OAI-ORE : Object Reuse and Exchange
an introduction
(www.openarchives.org/ore/)

UKOLN staff seminar
UKOLN, University of Bath, 26th July 2007

Julie Allinson <j.allinson@ukoln.ac.uk>
Repositories Research Officer

www.ukoln.ac.uk
A centre of expertise in digital information management
Order of things

- Acknowledgements
- ORE overview, background, structure
- Aims and objectives, scope
- ORE – a summary of work/thinking to date
  - compound information objects
  - publishing on the Web
  - using named graphs and resource maps
- Resource maps – requirements, serialisation, discovery
- Other issues
- Next steps for ORE
Preamble

acknowledgements and overview
Acknowledgements

- This presentation is heavily based on documents produced by Carl Lagoze and Herbert van de Sompel. Notably:
  - Compound Information Objects white paper
  - Herbert van de Sompel’s presentation at JCDL 2007 (particularly for the images!)

http://www.openarchives.org/ore/
ORE overview

- ORE ...
  - commenced October 2006
  - stands for ‘Object Reuse and Exchange’
  - falls in the remit of the Open Archives Initiative (creators of OAI-PMH)
  - funded by the Mellon Foundation, with support from the National Science Foundation in the U.S.
  - has an international focus and lots of interest
  - is a 2 year project, not the answer to all our problems
  - ends September 2008
Relationship to OAI-PMH

- OAI-ORE is NOT a replacement for OAI-PMH
- OAI-PMH will continue to exist as one approach to interoperability
  - OAI-PMH metadata-centric
- OAI-ORE will complement with richer functionality, when this is desirable
  - OAI-ORE is resource centric
Relationship to Pathways

- Pathways is about ‘rethinking scholarly communications’
- It proposes an ‘interoperability infrastructure’ comprising a shared data model, surrogate format and 3 shared services
- Pathways provides much of the ‘thinking’ for OAI-ORE and involves the same key players; there is much crossover between the two projects
- OAI-ORE is evolving its own models, concepts and definitions
- Current OAI-ORE work is focusing less on the idea of surrogates and lineage, keys concept in Pathways
ORE project organisation

• Coordinators:
  - Carl Lagoze (Cornell)
  - Herbert Van de Sompel (LANL)

• ORE Advisory Committee
  - UK representatives: Liz Lyon (UKOLN) and Peter Murray-Rust (Cambridge)

• ORE Technical Committee
  - 6 out of 14 members from the UK: Pete Johnston, Rob Sanderson (Liverpool), Richard Jones (Imperial), Les Carr (Soton), Leigh Dodds (Ingenta), Tony Hammond (Nature)

• ORE Liaison Group
  - UK representatives: me (sitting in for Rachel) and Andy Powell
What is ORE doing?

aims and objectives
“Facilitate Use and Re-Use of Compound Information Objects (and of their component parts)”

- OAI-ORE presentation at JCDL 2007

- Adhering to the first and second principles of
  - enriching the web graph with boundary information
  - adding meaning to the web graph through relationships and links

- Currently OAI-ORE is doing a lot of thinking ... about the scope, the problem space, the requirements, the abstract model
Aims and objectives

• Overall objective:
  - Identify, profile and develop 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.
Specific aims (1)

- To provide effective and consistent ways:
  - to facilitate **discovery** of objects,
  - to **reference** (link to) objects (and their parts),
  - to **obtain** a variety of disseminations of objects,
  - to aggregate and disaggregate objects,
  - to **harvest** and **deposit** (register, put) objects
  - to enable processing by automated agents
Specific aims (2)

• To 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

• In other words, to establish the basis of an ‘interoperability layer’
Where ORE is at

a summary of current work and thinking
Compound Information Objects

- Identified, bounded aggregations of distinct information units that when combined form a logical whole
- Examples vary according to:
  - Semantic type: book, article, software, dataset, simulation, ...
  - Media type: text, image, audio, video, mixed
  - Media format: PDF, HTML, JPEG, MP3, ...
  - Network location
  - Relationships: internal, external
Some examples

- Scholarly:
  - 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.

- Not-so scholarly
  - Flickr ‘sets’, comments/annotations etc.

- OAI-ORE presentation at JCDL 2007
Doing it on the Web

- information systems already share compound object components using the web architecture
- but the notion of boundary and typed relationships is lost
- enriching the web with this kind of additional information is a fundamental principle for OAI-ORE
Web architecture

• OAI-ORE is agreed that it must leverage the Web Architecture, through:
  – (HTTP) URIs that identify resources, which are “items of interest”, that,
  – when accessed through standard protocols such as HTTP, return representations of current resource state
  – and which are linked via URI references
  – content negotiation allows multiple representations to be served up from the same HTTP URI
Web architecture in pictures

URI

Identifies

Resource

Represents

Representation 2

Represents

Content Negotiation

Represents

Representation 1

Relationships/links are usually untyped
Publishing compound objects to the Web (1)

- Web graph without any explicit compound objects
- each information object identified with a URI
- and there are links between them
Publishing compound objects to the Web (2)

- Compound object and its parts are published to the Web with URIs
- Links indicate relationships but cannot show boundaries and true structure in a machine context
This time ... added layer is publishing the compound object and its parts with relationships and boundary as a ‘named graph’
The named graphs approach

- A named graph describes the compound object
- It consists of nodes and arcs where nodes are resources and arcs are typed relationships
- It is a web resource and can be referenced by any resource on the web
- It is not the compound object itself
- It is identified by a HTTP URI
- Which points to the Resource Map (ReM) - an encoded description (serialization) of the named graph
Named graph publishing (with added déjà vu)

- Named graph is published at a HTTP URI
- Resource Map is available from that HTTP URI
Named graph publishing on the Web

- Notice the compound object node (6)
- Typed relationships, e.g. between (1) and (5)
- Untyped relationships to other stuff on the Web
Everything together

- Compound object is published as a ‘named graph’ identified by a HTTP URI
- Represented by a Resource Map
Re-use and context

Can reference (i.e. re-use) as follows:

(U) : just the resource identified by (U)

(X) : just the named graph identified by (X)

(U) in the context of (X): the resource identified by (U), but as it exists in the context of named graph (X)
Resource maps
the what and the how
Resource map (ReM) requirements

- Resource Maps **must** allow for simply expressing the resources that are considered part of a compound object.
- Resource Maps **may**
  - Express resources that are not part of a compound object.
  - Distinguish between resources that are part of the compound object and those that are not.
  - Express the relationships among the resources referenced by the named graph.
  - Express the types of the relationships among the resources referenced by the named graph, i.e. label the arcs.
  - Express other information related to the named graph and to the resources that it references such as metadata, etc.
Serialising the ReM

- The ReM can be viewed as a splash page for machine consumption
- Work on a draft specification and serialisation for the ReM is imminent
- Candidate specifications include
  - RDF/XML
  - TriX
  - ATOM
  - YADS
Discovering the ReM

- Harvest discovery
  - e.g. by exposing via OAI-PMH, sitemaps or ATOM collections lists

- GET discovery
  - using an ORE-specific HTTP header, e.g.
    - X-ORE-Named-Graph:

- Linked data discovery
  - using HTTP link headers
  - embedding context information into link attributes
Almost finished

the what else and the what
next
Recap

- ORE is working towards facilitating a scholarly communications interoperability layer through the publication of named graphs which describe compound objects, their boundaries and relationships and in so doing make them re-usable and discoverable.
Outstanding issues

- Rootedness and connectedness – the need for a ‘containment’ or ‘compound object’ node
- More work needed on the scope and requirements
  - authored objects vs arbitrary ‘bags’ of resources
  - inclusion of non-information resources
- Defining an abstract model
- Thinking about the services that need to be supported, e.g. harvest, obtain (GET), but also deposit (PUT)
- Authority and ownership, assessment of trustworthiness of Named Graphs
- Development of vocabularies for expressing types of links between resources contained in a named graph
- Development of vocabularies for expressing properties of resources in a named graph, e.g. semantic type, media type and media format

not all are in scope for ORE!
Next steps

- ORE Acceleration Project will produce an alpha specification by 30 September 2007
- For use in an (almost funded) eChemistry project

and that’s it