Enhanced E-theses Project

Deliverable 10

Research roadmap: interoperability – Review of ORE HTTP implementation paper in the context of compound ETDs

Document Description

Project	
Title:	Enhanced E-theses
Start date:	1st December 2007
Funding Agency	Knowledge Exchange
Document	
Deliverable number:	D10
Deliverable title:	Research roadmap: interoperability – Review of ORE HTTP implementation paper in the context of compound ETDs
Actual Date of Delivery:	January 2009
Author:	Chris Awre (University of Hull)
Editor(s):	Keith Russell (Knowledge Exchange)
Workpackage:	WP10
Workpackage title:	
Workpackage leader:	All partners
Version/Revision:	1.1
Draft/Final:	Final
Total number of pages: (including cover)	7
File name:	Enhanced Etheses project_deliverable 10_3 v1.1.doc

This work is made available under a Creative Commons attribution 3.0 licence. For details please see http://creativecommons.org/licenses/by/3.0/

Section 1: Interoperability for compound ETDs

Interoperability

In order to consider the interoperability of compound ETDs it is first important to understand what is meant by interoperability, and for what purpose that interoperability is required. At the heart of interoperability, though, is the need for the use of standard ways of working. This may encompass technical standards for the structuring of the compound ETD through to implementation of standard processes for creating, capturing, sharing and/or preserving the compound ETD. The proposed use of ORE is an example of this, making use of this specification to provide a standard way of describing compound ETDs, and this paper looks at some of the issues relating to this and how they might be taken forward beyond the current project.

Interoperability can take place and be useful on different levels:

- End-user interoperability: this refers to establishing links between users and the services they need. This is commonly achieved by Web-based technologies and frameworks but might also include integration with end-user applications and desktop systems;
- Technical interoperability: this refers to the integration of services and computational resources with other services and also with end-user applications. This integration is usually achieved through network-facing interfaces, adapters and APIs but might also include more tight coupling;
- Institutional interoperability: there is an increasing need for institutional systems to interact with a wide range of resources, such as digital libraries, Virtual Learning and Research Environments, and security infrastructures. Sharing of data, information and knowledge is essential;
- International interoperability: when working across different countries, there are variations in standards, technical infrastructures, communication styles and practices. The creation of virtual organisations and communities could help resolve some of these differences.

Interoperability and compound ETDs

Placing compound ETDs in the context of the different levels of interoperability provides guidance for how compound ETDs might be shared most effectively.

- End-user interoperability: end-users need to be aware of services that can provide access to compound ETDs. These may include specialised portals such as the EThOS service in the UK or the DART-Europe gateway, or other resource discovery services on the Web. The latter is particularly relevant in the light of the potential use of ORE, as ORE aims at making the ETD

Aggregations discoverable through their identification on the Web and availability through the Resource Map serialisations. Where resource discovery tools are able to intelligently discover and interpret the Resource Maps, the compound ETDs will be able to be released into the Web as self-defining entities. Whilst such generic Web access is also currently available, this is usually limited to individual text files that cannot indicate how they might be related to associate files. Specialised portals are likely to remain of value to draw end-users in, but Web access is likely to be more fulfilling.

- Technical interoperability: the use of ORE is central to facilitating technical interoperability, though will need associated tools and services to be conformant with the specification to ensure it can be used for such a purpose. There have been attempts in 2008 to encourage experiments with technical interoperability through two competitions (see below for details), and it is hoped that such initiatives will prove fruitful in the long-term. The use of other metadata and packaging standards to facilitate the description and structuring of compound ETDs will also assist in parallel.
- Institutional interoperability: the institutional environment is perhaps not so immediately relevant to compound ETD interoperability. However, systems used for creating compound ETDs may draw on information and content from different systems, and the institutional management and preservation of the compound ETD may also involve multiple systems. Ensuring that these interoperate successfully will ease the processes involved in managing compound ETDs and facilitate other levels of interoperability
- International interoperability: the aim of this project has been, in part, to investigate how compound ETDs might be made interoperable across different countries within Europe. Agreement on the adoption of practices and standards across these countries will facilitate this level of interoperability. This project is a step towards informing this.

Section 2: Review of ORE HTTP implementation paper

This section of the report takes a focused look at one aspect of technical interoperability for compound ETDs, to support their discovery. It reviews the ORE paper on HTTP implementation that examines how best to make use of the standard protocol used for Web communication to ensure the most successful use of ORE in describing resources on the Web. This paper does not explain the ORE protocol itself, for which other documents made available by the Open Archives Initiative are better suited.

Scope of the paper

The paper is a user guide on how to make use of HTTP to facilitate the identification of ORE Aggregations and Resource Maps on the Web. ORE is not bound to the use of HTTP in the way that the OAI-PMH protocol is, but, recognising the very wide use of

HTTP by commonly used browsers, it is essential that ORE-described resources can make effective use of it to maximise exposure and access to the ORE Aggregations.

Within ORE there are two distinct entities that are generated aside from the resources being described themselves. The Aggregation is at the heart of the ORE specification, and is what specifies what is linked to what. The Resource Map describes the Aggregation. The Aggregation has its own URI for identification on the Web, but it does not have its own representation, a description of its make-up. Anyone accessing the Aggregation URI needs to be re-directed to the Resource Map describing that aggregation, which will have its own URI. Where there is more than one Resource Map describing an Aggregation, each having its own serialisation (e.g., ATOM, RDFa, etc.), each Resource Map will have its own URI: the client making the request should specify a MIME-type preference for which serialisation to have so that the correct Resource Map is presented.

The paper considers these issues and examines different circumstances and use cases in the use of HTTP that may arise to enable a correct link between the Aggregation and Resource Map. At the centre of the approach taken is the use of the 303 re-direction capabilities within HTTP to bounce the requesting server/user from the Aggregation to the Resource Map. Where this is not available a method using hash URIs is proposed to enable the link, though this will only work where there is a single Resource Map for the Aggregation.

The boundary between machine and human access – splash pages

The ORE HTTP paper makes clear that splash pages on the Web, where end-users will most often arrive at links to the resources within an Aggregation, are not the same as an Aggregation: they cannot be identified using a URI (the splash page URI is for that web page, not for the Aggregation) and they do not by themselves provide any information on how the different resources are related to each other (even if human interpretation might reveal this). However, the wide use of splash pages does require that there is an identified relationship between them and ORE Aggregations.

The 303 re-direction used to link an Aggregation to its Resource Map(s) can also be used to re-direct a request from an Aggregation to the splash page, where the client expresses a preference for HTML over, for example, ATOM as the preferred MIME-type. The ORE paper emphasises that where no client preference is expressed the default re-direction should be to a Resource Map rather than a splash page: this stresses the primary machine access role for ORE.

The splash page, being an HTML page and a resource in its own right, may, of course, be included within the Aggregation itself, and the decision on this is left to the implementer.

The boundary between machine and human access to information about an ORE Aggregation can be blurred by embedding the Resource Map within the splash page. Where the Resource Map is serialised as RDFa, this can be embedded within a splash page where this is structured as XHTML. The default Resource Map re-direction from the Aggregation can, in these circumstances, make available both machine and human-readable information at the same time.

Proxy URIs

Resources within an ORE Aggregation don't necessarily just belong to that Aggregation, e.g., an image may belong to two different collections, each described by a separate Aggregation and its associate Resource Maps. In order to allow a resource to be identified in the context of a specific Aggregation, a Proxy for that resource can be created that has its own URI and can be identified separately.

When a Proxy URI is dereferenced, it will identify the aggregated resource. It can also, though provide information on the Aggregation that the resource is part of. Identifying the Aggregation will lead to the Resource Maps and additional information about the resource, the Proxy and the context in which they are placed on the Web.

Section 3: Compound ETD interoperability and HTTP re-direction

The processes described in the previous section do not affect compound ETDs per se, as the use of HTTP applies to all resources described by ORE. However, the technical use of ORE in the ways described may have an impact on the end-user and other levels of interoperability of compound ETDs. The impact on compound ETDs will relate to the use case at hand.

- Where access by a machine is required, then the use of a Resource Map in the most useful representation, as specified by the client request, will enable flexible access for subsequent processing. It is advisable that ETD Aggregations make available a variety of Resource Map serialisations in this instance, guided by community best practice and adoption, to maximise this flexibility. Which serialisations are most useful will also be guided by future developments of services and applications that require, and can make the best use of, the Resource Maps. However, the HTTP re-direction makes access to compound ETD aggregations over the Web for subsequent processing relatively straightforward.
- Where access by end-users is required, the ability to use the Aggregation URI as a step-off point to the splash page means that a single URI can be issued for all requirements, human and machine-based. Where the splash page presented contains the Resource Map as well, intelligent clients able to interpret this can provide richer information about the compound ETD, a feat that would also be possible if an ATOM Resource Map were also available

using ATOM-compliant readers (which, of course, include modern Web browsers).

As indicated in the discussion on interoperability at the start of this paper, the potential of ORE is to enable richer discoverability of resources and how they are related via the Web. Where clients are able to interpret the information presented to them when retrieving Resource Maps these offer the potential to greatly facilitate and increase the use of general Web search engines to discover ETDs. The role of specific ETD portals and search engines will have organisational and other end-user interoperability purposes, but they will need to take note of this increased capability and focus their services accordingly to best meet user needs. For example, where a Web search engine will facilitate discovery through search, an ETD portal may be better placed to make use of the information within the Resource Map to provide richer access and functionality.

The use of Proxies and their associated URIs is of high potential benefit to compound ETDs. It allows resources to be referenced by an ETD that may not be exclusively part of that ETD, but be a resource referenced within it, and which may be part of a separate Aggregation elsewhere. Using the Proxy URI will allow a clear reference to the resource in the context of the ETD. This advantage to incorporating external resources within a compound ETD may be outweighed by the additional work that may be required to include Proxies for all resources within Resource Map serialisations, and a balance reached on best and most pragmatic practice.