

# **SCHEMAS**

**Contract: N° IST-1999-10100**

**Forum for Metadata Schema Implementers**

**METADATA WATCH REPORT #7**

## **D28**

**Document number:**

**SCHEMAS-PwC-WP2-D28-Final-20011217**

<b>General Information</b>
----------------------------

<b>Title</b>	Metadata Watch Report #7
<b>Creator</b>	Makx Dekkers
<b>Contributors</b>	Laurie Causton, Annemieke de Jong, Erik Duval, Michael Day, Marieke Napier
<b>Subject-Keywords</b>	Deliverable D28; WP2; Metadata Watch Report #7
<b>Description</b>	This document comprises Metadata Watch Report #7 and domain reports for the Publishing, Educational, Audiovisual, and Cultural Heritage domains
<b>Publisher</b>	PricewaterhouseCoopers
<b>Date</b>	17 December 2001
<b>Type</b>	Text Manuscript
<b>Format</b>	application/MSWord2000
<b>Identifier-</b>	
<b>Document Number</b>	SCHEMAS-PwC-WP2-D28-Final-20011217
<b>Language</b>	English
<b>Rights</b>	European Commission; External distribution via SCHEMAS Web site

**Dublin Core metadata for this document**

```
<?xml version="1.0"?>
<rdf:RDF xml:lang="en"
xml ns: rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xml ns: dc="http://purl.org/dc/elements/1.1/"
<rdf:RDF rdf:resource = " ">
<dc: title> Metadata Watch Report #6 and Standards Framework Report #3
</dc: title>
<dc: creator> Makx Dekkers </dc: creator>
<dc: contributor> Laurie Causton </dc: contributor>
<dc: contributor> Annemieke de Jong </dc: contributor>
<dc: contributor> Erik Duval </dc: contributor>
<dc: contributor> Michael Day </dc: contributor>
<dc: contributor> Marijke Napier </dc: contributor>
<dc: subject> Deliverable D28 </dc: subject>
<dc: subject> WP2 </dc: subject>
<dc: subject> Metadata Watch Report #7 </dc: subject>
<dc: subject> Metadata activity reports </dc: subject>
<dc: description> This report comprises Metadata Watch Report #7 and domain
reports for the Publishing, Audiovisual, Educational, and Cultural Heritage
domains </dc: description>
<dc: publisher> PricewaterhouseCoopers </dc: publisher>
<dc: date> 2001-12-17 </dc: date>
<dc: type> Text </dc: type>
<dc: format> application/MSWord2000 </dc: format>
<dc: identifier> SCHEMAS-PwC-WP2-D28-Final-20011217 </dc: identifier>
<dc: language> en </dc: language>
<dc: rights> European Commission; External distribution via SCHEMAS Website
</dc: rights>
</rdf:RDF>
```

## Table of Contents

<b>1</b>	<b>INTRODUCTION.....</b>	<b>5</b>
<b>2</b>	<b>OBJECTIVE .....</b>	<b>5</b>
<b>3</b>	<b>TOP-LEVEL SYNTHESIS .....</b>	<b>5</b>
	<b>APPENDIX A. PUBLISHING DOMAIN.....</b>	<b>7</b>
<b>1</b>	<b>ACTIVITY REPORT .....</b>	<b>7</b>
<b>2</b>	<b>REGISTRY ACTIVITIES .....</b>	<b>8</b>
	<b>APPENDIX B. AUDIOVISUAL DOMAIN.....</b>	<b>9</b>
<b>1</b>	<b>INTRODUCTION.....</b>	<b>9</b>
<b>2</b>	<b>SOME EXAMPLES OF AUDIOVISUAL METADATA REGISTRIES.....</b>	<b>10</b>
	<b>APPENDIX C. EDUCATIONAL DOMAIN .....</b>	<b>13</b>
<b>1</b>	<b>METADATA REGISTRIES IN EDUCATION .....</b>	<b>13</b>
<b>2</b>	<b>STANDARDIZATION WORK.....</b>	<b>14</b>
<b>3</b>	<b>CONSORTIA BASED WORK.....</b>	<b>14</b>
<b>4</b>	<b>REFERENCES.....</b>	<b>16</b>
	<b>APPENDIX D: CULTURAL HERITAGE DOMAIN.....</b>	<b>17</b>
<b>1</b>	<b>CURRENT STATE OF DOMAIN .....</b>	<b>17</b>
<b>2</b>	<b>METADATA REGISTRIES IN THE CULTURAL HERITAGE DOMAIN</b>	<b>18</b>
<b>3</b>	<b>REFERENCES.....</b>	<b>20</b>

## 1 Introduction

This deliverable is the seventh Metadata Watch Report from the SCHEMAS project.

This report contains domain reports on the sectors Publishing, Audiovisual, Education, and Cultural Heritage.

## 2 Objective

As specified in the project objectives, the purpose of the SCHEMAS Metadata Watch (MD Watch) is to provide a quarterly overview of world-wide progress in the metadata field, which includes work on metadata sets, schemas, frameworks, registries, and the tools needed to create and use all of these.

This seventh metadata watch report, apart from continuing assessment of ongoing activities in the various domains, has a special focus on Registry activities.

## 3 Top-level synthesis

Any standardisation activity leads to a point where the stable definition of the standard needs to be published. In traditional standardisation, this happens usually in the form of a paper document that can be ordered from a national standards body or maintenance agency.

Metadata standards are no different in this respect. They are also published for a human audience to read and understand. Examples<sup>1</sup> are the standards for the library standard MARC21, the Dublin Core Metadata Element Set, the publishers' ONIX standard, or the US National Geographic Data Committee's Content Standard for Digital Geospatial Metadata, the IEEE Learning Object Metadata standard or the ISO standard ISO/IEC 11179 for specification and standardization of data elements. Many of these standards nowadays are also published as Web documents for easier reference and wider distribution.

Apart from publication for a human audience, many metadata standardisation activities now see a need for publication of the standard in machine-readable format. In metadata terminology this is commonly referred to as a *schema*. The underlying notion is that it would be useful for software processes to find the definition of the standard in a format that makes it possible to change their behaviour accordingly. Examples are: a metadata generation tool that can use different standard schemas and configure the user interface accordingly, or a metadata harvester that can configure its indexing mechanism on the basis of knowledge of the schema that some harvested metadata is based on.

A metadata registry is the mechanism to publish such schemas, which allows their discovery, interpretation and re-use.

Another emerging use of registries is for the publication of Application Profiles, as is the case for example in the SCHEMAS Registry.

---

<sup>1</sup> For more information about the standards mentioned, see the Results section in SCHEMAS Standards Framework Report #2, <http://www.schemas-forum.org/stds-framework/second/section5.html>.

There are two main types of Registry formats: structured text presentations and XML-encoded files. The XML-encoded format can either be in plain XML or in RDF.

There are a number of registry activities that provide some form of structured text to describe schemas and elements within schemas. Examples are:

- The DESIRE registry at UKOLN, see: <http://desire.ukoln.ac.uk/registry/>
- ISO/IEC 11179-based registries, such as the ones listed at <http://hmrha.hirs.osd.mil/mrc/>

Registries with machine-readable schemas come in one of two forms:

- XML-based registries, such as the Registry maintained by XML.org (see: <http://www.xml.org/xml/registry.jsp>) or the one under development by the OASIS/ebXML Registry Technical Committee (see: <http://www.oasis-open.org/committees/regrep/>).
- RDF-based registries, such as the Open Metadata Registry work in the Dublin Core Metadata Initiative (see: <http://wip.dublincore.org:8080/registry/Registry>) or the Open Metadata Registry of the University of Library and Information Sciences in Japan (see: <http://avalon.ulis.ac.jp/registry/>).

As illustrated in the domain reports, many activities are now starting to realise the usefulness of Registries from an understanding that implementations need to interoperate, and that schemas may be shared by applications. Many are still in the phase where a schema is described for humans to read and understand.

The Semantic Web, in effect an open environment where intelligent agents can encounter any type of metadata, the machine-readable variant of Registries is a crucial aspect for interoperability across applications and across domains. The World Wide Web Consortium's Semantic Web Activity recommends the use of RDF-based Registries because of the capabilities of RDF to express relationships between semantics. This is also the approach chosen for the Open Metadata Registry work in the Dublin Core Metadata Initiative, as this is fundamentally a cross-domain activity.

In more closed environments, where there is no immediate need to interoperate between partners that are not know a-priori, the need for declaring relationships between semantics is less important. Therefore, in those environments the pure XML approach is dominant, as demonstrated in the OASIS/ebXML work.

In the wake of the trend towards increased co-operation between standards activities, as identified in the previous Metadata Watch Report, there will also be an increased interest in the interoperability between Registries.

CORES, the follow-up project to SCHEMAS, which is currently under negotiation with the European Commission, aims to address some of these aspects.

## Appendix A. Publishing domain

*Correspondent: Laurie Causton, Clearbay Limited*

### 1 Activity Report

#### 1.1 Current state of domain

While not always directly related to metadata aspects, activity in this period appears dominated by DOI and its related initiatives, and by the ISO TC46/SC9 initiatives.

The DOI-based Crossref organisation announced in October that seven new publishers and two new affiliates had recently joined, making a new total of 83 publishers, and 23 affiliates and agents now on board. There are further technical developments also under way (see later).

The International DOI Foundation has, as is usual these days, been very active. Phase 2 of its DOI-EB project was demonstrated at the Frankfurt Book Fair 2001 in October, and in November there were further announcements on the indecs2 rights data dictionary - more properly known as <indec>2RDD. There are now nine organisations in the consortium in which <indec>2RDD will be developed to enable the exchange of key information between content industries and ecommerce trading of intellectual property rights.

Despite the ISBN being a veteran in the identifier world, its community is not inactive, with ISO TC 46/SC 9 currently considering a proposal to revise the 1992 edition of its standard (ISO 2108). The objectives partly relate to the acknowledgement of the changing face of publishing prompted by the arrival of electronic books - e-books are “consuming the existing capacity of the ISBN system at a faster rate than originally anticipated when it was designed for numbering printed books in the late 1960's.” This need to embrace the growing digital market was noted in an earlier report commenting on the paper discussing the role of the ISBN in the digital world. And the objectives relate partly to the need to embrace metadata - “to specify the metadata to be associated with ISBN assignments and the method of its association”; the ISBN will no longer be only an intelligent identifier. The target date for implementation of changes to the ISBN system is January 2005.

The ISBN's ‘cousin’, the International Standard Textual Work Code (ISTC) is also making progress. ISO TC 46/SC 9 announced in November that a Committee Draft has been distributed for review and voting to its member bodies and liaison organisations, with voting scheduled to end on January 30 2002.

And still in the same family, in October ISO announced the ratification of the International Standard Musical Works Code (ISWC) as the unique standard for the world-wide identification of musical works. It is now referred to as ISO/15707, of which the first edition will be published in November 2001.

#### 1.2 Overlaps and gaps identification

Not much to report in this regard, and what there is reflects the trend towards overlap reduction and collaboration.

For ISBN, an A-liaison has been established with the International DOI Foundation (IDF). For those unfamiliar with the terminology, an A-liaison is with an organisation which makes “an effective contribution to the work of the technical committees or

subcommittees for questions dealt with by this technical committee or subcommittee. Such organisations are sent copies of all relevant documentation and are invited to meetings...”

Nevertheless, note that for example an ISWC identifies a musical work as “an intangible creation.” The manifestations of that work use different identifiers - to quote the ISWC committee, “the International Standard Recording Code (ISRC) for sound recordings, the International Standard Music Number (ISMN) for printed music, and the International Standard Audiovisual Number (ISAN) for audiovisual works.” This is not to say that in this case overlaps exist, but this serves to exemplify that many initiatives are unavoidably related to each other, and that overlaps may sometimes be a fact of life.

### **1.3 Trends**

One recent trend has been to seek to extend the scope of initiatives beyond their original remit. The IDF is looking to do this, with plans for development of the DOI in other sectors and applications. Crossref too is moving in that direction, with the implementation of developments to extend its services to books and conference proceedings.

### **1.4 Main issues**

No major new issues have presented themselves in this period. Commercial considerations appear still to dominate, with digital rights management remaining at the forefront.

## **2 Registry activities**

The particular emphases found in the publishing sector have not given rise to much in the way of registry activities. Indeed, an analysis of the various activities in this sector at the very start of the Metadata Watch project listed more than thirty possibly relevant initiatives, none of which was considered to be registry-related.

Perhaps the closest parallel nevertheless, where metadata considerations do arise, is in those schemes, typically for identifiers, which require registration agencies, in that metadata submission is an integral part of registration.

For example, DOI Registration Agencies provide the infrastructure to allow registrants to declare and maintain the metadata associated with a DOI. More than this, it is not accepted blindly, but a level of QA is expected to ensure that metadata is “consistent and complies with both DOI Kernel and appropriate DOI Application Profile standards.” Registration Agencies can also develop services exploiting the metadata, although Kernel metadata must be publicly declared and freely available. The DOI Namespace is in fact a data dictionary and registry, although the IDF recognise that “existing well-defined metadata schemes such as ONIX” can contribute the required metadata.

Likewise, an ISWC agency would allocate ISWC numbers and administer a database for the numbers and their corresponding descriptive metadata. ISTCs will be allocated by ISTC registration agencies, although the administration of the ISTC system is still under discussion, but similar procedures would no doubt be implemented.

## Appendix B. Audiovisual domain

*Correspondent: Annemieke de Jong, Nederlands Audiovisueel Archief*

### 1 Introduction

Information on metadata developments and initiatives in the domain of audiovisual production, archiving and distribution is provided on several public and restricted web sites. Not all of these web sites can be defined as 'official' metadata registries, but most of them do offer detailed, structured information on ongoing work on dictionaries, schema's and data models. These websites should mainly be seen as a publication context for work on attributes (names, definitions, usage, syntax) and metadata schema's (for storing, processing and/or exchanging metadata and essence) that are being developed for a specific domain, e.g. professional broadcast production; multimedia technology; audiovisual archiving; B2B exchange of metadata; B2C applications. In some cases the registries officially prescribe standards. Although a few sites also report about 'real world' experiences, implementation and practical usage of specific models and dictionaries, the emphasis generally lies on providing information on the theoretical work that is going on. Special kinds of metadata registries are being maintained by the partners of some of the audiovisual and multimedia projects within the Fifth Framework of the IST-programme of the European Commission. Web publication of the workpackage deliverables on metadata dictionaries and models developed for the project, here serve as internal metadata registries.

The metadata-websites are usually set up and maintained by the organisation that coordinates the work, in most cases being an official standardising community. Next to representatives of this standardizing body, this group may consist of broadcasters, audiovisual archives, software vendors, system integrators, telecommunication companies and academic partners. The general objective of the sites is to inform the members and other interested parties alike, on achieved and/or ongoing work and thus contribute to the overall synchronisation of metadata development on the various organisational, national and international levels. The sites may also serve as an interactive, virtual 'meeting and working area', that holds repositories of dynamic documents, each describing parts of the work-in progress. These documents may contain recent working output of working groups as well as formal recommendations. Comments and modifications to draft documents may be processed online. The contributions, sent in by members, are collected, structured and integrated in new versions of the publications on a regular basis.

The approach to public availability of the documents differs from registry to registry. Generally only part of the information on the sites can be publicly and freely accessed. Access to metadata work on project sites is usually a hundred percent restricted to partners of the project. The access level to the documents on the sites of standardising bodies may vary, usually depending on the stage the work is in. When documents are still subject to change, they may only be distributed for ballot. Only members may have access to these documents, in order to send in editorial responses, comments on outstanding issues and ballots. Some standardizing communities offer non-profit organisations and public broadcasters a free, online possibility to become an associate member. In other cases a company or individual has to become a paying member to access any document still under ballot.

## **2 Some examples of audiovisual metadata registries**

### **2.1 Media Exchange Format (MXF) Information centre**

<http://www.pro-mpeg.org/mxf.htm>

#### **Objective MXF metadata work**

Initiative of the Pro-MPEG forum, specifies a format for the transfer of programme material between equipment in the professional broadcast environment. MXF Files are intended for sequential writing and for sequential and random access reading. MXF Files may be directly converted to and from standardised streaming formats. Types of programme material include video, audio, data essence and associated metadata. MXF is meant for storing and forwarding finished work and serves as a 'wrapper' for exchange of various types of frame based audio and video essence, along with the metadata required to describe and use it. MXF uses the KVL metadata system, which may refer to a local dictionary or to external public registries. The combination of the normative and informative sections facilitates flexible television equipment that will be interoperable over a variety of user-specific applications, including those specified by relevant SMPTE Standards.

#### **Contents and function MXF site**

Platform for exchange of information; access to working output; providing general and specific instructions on how best to implement new standards in a multi vendor environment; unified approach to conformance approval. Latest version of MXF (no.8) is put on the MXF Information Centre site for ballot. SMPTE metadata sets can be seen 'in action' here.

#### **Access conditions and users**

Publicly accessible are working documents on engineering guidelines and recommended practices on format specification; descriptive metadata specification; label registry; various mappings. Separate sections for ballot responses and issue lists (restricted access). Documents under ballot restricted to principal members (paying commercial companies and manufacturers) and associate members (non-commercial provision, broadcasting, video service, scientific research).

### **2.2 EBU Project group P-Meta**

[http://www.ebu.ch/pmc\\_home.html](http://www.ebu.ch/pmc_home.html)

#### **Objective P-Meta work**

Building a data model for the exchange of programme material between European broadcasters, archives, producers and distributors; developing a standard approach to structuring information related to media items and their exchange between process stages and business entities. Project focuses on the development of metadata information standards and architecture; on unique identifiers in broadcast use; on technical metadata standards and architecture. P-Meta work builds on the seminal work of SMPTE on defining the Metadata Dictionary, Unique Material Identifiers SMPTE UMID 330, mapping of metadata into transport and the work to date by members.

## **Contents and function P-Meta site**

Virtual 'working area' for the ongoing work of the P-Meta group; FTP-site with documentation on the work packages, stage plans, presentations, progress reports, reference documents other standards (SMEF, BBC, RAI data model, Dublin Core, MPEG-7), information on the ongoing work on the unique identifiers, results of mapping work P-Meta attributes with Dublin Core, SMPTE-metadata and FIAT-IFTA Minimal Datalist; minutes of the meetings, attributes list and draft XML Schemas.

### **Access conditions and users**

FTP site with document repositories. Fully restricted area, site can only be accessed via a username and password to be owned exclusively by members of the P-metagroup (presumably for as long a period as the work is in progress, officially until the end of 2001).

## **2.3 TV-Anytime**

<http://www.tv-anytime.org>

### **Objective TV-Anytime work**

The TV-Anytime Forum seeks to develop specifications to enable audiovisual and other services based on mass-market high volume digital storage in consumer platforms (local storage). The Forum develops specifications for open interoperable and integrated secure systems, from content creators/providers, through service providers, to consumers. TV-Anytime has designed a framework based around a data model and a common metadata representation format. The Forum develops specifications for the following tools: content referencing, metadata, and rights management. As a result, the TV-Anytime metadata system shall allow development of competitive or complementary applications and services which support for example interactive TV, parental guidance systems, multilinguality, different views, indexes, identification and differentiation of the content, storage of content, various e-commerce models, personal annotations, links to other programmes, history information, synchronisation between content and metadata, and protection of personal data.

### **Contents and function of the TV-Anytime site**

General section: press releases, news, and information on Forum and the membership, announcements meetings and events, calls for contributions. Documents approved by the Forums Plenary meetings are the official documents of the TV-Anytime Forum. These include resolutions, reports of meetings, recommendations, work plans, working procedures etc. Descriptions and pointers to this extensive set of documents can be found at the site. Several separate TV-Anytime document repositories exist for the Forum's output and for the output of the working groups.

### **Access conditions and users**

General sections and FTP-site with document repositories. All published documents are publicly available, password is freely provided on the site. Input and participation of all interested parties is invited.

## **2.4 Society of Moving Pictures and Television Engineers (SMPTE)**

<http://www.smpte.org>

## **Objective metadata work SMPTE**

SMPTE means to develop and harmonise standards for the exchange of programme material; support global interoperability by defining and structuring metadata tags in a way that enables the interchange of SMPTE metadata with metadata from different sources and originated by other bodies.

### **Contents and function SMPTE site**

To provide general information on SMPTE, membership, conferences etc. Standards, recommended practices, engineering guidelines are published by number and can be purchased online in the SMPTE-store. SMPTE site frequently releases trail publications of proposed SMPTE Standards, recommended practices and engineering guidelines that are posted for review by the public. Site currently offers no information about the status of the SMPTE metadata dictionary that was out for ballot earlier this year. News on the new version-in-progress to be provided by Mr. Mike Cox (mirador\_techniques@compuserve.com)

### **Access conditions and users**

Apart from the general sections, all access to information SMPTE site is restricted, to enter interested parties are obliged to become a registered, paying member of the SMPTE society.

## **2.5 MPEG-7**

<http://mpeg.telecomitalia.com/standards/mpeg-7.htm>

### **Objective of the MPEG-7 work**

MPEG-7 will provide a set of standardised tools to describe multimedia content. The MPEG-7 standard will be subdivided into seven parts: systems, Description Definition Language (DDL), visual elements, audio elements; multimedia descriptions schemes; reference software; conformance: guidelines and procedures for testing conformance of MPEG-7 implementations. MPEG-7 application domains include audiovisual archives (for storage and retrieval of audiovisual databases, broadcasting (for media selection and distribution), Web-based services ('push and pull'), teleshopping, education, and biomedical services (such as surveillance). It concerns all types of multimedia: audio and speech; moving video, still pictures, graphics, and 3D models; and information on how objects are combined in scenes.

### **Contents and function of the MPEG-7 site**

Site provides a complete overview of the MPEG-7 standard version 5.0, explaining in detail which pieces of technology it includes and what sort of applications are supported by the technology. Document also includes an introduction to XML and the relation to MPEG-7, a FAQ section, as well as information on how to contribute and how to get involved in the MPEG-7 work. Also included are references to other MPEG-7 sites: <http://drogo.scelt.it/mpeg>, <http://www.mpeg-7.com> (both contain information on requirements applications, principal concept lists, many publicly available MPEG-7 documents and links to other MPEG-7 web pages.)

### **Access conditions and users**

Site contains one large document that reflects the status of the MPEG-7 work in March 2001. No restricted sections, open to all.

## Appendix C. Educational Domain

*Correspondent: Erik Duval, Katholieke Universiteit Leuven, Belgium*

### 1 Metadata registries in education

#### 1.1 Introduction

Generally speaking, it appears that the concept of metadata registries is still quite immature in the field of educational metadata. The DESIRE registry is the more developed one, but it remains unclear how intensive it is being used and for what purpose. The other initiatives in this area are less advanced.

Overall, most “registries” are focusing on the provision of information to human readers, rather than facilitating “automatic” crosswalks or mappings between metadata schemas.

#### 1.2 DESIRE

The DESIRE registry lists a number of application profiles that are oriented towards education [1]:

- ADL SCORM Raw, Content and Course Metadata (see 3.3)
- The DCMI-Education working group proposal (see 3.4)
- The EASEL metadata set [16];
- The IMS Learning Resource Metadata Core (see 3.2)
- The MILO metadata scheme (see 3.8).
- The National Curriculum Metadata Standard (see 3.9)
- The National Grid for Learning Scotland (see 3.10)
- The Virtual Teacher Centre (see 3.11)

Functionality of the DESIRE registry includes the ability to browse and search application profiles and other schemes. For each of those are listed: the status, a description, the registration authority and the date of last modification. For each element in a profile, the recommended scheme, if any, is listed. In the DESIRE registry, the latter kinds of schemes include vocabularies

#### 1.3 GMR

The German Metadata Registry (GMR) is quite Dublin Core centric, and includes a registry specific to the education domain [6]. This resource follows a much more “loose” idea of what a registry is, more akin to the idea of a portal, and including press releases, a literature list, web links, etc. Education related elements include:

- GradeLevel
- SubjectOfInstruction
- UserRestrictionsAndCost

- TechnicalRequirements
- PhysicalMedium

A mapping is defined between school levels and the German system of “Gymnasium”, “Realschule”, “Gesamtschule”, and “Berufsbildende Schule”.

#### **1.4 Metadata.net**

At metadata.net [7], a registry lists a number of schemas, including EdNA (see 3.5), IMS (see 3.2), GEM (see 3.7). Functionality is limited to references to the relevant web sites, as well as a specification of the elements that includes a description, a label and a syntax specification.

#### **1.5 Metaform**

The Göttingen registry includes information about the LOM (see 2.1) and GEM (see 3.7) specifications. Functionality includes the ability to identify manifestations of (mainly Dublin Core) elements in different schemas, to map between Dublin Core and other schemas

## **2 Standardization work**

### **2.1 IEEE LTSC LOM**

The IEEE working group on Learning Object Metadata is finalizing its second recirculation ballot. The next meeting of this group takes place in December 2001, in Hawaii. It is expected that a new ballot will be carried out later that month. Current estimate is that the current specification would be finalized as an IEEE standard in the first half of 2002.

The DCMI/IEEE LOM Memorandum of Understanding [5] has been re-activated at a meeting in Ottawa. As a concrete outcome of that meeting, a “metadata manifesto” is being developed that presents the common views on modular, interoperable metadata. It is expected that this “manifesto” will be finalized by end 2001.

### **2.2 CEN/CENELEC LTWS**

The CEN/CENELEC Learning Technologies WorkShop [9] will have its next meeting in Berlin, in December 2001. Work on so-called “Educational Modeling Languages” (EML) is progressing steadily, with a number of open events in October-December 2001. At this time, the main focus is on comparisons of different approaches, and on the separation from conceptual modeling and binding issues.

Work is also progressing on quality assurance (both product and process related), educational copyright, internationalization of LOM and a repository of taxonomies and vocabularies that are used in educational metadata.

## **3 Consortia based work**

### **3.1 ARIADNE**

The ARIADNE Foundation [10] held its first annual conference in November 2001, in Leuven. Some 15 presentations dealt with a variety of issues, including the

description of educational resources through the multilingual ARIADNE metadata toolset.

### **3.2 IMS**

The Instructional Management Systems (IMS) consortium [11] released an “errata update” of its LOM based metadata specification, in order to correct minor discrepancies between element descriptions in the binding and the conceptual schema.

### **3.3 ADL**

The Advanced Distributed Learning (ADL) initiative [12] holds a “plugfest” in November, with practical interoperability experiments, based on the SCORM specification. In October, SCORM v 1.2 was released, with a cleaner separation between the content aggregation model and the run-time environment. An application profile of the IMS content packaging specification has been integrated into SCORM. This profile also replaces the earlier Content Structure Format.

### **3.4 DCMI-Education**

This working group had a meeting in October, at the DC meeting in Tokyo, Japan. Important topics included:

- the inclusion of Education specific provisions in the DCMI specification (Audience element, Mediator qualifier for the Audience element, and a ConformsTo qualifier for the Relation element);
- the DCMI/IEEE LOM MoU (see 2.1);
- proposal for a Level qualifier for the Audience element;
- relevance of education specific resource type vocabulary, additional element qualifiers, an accessibility qualifier, capturing teaching processes and learning objectives.

### **3.5 EDNA**

There seem to be no new metadata related developments on the “EDucation Network Australia” [13].

### **3.6 EUN**

The EUropean Schoolnet [14] project now includes a search facility over educational repositories in Europe. This keyword based search facility relies on spider software, apparently not unlike the approach of the major web search sites.

### **3.7 GEM**

The Gateway to Educational Materials [15] remains one of the larger repositories of educational material. No new metadata related developments are apparent.

### **3.8 MILO**

MILO (Metadata and Information for Learning Opportunities) is an applied research project for learndirect, which delivers learning services for the UfI, the University for Industry [2]. MILO focuses on learning event information (courses). The scheme is

based on IEEE LTSC LOM (see 2.1) and DCMI-Education (see 3.4), and a local extension that covers the location where the event takes place.

### **3.9 National Curriculum Metadata Standard**

This is a metadata specification for the description of resources which refer to the “National Curriculum” in the UK, maintained by the Qualifications and Curriculum Authority (QCA) [3]. The specification contains 23 data elements, 11 from DCMI-Education (see 3.4) and 12 local ones that deal with expiry and publication date, publisher email, learning duration, end users, status, teaching subject, cross-curricular areas, etc.

### **3.10 National Grid for Learning Scotland**

This metadata specification for describing educational resources is based on the EUN specification (see 3.6) [4]. From that specification, 11 data elements are mandatory. Controlled vocabularies are defined for subject, resource type and user level. A template based on HTML META tags is also provided.

### **3.11 Virtual Teacher Centre**

The Virtual Teacher Centre (VTC) is a service for schools that includes a facility to search for resources across the U.K. National Grid for Learning [5]. The metadata specification is based on that of the National Curriculum (see 3.10), with additional elements for study level, study year, educational user, etc. Besides a “metadata handbook”, the VTC also provides an RDF implementation guide and defines “allowable values” for most elements.

## **4 References**

- [1] <http://desire.ukoln.ac.uk/registry/index.php3>
- [2] <http://www.openline.go-legend.net/MILO/index.htm>
- [3] <http://www.nc.uk.net/download/standard206.rtf>
- [4] <http://www.ngflscotland.gov.uk/ngflfocus/keypub/ngflmetadata.asp>
- [5] <http://standards.ieee.org/announcements/metaarch.html>
- [6] <http://www.mpib-berlin.mpg.de/dok/metadata/gmr/mdeden.htm>
- [7] <http://metadata.net/>
- [8] <http://ltsc.ieee.org/>
- [9] <http://www.cenorm.be/iss/Workshop/lt/Default.htm>
- [10] <http://www.ariadne-eu.org/>
- [11] <http://www.imsproject.org/>
- [12] <http://www.adlnet.org/>
- [13] <http://standards.edna.edu.au/metadata/index.html>
- [14] <http://www.eun.org/>
- [15] <http://www.thegateway.org/>
- [16] <http://www.fdgroup.co.uk/easel/>

## Appendix D: Cultural Heritage Domain

*Correspondents: Michael Day and Marieke Napier (UKOLN)*

### 1 Current state of domain

This is not a review of all existing metadata activities in the cultural heritage domain, but just points to some important current initiatives.

#### 1.1 Metadata Encoding and Transmission Standard (METS)

The Metadata Encoding and Transmission Standard (METS) is an initiative of the Digital Library Federation (DLF) and is being maintained in the Network Development and MARC Standards Office of the Library of Congress. The METS schema is a standard for encoding descriptive, administrative, and structural metadata about digital library objects and the complex links between these types of metadata within a repository. It does so by providing an XML document format that can be used for both the management of digital library objects within a repository and the exchange of such objects between repositories (or between repositories and their users).

Work on METS began in May 2001 and is a continuation of work undertaken as part of the Making of America II (MOA2) testbed project (a DLF project led by the University of California at Berkeley). The MOA2 project was one of the first to distinguish between descriptive, administrative and structural metadata, and it produced an XML Document Type Definition (DTD) to describe digitised resources (Hurley, *et al.*, 1999). The METS schema builds on the MOA2 DTD, but is based on the XML Schema language.

A METS document is made up of 4 sections: Descriptive Metadata, Administrative Metadata, File Groups and a Structural Map. Metadata can either be included within the METS document itself or referenced via an identifier or locator. The following descriptions are based on those published in the *METS Overview & Tutorial* provided on the initiative's Web pages (<http://www.loc.gov/standards/mets/>):

- Descriptive Metadata – This section points to descriptive metadata that is either external to the document (e.g., a Dublin Core (DC) or MARC record maintained on a separate server) or embedded internally, or both, using <dmdSec> (Descriptive Metadata Section) elements. External metadata is linked to by a <mdRef> element and internal by <mdWrap>.
- Administrative Metadata – This section provides information on how the files were created and stored (technical metadata), intellectual property rights metadata, metadata about the source object from where the digital object derives and digital provenance metadata. It also includes information on source/destination and master/derivative relationships between files and migration/transformation information. As mentioned previously this metadata can be held externally or internally.
- File Groups – This section lists all the files that make up the digital object e.g. thumbnails, master archival images, PDFs, etc. Each file is given a unique ID attribute that allows it to be referenced by other parts of the document. The

<GROUPID> attribute allows the linking of two versions of the digital object (e.g., a WAV file and MP3), that contain the same information.

- Structural map – The structural map outlines the hierarchical structure for the digital library object allowing users to navigate clearly. The map links to content files and metadata for each element. The <StructMap> element encodes this hierarchy as a nested set of <div> elements. Each div element may contain pointers and file pointers. METS pointers specify other documents as containing the relevant information, which is useful in keeping the size down when encoding large collections of material.

### **1.2 DC-Library Application Profile (DC-Lib)**

A DCMI Working Draft published in October proposed an application profile to help clarify the use of the DCMES in libraries and library-related applications and projects (Guenther, 2001). The DC-Library Application Profile (DC-Lib) defined how existing DC elements should be used (e.g. whether elements are mandatory or not, which schemes should be used, specific DC-Lib qualifiers, etc.) and proposed the adoption of a DC-Ed 'audience' element, and a new 'holding location' element. The draft DC-Lib application profile was discussed at a meeting held in conjunction with the IFLA Council and General Conference held in Boston (August 2001) and at the DC-2001 Conference in Tokyo (October 2001).

### **1.3 Encoded Archival Context (EAC)**

The Encoded Archival Context is being developed by an international group of archivists as an XML-based standard for the description of record creators (as individuals, families and corporate bodies). It is intended to compliment the existing Encoded Archival Description (EAD), maintained by the Network Development and MARC Standards Office of the Library of Congress in partnership with the Society of American Archivists (<http://www.loc.gov/ead/>). The EAC initiative differs markedly from libraries' traditional emphasis on name authorities (i.e. unique identification) in that it would provide more detail on the entities that bear such names. Pitti (2001) argues that a standard for describing record creators would help the documentation process, facilitate access and (potentially) lead to the creation of biographical and historical databases that could be used as resources in their own right. The development of EAC has been influenced by the *International Standard Archival Authority Record for Corporate Bodies, Persons and Families* (ISAAR(CPF)) published by the International Council on Archives and it is hoped that EAC will inform later versions of this standard. It is the opinion of Pitti (2001) that EAC could be useful in broader cultural heritage contexts, including genealogy. EAC could also attempt to accommodate existing information on persons or corporate bodies, e.g. data from the Library of Congress Name Authority File. It may also possible to link EAC data to the output of large-scale biography initiatives like the *New Dictionary of National Biography* (<http://www.oup.co.uk/newdnb/>), some of which are already available in digital form.

## **2 Metadata registries in the cultural heritage domain**

Metadata registries are services that support interoperability by providing authoritative definitions of element names and translations, information about the use of elements and schemes, and mappings of elements to other metadata standards. Serious discussion of metadata registries originated in the late 1990s with the

development of ISO/IEC 11179 (*Specification and standardisation of data elements*) and events like the Joint Workshop on Metadata Registries held at Berkeley in July 1997. The importance of metadata registries was noted by a 'metadata summit' organised by the RLG (Cromwell-Kessler & Erway, 1997) and by the report of the EU-NSF Working Group on Metadata (1998). The RLG metadata summit mostly concerned mappings or crosswalks, and concluded that there was a need for registries to "serve as authoritative records of the equivalence between the data elements of various metadata schemes, are essential to ensure their consistency." It cited crosswalk sites then maintained by the Library of Congress (for MARC21) and OCLC as examples of *de facto* registries.

More recently, metadata registries were the focus of one of the recommendations of the draft Library of Congress action plan for the *Bibliographic Control of Web Resources* that grew out of the Conference on Bibliographic Control in the New Millennium. The action plan recommended the identification and publicising of existing registries of metadata schemes "to establish points of convergence among them, to promote the consistent labelling of fields, and to facilitate mapping of fields" (Library of Congress, 2001).

Registry development started with simple implementations like the HTML-based registry created by the ROADS project for ROADS/IAFA templates (<http://www.ukoln.ac.uk/metadata/roads/templates/>). Phase II of the DESIRE project developed a more sophisticated prototype of a metadata registry and input several schemes: the existing Dublin Core standards, the BIBLINK Core, the eLib simple collection level description and the most popular types of ROADS template (Heery, *et al.*, 2000). Crosswalks were made via the generic terms defined by the ISO Basic Semantics Register (BSR).

Active metadata registries include the DESIRE registry, which currently includes a number of Metadata for Education Group (MEG) related schemas (<http://www.ukoln.ac.uk/metadata/education/registry/>) and MetaForm. MetaForm (<http://www2.sub.uni-goettingen.de/metaform/>) is a registry hosted by the Göttingen State and University Library (SUB) as part of the META-LIB project. It has a particular focus on implementations of Dublin Core and includes a number of DC-based 'crosswalks', information on the usage of DC (called 'crosscuts') and various mappings between DCMES and other formats.

Another registry is being set up by the Networked Knowledge Organization Systems/Services (NKOS) working group to help define controlled vocabularies, classification schemes, thesauri, etc. To date, the group has produced a draft *Reference Document for Data Elements* that defines a proposed element set for a registry of knowledge organisation systems. Each element is defined using a set of ten attributes from the ISO/IEC 11179 standard ([http://staff.oclc.org/~vizine/NKOS/Thesaurus\\_Registry\\_version3\\_rev.htm](http://staff.oclc.org/~vizine/NKOS/Thesaurus_Registry_version3_rev.htm)).

Metadata format maintainers sometimes publish official or semi-official mappings to other metadata standards. For example, the Network Development and MARC Standards Office of the Library of Congress publish and maintain mappings from MARC21 to (or from) Dublin Core and ONIX. The MARC documentation pages also link to third party crosswalks that map MARC to (or from) the FGDC Content Standards for Digital Geospatial Metadata and GILS (<http://lcweb.loc.gov/marc/marcdocz.html>).

Other registry-type services include the database of international metadata activity published through the Metadata Observatory (<http://www.sub.uni-goettingen.de/ssgfi/observatory/>) - a product of the CEN/ISSS Workshop on MMI-DC (Metadata for Multimedia Information - Dublin Core) - and the guides to electronic information exchange standards produced by the Diffuse project (<http://www.diffuse.org/>).

### 3 References

Cromwell-Kessler, W. & Erway, R. (1997). *Metadata summit organised by the Research Libraries Group, Mountain View, California, July 1, 1997: meeting report*. <http://www.rlg.org/meta9707.html>

EU-NSF Working Group on Metadata. (1998). *Metadata for digital libraries: a research agenda*. Draft 10. Le Chesnay: ERCIM. <http://www.ercim.org/publication/ws-proceedings/EU-NSF/metadata.html>

Guenther, R. (2001). *DC-Library Application Profile (DC-Lib)*. DCMI Working Draft, 12 October. <http://dublincore.org/documents/2001/10/12/library-application-profile/>

Heery, R., Gardner, T., Day, M. & Patel, M. (2000). *DESIRE metadata registry framework*, 31 March. <http://www.desire.org/html/research/deliverables/D3.5/>

Hurley, B.J., Price-Wilkin, J., Proffitt, M. & Besser, H. (1999). *The Making of America II Testbed Project: a digital library service model*. Washington, D.C.: Council on Library and Information Resources. <http://www.clir.org/pubs/abstract/pub87abst.html>

ISO/IEC 11179 (Parts 1 to 6), *Information technology -- Specification and standardisation of data elements*. Geneva: International Organization for Standardization.

Library of Congress. (2001). *Bibliographic control of Web resources: a Library of Congress action plan*, 25 July 2001: <http://lcweb.loc.gov/catdir/bibcontrol/draftplan.html>

Pitti, D.V. (2001). *Creator description: Encoded Archival Context*. Paper presented at Computing Arts 2001: Digital Resources for Research in the Humanities, University of Sydney, 26-28 September 2001. <http://setis.library.usyd.edu.au/drrh2001/papers/pitti.pdf>