5-2007

The Open Archives Initiative

Michael L. Nelson
Old Dominion University, mnelson@odu.edu

The Open Archives Initiative

Michael L. Nelson

Computer Science, Old Dominion University
www.cs.odu.edu/~mln/

www.openarchives.org

Open Archives Initiative
Protocol for Metadata Harvesting

- *data providers / repositories:*
  - "A repository is a network accessible server that can process the 6 OAI-PMH requests in the manner described in [the OAI-PMH document]. A repository is managed by a data provider to expose metadata to harvesters."

- *service providers / harvesters:*
  - "A harvester is a client application that issues OAI-PMH requests. A harvester is operated by a service provider as a means of collecting metadata from repositories."

Data Providers / Service Providers
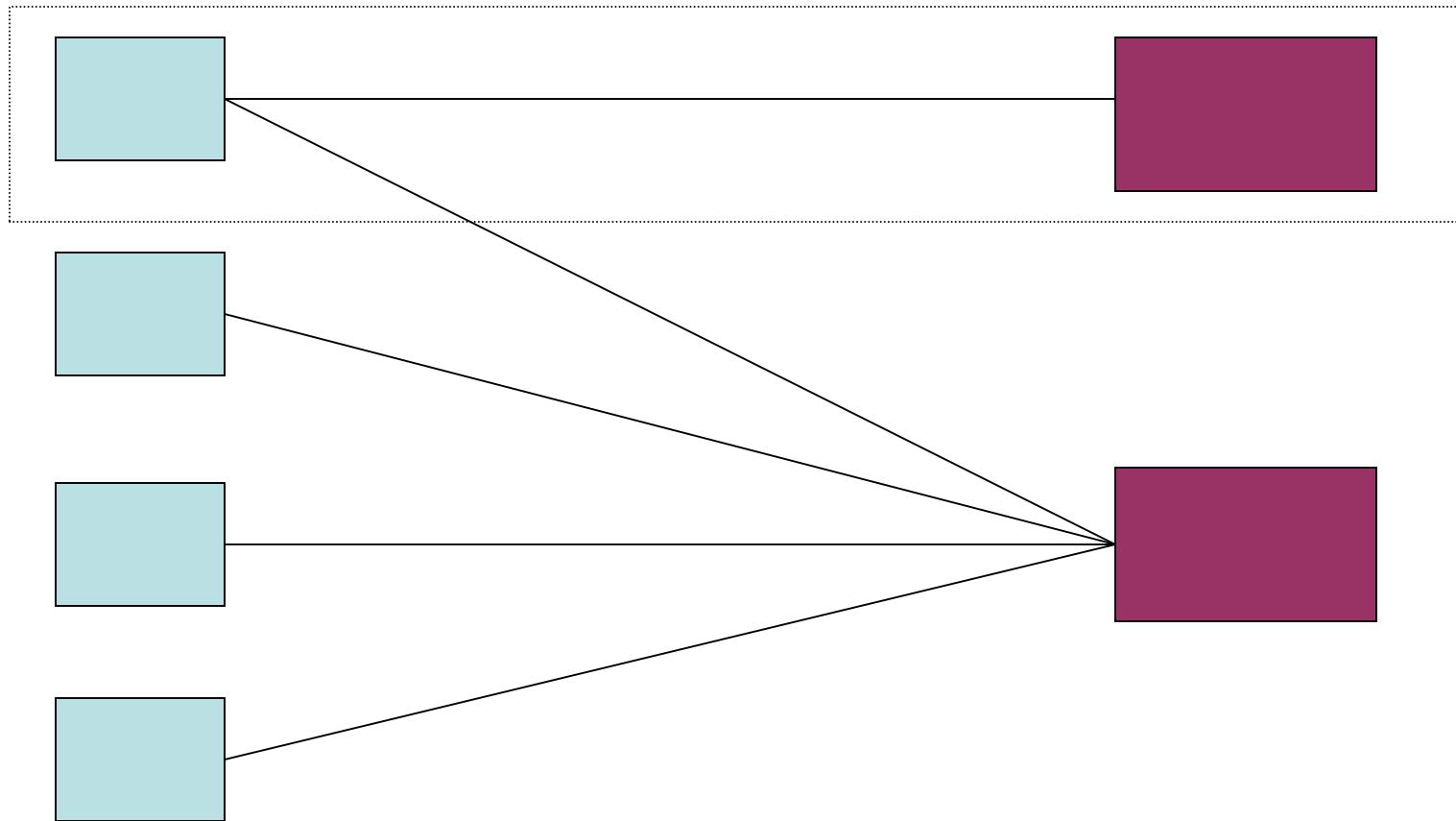
data providers
(repositories)

service providers
(harvesters)

Overview of OAI-PMH Verbs

Verb	Function
Identify	description of repository
ListMetadataFormats	metadata formats supported by repo
ListSets	sets defined by repository
ListIdentifiers	OAI unique ids contained in repo
ListRecords	listing of N records
GetRecord	listing of a single record

repository metadata

harvesting verbs

most verbs take arguments: dates, sets, ids, metadata formats
and resumption token (for flow control)

OAI-PMH data model

resource

OAI-PMH sets

OAI-PMH identifier

entry point to all records pertaining to the resource ← item

OAI-PMH identifier
metadataPrefix
datestamp

Dublin Core
metadata

MARCXML
metadata

← records

metadata pertaining
to the resource
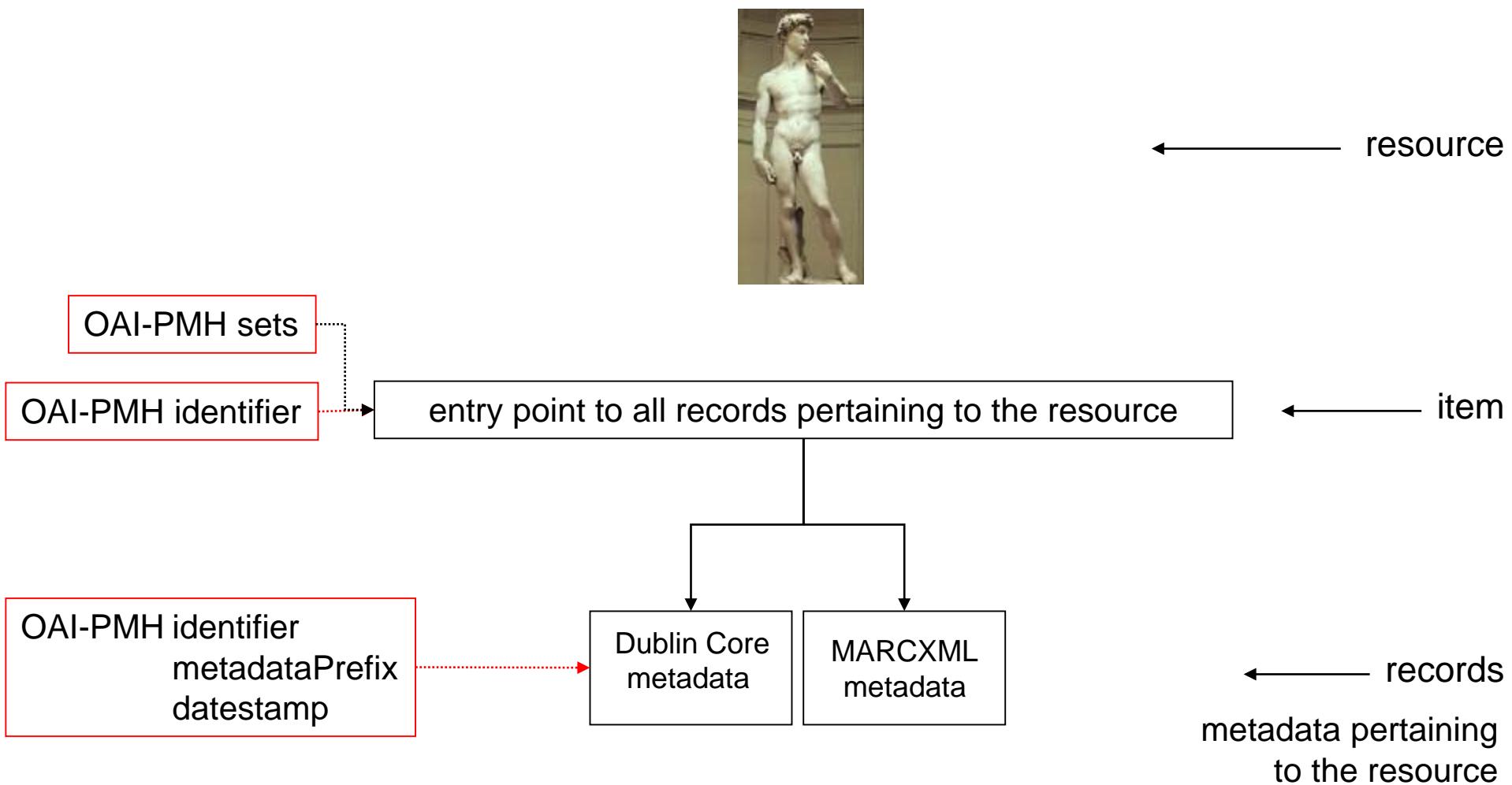
Complexity Comes to OAI-PMH…

- First noticed in how people would populate their Dublin Core records
    - people need the HTML splash page
    - crawlers need the PDF file
- Ad-hoc conventions and methods used to expose the repository's knowledge about the structure of the object
- Next three slides taken from "Resource Harvesting Within the OAI-PMH Framework"
    - http://www.dlib.org/dlib/december04/vandesompel/12vandesompel.html

Dublin Core Encoding Type 1

```
<oai_dc:dc>
    <dc:title>A Simple Parallel-Plate Resonator Technique for Microwave.
        Characterization of Thin Resistive Films</dc:title>
    <dc:creator>Vorobiev, A.</dc:creator>
    <dc:subject>ING-INF/01 Elettronica</dc:subject>
    <dc:description>A parallel-plate resonator method is proposed for
        non-destructive characterisation of resistive films used in
        microwave integrated circuits. A slot made in one ... </dc:description>
    <dc:publisher>Microwave engineering Europe</dc:publisher>
    <dc:date>2002</dc:date>
    <dc:type>Documento relativo ad una Conferenza o altro Evento</dc:type>
    <dc:type>PeerReviewed</dc:type>
    <dc:identifier>http://amsacta.cib.unibo.it/archive/00000014/</dc:identifier>
    <dc:format>pdf
      http://amsacta.cib.unibo.it/archive/00000014/01/GaAs_1_Vorobiev.pdf
    </dc:format>
</oai_dc:dc>
```

splash page

locator of resource

Dublin Core Encoding Type 2

```
…
   <dc:identifier>http://amsacta.cib.unibo.it/archive/00000014/</dc:identifier>
   <dc:relation>
     http://amsacta.cib.unibo.it/archive/00000014/01/GaAs_1_Vorobiev.pdf
   </dc:relation>
…
```

splash page

locator of resource

Dublin Core Encoding Type 3

```
…
  <dc:identifier> http://amsacta.cib.unibo.it/archive/00000014/</dc:identifier>
   <dc:relation>
      http://resolver.unibo.it/00000014/
      </dc:relation>
   <dc:relation>
      http://amsacta.cib.unibo.it/archive/00000014/01/GaAs_1_Vorobiev.pdf
      </dc:relation>
…
```

splash page

splash page

locator of resource

OAI Object Re-Use and Exchange

- Develop, identify, and profile extensible standards and protocols to allow *repositories, agents, and services to interoperate* in the context of *use and reuse of compound digital objects* beyond the boundaries of the holding repositories.

- Aim for more effective and consistent ways:
  - to facilitate discovery of these objects,
  - to reference (link to) these objects (and parts thereof),
  - to obtain a variety of disseminations of these objects,
  - to aggregate and disaggregate these objects,
 - Enable processing by automated agents

The Structure of Compound Objects is Obfuscated When Mapped to the Web

Repository$_1$

PDF

HTML

JPEG

Data

The common origination of URI1 and URI2 from the same compound object is not visible from the web graph.

URI$_1$

URI$_2$
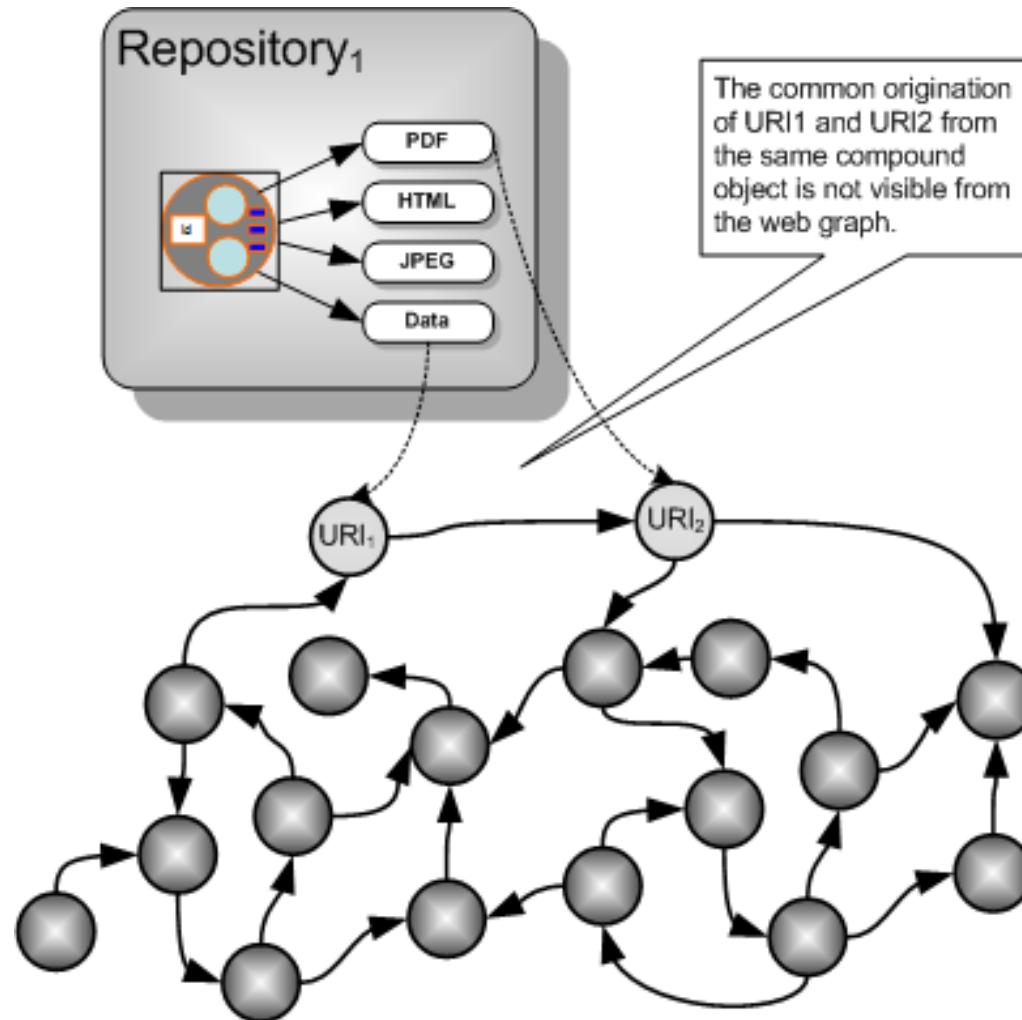
Useful for humans and useful for applications is often different
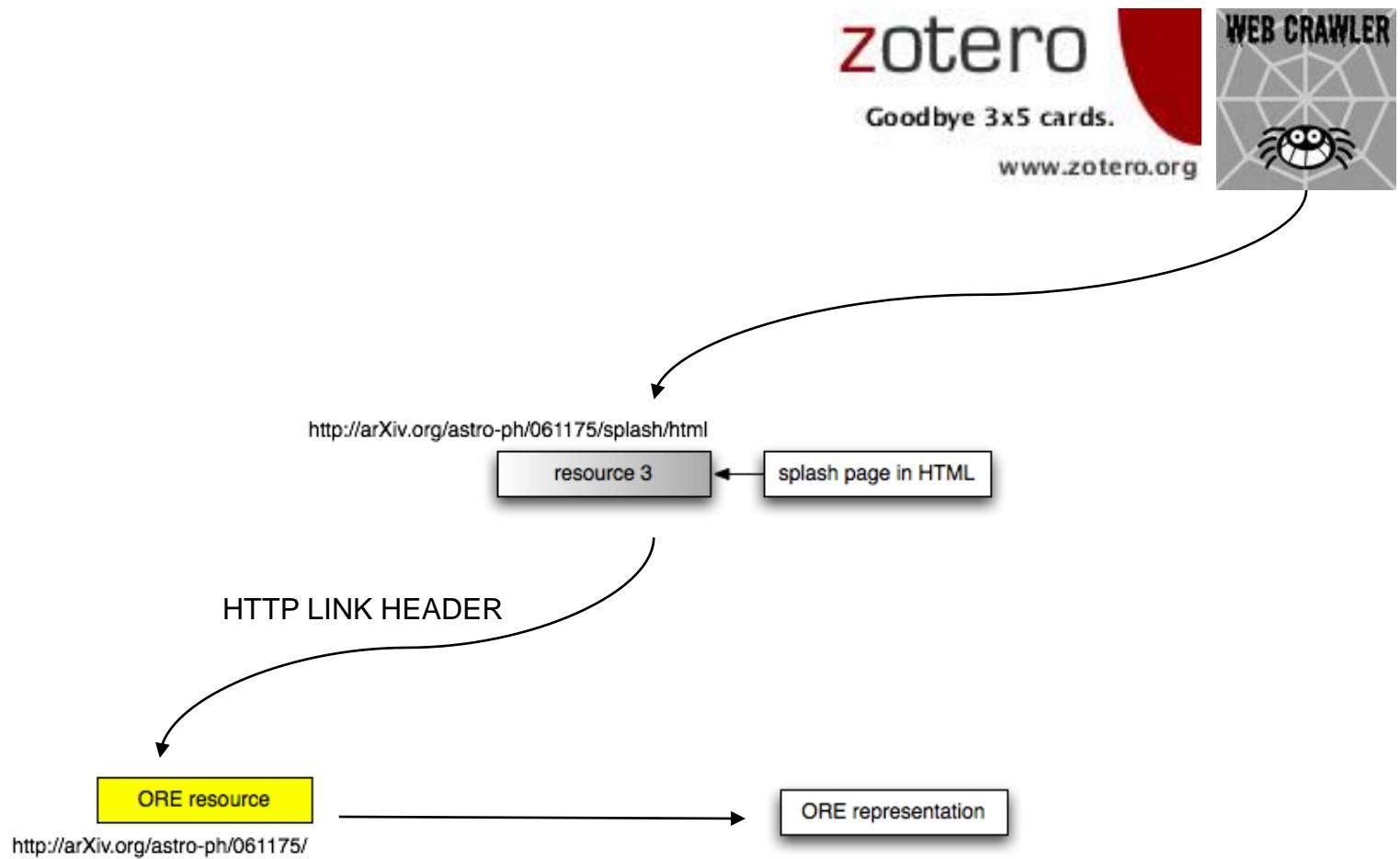
Through the Resource Map, the Web application sees the compound object
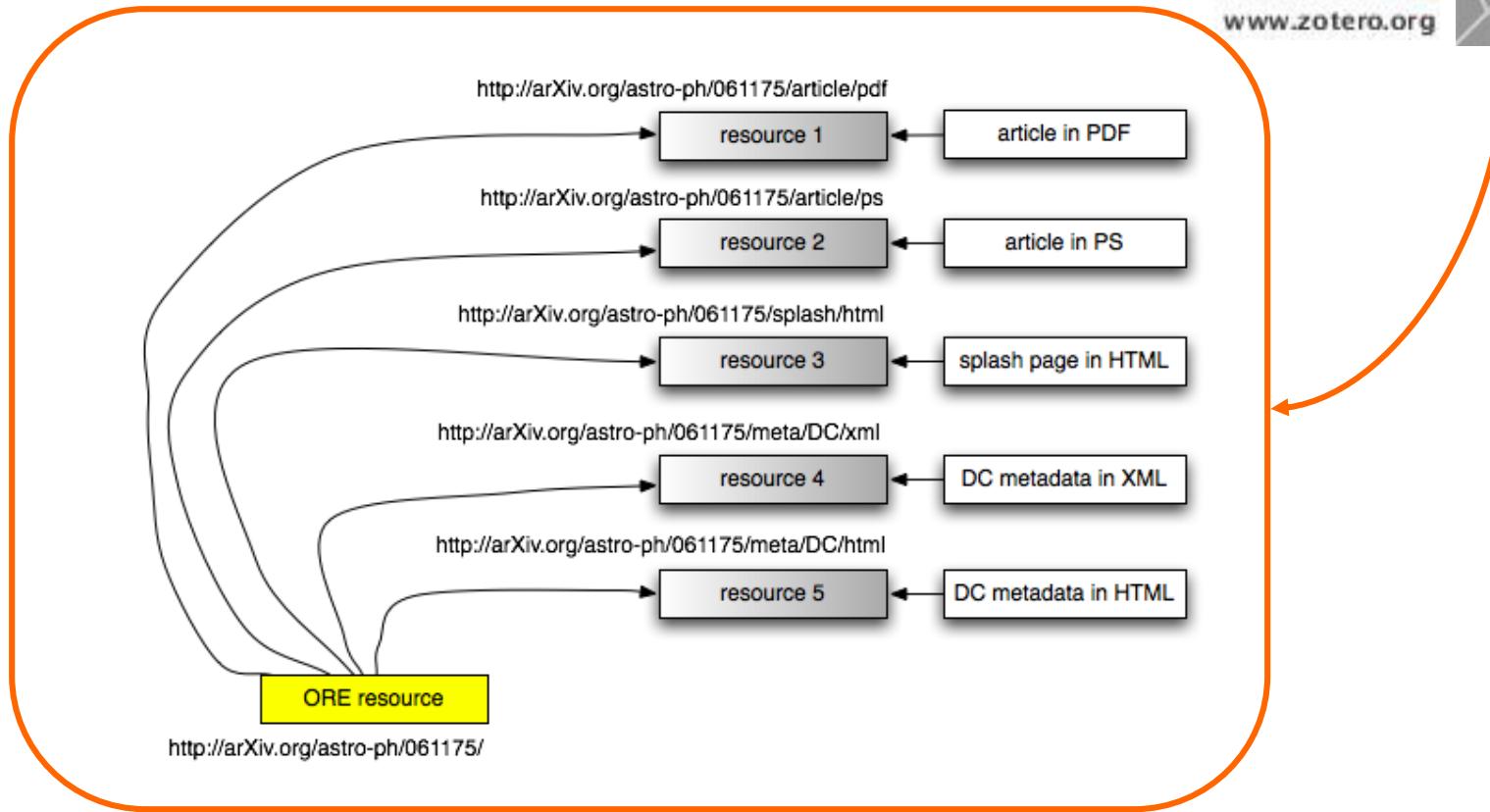
This approach reveals compound objects in the Web graph
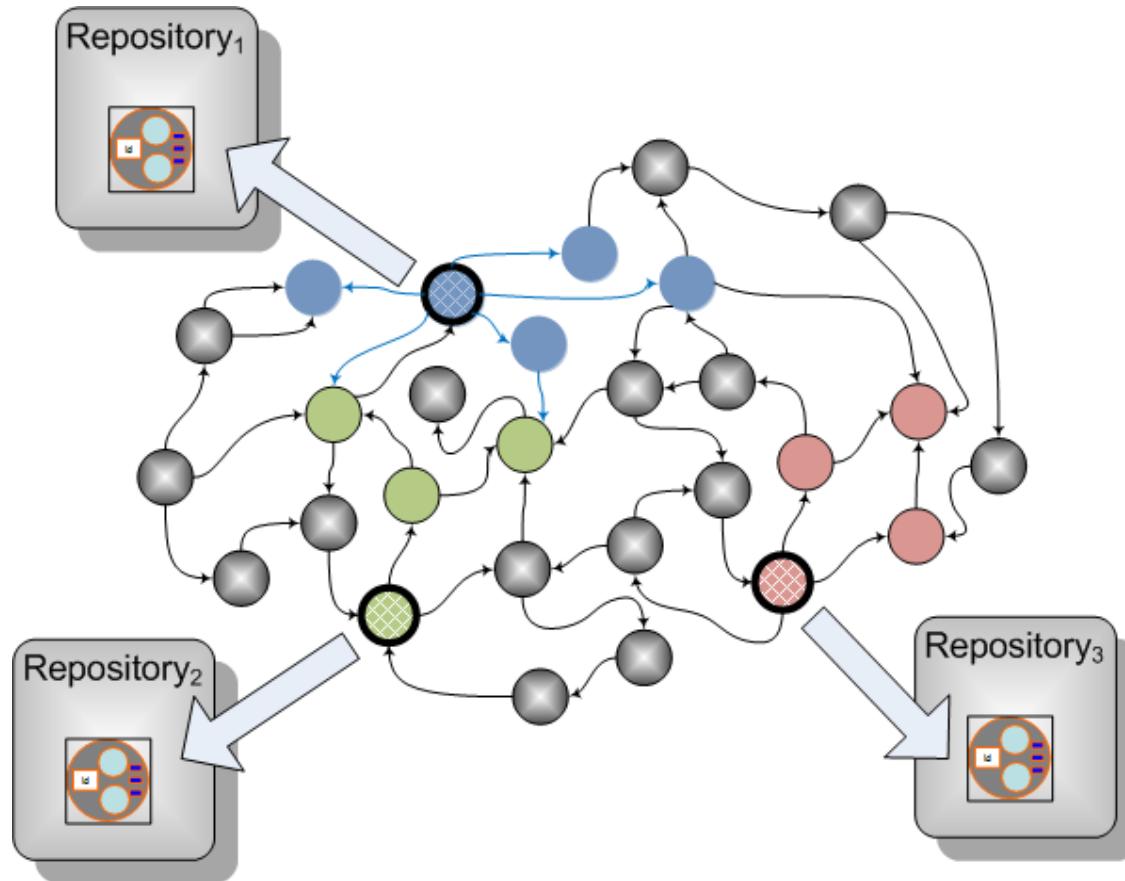
OAI: Its Not Just for
Metadata Harvesting Anymore...

OAI-PMH	OAI-ORE
Repository structure	Object structure
Metadata centric	Resource centric
Metadata harvesting	Object re-use (obtain, harvest, register)

OAI-PMH and OAI-ORE are complimentary;
- you can do one without the other
- you can do them together

OAI-ORE : Current Status

- Ongoing definition of the ORE framework
    - Reach joint problem statement
    - Issues regarding identification
    - Model for ORE resource
    - Publishing ORE resources to the Web
    - Discovering ORE resources

- Review of appropriate technologies for ORE Model and Resource Map
    - ATOM
    - DID/DIDL, IMS/CP, METS, Ramlet
    - RDF, RDF/XML
    - Dublin Core Abstract Model
    - ...

OAI-ORE : Current Status

- Explore demonstrators using these concepts in preparation of May 2007 ORE Technical Committee meeting

- Post May 2007 meeting:
  - Hopefully work towards alpha specs for ORE resource, Resource Map, discovery of ORE resource
  - Experimentation with alpha specs