben. Dabei legte die Europäische Kommission einen besonderen Wert auf die Einbringung europäischer Interessen in diese internationalen Bewegungen in der Hoffnung, motivierend auf Projekte und Institutionen einwirken zu können, sodass diese ihre digitalen und digitalisierten kulturellen Ressourcen einer bereiteren Öffentlichkeit zugänglich machen.

So zielte der Workshop dann auch darauf ab, Wissenschaftler, technische Entwickler und Projektmanager zusammenzubringen, die bereits mit den technischen Ergebnissen der OAI experimentiert oder sich damit auseinander gesetzt haben, quasi bereits in einem gewissen Sinne Experten darstellen. Die Expertise der einzelnen, aus den unterschiedlichsten Anwendungsbereichen kommenden Teilnehmer führte zu einem intensiven Informationsaustausch zu technischen Lösungen, organisatorischen Problemen und Erfahrungen über die Herangehensweise bei Projekten aus den Bereichen digitale Bibliotheken und offene Archive.

Workshop und Diskussionen

Am Workshop nahmen etwas mehr als fünfzig Interessierte, die Projektpartner sowie sechs Vortragende teil. Schaut man sich die Verteilung innerhalb Europas an, so waren folgende Staaten präsent: Belgien, Frankreich, Deutschland, Italien, Niederlande, Norwegen, Portugal, Spanien, Schweden, Schweiz und Großbritannien.

Dabei unterschieden sich die Teilnehmer nicht nur durch ihre Nationalität, sondern auch durch ihre fachliche und institutionelle Zugehörigkeit. Archive, Forschungseinrichtungen, Universitäten, National- und Forschungsbibliotheken, Rechenzentren, aber auch kommerzielle Einrichtungen aus den Bereichen Landwirtschaft, Astrophysik, Informatik, Rechtswissenschaft und Medizin entsandten ihre Mitarbeiter.

Zu Beginn des Tages gaben die Veranstalter einen Überblick über die bestehenden OAI-Aktivitäten in Europa. So stellte sich heraus, dass die **Bundesrepublik Deutschland** die größte Anzahl OAI-kompatibler Archi-

europäische Interessen
stärken

ve besitzt. Dies sind zum einen Dokumentenserver der Universitätsbibliotheken wie der Humboldt-Universität zu Berlin, der SuUB Bremen, der TU Chemnitz, der Universität Dortmund, der Universität Duisburg und der SLUB Dresden, der Server des Bibliotheksservice-zentrums Baden-Württemberg als Zugang zu den digitalen Dokumenten des gesamten Bibliotheksverbundes, zum anderen aber auch fachspezifische Suchmaschinen, wie PhysDoc bzw. Medienserver wie TIMMS in Tübingen oder sogar Bibliothekskataloge wie der in Oldenburg von oder das von Prof. Bayer, Institut für Informatik der TU München, als Bibliothekssystem entwickelte SISIS/Elektra. Service-Provider im Sinne der OAI besitzen die Humboldt-Universität zu Berlin, die ZDV der Universität Tübingen, das BSZ Baden-Württemberg in Konstanz und das Institut für Science Networking (Prof. Hilf) der Universität Oldenburg mit dem Physnet-Search-Service. Das europäische Projekt CYCLADES¹³, an dem Prof. Fuhr in Dortmund mitarbeitet, war als deutsches Projekt ebenfalls vertreten. Andere Staaten, wie **Großbritannien** haben weniger aktive OAI-Archive zu verzeichnen, sind jedoch aktiver an den technischen Entwicklungen beteiligt. Hier ist zuerst die University Southampton zu nennen, deren Eprints.org-Software international genutzt wird. UKOLN setzt eine OAI-basierte Suchmaschine im Projekt Ressource Discovery Network (RDN)¹⁴ ein. Auch **Italien** ist eines der führenden Länder gemessen an OAI-Projekten und Implementationen. Mehrere universitäre Data-Provider in Bologna, Florenz, Padua, Trient, Mailand und zwei europäische Projekte, CYCLADES und TORII (Tools for Innovative Publishing in Science) am SISSA-Institut in Triest, die auch Service-Provider bereitstellen, bestätigen dies. In den **Niederlanden** beschäftigt sich vor allem das ARNO Projekt¹⁵, durchgeführt von den Universitäten Tilburg, Amsterdam und Twente mit dem Aufbau eines OAI-Data- und Service-Providers. Im Projekt Roquade¹⁶ der Universitäten Utrecht, Delft und dem NiWi Institut (Netherlands Institute for Scientific Information Services) wird ebenfalls ein institutioneller OAI-Server aufgebaut. In **Belgien** gibt es bisher zwei Data-Provider, an der Université Catholique de Louvain und der Université Libre de Bruxelles. Die Université Catholique de Louvain betreibt ebenfalls einen Service-Provider. Einige **französische** Bibliotheken (wie Lyon ²¹⁷) experimentieren mit dem OAI-Ansatz, wobei es dort keine Koordination der Experimente gibt. Registriert sind einige Data- und Service-Provider bei der OAI: ATILF Resources, CCSD theses-EN-ligne, Institut National de Physique Nucleaire et de Physique de Particules und das Lacito Archive¹⁸. In **Schweden** ist ein neues Projekt zum Thema Vernetzung digitaler Archive in Planung, die Universität Uppsala betreibt ein OAI-

Archiv, welches auch über einen universitätsinternen Service-Provider bedient werden kann. In **Norwegen** wird mit der Bibliothekssoftware BIBSYS unter Führung der Norwegischen Nationalbibliothek experimentiert. In der **Schweiz** ist es das CERN in Genf, welches einen OAI-kompatiblen Dokumentenserver betreibt. Nicht alle diese Projekte, Data- oder Service-Provider registrieren sich in der OAI.

Protokolle

Nach einer kurzen Einführung in den Workshop und das dahinter stehende Projekt durch Donatella Castelli und Leona Carpenter gab **Michael Nelson** (Old Dominion University, USA), ein Vertreter des technischen Komitees der OAI, einen Überblick über die aktuellen Entwicklungen des OAI-Protokolls der Version 2. Dabei betonte er, dass die Version OAI 2.0 zwar auf den älteren Versionen basiert, jedoch nicht rückwärtskompatibel zur Version 1.1 sei. Allerdings gibt es einen Migrationszeitraum, in dem den OAI-Repositoryn die Möglichkeit eingeräumt wird, zu der neuen Version zu migrieren, und damit bleiben diese dann auch bis Dezember 2002 bei der OAI registriert. Die Notwendigkeit dieses Schrittes ergab sich aus der Absicht, so Michael Nelson, die Version 2.0 zu einem stabilen Protokoll zu machen. Die Grundphilosophie des OAI-Protokolls sei jedoch erhalten geblieben: es ist HTTP-basiert, enthält sechs Anfragetypen, nutzt Dublin Core als Minimalmetadatensatz, der über eine XML-Kodierung ausgetauscht wird. Nach einer detaillierten Darstellung des Protokolls ging Michael Nelson kurz auf die Alpha-Test-Phase ein, an der neben den US-amerikanischen Testern auch eine Reihe europäischer Aktiver teilgenommen haben, so die British Library, CERN, UKOLN, Danmarks Tekniske Videncenter, University of Southampton und die Humboldt-Universität zu Berlin. Die Vision des OAI Steering Committee ist es, dass die bei der OAI registrierten Repositorien von 77 (Mai 2002) auf eine Zahl anwachsen, die eine Registrierung nicht mehr notwendig macht, da, ähnlich wie beim HTTP-Protokoll in Zukunft sowieso jeder Server, jedes Archiv das OAI-Protokoll unterstützen wird.

Fabio Asnicar (International School for Advanced Studies, Italien) berichtete über TORII¹⁹; ein System, welches den direkten Zugang zu digitalen Forschungscommunities ermöglicht. Dabei ist es nicht nur wichtig, so genannte Repositorien mit ihren Informationen, Archiven und Datenbanken anzubieten, sondern auch entsprechende Dienstleistungen, die darüber gelegt werden können, wie z.B. Begutachtungsprozesse für wissenschaftliche Zeitschriften, Cross-Zitierung oder eben digitale Communities, die sich durch Synergieeffekte aus den Dienstleistungen

große Vielfalt an Projekten

OAI-Protokoll Version 2.0

TORII-Zugang zu digitalen Forschungscommunities

und den angebotenen Informationen herausbilden. TORII stellt dafür eine Plattform mit einer interaktiven Benutzerschnittstelle zur Verfügung, die derartige Dienste ermöglichen soll. Angelehnt an die Funktionalität von Groupwaresystemen kann der Nutzer seine Dokumente in Ordnern ablegen, um sie später erneut zu nutzen oder anderen zur Nutzung zu überlassen. Er kann in einem virtuellen Dokumentenbestand suchen und sich die Suchergebnisse in seinem Ordner abspeichern. Die Hauptfunktionalität der Integration dieser dynamischen Prozesse wird über eine OAI-Schnittstelle und einen speziellen XML-kodierten Metadatensatz realisiert. Eine Suchmaschine namens Okapi, die in der Lage ist, die Relevanz von Retrievalergebnissen einer Volltextrecherche über probabilistische Modelle auszuwerten, wird direkt über die TORII-Software angebunden. Eine weitere in TORII integrierte Software ist iCite, ein Werkzeug, mit dessen Hilfe Zitate aus den zum Archiv gehörenden Dokumenten extrahiert und beurteilt werden können, um Dokumente entsprechend ihrer Zitierhäufigkeit bewerten zu können.

Im Vortrag von **Françoise Genova** (Strasbourg Astronomical Data Center, Frankreich) kam bei der Vorstellung des WWW-Portals des Straßburger Astronomischen Datenzentrums zum Ausdruck, dass der Ansatz der OAI, so wie er heute existiert, nicht uneingeschränkt auf jede Art der digitalen Datensammlung angewendet werden kann. Die Astrophysik bildet dafür ein gutes Beispiel, da hier der Bedarf an speziellen Dienstleistungen bereits vor einiger Zeit die Etablierung von De-facto-Standards hervorgebracht hat, die sich nun schwer in OAI-Infrastrukturen einpassen lassen. Einige Abbildungen der genutzten De-facto-Standards auf echte Standards werden derzeit von der Community ausprobiert, da man sich dadurch einen größeren Zugriff und Nutzung der Dokumente erhofft. Bisher existieren jedoch noch keine festen Ergebnisse.

Der Vortrag von **Steve Hitchcock** (Southampton University, Großbritannien) »Revealing a new Dynamic: Interaction in an Open Access Archive« gab eine Einführung in die Überlegungen einer E-Prints-Community. Da es so genannte E-Prints bereits lange vor der Gründung der OAI-Initiative gab und die OAI in ihren Anfängen auch zuallererst das Ziel verfolgt hat, den Zugang zu E-Prints zu fördern, zielte sein Vortrag vor allem auf die Entwicklungen des Open-Citation-Projektes ab. Er stellte die dort entwickelte Software, die sowohl OAI als auch die BOAI (Budapest Open Access Initiative)²⁰ unterstützt vor. Unter Zuhilfenahme von Daten aus dem Open-Citation-Projekt demonstrierte Steve Hitchcock, wie »open access« im Sinne der freien, uneingeschränkten Zugänglichkeit zu Volltext-

ten und eine schnellere Interaktion zwischen Nutzern dem WWW und Publikationen eine neue Dynamik verleihen kann.

Umberto Straccia (IEI-CNR, Italien) berichtete über das Projekt Cyclades (Open Collaborative Virtual Archive Environment), dessen Hauptziel in der Entwicklung einer virtuellen Umgebung für ein kooperatives Arbeiten, einer Groupware, liegt. Die Basis dafür bildet das am Fraunhofer (IPSI) entwickelte BSCW-System, das um eine OAI-Schnittstelle für das Sammeln, Indexieren und Speichern von Informationen in einer lokalen Datenbank sowie um einen Mechanismus erweitert wurde, der es ermöglicht, Informationen und digitale Sammlungen von Objekten und Informationen dynamisch zu strukturieren und über eine nutzerspezifische Sicht zu nutzer-eigenen virtuellen Sammlungen zusammenzustellen.

Martin Veseley (CERN, Schweiz) gab einen Eindruck von der Implementierung des OAI-PMH am CERN-Dokument-Server. Dieser umfasst ca. 550.000 Metadatensätze. Für das CERN sei das Handling und der Dokumentenaustausch sehr wichtig, die Anwendung des OAI-Protokolls mit der Unterscheidung in Data-Provider und in Service-Provider kommt der dort arbeitenden Preprint-Community sehr entgegen, insbesondere durch seine geringen Anpassungskosten an bestehende Softwarelösungen und dadurch, dass es damit möglich wird, relativ einfach über syntaktisch heterogene Infrastrukturen Dienstleistungen zu erreichen. Die Hauptdienstleistung des CERN-Dokument-Server bildet die »Internet«-Suchmaschine für Partikelphysik und verwandter Gebiete. Die CERN-Dokument-Server-Software (CDSware) steht interessierten Nachnutzern zur Verfügung. Am Ende seines Vortrages wurde über einige Einzelheiten des OAI-Protokolls diskutiert, so z.B. darüber, auf welcher Ebene eine benutzerorientierte Anpassung des Protokolls an spezielle, communityspezifische Bedürfnisse erfolgen sollte, auf der Ebene der Anwendungssoftware oder auf der Protokollebene selbst. Diese Problembereiche umfassen die Flusskontrolle für HTTP-Fehlermeldungen, die semantische Heterogenität von inhaltlichen Einteilungen (Sets) für Repositorien und den Informationsverlust bei ausschließlicher Unterstützung des OAI-Dublin-Core-Metadatenformates als einziges Metadatenformat.

Neben den Vorträgen zu existierenden Standards, Projekten und Problemen wurde in insgesamt vier Diskussionsgruppen über unterschiedliche Themen nachgedacht.

Im Arbeitskreis **Organisational Validation** wurde diskutiert, was man aus den Erfahrungen der OAI-Erstanwender bereits lernen kann, ob es spezielle Ar-

gumente für oder wider den Aufbau eher fachspezifischer Repositorien oder institutioneller digitaler Sammlungen gibt und ob und wie die Nutzung eines einfachen Dublin-Core-Konzeptes innerhalb der technischen Spezifikation der OAI erweitert werden kann.

Die Diskussionsrunde **Glossary** besprach welche Begrifflichkeiten aus dem Bereich der digitalen Bibliotheken noch der Klärung bedürfen, da sie in unterschiedlichen Sprachen und Fachcommunities eine andere Bedeutung haben, und welche Definition in ein Glossar aufgenommen werden sollte.

Thema in der Parallelsitzung **Communities and Services** war die Frage, wer von den Ansätzen zur Öffnung digitaler Ressourcen profitiert und wie, ob und welche unterschiedlichen Anreize sowie technische und organisatorische Herausforderungen in den verschiedenen Bereichen existieren, welche Dienstleistungen auf der Basis des OAI-Protokolls möglich sein können und welche Erfahrungen dabei bisher von Erst Anwendern gemacht werden konnten.

Im Teil **Technical Validation** wurde besprochen, welche Software-Werkzeuge genutzt werden können, um ein Archiv im Internet zu öffnen und in welche technischen Lösungen und Infrastrukturen neue Ansätze, wie der der OAI einzupassen sind. Sind diese neuen Ansätze wirklich besser, wo liegen die Vor- und Nachteile? An welcher Stelle müssen community-spezifische Standardisierungen erarbeitet werden, wie können diese so verwaltet werden, dass Interessierte auf bereits existierende Software, fachspezifisch angepasste OAI-Metadatenstandards zugreifen oder Kooperationspartner finden können? Hier wurde auch seitens der Projektpartner an der Humboldt-Universität zu Berlin die Auswertung eines Fragebogens²¹ zur Nutzung von OAI vorgestellt.

Die Folien zu den Vorträgen und Abstracts finden sie unter:
www.oaforum.org/workshops/pisa_programme.php

Weitere Planungen

Geplant sind weitere Workshops, die jeweils auf unterschiedliche Themen und Nutzergruppen zugeschnitten sind:

- »Gaining Access to Hidden Resources« five. 7. Dezember 2002 in Lissabon. Hier werden vor allem Teilnehmer aus dem Bereich der Archive erwartet, um über Applikationen, Bedingungen und Erfahrungen bei der Öffnung konventioneller Archive über das Internet zu diskutieren.

- »Museums and Open Archives« Ende März 2003 in Berlin zielt darauf ab, mit Experten aus dem Bereich der Museen und Kunstarchive die Frage der Öffnung und Interoperabilität internetbasierter offener Archive zu erörtern.
- »OAI and Library Community«, wird im Spätsommer 2003 voraussichtlich in Bath stattfinden. Hier soll speziell auf die Probleme und Ansätze der Bibliotheken und E-Print-Communities eingegangen werden.

¹ www.openarchives.org

² www.oaforum.org

³ IST-2001-320015

⁴ Siehe www.cordis.lu/ist/

⁵ Siehe www.ukoln.ac.uk

⁶ Siehe www.iei.pi.cnr.it/

⁷ www.hu-berlin.de/rz

⁸ Open Archives Initiative Protocol for Metadata Harvesting – Version 2.2002, www.openarchives.org/OAI/2.0/openarchivesprotocol.htm (zuletzt besucht: 17.07.2002)

⁹ Lagoze, C. & Van de Sompel, H.: The Open Archives Initiative: Building a low-barrier interoperability framework. 2001, Vortrag auf der JCDL2001, www.openarchives.org/documents/oai.pdf (zuletzt besucht: 17.07.2002)

¹⁰ Susanne Dobratz, Friederike Schimmelpfennig and Peter Schirmbacher: »The Open Archives Forum«, in: ARIADNE, 31, 2001, www.ariadne.ac.uk/issue31/open-archives-forum/intro.html (zuletzt besucht: 17.07.2002)

¹¹ Susanne Dobratz, Birgit Matthaer and Dr. Peter Schirmbacher: »Open Archives Forum (OAF)« DLIB Magazine vol. 8 no. 2, February 2002 www.dlib.org/dlib/february02/ozinbrief.html#DOBRTATZ (zuletzt besucht: 17.07.2002)

¹² Donatella Castelli: »Open Archives Forum«, ERCIM News No. 48, January 2002, www.ercim.org/publication/Ercim_News/enw48/castelli.html (zuletzt besucht: 17.07.2002)

¹³ www.ercim.org/cyclades/

¹⁴ www.rdn.ac.uk/

¹⁵ <http://cf.uba.uva.nl/en/projects/arno/>

¹⁶ www.roquade.nl/

¹⁷ <http://eprints.univ-lyon2.fr>

¹⁸ <http://lacito.archivage.vjf.cnrs.fr/>

¹⁹ <http://tips.sissa.it/docs/booklet.pdf>

²⁰ www.soros.org/openaccess/index.shtml

²¹ Die Ergebnisse sind nachlesbar unter www.oaforum.org/questionnaire.php.

DIE VERFASSERIN

Susanne Dobratz, Humboldt-Universität zu Berlin, Universitätsbibliothek, AG Elektronisches Publizieren, 10099 Berlin,
susanne.dobratz@rz.hu-berlin.de