

# Open Archives Initiative Object Re-Use & Exchange

Herbert Van de Sompel <sup>(1)</sup> & Carl Lagoze <sup>(2)</sup>

Acknowledgments  
Michael L. Nelson <sup>(3)</sup>

<sup>(1)</sup> Digital Library Research & Prototyping Team, Research Library, Los Alamos National Laboratory  
herbertv@lanl.gov

<http://public.lanl.gov/herbertv>

<sup>(2)</sup> Information Science, Cornell University

<sup>(3)</sup> Computer Science, Old Dominion University

ORE is supported by the Andrew W. Mellon Foundation  
with additional support of the National Science Foundation



OAI Object Re-Use and Exchange  
OAI5, CERN, Switzerland, April 18th 2007  
Herbert Van de Sompel



# General information about OAI-ORE



OAI Object Re-Use and Exchange  
OAI5, CERN, Switzerland, April 18th 2007  
Herbert Van de Sompel



# OAI Object Re-Use and Exchange

- OAI-ORE is a new interoperability effort conducted under the umbrella of the OAI
- Supported by the **Andrew W. Mellon Foundation**; additional support from the **National Science Foundation**
- International effort; October 2006 - September 2008:
  - Coordinators: Carl Lagoze & Herbert Van de Sompel
  - ORE Technical Committee: 13 international members
  - ORE Liaison Group: 8 international members
  - ORE Advisory Committee: 16 international members
  - Representing: scholarly publishers and aggregators, eScience, eHumanities, education, search engines, various repository systems, digital library efforts, related standardization efforts, etc.
- See <http://www.openarchives.org/ore/>



# OAI is not just about metadata anymore

OAI-PMH	OAI-ORE
Repository structure	Object structure
Repository centric	Web centric
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



# Context of OAI-ORE Standards & Protocols



OAI Object Re-Use and Exchange  
OAI5, CERN, Switzerland, April 18th 2007  
Herbert Van de Sompel



# An Early Formulation of the Problem

- 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/vandesompe/12vandesompe.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

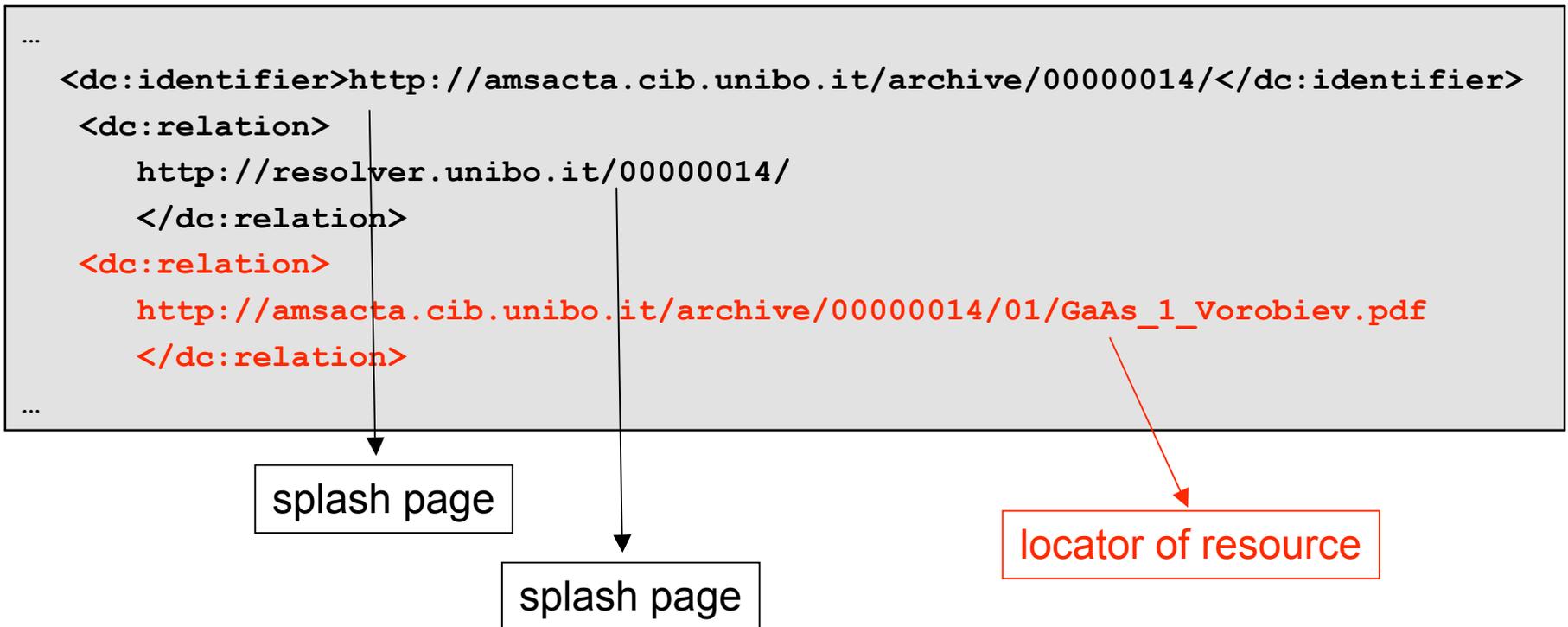
```
...  
<dc:identifier>http://amsacta.cib.unibo.it/archive/00000014/</dc:identifier>  
<dc:relation>  
  http://amsacta.cib.unibo.it/archive/00000014/01/GaAs_1_Vorobiev.pdf  
</dc:relation>  
...
```

splash page

locator of resource



# Dublin Core Encoding Type 3



## And more recently ...

"Are repositories successfully exposing the full-text of articles (the PDF file or whatever) to Google rather than (or as well as) the abstract page?"

"Are we consistent in the way we create hypertext links between research papers in repositories?"

(from Andy Powell's eFoundations blog)



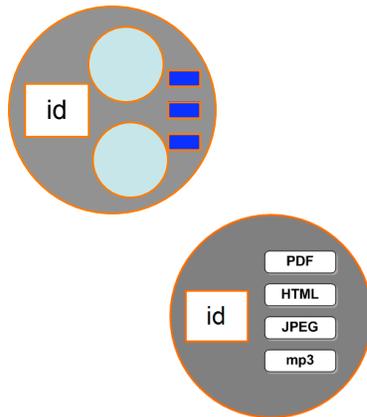
As the objects get more complex,  
things get worse

Rather than continue down that path,  
let's back up and restart...



# Compound Information Objects

Units of scholarly communication are compound information objects:



compound  
information  
objects

Identified, bounded aggregations of related information units that form a logical whole.

Components of compound object may vary according to:

- Semantic type: book, article, moving image, dataset, ...
- Media type: PDF, HTML, JPEG, MP3, .
- Internal relationship: parts, views, ...
- External relationships



# Scholarly Examples

The Open Archives Initiative: Building a low-barrier interoperability framework (2001) (Make Corrections) (22 citations)

Carl Lagoze, Herbert Van de Sompel  
ACM/IEEE Joint Conference on Digital Libraries

**CiteSeer** Home/Search/Context/Related  
Links: [DBLP](#)

(Enter summary)

View or download:  
[openarchives.org/documents/oai.pdf](#)  
Cached: [PS](#) [gz](#) [PS](#) [PDF](#)  
[Image](#) [Update](#) [Help](#)

From: [arXiv.org/abs/astro-ph/0611775](#) [quedon](#) (more) (Enter author homepages)

Rate this article: 1 2 3 4 5 (best) [Comment on this article](#)

**Abstract:** The Open Archives Initiative (OAI) develops and promotes interoperability solutions that aim to facilitate the efficient dissemination of content. The roots of the OAI lie in the E-Print community. Over the last year its focus has been extended to include all content providers. This paper describes the recent history of the OAI – its origins in promoting E-Prints, the broadening of its focus, the details of its technical standard for metadata harvesting, the applications of this... ([Update](#))

**Cited by:** [More](#)  
[Metadata Interoperability and Distributed - Information Search On](#) ([Correct](#))  
[Interoperability Adaptors For Distributed - Information Search On](#) ([Correct](#))  
[Notes from the Interoperability Front: - Progress Report On \(2002\)](#) ([Correct](#))

**Active bibliography (related documents):** [More](#) [All](#)  
[0.7: Integrating Hypermedia Functionality into Database.](#) - Bhaumik, Vaitis. (2001) ([Correct](#))  
[0.7: Smart Objects, Dumb Archives: Insuring the Long-Term Integrity of .](#) - Nelson (2000) ([Correct](#))  
[0.5: Journal of Government Information \(2001\), 28\(4\), pp. 389-394.](#) - Nasa Langley Research ([Correct](#))

**Similar documents based on text:** [More](#) [All](#)  
[1.1: Heterogeneity in Open Archives Metadata - Fischer, Fuhr \(2001\)](#) ([Correct](#))  
[0.9: The Open Archives Initiative: Realizing Simple and Effective.](#) - Suleman, Fox (2001) ([Correct](#))  
[0.9: Developing Services for Open Eprint Archives.](#) - Hitchcock, Carr. (2000) ([Correct](#))

**Related documents from co-citation:** [More](#) [All](#)  
[5: The Santa Fe Convention of the Open Archives Initiative \(context\) - Sompel, Lagoze - 2000](#)  
[4: Arc - An OAI Service Provider for Digital Library Federation \(context\) - Liu, Maly et al. - 2001](#)  
[4: A Spectrum of Interoperability: The Site for Science Prototype for the NSDL \(context\) - Arms, Hillmann et al. - 2002](#)

**BibTeX entry:** ([Update](#))

Carl Lagoze and Herbert Van de Sompel. 2001. The Open Archives Initiative: Building a low-barrier interoperability framework. <http://www.cs.cornell.edu/lagoze/papers/oai-jcdl.pdf>. <http://citeseer.ist.psu.edu/lagoze01open.html> [More](#)

```
@inproceedings{lagoze01open,
  author = "Carl Lagoze and Herbert Van de Sompel",
  title = "The open archives initiative: building a low-barrier interoperability framework",
  booktitle = "(ACM)/IEEE Joint Conference on Digital Libraries",
  pages = "54-62",
  year = "2001",
```

<http://citeseer.ist.psu.edu/lagoze01open.html>

[astro-ph/0611775] Accelerating cosmologies tested by distance measures

[arXiv.org](#) > [astro-ph](#) > [arXiv:astro-ph/0611775](#)

Search for  (Help | Advanced search)  
All papers

**Astrophysics**

## Accelerating cosmologies tested by distance measures

V. Barger, Y. Gao, D. Marfatia

(Submitted on 25 Nov 2006 (v1), last revised 23 Jan 2007 (this version, v3))

We test if the latest Gold set of 182 SNIa or the combined "Platinum" set of 192 SNIa from the ESSENCE and Gold sets, in conjunction with the CMB shift parameter show a preference between the LambdaCDM model, three wCDM models, and the DGP model of modified gravity as an explanation for the current accelerating phase of the universe's expansion. We consider flat wCDM models with an equation of state  $w(a)$  that is (i) constant with scale factor  $Sa$ , (ii) varies as  $w(a)=w_0+w_a(1-a)$  for redshifts probed by supernovae but is fixed at  $-1$  at earlier epochs and (iii) varies as  $w_0+w_a(1-a)$  since recombination. We find that all five models explain the data with comparable success.

**Comments:** 15 pages, 7 figures, 1 table. New ESSENCE SN data included  
**Subjects:** Astrophysics (astro-ph); General Relativity and Quantum Cosmology (gr-qc); High Energy Physics - Phenomenology (hep-ph); High Energy Physics - Theory (hep-th)

**Journal reference:** Phys.Lett. B648 (2007) 127-132  
**DOI:** [10.1016/j.physletb.2007.03.021](https://doi.org/10.1016/j.physletb.2007.03.021)  
**Cite as:** [arXiv:astro-ph/0611775v3](https://arxiv.org/abs/astro-ph/0611775v3)

**Submission history**  
From: Danny Marfatia [[view email](#)]  
[\[v1\]](#) Sat, 25 Nov 2006 20:26:32 GMT (313kb)  
[\[v2\]](#) Wed, 6 Dec 2006 00:24:00 GMT (450kb)  
[\[v3\]](#) Tue, 23 Jan 2007 21:45:01 GMT (923kb)

*Which authors of this paper are endorsers?*

Link back to: [arXiv](#), [form interface](#).

**Download:**

- PostScript
- PDF
- Other formats

**References & Citations**

- [SLAC-SPHRES HEP](#) (refers to, cited by, arXiv reformatted)
- [NASA ADS](#)
- [CiteBase](#)

1 [trackback \(?\)](#)

[previous](#) | [next](#)

<http://arxiv.org/abs/astro-ph/0611775>



OAI Object Re-Use and Exchange  
OAI5, CERN, Switzerland, April 18th 2007  
Herbert Van de Sompel

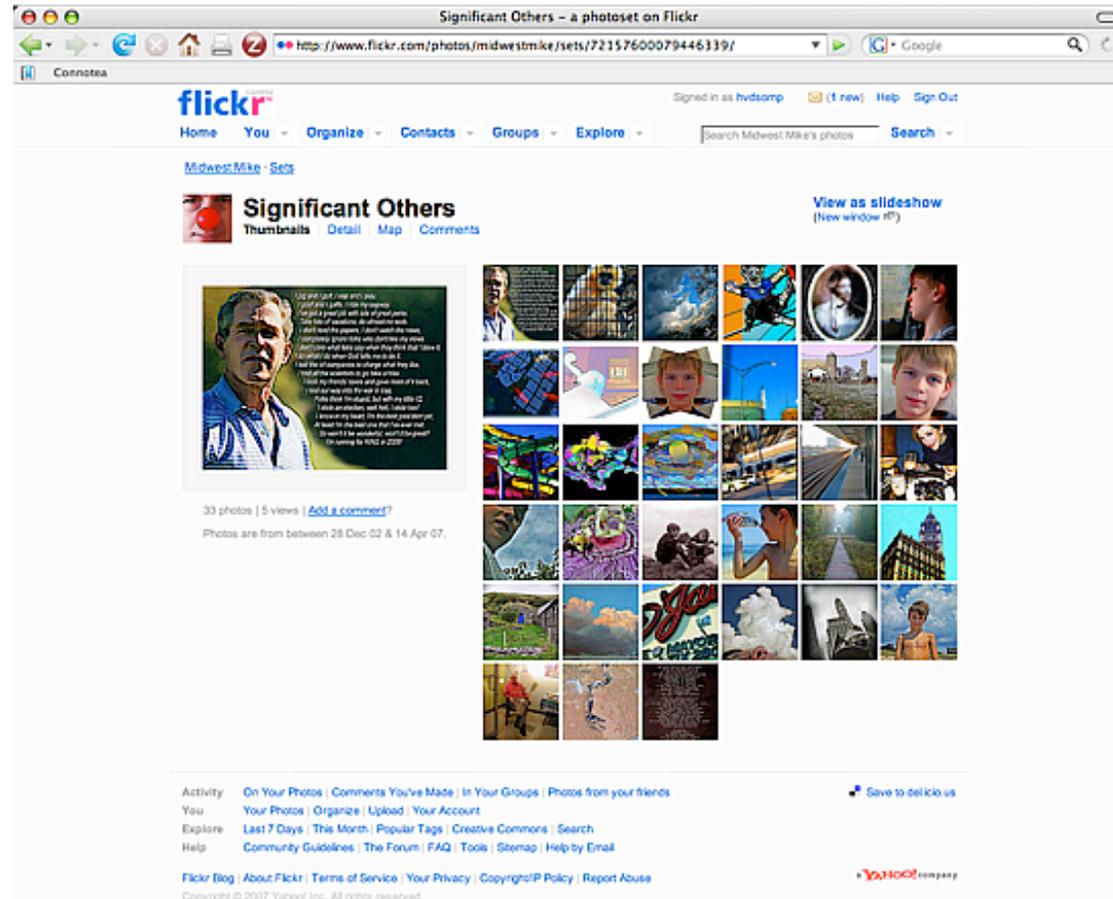


## And more scholarly examples ...

- Scholarly publication with an article and supporting information including dataset, video, etc.
- Digitized book with multiple chapters, each chapter containing multiple scanned pages.
- Archaeological assemblies of images, maps, charts, and find lists.
- An ARTstor image object that is the aggregation of various renderings of the same source image.
- ...



But these things are not only scholarly ...



<http://www.flickr.com/photos/midwestmike/sets/72157600079446339/>

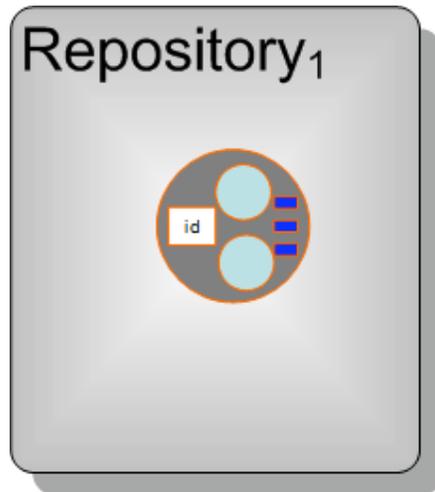


OAI Object Re-Use and Exchange  
OAI5, CERN, Switzerland, April 18th 2007  
Herbert Van de Sompel



# Access Repositories

Compound objects are made accessible by a variety of scholarly repositories:



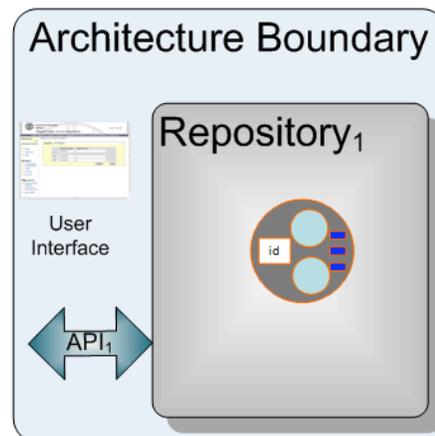
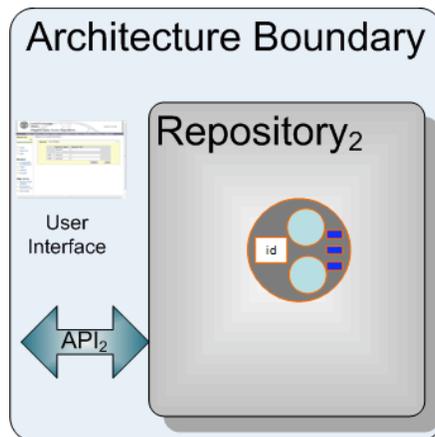
- Institutional repositories
- Discipline-oriented repositories
- Publisher repositories
- Dataset repositories
- Cultural heritage repositories
- Learning object repositories
- Digitized book and manuscript collections
- Research-group and managed personal (ePortfolio) repositories
- ...

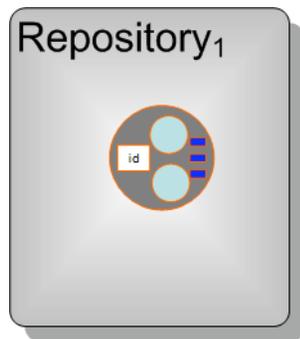


# Access Repositories

Repositories expose compound objects in manners specific to the repository architecture:

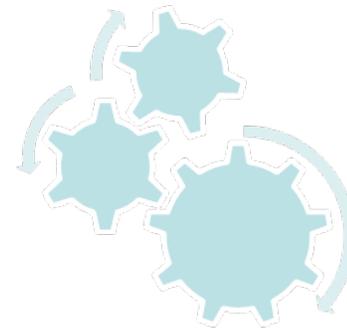
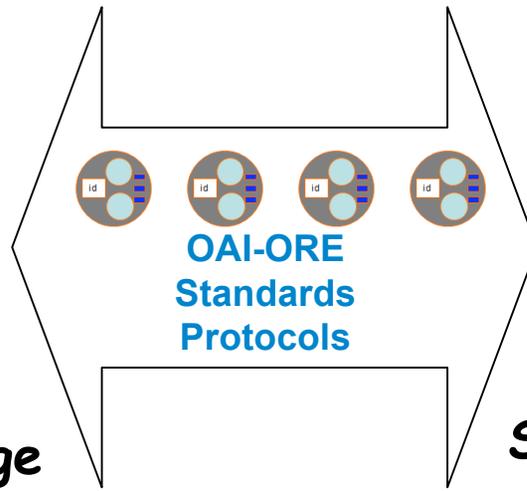
- Interfaces (API & user-oriented)
- Identification schemes
- Representation of compound objects
- Publication of compound objects and components to the Web





## Systems that manage digital objects

- Institutional repositories
- Discipline-oriented repositories
- Publisher repositories
- Dataset repositories
- Cultural heritage repositories
- Learning object repositories
- Digitized book and manuscript collections
- Image repositories
- ...



## Systems that leverage managed digital objects

- All repositories from left column
- Search engines
- Authoring tools
- Citation management tools
- Collaborative environments
- Social network applications
- Graph analysis tools
- Preservation services
- Workflow tools
- ...



# 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*



# Taking the Web perspective



OAI Object Re-Use and Exchange  
OAI5, CERN, Switzerland, April 18th 2007  
Herbert Van de Sompel

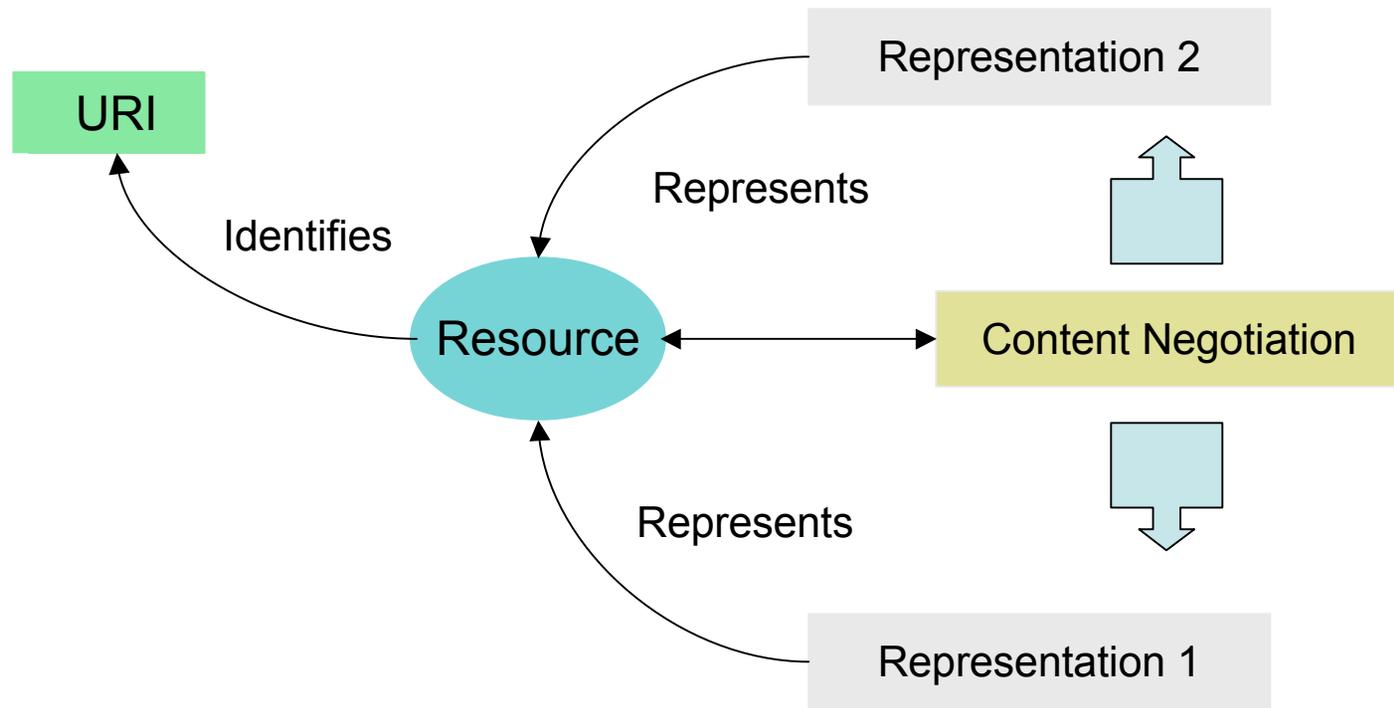


# Working with the web architecture

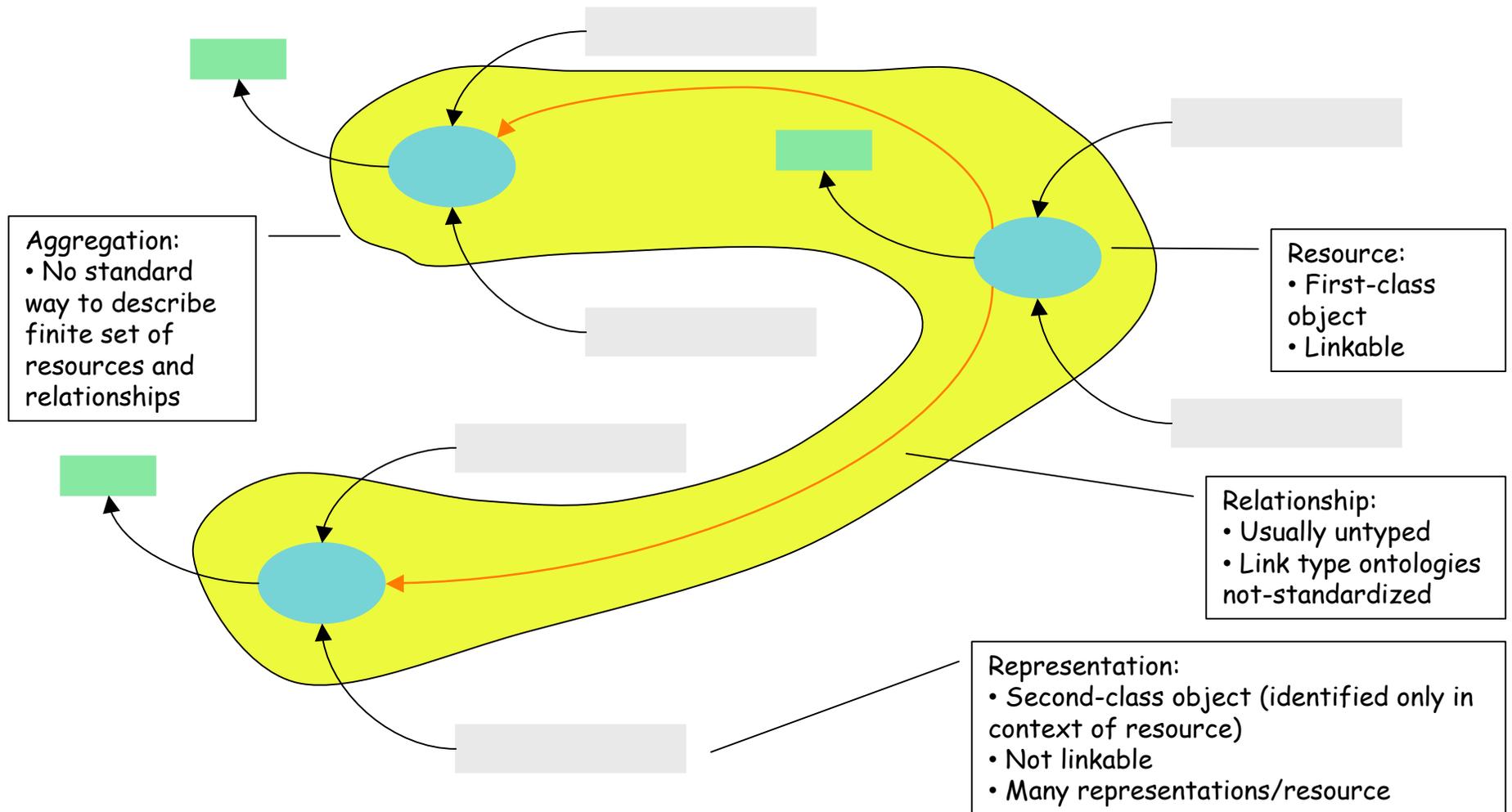
- Whatever we do must be **congruent with the web architecture**
  - Use existing capabilities where they are appropriate
  - Cleanly layer capabilities meeting the needs of our problem space
- Provide the infrastructure for web-based information systems that exploit/enhance and therefore overlay on the existing web.



# W3C Web Architecture

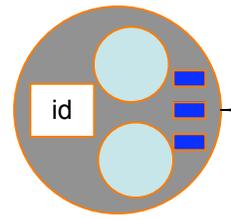


# W3C Web Architecture: more details



# Compound Object

astro-ph/0611775

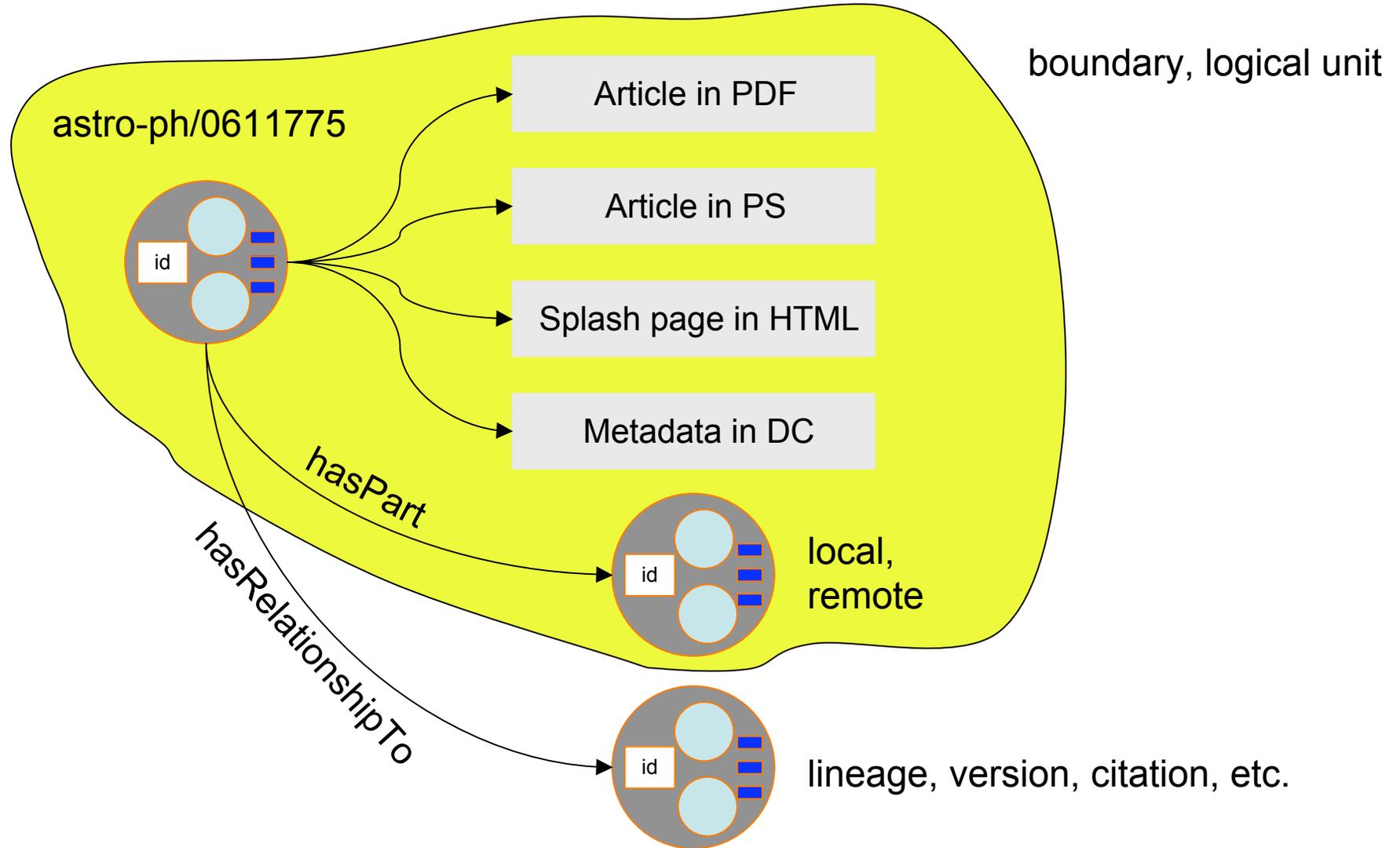


- Article in PDF
- Article in PS
- Splash page in HTML
- Metadata in DC

Multiple Views, diverging in media-type, format, and content-type

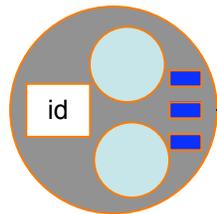


# More complexity ...



# Compound Object

astro-ph/0611775



Article in PDF

Article in PS

Splash page in HTML

Metadata in DC

Let's publish it to the Web



<http://arxiv.org/astro-ph/0611775/article/>

Resource 1

Article in PDF

Article in PS

<http://arxiv.org/astro-ph/0611775/splash/>

Resource 2

Splash page in HTML

<http://arxiv.org/astro-ph/0611775/meta/DC/>

Resource 3

DC meta XML

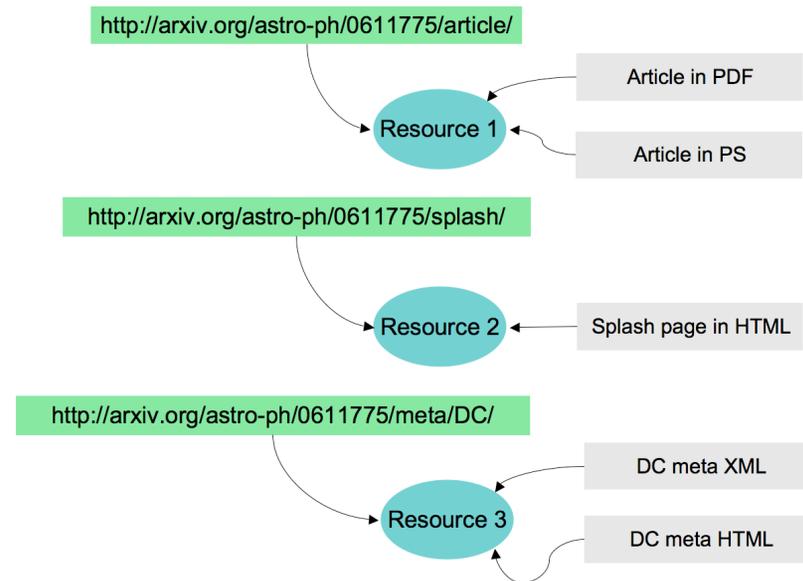
DC meta HTML



# Compound Object published to the Web

“Are repositories successfully exposing the full-text of articles (the PDF file or whatever) to Google rather than (or as well as) the abstract page?”

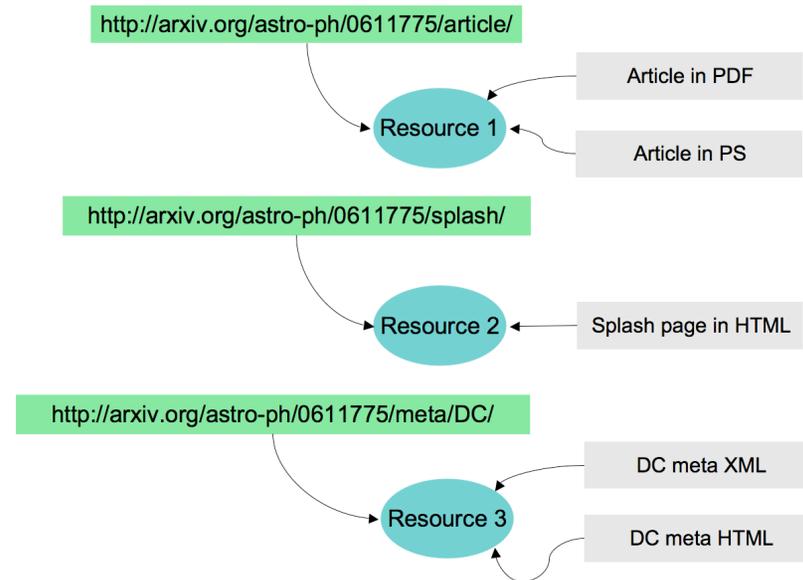
- **Discovery:** How does Google find all these resources that originate from the same digital object?
- **Boundary:** How does Google know these resources originate in the same digital object?



# Compound Object published to the Web

“Are we consistent in the way we create hypertext links between research papers in repositories?”

- Citation: Which Resource to link to?
- Citation: How to reference the PDF version (and not the PS version)?



# Thoughts about a possible approach

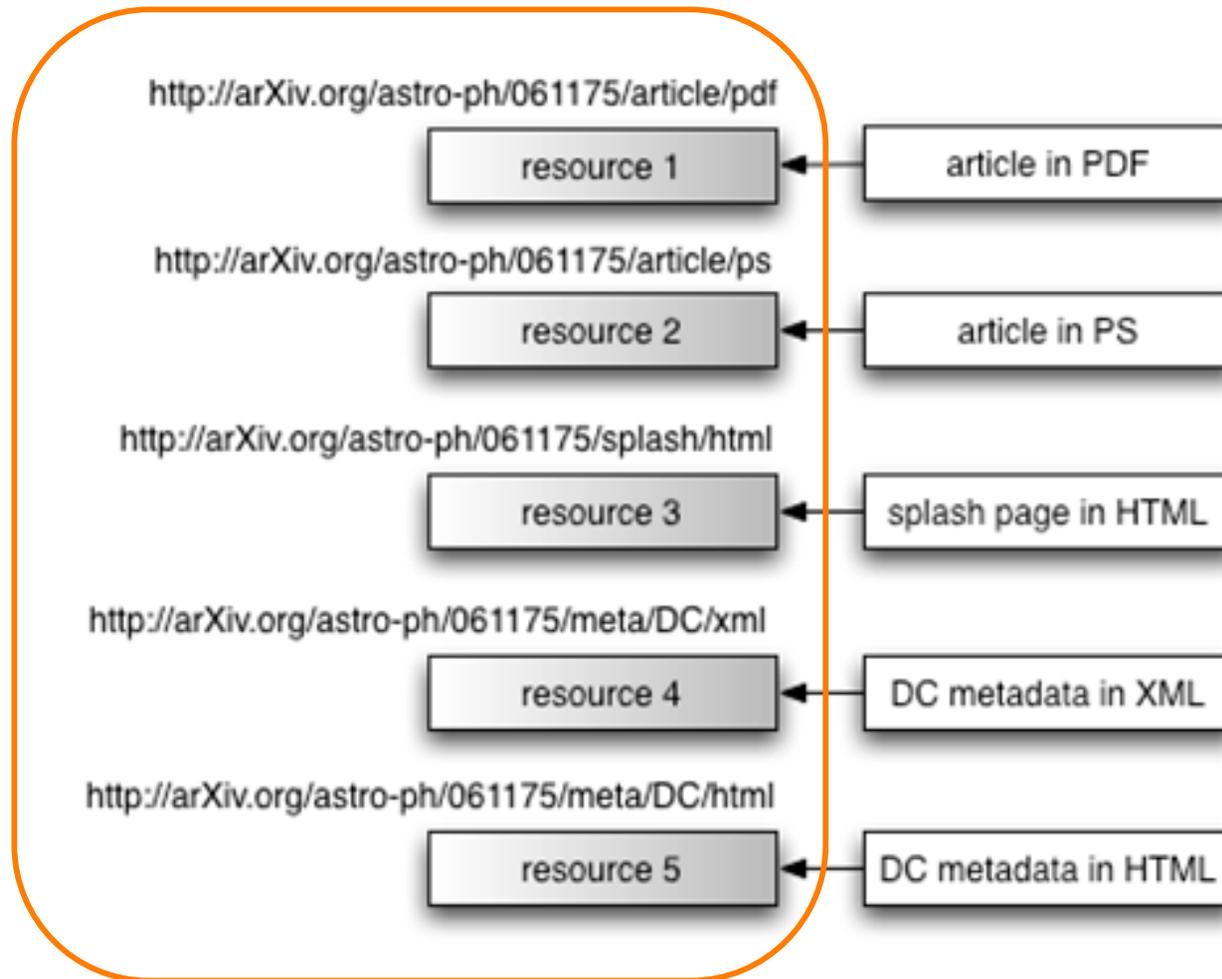


OAI Object Re-Use and Exchange  
OAI5, CERN, Switzerland, April 18th 2007  
Herbert Van de Sompel



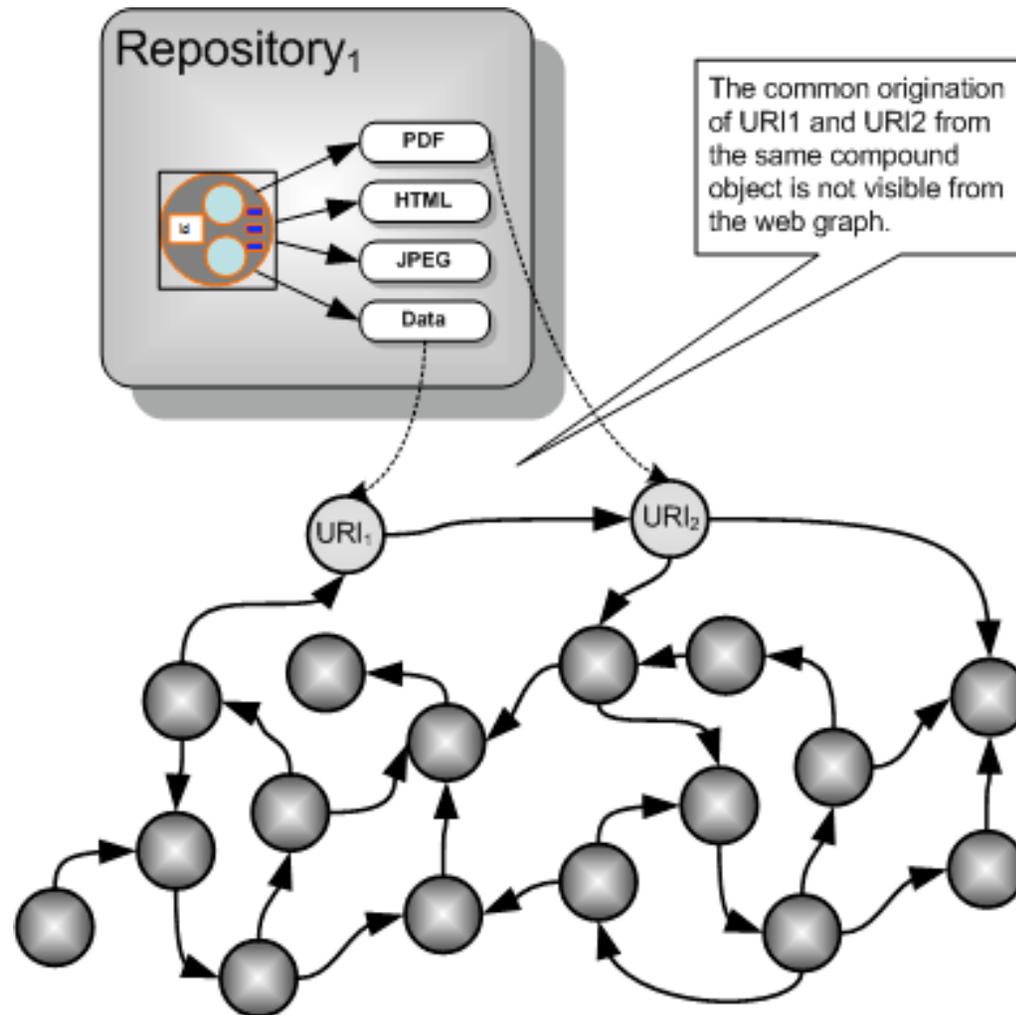
## Observation 1

Components of a compound object must be published as resources in order to be reference-able



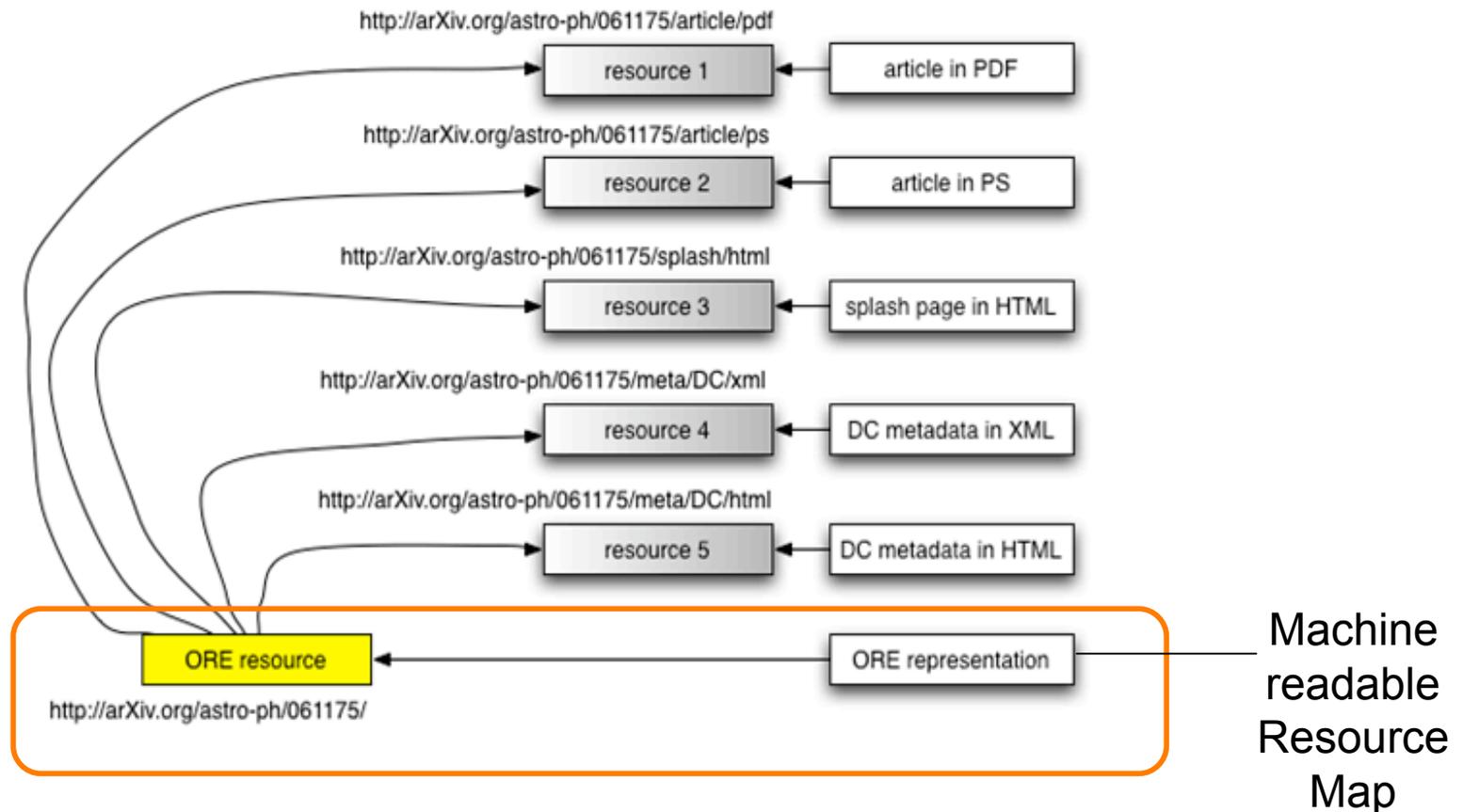
## Observation 2

The object "as such" (boundary, structure, relationships) is invisible to Web applications



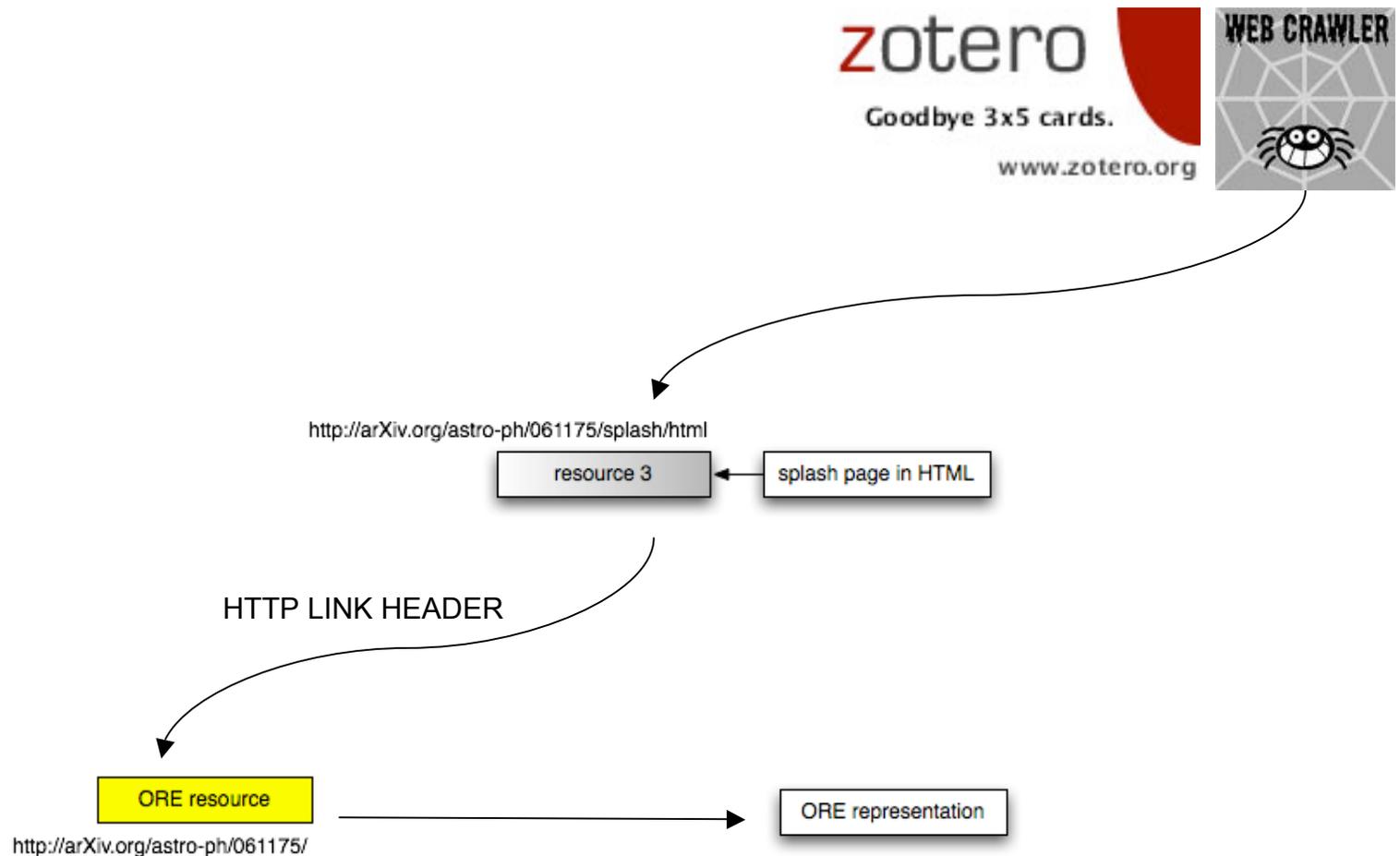
## Observation 2 bis

How about publishing a resource that makes a Resource Map available that formally expresses the boundaries of the object?



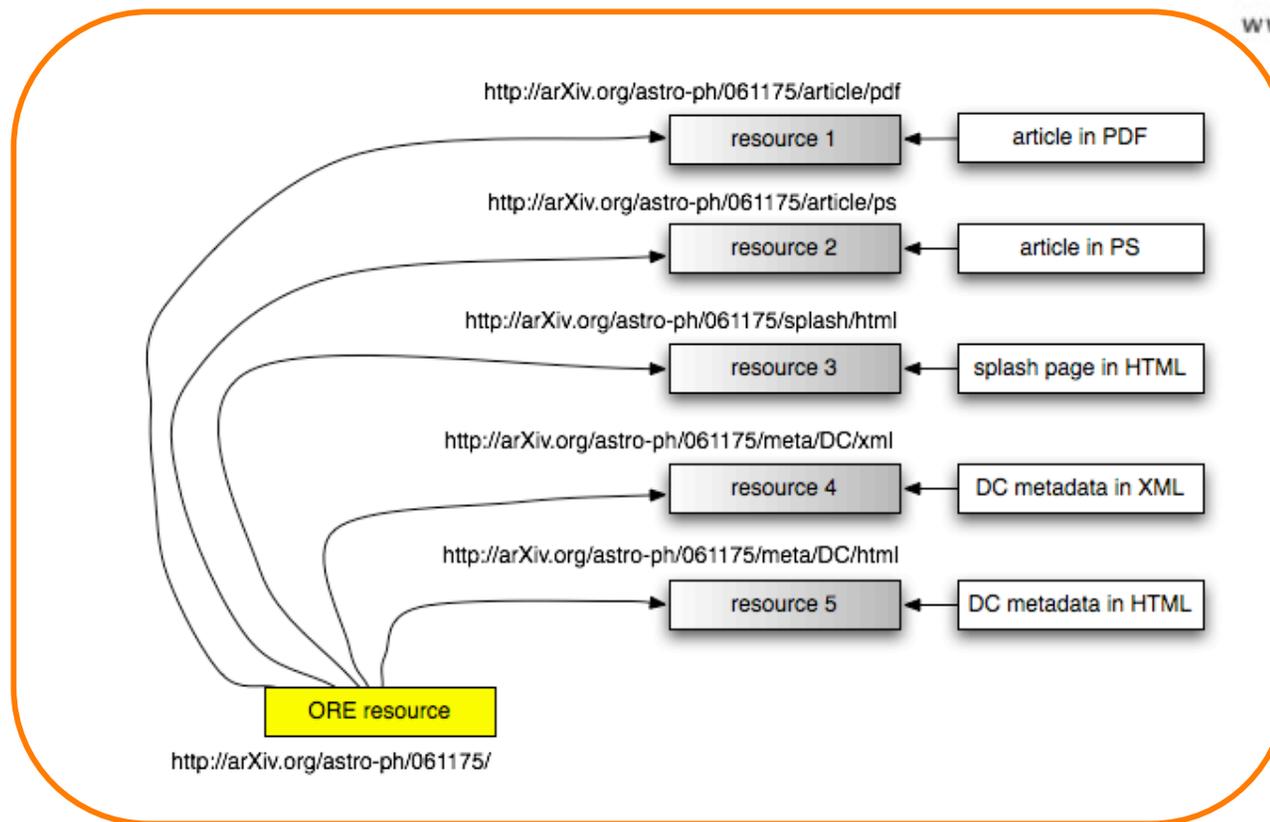
## Observation 3

And now facilitate discovery of the Resource Map (and hence of the compound object) by Web applications



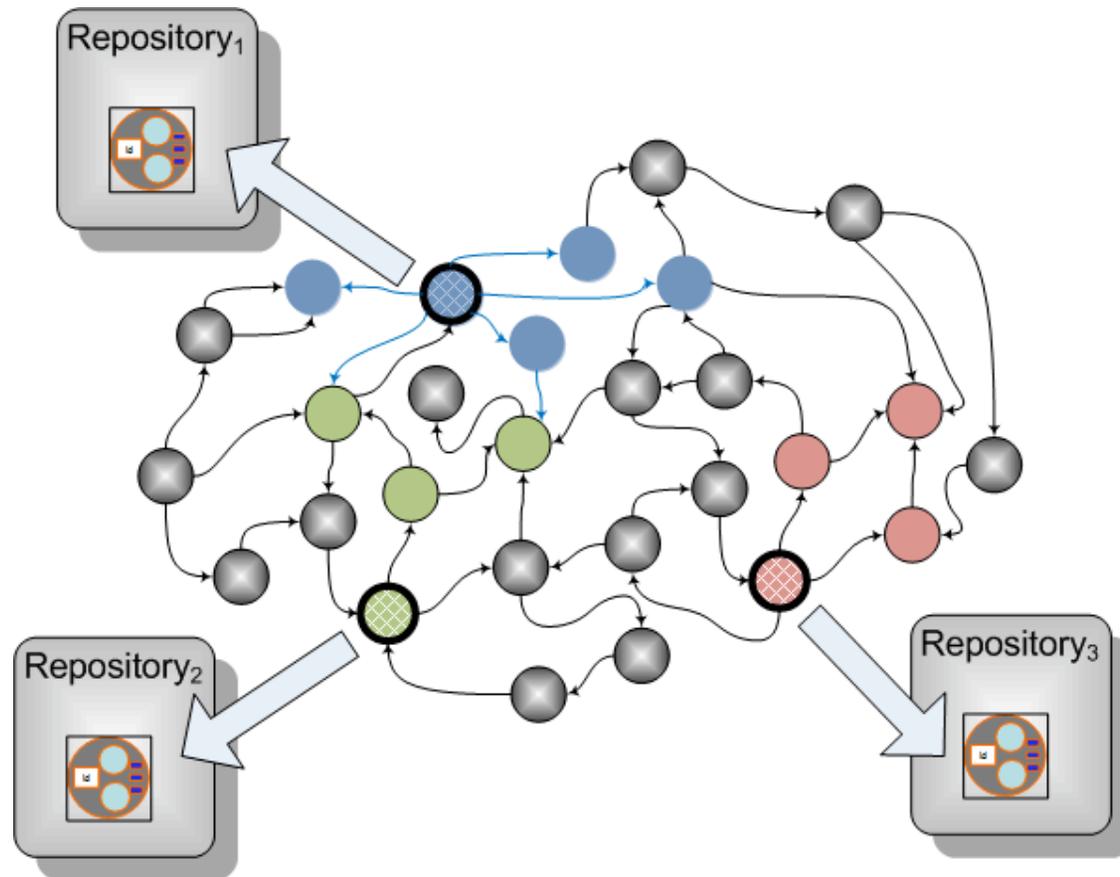
## Observation 3 bis

Through the Resource Map, the Web application sees the compound object



# Observation 4

## Resource Maps reveal compound objects in the Web graph



# Resource Map available from ORE resource

- Expresses an **aggregation of resources and relationships** in a machine-readable manner.
- Describes a **graph**:
  - finite set of resources and relationships among the resources
  - relationships among resources that are members of the aggregation and & resources are external to the aggregation
- Can be used to express:
  - Our scholarly compound objects
  - Whichever aggregation of resources and relationships
- Having a standardized format for Resource Maps opens the door to "graph publishing" (cf. Semantic Web notion).



# Use and Re-Use enabled by the ORE resource

- ORE resource has a URI:  $\text{HTTP}_{\text{ORE}}$
- $\text{HTTP}_{\text{ORE}}$  identifies a graph (cf. Semantic Web notion Named Graph)
- The Resource Map is available via HTTP GET on  $\text{HTTP}_{\text{ORE}}$
- $\text{HTTP}_{\text{ORE}}$  can become the key for object re-use: Obtain, Harvest, Register (cf. Web 2.0 mash-up)
- What is being transferred across systems is initially  $\text{HTTP}_{\text{ORE}}$  and/or the associated Resource Map.



So, where does ORE stand?



OAI Object Re-Use and Exchange  
OAI5, CERN, Switzerland, April 18th 2007  
Herbert Van de Sompel



# 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



# OAI-ORE : Afterwards

- Look into core services Obtain, Harvest, Register, in terms of ORE resource and Resource Map.
- Note:
  - It is expected that the result of ORE will largely be an aggregation of (profiles of) existing standards/specifications, not a from-scratch specification (cf. OAI-PMH).



# Questions

Further information

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



OAI Object Re-Use and Exchange  
OAI5, CERN, Switzerland, April 18th 2007  
Herbert Van de Sompel

