The OAI Object Re-Use & Exchange (ORE) Initiative

Carl Lagoze [1] & Herbert Van de Sompel [2]
Report on work by the OAI-ORE Technical Committee

[1] Information Science, Cornell University
[2] Research Library, Los Alamos National Laboratory

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

General information about OAI-ORE

OAI Object Re-Use and Exchange

- OAI-ORE is a new 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
- http://www.openarchives.org/ore/

Meeting in NYC, April 20-21 2006

- Supported by Microsoft, Mellon Foundation, Coalition for Networked Information, Digital Library Federation, JISC
- Representatives from institutional *Repository* projects, scholarly content *Repositories*, Registry projects, various projects that touch on interoperability
- See http://msc.mellon.org/Meetings/Interop/ for Agenda, Participants, Topics & Goals, Terminology, Presentations, Prototype demonstration.
- Report available since beginning of August 2006

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

OAI-ORE Scope

- Establish the basis for a *digital scholarly communication system* composed of:
  - systems that *manage content* such as institutional repositories
  - Systems and applications that *leverage managed content* such as search engines, productivity tools, and data and text analysis services.
- Of course, other application domains are possible

OAI Object Re-Use and Exchange

- OAI-ORE project organization:
  - Coordinators: Carl Lagoze & Herbert Van de Sompel
  - ORE Advisory Committee
  - ORE Technical Committee
  - ORE Liaison Group

ORE Technical Committee

- Les Carr - University of Southampton (UK)
- Leigh Dodds - Ingenta (UK)
- Tim DiLauro - Johns Hopkins University
- Dave Fulker - University Corporation for Atmospheric Research
- Tony Hammond - Nature Publishing Group (UK)
- Richard Jones - Imperial College (UK)
- Peter Murray - OhioLINK
- Michael Nelson - Old Dominion University
- Ray Plante - National Center for Supercomputing Applications
- Andy Powell - Eduserv Foundation (UK)
  - Represented by Pete Johnston at recent meeting.
- Rob Sanderson - University of Liverpool (UK)
- Simeon Warner - Cornell University
- Jeff Young - OCLC

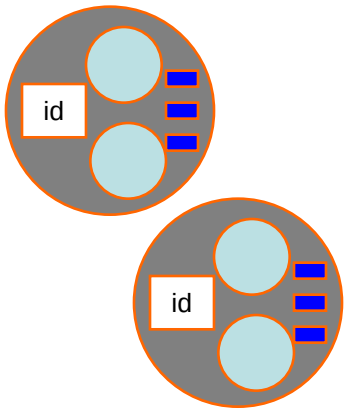
The Open Archives Initiative Protocol for Metadata Harvesting

Protocol Version 2.0 of 2002-06-14
Document Version 2004/10/12T15:31:00Z
http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

- OAI-ORE is resource centric in contrast to OAI-PMH metadata-centric approach.
- OAI-PMH will continue to exist as one approach to interoperability
- OAI-ORE will complement when richer functionality is desirable.

Context of OAI-ORE Standards & Protocols

Compound Digital Objects



Digital Objects

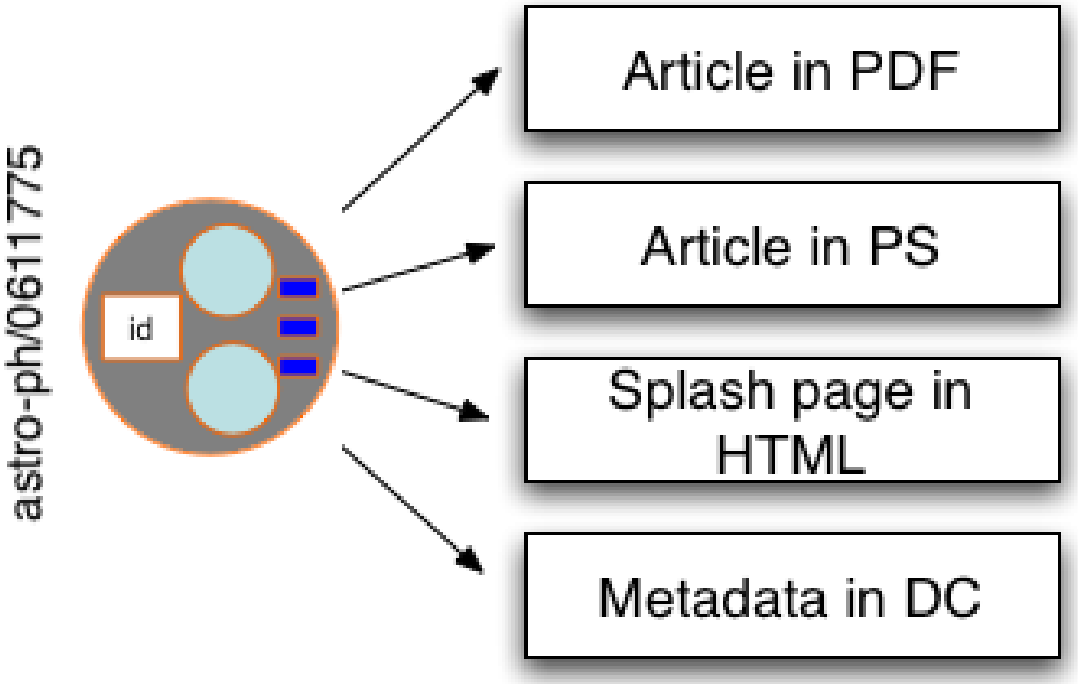
Digital content with multiple components varying on:

- **Content (semantic) types** including:
  - Text
  - Datasets
  - Simulations
  - Software
  - Dynamic knowledge representations
  - Machine readable chemical structures
  - Bibliographic and other types of metadata
- **Media types** including
  - IANA registered MIME types
  - Other type registries such as GDFR
- **Network locations** including content from:
  - Institutional repositories
  - Scientific data repositories
  - Social networking sites
  - General web
- **Relationships** including:
  - Lineage
  - Versions
  - Derivations

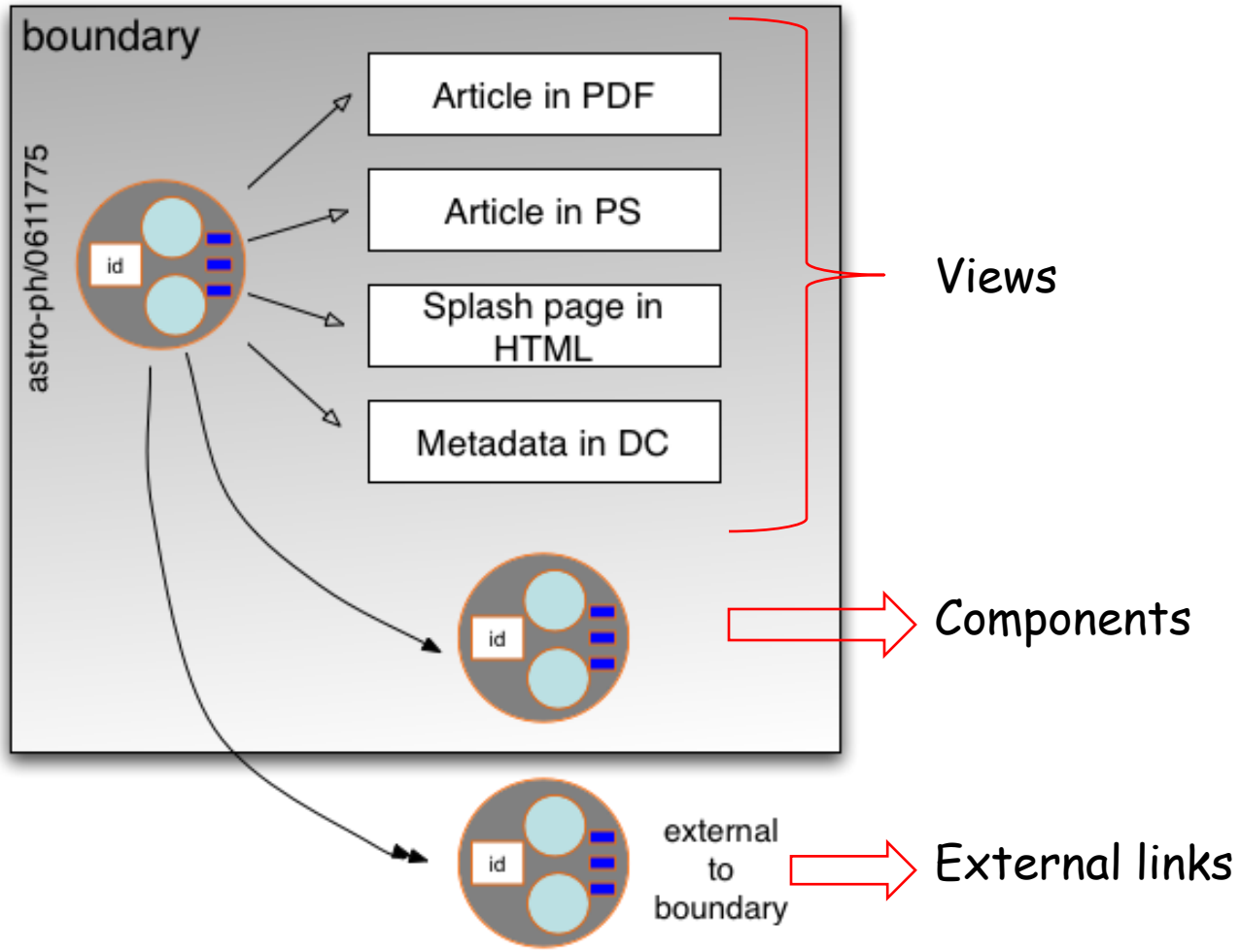
Examples of Compound Digital Objects

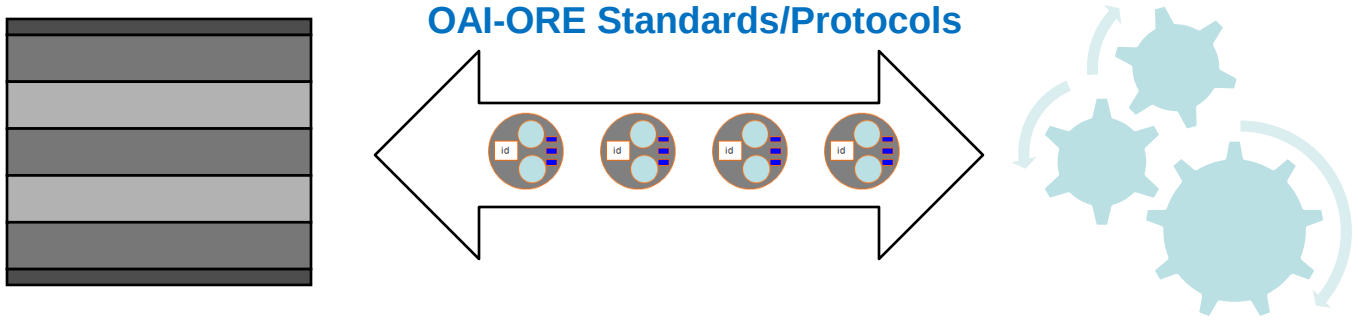
- arXiv paper with different disseminations
- An issue of an overlay journal built from distributed ePrints
- eScience publication combining text, data, simulations
- eHumanities resource combining primary and derived content

Simple Digital Object



More complexity…





Systems that manage content

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

Systems that leverage managed content

- Other repositories
- Search engines
- Authoring tools
- Citation management
- Collaborative environments
- Social network applications
- Data/Text mining applications
- Graph analysis tools
- Preservation services
- Workflow tools
- Report generation tools

This is not about asset transfer

- Many use cases do not require transfer of all digital object content
  - Preservation mirroring – of course
  - Overlay journal - no
- In many cases full transfer is not permitted, impractical (read terabytes) or superfluous

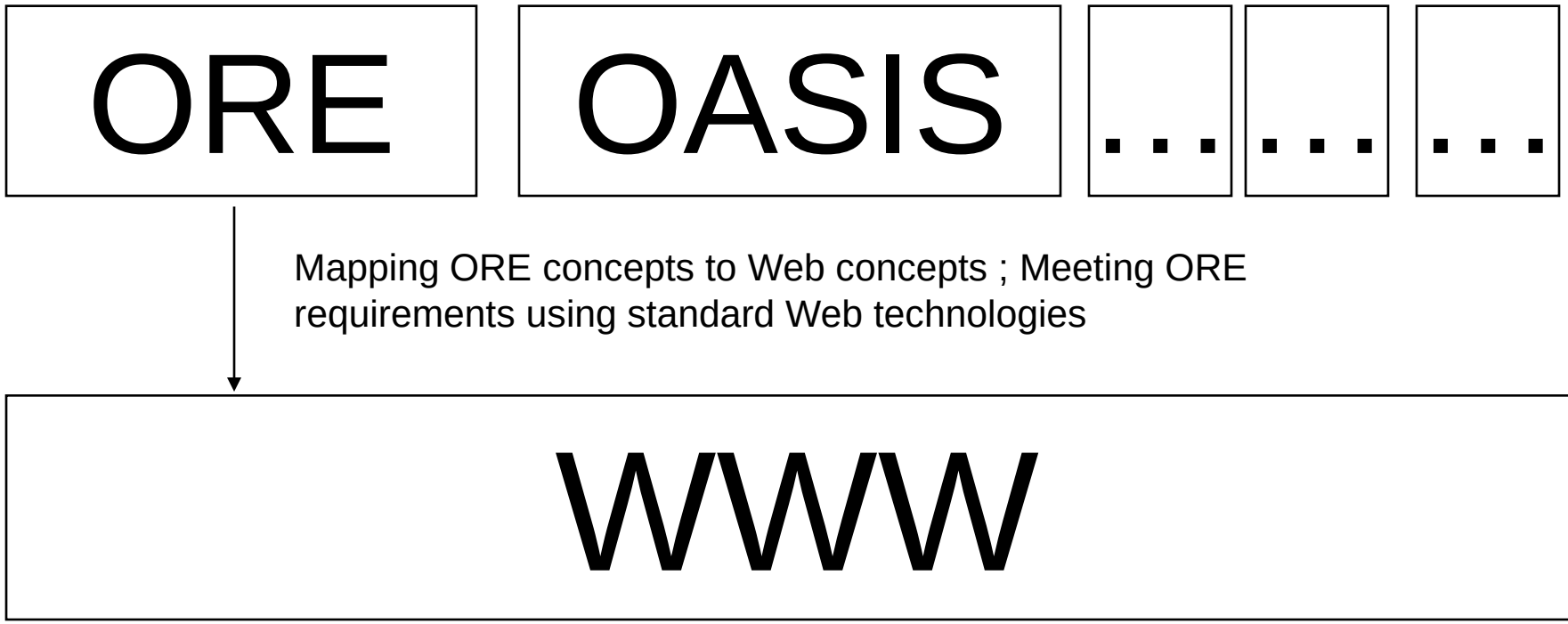
Some motivating use cases

- Find, collect, analyze, relate, and publish data-oriented scholarly objects
- Preserve compound digital objects
- Remote submission of compound digital objects
- Citation management
- Object equivalence recognition (de-duping) to aid resource discovery
- Graph-based quality assessment of data-centric scholarship

Working with the web architecture

- **Whatever we do it 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.

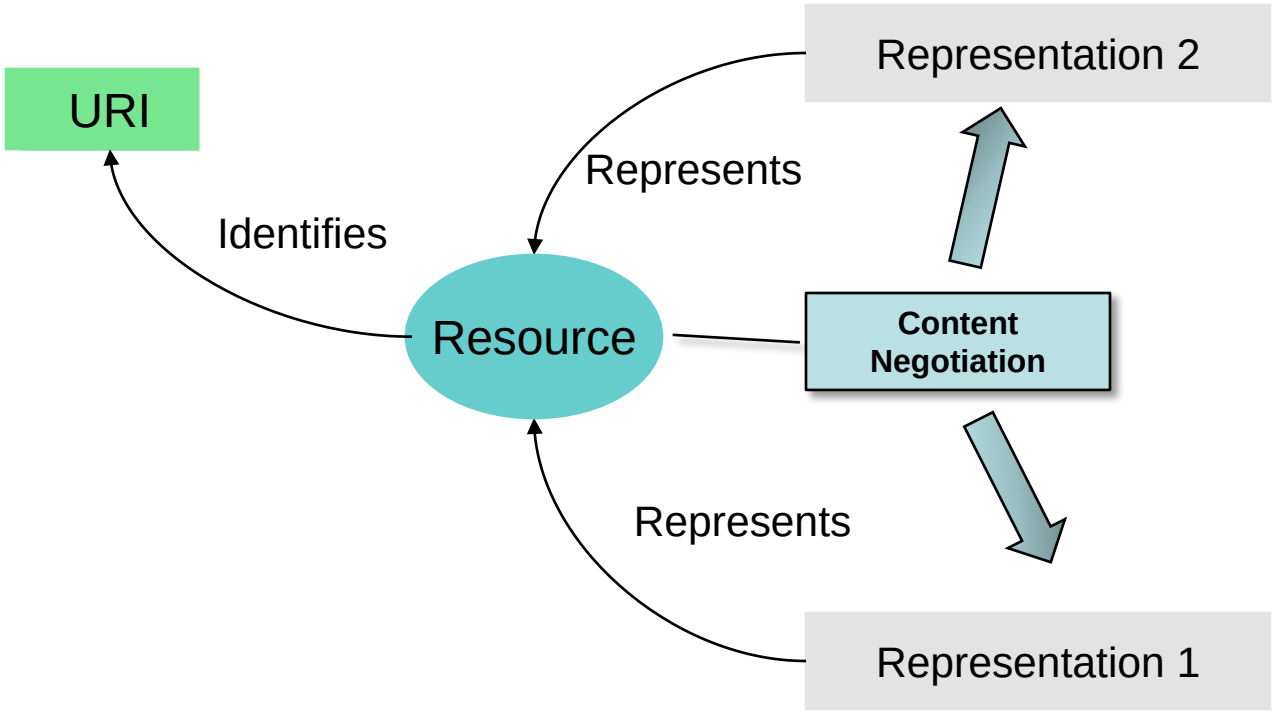
Layering the ORE world on the Web world



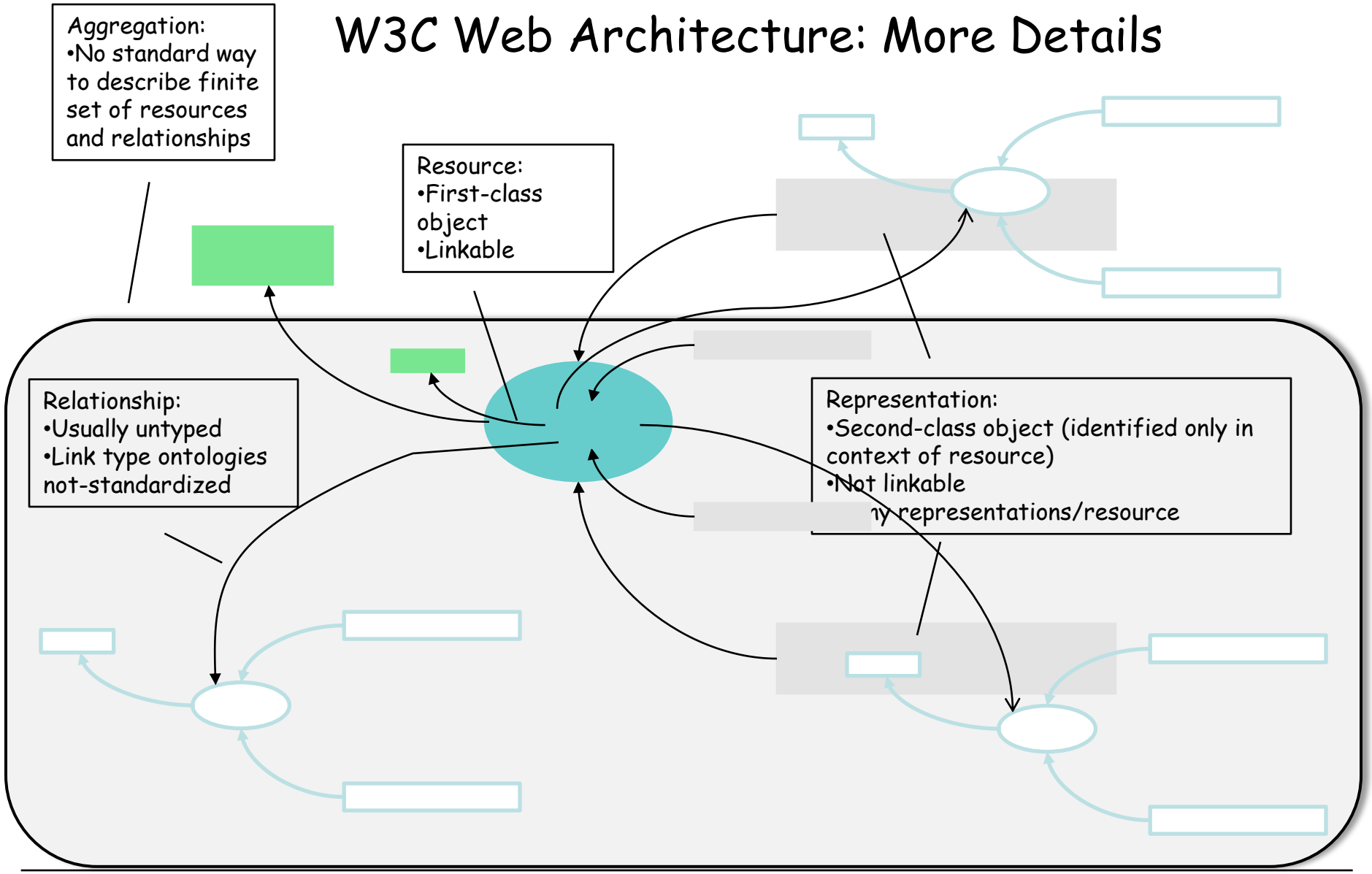
Layering the ORE world on the Web world

WWW

W3C Web Architecture



W3C Web Architecture: More Details

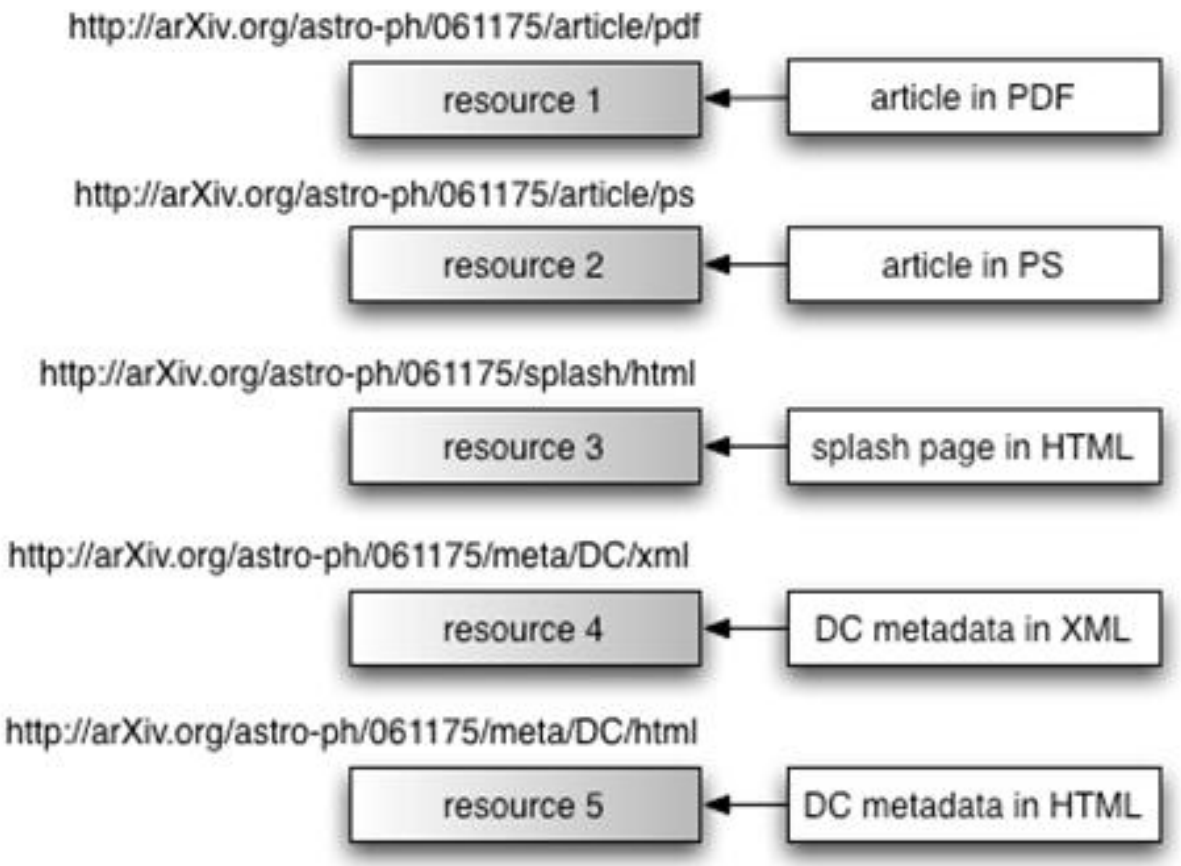


Layering the ORE world on the Web world

ORE

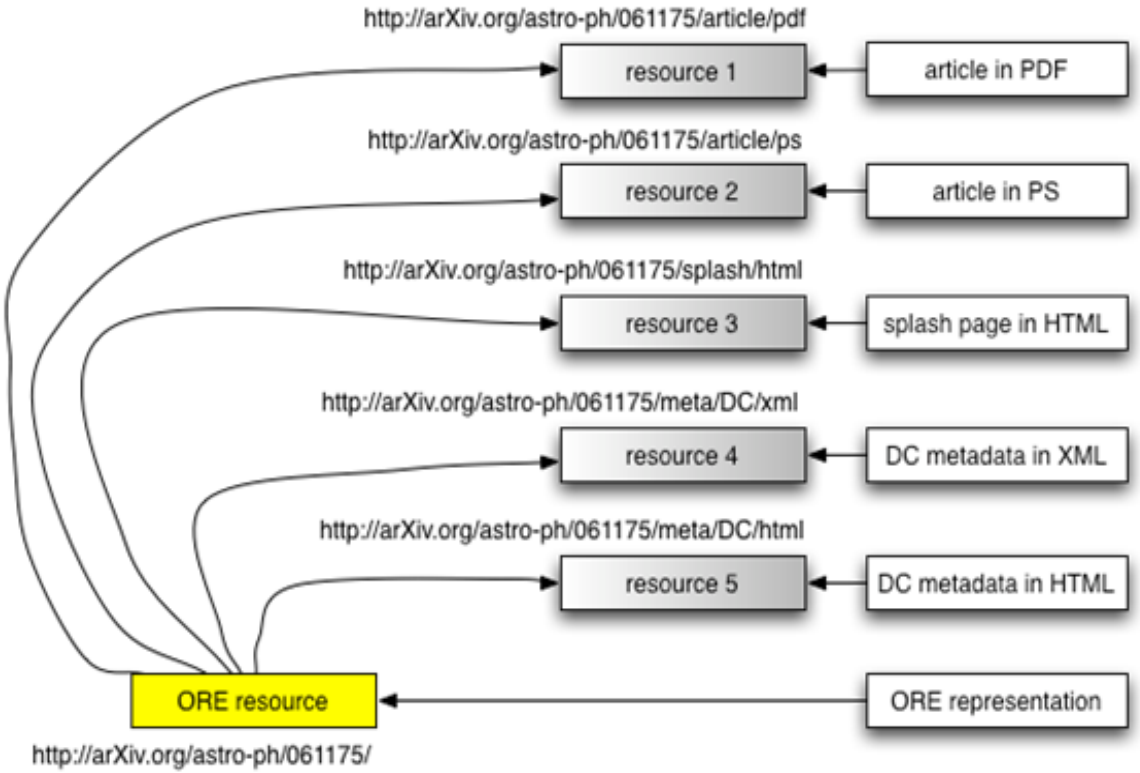
WWW

Observations (1)
Views of digital object must be bound to resources in order to be reference-able



Note: boundary of digital object is not expressed

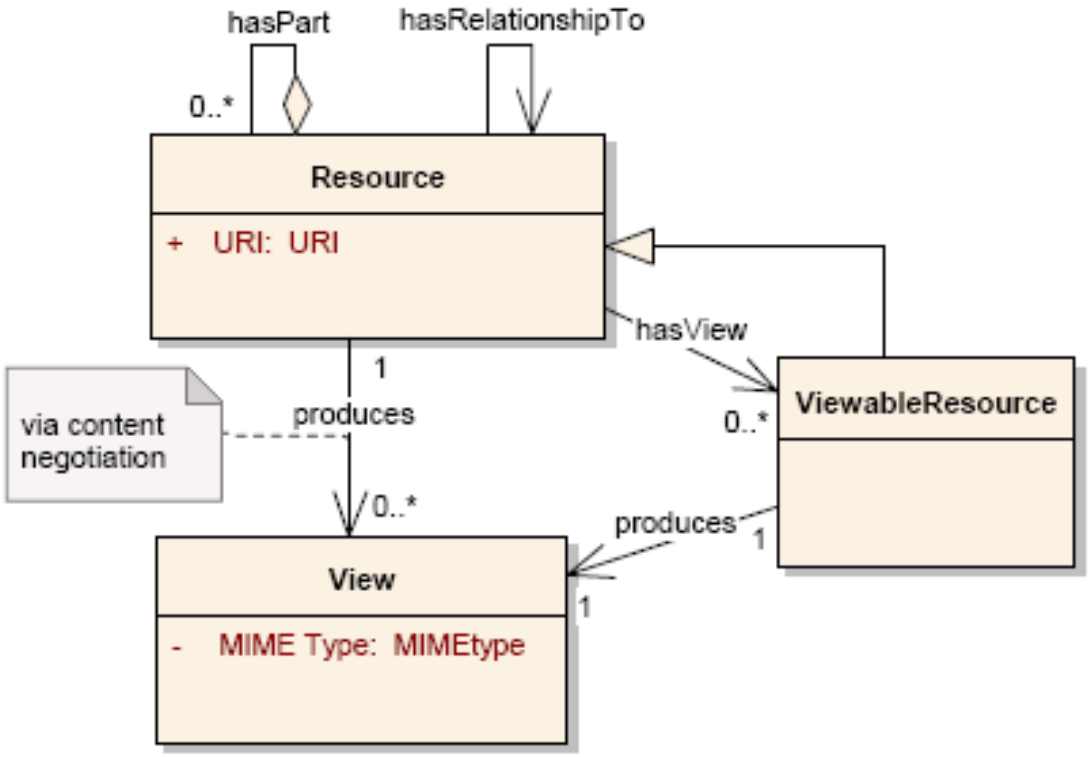
Observations (2):
Compound digital object must map to a resource with a representation that formally expresses the boundaries of the object



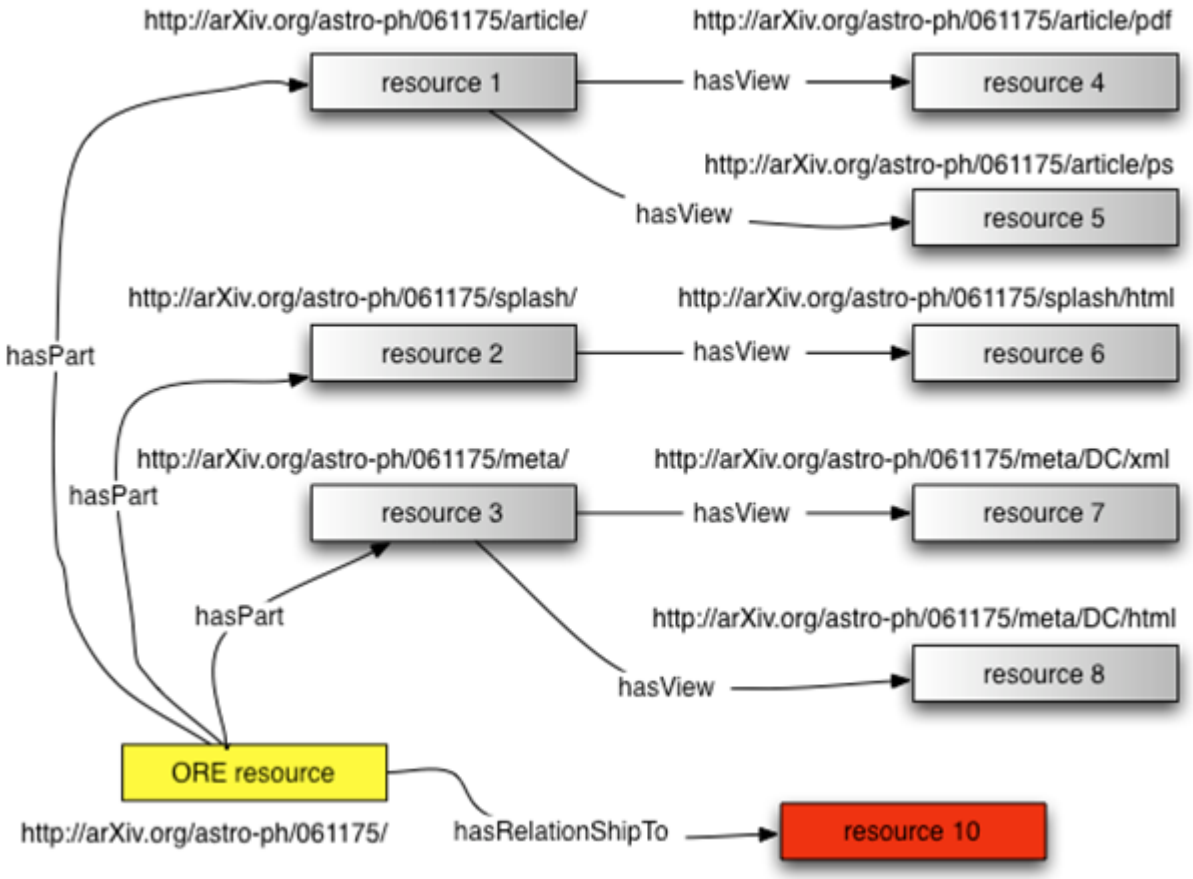
ORE Model

- Formally expresses a bounded aggregation of resources and relationships that corresponds to a compound digital object
- Describes a connected sub-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
- Intra-aggregation relationships: *hasPart*, *hasView*
- Inter-aggregation relationships: *hasRelationshipTo with* community specialization
- Sub-graph is rooted in an *ORE resource*:
  - URI identifies the compound digital object
  - Provides representations that are serializations of instances of the ORE model, which describes the corresponding *ORE Aggregation*.

Preliminary ORE Model



Compound digital object modeled according to preliminary ORE model



ORE Services

- Exchange instances of the ORE model
- ORE resources are access points for service requests
- Three classes:
  - *Harvest*: a request for a batch of instances that correspond to the ORE model from a set of ORE Resources.
  - *Obtain*: A request for an instance that corresponds to the ORE Model from a specific ORE Resource.
  - *Register*: A request to add new nodes or relationships to an ORE aggregation.

The plan from here

- Flesh out use cases as a tool for testing the model
- Review appropriate technologies
  - OAI-PMH
  - RSS/ATOM
  - openURL
  - DIDL
  - ...
- Next meeting late May 2007
  - From model to implementation

Questions