

SCHEMAS

Contract: N° IST-1999-10100

Forum for Metadata Schema Implementers

**METADATA WATCH REPORT #8
STANDARDS FRAMEWORK #4**

D29 / D35

Document number:

SCHEMAS-PwC-WP2/WP3-D29/D35-Final-20020131

General Information

Title	Metadata Watch Report #8 and Standards Framework Report #4
Creator	Makx Dekkers
Contributor	Laurie Causton, Michael Day, Erik Duval, Annemieke de Jong
Subject-Keywords	Deliverable D29/D35; WP2/WP3; Metadata Watch Report #8; Standards Framework Report #4
Description	This document comprises Metadata Watch Report #8 and Standards Framework Report #4 and domain reports for the Audiovisual, Cultural Heritage, Educational and Publishing domains
Publisher	PricewaterhouseCoopers
Date	31 January 2002
Type	Text Manuscript
Format	application/MSWord 2000
Identifier-	
Document Number	SCHEMAS-PwC-WP2/WP3-D29/D35-Final-20020131
Language	English
Rights	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/"
xml ns: smes="http://www.schemas-forum.org/registry/schemas/SCHEMAS/1.0/smes#">
<rdf:RDF rdf:resource = " ">
<dc: title> Metadata Watch Report #8 and Standards Framework Report #4
</dc: title>
<dc: creator> Makx Dekkers </dc: creator>
<dc: contributor> Laurie Causton </dc: contributor>
<dc: contributor> Michael Day </dc: contributor>
<dc: contributor> Erik Duval </dc: contributor>
<dc: contributor> Annemieke de Jong </dc: contributor>
<dc: subject> Deliverable D29/D35 </dc: subject>
<dc: subject> WP2/WP3 </dc: subject>
<dc: subject> Metadata Watch Report #8 </dc: subject>
<dc: subject> Standards Framework Report #4 </dc: subject>
<dc: subject> Metadata activity reports </dc: subject>
<dc: description> This report comprises Metadata Watch Report #5 and Standards
Framework Report #4 and domain reports for the Audiovisual, Cultural Heritage,
Educational and Publishing domains </dc: description>
<dc: publisher> PriewaterhouseCoopers </dc: publisher>
<dc: date> 2002-01-31 </dc: date>
<dc: type> Text </dc: type>
<dc: format> application/MSWord </dc: format>
<dc: identifier> SCHEMAS-PwC-WP2/WP3-D29/D35-Final -20020131 </dc: identifier>
<dc: language> en </dc: language>
<dc: rights> European Commission; External distribution via SCHEMAS Web site
</dc: rights>
</rdf:RDF>
```

Table of Contents

1. INTRODUCTION	5
2. FOCUS	5
3. DETAILED METADATA AND STANDARDS ACTIVITY REPORTS	5
4. SYNTHESIS	5
APPENDIX A: AUDIOVISUAL DOMAIN	7
APPENDIX B. CULTURAL HERITAGE DOMAIN	11
APPENDIX C: EDUCATIONAL DOMAIN	14
APPENDIX D: PUBLISHING DOMAIN	17

1. Introduction

This deliverable combines the eighth Metadata Watch Report and fourth Standards Framework Report from the SCHEMAS project. They form the last deliverables of this part of the work in the SCHEMAS project.

The purpose of the SCHEMAS Metadata Watch (MD Watch) is to provide a quarterly overview of worldwide 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. The Standards Framework is aimed at mapping this wide diversity of metadata standardisation to application areas and to provide information to implementers about what is going on, what they could use and where they can find information about how to use certain standards

This report contains domain reports from the SCHEMAS domain correspondents on the Audiovisual, Cultural Heritage, Education and Publishing domains.

2. Focus

In this combined final deliverable of the metadata watch and standards framework activities, as well as the domain reports contained herein, we have focused specially on the future outlook of metadata activities and standards, as far as was possible from analysing these activities.

3. Detailed metadata and standards Activity Reports

An important part of the work in the metadata watch and standards activities in the SCHEMAS project was the building of a set of so-called Activity Records, structured descriptions of metadata projects and standards, that are integrated in the SCHEMAS Registry. These records are not reproduced as part of this report, but the work of the domain correspondents in creating and updating these many records was a critical factor in the production of the results of the SCHEMAS project. All the Activity records that have been produced can be found in the SCHEMAS Registry (<http://www.schemas-forum.org/registry/desire/index.php3>).

4. Synthesis

At its inception, SCHEMAS worked from a perspective that there was confusion among implementers of metadata about which standards and approaches to follow. There was a perception that there were many standards with overlapping objectives and insufficient experience and knowledge across many domains of how the application of standards would meet local requirements in the short term and wider interoperability in the long term.

A changing metadata landscape

In the last two years, the metadata landscape has changed in several ways:

- Many people have started to realise that it is important to avoid or reduce duplication of effort – in most of the domains that were covered by the metadata watch, there is now increased co-operation between activities, looking for ways to either harmonise, align, map, or combine metadata approaches.

- In some domains, a dominant standard has emerged that federates other activities, e.g. ONIX in Publishing, SMPTE in Audiovisual, and LOM in Education.
- In several domains, the relationship between domain-specific standards and the Dublin Core is investigated with a view for wider, cross-domain, interoperability.
- There is a shifting emphasis from schemas design to more fundamental modelling approaches. Important activities that are being discussed in various domains are the IFLA FRBR model (concentrating on the various forms – conceptual and physical – of content and the relationships between those forms, see: <http://www.ifla.org/VII/s13/frbr/frbr.htm>) and the ABC Ontology and model (elaborating a model that looks specifically at interoperability between metadata approaches and concentrates on the activities that are associated with content, see: <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Lagoze/>).

We hope that SCHEMAS has contributed to some of these changes, in the sense that SCHEMAS provides information to metadata schemas designers that enables them to see and evaluate relationships between various approaches, with a longer-term goal to increase harmonisation efforts.

Standards and sustainability

Currently, the metadata standards area seems to be in a transition period:

- For many standards, the definition phase is now completed and a phase has started that will see proof-of-concept activities, wide implementation and evaluation of approaches.
- Standards activities are entering a phase that they need to consider their sustainability, and in some cases there are attempts to work closer together with others, and even mergers occur.
- With the creation of large metadata collections and development of services, domain boundaries are further blurring, leading to an increased interest in cross-domain interoperability. An important driver here is the Semantic Web vision expressed by the W3C, a concept that is fundamentally rooted in cross-domain contexts and that also introduces the notion of partial understanding, acknowledging that there may be different levels of understanding between approaches: deep understanding and interoperability in a limited area or group of organisations, and more shallow understanding in a wider context.

These factors contribute to an increased awareness that standards groups need to co-operate to reduce duplication of effort, thereby reducing the confusion for implementers and increasing the likelihood that wider interoperability can be achieved.

Future outlook

It can be expected that the role of modelling activities like IFLA FRBR and ABC, basic metadata element sets like Dublin Core, constructs such as application profiles (perhaps formalised along lines explored by the SCHEMAS project), and enabling technologies like RDF and XML will form the cornerstones for future interoperability.

An essential element of these further developments is human interaction among the designers of standards and increased awareness of user needs. This is one of the aspects covered in CORES, the follow-up project to SCHEMAS, that is expected to start in May 2002.

Appendix A: Audiovisual Domain

Correspondent: Annemieke de Jong, Nederlands Audiovisueel Archief

Current situation

A number of evolving metadata modelling standards for the audiovisual domain now exist in literature that may describe metadata dictionaries and/or data models. These standards are either designed for a specific 'audiovisual' application or use, or they focus on aspects of the production, archiving and distribution process as a whole. The models might be developed for storing content on a database and video server (metadata and essence), in order to process content and they might focus on the exchange of metadata between different systems, using various modelling techniques like SQL, object-orientation and/or XML. Each of these approaches can cover different aspects of content and metadata creation, management and delivery. Several global domains can be distinguished:

- Capture, creation and pre-production
- Content identification
- Post-production
- Archives, preservation and cultural heritage
- Content protection and transactions
- Distribution and delivery
- Search and selection, Interactivity, User experience

Information on 'audiovisual' models and standards is directed towards professional communities and is often publicly available on the web. Within the professional media organisations several proprietary dictionaries and models have now been built, most of which are currently being tested and implemented as application profiles in (pilot) systems. Technology providers have also been developing proprietary models for their content management systems and applications. These usually include (parts of) existing standards, to which specific metadata requirements of their customers are added. Commercial models and organisation models are often not fully documented and/or publicly available and accessible.

Communication and coordination

Interoperability among metadata in different systems is a fundamental requirement in the networked digital multimedia environment. All communities involved in the setting up of media-management system requirements acknowledge the time-consuming labour of defining, structuring and mapping metadata. It can be noticed that much of the same work is being carried out in a lot of different places. For these reasons organisations tend to want to share work and experience and preferably base their designs on existing reference models. Although contributions to the development of interoperable schemes and metadata dictionaries still stem from the respective professional qualifications and commercial interests, communication between the various initiatives is seen to improve. It can be observed that the developments are increasingly coordinated at European and international levels. Professional platforms have been established (workshops, conferences, web sites) where the research community can meet with the day-to-day

practice of media organisations and the developments within the industrial circles, in order to exchange information and synchronize activities. Several European projects within the context of the IST programme address metadata-related issues and enhance the communication and coordination on organisational, national and international levels. Contacts between the projects partners lead to overall harmonization and synchronisation of metadata vocabularies and models and have a favourable effect on efficiency and interoperability.

Standard models and dictionaries must be tailored for various sorts of content providers, technology providers, functionality providers and many categories of end-users and end-usage. Although it is recognized that much can and needs to be integrated into international standards, there will always remain elements of purely local importance. It has become clear that the more authoritative reference metadata models for audiovisual resources do by far not meet every requirement in the field. On the one hand these models tend to cover only one specific aspect of the audiovisual modelling domain and have no provisions for applicability to one another. On the other hand organizations have too many different specifications as to their required data structures and formats, workflows and user communities. An obstacle for the common use of generic standardized metadata models is also that a model often reflects the business model of an organization. Detailed specifications often cannot be offered as shareware. For these reasons the level of standardisation is limited. Organisations can only join forces to a certain extent. The question that rises here is then, to what extent? Which levels and elements can be standardized and what needs to be primarily geared to the specific requirements of an organisation? Answers to these questions still have to be found in the ongoing process of comparing and mapping the authoritative standards and initiatives to the various proprietary requirements of media organisations.

Examples of harmonization

Both national and international projects and 'local' media management initiatives profit from contacts and collaborations with standardizing communities and efforts. Standardizing communities themselves also collaborate on an internal level.

The IFLA-FRBR model

An example of a standard reference metadata structure that is being used for different 'audiovisual' and multimedia purposes is the IFLA model (Functional Requirements for Bibliographical Records). The international Federation of Library Associations and Institutions has in 1998 defined a model that includes a bibliographical record as an aggregation of data that are associated with the entities, described in library catalogues and national bibliographies. These data pertain to textual, music, cartographic, audiovisual, graphic and three-dimensional materials. The IFLA FRBR has its origin in a library context and is principally devoted to textual documents. From this point of view the model is limited. Nevertheless the model has a multi-layered and hierarchical structure that can be easily specialised to describe specific requirements. More over, applications devoted to different aspects of the described documents can work on different abstraction levels. Important feature is that IFLA FRBR permits to distinguish different aspects of the same work: the distinct intellectual or artistic creation, the intellectual or artistic realisation of a work, the physical embodiment of an expression of a work and a single exemplar of a manifestation. These features make the model very suitable for the digital audiovisual environment, where it can function as a conceptual framework for digital heterogeneous media resources and multi-dimensionally structure several types of digital and analogue audiovisual formats and distribution channels.

1. The ECHO-project, an IST-Fifth Framework project of the EC, builds a digital infrastructure based on the IFLA data model, meant for the online indexing and distribution of historical film collections located in archives in the Netherlands, France, Italy and Switzerland.
2. The National Library of Portugal uses IFLA-FRBR for their multimedia collection management system to be able to digitally bind multiple manifestations of the same works, to embed and relate several multimedia contents as well as to provide hierarchical structured access to these works.
3. The IMMIX system of the Netherlands Audiovisual Archive uses the IFLA-FRBR model as a basis for its object-oriented data structure, that models various audiovisual resources as well as online archive functionalities, within a professional broadcast production environment.
4. The ABC Ontology and Model, a foundation metadata model built as part of the Australian Harmony Project and designed for individual metadata communities and system builders, has used IFLA-FRBR as a reference model for the modelling of physical, digital and analogue objects held in libraries, archives, museums and the internet.

Mappings within the standardizing community

The SMPTE metadata dictionary has its origin in the working together of the European Broadcasting Union (EBU) and the Society for Motion Picture and Television Engineers (SMPTE) in the EBU / SMPTE Task Force. It was decided that standardisation of metadata within the media industry was needed. The important thing is to realise that the SMPTE metadata dictionary and EBU P/META are complementary; the first aims at support of metadata in systems in particular in the domain of programme making in the broadcast environment, the second focuses on Business-to-Business exchanges of programme material. Although this is the case, EBU P/META has decided to map, in the near future, the final P/META scheme into the SMPTE metadata dictionary and adding those P/META attributes not covered by the SMPTE metadata dictionary. TV-Anytime, metadata for Business-to-Consumer services, is also aware of the need of compliance and interoperability. This initiative closely looks into the work of other metadata developers (e.g. MPEG-7, SMPTE and EBU P/META) and makes use of their tools where possible. TV-Anytime uses the MPEG-7 Description Definition Language (DDL) to describe metadata structure as well as the XML encoding of metadata. In 2000 the MPEG Ad-hoc Group on Integration was established in the awareness of the need for bringing different metadata schemes in the audio-visual domain together. The first step to be taken is the mapping of SMPTE and MPEG-7 (Multimedia Content Description Interface) dictionaries. The urgency of this specific mapping activity - that is not yet on its way - is widely acknowledged. MPEG-21 also intends to focus on mapping between the various standards. MPEG-7 has further liaisons with TV-Anytime, EBU P/META ea. The Advanced Authoring Format (AAF), a professional file interchange format designed for post-production and authoring, has instead of inventing its own metadata implemented SMPTE metadata, so harmonising activities are not needed here. The Media Exchange Format (MXF), a file transfer format, has its own metadata solution with a reference to the SMPTE metadata dictionary. It is also possible to use other schemes as a plug-in, for example EBU P/META.

Outside the domain, the Dublin Core metadata standard is the most frequently looked at. Both for EBU P/META and MPEG-7 actual mappings have been made. The main problem that occurs is the difference in granularity. Dublin Core Element Set, being a simple content description model, has only 15 elements, both P/META and MPEG-7

many more. MPEG-7 is highly structured and provides detailed information. Because of the difference between MPEG-7 and Dublin Core, the International Organisation for Standardisation (ISO), of which MPEG is a working group, proposed to combine and to integrate the both, by working with XML Schemas and the XML Namespace mechanism. This will make it possible that both simple Dublin Core-based searching can be done as well as fine-grained content-based using MPEG-7. The P/META - Dublin Core mapping was made because of the work going on in the EBU Panel P/FRA (Future Radio Archives). This working group proposes a core metadata set for radio archives based on Dublin Core, which will eventually be incorporated in the final P/META Scheme.

Future work

In the next period an overall picture of the audio-visual and multimedia metadata landscape should be consolidated, in particular regarding the relations between standardization activities. Project and organisational implementations, relationships and possible combination of solutions need to be assessed. To support cross-fertilisation, the results of the different implementations must be more widely published. It is necessary to ensure that each player in the media-management field, be it media organisations, IST projects or commercial technology providers, is aware of the choices of other players and incorporates harmonisation as an important decision factor. The need of communication between organisations in different environments will push to further refine the way of performing a standardized media-related data interchange. Active participation in standardisation is an absolute requirement in a framework organised around the exchange of information. The difficulty will remain in the synchronization of the work within the different committees.

Besides, harmonisation of metadata has only value if the solutions are integrated into systems. Presently, more and more professional content- and media management systems are being set up and made operational by media organisations or as project deliverables, be it as pilot, proof of concept or as 'real life' systems. Within the next few years these implementations will have generated more practical information and experience as to the usability and applicability of both common and proprietary metadata specifications. The possibilities and limitations of the 'practical' integration of standard metadata dictionaries and models will be further exposed.

A number of issues will have to be more widely addressed and documented. Process models and working procedures within media organisations have to be synchronized in depth and structured connections will have to link and synchronize metadata models and workflow management systems. More work has to be done regarding the support and integration of audiovisual legacy material (metadata and essence) into the digital media environment. The requirements of different user communities will have to be analysed and harmonised and the distributed (re)use of audiovisual content should be further supported, including the improvement of existing models for the clearing and handling of copyrights. The domain of multilinguality is also an important metadata aspect that should deserve more attention. Another key metadata domain is the affordable capture of metadata. Next to the capturing of "technical metadata" (e.g. camera positions, shooting conditions, coding parameters) another important metadata segment is the capture of descriptive metadata. Harmonisation of metadata being used for statistical applications is also a domain that needs to be further studied as statistics are going to play an important role in the analysis of e.g. the metadata related to user preferences and the user consumption history.

Appendix B. Cultural heritage domain

Correspondent: Michael Day, UKOLN, University of Bath

The cultural heritage domain is one that has long experience of developing and implementing metadata systems, whether for library catalogues, archival finding aids or museum documentation. The domain has also been heavily involved in more recent metadata developments, e.g. DCMI. The library sector continues to have an interest in 'bibliographic control,' and some within the sector have sought to extend the scope of this to networked information resources. There is sometimes a tension between those who feel that this should best be done using the metadata standards and systems developed and used within the library community (e.g. OPACS), and those who argue that the scale of the Internet requires a fresh approach. Traditionally, the cultural heritage domain has taken the lead in the development of descriptive metadata standards like MARC or Encoded Archival Description (EAD) and controlled vocabularies for subjects and the names of people, organisations and geographical entities. It is to be hoped that these can help influence future metadata developments.

Co-operation within and outside domain

Traditionally, metadata initiatives within the cultural heritage domain have focused on the three major sectors: museums, archives and libraries. In the past, each sector (or part of it) developed its own standards and those developed in one sector were rarely used in others. In fact, until comparatively recently, co-operation within a single sector was probably the exception rather than the norm. Libraries were the first sector to develop consistent standards that allowed the exchange of catalogue records between systems. This was led by the development of standards for information interchange (e.g., ISO 2709 and the MARC formats) and agreement on content rules (e.g. the Anglo-American Cataloguing Rules (AACR) and the International Standard Bibliographic Description (ISBD) series). The other sectors had less need for metadata sharing, but have since developed standards like ISAD(G) - the General International Standard Archival Description, encoding standards like the EAD and SPECTRUM, the UK museum documentation standard.

Of late there has been a considerable increase in co-operation within the cultural heritage domain. Part of this may be attributed to wider political or professional factors, e.g. in the UK, the Museums & Galleries Commission and the Library and Information Commission were merged in the year 2000 to form Resource: the Council for Museums, Archives and Libraries (<http://www.resource.gov.uk/>). However, perhaps a more significant trigger for increased collaboration and co-operation within the cultural heritage domain - especially over metadata - is a common involvement in things like digital content creation and the delivery of services through networks.

- *Digital content creation* - digitisation initiatives abound in libraries, archives and museums and, with limited resources available, it makes sense for organisations to share expertise and to use common standards wherever possible. So for example, Abby Smith of the US Council on Library and Information Resources (CLIR) has noted that once they are digitised, museum objects and library items are indistinguishable from one another. She says that "a digital *Blue Boy* [a painting by Gainsborough] and a digital *Huckleberry Finn* [a novel by Mark Twain] share the same behaviours, demand the same creation of metadata and management tools, reside on the same network, and are retrieved onto the same computer screen."
- *Cross-sector and cross-domain information retrieval* - the increased use of networks - and in particular the World Wide Web - has begun to blur the perceptions of users as

to where 'information' should be found. This has led to the development of systems that can search across the holdings of multiple institutions, across different types of information object, regardless of domain. An example of this might be the pilot 'Gateway' developed for the UK Arts and Humanities Data Service (AHDS), which acts as a 'union catalogue' of the digital holdings of AHDS service providers (<http://www.ahds.ac.uk/>). Such systems often make use of the products of cross-domain initiatives like DCMI or the Bath Profile of Z39.50 (<http://www.ukoln.ac.uk/interop-focus/bath/>). These initiatives are major focuses of cross-sector and cross-domain co-operation.

- *Digital preservation* - in the same way that digitisation initiatives help blur the differences between library, archive and museum objects, digital objects across all sectors (and domains) are subject to similar digital preservation problems. There is evidence of co-operation within the cultural heritage domain in projects like NEDLIB (<http://www.kb.nl/coop/nedlib/>), which included partners from both the libraries and archives sectors. Also, the UK Digital Preservation Coalition (DPC) has members from both within the cultural heritage domain (e.g., libraries, archives and the cross-sector Resource) and outside (e.g., publishers' organisations and research funders). In the preservation metadata field, there are few official links between initiatives like the Recordkeeping Metadata Forum (<http://www.archiefschool.nl/amf/>) and the OCLC/RLG Preservation Metadata Working Group (<http://www.oclc.org/research/pmwg/>), but this is an issue that may be addressed in the future. There is co-operation between the library-based preservation metadata developments and other domains, in particular with the space data communities who have been developing the *Reference Model for an Open Archival Information System (OAIS)* (currently ISO DIS 14721).

It is to be hoped that co-operation, both within and outside the cultural heritage domain, will become a persistent factor in metadata initiatives. It is likely that different metadata standards and systems will co-exist within the domain for the foreseeable future. There is no indication that the various cultural heritage metadata standards are converging on a single standard. Instead, the domain persists in demonstrating the virtues of diversity. This is because metadata collectively helps support a wide range of functions, e.g. resource management, resource discovery, rights management, preservation or the support of recordkeeping regimes. There are also fundamental differences of purposes between things like library cataloguing and archival description, which is based on securely linking records to their context and use - part of the principle of *respect des fonds*. Diversity need not be a serious problem as long as there is an awareness of interoperability issues and a willingness to work together, both within the domain and outside.

One area where there is scope for more co-operation is at the metadata framework level. Over the past few years there has been a growing interest in developing metadata standards based on more formal modelling techniques. Examples of such initiatives might be the IFLA Functional Requirements for Bibliographic Records (FRBR), the CIDOC Conceptual Reference Model (CRM) and the models underlying the Australian Recordkeeping Metadata Schema (RKMS). A more generic development is the ABC model that has been developed by the Harmony project (Lagoze & Hunter, 2001). The principles underlying both the RKMS framework and the ABC model show some level of 'event-awareness,' i.e. that they are less concerned with describing fixed objects (resources) but with the activities associated with them. Thus Lagoze (2000) could say that the new networked context "requires a model that can flexibly express the relationships between resources, abstract concepts, and multiple descriptions of those resources and concepts." Following its bicentennial conference in November 2000, the

Library of Congress action plan on *Bibliographic control of Web resources* (revised, 19 December 2001) includes a proposal to conduct research into how AACR2 might incorporate greater 'event-awareness' into the catalogue (<http://lcweb.loc.gov/catdir/bibcontrol/actionplan.htm>).

One recent initiative that is attempting to look at cross-domain issues is a project at the University of Toronto entitled "Modelling a 'Human Understandable' Metalevel Ontology for Enhancing Information Seeking on the World Wide Web" (<http://www.fis.utoronto.ca/special/metadata/>). This is looking at defining a metadata *lingua franca* for metadata. To get this started, the project has used existing crosswalks between USMARC/MARC21 and eight metadata standards (from various domains). These standards are the CIMI Z39.50 profile, DCMES, EAD, the FGDC Content Standard for Digital Geospatial Metadata, GILS, ONIX, TEI and the VRA Core Categories. Initial results found very little overlap between the schemes (almost 70% of the elements stood alone) and a wide variation in terminology (Howarth, 2001). Projects like this help demonstrate the problems of ensuring complete interoperability between different sectors and domains.

References

Howarth, L. (2001). *Designing a metadata enabled namespace for enhancing resource discovery in knowledge bases*. Electronic Resources: Definition, Selection and Cataloguing, Rome, 26-28 November 2001.

Lagoze, C. (2000). *Business unusual: how "event-awareness" may breathe life into the catalog?* Bicentennial Conference on Bibliographic Control for the New Millennium, Library of Congress, Washington, D.C., 15-17 November 2000: http://lcweb.loc.gov/catdir/bibcontrol/lagoze_paper.html

Lagoze, C. & Hunter, J. (2001). The ABC ontology and model. *Journal of Digital Information*, 2 (2): <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Lagoze/>

Appendix C: Educational domain

Correspondent: Erik Duval, Katholieke Universiteit Leuven, Belgium

Standardization work

IEEE LTSC LOM

Since 1998, the IEEE Learning Technology Standards Committee (LTSC) is developing the Learning Object Metadata (LOM) standard (<http://ltsc.ieee.org/>), based on original work in ARIADNE and IMS. The so-called “Base Schema” that defines the overall structure has been stable for the last 2 years or so, and the standardisation process is now nearing completion, with a scheduled approval of the standard in the first half of 2002.

The LOM specification has been implemented in numerous projects and consortia, including many of those mentioned below. It is expected that commercial adoption of LOM will be widespread, as some of the vendors are actively involved in the standardisation process, and as many more have started implementing support for LOM in their products.

Generally speaking, the LTSC is entering a consolidation phase: the number of working groups is being reduced, so that work can advance faster in a smaller number of groups. Some of the work items are being transferred to the ISO/IEC JTC1 SC36 committee.

CEN/CENELEC LTWS

The CEN/CENELEC Learning Technologies WorkShop was set up in 1999, to make sure that European interests would be defended in the culturally sensitive area of learning technology (<http://www.cenorm.be/iss/Workshop/lt/Default.htm>). Much of the work taking place in this context is related to internationalisation and localisation of LOM: the LOM document itself is translated in European languages, a profile for alternative language versions of a learning object is being developed, a repository of vocabularies and taxonomies is being set up, etc.

Other work in this organisation relates to “Educational Modelling Languages” (EML), quality assurance, educational copyright, and the description of language capabilities.

Generally speaking, the LTWS has now passed its initial exploratory phase, and is beginning to produce so-called “CEN Workshop Agreements” that complement the LTSC work.

ISO/IEC JTC1 SC36

Started in 2000, this committee on “Standards for: Information Technology for Learning, Education, and Training” has groups working on vocabulary issues, collaborative technologies, and learner information (<http://jtc1sc36.org/>). Both LTSC and LTWS maintain active liaisons with SC36.

Specifically with respect to metadata, SC36 has invited the LOM working group to submit its work “when deemed appropriate”. In a more general way, the subcommittee is still in an initial phase, deciding on the work it will undertake. The number of national committees that participate in the subcommittee (and that typically mirror the committee as a national focus point) is growing.

Consortia based work

ARIADNE

The ARIADNE Foundation started in 2000, as a more sustainable follow-up of a number of projects (<http://www.ariadne-eu.org/>). Central to its goal of “share and reuse” is a metadata infrastructure that is based on the so-called “Knowledge Pool System”, a distributed repository of learning objects with associated multilingual LOM descriptions.

IMS

The Instructional Management Systems (IMS) consortium issues a set of public specifications and associated documents (guidelines, implementation documentation, etc.) (<http://www.imsproject.org/>). Specifically for LOM, IMS developed a “best practices and implementation guide, as well as an XML Schema and a DTD (the latter jointly with ARIADNE).

ADL

Since 2000, the Advanced Distributed Learning (ADL) initiative has been developing the “Sharable Content Object Reference Model (SCORM) (<http://www.adlnet.org/>), which includes the LOM specification to address its metadata needs. Regularly, so-called “plugfests” are organized, where partners engage in practical interoperability experiments between Learning Management Systems and content.

DCMI-Education

This working group develops extensions to the Dublin Core Metadata Initiative (DCMI) element set (<http://dublincore.org/groups/education/>). Work has concentrated up to now on a domain-specific audience element and a qualifier for the relation element. The intent is to develop a vocabulary for the type element and to work on the description of educational objectives.

EDNA

Education Network Australia relies on the DCMI-Education specification (<http://standards.edna.edu.au/metadata/index.html>). EDNA has added its own elements for audience (now further developed by DCMI-Education), approver, category codes, reviews, and metametadata. The EDNA site gives access to metadata on some 15.000 learning objects. A toolset for metadata authoring is available.

EUN

The EUROpean Schoolnet has developed the “European Treasury Browser” (<http://www.eun.org/>). Metadata element sets have been developed for the level of individual resources and larger scale collections, based on the DCMI element set. An extensive thesaurus has been developed that describes many aspects of learning objects.

GEM

The Gateway to Educational Materials is a project of the ERIC Clearinghouse (<http://www.thegateway.org/>). Its metadata set is based on the Dublin Core set and includes a number of controlled vocabularies for subject, resource type, pedagogy, audience and format. With around 25.000 learning objects, it is one of the largest freely accessible repositories at this moment.

Cooperation within the domain of education

Within the domain of standardization, it can be expected that there will be stabilisation and consolidation around the finalized LOM standard: many implementations are currently based on slightly different versions of LOM. Most of the organisations involved have committed themselves to “upgrading” to the final LOM standard.

As LOM includes provisions for controlled extensions (by adding metadata elements, explicitly identifying new vocabularies or using the specific category for classification structures), it is expected that more globally acceptable such extensions will become increasingly popular. In fact, it is explicitly intended that such extensions would “percolate up” into revisions or new versions of LOM.

There is increasing adoption of so-called “application profiles” to accommodate community specific requirements within the framework as defined by a standard such as LOM: ARIADNE and ADL, for instance, have both defined such profiles, making some of the elements mandatory, restricting the value space to a subset of the one defined in the standard, etc. It is expected that communities will collaborate on common application profiles, or that, at the least, the sometimes rather subtle differences between communities will become more explicit through the mechanism of application profiles.

Cooperation with other domains

Increasingly, avenues for interoperability between LOM and DCMI are being investigated. The two organisations issued a Memorandum of Understanding (<http://standards.ieee.org/announcements/metaarch.html>). A “metadata manifesto” is being developed that presents the common views on modular, interoperable metadata. Work is ongoing on the development of a set of fundamental principles for modular interoperable meta-data. Application profiles that involve both LOM and DCMI metadata are being investigated. There is some promise that the emerging Resource Description Framework (<http://www.w3.org/RDF/>) will facilitate this process, although there is also some apprehension to make solutions too dependent on one particular technology.

Appendix D: Publishing domain

Correspondent: Laurie Causton, Clearbay Limited

Activity Report

Crossref (<http://www.crossref.org>) continues to grow, now with almost one hundred publishers on board, more than 5,600 journals and article records in the database approaching four million.

ONIX (<http://www.editeur.org/onix.html>), which now effectively incorporates EPICS, also marches on. Korea and Russia are the latest countries to establish ONIX implementation groups, and the French and German book trades are currently implementing. Development of ONIX for Serials is under way, and there is further activity in adding ONIX elements to provide information on the availability of subsidiary rights; this latter enhancement will allow publishers to communicate this information in standard formats.

ISWC, now approved as ISO standard 15707, now has its own home at <http://www.iswc.org> rather than being part of the CISAC site, although of course CISAC are still in charge.

The Metadata Information Clearing House (MICI), previously hosted by John Wiley, now has a new home at <http://www.metadatainformation.org>. This is a well-intentioned initiative which showed promise as a source of publishing metadata information, but it appears to continue to suffer from a lack of maintenance resources to keep it up to date, with the most recent entry dated March 2001.

Future outlook

A sound future for Crossref (<http://www.crossref.org>) seems assured, judging by the steady - and impressive - growth of this initiative. The planned expansion into other reference sources such as encyclopaedias, textbooks and conference proceedings will further consolidate its position.

The Crossref community recognises that one issue to be addressed is establishing an interface with the library community, specifically to help libraries incorporate article-level linking in their records. This community nevertheless has itself addressed ONIX (<http://www.editeur.org/onix.html>), with the Library of Congress mapping of ONIX to MARC 21 and the British Library's equivalent mapping to UNIMARC. ONIX is well positioned for the future; version 2.0, released last July, supports multimedia products such as e-books. ONIX for Serials will allow structured information at any level of granularity (serial, issue, article etc), with a single XML-based record structure and transmission standard. As an example of the co-operation now more evident between the various initiatives, EDItEUR is working closely with CrossRef, ISSN and publishers on this development.

The DOI (<http://www.doi.org>) is of course now in full commercial operation, Crossref being a prime example, and more registration agencies are progressively being added. The International DOI Foundation (IDF) has close working relations with a number of other bodies both within and outside the publishing sector. On the metadata front these include the World Intellectual Property Organization, the ISO TC46SC9 subcommittee (responsible for, among others, ISBN and ISTC), and <indecs>, whose work is now being developed jointly by the IDF and EDItEUR. Also, as part of its plans to offer a solid operational foundation, the IDF recognises the need to provide robust metadata tools for commercial operations.

<indecs> (<http://www.indecs.org>) itself has of course provided a basis for metadata standards in both ONIX and DOI whose communities are now involved in the development of indecs2, the rights data dictionary.

The ISBN community recognised some months ago that the standard needs to change, not least due to the advent of electronic books. However, as with any standards development, this will be a long process, with a target date set by the relevant working group of TC46SC9 (<http://www.nlc-bnc.ca/iso/tc46sc9/wg4.htm>) of January 2005. Metadata issues are being addressed; a committee of the International ISBN Agency (<http://www.isbn.spk-berlin.de>) is being formed to determine the core metadata set to be collected.

ISWC (<http://www.iswc.org>), now in its ISO incarnation, will continue to build its infrastructure. The ISWC itself is essentially a "dumb" number, and will rely on a network of decentralised databases as the repository for the descriptive metadata associated with any ISWC. It will also offer online searching and provide for the immediate allocation of ISWC numbers by agencies. No online database is yet available, and there is no doubt much work to be done - since the first ISWC issue in 1996 (for those who are interested, to Abba's "Dancing Queen"), more than two million numbers have been assigned.

The various other initiatives in the publishing sector will continue to develop.

The Open eBook Forum (<http://www.openebook.org>) incorporates metadata into its Publication Structure Specification, and also sees it as critical in the Rights & Rules area. Its Metadata Working Group is gathering and assessing requirements.

The next major objective for the ISTC (<http://www.nlc-bnc.ca/iso/tc46sc9/wg3.htm>) will be starting the process of choosing the international Registration Authority to run the system. Note also that the responses to the Committee Draft of the standard will all be in by the end of January 2002.

The IPTC's news-related initiatives - NITF (<http://www.nitf.org>) and NewsML (<http://www.newsml.org>) are both expected to evolve as more feedback is received from implementers and users, and more specialised content mark-up solutions for use in NewsML, like SportsML, are also anticipated.

In general, co-operation seems more in evidence now than when the Metadata Watch reports began. There is increasing recognition of this, as witnessed by the examples noted above, and by the proposal for a Metadata Interoperability Forum being made to the EC.