



Tutorial 3

OAI and OAI-PMH for absolute beginners a non-technical introduction

Philip Hunter

IRIScotland Project, University of Edinburgh, United Kingdom

philip.hunter@ed.ac.uk

Tutorial development also generously supported by UKOLN
at the University of Bath



Overview of the morning

- Overview and Introductions
- Part I
 - History and overview
- Short break (10.15 am)
- Quiz
- Part II
 - Main Ideas of the OAI-PMH
- Part III
 - Implementation issues



Acknowledgements

- These slides have a long history!
- Many of them have been kindly donated by (taken from!)

Herbert Van de Sompel

Carl Lagoze

Michael Nelson

Simeon Warner

Andy Powell

Pete Cliff

Uwe Muller

Monica Duke

(and others probably!)



My Background in OAI PMH

- Long interested in Scholarly Communications issues (more later)
- Covered developments with Ariadne (SPARC onwards)
- Began association with the Open Archives Forum in 2001
- Project Manager for ePrints UK in 2003
- ePrints UK implemented a harvester using OAI PMH
- Now at IRIScotland project in Edinburgh



Tutorial 3

OAI and OAI-PMH for absolute beginners

**An introduction to the Open Archives Initiative
and the Protocol for Metadata Harvesting**

Part I: History and basic concepts



The Open Archives Approach

- Facilitates access to heterogenous web-accessible material
- A low-barrier interoperability solution
- Based on repositories supporting
 - Metadata sharing
 - Publishing
 - Archiving
- Arose out of the e-print community
- 2 main features
 - Open Archives Initiative
 - OAI Protocol for Metadata Harvesting (OAI-PMH)



The Open Archives Initiative

- **Mission**
 - "The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content."
- **Executive for management, steering and technical committees**
- **Funding**
 - Digital Library Federation (DLF)
 - National Science Foundation (NSF)
 - Coalition for Networked Information (CNI)
- **Participation of a world-wide community, especially Europe and North America**



OAI-PMH

- A mechanism for harvesting
- Data providers make metadata available for harvesting
- Service Providers harvest metadata
- Metadata can be centrally collected or “aggregated”
- That’s all it is: a way to bring metadata together in one place!



Open Archives Forum Tutorial

- Task List Page
- Task 1 Seven key definitions
- Web link

<http://www.oaforum.org/tutorial/english/page1.htm#section3>



A History Lesson - Roots of OAI

- Early activity: scholarly research (eprints archive)
 - XXX (arXiv) – high energy physics
 - CogPrints - psychology
 - NCSTRL – computer science technical reports
 - RePEc - economics
- Web interfaces for people
 - No machine interfaces
- Different interfaces for different archives
- End Users forced to learn diverse interfaces
- Little or no autonomous metadata sharing

Slide 10

P. J.1

Good place for demo of arXiv site

Hunter, 10/16/2005



Santa Fe Meeting

- “...the joint impact of these and future initiatives can be substantially higher when interoperability between them [e-print archives] can be established...”

[Ginsparg, Luce, Van de Sompel, UPS Call, July 1999]



The Problems

Two problems:

- End users were/are faced with multiple search interfaces making resource discovery harder.
- No machine based way of sharing the metadata



Cross Search?

- US Digital Library Experience suggests cross searching doesn't scale - $N > 100 = \text{bad!}$
- Collection description - knowing which target to use
- Query language and search attribute variation
- Rank merging problem
- Different size and type of target can skew results
- Performance - limited to slowest target
- Difficult to build a browse interface

SOLUTION: get all the metadata records in one place



The idea of Harvesting

- Harvest records out of archives into one place
- Universal Preprint Service Prototype

So:

- $N = 1$ most of the time...
- One query language, set of search attributes and ranking algorithm
- An awareness of the data makes browse structures easier to build
- UPS was quickly changed to OAI - the Open Archives Initiative



Terminological Pause

- The Open Archives Initiative name implies an archival aspect of the protocol – what characteristics would you generally expect of something which could be called an archive?
- The Open Archives Initiative also makes use of the term ‘repository’. What sort of thing would you expect of an entity called a repository?
- And.... could a repository also be an archive?



Terminological Surprise

- The Open Archives Initiative use of the term 'archive' in fact implies little that we normally associate with archives
- No preservation aspect is implied *whatsoever* (not what the protocol is about at all)
- No appraisal or provenance of eprints or digital objects is implied by this descriptive term
- Simply refers to a collection of digital objects (full text, learning objects, etc.) which *might* (only *might*) also have been harvested along with the metadata
- 'Archive' is a term within the OAI PMH which (strictly) serves to distinguish a collection of digital objects (the 'archive') from the collected metadata associated with these objects, described as 'repositories'
- In the context of the OAI PMH, these terms serve only to distinguish between these collections of data
- But....



Terminological surprise (2)

Repositories expose metadata about ePrints (strictly there are no metadata archives)

Archives hold ePrints (strictly there are no eprint repositories)

OAI definition of 'archive'

Repository metadata points to ePrints in Archives

Repository metadata can point to document locations where these items are not in digital format



Some Background to the Open Archives Initiative

- Origins of the OAI
 - Los Alamos Physics Pre-Print Archive
 - What is a Pre-Print?
 - About speeding up Scholarly Communication
 - Opening access to interested communities
 - Santa Fe Convention, 1999
 - Technical Goals of the OAI



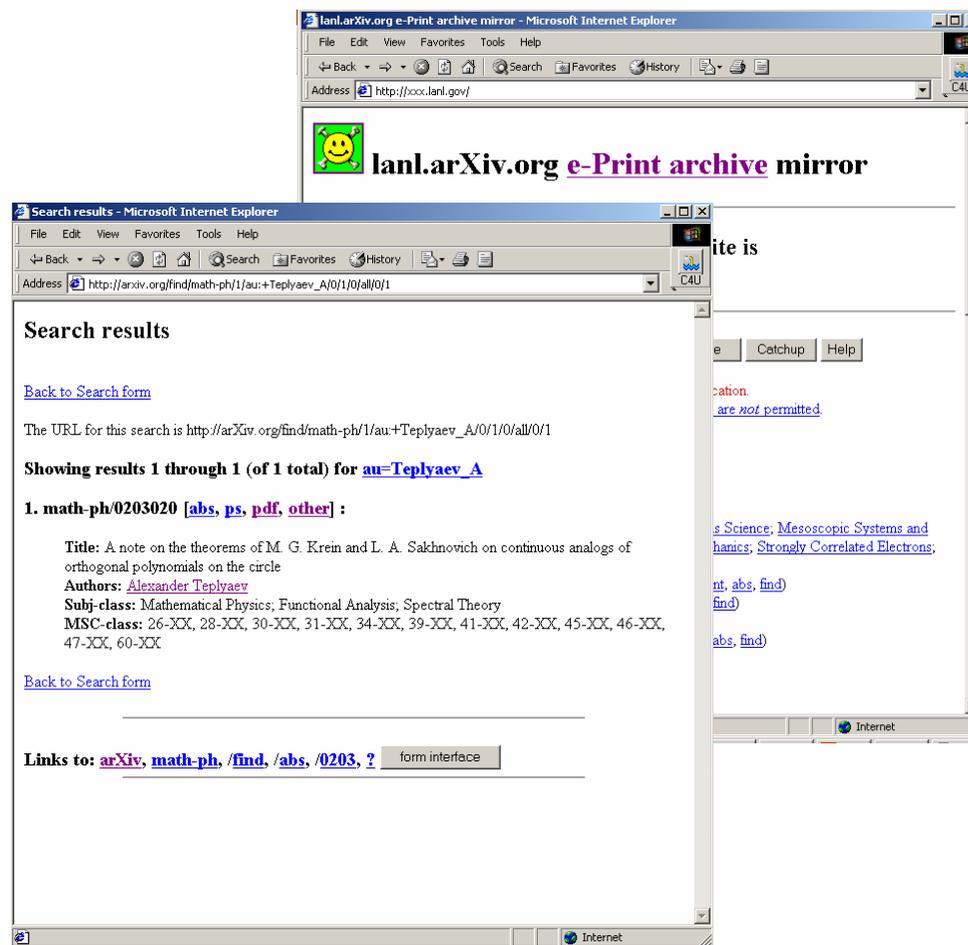
Some Background to the Open Archives Initiative (2)

- The Rising Cost of Scholarly Communication and the Response
 - The Harnad Analysis
 - The SPARC Initiative
 - The OAI as part of the response to the crisis in scholarly publishing
 - ePrint archives as a more equitable and efficient model for dissemination of research



The First Preprint Server

- Established 1991 at Los Alamos
- Moved to Cornell University 2001
- Papers available on the day of submission



<http://www.arxiv.org>



ePrints

- What is an ePrint?

Preprints

Postprints

ePrints (is there a clear definition?)

‘Unidentified Document-like Objects’, or something else?



Resources and the OAI PMH

- What kind of materials might be made available as ePrints?
 - Peer reviewed papers (postprints)
 - Grey literature (university reports, departmental documents)
 - Theses
 - Collections (as Collection Level Descriptions)
 - Images
 - Multimedia and e-Learning Objects
 - Virtually anything



Data and Service Providers

- **Data Providers**

Are creators and keepers of the metadata for objects (repositories) and (possibly but not necessarily) archives of resources

Handle deposit and publishing
- **Service Providers**

Are harvesters of metadata for the purpose of providing a service such as a search interface, peer-review system, etc.
- **One 'service' can play both roles**



The Dawn of a Protocol

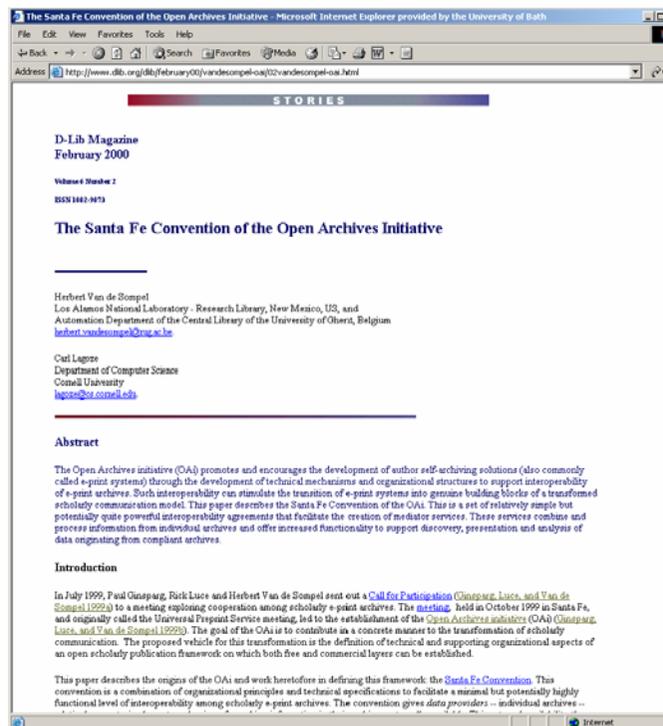
To facilitate metadata harvesting there needs to be agreement on:

- Transport protocol - HTTP or FTP or ...
- Metadata format - Dublin Core or MARC or ...
- Metadata Quality Assurance - mandatory element set, naming and subject conventions, etc.
- Intellectual Property and Usage Rights - *who* can do *what* with *what*?
- Agreement led to (fanfare): the Santa Fe Convention



Santa Fe Convention

- Santa Fe Convention, 1999
- An account of the meeting and the motives behind the proposal for the protocol published in D-Lib magazine, February 2000.



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



Santa Fe Convention



Richard Luce - Herbert Van de Sompel - Paul Ginsparg

[Picture of the Santa Fe Convention, 'borrowed' from a presentation by H. Van de Sompel]

CERN Workshop on Innovations in Scholarly Communications (OAI4) 20-22 October 2005



Santa Fe Convention

- First incarnation of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
- Drew upon:
 - The UPS Prototype
 - RePEc/SODA - the Service/Data provider model
 - the Dienst Protocol
 - Work of the Santa Fe group
- To “optimise the discovery of e-prints”



The Open Archives Initiative

OAI:

- Lagoze, Van de Sompel and lots of great people
- Santa Fe Convention [1999]
- OAI-PMH v.1 [2001]
- OAI-PMH v.2 [2002]

OAI-PMH:

- simple, generic protocol to harvest structured data
- HTTP based
- responses are valid XML instance documents
- unqualified Dublin Core as mandatory metadata format
- /use of other metadata formats encouraged



The OAI-PMH 1.0

- Introduced Dublin Core element set
- Drew upon:
 - Santa Fe Convention
 - Digital Library Federation meetings
 - Work at Cornell
 - Feedback from alpha-testers
- A new focus to facilitate the discovery of “document-like objects”



The OAI-PMH 1.0 - Summary

- Low barrier interoperability specification
- Based around metadata harvesting model
- Focus on “document-like objects”
- HTTP based
- GET / POST requests
- XML responses
- Uses unqualified Dublin Core
- Not a search protocol!
- **Experimental**



The OAI-PMH 1.1

- A revision of the 1.0 specification taking account of changes to the emerging XML Schema specification



The OAI-PMH 2.0

- Major revision - not compatible with 1.x
- Drew upon:
 - OAI-PMH 1.x
 - Feedback from OAI Implementers List
 - OAI tech deliberation
 - Feedback from alpha-testers
- “the recurrent exchange of metadata about *resources* between systems”



The OAI-PMH 2.0 - Summary

- Still a low barrier interoperability specification
- Based around metadata harvesting model
- **Metadata about resources**
- HTTP based
- GET / POST requests
- XML responses
- Uses unqualified Dublin Core
- Not a search protocol!
- **Stable - OAI has committed to making subsequent revisions of the protocol backwards compatible**



Santa Fe
convention

OAI-PMH
v.1.0/1.1

OAI-PMH
v.2.0

nature

experimental

experimental

stable

verbs

Dienst

OAI-PMH

OAI-PMH

requests

HTTP GET/POST

HTTP GET/POST

HTTP GET/POST

responses

XML

XML

XML

transport

HTTP

HTTP

HTTP

metadata

OAMS

unqualified
Dublin Core

unqualified
Dublin Core

about

eprints

document
like objects

resources

model

metadata
harvesting

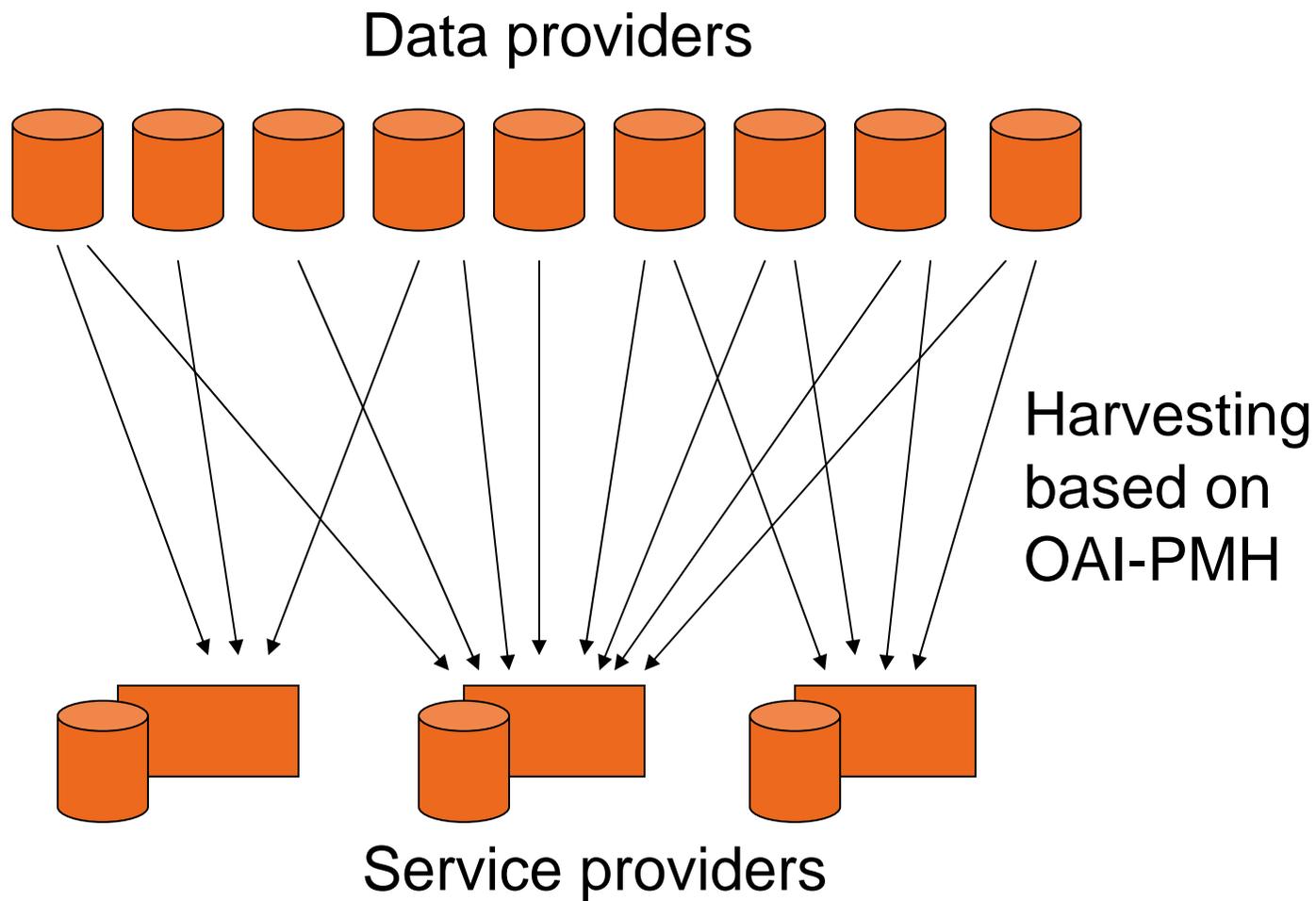
metadata
harvesting

metadata
harvesting

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

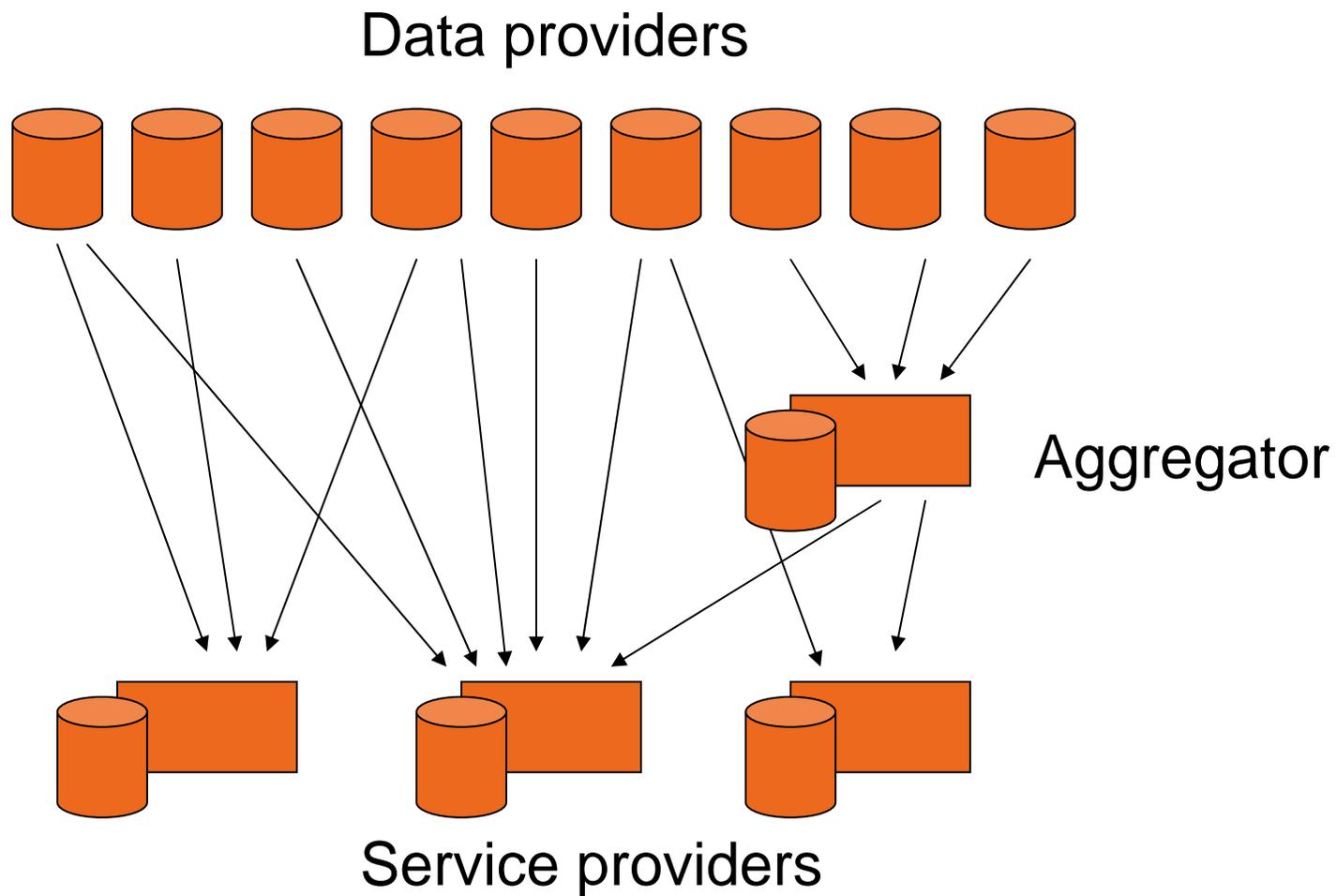


Multiple data and service providers



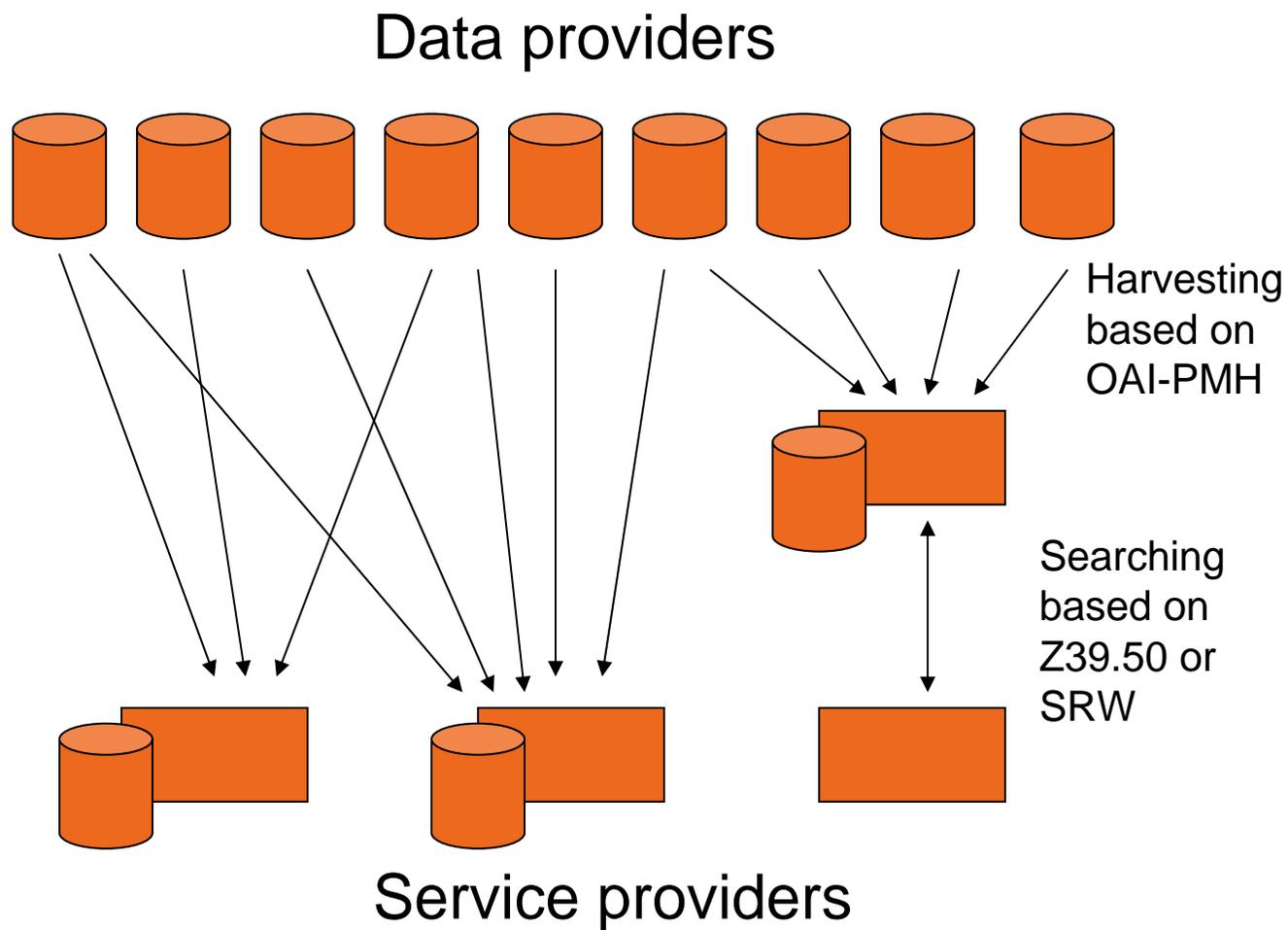


Aggregators





Can be mixed with x-searching





Accessing aggregated records

Task 2:

XXX (arXiv) – high energy physics

<http://www.arxiv.org/>

CogPrints – psychology

<http://cogprints.org/>

NCSTRL – computer science technical reports

<http://www.ncstrl.org/>

RePEc – economics

<http://repec.org/>

Slide 38

P. J.6

Good place for demo of arXiv site

Hunter, 10/16/2005



The Benefits of OAI-PMH

- Simple
- Web (and so firewall) friendly
- Access-control, compression, error codes, etc. based on HTTP
- Many toolkits - can hide the protocol from developers
- Multiple SPs can harvest from multiple DPs ensuring a wider spread of metadata
- A base layer to build other services on
- Complements search protocols like Z39.50



Summary So Far

- Early movers were developing separately
- Need for interoperability
- Santa Fe Meeting led to OAI
- OAI promotes interoperability via:
- OAI-PMH
 - Low cost
 - Harvesting model
 - Data Providers / Service Providers
 - Simple, easy and built on existing technology
 - An open standard



Open Archives Forum Tutorial

- Task Page
- Task 3: Sources of further information
- Web link

<http://www.oaforum.org/tutorial/english/page2.htm#section9>



Tutorial 3

OAI and OAI-PMH for absolute beginners

**An introduction to the Open Archives Initiative
and the Protocol for Metadata Harvesting**

Part II: Main Ideas of OAI-PMH



Open Archives Forum Tutorial

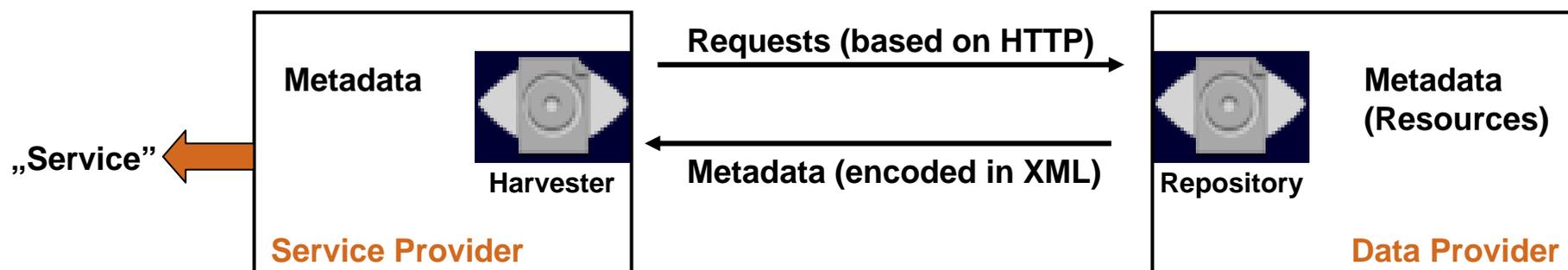
- Task Page
- Task 4: Quiz
- Web link

<http://www.oaforum.org/tutorial/english/page1.htm#section5>



The Open Archives Initiative (OAI)

- Main ideas
 - free access on the archives (at least: metadata)
 - consistent interfaces for archives and service provider
 - low barrier protocol / effortless implementation
 - based on existing standards (e.g. HTTP, XML, DC)
- Basic functioning of protocol



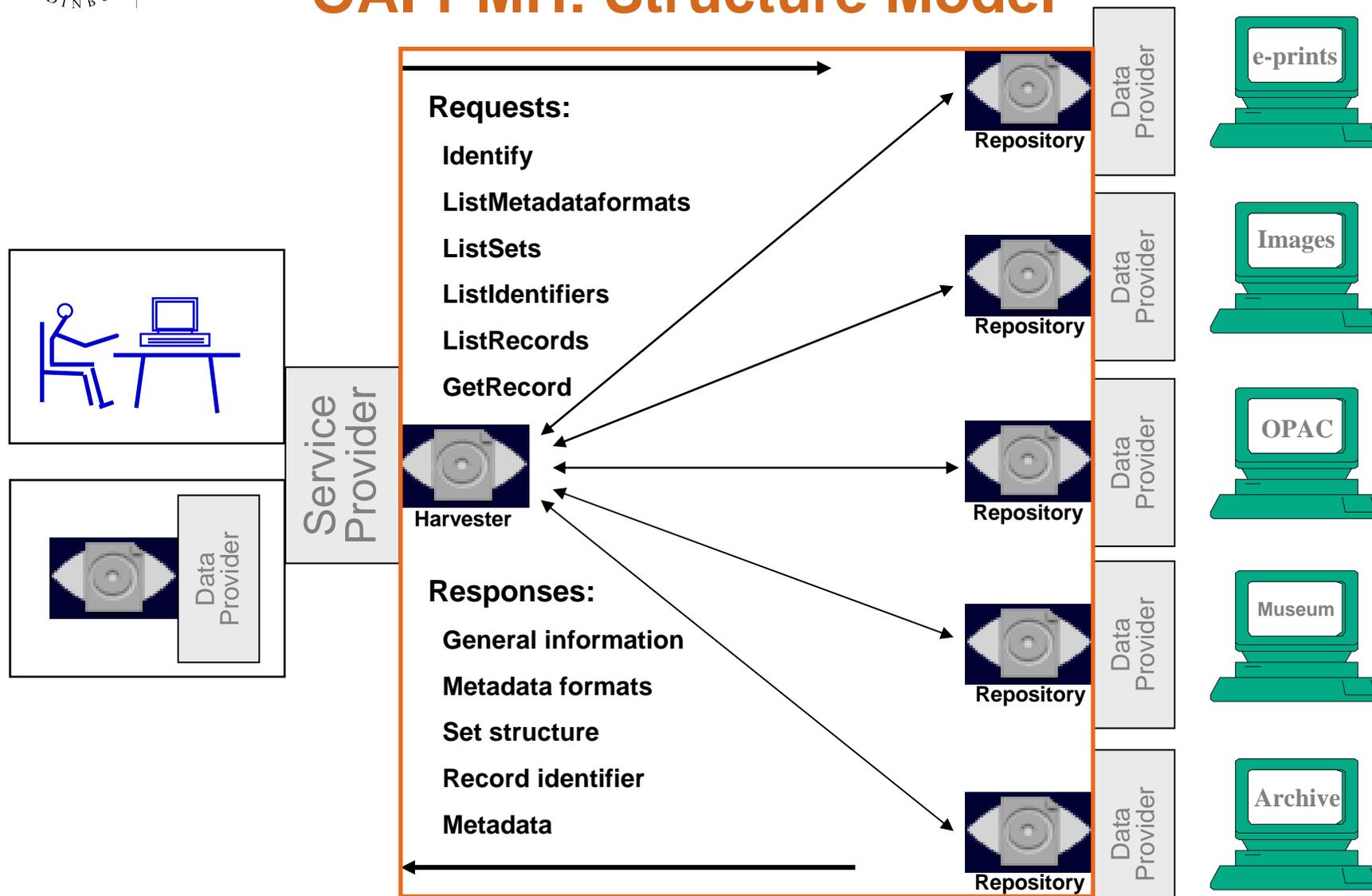


OAI: General Assumptions

- Two groups of 'participants'
- *Data Providers* (Open Archives, Repositories)
 - give free access to metadata
 - but - not necessarily give free access to full texts / resources
 - easy to implement, low barriers
- *Service Providers*
 - use OAI compliant metadata from *Data Providers*
 - harvest and store metadata (no live requests!)
 - may select certain subsets from *Data Providers*
(set hierarchy, date stamp)
 - may enrich metadata
 - offer (value-added) service on the basis of the metadata



OAI-PMH: Structure Model





OAI-PMH: Protocol Overview

- protocol based on HTTP
- request arguments as GET or POST parameters
- six request types
- e.g. `http://archive.org?verb=ListRecords&from=2002-11-01`
- responses are encoded in XML syntax
- supports any metadata format (at least: Dublin Core)
- logical set hierarchy (definition: data providers)
- date stamps (last change of metadata set)
- error messages
- flow control



Protocol Details: Definitions

Harvester

- client application issuing OAI-PMH requests

Repository

- network accessible server, able to process OAI-PMH requests correctly

Resource

- object the metadata is “about”, nature of resources is not defined in the OAI-PMH

Item

- component of an repository from which metadata about a resource can be disseminated
- has an unique identifier

Record

- metadata in a specific metadata format

Identifier

- unique key for an item in a repository

Set

- optional construct for grouping items in a repository



Protocol Details: Definitions (2)



← resource

item =
identifier

all available metadata
about *David*

← item

Dublin Core
metadata

MARC
metadata

SPECTRUM
metadata

← records



Protocol Details: Records

- metadata of a resource in a specific format
- three parts
 1. header (mandatory)
 - identifier (1)
 - datestamp (1)
 2. metadata (mandatory)
 - XML encoded metadata with root tag, namespace
 - repositories must support Dublin Core
 - May support other formats
 3. about (optional)
 - rights statements
 - provenance statements



Protocol Details: Metadata Schema

- OAI-PMH supports dissemination of multiple metadata formats from a repository
- properties of metadata formats
 - id string to specify the format (`metadataPrefix`)
 - metadata schema URL (XML schema to test validity)
 - XML namespace URI (global identifier for metadata format)
- repositories must be able to disseminate unqualified Dublin Core
- arbitrary metadata formats can be defined and transported via the OAI-PMH
- returned metadata must comply with XML namespace specification



Protocol Details: Metadata Schema (2)

- minimum standard: unqualified Dublin Core
 - <http://dublincore.org/>
 - Dublin Core Metadata Element Set contains 15 elements
 - elements are optional
 - elements may be repeated

The Dublin Core Metadata Element Set:

Title	Contributor	Source
Creator	Date	Language
Subject	Type	Relation
Description	Format	Coverage
Publisher	Identifier	Rights



Request Types

- six different request types
 1. Identify
 2. ListMetadataFormats
 3. ListSets
 4. ListIdentifiers
 5. ListRecords
 6. GetRecord
- Not obligatory for harvester to use all types
- repository must implement all types
- required and optional arguments
- depend on request types



Example: <http://edoc.hu-berlin.de/OAI-2.0?>

verb=ListIdentifiers&from=2002-01-06&until=2002-01-08&metadataPrefix=oai_dc&set=doctypes:dissertations

```
<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2002-10-22T17:49:49+01:00</responseDate>
  <request verb="ListIdentifiers" from="2002-01-03" until="2002-01-08" metadataPrefix="oai_dc"
    set="doctypes:dissertations">http://edoc.hu-berlin.de/OAI-2.0</request>
  <ListIdentifiers>
    <header>
      <identifier>oai:HUBerlin.de:3000819</identifier>
      <timestamp>2002-01-08</timestamp>
      <setSpec>doctypes</setSpec>
      <setSpec>doctypes:dissertations</setSpec>
      <setSpec>dnb</setSpec>
      <setSpec>dnb:dnb33</setSpec>
    </header>
    <header>
      <identifier>oai:HUBerlin.de:3000831</identifier>
      <timestamp>2002-01-07</timestamp>
      <setSpec>doctypes</setSpec>
      <setSpec>doctypes:dissertations</setSpec>
      <setSpec>dnb</setSpec>
      <setSpec>dnb:dnb27</setSpec>
    </header>
  </ListIdentifiers>
</OAI-PMH>
```



Protocol Details: Sets

- Logical partitioning of repositories
- Optional – archives do not have to define sets
- No recommendations
- Also support selective harvesting
- Useful sets are defined by the community where they are used:
 - publication types (thesis, article, ...)
 - document types (text, audio, image, ...)
 - content sets, according to DNB (medicine, biology, ...)



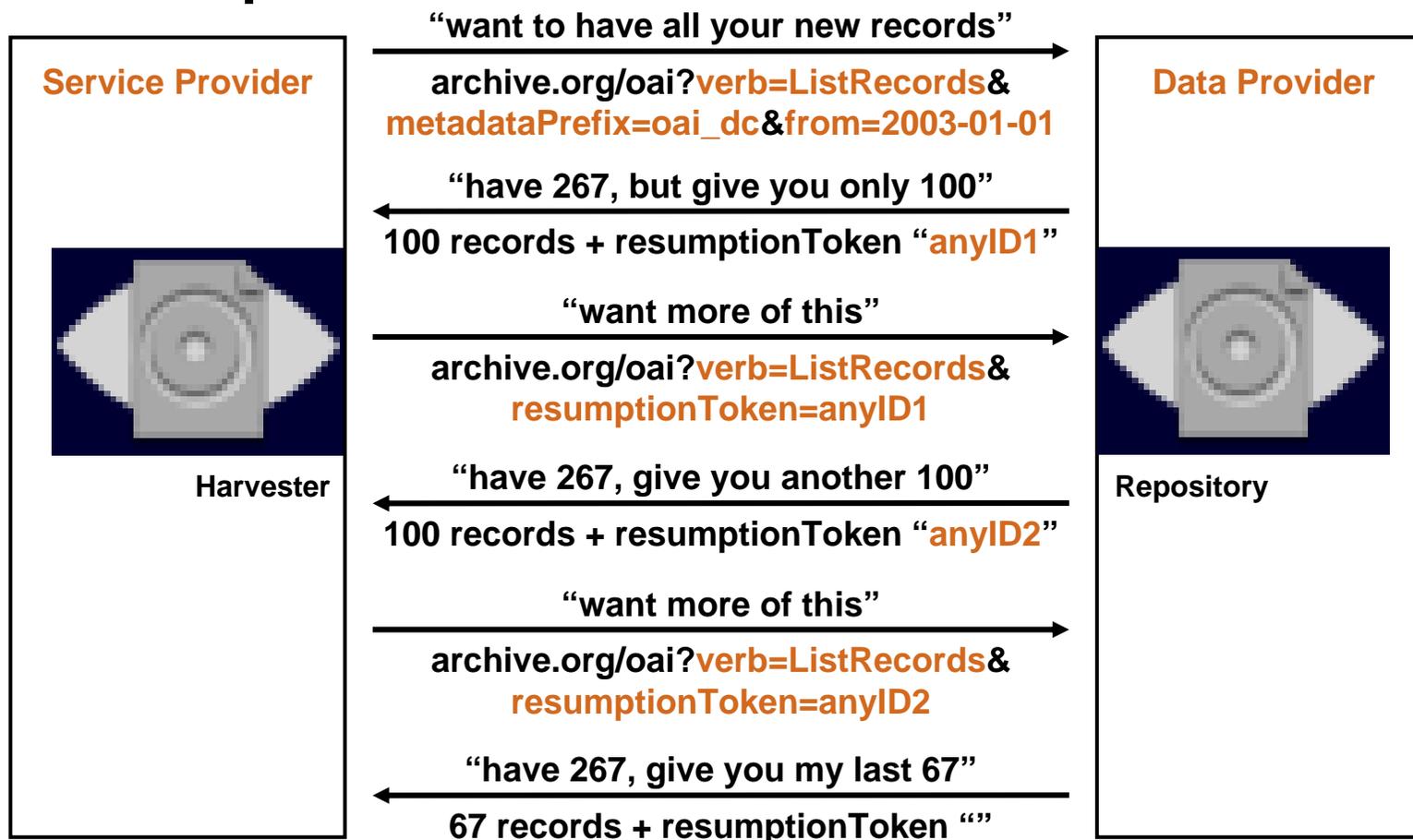
Protocol Details: Datestamps

- date of last modification of a metadata set
- mandatory characteristic of every item
- enables selective harvesting



Protocol Details: Flow control

Example





Task 5: Using Repository Explorer

- <http://oai.dlib.vt.edu/cgi-bin/Explorer/oai2.0/testoai>
- **Tasks**
 - Scroll down the alphabetical list to find the arXiv repository
 - Click on the Identify link in the Verbs box
 - Click on the list Metadata Formats link
 - Copy oai_dc into the MetadataPrefix box in the parameters section
 - Click on ListRecords
 - Copy the identifier from the header section of the first result, scroll to the bottom of the page and paste the identifier into the identifier box of the parameters section
 - Select raw XML in the display section and click GetRecord in the verbs section



Tutorial 3

OAI and OAI-PMH for Beginners

An introduction to the Open Archives Initiative
and the Protocol for Metadata Harvesting

Part III: Implementation Issues



Agenda

1. Data Provider or Service Provider
2. Metadata Records
3. Tools and services
4. Examples



General: First Questions

Data Provider

Which data do I want to deliver?

Which service providers do I want to provide with data?

Service Provider

Which Service do I want to provide?

From which data providers do I get the metadata?

In which way the metadata have to be processed?

Data Provider & Service Provider

Which aspects do we have to agree upon?



General: Metadata Formats / Sets

- required: unqualified Dublin Core
- special subjects / communities: other metadata specifications may be required
 - describe resources in a specialised way
 - definition of an XML schema (publicly available for validation)
- define set hierarchy
 - sensible partitioning for selective harvesting
 - agreement between data providers and between data and service providers



General: Organisational Structure

- aggregated data providers
 - if harvested by a service provider, “sub data providers” should not be harvested by same SP (duplication ...)
- subject gateways
 - selective harvesting if corresponding sets have been defined and implemented



Data Provider: Prerequisites

- metadata on resources (“items”)
 - should be stored in (SQL) database
 - possible in case of need: file system ...
 - unique identifier for each item
- web server, accessible via the internet
 - e.g. apache, IIS
- programming interface / API
 - e.g. Perl, PHP, Java-Servlet
 - web server extension
 - access to database (or filesystem)
 - not needed: session management



Data Provider: Prerequisites (2)

- archive identifier / base URL
- unique identifier for items
- metadata format (at least: unqualified Dublin Core)
- timestamps for metadata (created / last modified)
- logical set hierarchy (may have)
 - agreement within (subject) communities
- flow control / implementation of resumption token (optional, 'larger' archives should have that)



Service Provider: Prerequisites

- internet connected server
- database system (relational or XML)
- programming environment
 - can issue HTTP requests to web servers
 - can issue database requests
 - XML parser



Agenda

1. Data Provider or Service Provider
2. Metadata Records
3. Tools and services
4. Examples



The Basics

- OAI-PMH uses XML Schemas
- Schemas described what is allowed in an XML document
- Schemas have a 'name' (namespace)
- Schemas have a physical location (commonly on the web)
- Example

Namespace

http://www.openarchives.org/OAI/2.0/oai_dc/

http://www.openarchives.org/OAI/2.0/oai_dc.xsd

Location



The Basics (2)

- Any XML with an XML Schema = OK for OAI!
- OAI-PMH mandates 'oai_dc' schema
- OAI-PMH documentation includes schema for
 - RFC1807 metadata
 - MARC21 metadata (Library of Congress)
 - oai_marc metadata



Example: <http://edoc.hu-berlin.de/OAI-2.0?>

verb=GetRecord&identifier=oai:HUBerlin:3000819&metadataPrefix=oai_dc

```
<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2002-11-27T14:57:01+01:00</responseDate>
  <request verb="GetRecord" metadataPrefix="oai_dc"
    identifier="oai:HUBerlin.de:3000819">http://edoc.hu-berlin.de/OAI-2.0</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:HUBerlin.de:3000819</identifier>
        [...]
      </header>
      <metadata>
        <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
          xmlns:dc="http://purl.org/dc/elements/1.1/"
          xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
            http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
          <dc:title>Einfluß genetischer Variationen im Tumor Nekrose [...]</dc:title>
          <dc:creator>SchÄpfel, Antje</dc:creator>
          [...]
        </oai_dc:dc>
      </metadata>
    </record>
  </GetRecord>
</OAI-PMH>
```



oai_dc

- Mandatory ‘Lowest Common Denominator’
- Simple unqualified DC schema
- A Container schema is also required
 - OAI specific
- Locations:
 - Container schema hosted @ OAI Web site
 - Imports a generic DCMES schema (Metadata Element Set)
 - DCMES schema @ DCMI Web site



Other metadata formats

- oai_dc is a simple format providing baseline interoperability
- It may not be suitable:
 - Not enough (or the required) elements!
 - Not very precise - it is an “unqualified” MES
(not covered in this talk... Sorry!)
 - Not the metadata format you need ie. not:
 - IMS/IEEE LOM - eLearning metadata
 - ODRL - Open Digital Rights Language



oai_dc... not the MES I'm looking for

- Implement a different format eg. IMS/IEEE LOM
- Already agreed names, XML schema and namespaces
- Easier than creating your own schema
- Create test records and validate
- Modify repository (source code and/or configuration files) to support new format
 - e.g. listMetadataRecords response
- Test and validate new repository output



Extending a format

- Decide a name and some namespaces
- Develop XML schema for the container and the new elements
- Create test records and validate
- Modify repository (source code and/or configuration files) to support new format
- Test and validate new repository output



Summary

- OAI-PMH allows for any MES so long as...
- ...it is encoded in XML with an XML Schema
- All repositories *must* support oai_dc for...
- ...minimum level of interoperability
- If oai_dc is not enough - extend it!
- If oai_dc is not 'the one' - use something else as well!



Agenda

1. Data Provider or Service Provider
2. Metadata Records
3. Tools and services
4. Examples



Choosing tools

- Choice depends on
 - Technical skills available
 - Type of repository or service
- Evaluations and comparisons

Guide to institutional repository Software

<http://www.soros.org/openaccess/software/>

DAEDALUS: Initial experiences with EPrints and DSpace at the University of Glasgow

<http://www.ariadne.ac.uk/issue37/nixon/> (Ariadne)

DSpace vs. ETD-db: Choosing software to manage electronic theses and dissertations

<http://www.ariadne.ac.uk/issue38/jones/>



Available Tools

- Large choice see list at <http://www.openarchives.org/tools/>
- Most are open source
- Available for a variety of platforms
- Difference in emphasis
 - Metadata formats supported
 - Configurability
 - Use 'out of the box' or programming library



Tool Examples

- Dspace
<http://www.dspace.org/>
- CERN
<http://cdsware.cern.ch/>
- Eprints.org
<http://software.eprints.org/>
- ARC
<http://sourceforge.net/projects/oaiarc/>
- Net::OAI::Harvester
<http://search.cpan.org/~esummers/OAI-Harvester-0.94/lib/Net/OAI/Harvester.pm>
- Develop your own (if none of these meet your requirements)



How to advertise your service and find data providers

- Repository Explorer
<http://oai.dlib.vt.edu/cgi-bin/Explorer/oai2.0/testoai>
- OAISTER
<http://www.oaister.org/o/oaister/>
- Southampton
<http://archives.eprints.org/eprints.php>
- ARC
<http://arc.cs.odu.edu/> [appears broken]



Agenda

1. Data Provider or Service Provider
2. Metadata Records
3. Tools and services
4. Examples



Duke University

Portfolio @ Duke : About - Mozilla

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop <https://portfolio.oit.duke.edu/local/aboutpage.jsp> Search Print

Search: Go **Portfolio @ Duke**

Sunday, December 7, 2003

Browse By >>	File Name	Last Name	Date	All Resumes
Manage >>	View/Edit	Add To	Folders	Edit Profile

[Home](#)
[About](#)
[On-line Help](#)
[Duke Home](#)
[A&S Home](#)

About

Portfolio @ Duke was developed from a requirement of the Deans' Office to improve its ability to accurately assess the progression of intellectual growth at the University. One particular mode of assessment is the objective comparison of documents written by individual students over time as a part of the regular curriculum.

The Deans' Office was interested in an easy-to-use, Web-based interface which would allow undergraduates to upload writing samples to a centralized, secure repository which would be accessible to the office's assessment experts. In addition to this official assessment space, they were interested in providing unofficial personal space that undergraduates might use to publish information of their own choosing, such as Web pages of their own design, resumes and/or curricula vitae to facilitate educational and/or career aspirations, or image files to be shared with family and friends.

It was decided that programmers from Arts & Sciences Computing, specifically (CSILL), would work with representatives of OIT and the Deans' Office to come up with a viable solution that would meet the requirements.

The programmers found an open-source solution. DSpace is a newly developed (available Nov 2002) digital repository created to capture, distribute and preserve the intellectual output of MIT. As a joint project of MIT Libraries and the Hewlett-Packard Company, it provides stable long-term storage needed to house the digital products of MIT faculty and researchers.

The system was molded into the Portfolio @ Duke system that you are using today. It was beta-tested by student, faculty, and staff in the summer of 2003 and is slated to be broadly used in the fall of 2003.

<https://portfolio.oit.duke.edu/index.jsp>

CERN Workshop on Innovations in Scholarly Communications (OAI4) 20-22 October 2005



University of Oregon

A screenshot of a web browser displaying the DSpace at the University of Oregon website. The browser window title is "DSpace at the University of Oregon: Home - Mozilla". The address bar shows "https://ir.uoregon.edu:8443/dspace/index.jsp". The website header includes "UNIVERSITY OF OREGON LIBRARIES" and navigation links for "ABOUT THE LIBRARIES" and "SERVICES". The main content area is titled "DSpace at the University of Oregon >" and contains several sections: "ScholarBank news" with two entries from December 2003 and September 2003; a "Search" section with a text input field and a "Go" button; "Communities in ScholarBank" with a list of departments including Computer and Information Science, Economics, Planning, Public Policy, and Management, Psychology, and Sociology; and a "What is ScholarBank?" section explaining the repository's purpose. A footer note states "Last update: 6 June 2003 by JQJ" and provides the URL "https://ir.uoregon.edu:8443/dspace/index.jsp".

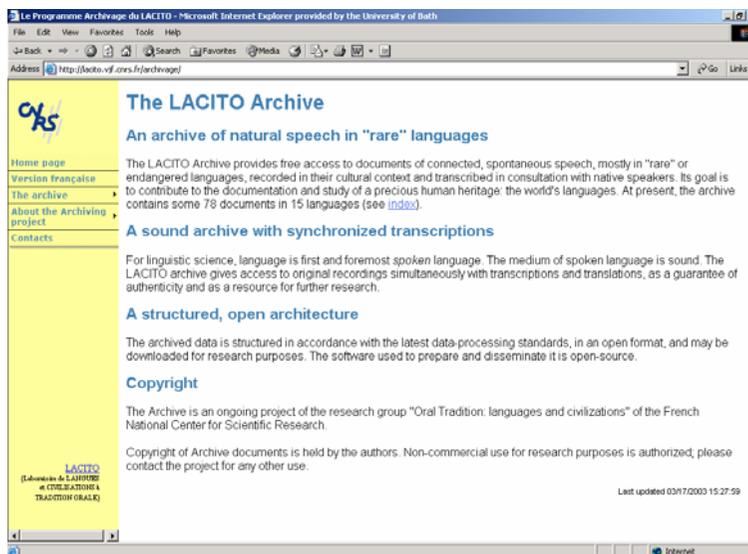
<https://ir.uoregon.edu:8443/dspace/index.jsp>

CERN Workshop on Innovations in Scholarly Communications (OAI4) 20-22 October 2005



The LACITO Archive

- The LACITO Archive
 - An archive of natural speech in “rare” languages
- Gives access to original recordings, with transcriptions and translations





The LACITO Archive

The screenshot shows the LACITO Archive website in a Microsoft Internet Explorer browser window. The browser title is "Le Programme Archivage du LACITO - Microsoft Internet Explorer provided by the University of Bath". The address bar shows "http://lacito.vjf.cnrs.fr/archivage/". The page content includes a navigation menu on the left with links like "Home page", "Version française", "The archive", "About the Archiving project", and "Contacts". The main content area features a search list with phonetic forms and glosses for six items. Item 1: "šallu ša n-nabi: imp.pl.prier loc. art.-Prophète". Item 2: "ʔalla mšalli wa sallam šala ʔa:tim an-nabiyy-i:n wa sayyid al-mursal-i:n Dieu partc.prier et acc.3sg.saluer loc. dernier art.-Prophète-suf.pl. et seigneur art.-envoyé-suf.pl.". Item 3: "ga:lu ka:n bih wa:hed ta:ʒer acc.3pl.dire exp.tps. cop.exist. un commerçant". Item 4: "ligiy ʔabi:r sa:r w šalla fi:l ʒa:miʃ w ligiy ʔabi:r-eh acc.3sg.trouver ami acc.3sg.aller et acc.3sg.prier loc-art. mosquée et acc.3sg.trouver ami-suf.". Item 5: "b-yitha:kaw w kayf ha:l-ak w min ha:na la hunak part.-inacc.3pl.parler et comment état-suf. et loc. déic.ici dir. déic.là-bas". Item 6: (partially visible). The search list is titled "Search list:" and includes a list of phonetic characters from "a" to "avn".

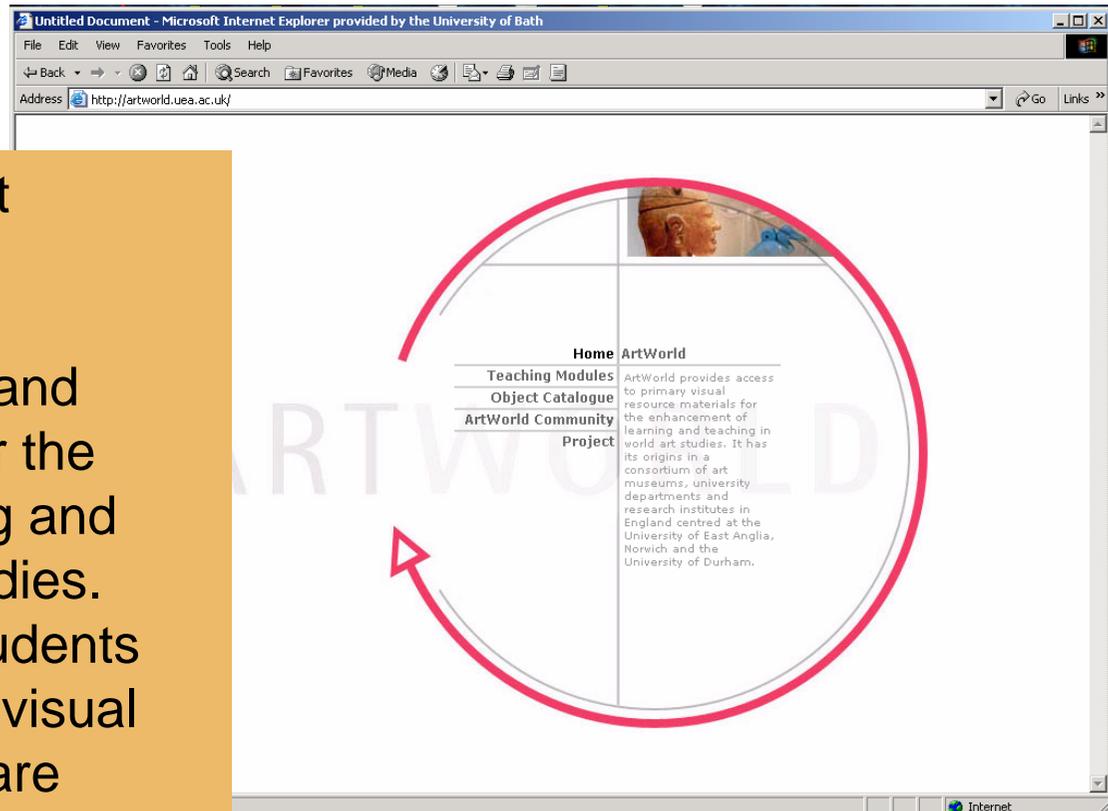
<http://lacito.vjf.cnrs.fr/archivage/index.html>

CERN Workshop on Innovations in Scholarly Communications (OAI4) 20-22 October 2005



ArtWorld

- A group of museums, art galleries and academic departments.
- Provides digital images and associated resources for the enhancement of learning and teaching in world art studies.
- Facilitates access for students and teachers to primary visual resource materials that are normally relatively inaccessible or widely scattered.



<http://artworld.uea.ac.uk/>



Summary

- During today's tutorial we hope that you have:
 - gained an overview of the history behind the OAI-PMH and an overview of its key features
 - acquired an understanding of how the protocol works
 - learned something about some of the main implementation issues
 - gained familiarity with the OAForum tutorial and learned where to look for more information
 - become comfortable with the terminology used
 - started thinking about how you will be using OAI in your institution



Resources

- Open Archives Initiative (OAI official Web site)
<http://www.openarchives.org/>
- Open Archives Forum (OA-Forum Web site)
<http://www.oaforum.org/>
- OAI-PMH protocol specification
<http://www.openarchives.org/OAI/openarchivesprotocol.html>
- Implementation guidelines:
<http://www.openarchives.org/OAI/2.0/guidelines.htm>
- OAI general mailing list
<http://www.openarchives.org/mailman/listinfo/OAI-general/>
- OA-Forum expert reports and reviews of organisational and technical issues
[Links from http://www.oaforum.org/documents/](http://www.oaforum.org/documents/)



Resources

- Repository explorer
<http://oai.dlib.vt.edu/cgi-bin/Explorer/oai2.0/testoai>
- Tools
<http://www.openarchives.org/tools/>
- Implementers mailing list
<http://www.openarchives.org/mailman/listinfo/OAI-implementers/>
- Dublin Core
<http://dublincore.org/>
- The Eprints User's Handbook
<http://software.eprints.org/handbook>



ePrint Archives

- ArXiv
<http://arXiv.org/>
- RePec
<http://www.repec.org/>
- Cogprints
<http://cogprints.ecs.soton.ac.uk/>
- NCSTRL:
<http://www.ncstrl.org>



Examples of Service Providers

- Citation Indexing
<http://icite.sissa.it>
- Printing on Demand Service
<http://www.proprint-service.de>
- Value added Search Engine
<http://www.myoai.com>
- DINI
<http://edoc.hu-berlin.de/oaisearch/>
- Physnet
<http://physnet.uni-oldenburg.de/oai/query.php>
- ARC
<http://arc.cs.odu.edu/>



Task Page

Task 1 Seven Key Definitions

<http://www.oaforum.org/tutorial/english/page1.htm#section3>

Task 3 Sources of Further Information

<http://www.oaforum.org/tutorial/english/page2.htm#section9>

Task 3 Quiz

<http://www.oaforum.org/tutorial/english/page1.htm#section5>

Task 4 Using Repository Explorer

<http://oai.dlib.vt.edu/cgi-bin/Explorer/oai2.0/testoai>

[Now at: <http://re.cs.uct.ac.za/>]

Task 5 Exploring some service interfaces: choose from

<https://portfolio.oit.duke.edu/index.jsp>

<https://ir.uoregon.edu:8443/dspace/index.jsp>

<http://artworld.uea.ac.uk/>

Or any of the service providers or archives listed under Resources



End of Tutorial 3

OAI and OAI-PMH for absolute beginners a non-technical introduction

Philip Hunter

IRIScotland Project, University of Edinburgh, United Kingdom

philip.hunter@ed.ac.uk

Tutorial development also generously supported by UKOLN
at the University of Bath