

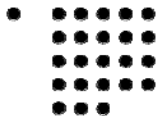
Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH)

Achim Oßwald

FH Köln

Institut für Informationswissenschaft

Wintersemester 2010 (Stand: 3.12.10)



Elektronisches Publizieren

OAI - PMH

FH Köln - Institut für Informationswissenschaft - Prof. Dr. Achim Oßwald

1 / 18

12 / 2010

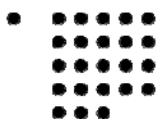
OAI-PMH im Überblick

OAI (Open Archives Initiative)

- Initiative zu Entwicklung und Verbreitung standardisierter Verfahren des freien, einfachen und weltweiten Zugriffs auf Archive / Repositorien digitaler Dokumente
- Will offene, konsistente Schnittstellen für Daten- u. Serviceprovider
- Nutzt ohnehin schon verfügbare Protokolle und Standards (http; XML, Dublin Core (DC))

PMH (Protocol for Metadata Harvesting)

- Einfaches Harvesting-Protokoll auf der Basis von http u. XML zum Einsammeln von DC-basierten Metadaten digitaler Objekte



Elektronisches Publizieren

OAI - PMH

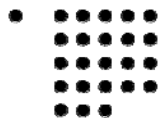
FH Köln - Institut für Informationswissenschaft - Prof. Dr. Achim Oßwald

2 / 18

12 / 2010

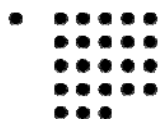
Open Archives Initiative: Entstehung (1)

- Seit Mitte der 90er Jahre entstehen weltweit Archive digitaler Publikationen (Pre- u. Postprints)
- Die verschiedenen Dokumentserver werden mit unterschiedlicher SW betrieben
 - => unterschiedliche IR-Oberflächen
 - => unterschiedliche Abfragesprachen
 - => interessierte Wissenschaftler müssen verschiedene Suchoberflächen lernen
- ➔ Wunsch nach EINER Suchoberfläche mit EINER Abfragesprache
damals bezeichnet als „UPS / Universal Preprint Service“, jedoch schnell umbenannt in Open Archives Initiative
<http://www.openarchives.org/news/ups1-press.htm>



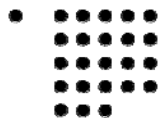
Open Archives Initiative: Entstehung (2)

- Zwei Konzepte bieten sich hierfür an:
 1. Suche über die verschiedenen Archive mittels einer DB-übergreifenden Suchtechnik
= Z39.50 / SRW (Search and Retrieve Web Service) = sog. „cross database search“
 2. Metadaten-Harvesting und verlinken auf die Original-Speicherorte
- ➔ Harvesting-Ansatz war perspektivisch realistischer und tragfähiger



Open Archives Initiative: Entstehung (3)

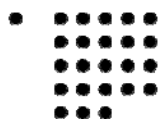
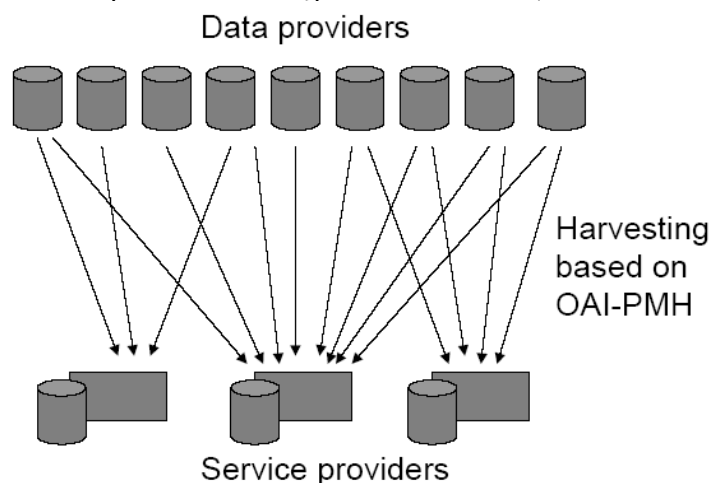
- Erstmalige Präsentation eines Prototyps für ein derartiges Konzept in Santa Fe, New Mexico, Oktober 1999 (Santa Fe Convention)
siehe: D-LIB Magazine 6 (2) 2000
<http://www.dlib.org/dlib/february00/vandesompe/oai/02vandesompe/oai.html>
- Metadaten-Archiv, dessen Inhalt aus den Metadaten mittels Harvesting von vielen verschiedenen Archiven zusammen "geerntet" wird



OAI: Daten- u. Serviceprovider (1)

- *Datenprovider / data provider*
„A data provider maintains one or more repositories (web servers) that support the OAI-PMH as a means of exposing metadata.“

<http://www.openarchives.org/documents/FAQ.html> v. 10.6.2002



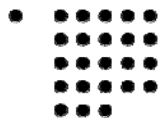
OAI- Daten- u. Serviceprovider (2)

➤ *Serviceprovider / service provider*

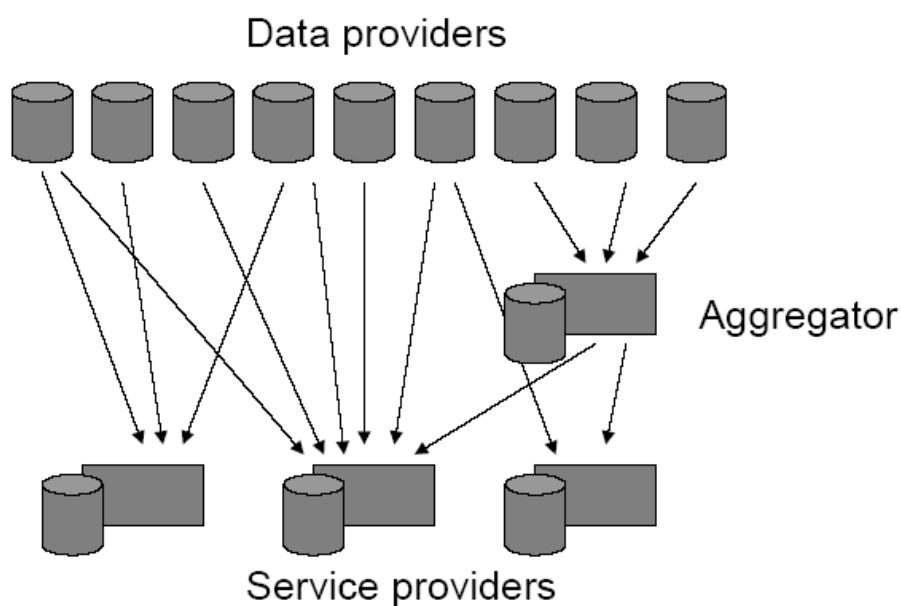
„A service provider issues OAI-PMH requests to data providers and uses the metadata as a basis for building value-added services.“

<http://www.openarchives.org/documents/FAQ.html> v. 10.6.2002

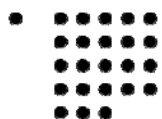
- Aber auch ein Datenprovider kann den Zugriff auf die von ihm angebotenen (Meta-)Daten mittels einer Suchoberfläche anbieten



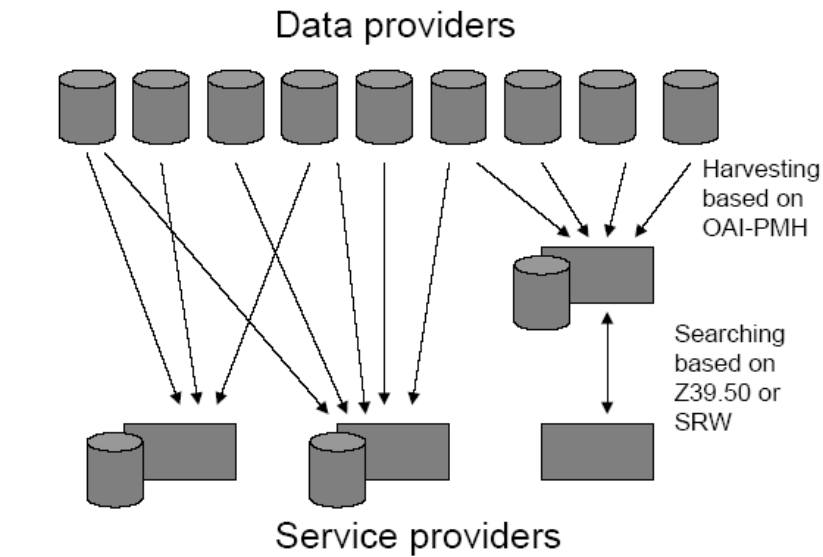
OAI: Daten u. Service Provider / Aggregatoren



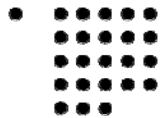
<http://www.oaforum.org/tutorial/english/page2.htm>



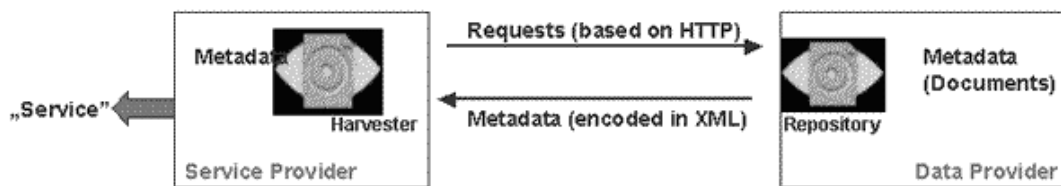
OAI: Kombination von Harvesting und Suche via Z39.50 / SRW



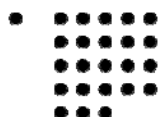
<http://www.oaforum.org/tutorial/english/page2.htm>



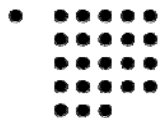
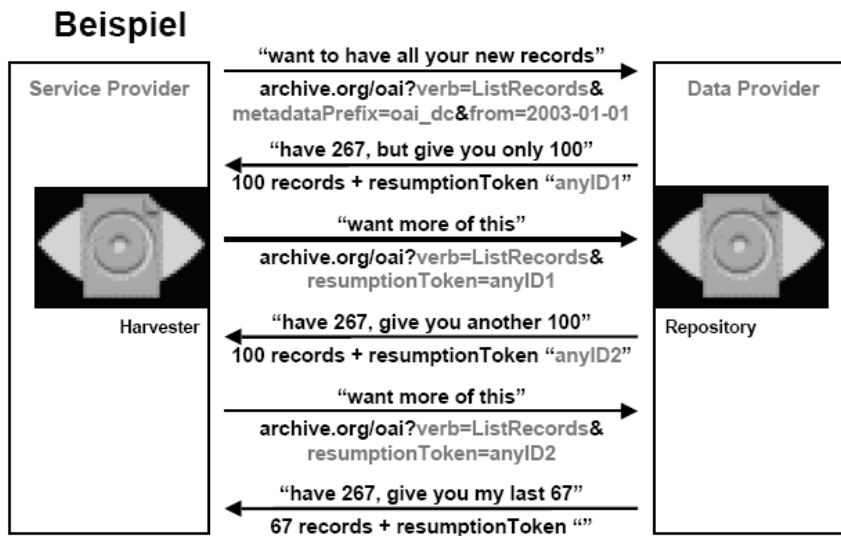
OAI: Basis-Funktionalität



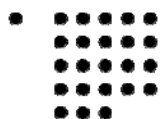
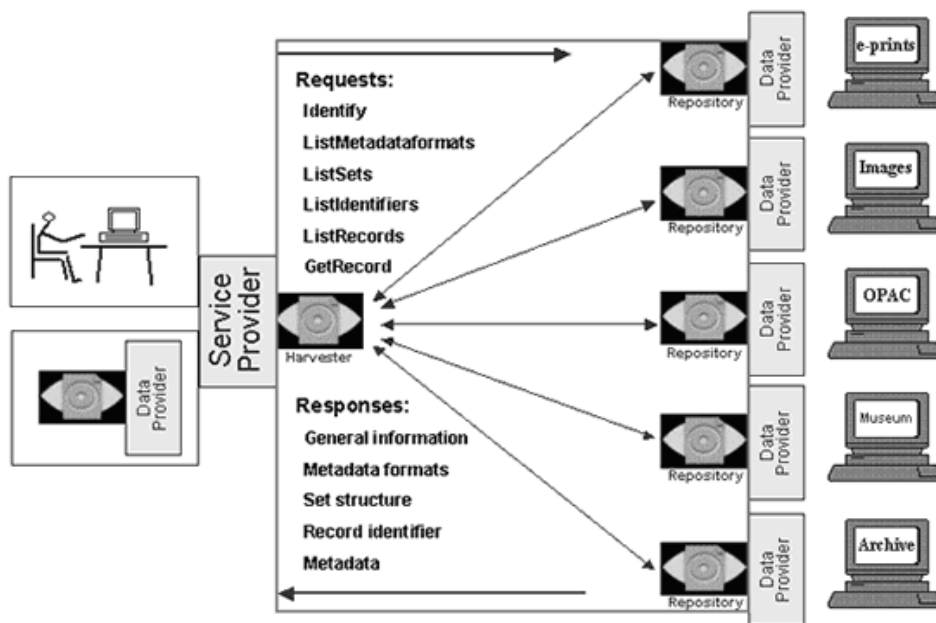
<http://www.oaforum.org/tutorial/english/page3.htm>



Beispiel für OAI-PMH-Dialog

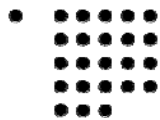


OAI: Funktions- und Sprachumfang



Object Reuse and Exchange

- „Open Archives Initiative Object Reuse and Exchange (OAI-ORE) defines standards for the description and exchange of aggregations of Web resources. These aggregations, sometimes called compound digital objects, may combine distributed resources with multiple media types including text, images, data, and video. The goal of these standards is to expose the rich content in these aggregations to applications that support authoring, deposit, exchange, visualization, reuse, and preservation.“
<http://www.openarchives.org/ore/>

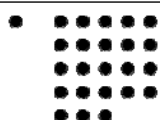


Weiterentwicklung im OAI-Bereich: Object Reuse and Exchange

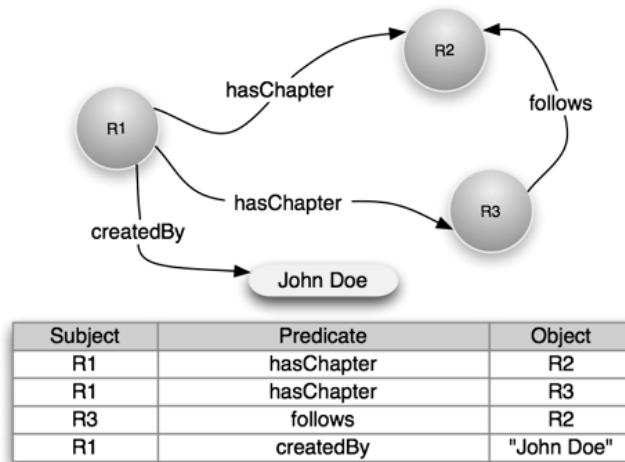
The screenshot shows the arXiv interface for the document 'Parametrization of K-essence and Its Kinetic Term'. Numbered callouts point to the following elements:

- 1: Browser address bar showing the URL <http://arxiv.org/abs/astro-ph/0601007>.
- 2: 'Download:' section with options for PostScript, PDF, and Other formats.
- 3: Document title: 'Parametrization of K-essence and Its Kinetic Term'.
- 4: Author names: 'Hui Li, Zong-Kuan Guo, Yuan-Zhong Zhang'.
- 5: Submission information: '(Submitted on 31 Dec 2005 (v1), last revised 18 Jan 2006 (this version, v2))'.
- 6: Metadata section including 'Comments', 'Subjects', 'Journal reference', 'DOI', and 'Cite as'.
- 7: 'Submission history' section showing the document's revision dates and sizes.
- 8: 'Current browse context:' section showing the document's location within the 'astro-ph' category.
- 9: 'References & Citations' section with links to SLAC-SPHRES HEP, NASA ADS, and CiteBase.

„Figure 1: Human start page for an arXiv document“ aus: ORE User Guide – Primer, 17 October 2008 <http://www.openarchives.org/ore/1.0/primer.html>



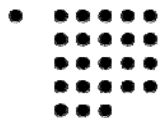
Object Reuse and Exchange: Resource Description Framework (RDF)



„Figure 2: Four RDF triples and their graph representation “

aus: ORE User Guide – Primer, 17 October 2008

<http://www.openarchives.org/ore/1.0/primer.html>



Elektronisches Publizieren

OAI - PMH

FH Köln - Institut für Informationswissenschaft - Prof. Dr. Achim Oßwald

15 / 18

12 / 2010

Object Reuse and Exchange: Aggregation und Resource Map

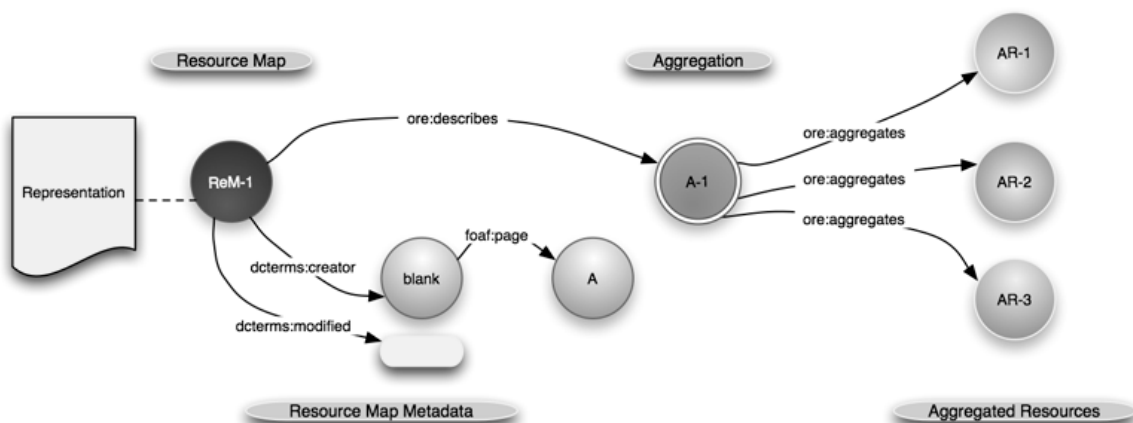
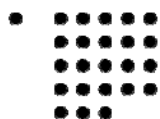


Figure 3:

The Aggregation A-1 aggregates three Resources and is described by Resource Map ReM-1

aus: ORE User Guide – Primer, 17 October 2008; <http://www.openarchives.org/ore/1.0/primer.html>



Elektronisches Publizieren

OAI - PMH

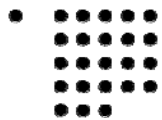
FH Köln - Institut für Informationswissenschaft - Prof. Dr. Achim Oßwald

16 / 18

12 / 2010

Object Reuse and Exchange

- Resource „Aggregation“ als Zusammenfassung mehrerer Webressourcen („stands for a set or collection of other Resources“), die z.B. über eine Webseite zugänglich gemacht werden
- Aggregation ist ein „conceptual construct“ entsprechend dem Semantic Web Konzept, das mit einem URI adressiert wird
- Dieser Aggregation wird eine sog. „Resource Map“ zugeordnet, die maschinenlesbare Information über die Aggregation bereitstellt



OAI-PMH: Einführende Quellen

- Homepage der OAI-Initiative
<http://www.openarchives.org/>
- OAI-Tutorial
<http://www.oaforum.org/tutorial/>
- DINI: Online-Publikationen in Deutschland (OAI-Suche)
<http://miles.cms.hu-berlin.de/oaisuche/oai.php>
- Object Reuse and Exchange
<http://www.openarchives.org/ore/> bzw.
<http://www.openarchives.org/ore/1.0/primer.html>

Alle URLs wurden zuletzt am 3.12.2010 überprüft.

