

ENHANCED SCIENTIFIC COMMUNICATION BY AGGREGATED PUBLICATIONS ENVIRONMENTS (ESCAPE)

Maarten van Bentum¹; Arjan Hogenaar²; Dennis Vierkant¹.

¹Library and Archive, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands
e-mail: m.vanbentum@utwente.nl; d.vierkant@utwente.nl

²Royal Netherlands Academy of Arts and Sciences
P.O. Box 19121, 1000 GC Amsterdam, The Netherlands
e-mail: arjan.hogenaar@bureau.knaw.nl

Abstract

The ESCAPE-project aims at extending the existing infrastructure of repositories of scientific publications in such a way that it will be possible to identify, describe, preserve and present aggregations of related objects (documents, videos, datasets, etc.), not necessarily produced by an individual author or group of authors.

To this end a repository for OAI-ORE resource maps will be developed as well as an editor tool for creating and changing resource maps. The repository will be based on Fedora 3.1, reusing its built-in RDF support.

The resource maps repository form the basis for the so-called Aggregated Publications Environments. These APE's provide a resource map editor and a tool for browsing and searching aggregated resources. APE's will be developed for three research groups involved in this project. For the description of content relations, besides the use of "Proxy" in OAI-ORE we introduce the "Relation Annotation" object. This is an object which describes the relationship between two resources.

Keywords: resource maps; RDF; OAI-ORE; scientific communication.

1. Introduction

Scientific communication concerns the formal and informal exchange of information about research among scientists. Not only the scientists but also the scientific publisher is an important actor in this process. Around and in relation with the

world of scientific communication, other media publish information about scientific research aimed at a variety of target groups like policy makers, companies and the general public. This information often has another form than the 'traditional' scientific publication, but its content is clearly related to it. In many cases it concerns application oriented publications, policy documents, guidelines, comments/reviews, but also non-specialist or newspaper documents, visual material, etc. In many cases this derived material is made by non-scientists and is not published by scientific publishers.

Improved access to the results of scientific research leads to increasing interest and production of derived material, both scientific and non-scientific. At the same time there will be a growing need to present this material on the web in connection with the original scientific publication [1]. A good presentation of original and derived material improves the quality of communication not only within the scientific community but also between science and society.

To meet the above mentioned need for 'enhanced' scientific communication, the ESCAPE-project has been initiated in order to extend the existing infrastructure of repositories of scientific publications in such a way that it will be possible to identify, describe, preserve and present aggregations of related objects (documents, videos, datasets, etc.), not necessarily produced by the same individual author or group of authors. The applications to be developed, the aggregated publications environments, can be seen as elements in the service layer of the OAI data/services model, without requiring complex changes in the mode of operation of traditional repositories to deliver objects as an item in an aggregation with other objects.

The ESCAPE project is a collaboration of libraries, in their role of data and service provider, and scientific research groups as the demanding parties in this project. The participants are two Dutch university libraries (University of Groningen and University of Twente), three research groups (Psychology and Physics of Fluids both from University of Twente and Law/University of Groningen) and the Royal Netherlands Academy of Arts and Sciences.

2. Methodology

The methodology of the project differs from the way in which complex objects are realized by means of XML-containers, like METS, MODS and MPEG21-DIDL. The ESCAPE-project is based on the web architecture principles [2], mainly using individual resources (objects) and unique identification of the resource by means of a URI (in most cases a URL). In FRBR terminology the core of description is at item

level [3]. Identified objects become part of a structure of related individual objects. These objects can be taken into different structures. The structures may be inspired by library principles (like in FRBR or in a more traditional metadata format like MARC 21), but can also be based on content based relations between documents. The kind of content relations depends on the scientific discipline.

The OAI-ORE model is completely compatible with the web architecture principles and is more flexible than the use of XML-containers. Moreover, it offers the possibility for third parties to choose their own way of structuring a group of related objects.

Another topic is the discovery of resource maps from the individual objects. OAI-ORE describes a few possible methods to realize resource map discovery; in this project we test the feasibility of these suggested methods.

Considerable attention is given to the dissemination of the acquired expertise and knowledge about the application of the tools developed in this project. Guidelines will be devised that make it easy for repository managers to make known if and how their holdings are part of a resource map. At the end of 2009 a working prototype will be delivered, which can be used in a wide variety of contexts and will be in use by the three research groups participating in this project.

3. Results

The results of the ESCAPE project are twofold: on the one hand, technical infrastructure elements based on resource maps as described in the OAI-ORE model [4] and on the other applications based on this infrastructure, the so-called aggregated publications environments (APE's) integrated in specific websites of scientific groups.

Resource Maps repository

In this project a repository has been developed especially for the archival of resource maps. Resource maps define a group of objects and determine the nature of the relations that exist between them. Additionally, resource maps offer access to these objects and can be edited by specified library personnel or scientists themselves. The repository is based on Fedora 3.1, reusing its built-in RDF support. In order to be useful to the end-users, the system needs to be able to describe various types of relations. For this purpose we investigate whether the existing vocabularies are suitable or new vocabularies need to be developed. The use of existing vocabularies is preferred in order to maximize interoperability with other systems.

The repository stores the individual entities of the OAI-ORE data model [5] and exposes these through OAI-ORE resource maps. These resource maps are harvestable making use of the OAI-PMH protocol [6]. They can be harvested by other parties, such as libraries with scientific publication repositories, and presented together with the concerned document. In this way it will be possible to offer a view of a network of content related objects, starting from a particular publication or information need. The repository also deals with issues such as generating OAI-ORE Proxy objects [7] and providing access to the resource map from the aggregation URI through content negotiation [8].

Resource Map editor

During the project a special editor will be developed to produce or change resource maps in a user friendly manner. The resource map editor is a tool for defining and describing relations between objects, by authorized professionals using standardized descriptions and by unauthorized users using free descriptions. Both types of descriptions can be viewed and edited separately.

Resource Map vocabularies

There are two kinds of relation descriptors. Formal relations like title, author, publisher, can be described using standard vocabularies such as DCMI Metadata Terms and FOAF. However, these vocabularies hardly consist of terms to describe content relations, for example relating a publication as an application of another publication. In general, relating the content of publication in a standardized manner is a complicated semantic issue. For example, the involved scientist groups have a rather different idea about the way they want to describe content relations between objects, depending on the discipline and objectives of these groups.

To solve the problem of standardizing the description of content relations, besides the use of "Proxy" in OAI-ORE [9] we introduce the "Relation Annotation" object. This is an object which describes the relation between two resources. If there is content relation between two resources this can be described by means of a standard like dcterms:relation. In order to explain this relation one can add a relation annotation object which refers to both resources. The relation annotation object makes the relation identifiable by means of a URI, this allows making further RDF statements to describe the content relation, such as dcterms:description.

Aggregated Publications Environments

The above mentioned results are the basis for the Aggregated Publications Environments to be developed. These APE's provide a resource map editor and a tool

for browsing and searching aggregated resources. The browse and search interface presents the relations using commonly accepted web User Interface concepts. In the ESCAPE project the APE's will be developed for the following cases:

- 1) Center for Risk and Safety Perception (CRiSP). An APE as part of the website CRiSP where other (scientific) publications will be made available in connection with additional content related documents about risk and safety perception;
- 2) Brandaris128. An APE as part of the website Brandaris128 where other results and derived material (like television programs) will be presented which originates from research performed with the high-speed camera Brandaris128;
- 3) Private law and notarial law. An APE for making content relations between verdict and jurisprudence.

4. Discussion

Perhaps the main relevant feature of OAI-ORE and the semantic web is that it forces information taxonomists, including librarians, to identify each and every object with a URI. Objects include not only documents, but also authors, institutes and even parts of objects as for example, sections in a scholarly paper. This means identification on the item level. Aggregations of items result from a description of these items and their interrelations. OAI-ORE allows such an aggregation to be handled as one identity using a single URI.

Adequate relation descriptions require unique identifiers of the aggregated resources (items, objects). However, the identification of a particular person, institution or even publication is not unique because in most cases it relates to different items (objects), in a different context (or aggregation). For instance, a person can be identified on the web in his private context (or aggregation of family members), or in his work context (or aggregation of colleagues). A URI of a person, institution or publication in fact refers to a proxy (in OAI-ORE terms) being the person, institution or publication in a certain context.

Some issues need to be solved in the identification of items. One issue is that in the case of identifying publications within one context, there may be different representations (such as pdf, doc, html) with possibly different access restrictions (publisher, journal or institutional repository); therefore, different items of the same publication each have a unique URI. If a single publication can have multiple different URIs, how can a meaningful connection then be made between two resource maps which describe aggregations containing the same publication but with a different URI?

Notes and References

- [1] Several scientist groups both at the University of Twente and the University of Groningen, The Netherlands, expressed this need. Some of these groups are involved in the ESCAPE-project.
- [2] DAVIS, IAN and RICHARD NEWMAN. *Expression of Core FRBR concepts in RDF*. 2005. Available at <http://vocab.org/frbr/core> (April 2009).
- [3] FRBR (Functional Requirements for Bibliographic Records) distinguishes different description levels:
 - work** creation of one or more persons
 - expression** a translation, version, realization, et cetera of a work
 - manifestation** refers to a particular form (print, electronic, cd, dvd, et cetera)
 - item** exemplar of a manifestation (specific book or cd in the library, an auto-graphed book, et cetera).
- [4] See ORE specifications: <http://www.openarchives.org/ore/>.
- [5] See <http://www.openarchives.org/ore/1.0/datamodel.html>.
- [6] See <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>.
- [7] See <http://www.openarchives.org/ore/1.0/http.html#Proxy>.
- [8] See <http://www.openarchives.org/ore/1.0/http.html#RedirectCN>.
- [9] See <http://www.openarchives.org/ore/1.0/datamodel.html#Proxies>.

June 2009
Printed on demand
by "*Nuova Cultura*"
www.nuovacultura.it

Book orders: ordini@nuovacultura.it