

## Die Open Archives Initiative (OAI)

### Neue Zugangsform zu wissenschaftlichen Arbeiten?



*Nach einer Erläuterung der Ereignisse und Gründe, die zur Entwicklung der Open Archives Initiative (OAI) geführt haben, wird die Entstehungsgeschichte der OAI sowie ihre Organisationsstruktur dargestellt. Der Input und Diskussionen von Interessierten und Erstanwendern bei mehreren öffentlichen Workshops trugen zur Optimierung der Alpha-Version des Protokolls bei, mit dem Ergebnis, dass die erste offizielle Version im Januar 2001 vorgestellt wurde. Darauf hin folgten zwei Workshops in den USA und Europa zur Darlegung des Protokolls und Vorstellung exemplarischer Implementierung des Protokolls. Das OAI Metadata Harvesting Protokoll wird in seiner Zielsetzung und Möglichkeiten detailliert dargestellt. Auch die vom Protokoll erforderlichen Metadaten werden erläutert. Die Basisbefehle zur Abfrage in verschiedenen OAI-kompatiblen Repositories, auch mit Unterscheidung zwischen besonderen Sammlungen innerhalb der Repositories werden erklärt. Zum Schluß werden die Aussichten für die OAI im Rahmen der Experimentierphase der 1. Version des Protokolls sowie im Rahmen der gegenwärtigen Entwicklungen im wissenschaftlichen Publikationswesen diskutiert.*

The Open Archives Initiative – New acces to scientific publications?

*After describing the events and reasons which led to the development of the Open Archives Initiative (OAI), the history of the OAI and its organizational structure are delineated. The input and discussions among interested persons and early adopters contributed to optimizing the alpha-version of the OAI Protocol with the result that the first official version of the Protocol was introduced in January 2001. This was followed by two Workshops in the United States and in Europe to present the Protocol and introduce exemplary implementations of the Protocol. The OAI Metadata Harvesting Protocol is then described in detail, giving its mission and possibilities. The OAI-compatible metadata needed for using the Protocol are delineated. The basic commands for requests to be sent to the OAI compatible repositories, also with distinction between sets within repositories, are explained. Finally, the future possibilities for the OAI are discussed within the context of the current experimental phase for the first version of the OAI Metadata Harvesting Protocol as well as in context of the current developments within the field of scientific publications.*

Open Archives Initiative (OAI) – Accès nouveau aux travaux scientifiques?

*Tout d'abord l'auteur décrit les incidents et les causes qui menaient à la naissance de Open Archives Initiative. Puis elle présente l'historique et la structure organisatrice de OAI. A l'aide des personnes intéressées et des applicateurs de la version alpha, qui contribuaient avec leurs idées et discussions à perfectionner le protocole, il était possible de présenter en janvier 2001 la première version officielle du protocole OAI. Les deux colloques suivants aux Etats-Unis et en Europe visaient à présenter le protocole et à montrer plusieurs implémentations exemplaires. L'auteur décrit en détail le OAI Metadata Harvesting Protocol montrant ses objectifs et ses possibilités. Elle décrit de même les métadonnées dont on a besoin. Elle explique les commandes fondamentales d'une recherche dans les repositories qui sont compatibles avec OAI vu la distinction entre plusieurs fonds dans un repository. Finalement l'auteur discute les perspectives de OAI dans la phase expérimentale de sa première version et de même dans le cadre du développement actuel dans le domaine des publications scientifiques.*

### Einführung

Die Open Archives Initiative ist ein Ergebnis von drei Entwicklungen Anfang der 90er Jahre, aus denen sie auch ihre Zielsetzung definiert hat.

1. Die Wurzeln der Open Archives Initiative liegen in den Bemühungen, Zugang zu den neuesten Forschungsergebnissen zu verschaffen, indem eingereichte Artikel und Abhandlungen der wissenschaftlichen Community über Preprint-Server schnell und ohne Zugangsbeschränkungen zur Verfügung gestellt wurden.

2. Die Forschungscommunity hatte schnell nach Öffnung des Internets für die Universitäten erkannt, welche Möglichkeiten der digitale Austausch von Daten für die

Informationsversorgung aller Bereiche von Forschung und Lehre geben würde. So entstand die Forschungsinitiative „Digital Libraries Initiative“ der US-National Science Foundation (1994-1998, Phase II 1998-2001), die Förderprogramme zur Entwicklung nicht nur von digitalen Bibliotheken sondern von vielen weiteren innovativen Einrichtungen für die schnellere Verbreitung von Inhalten und neuen Informationsformen erarbeitete. Aus diesen Initiativen kamen die Erkenntnisse der Bedeutung der Interoperabilität im technischen, formalen und organisatorischen Sinne sowie der Bedeutung von Standardisierungen bei Formaten, Metadaten und Datenaustausch- und Transportprotokollen.

3. Forscher und Autoren von Zeitschriftenartikeln wurden zunehmend unzufrieden mit der zeitlichen Verzögerung

rung zwischen Abgabe eines eingereichten Aufsatzes bis zur Veröffentlichung, mit den hohen Zeitschriftenpreisen, die viele Bibliotheken veranlasst haben, Abonnements abzubestellen und Bestände zu reduzieren, sowie mit der üblichen Übertragung der Autorenrechte an den Verlag, der dann unter anderem das alleinige Vertriebsrecht hatte und durch seine Publikations- und Preispolitik die Verbreitung seiner Forschungsergebnisse einschränken konnte.

So entstand eine „grass-roots“-artige demokratische Bewegung unter den Autoren und Wissenschaftlern, die Preprint-Server aufbauten, wie denjenigen für Hochenergiephysikaufsätze beim Los Alamos National Laboratory von Paul Ginsparg. Die Gefahren und Risiken für die Zeitschriftenverleger und die Autoren, deren berufliche Weiterentwicklung oft von den Veröffentlichungs- und Zitierungsraten in den hochwirksamen wissenschaftlichen Zeitschriftentiteln (mit hohem „Impactfaktor“) und vom Peer-Review-System abhing, führten jedoch dazu, dass manche Wissenschaftler oder sogar ganze Wissenschaftsgebiete, wie z.B. bis in die jüngste Zeit die Psychologie, sehr zögerten, ihre Aufsätze vor Veröffentlichung in einer Peer-reviewed-Zeitschrift auf einen Eprint-Server zu legen, oder dass ihnen sogar durch den Verlag verboten wurde, den Aufsatz in irgendeiner Form vor dem Druck in der akzeptierenden Zeitschrift zu „Ver-Öffentlich-en“<sup>1</sup>.

So hat die Open Archives Initiative eine inhaltsunabhängige Aufgabe, nämlich einen Modus zu finden und die operationellen Tools dafür zu entwickeln, einzelne Server mit elektronischen Informationsmedien aus der wissenschaftlichen Community recherchierfähig zu machen und die Ergebnisse einer Recherche sinnvoll für wissenschaftliche Zwecke darzustellen. Im Folgenden werden die Entstehungsgeschichte, die Organisationsstruktur, die teilnehmenden Institutionen, die Spezifizierungen für das erste Protokoll inkl. der OAI Empfehlungen für Metadaten, Darstellungen der ersten Anwender bei den OAI-Tagen in den USA und in Europa, sowie einen Ausblick für die Zukunft dargelegt.

Um einige Begrifflichkeiten klarer zu stellen, werden sie unten in einem Glossar in der Bedeutung, die für die OAI genutzt wird, erklärt.

## Entstehungsgeschichte der OAI<sup>2</sup>

Juli 1999 ging von Paul Ginsparg, Rick Luce und Herbert Van de Sompel<sup>3</sup> eine gezielte Einladung an Leiter von Preprint-Servern in den USA und andere Projektleiter bedeutender digitaler Bibliotheksprojekte zu einer Diskussionsrunde über die Kooperations- und Entwicklungsmöglichkeiten dieser Art von Veröffentlichungs- und Selbstarchivierungsmethoden für wissenschaftliche Forschungsergebnisse<sup>4</sup>. Das Treffen fand am 21.-22. Oktober 1999 in Santa Fe im Bundesstaat New Mexiko (USA) mit der Unterstützung mehrerer richtungsgebender bibliothekarischer Fachgesellschaften und Veröffentlichungsorgane statt<sup>5</sup>. 26 Wissenschaftler, Informatiker und führende Bibliothekare haben am Treffen teilgenommen – alle eingeladenen Institutionen außer einer hatten einen Vertreter zum Treffen geschickt. Ziel des Treffens war es, die technischen und organisatorischen Voraussetzungen für einen Service zu schaffen, der die Metadaten aller Preprint-Server, fachlich unabhängig von einander, in ein virtuelles, universelles Preprint-Sy-

stem einbindet und dabei ein neues wissenschaftliches Kommunikationsmodell zusammenführt. Aus diesem Treffen, moderiert von Cliff Lynch (Coalition for Networked Information – CNI) und Don Waters (Digital Library Federation – DLF), kamen zwei wichtige Entscheidungen: Es sollte zum Zweck der allgemeinen Kommunikationen zwischen den Preprint-/Eprint-Servern

1. ein gemeinsamer Satz von Metadaten angesammelt werden, der relativ einfach und handhabbar ist, und
2. ein Navigationsprotokoll entwickelt werden, das auf der Basis eines Dienstprotokolls mittels dieses Metadatensatzes die wichtigsten Informationen aus den Preprint-Servern einsammelt und für eine Recherche in einen Harvesterindex zusammenstellt. Dieser Service wurde zunächst „Universal Preprint Service“ und die Vereinbarung, die sich aus diesem Treffen ergeben hat, die „Santa Fe Convention“<sup>6</sup> genannt und umfasste sowohl organisatorische Prinzipien als auch technische Bestimmungen, um die Funktionalität des Austausch und der Kommunikationen zwischen den Preprint-Servern zu gewährleisten. Innerhalb von wenigen Monaten entstand das erste Protokoll, das diese Interoperabilität in Alpha-Test-Format für die Frühanwender anbot. Während dieser Zeit entwickelte sich der Name zur „Open Archives Initiative“.

- 1 Stevan Harnad, Kognitionspsychologe und Verfechter des Selbstarchivierungssystems und Eprint-Server, hatte diese Prozedur anhand seiner Auflösung der englischen Entsprechung dargestellt: PUBLIC-ation, d.h. etwas öffentlich (publik) machen. Siehe Stevan Harnad: Scholarly Skywriting and Prepublication Continuum of Scientific Inquiry. In: Psychological Science 1, 6 (1990) 342-344 (<<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.skywriting.html>>); Stevan Harnad: The Future of Scholarly Skywriting. In: i in the Sky: Visions of the information future. Hrsg. v. A. Scommel, ASLIB Nov 1999 (<<http://www.cogsci.soton.ac.uk/~harnad/papers/harnad99.aslib.html>>); Stevan Harnad: Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge. In: The Public Access Computer Systems Review 2, 1 (1991) 39-53. Harnad und Andrew Odlyzko, Mathematiker, haben beide einiges zu diesem Thema des sich im Wandel befindlichen Veröffentlichungssystems geschrieben.
- 2 Ein zusammenfassender Überblick über die Entstehungsgeschichte der OAI ist über die URL <[http://www.openarchives.org/meeting\\_history.htm](http://www.openarchives.org/meeting_history.htm)> einzusehen.
- 3 Seinerzeit waren alle am Los Alamos National Laboratory (LANL): Paul Ginsparg ist Physiker, der 1991 den LANL Preprint Server (nun arXiv.org) gegründet hat; Rick Luce ist Bibliotheksdirektor der Forschungsbibliothek der Los Alamos National Laboratory (LANL), Los Alamos, New Mexiko/USA, und Herbert Van de Sompel hatte während seiner Doktorarbeit am Ausbau und der Weiterentwicklung des LANL-Servers bis Juli 2000 gearbeitet.
- 4 Paul Ginsparg, Rick Luce, and Herbert Van de Sompel: „Call for participation in the UPS initiative aimed at the further promotion of author self-archived solutions“ July 1999. <<http://www.openarchives.org/ups-invitation-ori.html>>.
- 5 Unter den Sponsoren des Treffens befanden sich die Council on Library and Information Resources (CLIR), die Digital Library Federation (DLF), die Scholarly Publishing & Academic Resources Coalition (SPARC), die Association of Research Libraries (ARL) und Los Alamos National Laboratory (LANL).
- 6 Herbert Van de Sompel, Carl Lagoze: The Santa Fe Convention of the Open Archives Initiative. In: D-Lib Magazine February 2000, 6, 2 <<http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>>.

Der Name „Open Archives Initiative“ ist nicht ohne eine gewisse Konfliktrichtigkeit. Das Wort „Open“ (zu deutsch „offen“) hatte ursprünglich die Bedeutung, dass diese Preprint-Server frei zugänglich waren sowie auch kostenfreie Möglichkeiten zum Zugriff auf die wissenschaftlichen Forschungsergebnisse boten und nicht gebunden an einem Zeitschriftenpreis waren. Es wurde jedoch den Gründern der Open Archives Initiative klar, dass die OAI nicht über die individuelle Serverpolitik und über Kostenfragen bestimmen wollten oder konnten. „Offen“ hat hier die Bedeutung von „Offenheit“ im Rahmen der technischen Architektur des Protokolls, das den Zugang zu verschiedenen Inhalten diverser Anbieter bzw. zu ihren Metadaten gewährleistet. Es hat nicht mehr den Sinn, dass der Zugang zu elektronischen Dokumenten oder Dienstleistungen kostenfrei ist. Dass langfristig solche Dienste von manchen Stellen nicht umsonst angeboten werden könnten, sollte nicht dazu führen, dass sie als Teil des Angebots, das über das OAI-Protokoll abgesucht wird, ausbleiben sollten. Dies würde den Sinn der Open Archives Initiative, wissenschaftliche Informationen besser zugänglich zu machen, widersprechen. So sind OAI-kompatible Web-Server offen im Sinne, dass sie die wichtigsten Informationen über die Inhalte der darin gespeicherten Dokumente und andere Informationsressourcen durch standardisierte Metadatenformate offen für den Harvestingdienst und Indexaufbau zwecks Recherche in diesen Preprint/Eprint Servern zur Verfügung stellen<sup>7</sup>.

Aber auch das Wort „Archives“ hatte zunächst für Verwirrung gesorgt. Die Preprint- mittlerweile auch Eprint-Server hatten in gewisser Weise den „Selbstarchivierungs-Zweck“, weil die meisten Autoren dies als eine Art Archivversion des an eine Zeitschriftenredaktion abgegebenen Aufsatzes ansahen und so eventuell vor einer allgemeinen Verbreitung geschützt werden konnten. Durch die langen Veröffentlichungszeiten, die sich oft auf ca. 2 Jahre erstreckten, aber auch hin und wieder bis 3 Jahre dauerten, wollten die Autoren eine Möglichkeit haben, ihre Ergebnisse und Ideen möglichst schnell zu veröffentlichen, um nicht Gefahr zu laufen, dass andere ihre Ideen „klauen“ bzw. als „erster“ öffentlicher Anmelder der Erkenntnisse in der wissenschaftlichen Welt durch die Publikation der Ergebnisse bekannt wurden. So funktionierten die sogenannten „Preprint“-Server gleich als Eprint-Archiv-Server, indem die dort archivierten Aufsätze und Berichte datiert, zitiert und versioniert wurden. In gewisser Weise fungierte mancher dieser Archivserver auch als Gewährleistung des freien Zugangs zu demselben oder demselben mit lediglich geringen Abweichungen von einem veröffentlichten Text in einer kauf- oder lizenzierungspflichtigen Zeitschrift. Der Nachteil bzw. der Zwiespalt für den geographisch und/oder wirtschaftlich benachteiligten Interessenten, Studenten oder Forscher an diesem Inhalt ist wohl jedem Leser klar. So kam es – trotz der langen Diskussionen über mögliche Mißverständnisse, die aus dem Namen erwachsen sind – dazu, dass der Name in die „Open Archives Initiative“ übergang.

Zwei Workshops zur Vorstellung der Ziele der Open Archives Initiative wurden im Juni und September 2000 im Rahmen zweier Tagungen zu digitalen Bibliotheken durchgeführt: die ACM-Jahrestagung im Juni in San Antonio und im September an der European Conference on Digital Libraries (ECDL) in Lissabon. In beiden Work-

shops gab es nach der Vorstellung der Ziele und der derzeitigen Arbeit am OAI-Metadatenatz und am OAI-Protokoll die Möglichkeit für Fragestellungen. Einige Vertreter der Erstanwender waren bei beiden Workshops und konnten zu der Implementierung des OAI-Protokolls Fragen beantworten. In beiden Workshops war der Ertrag der Diskussionen eine größere Klarheit in der Digitalen Bibliothekscommunity über die Grenzen und Möglichkeiten der OAI. Diese Workshops dienten auch den Veranstaltern und den Organisatoren des technischen Komitees der OAI dazu, sich über ähnliche Projekte und mögliche Kooperationspartner zu informieren. Daraus entstand die OAI-Mailingliste *oai-general*. Im August 2000 wurde eine erste finanzielle Unterstützung durch die Coalition for Networked Information (CNI) und die Digital Libraries Federation (DLF) in den USA gesichert. Die Arbeit am Protokoll war schon teilweise finanziert durch Fördermittel von der US-National Science Foundation und der Defense Advanced Research Project Agency (DARPA). So konnten die ersten formalen Strukturen für die Open Archives Initiative als Organisation geschmiedet werden, indem ein Steering Committee (s.u.) einberufen wurde. Gleichzeitig wurden Anfang September 2000 ca. 18 Informatiker und technische Spezialisten aus der Erstanwendergruppe zu einem Arbeitstreffen an die Cornell Universität (Ithaca, New York) eingeladen, um die Voraussetzungen und Bestimmungen für die Rahmenarchitektur der OAI und dessen Protokoll auszuarbeiten. Die Ergebnisse dieser Arbeitssitzung wurden in eine erste schriftliche Beschreibung des Protokolls eingearbeitet und im Workshop an der ECDL (s.o.) in Lissabon vorgestellt. Hier ging es um eine Präzisierung des Metadatenatzes und die Aufnahme einiger noch nicht eingebauter Eigenschaften ins Protokoll. Vom September bis Dezember verfeinerte das Technical Committee das Protokoll und testete es in ihren eigenen Einrichtungen aus.

### Offizielle Vorstellung der OAI-Spezifizierungen und des OAI-Metadaten Harvesting-Protokolls

Nach einigen Präsentationen der Alpha-Version des OAI-Metadaten-Harvesting-Protokolls (Mai, Juni 2000) wurde im Juli 2000 die Beta-Version für Erstanwender zur Verfügung gestellt. Das Ziel der Architektur des Protokolls ist es, einen leichten Zugang zu den vereinheitlichten Metadaten zu ermöglichen, mit denen die Datenanbieter (Eprint-Serverbetreiber) ihre Volltexte versehen haben, den Server-Providern den Zugang zu diesen Metadaten zu ermöglichen, damit sie diese um ihre Mehrwertdienste erweitern können.

Die Beta-Version wurde von ca. 8 Eprint-Serverbetreibern (darunter auch LANL) getestet und zwischen September und Dezember 2000 hatte das Technische Komitee der OAI weitere Erfahrungswerte eingearbeitet, um die Version 1.0 im Januar mit Erfahrungsberichten der Erstanwender in der Öffentlichkeit vorzustellen. Dafür waren zwei offene „Open Meetings“ geplant: eines am 23. Januar 2001 in Washington, D.C. (USA) und ein europäisches Treffen am 26. Februar 2001 in Berlin (Deutschland). Die eintägige Vorstellung enthielt einen

7 Open Archives Initiative: Frequently Asked Questions <<http://www.openarchives.org/faq.htm>>.

Überblick über die Entstehung und die Organisation der Open Archives Initiative, eine detaillierte Vorstellung der Architektur des OAI Metadata Harvesting Protokolls, einige Darstellungen der Erstanwender und Diskussionen sowohl über die dargestellten Implementierungen als auch über die weiteren Kooperationsmöglichkeiten unter den Interessenten und Teilnehmern mit Projekten im Aufbau. Da die Spezifizierungen für das Protokoll (<http://www.openarchives.org/OAI/openarchivesprotocol.htm>) bereits am 15. Januar 2001 auf der OAI Webseite öffentlich zur Verfügung gestellt wurden, konnten sich die Teilnehmer auf die Details vor der Tagung vorbereiten, was eine gezieltere Diskussion ermöglichte. Diese Open Meetings wurden von OAI in Zusammenarbeit mit anderen Sponsoren kostenlos für die Teilnehmer angeboten. Bei dem Treffen in Washington, D.C. nahmen 140 Personen – fast ausschließlich Amerikaner – teil, in Berlin nahmen 127 Personen aus 18 Ländern teil. Das europäische Treffen wird hier etwas mehr in Detail dargestellt. Beide Tagungen werden auf der Webseite von OAI entsprechend mit Bildern, Texten und Folien der Referenten dokumentiert (siehe <http://www.openarchives.org/DC2001/OpenMeeting.html> und <http://www.openarchives.org/Berlin2001/OpenMeeting.html>).

### Der Open Archives Initiative European Day

Das europäische Pendant zur offiziellen Vorstellung des OAI Metadata Harvesting Protokolls in den USA mit Demonstrationen durch die Erstanwender des Protokolls fand am 26. Februar 2001 in Berlin in der Staatsbibliothek Preussischer Kulturbesitz statt. Der European Open Archives Day wurde in englischer Sprache abgehalten, um eine möglichst große Beteiligung aus vielen europäischen Ländern anzuregen. Unter den TeilnehmerInnen am europäischen OAI-Tag waren Wissenschaftler fast aller Fachrichtungen (Physik, Mathematik, Chemie, Sprachwissenschaft, Informatik, Germanistik, Geschichte, Sozialwissenschaften, Erziehungswissenschaft und Biologie), sowie Bibliothekare, Informatiker, Rechenzentrenleiter und -mitarbeiter, Software-Produzenten, Zeitschriftenhändler und Subskriptionsagenturen, Wissenschaftsjournalisten sowie Buchhändler und Verleger. Auch mehrere Vertreter der Förderpolitik für Bibliotheks- und Informationsprojekte in Deutschland sowie Großbritannien nahmen daran teil. Viele der Teilnehmer waren ebenfalls an einem „digitalen Bibliotheksprojekt“ oder einem „Projekt zur Erforschung des elektronischen Publikationswesens“ beteiligt, das von der EU, der DFG, dem Global-Info oder einem ähnlichen Förderprogramm in ihrem eigenen Land unterstützt wird.

Der Ablauf des europäischen OAI-Tages war weitestgehend an den Ablauf des US-OAI-Tages in Washington angeglichen. Nach einer kurzen Einführung und Begrüßung wurde über die Entstehungsgründe und -geschichte sowie die Ziele der OAI von Carl Lagoze (Cornell Universität) informiert. Lagoze betonte den „experimentellen“ Aspekt der Initiative, die primär organisatorische Zusammenarbeit umfasst, um die Metadaten der abgelegten elektronischen wissenschaftlichen Texte und anderen Forschungsergebnissen in Eprint-Servern um die Welt durch das OAI-Protokoll zur Verfügung zu stellen und dadurch den Zugang zu wissenschaftlichen

Ressourcen zu verbessern, wo Standardisierungsprozesse noch keine Formatvereinheitlichung erreicht haben. Lagoze fasste die Kernkonzepte<sup>8</sup> der OAI (Version 1.0) zusammen:

- „Low-barrier Interoperability“ – Interoperabilität durch geringst möglichen Aufwand,
- „Data Provider and Service Provider Model“,
- „Metadata as Consumption Basis“ – ein Metadatenharvester-Modell, das http-basiert ist, eine Abwandlung vom Programm „Dienst“ mit Antworten in XML-Schema,
- gemeinsames Metadatenformat und ggf. parallel dazu Domäne-spezifische Metadatenätze,
- akzeptierte Nutzung von Metadaten, mit dem Fair-Use-Prinzip, dass die Nutzung der Metadaten durch andere nicht zum Nachteil des Metadatenersellers ist,
- „Flexibility“.

Diese Kernkonzepte sind als Ergebnis des öffentlichen Inputs der verschiedenen Workshopteilnehmer, des Feedbacks der Erstanwender und der Arbeitssitzungen sowohl des OAI-Technical Committee und des OAI-Steering Committees anzusehen.

Nach dieser perspektivischen Einleitung in die OAI sprach Paul Ginsparg, Gründer des ersten Preprint-Servers „LANL“ (nun unter [www.arXiv.org](http://www.arXiv.org) im Wesentlichen erweitert), über den Wandel im Bereich des wissenschaftlichen Kommunikations- und Publikationswesens. Mit dem Hintergrund der hochbezugssussten Forschungsszene der Hochenergiephysik während der Entwicklung des Internets konnte Ginsparg die Wandlungen in der Haltung der Autoren gegenüber den Verlagen und gegenüber den Karriere-Benchmarking des Peer-Review und Zitierungsfrequenzen zeigen. Ginsparg betonte, dass eine neue Basis für das Messen der Qualität eines Artikels die Zitierfrequenz ersetzen muss, weil diese nicht der eigentlichen Nutzung und dem Lesen der Artikel entspricht. Er zeigte das exponentielle Wachstum seines Preprint-Servers und demonstrierte die Bedeutung dieses Dienstes, der den eigentlichen Vertrieb von wissenschaftlichen Publikationen ergänzt. Am Ende zeichnete er spontan die intrinsischen Verbindungen zwischen Datenanbietern, Informationsdiensten und „Wissens“-anbietern („Data, Information and Knowledge“), die in einer komplexen Wechselwirkung zu einander stehen. (vgl. hierzu seine handschriftliche Folie auf der Tagungswebseite, s.o.)<sup>9</sup>.

Carl Lagoze referierte danach detaillierter über die technischen Voraussetzungen zur Implementierung des OAI-Metadata Harvesting Protokolls. Nach der Mittagspause berichteten Erstanwender, Projektleiterinnen und OAI-Dienstleistungsentwickler über ihre Erfahrungen mit dem OAI-Protokoll und über ihre Projekte. Darunter berichtete Kurt Maly (Old Dominion University, USA)

8 Wo das englische Konzept verständlich ist, hat die Verfasserin den englischen Begriff gelassen, wo evtl. mehrere Interpretationen möglich sind, wurden die englischen Konzepte ins Deutsche übersetzt.

9 Siehe auch Paul Ginsparg: *Winners and Losers in the Global Research Village*. Invited Contribution for UNESCO Conference, 19.-23. Februar 1996 (<http://xxx.lanl.gov/blurg/pg96unesco.htm>). Zum Schluß gibt Ginsparg eine beispielhafte Erklärung zu Autorenrechten und „Fair Use“.

über den von seinem Projekt entwickelten Dienst „Arc“ (siehe Anm. 14), eine Dienstleistung basierend auf den Metadaten in OAI-kompatiblen Servern, die eine gleichzeitige Recherche in allen Servern durchführt. Heinrich Stamerjohanns (Universität Oldenburg) und Susanne Dobratz (Humboldt Universität zu Berlin) berichteten über die Implementierung von Version 1.0 mit deutschsprachigen Dokumenten und anderen Hürden zur Implementierung des Protokolls in den beiden OAI-kompatiblen Server in Deutschland, den PhysNet-Server in Oldenburg und den Dokumentenserver der Humboldt Universität zu Berlin.

Jean-Yves Le Meur zeigte Beispiele aus dem großen CERN Document Server (CDS) in Genf, der vollständig auf XML umgewandelt ist. Die CERN Politik sieht die elektronische Archivierung aller Forschungsberichte und Abhandlungen im Document Server vor. Yohanan Spruch (Israel) überließ den Vortrag Frau Eva Krall (Ex Libris), die von der OAI-Kompatibilität des Ex Libris-Systems sprach und zeigte, wie man auch das OAI-Protokoll zum Aufbau eines Verbundsystems nutzen kann. Andy Powell (UKOLN) beschrieb die Erfahrungen mit der Implementierung des Protokolls beim Arts and Humanities Database Service (AHDS) in Großbritannien. Les Carr (Abteilung Elektronische Dienstleistungen und Informatik der University of South Hampton, Großbritannien) stellte in seinem Vortrag die Frage nach den weiteren Dienstleistungen und Erkenntnissen im wissenschaftlichen Publikationsbereich, während sich diese Umstellung auf elektronische Fachinformation vollzieht. Er beschrieb erste Auswertungen von Untersuchungen über die Akzeptanz und Nutzung von Eprint-Servern und sprach einige der wesentlichen Fragen an, die Bibliothekare und Wissenschaftler zugleich beschäftigen. Donatella Castella (Universität Pisa, Italien) schilderte die Vorgehensweisen im Cyclades-Projekt, einem EU-Projekt, das das OAI-Protokoll als einer der Erstanwender implementiert hat (<<http://www.ics.forth.gr/proj/iss/Projects/cyclades.html>>). Sie sprach auch unter anderem die Schwierigkeiten der Inhaltserschließung innerhalb des Projekts an. Der Cyclades-Dienst zielt auf eine Untermenge der OAI-Server, dessen Metadaten inhaltlich vom Cyclades abgesehen und verarbeitet werden können. Jeff Young demonstrierte die Anwendung vom OAI-Protokoll beim OCLC vor allem in Kooperation mit dem NDLTD-Dienst für elektronischen Dissertationen und Thesen. Auszüge aus dem OCLC WorldCat konnten in Dublin Core Metadaten umgewandelt werden, um eine Basis für die Erprobung vom OAI-Metadata Harvester Protokoll in bibliothekarischen Daten zu stellen. Die Bedeutung eines solchen Konvertierungstools von herkömmlichen Bibliotheksdaten in OAI-kompatible Metadaten liegt auf der Hand.

Abschließend leitete Carl Lagoze die Diskussion. Es war offensichtlich, dass die meisten der Teilnehmer zwar noch keine Implementierung des OAI-Protokolls bisher in ihren Einrichtungen hatten, aber großes Interesse daran hatten. Gerade die Fragestellungen aus Ungarn und aus Skandinavien zeigten reges Interesse an der Weiterentwicklung des OAI-Einsatzes in ihren Projekten und Ländern.

Das Treffen wurde von Diann Rusch-Feja unter Mitwirkung von Friederike Schimmelpfennig (DINI Geschäftsstelle), Sarah Aerni (MPI für Bildungsforschung) und Uta Siebekz (Fritz-Haber-Institut) organisiert; Sponso-

ren des europäischen OAI Tages waren die Staatsbibliothek Preußischer Kulturbesitz zu Berlin sowie die Stiftung Preußischer Kulturbesitz, das Joint Information Systems Council (JISC) von Großbritannien und das Distributed National Electronic Resource (DNER) sowie die Deutsche Initiative für Netzwerkinformation (DINI). Weitere Berichte über den europäischen OAI Tag sind in *c't*<sup>10</sup> und in *Bibliotheksdienst*<sup>11</sup> nachzulesen.

## Die Organisation der Open Archives Initiative

Aktivitäten der Open Archives Initiative werden zum Teil von der Digital Library Federation der USA (DLF) unterstützt, die beim Council for Library and Information Resources (CLIR) angesiedelt ist, und von der Coalition for Networked Information (CNI, in Reston, Virginia). Fördergelder für die Entwicklung des OAI-Metadata Harvesting Protokolls können von der US National Science Foundation (NSF) sowie von der US Defense Advanced Research Projects Agency (DARPA) erhalten werden. Politikentscheidungen und richtungsweisende Diskussion leistet das OAI „Steering Committee“, das aus 12 Vertretern wissenschaftlicher Einrichtungen, wissenschaftlicher Bibliotheken, Fachgesellschaften, Forschungseinrichtungen und wissenschaftlicher Eprint-Servern besteht<sup>12</sup>. Die Aktivitäten des Steering Committee werden von Daniel Greenstein koordiniert. Die Kernarbeit am Protokoll und an den Spezifizierungen des OAI-Protokolls und des Metadatenatzes wird vom „OAI Technical Committee“ durchgeführt, das auch die Implementierungen des Protokolls im Hinblick auf die Einarbeitung von weiteren Ergänzungen begleitet. Eine Art von Geschäftsstelle und Koordinierungsstelle für OAI-Aktivitäten leiten Carl Lagoze und Herbert Van de Sompel in der Informatik Abteilung der Cornell University. Die Mailinglisten *OAI-general* und *OAI-Implementers* sind beide über die Open Archives Initiative Homepage zu erreichen. Die beiden Diskussionslisten werden auch archiviert.

## Teilnehmende Institutionen und Sammlungen

Ab Anfang 2001 konnten Institutionen, die OAI-kompatible Metadaten-Repositories haben, sich bei OAI anmelden und ihre Sammlungen auch für weitere Anwendungsexperimente zur Verfügung stellen. Bis Anfang Mai 2001 meldeten sich 26 Institutionen<sup>13</sup> mit Eprint-Servern, darunter auch die Humboldt-Universität zu Berlin, die europäische Gesellschaft für Sprachenres-

10 Richard Sietmann: Napster für die Wissenschaft: Ein neues Protokoll für die virtuelle Fachbibliothek. In: *c't* 6/01 (2001) S. 78-79; auch <<http://www.heise.de/ct/01/06/078/default.shtml>>.

11 Inka Trappenbeck: Die Open Archives Initiative. Ein Bericht vom „Open Meeting“ in Berlin. In: *Bibliotheksdienst* 35 (2001) S. 443-446.

12 Das Steering Committee besteht aus Caroline Arms (LoC), Lorcan Dempsey (DNER, UK), Dale Fletcher (Harvard), Ed Fox (Virginia Tech), Paul Ginsparg (LANL), Daniel Greenstein (CLIR), Carl Lagoze (Cornell University), Cliff Lynch (CNI), John Ober (California Digital Library), Diann Rusch-Feja (MPI für Bildungsforschung), Herbert Van de Sompel (Cornell University), Don Waters (Carnegie Mellon Foundation).

13 Die vollständige und aktuelle Liste ist über die URL <<http://oaisrv.nsd.cornell.edu/Register/BrowseSites.pl>> erhältlich.

sourcen und CogPrints an der South Hampton University (England). Diese Liste der „Frühanwender“ zeigt, daß die OAI längst keine ausschließlich amerikanische Initiative mehr ist. Darunter sind Repositories, die mehrere fachbezogene Sammlungen nachweisen, und einige, die eine sehr fachspezifische Sammlung aufweisen.

Bis Mitte Mai hatten lediglich zwei Institutionen Services registriert, die auf den Metadaten aufbauen, die durch das OAI Metadata Harvesting Protokoll gesammelt werden. Diese sind „Arc“<sup>14</sup>, ein Recherche-Service über mehrere OAI-kompatible Repositories, der von Informatikern an der Old Dominion University entwickelt worden ist, und „Repository Explorer“, der an der Virginia Tech zur Validitätsprüfung der Repositories über das OAI-Protokoll entwickelt wurde.

Webserver, deren Daten OAI-kompatibel sind und die Voraussetzungen für die Implementierung des OAI-Protokolls erfüllen, können sich als OAI-Repositories anmelden<sup>15</sup>. So werden OAI-kompatible Datensammlungen den Implementierern bekannt, die OAI-kompatible Dienstleistungen anbieten, bzw. die Metadaten von OAI-kompatiblen Webservern bei Abfragen verarbeiten können.

### Das OAI Metadata Harvesting Protokoll

Zielsetzung des OAI Metadata Harvesting Protokolls ist es, einen applikationsunabhängigen Rahmen für Interoperabilität zu bieten, der in verschiedenen Fachcommunities für das Auffinden und die weitere Bearbeitung webbasierter wissenschaftlicher Veröffentlichungen anwendbar ist. Dieser Interoperabilitätsrahmen umfasst zwei Arten von Teilnehmern: Datenanbieter, die die Inhalte so aufbereiten, dass ihre Metadaten vom OAI Protokoll erfasst werden können, und Serviceprovider, die verschiedene Dienstleistungen und Mehrwert-Services auf der Basis des OAI-Netzes und enthaltener OAI-Metadaten aufbauen und anbieten. Letztere können kostenpflichtig, lizenziert oder frei und kostenfrei verfügbar sein. Dies sind Dienstleistungen, die über die Such- und Harvestingprotokollergebnisse des OAI-Protokolls hinausgehen und einen besonderen Service anbieten.

Entwickelt wurde das OAI Protokoll vom Technical Committee der OAI, das von Herbert Van de Sompel und Carl Lagoze geleitet wurden. Dieses Technical Committee hat im Zeitraum vom September 2000 bis zur Fertigstellung der Version 1.0 des OAI-Metadata Harvesting Protokolls im Januar 2001 mit dem Zweck existiert, diese Version 1.0 zu erstellen und testen. Seit Januar 2001 wird im Rahmen der jetzt anstehenden „Experimentierphase“ mit Version 1.0 ein neues Technical Committee zusammengestellt, das eine größere internationale Beteiligung hat, um die Entwicklungen in den Projekten widerzuspiegeln. Dieses Committee hat dann den Zweck, „to evaluate our current technical approach and understand how it might be modified or enhanced. At this point, however, major changes are not expected.“<sup>16</sup>

So wurde das Open Archives Initiative Protocol for Metadata Harvesting (<<http://www.openarchives.org/OAI/openarchivesprotocol.htm>>), Version 1.0, am 18. Januar 2001 auf der Open Archives Website veröffentlicht. Vorher konnte die Beta-Version des Protokolls ab Anfang Juli 2000 heruntergeladen und ausgetestet wer-

den, wenn ausreichende Dokumente mit OAI-verträglichen Metadaten im XML-Format vorhanden waren. Diese „Early Adapters“ (Frühanwender) des Protokolls waren meist diejenigen, die bereits Preprint-Server oder ähnlichen Server mit elektronisch verfügbaren wissenschaftlichen Dokumenten betrieben hatten oder mit an der Entwicklung des Protokolls beteiligt waren. Große Sammlungen von elektronischen Dokumenten, die mit OAI-kompatiblen Metadaten versehen sind, werden als „repositories“ behandelt. Ein „Repository“ („Lager“, „Behälter“, „Sammlung“, hier hat es den Sinn „elektronisches Textarchiv“ kann aber auch andere elektronische Datenformate enthalten) ist im OAI-Sinn ein über Internet zugänglicher Server, der OAI-Protokoll-Anfragen, die in HTTP eingebettet<sup>17</sup> sind, aufgrund der vorhandenen OAI-kompatiblen Metadaten bearbeiten kann und Recherche-Ergebnisse zur Anfrage zurückschicken kann. Das Ergebnis besteht aus Metadatenätzen, die auf die verzeichneten Objekte im Repository zurückgreifen. Wo aber das Objekt wirklich ist, ist irrelevant. Auch wie (z.B. automatisch „on-the-fly“) oder von wem die Metadaten erstellt werden, ist für die Funktion des Protokolls nicht von Belang.

Das Ziel der Arbeit in der Open Archives Initiative beruht auf der grundlegenden Funktion der „digitalen Bibliotheken“, wie sie im Digital Libraries Förderprogramm der US-National Science Foundation definiert worden ist: Such- und Navigationsmöglichkeiten sollen über heterogene Datenbestände mittels interoperabilitätsbefähigender Datenverarbeitungstools verfügen<sup>18</sup>. Um in mehreren Sammlungen von elektronischen Werken mit einem Protokoll zu Anfragenergebnissen zu kommen,

14 arc – Cross Archive Searching Service <<http://arc.cs.odu.edu/help/archives.htm>>.

15 Die OAI-Anmeldungsseite ist: <<http://oaisrv.nsd.l.cornell.edu/Register/RegisterSite.pl>>.

16 Aussage von Carl Lagoze im Gespräch am 7.5.2001 [zitiert mit Erlaubnis].

17 OAI Protokoll-Anfragen werden als HTTP Anfragen ausgedruckt (siehe die Beschreibung des Protokolls). Der BASE-URL bezeichnet den Host und Port eines Internetservers, der die Repository enthält. In den meisten Fällen wird der BASE-URL auch den Pfad enthalten, der von diesem HTTP-Server als Händler für OAI-Protokollanfragen dienen sollte. Die Zusammensetzung des Pfads, sowie dessen Vorhandensein oder Abwesenheit, wird in der Konfigurierung des HTTP-Servers bestimmt, der die Repository enthält. Neben dem BASE-URL enthalten alle OAI-Protokollanfragen eine Liste der „keyword arguments“, die als „key=value“ Paar ausgedruckt werden. Diese Anfragen („arguments“) können in beliebiger Reihenfolge angegeben sein und mehrfache Anfragen („arguments“) müssen mit dem Zeichen „&“ getrennt werden. Jede OAI-Protokollanfrage enthält mindestens ein key=value Paar, das die Anfrage des Klienten definiert. Der „key“-Teil enthält den String für „verb“. Der „value“-Teil enthält einen der OAI-Protokollbefehle/Anfragen. Die Zahl und Art von zusätzlichen „key=value“ Paaren hängt von den Bestandteilen („arguments“) für die einzelnen Protokollanfragen ab. Hier sind die Befehle „GET“ und „POST“ sehr wichtig.

18 Vgl. Andreas Paepke, Chen Chuan Chang, Hector Garcia-Molina, and Terry Winograd: Interoperability for Digital Libraries Worldwide. In: Communications of the ACM, 41, 4 (1998); Carl Lagoze: Defining Collections in Distributed Digital Libraries. In: D-Lib Magazine 5, no. 11 (1999) <<http://www.dlib.org/dlib/november98/lagoze/11lagoze.html>>; William Y. Arms: Digital Libraries. Boston: MIT Press 2000.

müssen Strukturen erreicht werden, die mit komplexen Umsetzungsformeln und Konkordanzen etc. umgehen können. So ist es am leichtesten, mittels eines gemeinsamen Metadatenformats und Metadatenharvester die Interoperabilität zu schaffen, die Suchmöglichkeiten und Retrievalschlüssel schnell, gezielt und ohne umständliche Umwandlung von inhaltlichen Informationsquellen oder Konvertierung gesamter Datenbanken und Datensammlungen zu erreichen. So spricht man von der Suche quer durch multiple repositories, cross-repository Suche etc. Es wird anerkannt, dass bestimmte Fachcommunities ggf. bestimmte Bedürfnisse haben, deren Erfassung außerhalb der Möglichkeiten des OAI liegen und daher sind mehrere Metadatenätze nicht nur möglich, sondern auch hilfreich, insbesondere, wenn solche Fachcommunities ihre eigenen Dienste zusätzlich zum OAI Protokoll erarbeiten.

Das OAI-Protokoll zu implementieren, setzt zunächst das Vorhandensein von OAI-kompatiblen Metadaten voraus. Die eigentliche Implementierung des Protokolls bedarf jedoch Kenntnisse von Perl und CGI-Skripten, um den Web-Server so zu konfigurieren, dass OAI-Protokollanfragen verarbeitet werden können. Angesichts des Interesses, das dem OAI in der wissenschaftlichen Community begegnet ist, ist zu erwarten, dass generische „Front-Ends“ zur Verarbeitung der Protokollanfragen über die OAI-HomePage oder über die Diskussionsliste „OAI-Implementers“ angeboten werden<sup>19</sup>.

### OAI-kompatible Metadaten

Das Protokoll erfordert OAI-kompatible Metadatenätze im XML-Format. Diese Sätze befinden sich in Dateien („records“), die aus folgenden Teilen bestehen:

Der <Header> enthält eine eindeutigen Identifikation (<Identifier>), der als Schlüssel für eine unverfälschte Identifikation eines Metadatenatzes zu einem bestimmten Objekt in der Sammlung notwendig ist. Diese Identifikation enthält den Namen der Sammlung („repository“) und eine eindeutige Bezeichnung des Metadatenatzes. Ein Präfix vor der Identifikation bezeichnet das Metadatenformat, falls die Metadaten in mehreren Metadatenformaten vorliegen (oai\_dc; oai\_marc; oai\_rfc1807 etc.). Der <Datestamp> im Header gibt das Erstellungsdatum, Lösungsdatum und Datum der neuesten Änderungen der Metadaten an. Der <Datestamp> liefert wichtige Informationen, die bei Anfragen z.B. nach allen neuen Metadatenätzen zu einem Thema oder zu einem Autor nach einem bestimmten Datum (z.B. die letzte Suchanfrage nach diesem Thema, Namen), und zur weiteren Verarbeitung der Metadaten erforderlich sind. Das Datum im <Datestamp> unterliegt der Regel für die Darstellung von Kalenderdaten in maschinenverarbeitbarer Form (siehe Punkt 2.4 im Protokoll).

Die Kategorie Metadaten <Metadata> umfasst eine Reihe von Metadatenelementen nach dem Schema Dublin Core und die dazugehörigen inhaltliche Werte („values“), die zusammen *ein* Objekt beschreiben. Das Protokoll unterstützt ebenfalls mehrfache Metadatenätze (z.B. von verschiedenen Personen, Institutionen etc.) auch in verschiedenen Formaten für ein Objekt. Die Mindestanforderung an die OAI-kompatibeln Metadaten ist, dass sie im Dublin Core Metadatenformat ohne Qualifier sind<sup>20</sup>. So ist das Präfix „oai\_dc“ erforderlich für alle

Metadatenätze. Ausserdem können im Repository weitere Metadatenätze zu einem einzelnen Objekt stehen, die auch in anderen Metadatenformaten ausgedruckt sind als in Dublin Core. Diese haben denselben <identifier> für den Metadatenatz jedoch mit einem Präfix, das den Metadatenformat bezeichnet (oai\_marc; oai\_rfc1807 etc.). Dieser Identifier ist jedoch nicht gleichzusetzen mit dem Dublin Core (dc) Element <identifier>, der dann den URL (JRN, URI) enthält, die eindeutige Adresse für das beschriebene Objekt. Da Nutzer elektronischer Archivdienste wie OAI meistens auch über die angezeigten Metadaten zum Objekt gelangen möchten, kann die Verlinkung zum Objekt über dieses Element erfolgen. Die Semantik der Werte der Metadaten folgt den Dublin Core Best Practice Guide (<<http://www.dublincore.org/documents/2001/04/12/userguide/>>). Syntaktische Regeln ergeben sich nur als Best Practice Empfehlungen, oder folgen üblichen bibliographischen und bibliothekarischen Formeln (z.B. Nachname, Vorname oder Nachname, Initialen etc.). Regeln zur Syntax und zur Semantik der Metadatenwerte liegen zwar außerhalb der Regelungsmöglichkeiten von OAI, jedoch empfiehlt es sich, dass die gängigen Regeln für Informationsretrieval hier benutzt werden, um semantische Interoperabilität auch zu erreichen und die informationstechnischen Möglichkeiten des OAI Protokolls am sinnvollsten auszuschöpfen.

Die ersten Angaben in <Metadata> im Metadatenatz („record“) beschreiben die Attribute, die den Namespace, Schema und Validierungsangabe für das benutzte Metadatenformat identifizieren: Diese sind im Idealfall

xmlns – die XML Namespace Spezifizierung für die Metadaten, die folgen,

xmlns:xsi – enthält die URI für das XML-Schema. Dies folgt dem W3 Consortium XMLSchema: <<http://www.w3.org/2000/10/XMLSchema-instance>>,

xsi:schemaLocation – enthält den Namespace URI (definiert durch den XML Namespace Spezifizierung) für die Metadaten, die folgen, mit dem URL der XML Schema zur Validierung der Metadaten: z.B. für Dublin Core im OAI-Format: <<http://purl.org/dc/elements/1.1/>> <<http://www.openarchives.org/OAI/dc.xsd>>.

Ausserdem gibt es im Metadatenatz („record“) die Kategorie <about>, die die Erfassung zusätzlicher Informationen über die Metadatenwerte ermöglicht. In <about> können z.B. Angaben zum Urheber der Metadaten bzw. Rechtebestimmungen für die weitere Nutzung der Metadaten etc. stehen. Die Struktur dieser Kategorie ist vom OAI-Protokoll nicht vordefiniert. Gegebenenfalls können einzelne Fachcommunities über die Syntax und Semantik dieser Kategorie einzelne Regelungen bestimmen. Um diesen Teil jedoch in einheitlicher Form abzurufen und durchzusuchen, empfiehlt es sich auch hier, wie oben im „Metadata“ ähnliche Regelungen für die Identifi-

19 Vgl. „How long will it take me to implement the protocol?“ In: Open Archives Initiative. Frequently Asked Questions <<http://www.openarchives.org/faq.htm>>.

20 Zur Erläuterung der zugrunde gelegten Dublin Core Elemente ohne Qualifikation, siehe <<http://www.openarchives.org/OAI/dc.xsd>>, die auch als DC Schema Qualifikation angewendet wird.

kation von validierbaren Namespace und Spezifizierungen für diese Kategorie (siehe Beispiel Punkt 4 des Protokolls) zu verschaffen. Hier können z.B. Links zu Rechtsaussagen gelegt werden, die für eine gesamte Repository gelten und daher nicht immer in jedem Metadatenatz wiederholt werden müssen.

Zur Zeit basiert die Interoperabilität des OAI-Protokolls auf den dafür vorgesehenen Metadaten. Dies entspricht dem Ziel der OAI, einen Rahmen mit sehr geringen Hindernissen und möglichst weitwendbar zu schaffen, um zu ermöglichen, dass mit diesem Tool verschiedenartige Eprint-Server mit einer Abfrage abgesehen werden könnten. Zur Erklärung der OAI-Abfragen und zu erwarteten Antwortformen wird auf das OAI-Protokoll verwiesen.

### Implementierung des OAI-Protokolls

Datenanbieter können ihre Daten anbieten, wenn bestimmte Konventionen eingehalten werden, die zur Interoperabilität beitragen. Nach dem OAI Protokoll und der Konformität mit den OAI-Metadaten müssen die Identifier der Eprints den Bestimmungen des URI-Syntax entsprechen. Durch den Präfix „oai“ wird erklärt, daß die Eprints in diesem Server die OAI-Standards einhalten. Interne Namensgebungen können auch Versionsbezeichnungen eines Eprints identifizieren und haben oft eine Datumsangabe im Namen der Eprints eingebettet. Die Regeln des „oai-identifier“ wurden oben erwähnt. Das Format für die Art des Inhalts basiert hauptsächlich auf text/xml. Der OAI-Namespace für das XML Schema wird in der URI-Namespace der Protokollanfrage als erstes angegeben und als zweite die URL des XML Schemas zur Validierung der Anfrage.

Der Zusatz „oai-identifier.xsd“ ermöglicht die Erfassung mehrerer Kommentare, die Formate erklären, wie z.B. qualifizierte Dublin Core Metadatenelemente oder weitere Metadaten, die im Archiv benutzt werden bzw. Kommentare, die zusätzliche Informationen zu bestimmten Metadatenätzen enthalten (z.B. oai\_marc.xsd).

In der Protokollanfrage befindet sich ein Zeit/Datum Anhang, der die Uhrzeit und Datum der Anfrage mitnotiert. Dies dient der Versionskontrolle bei der automatischen Auswahl der neuesten Version. Zusätzlich wird in der Protokollanfrage ein „http-status-code“ angehängt, um den Erfolg oder Fehlergrund zu kennzeichnen, wenn die Anfrage erfolglos ist. Status-Codes können auch dazu dienen, im Falle einer Fehlantwort die Anfrage an einen anderen Server weiterzuleiten bzw. eine Wiederholung der Anfrage nach einem bestimmten Zeitraum zu fordern. Der „http-reason-phrase“ gibt dem menschlichen Nutzer mehr Informationen über eine Fehlermeldung, die für den suchenden Nutzer gegebenenfalls nützlich ist.

OAI-Protokollbefehle haben die Form von Verben. Die wichtigsten vorgesehenen Befehle sind:

- Identify
- ListIdentifiers
- ListRecords
- ListMetadataFormats
- ListSets
- GetRecord

Die Antwort eines OAI-befähigten Servers auf den Verbbefehl schließt den „repositoryName“ des Archivs („repository“), den „baseURL“ der Repository, die „protocolVersion“ des OAI-Protokolls, der das Archiv unter-

stützt, und die „adminEmail“, die Emailadresse des Serveradministrators mit ein. Weiterhin kann eine Liste der sogenannten „description“-Behälter („description containers“) zur Beschreibung von domäne- bzw. fachspezifischen Sammlungen auf der Sammlungsebene benutzt werden.

Bei nicht vollständigen Antworten fungiert ein „resumptionToken“ als Antwortkontrolle und enthält einen Wert, der den „flow control token“ des letzten „ListIdentifiers“- , „ListRecords“- oder „ListSets“-Befehls, der nur partiell beantwortet wurde, wieder einsetzt. Somit kann man mit dem „resumptionToken“ Antworten, die vorher aus derselben Sammlung zu einer bestimmten Anfrage geliefert wurden, mit Antworten aus der abgebrochenen Liste von Antworten aus dieser Sammlung zusammenführen, damit die Antworten aus der ursprünglichen Abfrage komplett sind.

### Nutzungshinweise zum OAI-Protokoll

Einige Nutzungshinweise zeigen die flexiblen Einsatzmöglichkeiten des OAI-Protokolls:

#### 1 Deleted records

Das OAI-Protokoll unterstützt eine Erkennung von gelöschten Metadatenätzen („deleted record“), wenn zum Beispiel das Objekt nicht länger in dieser Form erhältlich ist (z.B. Entfernung eines Preprints, nachdem der Text veröffentlicht ist, was in manchen Fachgebieten oder bei Veröffentlichung in manchen Zeitschriften vom Verleger erforderlich ist). In diesem Fall wird eine Statusbezeichnung „deleted“ bei der Antwort auf ListRecords und ListIdentifiers darauf hinweisen, dass keine Metadaten für diesen Metadatenatz geliefert werden können. Politische Entscheidungen, z.B. wie lange noch der „deleted“ Status im Repository bleibt, oder ob überhaupt dieser Status im Metadatenatz festgehalten wird, bleiben im Ermessen der einzelnen Sammlungsbetreiber („repository administrator“). Ebenso liegt es bei den Service-Providern, ob eine solche Statusbezeichnung mit erfasst wird, ob sie erforderlich ist, oder ob auf die entsprechenden Anfragen und Befehle (wie auch GetRecord) geeignete Antworten mit diesen Informationen geliefert werden.

#### 2 Gruppierungen innerhalb einer Repository

Mit der Bezeichnung „set“ (Sammlung) können beteiligte Repositories hierarchische oder themenorientierte Sammlungen kennzeichnen, die zur Vorselektion von Metadatenätzen dienen. Mögliche Sammlungsbezeichnungen für eine Repository sind „institution“ oder „subject“, die dann beide unterschiedlich mit hierarchischen Unterteilungen („nodes“) weiter differenziert werden können:

Diese Sammlungsbezeichnungen werden als „Tag“ gekennzeichnet, weitere Spezifizierungen mit „Spec“:

```
setTag = institution
setSpec = institution:humboldt (für Humboldt Universität zu Berlin)
```

oder

```
setTag = subject
setSpec = subject:computersci (für Informatik = computer science)
```

Je nach inhaltlicher und institutioneller Zuordnung könnte ein Metadatenatz mehreren Sets angehören.

„setName“ erlaubt die genaue Buchstabenfolge eines Namens, der für die Anzeige gegebenenfalls wichtig ist, insbesondere wenn sonst mehrere institutionelle Sammlungen innerhalb einer Repository vorhanden sind und der aussenstehende Nutzer des OAI-Protokolls die internen Abkürzungen nicht erraten kann. Zum Beispiel wären folgende Möglichkeiten für eine fiktive Repository der Max-Planck-Gesellschaft unter anderem mit folgenden Sets denkbar:

```
setTag = institution
setSpec = mpib-berlin
setName = Max Planck Institute for Human
         Development, Berlin
```

```
setTag = institution
setSpec = mpib-munich
setName = Max Planck Institute for Biochemi-
         stry, München
```

Auf die Zuordnung oder hierarchische Gliederung von Institutionen oder Fachgebieten will OAI keinen Einfluss nehmen. Sie verweist auf die Fachcommunities, die möglicherweise bereits über einen vereinheitlichten Gebrauch eines kontrollierten Vokabulars zur Klassifizierung der Verzweigungen ihres Fachgebiets verfügen. In einigen Fällen sind geeignete Begriffe von Service-Providern wie Indexierungsdiensten (Abstrakt- und Indexierungsdatenbankbetreibern, bibliographischen Diensten etc.) entwickelt worden und in ständiger Verwendung. Dieser Dienst des OAI-Protokolls ersetzt keinen formalen Thesaurus, in dem z.B. die Hierarchie nach unten und nach oben gegebenenfalls auch in horizontalen Vergleichen verfolgt werden kann. Dieses Attribut des OAI-Protokolls dient lediglich dazu, logisch selektierte Sammlungen innerhalb einer Repository bei einer Anfrage anzusprechen, statt der gesamten Repository.

### Aussichten: Was man von OAI in der Zukunft erwarten kann

Die Entwicklungen von OAI und insbesondere dieses erste Jahr der Testphase für die Version 1.0 des OAI-Metadata Harvesting Protokolls dienen zunächst einer Experimentierphase. Obwohl das Protokoll in verschiedenen Einrichtungen getestet wurde, fehlen noch die Erfahrungen mit dem Protokoll und einem vereinheitlichten Metadatensatz, um festzustellen, welche Dienste bei einer „cross-disciplinary“ (quer-durch-verschiedene Fachrichtungen) Abfrage benötigt werden und ob die Ergebnisse davon wirklich nützlich sind. Welche Mehrwertdienste können auf dem Protokoll und der einfachen Abfrage sinnvoll aufgebaut werden? Weitere Fragen stellen sich ebenfalls, die durch Erfahrungswerte im Laufe des bestehenden Jahres gelöst oder beobachtet werden können, wie vertragen sich z.B. verschiedene Metadatensätze miteinander?

Im Vergleich mit dem bibliotheksorientierten Protokoll Z 39.50 ist das OAI-Protokoll viel weniger kompliziert und leichter implementierbar. Das Z 39.50 Protokoll entspricht fast dem vollen bibliothekarischen Austauschformat und hat dadurch verhältnismäßig viele Kategorien. Das Z 39.50 Protokoll arbeitet auf der Basis von Verarbeitungssessionmanagement mit Ergebnismengen und kann auch Filtermechanismen einbauen. Das OAI-Protokoll soll das Z 39.50 Protokoll nicht unbedingt ersetzen. Stattdessen wurde das OAI-Protokoll eher entwik-

kelt, um die Menge an Metadaten, die zur präzisen Metadatenrecherche als Suchstrategie dienen, klein zu halten, damit die Funktionalität und Interoperabilität am besten zu erreichen ist. Auch die Leistungsfähigkeit des OAI-Protokolls wird im Vergleich mit der Leistungsfähigkeit des Z 39.50 Protokolls während der sogenannten „Experimentierphase“ beobachtet und bewertet.

Im Rahmen der ersten Anwendungen des OAI-Metadata Harvesting Protokolls müssen auch Fragen nach dem geistigen Eigentum und nach den Autorenrechten untersucht werden. Wenn mehrere Metadatensätze von verschiedenen Stellen erhältlich sind, wo und inwieweit wird ein Metadatensatz bevorzugt, oder haben sie gegebenenfalls in verschiedenen wissenschaftlichen Communities unterschiedliche Stellenwerte? Im Gegensatz zu bibliothekarischen Katalogisaten nach bestimmten Regelwerken, können Metadaten als geistiges Eigentum erkannt werden und wie konkurrieren verschiedene Metadatensätze miteinander? Wie wird eine solche Frage gerade bei einem Rechercheprozess über das OAI-Protokoll abgewickelt? Könnten Kosten- und Urheberrechtsfragen zwischen Abfrage und Anzeige der ersten Ergebnisse ein großes Hindernis sein? Indem das OAI-Protokoll das „About“-Feld für Informationen über die Metadaten selbst definiert hat, besteht eine Möglichkeit, auch diese Rechte sowie die Zugangsrechte zu den Servern darzustellen.

Die sogenannte „Experimentierphase“ mit dem OAI-Protokoll Version 1.0 hat bereits zur Planung von mehreren Projekten zwischen institutionellen Anbietern von elektronischen Ressourcen geführt.

Des Weiteren sind einige Veranstaltungen geplant, bei denen das OAI-Protokoll und seine Anwendung im Mittelpunkt stehen. Darunter war z.B. die im März 2001 durchgeführte Konferenz über die Open Archives Initiative and Peer Review journals in Europa, die beim CERN in der Schweiz stattgefunden hat <<http://documents.cern.ch/AGE/current/fullAgenda.php?ida=a01193>>. Die Diskussionen zeigten, dass trotz ihrer Zielsetzung der Einfachheit und des effizientesten Weges zur Interoperabilität die OAI die Entwicklungen des RDF nicht ausser Acht lassen können. Für den 11. Juli 2001 ist ein OAI-Workshop in Großbritannien geplant.

Im September 2001 findet ein Open Archives Workshop im Rahmen der European Conference on Digital Libraries (ECDL) in Darmstadt statt, in dem einige Projekte dargestellt werden, aber auch geplant ist, über die bis dahin gemachten Erfahrungen mit dem OAI-Protokoll zu diskutieren und sich auszutauschen. Ebenfalls für September ist ein Treffen OAI-Interessierter und -Anwender in Deutschland koordiniert durch DINI geplant. Im November 2001 findet ein Workshop von der Arbeitsgemeinschaft der Spezialbibliotheken zum Thema Elektronisches Publizieren in Jülich statt, in dem einige Sitzungen über Preprint-Server, die OAI und die neuen Anforderungen und Formen elektronischen Publizierens angeboten werden. Sicherlich wird der Veranstaltungskalender bis zum Jahresende einige weitere Workshops und Konferenzen mit der Darstellung von OAI-Projekten und -Anwendern enthalten.

Zur Zeit werden mehrere Anträge bei der EU, NSF und anderen Fördereinrichtungen zur Erforschung der Grundlagen für die gemeinsame Recherche und Dienstleistungen von Eprint-Server gestellt. Sinnvoll ist es, dass die verschiedenen Experimente mit dem OAI-Pro-

tokoll und Erstellung von neuen Dienstleistungen koordiniert werden. Dieser Antrag will vor allem versuchen, die OAI-orientierten Projekte zusammenzuführen, um Doppelarbeit zu vermeiden und die Erfahrungswerte besser und schneller untereinander auszutauschen. Weitere international angelegte Projektanträge wollen das OAI-Protokoll und Metadatenvoraussetzungen in der mehrsprachigen Umgebung und mit kombinierten Metadaten über nicht-textorientierte Datenformate beobachten.

Die OAI selbst wird Ende 2001 und Anfang 2002 einige Workshops veranstalten, in denen mit Hilfe der Erfahrungsberichte überlegt wird, in welche Richtung weitere Bemühungen und Entwicklungen der OAI, ihre Tools und die Definition ihrer Rahmenarchitektur gehen sollten. Diese Veranstaltungen werden auf der Open Archives Initiative HomePage sowie in den Open Archives Mailinglisten bekannt gegeben werden.

### Glossar für einige OAI-Terminologien

Die folgende Begriffe werden hier kurz in dem Sinne definiert, wie sie in den Open Archives Initiative Spezifizierungen und Abhandlungen über OAI benutzt worden sind.

„Archiv“ („open archive“) – Für OAI-Zwecke hat das Wort „archive“ nicht denselben Stellenwert wie „Archivierung“ oder „Archiv“, wie sie im deutschen Sprachgebrauch benutzt werden. Stattdessen trägt der Begriff „archive“ hier die Bedeutung eines Web-Servers, der für die Speicherung und dauerhafte zur Verfügungstellung von elektronischen Informationsressourcen benutzt wird.

Datenanbieter („data provider“) – Eine Institution, die einen oder mehreren Web-Server mit Volltextdaten mit OAI-Metadaten ergänzt und die die Nutzung des OAI-Protokolls unterstützt, um Informationen über ihre Inhalte zur Verfügung zu stellen.

Dienstleistungsanbieter („service provider“) – Ein Informationsdienst, der das OAI-Protokoll benutzt, um Anfragen an OAI-kompatible Datenanbieter zu schicken, und der die sich daraus ergebenden OAI-kompatiblen Metadaten verwendet, um Mehrwertdienste anzubieten.

Eprint-Server – Ein Web-Server, der elektronische Volltexte aber auch andere Formate elektronischer Informationen anbietet.

Preprint-Server – Ein Web-Server, der vor allem Texte in elektronischer Form meistens zu einem bestimmten Fachgebiet speichert und anbietet. Ursprünglich waren dies noch nicht veröffentlichte, aber oft zur Veröffentlichung akzeptierte Texte.

Repository – Ein Web-Server, der Informationen in elektronischer Form speichert und mittels Metadaten über diese Informationen den Zugang zu den Inhalten für den Zweck verschiedenartiger Abfragen verschafft.

### Literatur

- Bowman, C.M., et al.: The Harvest Information Discovery and Access System. In: *Computer Networks and ISDN Systems* 28, 1&2 (1995) S. 119-125.
- Building a „GenBank“ of Published Literature. In: *Science* Mar 23 2001. S. 2318-2319.
- Buck, Anne M.; Flagan, Richard C.; Coles, Betsy: *Scholars' Forum: A New Model for Scholarly Communication*. March 1999 <<http://library.caltech.edu/publications/scholarsforum/>>.

Dehamothe, Tony, et al.: Netprints: the next phase in the evolution of biomedical publishing. In: *British Medical Journal* 319 (1999) S. 1515-1516 <<http://www.bmj.com/cgi/content/full/319/7224/1515>>.

Harnad, Stevan: For Whom the Gate Tolls? How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving, Now. In: <<http://www.cogsci.soton.ac.uk/~harnad/Tp/resolution.htm>>.

Harnad, Stevan: Free at Last: The Future of Peer-Reviewed Journals. In: *D-Lib Magazine* 5, 12 (1999) <<http://www.dlib.org/dlib/december99/12harnad.html>>.

Harnad, Stevan: Implementing Peer Review on the Net: Scientific Quality Control in Scholarly Electronic Journals. In: Peek, R.; Newby, G. (Hrsg.): *Scholarly Publishing: The Electronic Frontier*. Cambridge, MA 1996 S. 103-119.

Harnad, Stevan: Learned Inquiry and the Net: The Role of Peer Review, Peer Commentary and Copyright. In: *Learned Publishing* 11, 4 (1997) S. 283-292.

Harnad, Stevan; Carr, Les: Integrating, Navigating and Analyzing Eprint Archives Through Open Citation Linking (the OpCit Project). In: *Current Science* 79, 5 (2000) S. 629-638.

Judson, Horace Freeland: Structural Transformations of the sciences and the end of peer review. In: *Journal of the American Medical Association* 272, 2 (1994) S. 92-94.

Lagoze, Carl; Van de Sompel, Herbert: The Open Archives Initiative: Building a low-barrier interoperability framework. Paper in press and to be presented at the JCDL, 24.-28. Juni 2001 in Roanoke, Virginia (USA). (in press).

Leiner, B.M.: The NCSTRL Approach to Open Architecture for the Confederated Digital Library. In: *D-Lib Magazine* 5, 12 (1998) <<http://www.dlib.org/dlib/december98/leiner/12leiner.html>>.

Lucier, Richard; Ober, John: Scholar-led Innovation in Scholarly Communication University ePub: An Initiative in Electronic Scholarship. October 1999. In: <<http://www.cdlib.org/eschol/summary.html>>.

PubMed Central: An NIH-Operated Site for Electronic Distribution of Life Sciences Research Reports. August 1999. In: <<http://www.nih.gov/welcome/director/pubmedcentral/pubmedcentral.html>>.

Rötzer, Florian: Wissenschaftler fordern eine zentrale Datenbank für alle Veröffentlichungen. In: c't 26.03.2001. <<http://heise.de/tp/deutsch/inhalt/on/7222/1.html>> [Zugriff am 2.04.01].

Schauder, Don: Electronic publishing of professional articles: attitudes of academics and implications for the scholarly communication industry. In: *Journal of the American Society for Information Science* 45, 2 (1994) S. 73-100.

Stubley, Peter: Clumps as Catalogues. In: *Ariadne* 22 (1999) <<http://www.ariadne.ac.uk/issue22/distributed/distukcat2.html>>.

Van de Sompel, Herbert; Krichel, Thomas; Nelson, Michael L.; et al.: The UPS Prototype: An Experimental End-User Service across E-Print Archives. In: *D-Lib Magazine* 6, 2 (2000) <<http://www.dlib.org/dlib/february00/vandesompel-ups/02vandesompel-ups.html>>.

Varmus, Harold: E-BIOMED: A Proposal for Electronic Publications in the Biomedical Sciences. May 1999. In: <<http://www.nih.gov/welcome/director/pubmedcentral/ebiomedarch.htm>>.

Warner, Simeon: Open Archives Initiative protocol development and implementation at arXiv. 23 January 2001. In: <<http://www.openarchives.org>>.

Wissenschaftler fordern freien Zugriff auf Veröffentlichungen. In: c't 26.03.2001.

### Anschrift der Autorin:

Diann Rusch-Feja  
Max-Planck-Institut für Bildungsforschung  
Lentzeallee 94  
D-14195 Berlin