2009

Object Reuse and Exchange

Michael L. Nelson
Old Dominion University

Carl Lagoze
Cornell University

Herbert Van de Sompel
Los Alamos National Laboratory

Pete Johnston
Eduserv Research Programme, Bath, UK

Robert Sanderson
Los Alamos National Laboratory

See next page for additional authors

Original Publication Citation

Nelson, M. L., Lagoze, C., Van De Sompel, H., Johnston, P., Sanderson, R., & Warner, S. (2009). Object reuse and exchange. In J. Sieck & M. A. Herzog (Eds.), *Wireless Communication and Information : Radio Engineering and Multimedia Applications*. Verlag Werner Hülsbusch.

Object Reuse and Exchange

Michael L. Nelson[1], Carl Lagoze[2], Herbert Van de Sompel[3], Pete Johnston[4], Robert Sanderson[3], Simeon Warner[2]

[1]Department of Computer Science
Old Dominion University
Norfolk, VA 23529 USA

[2]Information Science Program
Cornell University
Ithaca, NY 14850 USA

[3]Research Library
Los Alamos National Laboratory
Los Alamos, NM 87545 USA

[4]Eduserv Research Programme
Bath, UK

Abstract

The Open Archives Initiative Object Reuse and Exchange (OAI-ORE) project defines standards for the description and exchange of aggregations of Web resources. The OAI-ORE abstract data model is conformant with the Architecture of the World Wide Web and leverages concepts from the Semantic Web, including RDF descriptions and Linked Data. In this paper we provide a brief review of a motivating example and its serialization in Atom.

1 OAI-ORE Motivation

The Architecture of the World Wide Web is built on the fundamental concepts of *URIs*, which identify *resources*, that when dereferenced return *representations*. While resources form the basic unit of granularity, the web is often experienced in terms of *websites*, loosely defined to be a collection of resources at on or more web servers. While the boundary of websites can be difficult to determine, formats such as the Sitemap Protocol allow webmasters to explicitly list the resources that should be considered as part of a website. In between the granularity of web resource and website are syndication feeds. Expressed in formats such as RSS and Atom, syndication feeds are a significant catalyst for what is known as "Web 2.0". A single website frequently has several *syndication feeds* that list web resources according to some criteria. For example, CNN.com has many syndication feeds corresponding to various news categories: latest, most popular, travel, sports, etc. By the nature of the RSS and Atom formats, feeds unambiguously enumerate and describe their members.

The OAI-ORE project defines a new level of web granularity: the *Aggregation*. An Aggregation contains one or more web resources, but is "smaller" than a syndication feed (and thus, smaller than a website as well). Aggregations have always implicitly existed on the web, but there has not been a format for unambiguously expressing their members. For example, a book on the web can consist of many web resources: various chapters and pages, inlined images, appendices, etc. In part because we are familiar with the meme "book", we intuitively know which web resources comprise the book and which web resources are not part of the book (e.g., links to Adobe Acrobat Reader). But even in this scenario, there are conditions that can be difficult to discern: are errata, translations and different file formats "part" of the book, or simply "related" to the book?

It is common for agents (people and robots) to use the URI of the "splash page" to represent a single web resource as well as the logical collection of resources accessible from that splash page. For example, it is common to alternatively use the URI: http://www.youtube.com/watch?v=0PKDJrIMJFs

to refer both to the HTML splash page embedding the video as well as the "video" itself. This is similar to the concept of "indirect identification" from the Architecture of the World Wide Web [Jac04]. While it is possible to configure a robot to correctly parse certain types of splash pages to correctly infer the logical collection of web resources, this approach cannot be generalized. For example, although both are represented as web resources, the errata of a book is semantically different than a PDF version of the book.

2 arXiv.org Example

Figure 1 shows the "splash page" of a scholarly eprint from arXiv.org with the following components:

1. The URI http://arxiv.org/abs/astro-ph/0601007 of the human start page.
2. The formats in which the document is available, i.e. PostScript, PDF, etc. These are effectively the constituents of the aggregation that is the arXiv document. For the remainder of this example we will consider this human start page, the splash page, as also a constituent of the aggregation
3. The title of the arXiv document.
4. The authors of the arXiv document.
5. The creation and last modification date of the arXiv document.
6. Identifiers of entities that are in some manner comparable to this arXiv document. For example, a version of this document was later published as an article in a peer-reviewed journal, and the Digital Object Identifier of that article is shown.
7. The versions of this document.
8. Links to other arXiv documents in the same collection (i.e., astro-ph).
9. Citations made by this arXiv document, and citations it received from other documents.

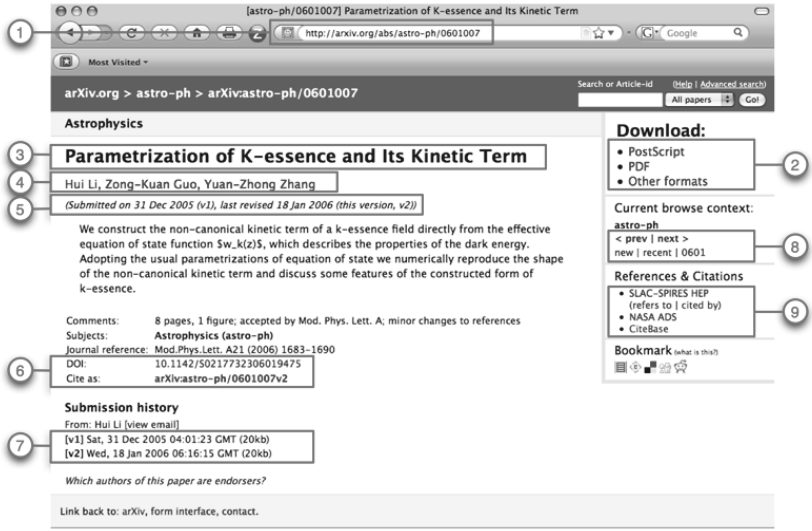

Fig. 1: Human start page for an arXiv document.

3 ORE Data Model Basics

In order to be able to unambiguously enumerate and describe an aggregation of Web resources such as shown in Figure 1, a new Resource is introduced that stands for a set or collection of other Resources. This new Resource, named an Aggregation, has a URI just like any other Resource on the Web does. And, since an Aggregation is a conceptual construct, it qualifies as Semantic Web Resource that does not have a Representation.

Following the Linked Data guidelines [Biz07], another Resource is introduced to make information about the Aggregation available. This new Resource, named a Resource Map, has a URI and it has a machine-readable Representation that provides details about the Aggregation. In essence, a Resource Map expresses which Aggregation it describes, and it lists the resources that are part of the Aggregation. But, a Resource Map can also express relationships and properties pertaining to all these Resources, as well as metadata pertaining to the Resource Map itself, e.g. who published it and when it was most recently

modified. Resource Maps can be expressed in different formats including Atom, RDF/XML, RDFa, n3, turtle, and other RDF serialization formats.

Finally, the resources aggregated by an Aggregation (which is serialized by a Resource Map) are called Aggregated Resources. For example, part 2 of Figure 1 constitutes the Aggregated Resources of that Aggregation. An abstracted example is shown in Figure 2, where Aggregation A-1 aggregates three Aggregated Resources (AR-1, AR-2, AR-3) and A-1 is described by a Resource Map (ReM-1). The Resource Map contains metadata about both A-1 and ReM-1.

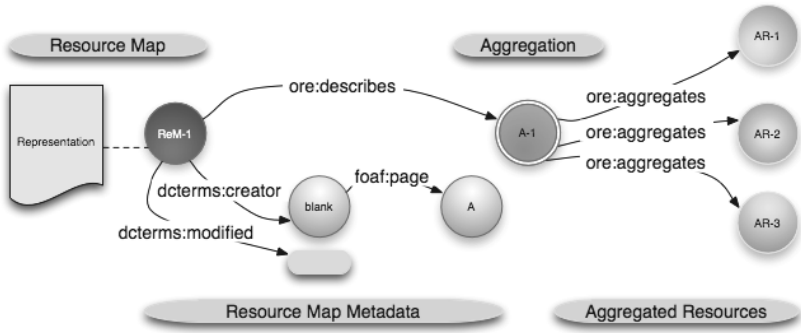

Fig. 2: ORE Data Model Basics: Aggregation, Resource Map, Aggregated Resources.

4 A Simple Resource Map

In Figure 3, an Atom serialization of a Resource Map based on the example in Figure 1 is presented. The Atom entry document corresponds to an Aggregation; the Atom feed document has no ORE semantics and is simply a method for publishing Aggregations. The /entry/link[@rel="self"]/@href element corresponds to the URI of the Resource Map, and the /entry/link[@rel="http://arxiv.org/aggregation/astro-ph/0601007 "]/@href element corresponds to the URI of the Aggregation. The /entry/link[@rel="http://www.openarchives.org/ore/terms/aggregates "]/@href elements correspond to the URIs of the various Aggregated Resources – in this case the PDF and PostScript version of the eprint as well as the HTML splash page. The /entry/link[@rel="alternate"]/@href and /entry/id elements have no ORE semantics and exist only for Atom

purposes. The rest of the Atom elements describe metadata for the Resource Map and Aggregation.

```
– <atom:entry>
    <atom:id>tag:arxiv.org,2008:astro-ph:0601007</atom:id>
    <!-- About the Aggregation for the ArXiv document -->
  – <!--
      This Atom entry describes a specific ORE Aggregation
    -->
    <atom:link href="http://arxiv.org/aggregation/astro-ph/0601007" rel="http://www.openarchives.org/ore/terms/describes"/>
    <atom:category term="http://www.openarchives.org/ore/terms/Aggregation" scheme="http://www.openarchives.org/ore/terms/"
    label="Aggregation"/>
    <!-- The Aggregation aggregates ... -->
    <atom:link href="http://arxiv.org/abs/astro-ph/0601007" rel="http://www.openarchives.org/ore/terms/aggregates" title="[astro-
    ph/0601007] Parametrization of K-essence and Its Kinetic Term" type="text/html"/>
    <atom:link href="http://arxiv.org/ps/astro-ph/0601007" rel="http://www.openarchives.org/ore/terms/aggregates" title="Parametrization of
    K-essence and Its Kinetic Term" type="application/postscript" hreflang="en"/>
    <atom:link href="http://arxiv.org/pdf/astro-ph/0601007" rel="http://www.openarchives.org/ore/terms/aggregates" title="Parametrization
    of K-essence and Its Kinetic Term" type="application/pdf" hreflang="en"/>
  – <!--
      Metadata about the Aggregation: title and authors
    -->
    <atom:title>Parametrization of K-essence and Its Kinetic Term</atom:title>
  – <atom:author>
      <atom:name>Hui Li</atom:name>
      <atom:email>lihui@somewhere.cn</atom:email>
    </atom:author>
  – <atom:author>
      <atom:name>Zong-Kuan Guo</atom:name>
    </atom:author>
  – <atom:author>
      <atom:name>Yuan-Zhong Zhang</atom:name>
    </atom:author>
  – <!--
      About the Resource Map (this Atom XML entry document) that describes the Aggregation
    -->
    <!-- The HTTP URI of this Resouce Map -->
    <atom:link href="http://arxiv.org/rem/atom/astro-ph/0601007" rel="self" type="application/atom+xml"/>
  – <!--
      Metadata about the Resource Map: datetimes, rights, and author
    -->
    <atom:updated>2008-10-03T07:30:34Z</atom:updated>
    <atom:published>2008-10-01T18:30:02Z</atom:published>
  – <atom:rights>
      This Resource Map is available under the Creative Commons Attribution-Noncommercial Generic license
    </atom:rights>
    <atom:link href="http://creativecommons.org/licenses/by-nc/2.5/rdf" rel="license" type="application/rdf+xml"/>
  – <atom:source>
    – <atom:author>
        <atom:name>arXiv.org e-Print Repository</atom:name>
        <atom:uri>http://arxiv.org</atom:uri>
      </atom:author>
    </atom:source>
  – <!--
      About the human start page that is part of the Aggregation
    -->
    <atom:link href="http://arxiv.org/abs/astro-ph/0601007" rel="alternate"/>
</atom:entry>
```

Fig. 3: An Atom Resource Map for the Aggregation in Figure 1.

5 ORE Summary

This paper has presented only the simplest example of an Aggregation and a corresponding Resource Map, in which Resource Maps function as a sort of unambiguous, machine-readable splash page to explicitly describe the kinds of implicit Aggregations that already exist on the web. Additional semantics, serialization examples, http implementation and discovery issues, and a

review of early adopting projects are discussed in further detail in [ORE08] and [Van09].

Literature

[Biz07] Bizer, C., Cyganiak, R., Heath, T., How to Publish Linked Data on the Web, 2007. http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/

[Jac04] Jacobs, I. and Walsh, N. (eds.), Architecture of the World Wide Web, Volume One, W3C Recommendation 15 December 2004. http://www.w3.org/TR/webarch/

[ORE08] ORE Specifications and User Guides - Table of Contents. http://www.openarchives.org/ore/1.0/toc.html

[Van09] Van de Sompel, H., Lagoze, C., Nelson, M. L., Warner, S., Sanderson, R., Johnston, P., Adding eScience Assets to the Data Web, Proceedings of Linked Data on the Web 2009 (LDOW2009). http://arxiv.org/abs/0906.2135