

## Die Open Archives Initiative

### Ein Bericht vom „Open Meeting“ in Berlin

#### Inka Tappenbeck

Die *Open Archives Initiative (OAI)*<sup>1</sup> ist eine Organisation, die Interoperabilitäts-Standards für den Zugriff auf Daten entwickelt, die auf Dokumentenservern verschiedenster Art über das Internet angeboten werden. Sie hat ihre Wurzeln im Preprint-Bereich und verfolgt das Ziel, die Bedingungen der wissenschaftlichen Kommunikation auf globaler Ebene zu verbessern. Die OAI baut also kein „Archiv“ im Sinne einer Einrichtung zur Langzeitarchivierung von Dokumenten auf, sondern sie entwickelt eine technische und organisatorische Infrastruktur, mittels derer die Recherche in verschiedenen dezentralen Datenbanken über eine definierte Schnittstelle möglich ist. „Offen“ ist die OAI sowohl in Bezug auf ihre Standards als auch in Bezug auf den Kreis der möglichen Teilnehmer. Diese Offenheit ist aber nicht gleichbedeutend mit einer kostenfreien Zugreifbarkeit der gesuchten Dokumente. Die Schnittstellendefinition, die von der OAI entwickelt wurde, um die Interoperabilität zwischen verschiedenen Dokumentenservern zu unterstützen, ist ebenso unabhängig von dem Inhalt der Dokumente wie von den rechtlichen und ökonomischen Bedingungen, die mit ihrer Nutzung verbunden sind. Die OAI ist also keineswegs auf kostenfrei zugreifbare Dokumente beschränkt. Ihr primäres Ziel ist vielmehr die globale Verbreitung von Information über Information. Sie ist daher ebenso für kommerzielle wie für nichtkommerzielle Informationsanbieter interessant.

Seit ihrer Gründung im Oktober 1999 in Santa Fe, New Mexico,<sup>2</sup> fand das erste europäische Treffen der OAI anlässlich der ECDL<sup>3</sup> im September 2000 in Lissabon statt. Die erste offizielle europäische OAI-Tagung wurde dann am 26. Februar 2001 von der Staatsbibliothek zu Berlin<sup>4</sup> ausgerichtet. Sie sollte dazu dienen, die von der OAI in den vergangenen Jahren entwickelte Interoperabilitäts-Architektur in Europa weiter bekannt zu machen. Diese Architektur basiert auf der Technik des „Metadata Harvesting“ und soll den Anbietern und Vermittlern wissenschaftlicher Information eine einfache Möglichkeit bereitstellen, ihre eigenen Metadaten zu veröffentlichen und auf die Metadaten anderer Anbieter zuzugreifen.

---

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

2 <<http://www.openarchives.org/ups1-press.htm>>

3 <<http://www.bn.pt/org/agenda/ecdl2000/>>

4 <<http://www.openarchives.org/Berlin2001/OpenMeeting.html>>

Auf der Tagung in Berlin gaben die Initiatoren der OAI dem überwiegend europäischen Publikum einen Überblick über die Entstehungsgeschichte und organisatorische Struktur ihrer Initiative und stellten das „OAI Protocol for Metadata Harvesting“<sup>5</sup> vor. Daran schlossen sich Präsentationen einiger OAI-Alpha-Tester und früher Anwender an, die bereits mit dem OAI Protokoll arbeiten oder planen, dies in nächster Zukunft zu tun.

Die Eröffnungsansprache hielt *Diann Rusch-Feja* (Max-Planck-Institut für Bildungsforschung, Berlin<sup>6</sup>), die als Gründungsmitglied der OAI deren Entwicklung mit auf den Weg gebracht und bis zum heutigen Tage begleitet hat. Danach gab *Carl Lagoze* (Cornell University, Ithaca / N.Y.<sup>7</sup>), einer der Initiatoren der Open Archives Initiative, einen Überblick über die Arbeit der letzten Jahre. Er betonte, dass die OAI eine sehr experimentelle Initiative sei und einen organisatorischen Rahmen für verschiedene Metadaten-Initiativen bereit stelle. Der Schwerpunkt liegt dabei auf der Entwicklung von Interoperabilitäts-Standards für Preprint-Server. Auf diese Weise soll auch die Akzeptanz von Preprints als seriöse Alternative zu den etablierten, kostenintensiven Fachzeitschriften innerhalb der Wissenschaften erhöht werden. Dass sich Preprints bis heute noch nicht als eine solche Alternative durchgesetzt haben, liegt Lagoze zufolge daran, dass die Nutzer ihre Recherchen bisher zumeist in verschiedenen, untereinander unverbundenen Preprint-Datenbanken mit je eigenen Zugangs- und Suchmodalitäten durchführen mussten, was insbesondere in interdisziplinären Forschungsgebieten einen sehr hohen Aufwand bedeutet. Einen Ausweg aus dieser Situation sieht er daher in der Entwicklung einer einheitlichen Schnittstellendefinition für Preprint-Angebote. Über diese Schnittstelle werden die vereinzelt und zum Teil sehr heterogenen Angebote zu einem „virtual digital space“ verbunden, in dem der Forscher nun fachübergreifend recherchieren kann. Diese Zielperspektive ist in der sog. „Santa Fe Convention“<sup>8</sup> als zentrales Ziel der OAI festgeschrieben worden.

Die von der OAI zur Realisierung der angestrebten „low barrier interoperability“ entwickelte Schnittstellendefinition wurde zum ersten Mal im Frühjahr 2000 im Los Alamos National Laboratory (Santa Fe, New Mexico) implementiert. Schnell weckte dieses erfolgreiche Experiment das Interesse von Bibliotheken und Verlegern. Anlässlich des Treffens der OAI in Lissabon<sup>9</sup> wurde daher beschlossen, technische und organisatorische Konsequenzen zu ziehen und über den Bereich der Preprints hinaus zu gehen. Künftig sollen auch die Be-

---

5 <<http://www.openarchives.org/OAI/openarchivesprotocol.htm>>

6 <<http://www.mpib-berlin.mpg.de/DOK/ehome.htm>>

7 <<http://www.cs.cornell.edu/>>

8 <<http://www.dlib.org/dlib/february00/vandesompele-oai/02vandesompele-oai.html>>

9 <<http://www.bn.pt/org/agenda/ecdl2000/>>

dürfnisse anderer Informationsanbieter in die Entwicklung der Spezifikationen einbezogen und so der Kreis der möglichen Anwender vergrößert werden.

Im Januar 2001 wurde die Version 1.0 des „OAI-Protocol for Metadata Harvesting“<sup>10</sup> der Öffentlichkeit präsentiert. Diese Spezifikation ist nun für eine Testphase von 12 bis 19 Monaten freigegeben. Während dieser Zeit werden daran von der OAI keine Modifikationen durchgeführt, um den experimentellen Einsatz bei den verschiedenen Testanwendern nicht zu destabilisieren. Nach diesem Zeitraum wird die Spezifikation auf der Grundlage der Testergebnisse weiter optimiert werden.

Im Anschluss sprach *Paul Ginsparg* über den Zusammenhang zwischen „OAI and scholarly communications“. Er stellte dar, wie sich die Nutzung von E-Prints in den Naturwissenschaften seit den achtziger Jahren entwickelt hat. Dabei ging er insbesondere auf das seit 1991 von ihm zunächst für den Bereich der Physik aufgebaute E-Print-Archiv „arXiv“<sup>11</sup> ein, in dem die Nutzer über ein gemeinsames Interface in den Beständen verschiedener E-Print-Datenbanken suchen können. Dieses virtuelle elektronische Gesamtarchiv hat kontinuierlich steigende Nutzungsraten zu verzeichnen. Heute liegt die wöchentliche Nutzung zwischen 800.000 und 900.000 Zugriffen. Dabei liegen die Zugriffe aus deutschen Forschungseinrichtungen im weltweiten Vergleich mit an der Spitze.

An diese Beiträge zur Entwicklungsgeschichte und zu den Zielperspektiven der OAI schloss sich eine Einführung in die technische Infrastruktur der OAI-Architektur an. *Paul Lagoze* erläuterte, in welcher Weise über die OAI-Schnittstelle Anfragen (HTTP) an ein sog. „repository“, einen Dokumentenserver, gestellt werden können und diese durch die Lieferung eines Metadatensatzes (XML) beantwortet werden. Dieser enthält sowohl „allgemeine“ Metadaten (Dublin Core Metadata Element Set<sup>12</sup>), die das Dokument selbst beschreiben, als auch anbieterspezifische Metadaten (Nutzungsbedingungen, rechtliche Modalitäten etc.). Die Dublin Core Metadaten werden dafür anhand bestimmter Heuristiken automatisch aus den anwenderspezifischen Metadaten generiert. Ferner hat jeder Metadatensatz einen OAI-spezifischen, weltweit eindeutigen Identifier und einen Zeitstempel, der eine eindeutige Information über den Zeitpunkt der letzten Modifikation der Ressource gibt.

Die Vorträge wurden ergänzt durch mehrere Präsentationen von Alpha-Testern und ersten OAI-Anwendern aus dem In- und Ausland, in denen diese ihre Erfahrungen darstellten. Dabei wurde einhellig betont, dass die Qualität der

---

10 <<http://www.openarchives.org/OAI/openarchivesprotocol.htm>>

11 <<http://arXiv.org/>>

12 <<http://dublincore.org/documents/dces/>>

von den Informationsanbietern bereitgestellten Metadaten von zentraler Bedeutung für die Qualität der Suchergebnisse ist. In diesem Zusammenhang wurde für den Einsatz eines kontrollierten Metadatenvokabulars (Dublin Core Metadata Element Set) mit XML als Syntax plädiert, um die Suche über alle OAI-Archive zu erleichtern. Zugleich war jedoch deutlich, dass auch der Einsatz von Dublin Core nicht weltweit einheitlich ist und eine Standardisierung daher auch in diesem Bereich anzustreben ist. Weiterhin sollten sich zukünftige Bestrebungen nach dem Wunsch der OAI-Erstanwender auf die Unterstützung der Bereitstellung von Metadaten richten, etwa durch die Entwicklung entsprechender Autoren-Tools, sowie auf die Verbesserung von Werkzeugen zur automatischen Metadaten-Konversion.

Die erste OAI-Implementierung in Deutschland ist für den Dokumentenserver der Humboldt Universität in Berlin<sup>13</sup> durchgeführt worden, der seit Februar 2001 OAI-kompatibel ist. Auch die verteilte Datenbank PhysNet<sup>14</sup> soll in diesem Jahr an OAI angeschlossen werden.

Das Open Meeting in Berlin hat als erste eigene Veranstaltung der OAI auf europäischem Boden ihr Ziel, diese Initiative in Europa weiter bekannt zu machen und neue Anwender zu werben, sicher erreicht. Dass die Bedeutung der OAI für die Zukunft des globalen Informationsmarktes auch in Europa inzwischen erkannt worden ist, zeigte sich nicht zuletzt daran, dass jedes europäische Land durch mindestens einen Tagungsteilnehmer in Berlin repräsentiert war.

Informationen über die Aktivitäten der OAI sind über die Homepage der Initiative<sup>15</sup> zugänglich. Dort gibt es auch die Möglichkeit, die laufenden Diskussionen innerhalb der OAI über die Listen „OAI general“ und „OAI implementers“ zu verfolgen.



---

13 <<http://www.hu-Berlin.de/>>

14 <<http://physnet.uni-oldenburg.de/PhysNet/>>

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