Interoperability for the Discovery, Use, and Re-Use of Units of Scholarly Communication

CTWatch QUARTERLY
August 2007

Herbert Van de Sompel, Los Alamos National Laboratory
Carl Lagoze, Cornell University

1. Introduction

Improvements in computing and network technologies, digital data capture, and data mining techniques are enabling research methods that are highly collaborative, network-based, and data-intensive. These methods challenge existing scholarly communication mechanisms, which are largely based on physical (paper, ink, and voice) rather than digital technologies.

One major challenge to the existing system is the change in the nature of the *unit* of scholarly communication. In the established scholarly communication system, the dominant communication units are journals and their contained articles. This established system generally fails to deal with other types of research results in the sciences and humanities, including datasets, simulations, software, dynamic knowledge representations, annotations, and aggregates thereof, all of which should be considered units of scholarly communication.[1]

Another challenge is the increasing importance of machine agents (e.g., web crawlers, data mining applications) as consumers of scholarly materials. The established system by and large targets human consumers. However, all communication units (including the journal publications) should be available as source materials for machine-based applications that mine, interpret, and visualize these materials to generate new units of communication and new knowledge.

Yet another challenge to the existing system lies in the changing nature of the social activity that is scholarly communication. Increasingly, this social activity extends beyond traditional journals and conference proceedings, and even beyond more recent phenomena such as preprint systems, institutional repositories, and dataset repositories. It now includes less formal and more dynamic communication such as blogging. Scholarly communication is suddenly all over the web, both in traditional publication portals and in new social networking venues, and is interlinked with the broader social network of the web. Dealing adequately with this communication revolution requires fundamental changes in the scholarly communication system.

Many of the required changes in response to these challenges are of a socio-cultural nature and relate directly to the question of what constitutes the scholarly record in this new environment. This raises the fundamental issue of how the crucial functions of scholarly communication [2] – registration, certification, awareness, archiving, rewarding – should be re-implemented in the new context. The solutions to these socio-cultural questions rely in part on the development of basic technical infrastructure to support an innately digital scholarly communication system.

This paper describes the work of the Object Re-Use and Exchange (ORE) project of the Open Archives Initiative (OAI) to develop one component of this new infrastructure in order to support the revolutionized scholarly communication paradigm – standards to facilitate discovery, use and re-use of new types of *compound* scholarly communication units by networked services and applications. *Compound units* are aggregations of distinct information units that, when combined, form a logical whole. Some examples of these are a digitized book that is an aggregation of chapters, where each chapter is an aggregation of scanned pages, and a scholarly publication that is an aggregation of text and supporting materials such as datasets, software tools, and video recordings of an experiment. The ORE work aims to develop mechanisms for representing and referencing compound information units in a machine-readable manner that is independent of both the actual content of the information unit and nature of the re-using application.

2. Compound Information Objects

The new units of communication that are emerging from the modern research environment have a compound nature that does not have a direct parallel in traditional, paper-based publications or in the digital versions thereof (e.g., pdf, LaTex). They are aggregates of multiple distinct components that can vary according to semantic type (article, simulation, video, dataset, software, etc.), media type (text, image, audio, video, mixed), media format (PDF, XML, MP3, etc.), and network location (different components made accessible by different repositories). In addition, each aggregate carries an identifier associated with it by the information system that composed the aggregation, thereby establishing it as a logical unit of scholarly communication. In the remainder of this paper, we will refer to these aggregates as either *compound information objects* or *compound objects* (Figure 1).

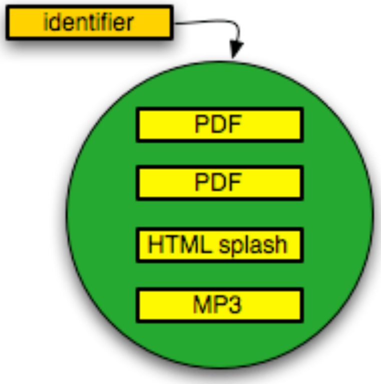


Figure 1. A compound information object composed by an information system.

These compound objects are a fundamental building block of eScience and eScholarship, and support for them is an essential aspect of cyberinfrastructure.[3] For example, the ImageWeb [4] activity led by David Shotton's BioInformatics Research Group at the University of Oxford explores the creation of so-called *image webs* that integrate cellular images held by publishers, research organizations, museums, and institutional repositories. Also, Gregory Crane, a leading scholar in the humanities, envisions the notion of *recombinant documents*.[5] These documents have a number of features that differentiate compound documents from physical documents or their digital incunabula.[5] They aggregate new information and existing fine-grained digital information. The aggregation can be human-author based, for example, as the result of a workflow within a so-called *scholarly workbench*,[6] or machine-generated based, for example, on machine learning techniques and web crawling.[7] The aggregation of an existing information unit into a compound object (re-use) is not due to the inherent nature of the aggregated unit, but is the result of the algorithmic design or the intention of the human that composed the compound object. Finally, these objects may be dynamic and grow over time based on usage patterns as well as social activity that provide additional context for the information within them.[8]

3. Publishing Compound Objects to the Web

The layer cake metaphor is commonly used to describe information infrastructure, where new layers of functionality build upon existing layers. Tim Berners-Lee has, for example, used this model to describe the semantic web that consists of functionality built on the base web architecture.[9] The work of ORE follows the same paradigm. It presumes the web architecture [10] as the de facto foundation for interoperability and positions the ORE standards as a layer over this web foundation. Thereby the ORE work leverages the facilitites provided by the web architecture, adding functionality related to compound objects that is not present in the web foundation layer.

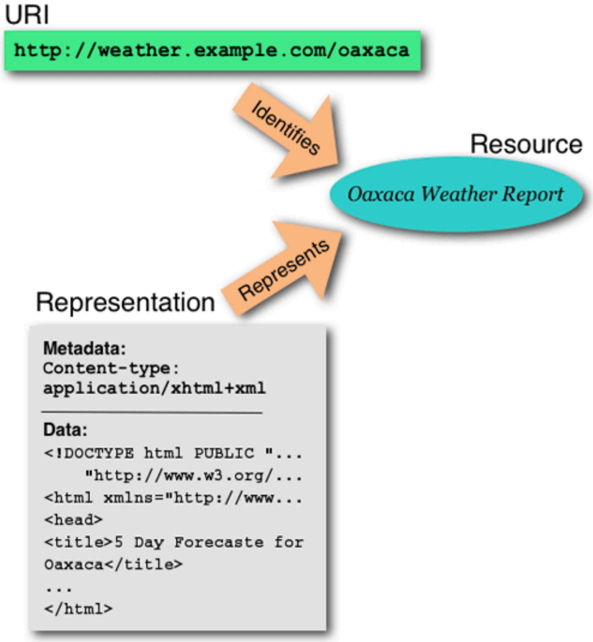


Figure 2. Web Architecture (taken from www.w3.org/TR/webarch/).

This web foundation layer (Figure 2) defines architectural notions that allow information systems that compose compound objects to publish them to the web by associating a URI with each of the components of a compound object, thereby making the components URI-identified *resources*. Web services and applications, such as browsers and crawlers, can use these URIs to obtain representations of the resources via *content negotiation*.

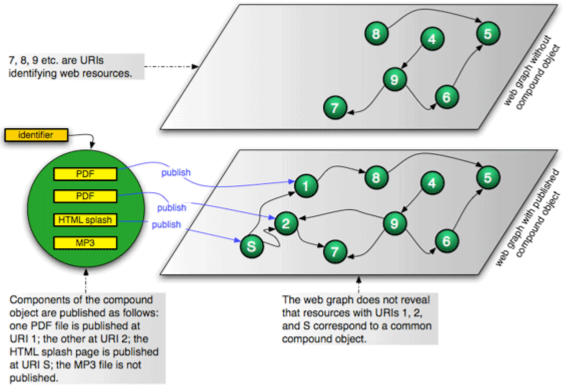
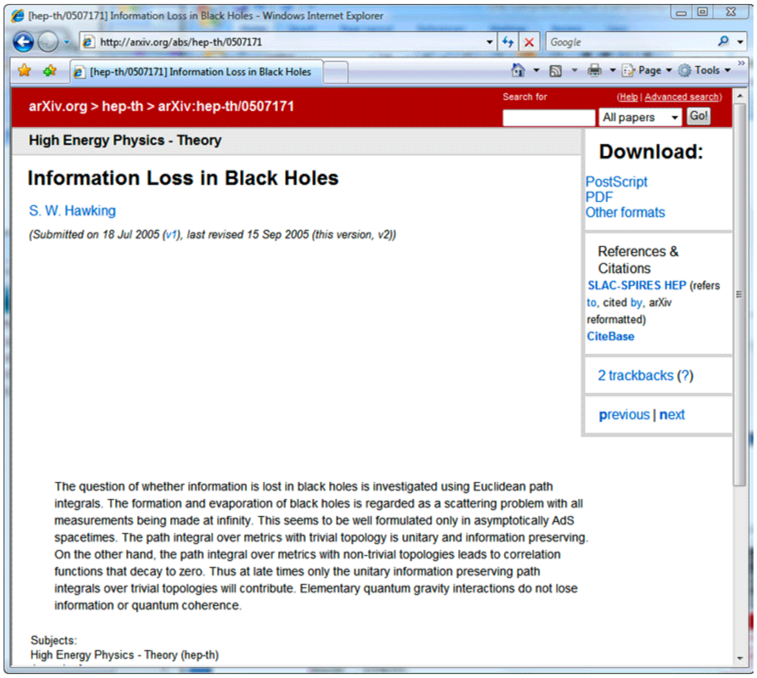

7, 8, 9 etc. are URIs identifying web resources.
8
4
5
9
7
6
web graph without compound object
identifier
PDF
PDF
HTML splash
MP3
publish
publish
publish
1
8
4
5
2
9
S
7
6
web graph with published compound object
Components of the compound object are published as follows: one PDF file is published at URI 1; the other at URI 2; the HTML splash page is published at URI S; the MP3 file is not published.
The web graph does not reveal that resources with URIs 1, 2, and S correspond to a common compound object.

Figure 3. Publishing a compound object to the web.

When the components of a compound object are published as resources on the web, they may link to each other (e.g., S links to 2 in Figure 3), they may link to other resources (e.g., 1 links to 8 in Figure 3), and other resources may link to them (e.g., 9 links to 2 in Figure3). These links are the basis of the rich information environment that is the web. But, because they are generally *un-typed* (they are standard hyperlinks), or their types do not conform to any general standard, the links do not define the boundary relationship that exists among the resources that are components of a compound object (Figure 3). The logical whole that is the compound object disintegrates into a set of distinct resources that are indistinguishable from the other resource in the web graph.

Many information systems address this problem by expressing the compound object via a user-oriented html "splash" or "jump-off" page that lists links to all components of the compound object and to a variety of related resources. This is illustrated by Figure 4, where a splash page in the arXiv provides access to the various formats in which a document is available and also to external resources, such as citations. This splash page resource is also shown in Figure 3 as resource S.


[hep-th/0507171] Information Loss in Black Holes - Windows Internet Explorer
http://arxiv.org/abs/hep-th/0507171
Google
[hep-th/0507171] Information Loss in Black Holes
Page
Tools
Search for
(Help | Advanced search)
All papers
Go!

arXiv.org > hep-th > arXiv:hep-th/0507171

High Energy Physics - Theory

Information Loss in Black Holes

S. W. Hawking

(Submitted on 18 Jul 2005 (v1), last revised 15 Sep 2005 (this version, v2))

Download:

PostScript
PDF
Other formats

References & Citations
SLAC-SPIRES HEP (refers to, cited by, arXiv reformatted)
CiteBase

2 trackbacks (?)

previous | next

The question of whether information is lost in black holes is investigated using Euclidean path integrals. The formation and evaporation of black holes is regarded as a scattering problem with all measurements being made at infinity. This seems to be well formulated only in asymptotically AdS spacetimes. The path integral over metrics with trivial topology is unitary and information preserving. On the other hand, the path integral over metrics with non-trivial topologies leads to correlation functions that decay to zero. Thus at late times only the unitary information preserving path integrals over trivial topologies will contribute. Elementary quantum gravity interactions do not lose information or quantum coherence.

Subjects:
High Energy Physics - Theory (hep-th)

Figure 4. Splash page for an arXiv document (arxiv.org/abs/hep-th/0507171).

These splash pages have come to de facto represent the compound object "as a whole" on the web, and, as a result, a convention has emerged to use the URI of the splash page as the URI of the compound object itself. While this approach is useful for human users, it is problematic from a machine re-use perspective because:

- Machine interpretation of splash page information is difficult or impossible due to the lack of standards for its structure; and
- According to the Web Architecture, the URI of the splash page merely identifies the resource that is the splash page, which is actually only a component of the compound object and not the compound object itself.

In addition to these problems related to identifying the compound object and defining its structure, the web architecture has no method to explicitly reference a resource within the context of a compound object. This functionality is important for scholarly communication because an existing resource can be re-used as a component of any number of compound objects. For example, a specific cellular image may be part of one image web illustrating confocal microscopy techniques and of another concerned with cancer therapy. For provenance tracking and citation, it is important to have the ability to reference a resource as it exists as a component of one or other specific compound object (e.g., the cellular image as a part of the cancer therapy image web), because the exact meaning of the resource or of a reference to it can be dependent on the context provided by such object.

The goal of the ORE work is to address this shortcoming and define an infrastructure layer over the web architecture allowing interoperable use and re-use of and reference to compound information objects across a variety of networked applications. This ORE layer explicitly does *not* replace or redefine any core web architecture concepts. Indeed, it leverages them fully as part of the solution to the problem of expressing compound objects.

The ORE layer expresses the boundaries of a compound object in a manner that can be processed by machines and agents. This can be viewed as the specification of a machine-readable splash page that lists the components of a compound object, as well as their internal and, optionally, external relationships. Such a specification would support:

- Re-use of a compound object and its components across web-savvy applications.
- Reference to a compound object and its components in a manner that supports an understanding of their "compound object context." This would allow reference to a resource as it exists in the context of a specific compound object, distinguishing that from a reference to the same resource as it exists in the context of another compound object, or as just a resource in its own right.
- Machine discovery of compound object information.

4. Towards a Solution for Compound Information Objects

The work to date on the formulation of an ORE interoperability layer over the web architecture consists of three components:

- Using Resource Maps for describing compound objects;
- Referencing compound object resources; and
- Discovering Resource Maps.

Similar to other efforts in the scholarly community to develop interoperable, machine-readable representations of research artifacts, this ORE work is inspired by activities in the semantic web community. Two notable influences are Linked Data [11] and named graphs. We note however that the semantic web influence on ORE work does not mandate an implementation built on semantic web technologies such as RDF, RDFS, RDF/XML, and OWL. Rather, OAI-ORE might employ more lightweight technologies that are easier to adopt and that can be used in applications that typically do not require explicit semantic information (e.g., existing popular search engines, blogs). These simpler formats could then be transformed by mechanisms such as GRDDL [14] to extract data for more complex semantic applications.

4.1 Using Resource Maps to Describe Compound Objects

As explained above, a major issue with representing compound objects on the web is the loss of the notion of the logical whole because the web architecture has no standardized facility for representing aggregations of resources. As such, a major focus of OAI-ORE is to add this missing logical boundary information and to do so in a machine-readable manner.

We are examining the use of named graphs [12],[13] as a model for expressing this boundary information. A named graph is a set of nodes and arcs. In this context, a node is a URI-identified web resource, and an arc is a directional link between two such resources typed by a URI that indicates a relationship type. The named graph itself is identified by means of a URI. This URI-identification means that the named graph itself is a logical unit and is an addressable resource on the web. As such, named graphs provide a mechanism for describing the basic relationship between a compound object and its components that is missing when a compound object is published to the web. In addition, named graphs provide a means of associating a proxy URI identity (the URI of the named graph) with the compound object. Finally, in addition to expressing this basic boundary-type information, these graphs can express semantically richer information because they may contain arbitrarily typed resources (nodes) and relationships (arcs) that, for example, meet the requirements of a specific application domain.

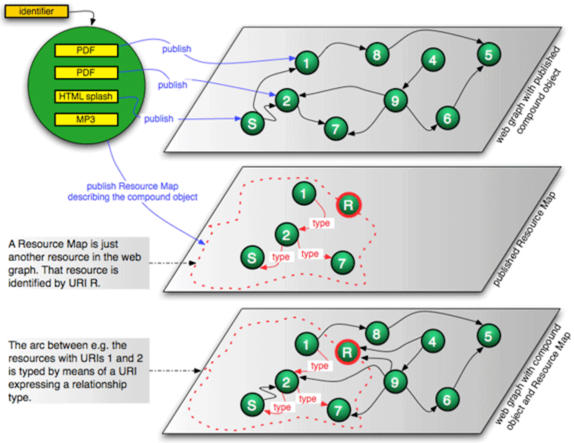

identifier
PDF
PDF
HTML splash
MP3
publish
publish
publish
1
8
4
5
2
9
S
7
6
web graph with published compound object
publish Resource Map describing the compound object
A Resource Map is just another resource in the web graph. That resource is identified by URI R.
R
type
type
type
published Resource Map
The arc between e.g. the resources with URIs 1 and 2 is typed by means of a URI expressing a relationship type.
web graph with compound object and Resource Map

Figure 5. Publishing a Resource Map to expose logical boundaries in the web graph.

The ORE interoperability layer intends to leverage named graphs by publishing *Resource Maps* that describe compound objects. A Resource Map is a named graph in which the nodes are resources that correspond with a compound object and its components, as well as resources that are related to these (e.g., the citations of a scholarly paper). A Resource Map must unambiguously distinguish between those “internal” and “external” resources. The arcs of a Resource Map are typed relationships between those resources. We envision a core relationship ontology and the ability to extend this core with discipline-specific ontologies. For practical reasons, a Resource Map is identified by means of a protocol-based URI (e.g., HTTP URI). This makes it possible to obtain machine-readable representations of the Resource Map through content negotiation. As a result, a Resource Map is considered an information resource as per [15].

Published Resource Maps overlay the web graph and effectively become part of (are merged into) it. This is illustrated in Figure 5, where the top pane shows the web graph without the information contained in the Resource Map and the bottom pane illustrates how the boundary of the object and relationships among the components are now visible in the web graph. The URI of the Resource Map (R in Figure 5) provides a web-based handle to the aggregate of multiple resources and their inter-relationships in the Resource Map. This URI can be referenced by standard web applications.

4.2 Referencing Compound Object Resources

Re-use of a web resource depends on the ability to reference it. As explained earlier, referencing issues exist when publishing compound objects to the web. First, there exists no web-parallel to the identifier of the compound object, shown in Figure 1. Second, there is a need to reference a specific resource not just in its own right (i.e., by means of its URI), but in the manner that it appears in the context of a certain compound object.

Referencing the Compound Object as a Whole

The protocol-based URI of the Resource Map identifies an aggregation of resources (components of a compound object) and their boundary-type inter-relationships. While this URI is clearly not the identifier of the compound object itself, it does provide an access point to the Resource Map and its representations that list all the resources of the compound object. For many practical purposes, this protocol-based URI may be a handy mechanism to reference the compound object because of the tight dependency of the visibility of the compound object in web space on the Resource Map (i.e., in ORE terms, a compound object exists in web space *if and only if* there exists a Resource Map describing it).

We note, however, two subtle points regarding the use of the URI of the Resource Map to reference the compound object. First, doing so is inconsistent with the web architecture and URI guidelines that are explicit in their suggestion that a URI should identify a single resource. Strictly interpreted, then, the use of the URI of the Resource Map to identify both the Resource Map and the compound object that it describes is incorrect. Second, some existing information systems already use dedicated URIs for the identification of compound information objects "as a whole." For example, many scholarly publishers use DOIs,[16] whereas the Fedora [17] and aDORe [18] repositories have adopted identifiers of the info URI scheme.[19] These identifiers are explicitly distinct from the URI of the Resource Map.

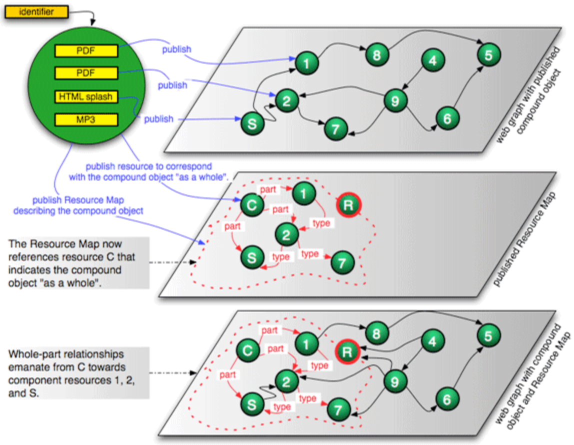

identifier
PDF
PDF
HTML splash
MP3
publish
publish
publish
1
8
4
5
2
9
S
7
6
web graph with published compound object
publish resource to correspond with the compound object "as a whole".
publish Resource Map describing the compound object
The Resource Map now references resource C that indicates the compound object "as a whole".
C
part
part
part
type
type
type
R
published Resource Map
Whole-part relationships emanate from C towards component resources 1, 2, and S.
web graph with compound object and Resource Map

Figure 6. Publish a Resource Map with a resource that indicates the compound object "as a whole."

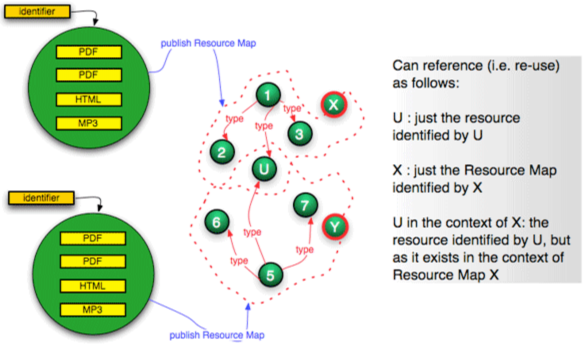
These issues suggest that it should be possible to express in the ORE specifications, and therefore in the Resource Map and its representations, an additional URI – the identifier of the compound object itself. Once a URI to identify the compound object "as a whole" is introduced, it would play the prominent role in the Resource Map of identifying the resource that corresponds with the compound object and that has component parts (resource C in Figure 6).

Referencing Resources in Context

On the web, resources can unambiguously be referenced by means of their URI. As a result, each published component of a compound object, as well as a published Resource Map describing a compound object, can be referenced. As mentioned in the previous section, a compound object "as a whole" can be referenced if it is assigned a dedicated URI. However, as explained earlier by means of the cellular image example, there is often the need in scholarly communication to reference a resource as a component of a specific compound object.

Figure 7 illustrates a scenario in which resource U is used as part of two compound objects. To reveal boundary information regarding the upper compound object, Resource Map X is published; Resource Map Y is published to do the same for the lower compound object. Because U is part of both compound objects, both

Resource Maps X and Y reference resource U. In order to accommodate the need to reference U as part of a specific compound object, OAI-ORE proposes to use an identifier pair that consists of the identifier of the resource itself and the identifier of the Resource Map that corresponds with the desired compound object, i.e. (U,X) to reference resource U as part of the upper compound object and (U,Y) to reference it as part of the lower compound object.



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

U : just the resource identified by U

X : just the Resource Map identified by X

U in the context of X: the resource identified by U, but as it exists in the context of Resource Map X

Figure 7. Resource U used in two compound objects.

4.3 Discovering Compound Objects on the Web

Exposing compound objects on the web via Resource Maps is only part of the solution; the Resource Maps and its referenced resources need to be discovered to really become part of the web graph. OAI-ORE proposes two complementing approaches with this regard:

- Harvest type discovery, which consists of making batches of Resource Maps available through existing mechanisms such as RSS, Sitemaps, and OAI-PMH.
- Linked Data [11] type discovery, which uses HTTP headers received in response to dereferencing the URI of a component of a compound object to point at the Resource Map(s) that correspond(s) with the compound object. This is shown in Figure 8 where a crawler lands upon splash page S and is pointed at the Resource Map R by means of a HTTP LINK header contained in the response to an HTTP GET request issued against S. From R, the crawler can obtain a Resource Map representation, and hence a list of all resources that are part of the compound object, as well as further internal and external relationships.

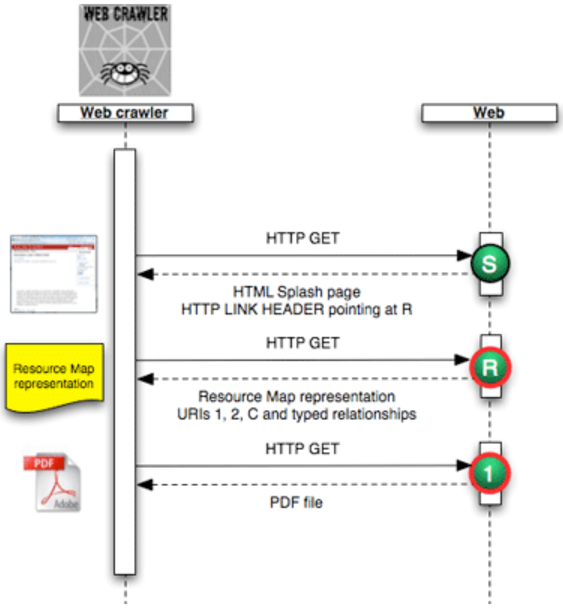

WEB CRAWLER
Web crawler
Web
HTTP GET
S
HTML Splash page
HTTP LINK HEADER pointing at R
HTTP GET
R
Resource Map representation
Resource Map representation
URIs 1, 2, C and typed relationships
HTTP GET
1
PDF
PDF file

Figure 8. A web crawler discovering the Resource Map via HTTP LINK HEADER.

5. Conclusion

Compound information objects are becoming the norm rather than the exception in the new scholarly communication environment. As a result, it is essential to augment the existing technical communication infrastructure with an interoperable approach that allows using, re-using, referencing, and discovering them across the borders of scholarly disciplines and applications. The international OAI-ORE effort works towards a solution that fully leverages the web architecture and that consists of publishing Resource Maps that describe compound objects, referencing resources in their compound object context, and mechanisms to facilitate discovery of Resource Maps.

Although OAI-ORE has made significant conceptual progress since it started in September 2006, important questions remain unanswered. How will the solution deal with versioning? How can the trustworthiness of Resource Maps be assessed? Which kinds of relationship types should OAI-ORE define to support bootstrapping adoption, and which should be left to individual communities? Which technologies should be used to represent Resource Maps, and how does a choice affect potential adoption? Some of these questions will receive at least a preliminary answer by the end of September 2007, which is the deadline that OAI-ORE has set itself for the release of a public alpha specification. Following that release, OAI-ORE will encourage experimentation by various scholarly communities and solicit feedback from potential stakeholders worldwide. The insights gained from those activities will be taken into account for a version 1 specification that is planned for September 2008.

Appendix

In the course of May 2007, the Digital Library Research & Prototyping Team of the Los Alamos Laboratory launched an experiment to explore the notion of Resource Map publishing as a means to expose compound object boundary-type information to the web. More particularly, the experiment explored whether an existing web application would be able to take advantage of published Resource Maps, without requiring any modifications to the application itself. The experiment pertained to archiving compound information objects as they evolve over time and the applications that were used were the Internet Archive's Heritrix toolkit that contains a web crawler and its Wayback Machine user interface.

The experiment's optimistic scenario assumes that Resource Map publishing has become so commonplace that the Internet Archive starts to actively collect them. The experiment zooms in on two publishers that make Resource Maps discoverable via dedicated Sitemaps. When a Resource Map listed in a SiteMap changes, its associated Sitemap date-time is changed. When a new Resource Map is published, it is added to the SiteMap. The Internet Archive uses these Sitemaps and their contained date-times as a trigger to collect and archive Resource Maps as well as the resources they reference. As a result, the Wayback Machine now allows searching for a specific Resource Map of a specific date and for immediately seeing the version of the resources referenced by that Resource Map as they existed on that same date. Understanding that Resource Maps expose the boundaries of compound objects, the net result is in effect an archive of evolving compound objects, versioned by the date-time of the Resource Map that describes them.

The screencast below shows a walk-through of the various components involved in the experiment and follows the evolution of some Resource Maps over time.

Compound Information Object
Archive Prototype Demonstration

Herbert Van de Sompel
Los Alamos National Laboratory

Acknowledgments

OAI-ORE is supported by the Andrew W. Mellon Foundation, the Coalition for Networked Information, Microsoft, and the National Science Foundation (IIS-0430906).
The authors acknowledge the contributions to the OAI-ORE effort from the ORE Technical Committee, Liaison Group and Advisory Committee. The authors also acknowledge the contributions of John Erickson (HP Labs) and Sandy Payette (Cornell Information Science).
Many thanks to Lyudmila Balakireva, Ryan Chute, Stephan Dresher, and Zhiwu Xie of the Digital Library Research & Prototyping Team of the Los Alamos Laboratory for their work on the prototype described in the Appendix.

[1] Van de Sompel, H., Payette, S., Erickson, J., Lagoze, C., Warner, S. "Rethinking Scholarly Communication: Building the System that Scholars Deserve," D-Lib Magazine, September 2004.

[2] Roosendaal, H. E., Guerts, P. A. T. M. "Forces and functions in scientific communities: an analysis of their interplay," in CRISP 97: Cooperative Research Information Systems in Physics, Oldenburg, Germany, 1997.

[3] National Science Foundation Cyberinfrastructure Panel, "Cyberinfrastructure Vision for 21st Century Discovery," National Science Foundation, Washington, D.C. 2007, http://www.nsf.gov/od/oci/CI_Vision_March07.pdf.

[4] "ImageWeb server," http://imageweb.zoo.ox.ac.uk/. Accessed June 29, 2007.

[5] Crane, G. "What Do you Do with a Million Books?," D-Lib Magazine, Vol. 12, March 2006.

[6] Razum, M. "eSciDoc - A Scholarly Information and Communication Platform in the Age," in Digital Library Goes e-Science (DLSci06), Alicante, Spain, 2006.

[7] Dmitriev, P., Lagoze, C., Suchkov, B. "As We May Perceive: Inferring Logical Documents from Hypertext," in HT 2005 - Sixteenth ACM Conference on Hypertext and Hypermedia, Salzburg, Austria, 2005.

[8] Lagoze, C., Krafft, D., Cornwell, T., Eckstrom, D., Jesuroga, S., Wilper, C. "Representing Contextualized Information in the NSDL," in ECDL2006, Alicante, Spain, 2006.

[9] Berners-Lee, T. "Semantic Web Road Map," W3C, http://www.w3.org/DesignIssues/Semantic.html.

[10] Jacobs, I., Walsh, N. "Architecture of the World Wide Web," W3C, Proposed Recommendation April 2004, http://www.w3.org/TR/2004/PR-webarch-20041105/.

[11] Berners-Lee, T. "Linked Data," W3C 2006, http://www.w3.org/DesignIssues/LinkedData.html.

[12] Carroll, J. J., Bizer, C., Hayes, P., Stickler, P. "Named Graphs, Provenance and Trust," in WWW 2005 Chiba, Japan: ACM, 2005.

[13] Carroll, J. J., Bizer, C., Hayes, P., Stickler, P. "Named Graphs," 2005, http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/NamedGraphs-WebSemanticsJournal.pdf.

[14] Davis, I. "GRDDL," W3C October 2006, http://www.w3.org/TR/grddl-primer/.

[15] R. Lewis, "Dereferencing HTTP URIs " W3C, http://www.w3.org/2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14.

[16] "The Digital Object Identifier System Home Page," International DOI Foundation (IDF), http://www.doi.org/.

[17] Lagoze, C., Payette, S., Shin, E., Wilper, C. "Fedora: An Architecture for Complex Objects and their Relationships," International Journal of Digital Libraries, Vol. 6, pp. 124-138, April 2005.

[18] Van de Sompel, H., Bekaert, J., Liu, X., Balakireva, L., Schwander, T. "aDORe: a modular, standard-based Digital Object Repository," http://www.arxiv.org/abs/cs.DL/0502028.

[19] Van de Sompel, H., Hammond, T., Neylon, E., Weibel, S. "The "info" URI Scheme for Information Assets with Identifiers in Public Namespaces," *IETF RFC 4452*, 2006, http://www.rfc-editor.org/rfc/rfc4452.txt.

URL to article: http://www.ctwatch.org/quarterly/articles/2007/08/interoperability-for-the-discovery-use-and-re-use-of-units-of-scholarly-communication/