

# **SCHEMAS**

**Contract: N° IST-1999-10100**

**Forum for Metadata Schema Implementers**

**METADATA WATCH REPORT #6**

**STANDARDS FRAMEWORK # 3**

**D27 / D34**

**Document number:**

**SCHEMAS-PwC-WP2/WP3-D27/D34-Final-20011111**

**General Information**

<b>Title</b>	Metadata Watch Report #6 and Standards Framework Report #3
<b>Creator</b>	Christian Eilert, Makx Dekkers
<b>Subject-Keywords</b>	Deliverable D27/D34; WP2/WP3; Metadata Watch Report #6; Standards Framework Report #3
<b>Description</b>	This document comprises Metadata Watch Report #6 and Standards Framework Report #3 and domain reports for the Educational, Government, and Publishing domains
<b>Publisher</b>	PricewaterhouseCoopers
<b>Date</b>	11 November 2001
<b>Type</b>	Text Manuscript
<b>Format</b>	application/MSWord 2000
<b>Identifier-</b>	
<b>Document Number</b>	SCHEMAS-PwC-WP2/WP3-D27/D34-Final-20011111
<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"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:smes="http://www.schemas-
forum.org/registry/schemas/SCHEMAS/1.0/smes#">
<rdf:RDF rdf:resource = " ">
<dc:title> Metadata watch Report #6 and Standards Framework Report #3
</dc:title>
<dc:creator> Christian Eilert </dc:creator>
<dc:creator> Makx Dekkers </dc:creator>
<dc:subject> Deliverable D27/D34 </dc:subject>
<dc:subject> WP2/WP3 </dc:subject>
<dc:subject> Metadata Watch Report #6 </dc:subject>
<dc:subject> Standards Framework Report #3 </dc:subject>
<dc:subject> Metadata activity reports </dc:subject>
<dc:description> This report comprises Metadata Watch Report #5 and
Standards Framework Report #3 and domain reports for the Educational,
Government, and Publishing domains </dc:description>
<dc:publisher> PricewaterhouseCoopers </dc:publisher>
<dc:date> 2001-11-11 </dc:date>
<dc:type> Text </dc:type>
<dc:format> application/MSWord </dc:format>
<dc:identifier> SCHEMAS-PwC-WP2/WP3-D27/D34-Final-20011111 </dc:identifier>
<dc:language> en </dc:language>
<dc:rights> European Commission; External distribution via SCHEMAS web site
</dc:rights>
</rdf:RDF>
```

--

## **Table of Contents**

<b><u>1. INTRODUCTION</u></b>	<b>6</b>
<b><u>2. OBJECTIVE</u></b>	<b>6</b>
<b><u>3. METHODOLOGY</u></b>	<b>6</b>
<b><u>4. RESULTS</u></b>	<b>6</b>
<b><u>5. CONCLUSIONS</u></b>	<b>9</b>
<b><u>APPENDIX A: EDUCATIONAL DOMAIN</u></b>	<b>10</b>
<b><u>APPENDIX B: GOVERNMENT DOMAIN</u></b>	<b>13</b>
<b><u>APPENDIX C: PUBLISHING DOMAIN</u></b>	<b>19</b>

## 1. Introduction

This deliverable combines the sixth Metadata Watch Report and third Standards Framework Report from the SCHEMAS project.

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.

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 on the sectors Education, Government and Publishing sectors.

## 2. Objective

This sixth metadata watch report is different from the previous five in that it combines the metadata watch report with the standards framework report. Since considerable overlap has been identified between both in the past it was decided to merge the two reports to form one in the future.

## 3. Methodology

The Web sites of the activities that were described in the SCHEMAS Standards Framework Report #2 were visited. Recent news and announcements were used in the descriptions in section 4 below.

## 4. Results

Only the standardisation activities that had recent news on their Web sites are listed below. For a fuller description of these and other standardisation activities, please refer to the SCHEMAS Standards Framework Report #2. Fuller descriptions of some of the standards activities listed are included in the domain-specific appendixes.

### Models

#### *CIDOC CRM MODEL*

In July 2001, the CIDOC Conceptual Reference Model Special Interest Group published a document “CRM Scope Definition” as a proposal of the Steering Committee of the CIDOC CRM SIG.

URL: <http://cidoc.ics.forth.gr/>

Contact: Martin Doerr, [martin@ics.forth.gr](mailto:martin@ics.forth.gr)

### ***OAIS (Reference Model for an Open Archival Information System)***

In May 2001, the Twelfth International OAIS Workshop was held in San Francisco.

In June 2001, a new Draft Recommendation with version 2 of the Reference Model for an Open Archival Information System (OAIS) – the Red Book – was released on the OAIS Web site.

URL: [http://ssdoo.gsfc.nasa.gov/nost/isoas/ref\\_model.html](http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html)

Contact: Don Sawyer, sawyer@ncf.gsfc.nasa.gov)

### **Basic**

#### ***ISO 11179***

In May 2001, a final committee draft became available for ISO 11179-3 Registry Metamodel for ballot by national standards bodies until October 2001.

URL: <http://www.sdct.itl.nist.gov/~ftp/L8/11179/>

Contact: Douglas D. Mann, mannd@battelle.org

#### ***RDF (Resource Description Framework)***

RDF was discussed in the Semantic Web Track at this year's World Wide Web conference in Hong Kong in May 2001.

In September 2001, three new working drafts became available on the Semantic Web Activity's Web site: Refactoring RDF/XML Syntax, RDF Test Cases, and RDF Model Theory.

URL: <http://www.w3.org/RDF/>

Contact: Eric Miller, em@w3.org

### **Cross-domain**

#### ***Dublin Core***

One new DCMI Recommendation was released: version 2 of Using Dublin Core (June 2001). Three architectural documents were released as Proposed Recommendation in September 2001 for public comment until October 2001: The DCMI Namespace Policy, RDF/XML Expression of Simple Dublin Core and RDF/XML Expression of Qualified Dublin Core.

In October, NISO, the US National Information Standards Organization and the Dublin Core Metadata Initiative (DCMI) announced the approval by ANSI of the Dublin Core Metadata Element Set (Z39.85-2001) as an US National Standard.

In October, the annual Dublin Core meeting was held in Tokyo, Japan, with a workshop and a conference under the title "DC-2001: International Conference on Dublin Core and Metadata Applications 2001".

URL: <http://dublincore.org/>

Contact: Stu Weibel, [weibel@oclc.org](mailto:weibel@oclc.org)

### **Domain-specific**

#### **a) Publishing**

##### *ONIX (Online Information eXchange)*

ONIX Release 2.0 appeared over July and August. See the Appendix on the Publishing domain.

URL: <http://www.editeur.org/onix.html>

Contact: Brian Green, [brian@bic.org.uk](mailto:brian@bic.org.uk)

#### **b) Audio-visual**

##### *MPEG (Moving Pictures Expert Group)*

In July 2001, an overview of MPEG-21 was published, including a status report and a timetable for standardisation. At the same time, a document specifying MPEG-21 Requirements for a Rights Data Dictionary and a Rights Expression Language was issued, together with a Call for Proposals for a Rights Data Dictionary and a Rights Expression Language with a submission deadline of 21 November 2001.

URL: <http://www.cselit.it/mpeg/>

Contact: Rob Koenen, [rkoenen@intertrust.com](mailto:rkoenen@intertrust.com)

#### **c) Education**

##### *IMS*

In August 2001, the Ottawa communiqué (“Harmonization of Meta-Data for Education and Training Communities”) was released as a result of the Ottawa Metadata Summit with representatives from IEEE/LOM, IMS and DCMI. This is described in more detail in the Appendix on the Educational Domain.

Also in August, version 1 of the Digital Repositories White Paper was published. See the Appendix on the Educational Domain.

In September 2001, version 1.2.1 of the IMS Learning Resource Meta-Data Information Model was released, which corrects errata from the previous version.

URL: <http://www.imsproject.org/>

Contact: Kevin Riley, [kriley@eloki.fsnet.co.uk](mailto:kriley@eloki.fsnet.co.uk)

#### **d) Research, academic**

##### *CERIF (Common European Research Information Format)*

Due to limited resources at the European Commission for the follow up of the CERIF2000 initiative, the core implementation effort of the latter is now in the hands of EuroCRIS (the European Current Research Information Systems).

URL: <http://www.eurocris.org/>

Contact: Keith Jeffery, [K.G.Jeffery@rl.ac.uk](mailto:K.G.Jeffery@rl.ac.uk)

#### **e) Government, geospatial, environment**

*GILS (Global Information Locator Service)*

In June 2001, GILS published documents describing the GILS Topic Tree. One hierarchical sample is available as well as full alphabetic and hierarchical versions.

URL: <http://www.gils.net/>

Contact: Eliot Christian, [echristi@usgs.gov](mailto:echristi@usgs.gov)

## **5. Conclusions**

Many of the standardisation activities described above have been quite active over the summer months, releasing new versions of specifications and addressing longer-term strategic issues.

In the last year, many activities have started to think about their own future and the sustainability of their activities over a longer period of time. Many standardisation activities that have started out on the basis of enthusiasm from a small group of volunteers are now faced with the challenge to build sustainable organisations with sufficient funding to continue their tasks, thereby providing a stable basis for their growing user communities.

Perhaps partly because of this tendency towards finding a stable basis, and partly because of the growing interest in the Semantic Web vision, more and more initiatives are beginning to see the need for co-operation. Although the Semantic Web is far from realised and urgently needs to solve some of the technical issues (e.g. deployment of RDF), it indicates a strong long-term need for interoperability across various standards and applications.

In that context, many activities have started work on building metadata registries, such as in the Dublin Core and DOI. This issue will be taken on in the next Metadata Watch Report.

## Appendix A: Educational domain

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

### 1. Standardization work

#### 1.1 IEEE LTSC LOM

At this time, the IEEE LTSC (Learning Technologies Standardization Committee) LOM (Learning Object Metadata) group is focusing mainly on the ballot of the document that defines the so-called Base Scheme. This is the scheme that specifies the data elements, their value spaces and data types. The ballot of the document was successful, in that both sufficient ballots were returned, and that a sufficient number of them were positive. In order to maximize consensus, the 180 ballot comments were discussed, and a new version of the document underwent a so-called 'recirculation ballot'. In that process, both the number of positive ballots increased further, and some 75 new or follow-up comments were submitted. The working group is currently processing those comments and will probably go through at least one more recirculation ballot, after which the LOM specification will probably be finalized as a full IEEE standard.

→ Web: <http://ltsc.ieee.org/>

→ Next meeting: December 3-7, 2001, Hawaii, US

#### 1.1.bis The Ottawa Metadata Summit

The December 2000 memorandum of understanding (MOU) between the Dublin Core Metadata Initiative (DCMI) and the LOM Working Group entered a more 'operational' phase at a meeting in Ottawa, in August 2001. Agreement was reached to work on the following problems:

- Increasing the understanding in the broad community about the various functions of metadata, including cross-domain resource discovery (DCMI) and management and deployment of learning objects (LOM);
- Dispelling misplaced perceptions that DCMI and LOM are competing;
- Providing solutions to issues of technical interoperability between the DCMI and LOM schemas through the development and application of so-called application profiles (see earlier SCHEMAS reports).

#### 1.2 CEN/CENELEC LTWS

The CEN/CENELEC (Centre Europeen de Normalisation) ISSS (Information Society Standardization System) LTWS (Learning Technologies WorkShop) is working on internationalization and localization of the LOM specification. At the moment, first drafts are available for the LOM document in French, German, Spanish, Catalan and Italian. Related activities deal with the availability of alternative language versions of a learning object, through a small and targeted LOM profile and with the description of language capabilities of a user. More recent projects within LTWS are working on:

- Taxonomies and vocabularies (so as to increase semantic interoperability);
  - Educational Modelling Language (EML), which supports the arrangement of learning objects in educational meaningful constellations.
- Web: <http://www.cenorm.be/iss/Workshop/lt/Default.htm>
- Next meeting: October 3-5, 2001, Turino, I

## **2. Consortia based work**

### **2.1 ARIADNE**

The ARIADNE foundation is deploying version 3 of its Knowledge Pool System, a LOM based distributed repository of reusable learning objects. This version includes multilingual support for the currently balloted version of LOM. The associated course environment is also being upgraded: the new version supports deployment of learning objects from several knowledge pools.

- Web: <http://www.ariadne-eu.org/>
- Next meeting: November 21-24, 2001, Leuven, B

### **2.2 IMS**

The IMS consortium recently released a 'white paper' on digital repositories. That paper sketches the 'problem scope' of this new working group that will evaluate the appropriateness of existing or newly proposed specifications and standards. There are strong relationships with many metadata issues.

- Web: <http://www.imsproject.org/>
- Continuous meetings of various groups and subgroups

### **2.3 ADL & SCORM**

The ADL Sharable Course Object Reference Model (SCORM) is currently being updated to version 1.2, after the so-called 'plugfest' meeting of May-June 2001.

- Web: <http://www.adlnet.org/>
- Next Meeting: Plugfest 5, November 26-29, 2001, Orlando, Florida

### **2.4 EDNA**

There seems to be no recent evolution on the Dublin Core based metadata specification of the Australian EdNA organization.

- Web: <http://standards.edna.edu.au/>

## **2.5 EUN**

The European Treasury Browser (ETB) project of EUN is organizing a three-week on-line discussion on the Quality of Web-based Learning Resources. The ETB thesaurus is in alpha version: it includes 1053 descriptors in 5 languages.

→ Web: <http://www.eun.org/>

## **2.6 Gateway to Educational Materials**

The Gateway to Educational Materials currently contains 23,694 learning objects, accessible through a ('Bobby approved', i.e. 'accessible') web site.

→ Web: <http://www.thegateway.org/>

## Appendix B: Government domain

*Correspondents: Makx Dekkers, Christian Eilert, PricewaterhouseCoopers*

### **State of domain**

A specific characteristic of the government domain is that there are two main approaches, the Dublin Core approach (see: <http://dublincore.org/>) and the GILS approach (see: <http://www.gils.net/>). Additional approaches can be found in specific areas that can be considered part of the government domain, such as FGDC for geographic data.

In broad terms, the GILS approach is used in the US government domain, whereas the Dublin Core approach is used in many countries and regions in the rest of the world. Many countries in Europe and around the world have adopted the Dublin Core approach as their national standard for e-Government applications, or are in the process of doing so, e.g. the UK, Ireland, Denmark, Finland, Australia, New Zealand, and Canada. Apart from national programmes, the Dublin Core approach has also been adopted by trans-national organisations, such as the European Environmental Agency, the United Nations Environment Program and the Food and Agriculture Organisation of the United Nations.

A major development in recent months has been the drafting of a proposal for extensions to the Dublin Core for e-Government applications. This proposal was prepared in the months July through September 2001, and has been evaluated by the Usage Board of the Dublin Core Metadata Initiative. A revised draft of this proposal is expected to be finalised early in 2002.

### **Activities**

#### **1) European co-operation**

##### ***MIReG – Managing Information Resources for e-Government***

Up to the beginning of 2001, there was no co-ordinated approach to metadata for government applications in Europe. Various activities were taking place, e.g. in the UK, Ireland, Denmark and Finland, but the European Commission was reluctant to take the lead until there were clear signals from the member states that it needed to become involved.

In early 2001, these signals were beginning to appear, especially from the UK and Denmark. This led to a meeting that took place in Brussels on 21 and 22 June 2001 under the title “Managing Information Resources for e-Government”. This meeting was co-hosted by the European Commission’s IDA programme and the Dublin Core Metadata Initiative Government Working Group. Representatives from e-Government programmes, mainly from the Scandinavian countries, the UK, Ireland, Hungary and European institutions in Brussels, and several non-European participants came together to explore the possibility of developing a framework for the use of metadata across European governments and parliaments. The objectives were (a) to determine

the feasibility of such an activity and (b) to identify the resources - both technical and non-technical - that would be required to support it.

Common issues that emerged from the presentations and the discussions were:

- All activities represented had selected the Dublin Core as the basis for their metadata standard:
- All activities were looking to establish ontologies (controlled vocabularies, category lists, or thesauri) for use by government agencies, e.g. for subject classification, document types and geographic information, to increase interoperability between agencies;
- Most activities had as their first priority to look into issues related to disclosure and discovery of government information; in the future however, issues related to the management of information resources and issues related to the disclosure and discovery of government services needed to be considered.

At the end of the meeting, it was decided by the participants that it would be valuable to aim at the establishment of a EU Government Metadata Framework to help establish interoperability between government activities on a Europe-wide scale

Very soon after the meeting, the TAC (IDA Member States Committee) agreed the establishment of an IDA Work Item, under the name of MIREG (Management of Information Resources for e-Government). It was also agreed that IDA would set up a working group of national experts on metadata.

MIREG is now part of the revised IDA work programme 2001, covering the government metadata framework, vocabulary control and other encoding schemes, ontologies and topic maps, software interfaces and best practice guidelines.

→ Web: <http://ag.idaprogram.org/Indis35prod/doc/312>

## 2) United Kingdom

### *United Kingdom Cabinet Office, Office of the e-Envoy; UK Government Metadata Framework (UK-GMF)*

UK GovTalk is part of the implementation strategy for the e-Government Interoperability programme aimed at achieving seamless electronic government. The public sector will now begin using the same standards and descriptors, or metadata, to tag its information resources. This will help citizens locate documents and web pages. The extensions to Dublin Core for government are now under discussion in the Usage Board of the Dublin Core Metadata Initiative. The proposed extensions are based on the draft UK standard, with records management elements removed, and a few other changes. However, the UK standard (e-GMS) is not yet on the site since it is subject to considerable change.

The purpose of the site is to enable the Public Sector, Industry and other interested participants to work together to develop and agree policies and standards for e-government. This is achieved through the UK GovTalk consultation processes. The site covers all aspects relating to the e-Government Interoperability Framework (e-

GIF) and the e-Government Metadata Framework (e-GMF). It provides repositories for draft and agreed XML schemas, best practice and case studies, and also advice on toolkits and other relevant information. In addition, the site allows for Communities of Interest to be established and consultations to be conducted by those communities. These communities can relate to any aspect of introducing e-services into the UK public sector.

→ Web: <http://www.govtalk.gov.uk/>

### **3) Australia**

#### *National Archives of Australia*

Originally conceived as a variant of the U.S. Government Information Locator Service (GILS), the Australian Government Locator Service (AGLS) abandoned the GILS framework and instead based the online locator service on the Dublin Core metadata standard.

The AGLS standard was developed under the auspices of the Government Technology and Telecommunications Committee, a cross-jurisdictional committee of officials who reported to the Online Council of Ministers. The ongoing development of the AGLS relies upon continued cooperation amongst the National Archives of Australia as lead agency for the initiative, the Office of Government Information Technology, Online Council Officials and, most importantly, all of those experts inside and outside of government who have an interest in promoting online resource discovery.

It was recognised that the prime purpose of assigning AGLS metadata, namely enabling resource discovery and resource retrieval by authorised users, is also one of the requirements of a record keeping system. Hence, AGLS metadata assigned to records should theoretically be a subset of any standardised metadata set specified for electronic record keeping purposes.

The National Archives of Australia is currently developing a government record keeping metadata standard for Commonwealth Government use. It is envisaged that the draft government record keeping metadata standard that is currently being developed by the National Archives will dovetail neatly with AGLS. Such compatibility would ensure that at document creation, AGLS metadata could be captured as part of the record keeping metadata capture process.

→ Web: [http://www.naa.gov.au/recordkeeping/gov\\_online/agls/summary.html](http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html)

### **4) Canada**

#### *Canada Treasury Board Secretariat*

A major driver in the Government of Canada in giving access to government information is a strong commitment to citizen-centric service delivery. Clients want more precise search access to Canadian government Internet information, while departments need information management standards and guidelines to support

implementation of Common Look and Feel standards and Portal, Gateways and Cluster managers need standards and guidelines to support managing and sharing information across departments.

In 1995-1999, attempts have been made to define a Canadian implementation of GILS. This was never adopted as a Government Standard. In the context of the Common Look and Feel standard which evolved from the GILS work, all government Web sites must adopt five metadata elements for description of web resources: Title, Originator, Language of Resource, Date and Controlled Subject.

In the last two years, work has been going on to define Record Keeping Metadata Requirements expanded from Dublin Core and GILS.

Early adopters of metadata solutions in the Government of Canada are:

- Environment Canada (implemented GILS in 1995, converted to Dublin Core in 2000)
- Natural Resources Canada - geospatial metadata standards
- Cultural agencies - Dublin Core for digital content
- SchoolNet - Dublin Core for educational resources

Most departments are awaiting demonstrated use of metadata in government search engines, clearer guidance from the Treasury Board and adoption of a standard by the government.

In the course of 2001, the following proposals have been developed:

- Government of Canada Metadata Framework for Information Resources
- Adoption of Dublin Core as Government of Canada (GoC) core metadata standard for resource discovery
- Adoption of a Controlled Vocabulary Standard and GoC Core Subject Thesaurus

## **5) United States**

### ***National Biological Information Infrastructure***

The NBII is the biological component of the Administration's effort to develop a National Information Infrastructure (NII), a network of distributed databases and information sources on biological resources data and information sources. The goal of the NBII is to provide swift user access to biological databases, information products, directories, and guides maintained by Federal, State, and local government agencies, non-government institutions, and private sector organizations in the United States and around the world. Typically, users come from both the public and private sectors--scientists, planners, decision makers within the Department of the Interior (DOI) and other Federal agencies, State and local governments, industry, international entities, teachers and students, and private citizens.

The NBII Clearinghouse lets users search descriptive information (metadata) about various biological databases on a network of computers. A key to implementing the Clearinghouse has been the development of an accepted set of metadata standards by BRD and partners (see below).

BRD has developed a national metadata standard for documenting biological data and information. The standard is compatible with other data cataloguing standards, including the Federal Geographic Data Committee (FGDC) metadata standard for geospatial data and the U.S. Machine Readable Cataloguing (USMARC) standard for published materials, books, and journal articles. For a graphical representation of their metadata standard, see <http://www.nbii.gov/metadata/standards/current.status.html>.

→ Web: <http://www.nbii.gov/>

### ***National Environmental Data Index***

The overall goal of the NEDI, within the National Information Infrastructure (NII), is to facilitate the use of the widest possible range of environmental data and information to support the ability to protect human health, safety, and welfare; to maintain and restore ecological integrity; and to sustain economic stability and growth.

NEDI is intended to provide distributed access to existing environmental information locator systems that describe data holdings (metadata databases). Providing one-stop access to separate indexes for environmental data and information queries is a high priority task in the design and implementation of the NEDI, which will increase the effectiveness of the collective system.

The design leverages existing standards and proven technologies, supporting multiple metadata standards, using the Internet and other communication links, support distributed searches (Z39.50) and allowing multiple access points to NEDI. The design is intended to be compatible with agency systems and with issue-specific systems such as the Government Information Locator Service (GILS), the Federal Geographic Data Committee (FGDC) Clearinghouse, the Global Change Data and Information System (GCDIS), the Global Change Master Directory based on the Data Interchange Format (DIF) and the USMARC used in the library community.

→ Web: <http://www.nedi.gov>

### ***National Archives and Records Administration, Strategic Plan***

NARA is an independent Federal agency that helps preserve US history by overseeing the management of all Federal records. Over the ten years of this plan, they want to stem the losses already occurring in their nation's recorded history, prepare to document the 21st century fully and efficiently, and take advantage of new technologies to extend our rich resources to every office, school, and home.

Its mission is to ensure ready access to the essential evidence that documents the rights of American citizens, the actions of Federal officials, and the national experience. One of the goals is to "expand current efforts to build a nationwide, integrated online information-delivery system that educates citizens about NARA and its facilities, services, and holdings."

One such project is the NARA Archival Information Locator (NAIL), a pilot database of descriptions of various NARA holdings nationwide. In addition, some of the descriptions have links to digital copies of the documents being described. NAIL is the cornerstone of a larger Electronic Access Project through which they plan to develop an online catalogue of all of their holdings nationwide.

→ Web: <http://www.nara.gov/nara/vision/naraplan.html>

## **Appendix C: Publishing domain**

*Correspondent: Laurie Causton, Clearbay Limited*

### ***Current state of domain***

There has been perhaps less activity than usual since the last report, maybe in part due to the summer break, or perhaps that a number of the metadata initiatives have reached a level of maturity and more are now in, or entering, operational life.

CrossRef certainly is up and running. It enlarged its membership in July, with five new publishers who between them offer nearly three hundred journals, bringing the total membership to 77 publishers. The total number of CrossRef-enabled journals now stands at well over 5,000. Four new affiliates also joined.

The International DOI Foundation also has moved into commercial application, but remains active on the 'development' front, making progress on collaboration with other bodies in more than one area (more on this below).

Also, the IPTC have released NITF Version 3.0, with more intelligent handling of tables, a cleaner DTD, and improved metadata support.

ONIX Release 2.0 appeared over July and August. This is a major new release introducing coverage of electronic books, adding many new elements and codes, making structural changes to enable coverage to be extended more widely to non-book media, and facilitating the structured description of product content, including book tables of contents. Because of the speed of change in this area, the e-book aspects are being maintained and updated separately from the rest of Release 2.0.

Indecs2 (the rights data dictionary) is under way, funded by the record and film industries, Accenture, and Microsoft among others, with the IDF and EDItEUR/ONIX as partners.

### ***Overlaps and gaps identification***

The trend towards greater co-operation to minimise overlaps continues. The International DOI Foundation and the Content ID Forum both develop specifications for content identification and metadata to enable e-commerce and rights transactions for copyrighted information. In August, they made an agreement to collaborate on building an infrastructure for the management of digital intellectual property. Norman Paskin, Director of the IDF, and Hiroshi Yasuda, President of cIDf, stated: "Convergence rather than divergence of the two systems will benefit the wider community of users. The similarity in approach of these two major initiatives was recognized as ground to reconcile differences to create interoperable infrastructure. IDF and cIDf will share information on system development and work with open collaborations such as participation in MPEG-21 (Moving Picture Experts Group)."

The Annual IDF Meeting equally stressed the need for information exchange between the media industries; organizations from the publishing, music, television, library and

technology industries attended this meeting to share information on the development of policy and technical infrastructure for digital copyright management.

### ***Trends***

The idea of the 'processable' digital object, and by consequence the DOI, appears to continue to grow in significance; at the Annual IDF meeting, a number of very definite views of its future were expressed. Robert Kahn, founder of the Corporation for National Research Initiatives, explained how digital object approaches are vital to business and society in the information age, as is reconceptualizing the Internet from the movement of data packets to the management of information. Doug Armati, author of a 1995 report on Information Identifiers, endorsed the wider remit of DOI, calling in fact for widening the DOI Foundation's operations "to facilitate policy and organizational infrastructure globally that will result in DOIs being used to identify EVERY possible Digital Object -- not just in the media industries."

### ***Main issues***

While not strictly a metadata issue, a recent US court ruling is of interest, given the attention currently being paid in the publishing metadata domain to e-books and rights management issues. Random House recently took the digital bookseller RosettaStone to court, because they sold in digital format eight books that are published on paper by Random. The court however ruled that RosettaStone may sell these electronic versions, with the judge saying that Random's right to "print, publish and sell the work in book form" doesn't apply to e-books. This recalls the New York Times Supreme Court case over electronic rights where the court ruled that, unless specified in contracts, newspaper and magazine publishers do not automatically own the right to resell freelance contributors stories to digital database companies. Like some of those freelance contracts, the Random House book contracts were signed before e-books or the Internet were an issue. One view is that, if the RosettaBooks stance prevails, e-rights to thousands of old titles conceivably would become available.

### ***Special reviews***

The central bodies of all of the publishing metadata initiatives covered in Metadata Watch were contacted recently to enquire about status, achievements and the future. The commentary below is based upon the replies received.

### ***CrossRef***

The continuing progress of CrossRef has been noted earlier in this report, and the organisation feels that it has been quite exceptional in managing to involve a large number of leading publishers in a highly collaborative endeavour. Currently it is in the process of extending the XML schema to also accommodate book and conference proceeding records, and metadata is seen as very central to its mission of providing an infrastructure for citation linking.

The biggest challenge at present is interfacing with the library community, and specifically helping libraries, whose records do not usually contain article-level metadata, incorporate article-level linking. Another challenge is providing a flexible

interface when it comes to the type of metadata that can be accepted in queries to the system to retrieve DOIs.

CrossRef is actively involved in the setting of identifier standards - in particular, for DOI deposit, resolution, and extended services, all of which involve metadata, and in working with other standards organizations such as EDItEUR to achieve compatible metadata requirements across a variety of publication types.

Some overlap is recognised, with EDItEUR on the ONIX-for-serials front, and with the ISTC initiative on standard identifiers.

## **DOI**

The DOI has moved from a development concept to full commercial applications, with several Registration Agencies now appointed, such as CrossRef. There is strong interest from supporters, partners and liaison in US, Europe and Asia, and more links with other sectors and initiatives. Currently there are around 4 million DOIs, with 8 million resolutions in September 2001. Future plans focus on consolidation, promotion and collaboration:

- More marketing to get the DOI message out;
- Firming up, and scaling up, the operational foundations, and providing robust tools for commercial operations, such as metadata (namespace) tools;
- “Business development” – development of DOI interest and applications by and in other sectors and applications;
- Working with other activities which are part of the bigger picture of e-commerce, rights, etc., as much as possible by leveraging other efforts rather than by the International DOI Foundation doing it all themselves.

Metadata is seen as central to the DOI; its functionality is predicated on the concept of structured interoperable metadata framework, and is based on indecs concepts. In fact, a major achievement of DOI is seen as the bringing together of “the techniques of persistent actionable identification and structured metadata.”

There are nevertheless issues to be resolved:

- Funding: Standards must be developed for the long term, and voluntary effort alone is not enough. The current economic climate is “now biting.”
- Coverage: The IDF feels that coverage is not deep enough (text and technology sectors have recognised the need for DOI in part but some large companies are not participating), nor is it broad enough (non-text sectors should be involved, but while significant interest has been shown, this is not being matched by funding or participation).
- Perceptions: The IDF is seen as neither “a standards organisation” nor “a commercial solution”, while in fact it sees its main role as infrastructure creation,

for digital commerce of intellectual property, and such “killer plumbing” infrastructure is hard to sell; “killer applications are more easily understood.”

The IDF sees an increased recognition of the need for tools and techniques to deal with structured metadata, ranging from Semantic Web through to individual application offerings like Adobe’s XMP. In the case of the DOI, overlap with other metadata initiatives is deliberate, relating to “those metadata efforts which have adopted the same indecs-like view of metadata, such as ONIX ([www.editeur.org](http://www.editeur.org)).” At the same time, there is increasing recognition that several activities developed in different sectors may find it useful to interoperate and share techniques, tools and avoid re-invention of the wheel through mappings, common principles, etc.

## **ISTC**

The major achievement of ISTC is seen as the provision to the text supply chain of a tool for managing e-commerce and rights processes at the work level.

The Committee Draft of ISTC has been submitted to the Secretariat, and will be sent out for comments in the next few months. The next big step will be the process for choosing the international Registration Authority to run the system; setting up the international management structure and implementing the standard in the supply chain are the current major issues for ISTC.

Metadata is a central part of ISTC, with the submission of core metadata required before any ISTC is issued. Concurrent standards are recognised, in that the ISTC metadata scheme is a stream of ONIX.

Overlaps exist; the ISTC overlaps closely with the other related ISO initiatives – the ISWC music work code and the ISAN audiovisual work code – and it heavily references the ISBN system; with regard to the latter, it is expected that the two systems will be linked at the identifier level.

However, convergence is important, particularly with regard to interoperability of metadata, both for e-commerce and rights management.

## **NewsML and NITF**

NewsML is at version 1.0 and is being widely implemented, with a Schema version to be released shortly. NITF is now at version 3.0, with an enhanced table model, as noted earlier. Both are expected to evolve as more feedback is received from implementers and users. More specialised content markup solutions are also expected for use in NewsML, like SportsML.

The IPTC see the major achievement of NITF as opening the news domain to XML markup for text content in a non-proprietary way. NewsML is an open content management and exchange envelope that can be used across many information domains; it is now being used in the news and finance areas but other implementations are also being progressed.

There are issues to be addressed. NITF requires XML-aware systems, but these are becoming more available. NewsML is a complex standard with considerable power,

but needs expertise to implement as well as advanced XML-aware and configurable content management systems; nevertheless, these are just emerging onto the market.

Both initiatives rely on metadata to populate parts of the data structure. While NITF is more concerned with content markup, NewsML is a package for all types of data and is a carrier of rich metadata. The IPTC feels that NewsML has also made a useful contribution to the metadata scene - the TopicSet constructs that feature in NewsML are becoming accepted as a good way to store and use metadata.

Overlaps do not appear to be an issue; currently there do not appear to be any other XML-based open standards that compete directly with NITF and NewsML, although there are on the other hand opportunities for collaboration - the PRISM work on metadata for publishing could be made into TopicSets for use with NewsML.

Collaboration in fact is an objective to an extent; the IPTC is trying to work with "other parties who are prepared to make their work open to achieve convergence." Even so, this necessarily focuses on that which is relevant to its members' needs.

### **Open eBook Foundation**

The Publication Structure version 1.01 for Open eBook was released in July 2001, with version 2.0 in development.

On the metadata front, a Metadata/Identifiers Working Group has been established to assess requirements, evaluate relevant work by other standards organisations and recommend a solution to meet the requirements.

Metadata is seen as very important for all activities. It is nominally part of the Publication Structure Specification, and will also be critical to any standards in the Rights and Rules area.

The Publication Structure is seen as one of the major achievement of Open eBook, together with getting international companies and organisations together to discuss, plan and create for the benefit of the consumer and a competitive marketplace.

The current main issue is maintaining momentum and participation in an extremely uncertain economic and political time.

While there are no known significant overlaps between Open eBook and other standards or metadata initiatives, the avoidance of such overlaps is important.

### **Commentary**

The responses received reflected the trends which have been evident over the period of the Metadata Watch reports:

- Metadata, in the initiatives which the reports are monitoring, is more than purely descriptive; it is increasingly oriented towards rights management issues, and to e-commerce requirements in general.
- Collaboration and moves to address overlaps are increasing. At the very least, overlaps are recognised and taken into account, and in many cases their

avoidance is considered to be key. This is evidenced by the approach taken by Open eBook where, in evaluating metadata aspects, assessment of relevant work by other standards organisations is a central activity, and by the recent announcements by the International DOI Foundation and the Content ID Forum.