ECP-2005-CULT-038137

TOWARDS AN ACCESSIBLE SCIENCE: FACILITATING ACCESS TO SCIENTIFIC DIGITAL RESOURCES FOR VISUALLY IMPAIRED STUDENTS

D4.3 Tools to Enable Communication Among University Libraries and Distributed Digital Library Implementation

Deliverable number	*D4.3*
Dissemination level	*Public*
Delivery date	*7th November 2008*
Status	*Final*
Author(s)	*@Science Thematic Network Consortium*

eContent***plus***

This project is funded under the *e*Content*plus* programme[1],
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

1 OJ L 79, 24.3.2005, p. 1.

EXECUTIVE SUMMARY

This document describes some tools that can facilitate exchange of digital objects among digital libraries based on digital repositories. These tools are the Open Archives Initiative protocols and practices. In the last section the digital repository of the @Science Network is introduced.

TABLE OF CONTENTS

1 INTRODUCTION

Over the years university libraries have more and more undertaken projects to make available their digital contents through digital repositories. These repositories are an online, searchable, web-accessible collection of digital objects (in some cases, grouped in collections) which are developed in the academic context (e.g., works of research, learning objects such as lecture notes and slide presentations prepared by university professors for their courses, books made available in digital format, research papers, conference proceedings, journals, etc.).

These repositories aim both at increasing the access to scientific content (i.e., to facilitate reading of research papers), and also at making available to the public (e.g., to students, researchers, etc.) reusable study material.

Digital repositories are often built to serve a community of users in a certain institution (they are known as institutional repositories). In other circumstances, repositories are built to serve scholars in a specific discipline (e.g., the arXiv repository for physics, mathematics and theoretical computer science).

During these years, many institutional repositories and discipline oriented repositories have been built by using several different conventions for metadata schemata, protocols for exchange of digital objects, admissible formats for files to be archived in the repository and submission and distribution policies. Thanks to the evolution of the open archive model, in particular since 2002 (Budapest Open Access Initiative: http //www.soros.org/openaccess/read.shtml), standards for metadata and protocols for metadata harvesting have been developed, and they have been spreading more and more all over the World in the design and implementation of digital repositories. Compliance with these standards will ensure better communication between those university libraries (or, generally speaking, content providers) which will maintain a digital repository of scholarly material.

Section 2 will illustrate the basic open archives practices which can ensure better access, use, reuse and exploitation of scholarly content stored in digital repositories. Blind and partially sighted people who need to access scientific content will widely take advantage of the open access initiative practices. Some basic reasons account for that:

- Open access initiative is making available more and more scientific content in open formats, which are far more adaptable for assistive technology;
- Federated searches in multiple repositories give the opportunity to discover related resources, which may be available in different formats. Therefore, for a visually impaired person, it is more likely to find a resource in accessible format concerning a certain topic;
- Metadata harvesting by using the OAI-PMH protocol is independent of the user interface. Therefore, if the user interface to search, retrieve and display resources in repositories is not accessible for visually impaired, an accessible user interface which exploits OAI-PMH to harvest metadata can be developed by respecting W3C Web Content Accessibility Guidelines.

2 OPEN ARCHIVE PRACTICES

2.1 Introduction

The open access model has been developed a lot through the years. In 1990, the open access arXiv for physics papers was set up. In 1995, two remarkable open access initiatives were launched: WoPEc for economics, NCSTRL for computer science. In 1996, PubMed included Medline as an open access archive. In 2001, the Public Library of Science (PLOS) was founded. In 2002, the Budapest Open Access Initiative was issued. In 2003, the Directory of Open Access Journals was released. In 2006, the European Commission issued a report in order to recommend open access for public research. That same year, in the United States the Federal Research Public Access Act of 2006 was approved, which recommended open access in federal research, and, at CERN, SCOAP3 was launched to convert to open access the particle physics journals. In 2007, the European Research Advisory Board issued a recommendation in using the open access model for the European Union funded research. In addition, the European University Association recommended open access for collecting research results. The European Research Council issued the ERC Scientific Council Guidelines for Open Access which requires the deposit of papers and scientific data in open access repositories. In 2007, 427 scientific societies were publishing 496 scientific journals based on the open access model. All over the World, graduation works and Ph.D. theses are more and more being archived in open access repositories.

Since 2002, best practices and standards have also started being widespread, thus ensuring better access, use, reuse and exploitation of digital objects stored in repositories. These practices and protocols will be described in the following sections.

The open archives model is based on two classes of participants:

- Data Providers. Data providers archive resources in digital repositories, describe resources through metadata sets, and make available metadata to harvesters through the OAI-PMH protocol (Open Archives Initiative Protocol for Metadata Harvesting: http://www.openarchives.org/OAI/openarchivesprotocol.html). Data providers can be, for example, university libraries which grant access to their content in digital format in a repository;
- Service Providers. The Service Providers harvest metadata from the Data Providers through the OAI-PMH protocol, in order to make available value-added services. Examples of Service Providers are the search engines specialized for a certain knowledge domain.

2.2 OAI-PMH

OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting, http://www.openarchives.org/OAI/openarchivesprotocol.html) is the protocol which Data Providers and Service Providers must comply with in order to make available content and achieve metadata harvesting.

This section introduces briefly the characteristics of the protocol:

- OAI-PMH requests are expressed as HTTP requests (through HTTP GET or HTTP POST methods). A typical implementation uses a standard web server that is configured to dispatch OAI-PMH requests to the software handling these requests;
- All responses to OAI-PMH requests must be well-formed XML instance documents. Encoding of the XML must use the UTF-8 representation of Unicode. Character references, rather than entity references, must be used. Character references allow XML responses to be treated as stand-alone documents that can be manipulated without dependency on entity declarations external to the document;
- dates and times are uniformly encoded using ISO8601 and are expressed in UTC throughout the protocol;
- OAI-PMH supports the dissemination of records in multiple metadata formats from a repository. All metadata available from a repository can be required through the ListMetadataFormats request. For purposes of interoperability, repositories must disseminate Simple Dublin Core metadata. Further metadata formats can be added;
- OAI-PMH makes possible to retrieve lists of records, sets or identifiers through list requests.

2.3 Metadata

Metadata formats are groups of defined elements or fields that can be used to describe a resource (i.e., administrative information, technical information, etc.). Making metadata sharable is the basic characteristic a data provider should implement. The most significant advantage to creating shareable metadata is that the metadata will be interoperable or meaningful when combined with metadata from other sources. By using metadata schemata and rules for creating metadata values similar to those used by others, the resources collected in a certain repository can meaningfully appear in search results together with related resources from other repositories. Therefore, the resources will receive more exposure, and end-users will have the opportunity to make previously unseen connections between related resources stored in different repositories. Data providers do not need know in advance which repositories contain related resources: if metadata are sharable, the metadata harvesters will be able to find relations between resources.

Based on the book "Metadata in Practice" (ed. Diane I. Hillmann and Elaine L. Westbrooks, Chicago: American Library Association, 2004, Thomas R. Bruce and Diane I. Hillmann) seven characteristics must be satisfied to have good metadata:

- Completeness. That means, in particular, choosing an element set which allows the resources to be described as completely as is economically feasible, and applying that element set as completely as possible;
- Accuracy. The metadata must be correct and factual, and conforming to syntax of the element set in use;
- Provenance. Here provenance refers to the provision of information about the expertise of the persons creating the original metadata, and its transformation history;
- Conformance to expectations. Metadata elements, use of controlled vocabularies, and robustness should match the expectations of a particular community. Since the resource can be retrieved by metadata harvesters, the expectations are not often known in advance in any circumstance;
- Logical consistency and coherence. This characteristic is defined as element usage matching standard definitions, and consistent application of these elements;
- Timeliness. That refers in particular to keeping up-to-date metadata concerning resources which change over the time and making available metadata describing actually available resources;
- Accessibility. It refers to proper association of metadata with the resource it describes and readability by target users.

Metadata may be of high quality within a certain local context, not in a broader context. There are other characteristics which make quality metadata also sharable in a distributed environment:

- Proper context. In a shared environment, metadata records will become separated from any high-level context applying to all records in a group, and from other records presented together in a local environment. It is therefore essential that each record contains the context necessary for understanding the resource the record describes, without relying on outside information;
- Content coherence. Enough information must be contained in metadata records in order to retrieve resources even if outside the environment the resource belongs to;
- Use of standard vocabularies. The use of standard vocabularies enables the better integration of metadata records from one source with records from other sources;
- Consistency. All decisions made about application of elements, syntax of metadata values, and usage of controlled vocabularies, should be consistent within an identifiable set of metadata records so those using this metadata can apply any necessary transformation steps without having to process inconsistencies within such a set;
- Technical conformance. Metadata should conform to the specified XML schemata and should be properly encoded.

The Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) enables service providers to harvest metadata exposed by data providers. Metadata harvesting through OAI-PMH makes possible all the advantages of open access repositories, therefore, for a repository to be OAI-PMH compliant it is indispensable. To this purpose:

- a Simple Dublin Core metadata record (Dublin Core Metadata Element Set 1.1, http://dublincore.org/documents/dces/) must be available for every item in the repository. To this purpose, the Open Archives Initiative makes available an XML schema at:
  http://www.openarchives.org/OAI/2.0/oai_dc.xsd
  for simple Dublin Core, and reserves the metadata prefix oai_dc for this schema;
- In addition to the Simple Dublin Core record, it is advisable to provide alternative or richer metadata records in further metadata formats. Actually, Simple Dublin Core is not able to describe details of many resources, which could be present in a certain repository. Moreover, Simple Dublin Core does not convey information about the use of controlled vocabularies or encoding schemes used. By supplying additional metadata formats which have the semantic richness to more clearly express meaning, data providers can help service providers make better use of their metadata. Metadata formats used must have an XML schema available for validation (see http://www.openarchives.org/OAI/openarchivesprotocol.html#MetadataNamespaces).

Some of the most used metadata formats are listed below. For a complete list, consult the Distinct Metadata Schemas report maintained by University of Illinois at Urbana-Champaign (on the web at: http://gita.grainger.uiuc.edu/registry/ListSchemas.asp).

MODS: Metadata Object Description Standard

MODS may be a good choice for data providers which:

- Locally depend of metadata definition practices bound to resource description standards commonly used in libraries;
- Address to a community well-versed in library descriptive practices, yet also want robust records in a format accessible to service providers outside the core library community.

The MODS v.3.0 XML Schema is available at
http://www.loc.gov/standards/mods/v3/mods-3-0.xsd.

Qualified Dublin Core

Qualified Dublin Core may be a good choice for data providers which:

- Need for more granularity of description than that available in simple Dublin Core but not a fundamentally different approach to resource description as the library community may wish;
- Use controlled vocabularies to be specified within the metadata records;
- Address many different communities which may have their own metadata practices.

A XML schema for Qualified Dublin Core can be created through the importation of the necessary namespaces and schemas:

- http://dublincore.org/schemas/xmls/qdc/2003/04/02/dc.xsd
- http://dublincore.org/schemas/xmls/qdc/2003/04/02/dcterms.xsd
- http://dublincore.org/schemas/xmls/qdc/2003/04/02/dcmitype.xsd

MARCXML

MARCXML may be a good choice for data providers which:

- Locally describe resources in MARC (e.g. many libraries);
- Address mainly the library community.

The MARCXML XML Schema is available at:
http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd.

ETD-MS: Electronic Theses, and Dissertations Metadata Standard

ETD-MS may be a good choice for data providers which:

- Mainly archive theses (e.g. graduation or PhD theses) and dissertations in digital format;
- Aim at contributing to aggregations of electronic theses and dissertations such as the Networked Digital Library of Theses and Dissertations (on the web at: http://www.ndltd.org).

The ETD-MS XML Schema is available at:
http://www.ndltd.org/standards/metadata/etdms/1.0/etdms.xsd.

IMS Learning Resource Meta-data Specification

It may be a good choice for data providers which:

- Archive different sorts of learning objects;
- Aim at making learning objects available to the educational community.

The XML schema can be found at:
http://www.imsglobal.org/xsd/imsmd_v1p2p4.xsd

2.4 OAI Repositories of Learning Objects: Success Experiences

This section introduces some repositories of digital learning objects compliant with the Open Archives Initiative specifications.

Ariadne

Website: http://www.ariadne-eu.org

Ariadne is a European association for sharing knowledge and fostering international cooperation in teaching that is open to the World. Ariadne was initiated in 1996 by the European Commission's telematics for education and training program. Since then, an infrastructure has been developed in Belgium and Switzerland for the production of reusable learning content, including distributed storage and discovery, as well as its exploitation in structured courses. The core of this infrastructure is a distributed library of digital, reusable educational components called the Knowledge Pool System (KPS) now actively used in both academic and corporate contexts. End users interact with the KPS through client tools. Java and web applications allow users to insert documents and their associated metadata into the KPS, search for relevant documents, and download them from the KPS.

EducaNext

Website: www.educanext.org

EducaNext is a service supporting the creation and sharing of knowledge for higher education. It is open to any member of the academic or research community. EducaNext fosters collaboration among educators and researchers, allowing to participate in Knowledge Communities, to communicate with experts in a certain knowledge domain, to exchange Learning Resources, to work together on the production of Educational Material: Textbooks, lecture notes, case studies, simulations, etc., to deliver distributed Educational Activities: lectures, courses, workshops, case study discussions, etc., to distribute electronic content under license.

JEM network

Website: www.jem-thematic.net

The JEM network is a European thematic network for Joining Educational Mathematics. It started on August 1st, 2006. The goal of the JEM thematic network is to pool together the required expertise and to contribute to the coordination of content enrichment activities in the area of mathematics, to the maintenance of agreed standards and to the delivery of powerful synoptic high-quality user information and support pages, invoked in e-learning platforms operated by the partners. These activities will have tremendous long-term benefits for the quality of e-learning in mathematics. JEM network makes available a rich repository of digital learning objects about mathematics and science.

PlanetDR-URV

Website: http://planet.urv.es/planetdr/

PlanetDR is an open source Content repository developed under the umbrella of the Spanish research project PLANET (Augmented Collaborative Plataform for content distribution and advanced training). PlanetDR supports standards such as IEEE LOM, IMS Content Packaging (IMS CP), IMS DRI (EduSource Communication Language-ECL) and LORI SQI. PlanetDR also provides a federation mode between PlanetDR servers and plans to support the FIRE protocol.

SMETE

Website: http://www.smete.org/

The SMETE Digital Library is a dynamic online library and portal of services by the SMETE Open Federation for teachers and students. SMETE makes available a wealth of teaching and learning materials about science, math, engineering and technology which are suitable for many ages.

3 THE @SCIENCE REPOSITORY

3.1 Introduction

The @Science Thematic Network aims at making available through a digital repository:

- Learning objects about scientific university courses accessible to blind and partially sighted students;
- Learning objects about guidelines, best practices, case studies concerning how to enable blind and partially sighted students to access scientific subjects.

Furthermore, the @Science thematic network aims at collecting contributions from institutions willing to share their learning objects. In order to achieve these goals, a digital repository based on OAI practices and standards was integrated in the @Science website. This repository can archive learning objects made available by @Science participants, it can import learning objects collections from other repositories, it can expose metadata for harvesters through the OAI-PMH protocol. Since at the moment there is no automatic accessibility validator for many file formats which can be used to make available accessible scientific content, contributions, before being archived in the repository are checked by @Science members to match accessibility characteristics.

3.2 Repository Features

The following is a list of the basic features of the @Science repository:

- Learning objects are archived according to the metadata schema IMS LOM (IMS Learning Resource Meta-data Specification v1.3, available at: http://www.imsglobal.org/metadata/). This metadata specification turns out to be especially suitable for learning objects. In addition, it is widely used by many repositories making available learning objects (see section 2.4). In particular, it is used by the Joining Educational Mathematics (JEM) thematic network.

  This metadata specification makes available fields to specify intellectual property rights for learning objects. The @Science repository will make available not only free learning objects (i.e., learning objects which are licensed under Creative Commons share-alike license), but also learning objects which cannot be distributed to everyone because of copyright restrictions. This feature will be managed by comparing user profiles, namely user permissions, with the IPR metadata field for each learning object. For example, according to national transpositions of EUCD directive about copyright (see Deliverable 4.1) some special conditions apply for people with disability. If a registered member is registered as visually impaired and gives to one of the @Science members a legal declaration of disability (i.e., a copy of certificate of disability), some learning objects could be accessible and could be downloaded by this registered user, not by all users;

- OAI-PMH Data Provider support. The repository can expose metadata about learning objects to be harvested by OAI-PMH v.2 compliant metadata harvesters (e.g., OAIster: www.oaister.org/);
- Classification of the learning resources on basis of Living Taxonomy for Mathematics. The Living Taxonomy Project is a collaborative effort aimed at creating a global set of open source, standards-based taxonomies for education. The purpose of these taxonomies will be to provide a free cataloging structure for the collection and sharing of education materials around the world. (further information at: http://www.livingtaxonomy.org/);
- LOM records can be imported both from IMS LOM v1.2 and IEEE LOM v1.0 formats. That facilitates importing of already available learning objects which can be shared with other repositories;
- LOM records can be exported in LOM XML to be exposed to harvesters or to be imported in other repositories;
- Metadata records can be edited either through an accessible XHTML form or through a separate module;
- The system automatically detects the format of the record on export;
- Real time download counter on basis of exports, available at the node view page;
- Terms in the vocabularies are linked to views which show all the records that contain the same values in the field;
- Author names are linked to user profiles if the author is a registered member.

3.3 Implementation

The repository system integrated in the @Science website is the one used by the Joining Educational Mathematics (JEM) thematic network (available at: http://www.jem-thematic.net/en/node/716).

This repository is open source, it can be integrated in websites developed using Drupal 5.x, and it is based on PHP 5.1 or later.

This repository specifically addresses the needs of a community which works with learning objects of scientific content.