

SCHEMAS

Contract: N° IST-1999-10100

Forum for Metadata Schema Implementers

METADATA WATCH REPORT #2

D23

**Document number:
SCHEMAS-PwC-WP2-D23-Final-20000907**

General Information

Title	Metadata Watch Report #2
Creator	Elise Sfeir
Subject-Keywords	Deliverable D23; WP2; Metadata Watch Report #2
Description	This document comprises the introduction and top-level synthesis for D23 Metadata Watch Report #2 plus the domain reports
Publisher	PricewaterhouseCoopers
Contributor	Michael Day, Erik Duval, Annemieke de Jong, Elise Sfeir, Laurie Causton
Date	7 September 2000
Type	Text Manuscript
Format	application/msword
Identifier-URL	www.schemas-forum.org/folder/filename
Identifier-Document Number	SCHEMAS-PwC-WP2-D23-Final-20000907
Language	English
Rights	European Commission; Internal circulation within project; External circulation via SCHEMAS Web site

Dublin Core Metadata for this document

```

<META NAME="DC.Title" CONTENT="Metadata Watch Report #2">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#title">

<META NAME="DC.Creator" CONTENT="Elise Sfeir">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">

<META NAME="DC.Subject" CONTENT="Deliverable D23">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">

<META NAME="DC.Subject" CONTENT="WP2">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">

<META NAME="DC.Subject" CONTENT="Metadata Watch Report #2">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">

<META NAME="DC.Description" CONTENT="This document comprises the introduction and top-
level synthesis for D23 Metadata Watch Report #2 plus the domain reports">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#description">

<META NAME="DC.Publisher" CONTENT="PricewaterhouseCoopers">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#publisher">

<META NAME="DC Contributor" CONTENT="Michael Day">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#contributor">

<META NAME="DC Contributor" CONTENT="Erik Duval">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#contributor">

<META NAME="DC Contributor" CONTENT="Annemieke de Jong">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#contributor">

<META NAME="DC Contributor" CONTENT="Elise Sfeir">

```

```
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#contributor">
<META NAME="DC.Contributor" CONTENT="Laurie Causton">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#contributor">

<META NAME="DC.Date" CONTENT="(SCHEME=ISO8601) 2000-09-07">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#date">

<META NAME="DC.Type" CONTENT="Text.Manuscript">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#type">

<META NAME="DC.Format" CONTENT="(SCHEME=IMT) application/msword">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#format">
<LINK REL=SCHEMA.imt HREF="http://sunsite.auc.dk/RFC/rfc/rfc2046.html">

<META NAME="DC.Identifier" CONTENT="http://www.schemas-forum.org/folder/filename">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#identifier">

<META NAME="DC.Identifier" CONTENT="(SCHEME=URN) SCHEMAS-PwC-WP2-D23-Final-20000907">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#identifier">

<META NAME="DC.Language" CONTENT="(SCHEME=ISO639-1) en">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#language">

<META NAME="DC.Rights" CONTENT="European Commission; Internal circulation within project; External circulation via SCHEMAS Web site">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#rights">

<META NAME="DC.Date.X-MetadataLastModified" CONTENT="(SCHEME=ISO8601) 2000-09-07">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#date">
```

Table of Contents

1. INTRODUCTION.....	5
2. OVERALL REVIEW.....	5
3. DOMAIN REPORTS.....	7
3.1 AUDIO-VISUAL SECTOR.....	7
3.2 EDUCATIONAL SECTOR.....	9
3.3 ACADEMIC SECTOR.....	11
3.4 GEOGRAPHICAL INFORMATION SECTOR.....	15
3.5 PUBLISHING SECTOR.....	17

1. Introduction

This deliverable is the 2nd Metadata Watch Report. The first one has been submitted in July and gave many useful details on metadata initiatives in various market segments.

As already specified 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 things.

This second Metadata Watch Report completes the First Metadata Watch Report and therefore is less detailed. Actually, there are fewer things to report since the last few months.

The present deliverable is therefore a review report. Some correspondents have brought some additions to their own market segment, other did not bring any new addition. Therefore, only five domains have been explored here:

- The audio-visual sector
- The educational sector
- The academic sector
- The geographical information sector
- The publishing sector

2. Overall review

Little seems to have happened since the first Metadata Watch Report, a few months ago. Some new activities have been identified but no major changes in the domains have occurred.

The top-level synthesis that has been made in the First Metadata Watch Report is still up-to-date. The information provided here is only a small addition to this analysis.

The audio-visual sector is an update of D22. No real changes and innovation occurred in that sector. Three new initiatives have been identified:

- The ISAN (International Standard Audiovisual Number ISO 46/sc 9 N252) and the UMID (Unique Material Identifier), on digital identifiers.
- The AAF (Advanced Authoring Format) which facilitates the exchange of images, sound and metadata across platforms and applications.

In the educational sector, no new initiatives have been identified. The main standards-related activities are still the same: IEEE LTSC LOM, DC-Education, CEN/CENELEC ISSS LTWS and ISO/IECJTC1 SC36.

In all these activities, some substantial progress and developments have been made on the scope of the work. Around the four main standardisation activities, a large number of projects, consortia and other initiatives are actively developing specifications, tools and infrastructures.

In the academic sector, the correspondent updated on topics such as World Wide Web Consortium metadata developments, Internet information gateway initiatives and the specification of metadata for recordkeeping and digital preservation. A more detailed description is given on metadata standards developed for electronic literacy and linguistic texts (the Text Encoding Initiative header), e-print services (the Open Archives initiative) and the biological diversity information.

In this domain, the main trends are:

- The need towards developing standardised metadata schemes is based on a need to share metadata and interoperate.
- The successful academic sector metadata initiatives will need to be collaborative ventures between academic institutions and other organisations.
- It is becoming clear that short-term research and development projects will often need to evolve into something more sustainable.

For the geographical information sector, 3 main initiatives have been identified:

- The GI and GIS Interoperability project
- The Australian Spatial Data Infrastructure (ASDI)
- Building a 4D Geodynamic Model of Australia

In this report, the Australian initiatives have been identified. They did a lot of work in metadata related to geographical information. This report is mainly an update of the work regarding the geographical areas covered. In the first one, it was mainly focused on Europe and America. Here it is mainly focused on Australian activities.

Finally, in the publishing sector, no new initiatives have been identified but an update on the progress of the existing ones have been done. In this report, the correspondent detailed the advancement of some initiatives. He dealt with the digital domain, the identifier watch and the rights protection. Under each area he developed the advancement of the work: like the MPEG calling for proposals for Intellectual Property Management and Protection (IPMP) solutions for multimedia applications, the EBX Working Group which have released their latest draft technical specification, the development of the EPICS e-book metadata set which continues or the release by the INDECS project of the version 1.1 of the ONIX standard for the exchange of bibliographic metadata.

Development in the rights arena continues to focus on protection, ownership and intellectual property rather than rights transactions.

3. Domain reports

3.1 *Audio-visual sector*

Correspondent: Annemieke de Jong, Nederlands Audiovisueel Archief

1. REVIEW AND UPDATE OF ACTIVITIES

Three main initiatives have been identified.

- The AAF: Advanced Authoring Format

It is a joint initiative of a number of large media organisations and software companies. It aims at a new media industry standard file format, designed to meet information interchange needs. The media industry uses a wide range of source materials, as well as a set of highly varied capture tools with very different constraints (camera, keyboards, audio input sources, scanners). This wide variety leads to a great deal of time and effort spent converting media into formats, that can be used by the wide variety of authoring applications. AAF is a software implementation of SMPTE metadata and SMPTE labels, designed particularly to make it easy to work with large collections of interrelated sets of metadata and essence. AAF defines a base set of the built in SMPTE classes. These classes can be used to interchange a broad range of data between applications, but applications may have additional forms of data that cannot be described by the built-in classes. AAF provides a mechanism to define new classes that allow these applications. Besides the ability to format and manipulate metadata itself, the AAF software toolkit provides added capabilities for management of metadata sets, for user extensions and for pluggable modules. Authoring is being described as the process of creating multimedia content, including the related metadata. The AAF is specifically meant to function within this part of the production process and does not support the delivery phase of multimedia content (e.g. play-out, broadcast).

Expert opinion:

It is an important initiative with high industrial support. Picks up in a very constructive way on the work of the SMPTE Metadata dictionary. Currently standardising committees like the EBU-project group P/Meta are studying ways to relate to AAF.

- The ISAN: International Standard Audiovisual Number ISO 46/sc 9 N252.

The ISAN will distinguish one audio-visual work from the other. It is a unique number that is permanently assigned to an audio-visual work, it will identify that work across national boundaries and language barriers. It can be used in computerised applications, particularly those, which involve databases or the exchange of information about audio-visual works. Possible application domains: collecting societies to assist in the allocation of royalties; track the use of audio-visual works; anti-piracy purposes. For audio-visual works in digital form the ISAN will be embedded into the appropriate master copies of the work and transferred to subsequent copies made from that master. For analogue formats (film) ISAN should

be affixed to the master negative. ISAN is a voluntary identification number without legal implication or meaning. The number will be assigned by an International ISAN Agency, appointed by ISO. The entity or person to whom an ISAN is given should have the capacity to permanently attach or link that ISAN to the specific audio-visual work that it identifies.

Expert opinion:

It is an ISO standard. Mute number. It can only be attached to complete productions, but does not support granularity. Should have added value to other standards and proprietary identifiers in the digital environment. MPEG2 and MPEG4 standards for the coded representation of multimedia and audio-visual objects already provide a space for the ISAN identifier in the MPEG format. The SMPTE Metadata dictionary has defined ISAN as a metadata entry under the class Identifiers and Locators. The EBU Projectgroup P/meta is currently looking into the viability of ISAN as a broadcast identifier.

- The UMID: Unique Material Identifier (UMID)

A Unique material Identifier provides a reference for all captured audio-visual content units in a clip or shot, so that particular a content unit can always be located either locally, remotely or on a archive storage medium. The UMIDS provides for the link between the essence (video, audio, graphics, stills etc.) and the metadata and generates a time code and date (time-axis) for synchronising this data. UMIDs can also be used to track copyright information about the essence. UMIDs are automatically generated by the systems and support granularity up to the level of shots and video and audio frames. The UMID structure contains a set of components that each represents a key-aspect of the audio- or video essence. There are two sorts of UMIDs: the basic UMID, with the minimal components necessary for the unique identification (the essential metadata) and the extended UMID that provides information on the creation time and date, recording location and the name of the organisation and the maker (the signature metadata). Together these basic and extended UMID have a defined length of 64 bytes.

Expert opinion:

Standard status. Intelligent number that in itself includes some of the most important metadata about a content clip. Important identification standard for the audio-visual production environment. Supports granularity (important in the process of creation and authoring). Provides a link between metadata and essence. High industrial support by some of the most important manufacturers of audio-visual equipment and media-management systems. Already included in most of the (forthcoming) metadata standards for the audio-visual production environment.

3.2 Educational sector

Correspondent: Erik Duval, Katholieke Universiteit Leuven

1. REVIEW AND UPDATE OF ACTIVITIES

The four main initiatives in the domain of education, identified in the report for April, were:

- a. IEEE LTSC LOM
- b. DC-Education
- c. CEN/CENELEC ISSS LTWS
- d. ISO/IECJTC1 SC36

These are still the main standards-related activities. Little seems to have happened over the last three months in both DC-Education and ISO/IECJTC1 SC36.

The IEEE LTSC LOM group held a meeting in Montreal in June, where substantial progress was made, mainly on the scope of the work to be undertaken for a 'version 1.0' standard, and on the proper notation for the specification. Both these issues were resolved with a large degree of consensus, and it seems like the document may be forwarded for ballot in the Fall of 2000.

The CEN/CENELEC ISSS Learning Technologies Workshop accepted a final workplan with recommendations for further work to be funded by the European authorities. A substantial part of the proposed work is concerned with metadata, including work on vocabularies and taxonomies, profiles, bindings and internationalisation of metadata. The second phase of this workshop will be launched in October, at an open meeting in Brussels.

Around the four standardisation activities mentioned above, a large number of projects, consortia and other initiatives are actively developing specifications, tools and infrastructures. As was mentioned in the previous report (April 2000): 'There is also considerable overlap in the work of the different consortia (ADL, AICC, ARIADNE, IMS, etc.) that work on actual implementations of the metadata related and other standards. This does not seem to be a problem, the less so as all these consortia contribute to the development of the standards, and then adapt them to the needs of their constituencies, a process referred to as 'profiling' in the standards world. In fact, having several independent implementations of the metadata standards is a good thing, as it increases the probability that problems or shortcomings will be identified early on. It would be most useful though to have some interoperability development taking place now that the specifications and implementations are maturing, so as to prove that the standard does indeed serve its ultimate goal.'

Below, we briefly review some recent developments in a number of the more relevant such organisations:

- The ARIADNE project (1995-2000) recently transformed into a permanent Foundation, which should become self sustainable, after some initial support from the European Commission. One of the most important aims of the Foundation is to

organise the further development and exploitation of the Knowledge Pool System, a distributed database of reusable learning components, with associated metadata that describe them. The ARIADNE infrastructure includes tools for indexation, query, course building, etc. ARIADNE and IMS (see below) jointly submitted their metadata specification to the IEEE LTSC and continue to collaborate on this issue.

- The IMS 'global learning consortium' released in May a version 1.1 of its metadata specification, a minor update to the previous version (changing all element names to lower case) 'conducted in close cooperation with the IEEE Learning Objects Metadata working group'. Also relevant in this context is the IMS Content Packaging specification, which enables the combination of learning materials into interoperable, distributable packages, and the IMS Content Management specification, which lays out the management and run-time interactions required of learning content.

- The Advanced Distributed Learning initiative was launched in 1997 by the US Department of Defense. In January, ADL released version 1.0 of the 'Sharable Course Object Reference Model (SCORM)'. This reference model includes an XML based Course Structure Format, an API for the runtime environment, and different profile of the IEEE LTSC LOM. Since January, developers have started to implement the model, which culminated in the so-called plug-fest in May. In parallel, a test suite for SCORM is being developed. ADL explicitly refers to IMS, and IEEE LTSC.

- MICROSOFT: In February, Microsoft released the Learning Resource iNterchange (LRN) toolkit. LRN is an XML based content interchange specification, based on the content packaging specification developed by IMS. An 'LRN-Viewer' can be downloaded. This software relies on XSL to present an LRN resource as a DHTML document, with a navigation panel to the left, and the content itself in a right panel. Microsoft explicitly refers to IMS, ADL, IEEE LTSC and ARIADNE.

3.3 Academic sector

Correspondent: Michael Day, UKOLN

1. CURRENT STATE OF DOMAIN

The scope of the first academic domain metadata watch report included World Wide Web Consortium metadata developments, Internet information gateway initiatives and the specification of metadata for record keeping and digital preservation. This second academic domain report will (briefly) update one of these topics, but will describe in more detail metadata standards developed for electronic literary and linguistic texts (the Text Encoding Initiative header), e-print services (the Open Archives initiative) and the more complex topic of biological diversity information.

2. MAIN ISSUES

1. Digital Preservation

The collaboration between the Research Libraries Group and OCLC on digital preservation has resulted in the creation of a small, but international, Digital Archive Attributes Working Group. This working group started work in June 2000 and initially will review the OAIS Reference Model, the various preservation initiatives of the National Library of Australia and the British Library, and the outcomes of research projects like NEDLIB and Cedars.

2. Literary and linguistic texts

Humanities scholars have used electronic texts in their research since the 1950s, when Roberto Busa began to compile a word index and concordance to the complete works of Thomas Aquinas. Over the succeeding years, scholars have created a large quantity of electronic texts - both for literary and linguistic research. Electronic text centres, e.g. the University of Virginia Electronic Text Center and the Oxford Text Archive, have been set up to serve this constituency of scholars, and to encourage the long-term retention and reuse of texts. It has long been realised that the easy reuse of electronic texts is dependent upon the application of a standardised encoding (or markup). Defining this standardised encoding has been the main goal of the Text Encoding Initiative (TEI).

The TEI is an ongoing collaborative research effort concerned with developing generic guidelines for the representation of textual materials in electronic form in order to facilitate the preparation and interchange of electronic texts for scholarly research. Since 1990, the TEI has published several editions of its "Guidelines for Electronic Text Encoding and Interchange", the encoding scheme of which is formulated as an application of the Standard Generalized Markup Language (SGML). TEI-conformant Document Type Definitions (DTDs) can be built up from the tag-sets documented in the Guidelines.

Metadata - typically bibliographic-type information - is stored in the TEI header. The elements in the header are divided into four parts: file description (where most of the

bibliographic information would be stored), encoding description, text profile and revision history. The file description element of the TEI header was broadly based on library cataloguing principles (e.g. the International Standard Bibliographic Description (ISBD) series) but did not mandate any particular content rules. In practice, however, electronic text centres creating TEI headers have often developed their own cataloguing guidelines. These tend to be compliant with library standards like the ISBD(ER) - the ISBD for electronic resources - and the 2nd edition of the Anglo-American Cataloguing Rules (AACR2). A good example of such rules is included in the University of Virginia Library's "Cataloging Procedures Manual".

A membership consortium known as the TEI Consortium is now responsible for the development, maintenance and promotion of TEI. This consortium will be responsible for ensuring the future sustainability of the TEI (e.g., through developing training and consulting services) and for ensuring that the TEI Guidelines have a role to play in the development of XML-based tools.

3. E-print archives

Another metadata-related initiative with its roots in the academic sector is the interoperable system being developed by the Open Archives initiative (OAI). The Open Archives initiative is working towards the development of a universal service that will give access to author self-archived scholarly literature (e-prints). Following the early example of the 'e-print archive' hosted by the Los Alamos National Laboratory (now called arXiv.org), a large (and growing) number of e-print services have been set-up and they increasingly form an important part of scholarly communication in some disciplines.

The Open Archives initiative aims to offer interoperability between these – quite diverse - e-print services. A meeting of the OAI in Santa Fe resulted in the publication of the "Santa Fe Convention", a set of interoperability agreements that are intended to aid the creation of e-print 'mediator services'. The Convention recognises the existence of format diversity but suggests that interoperability will depend upon the existence of a shared format for exchanging metadata. The proposed basic metadata set is called the Open Archive Metadata Set (oams) - the semantics of which has deliberately been kept simple in the interest of easy creation and widest applicability.

4. Biological diversity information

A different area where there is a pressing need for interoperability and standardisation is that of biological information. This need has been emphasised by the growing perception of the importance of biological diversity (biodiversity) and the creation of Internet-based services like the Clearing House Mechanism (CHM) of the United Nations Convention on Biological Diversity (1992).

One of the challenges of integrating biological information is that there are many different types of it. For example, curatorial institutions like museums, herbaria, botanical gardens and zoological gardens would typically have large amounts of descriptive data about biological specimens and artefacts. There is no single standard for these descriptions and they are often not in any machine-readable form, e.g. they could be ancient hand-written labels stuck on glass jars or placed next to insects on pins. Other information is concerned with geographical distribution, biological

nomenclature or publications. One key to any proposed biodiversity 'database' will be the digitisation and integration of this wide range of information types.

Digitising and integrating all of this information, however, will take a very long time. Even finding out what has been previously described is difficult - there is currently no master inventory of all of the species that have been described by taxonomists. Progress has been made by biologists in developing particular areas; e.g. the creation of nomenclatural databases like the International Legume Database & Information Service (ILDIS) or the Missouri Botanical Garden's VAST (VAScular Tropicos) database.

A variety of standards have been developed to support the interchange of biological information. These include the Association of Systematics Collections (ASC) Reference Model for Biological Collections and the Herbarium Information Standards and Protocols for Interchange of Data (HISPID). The DELTA (DEscription Language for TAXonomy) format can be used to record taxonomic descriptions and has been adopted as a standard for data exchange by the International Taxonomic Databases Working Group.

Generic approaches to taxonomic information have been taken by services like the Index to Organism Names (maintained by BIOSIS), the Integrated Taxonomic Information System (ITIS), and by the Species 2000 programme. Further developments might be facilitated by the 1999 agreement of the OECD's Global Science Forum (previously the Megascience Forum) to create a Global Biodiversity Information Facility (GBIF).

Smaller-scale progress is being made in two important areas: developing standardised metadata formats for biological information; and the Species 2000 programme - the production of a uniform and validated index of the names of all known species.

Some progress in the metadata format area has been achieved by the agreement of the Federal Geographic Data Committee's (FGDC) Biological Metadata Profile. This is an enhancement of the FGDC's Content Standard for Digital Geospatial Metadata (CSDGM). It includes all of the CSDGM elements but adds other elements that can be used to document biological information about taxonomy and nomenclature. Metadata created according to this profile can be added to biological metadata clearinghouses like the US NBII (National Biological Information Infrastructure) Metadata Clearinghouse.

The Species 2000 programme aims to create an index to all of the world's known species that could be used as a tool in inventorying and monitoring biodiversity worldwide. It is an international initiative, based on a federation of existing taxonomic databases who will create a range of global species databases that will cover all of the major groups of organisms. The Species 2000 programme is likely to take a long time. In the interim, regional approaches to indexing species are being developed, of which ITIS (a partnership of U.S., Canadian, and Mexican agencies) covers North America. In the UK, the National Biodiversity Network (NBN) will be developing relevant data (and metadata) standards and a dictionary of species.

3. TRENDS

Some trends can be identified:

First, to state the obvious, much of the impetus towards developing standardised metadata schemes is based on a need to share metadata or otherwise interoperate. Previous standardisation, however, has tended to be based within one particular sector; e.g. the sharing of MARC records between libraries. Increasingly, there is a need for some standardisation of data across sectors. Good examples are the way in which electronic text centres have developed content rules for TEI headers based on existing library standards and the interaction between biological and geographical information made possible by the development of the FGDC Biological Metadata Profile.

Secondly, successful academic sector metadata initiatives will need to be collaborative ventures between academic institutions and other organisations. This will be especially important in the biological diversity information area, where universities, curatorial institutions, government agencies, supra-national organisations, commercial database providers, professional societies, etc. will all have an interest in the development of standards.

Thirdly, it is becoming clear that short-term research and development projects will often need to evolve into something more sustainable. A good example of this is the development of the TEI from a development project funded by professional societies and other funding organisations into a non-profit membership consortium.

3.4 Geographical information sector

*Correspondent: Elise Sfeir, PricewaterhouseCoopers
With contributions of Roger Longhorn*

1. CURRENT STATE OF DOMAIN

Not much can be added in this report since the last correspondent report made 3 months ago.

The previous list of activities have been extended with 3 very interesting GI metadata activities:

- The GI and GIS Interoperability Project
- The Australian Spatial Data Infrastructure (ASDI)
- Building a 4D Geodynamic Model of Australia

The previous report focused mainly on American and European projects. This time, Australian initiatives are at the center because they are really active in that domain.

The ASDI initiative is interesting because they have a metadata working group who developed guidelines on metadata as well as a metadata entry tool. The guidelines provide details for the recommended metadata elements. This working group can play a key role in centralising or unifying the GI metadata schemas.

"GI and GIS Interoperability Project" of the JRC's Space Application's Institute (SAI), is also essential in the GI metadata field because it is participating in several other GI-related / metadata-related projects, such as ETeMII. In fact, the JRC's GI/GIS Project will probably have a bigger impact in unifying the European GI community around GI standards, including those for metadata, than will separate projects like ETeMII, because the JRC is an actual Directorate General of the EC and has long-term funding available, that is not linked to any specific calls, etc. from FP5.

2. MAIN ISSUES

First, Rick Pearsall, FGDC Metadata Co-ordinator (Federal Geographic Data Committee), sent an announcement on June 9, 2000 to say that the International Organisation for Standardisation (ISO) Technical Committee 211 (TC211) has just released for review the latest version of the ISO Metadata Standard (19115.3).

Numerous organisations plan to use the ISO Metadata Standard once it has been approved by the ISO Standards Committee. The FGDC is committed to harmonise the CSDGM with the ISO Metadata Standard and has numerous activities ongoing to assure that the harmonisation is successful. To protect the significant already existing metadata investment, it is important to assure that the proposed ISO Metadata Standard allows the maximum compatibility with existing FGDC compliant metadata records.

To establish a FGDC position on the ISO Metadata Standard, the FGDC is conducting a registered review of the ISO Metadata Standard Committee Draft (Version 3).

Anyone interested in helping to establish the FGDC position by reviewing and commenting on the ISO Metadata Standard and/or by helping to adjudicate the comments, the FGDC is interested in hearing from them. The registered review was closed on August 1, 2000.

For more info: http://www.fgdc.gov/metadata/iso_regrvw_v3r.html

Some members of the GI community fear that Dublin Core as a general metadata proposal for digital geospatial data is inadequate. Apparently it would be inadequate for geospatial documentation and does not support catalogue searching better than full-text search in its current form. This is because the Dublin Core elements, or metadata "fields" are general purpose containers into which can be put any text content. Although this is interesting for text fields like the equivalent of title and author, it fails completely on important fields such as Coverage which is meant to contain both temporal and spatial extent.

Furthermore, people use Dublin Core in their own way which is not always interoperable because of a lack of guidance.

This unsatisfaction of some part of the GI community is unfortunate because it will discourage other GI people to use Dublin Core, when it is actually possible to use it together with other metadata schemas creating thus application profiles.

3.TRENDS

The cartographic/mapping agency sectors of the much wider "GI community" are aware of "metadata" - but as Roger Longhorn said, his direct and personal experience of attending conferences and speaking at meetings of sectoral communities who use GI, such as agriculture and coastal zone managers/marine community, is that they are very unaware of metadata or are only now (in 2000!) becoming aware of its importance.

Much of the real value of GI resides in/with these sectors and not in the mapping/cartographic producers of GI! Therefore, their metadata experience is just only beginning.

4.OVERLAPS AND GAPS

The gaps are still the same as identified in the correspondent report 1, that is that there are no real development of metadata for the Web. Too many projects are dealing with metadata catalogues, queries and indexing. Furthermore, projects mainly develop metadata for their own use.

3.5 Publishing sector

Correspondent: Laurie Causton, Clearbay Limited

1. UNITED WE STAND, DIVIDED WE STAND?

The diversity of initiatives and schemes continues, some sharing common ground but with different objectives, and so the scene remains quite complex. While it is true that some initiatives indicate that they take into account others with some degree of commonality, taking something into account can simply mean that it is considered, then discarded. How much convergence and interoperability will result in the longer term remains to be seen.

Some more central and established developments seem to be more generally embraced; Dublin Core, for example, finds itself in the Electronic Book Exchange (EBX) metadata proposals, in PRISM, and in SantaFe. Of course, whatever its strengths and weakness, DC has the useful advantage that it is already here, which has an undeniable attraction for any initiative looking for a structure for its metadata. The DOI appears to continue to make headway, perhaps not surprisingly in view of its more commercial orientation; business interests have a way of promoting progress.

2. THE DIGITAL DOMAIN

As one of the more visible growth areas in publishing, the digital world has seen a good level of activity.

MPEG are calling for proposals for Intellectual Property Management and Protection (IPMP) solutions for multimedia applications, evidence perhaps of the current focus in the rights domain on content protection and delivery; as noted in the previous report - more on this later. And shortly to kick off is the grandly named CUIDADO (Content-based Unified Interfaces and Descriptors for Audio/music Databases available Online - which surely merits an award for an inventive acronym), an Esprit project co-ordinated by Ircam and aiming to employ the MPEG7 standard.

There is certainly growth in the emerging electronic book market. The EBX Working Group have released their latest draft technical specification, which introduces a Dublin Core-based format for what they term "concise" metadata, "used for such purposes as identifying available and preferred content formats, logging a transaction, identifying an e-book in an error message, and displaying the progress of downloading an e-book." For "extended" metadata, "describing an e-book on a bookseller's Web site and cataloguing an e-book in a library," they nod towards ONIX and MARC.

Development of the EPICS e-book metadata set continues, and America's Book Industry Study Group is starting to take on the challenges ahead, setting itself the task of developing standards for emerging technologies. "Electronic books, metadata, and XML are poised to change the book industry," comments the BISG. The work to be done will be co-ordinated by a newly appointed Executive Director and, although not

yet announced in detail, appears once again to be focussed at least in part on rights, noting "the increased emphasis on standards to both describe the content of ... transactions and to protect the intellectual property rights of the creators."

Equally, the US ISBN Agency has recently issued a paper discussing the role of the ISBN in the digital world, having already identified a number of formats of digital editions which should receive ISBNs. Moreover, while there has been no formal metadata for ISBN, but in its place for some time now a form of *de facto* metadata context, the paper reminds us that the International ISBN Agency is currently establishing the core metadata needed for both print-on-paper and digital ISBN products, in conjunction with EDItEUR International.

3. IDENTIFIER WATCH

Identifiers remain closely linked to many emerging metadata initiatives, and vice versa. Understandable perhaps, given the importance of an effective identification of the object or resource to which the metadata will apply - how do you describe something if it is not clearly identified? In this way, the identifier can become an integral part of the metadata element set. And conversely, some developers of identification schemes are recognising the importance of attached metadata.

Intelligent identifiers, of course, can be said to embed a certain level of metadata by virtue of their structure. Even so, the ISBN community, already considering metadata needs as noted above, have been discussing the removal of both the country/regional and publisher identifiers from the ISBN, thus turning it into a completely "dumb" number. In this eventuality, the corresponding metadata structure would grow in significance.

The draft standard for the International Standard Textual Work Code (ISTC) has been issued, for discussion at the working group's next meeting in September. This initial draft indicates that the proposed International ISTC Agency will have among its responsibilities the maintenance of a central database of ISTC numbers and their accompanying metadata. That said, as yet there is no definition of what form this metadata will take, although indications are that the EPICS Data Dictionary might be involved, but these are early days; the draft in fact focuses more on administrative issues, and has yet to address even the ISTC format.

A call has been issued for a 5-year evaluation of the Serial Item and Component Identifier (SICI). The structure and syntax of the SICI remain problematic and may be an obstacle to useful implementation, and in any event it is possible that part of its target area of application - article and article/component level - may be usurped by the DOI. NISO have released a draft standard for trial use of its cousin, the Book Item and Component Identifier (BICI), with the review period ending in January 2002.

4. RIGHTS - PROTECTING ONE'S OWN

Development in the rights arena continues to focus on protection, ownership and intellectual property rather than rights transactions. This is perhaps an inevitable result of the rapid proliferation of digital dissemination and the consequent response of the commercial, profit-oriented, parties involved - witness the recent MP3 saga.

While relevant developments continue, such as MPEG's IPMP and XrML, recent activity seems more concerned with combating piracy. Presumably mindful that an MP3-style revolution could happen in the incipient e-book industry, the Association of American Publishers (AAP) are to collaborate with Microsoft "on a broad educational and enforcement initiative" to fight e-book piracy. Equally a glance at the current home page of the IFPI's Web site might lead one to believe that piracy is the only issue on the agenda.

5. SO WHAT ELSE IS NEW?

The beta of version 1.0 of NewsML, the XML-based standard for news creation, storage and delivery, has been released by the IPTC.

The DOI-based CROSSREF service has "gone live", aiming to provide persistent, efficient reference links between citations in articles in scientific, technical and medical (STM) journals. Membership has grown to thirty-three publishers and the metadata database now embraces more than one million article records.

The INDECS project has now formally closed, but the work does not stop there. The partners have established a not-for-profit membership organisation, called the Indecs Framework Ltd, to carry forward the work. Close co-operation with the International DOI Foundation (IDF) is planned.

One of the most significant developments arising out of INDECS, the ONIX standard for the exchange of bibliographic metadata, has been released in version 1.1, receiving early endorsement from major online suppliers such as Amazon and Barnes&Noble.