

SCHEMAS

Contract: N° IST-1999-10100

Forum for Metadata Schema Implementers

METADATA WATCH REPORT #3

D24

Document number:

SCHEMAS-PwC-WP2-D24-Final-20001120

General Information

Title	Metadata Watch Report #3
Creator	Makx Dekkers
Subject-Keywords	Deliverable D24; WP2; Metadata Watch Report #3; Application profiles
Description	This document comprises the introduction and top-level synthesis for D24 Metadata Watch Report #3 plus the domain reports
Publisher	PricewaterhouseCoopers
Contributor	Michael Day, Erik Duval, Annemieke de Jong, Elise Sfeir, Laurie Causton, Walter Koch, Tom Baker, Rachel Heery, Manjula Patel
Date	20 November 2000
Type	Text Manuscript
Format	application/msword
Identifier- Document Number	SCHEMAS-PwC-WP2-D24-Final-20001120
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 #3">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#title">

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

<META NAME="DC.Subject" CONTENT="Deliverable D24">
<META NAME="DC.Subject" CONTENT="WP2">
<META NAME="DC.Subject" CONTENT="Metadata Watch Report #3">
<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 D24 Metadata Watch Report #3 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">
<META NAME="DC Contributor" CONTENT="Erik Duval">
<META NAME="DC Contributor" CONTENT="Annemieke de Jong">
<META NAME="DC Contributor" CONTENT="Elise Sfeir">
<META NAME="DC Contributor" CONTENT="Laurie Causton">
<META NAME="DC Contributor" CONTENT="Tom Baker">
<META NAME="DC Contributor" CONTENT="Rachel Heery">
<META NAME="DC Contributor" CONTENT="Manjula Patel">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#contributor">

<META NAME="DC.Date" CONTENT="(SCHEME=ISO8601) 2000-11-20">
<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">
<META NAME="DC.Identifier" CONTENT="(SCHEME=URN) SCHEMAS-PwC-WP2-D24-Final-
20001120">
<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) 2000-11-
20">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#date">
```

Table of Contents

1	INTRODUCTION	5
2	THE CONCEPT OF APPLICATION PROFILES	5
3	THE USE OF APPLICATION PROFILES	7
4	THE FUTURE OF APPLICATION PROFILES.....	7
5	DOMAIN REPORTS.....	8
5.1	INDUSTRY SECTOR.....	8
5.2	PUBLISHING SECTOR.....	11
5.3	AUDIO-VISUAL SECTOR	13
5.4	CULTURAL HERITAGE SECTOR	16
5.5	EDUCATIONAL SECTOR	18
5.6	ACADEMIC SECTOR	19
5.7	GEOGRAPHICAL INFORMATION SECTOR.....	21
6	REFERENCES	28
	APPENDIX A. APPLICATION PROFILES FOR AUDIO-VISUAL SECTOR.	29
	APPENDIX B. APPLICATION PROFILES FOR EDUCATIONAL SECTOR.	30
	APPENDIX C. APPLICATION PROFILES FOR GEOGRAPHICAL INFORMATION SECTOR	32

1 Introduction

This deliverable is the third Metadata Watch 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 things.

This third report concentrates on the issue of the development and use of Application Profiles in the various domains covered by the SCHEMAS Metadata Watch.

Contributions are included from the SCHEMAS partners PricewaterhouseCoopers, UKOLN and GMD, as well as from a number of correspondents in the following domains:

- Industry
- Publishing sector
- Audio-visual sector
- Cultural heritage sector
- Educational sector
- Academic sector
- Geographical information sector

There is also a small section on developments in the Government sector, a sector that is emerging in our work as an important addition to the metadata landscape.

2 The concept of application profiles

The concept of application profiles has emerged in discussions on metadata schemas in the last year, in relation to work that is being done on metadata registries, specifically in the Dublin Core Metadata Initiative. The partners in the SCHEMAS project, and specifically Thomas Baker of GMD and Rachel Heery and Manjula Patel of UKOLN, have made major contributions to this discussion.

Baker, in a “strawman proposal” to the Dublin Core Registry working group [1] defines application profiles as entities that declare which elements from which namespaces underlie the local schema used in a particular application or project. In his view, application profiles “re-use” semantics from namespaces and repackage them for a particular purpose. This is in line with Heery and Patel [2] who define application profiles as schemas consisting of elements drawn from one or more namespaces, combined together and optimised for a particular application. They suggest that a distinction can be made between a *namespace schema* (containing all those elements defined for a particular namespace) and an *application profile schema* (containing combinations of sub-sets of one or more namespace schemas).

It needs to be pointed out that the term *namespace* in these definitions should be read as the metadata element definitions and semantics defined within those namespaces. As an example, the namespace for the Dublin Core Metadata Element Set, version 1.1 can be referred to (in XML) as:

```
xmlns:dc= "http://purl.org/dc/elements/1.1/
```

At the location specified by the URL, the 15 Dublin Core elements and their semantics are defined.

In his "strawman proposal" and in subsequent discussions, Baker laid out a number of functional requirements for application profiles:

- to declare which elements from which namespaces an application or application community is using in its project- or domain-specific schema.
- to override the default (i.e., standard) definition or label of a particular element (from the namespace) with a definition or label that makes sense in the local context. For example, a profile for an application for describing collections might redefine the Dublin Core element Title -- officially, "A name given to the resource" -- with a more restricted local definition: "A name given to the collection".
- to describe the project or application to which it refers. At a minimum, this description should give the name and mission of a project or application. Ideally, it would provide further technical detail about the application environment in which the schema is used.
- to specify dependencies among vocabulary terms -- for example (in profiles that use Dublin Core), that particular encoding schemes may be used with particular elements.
- to define multiple entity classes in a particular data model. For example, a given application may need to describe both "people" (with attributes like name, occupation, and weight) and "resources" (with attributes like title, subject, and format).
- to provide guidelines, in free text, about the local use of particular terms.
- to express controlled vocabularies in a machine-understandable way. For example, if the values of the term "Economic Sector" are restricted to just "Agriculture", "Industry", or "Services", the profile should be able to declare this.
- to specify the language and the character encoding (e.g. Big-5 encoding for Chinese) of the human-readable parts well enough to allow these to be rendered in the correct font.

It needs to be noted that this is very much 'work in progress' and that these requirements may evolve over time, before there is a general agreement.

The SCHEMAS project adopted the following definition at the occasion of the second workshop [3]:

Implementation projects generally find that no one metadata standard will completely meet their descriptive needs. General standards such as Dublin Core must often be used alongside domain- or sector-specific standards such as MPEG-7 for multimedia and IEEE/LOM for educational resources; and new elements may be needed for local needs not covered by any of the existing standards. Recent practice distinguishes between the definition of semantics in "namespaces" (i.e. official standards) and the reuse and interpretation of those semantics in "application profiles". Application profiles are schemas that combine elements from multiple standards, perhaps with application-specific constraints such as the use of specific controlled vocabulary.

3 The use of application profiles

The concept of application profiles is rather new. What is not new is that many activities and projects have been mixing and matching metadata elements sets, and have added elements to existing sets and modified the semantics of existing elements (in the sense of defining them in the context of specific applications).

A number of examples is mentioned in the domain reports in section 5, for example:

- the UK's Resource Discovery Network uses six out of the fifteen Dublin Core elements with recommended data entry guidelines
- the European Renardus project uses eight Dublin Core elements are used and one project-specific element.
- the Australian Government Locator Service adds four elements to the 15 Dublin Core elements.
- various projects (EDNA, Schoolnet and GEM) in the educational sector use the fifteen Dublin Core elements with a number of additional elements. In practice, these projects make different choices to profile the same specification for their communities, although some specific elements appear in more than one.
- other educational projects build profiles on the basis of the LOM standard.
- the US Federal Geographic Data Committee, apparently being aware that implementers would need local adaptations, provides guidelines for creating application profiles for the content standard for Digital Geospatial Metadata

It is not surprising that in sectors where standardisation of metadata element sets is not well advanced or where there is little co-ordination between standardisation activities (such as the industry, publishing and audiovisual sectors) the use or even awareness of application profiles is low.

4 The future of application profiles

Looking at the list of functional requirements formulated by Baker, it can be seen that these cover a number of the questions for which implementers may be looking for answers. It supposes, however, that these implementers have an objective to re-use the experience and approach from others and that they have the aim to increase the potential interoperability between the collection they describe and other collections.

The work on application profiles is, however, in its early stages. A number of fundamental questions have only begun to be asked and answers need to be found through further research and experimentation.

Much is dependent on the emergence of registries where application profiles can be published and found by others. Work in the area of registries is underway in various places, such as the Dublin Core Metadata Initiative, the Indecs project, XML.org, and, indeed, the SCHEMAS project itself.

Based on the experiences gained in these various activities, conclusions can be reached on how application profiles can help implementers to make the best use of experiences from other activities, thereby reducing the resources in the design and implementation phase, as well as helping further harmonisation to take place.

5 Domain reports

5.1 Industry sector

Correspondent: Elise Sfeir, PricewaterhouseCoopers

5.1.1 Introduction

Regarding metadata, the industry domain is a little bit apart. Actually, industry can cover a lot of very various sub-domains which do not necessarily have the same needs. Therefore, it is more difficult to find a common ground.

Yet, and as already specified in the first industry correspondent report, regarding schemas, XML is popular amongst the industrial world.

All the activities that have been looked at actually use XML schemas and namespaces. Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML. Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web.

More information on XML Schemas is hosted by Oasis and can be viewed on: <http://www.oasis-open.org/cover/schemas.html>.

Yet, it is difficult to look at the application profiles that have been built by the various activities because such information is not provided on their web sites and documentation. For instance, on the BizTalk web site, you can download your own schemas, but you cannot view how they have been built. No information is available on how schemas look like. It appears that XML schemas are used as a basis for the development of application profiles, but no information is provided on these latter. Therefore, the present report might give only very poor details on the construction of the industry application profiles.

The initiatives developed here are all supported by important IT companies such as Netscape, Microsoft or Sun Microsystems.

5.1.2 RSS or RDF Site Summary - Netscape

All the following information has been taken from the RSS Web site: <http://www.egroups.com/files/rss-dev/specification.html>

RSS is a lightweight multipurpose extensible metadata description and syndication format. It is an XML application, conforming to the W3C's RDF Specification. It is extensible via XML-namespace and/or RDF based modularization.

An RSS summary, at a minimum, is a document describing a "channel" consisting of URL-retrievable items. Each item consists of a title, link, and brief description. While items have traditionally been news headlines, RSS has seen much repurposing in its short existence.

RSS 0.9 was introduced in 1999 by Netscape as a channel description framework / content-gathering mechanism for their My Netscape Network (MNN) portal. By providing a simple snapshot-in-a-document, web site producers acquired audience through the presence of their content on My Netscape.

As RSS continues to be re-purposed, aggregated, and categorized, the need for an enhanced metadata framework grows. Channel- and item-level title and description

elements are being overloaded with metadata and HTML. Some producers are even resorting to inserting unofficial ad hoc elements (e.g., <category>, <date>, <author>) in an attempt to augment the sparse metadata facilities of RSS.

There are 2 solutions for this augmentation: one solution is the addition of more simple elements to the RSS core. A second one, is the compartmentalization of specific functionality into the pluggable RSS modules. This is the approach adopted in the RSS specification that can be viewed on the web site. Modularization is achieved by using XML Namespaces for partitioning vocabularies. Adding and removing RSS functionality is then just a matter of the inclusion of a particular set of modules best suited to the task at hand.

At the last Dublin Core meeting in Ottawa, a RSS expert said that they want to strengthen the connection between the Dublin Core community and developers of RDF Site Summary (RSS). In many ways, RSS has already proved useful as a metadata testbed and validates many of the assumptions implicit in the Dublin Core efforts. RSS was meant to support Dublin Core, but Netscape dropped it from the specification at the last moment to the dismay of the DCMI community.

The new RSS 1.0 proposal provides a way to utilize Dublin Core as a common framework for sharing richer metadata. The goal is to bridge these two efforts so that Dublin Core can benefit from the experience of RSS developers and their tools, and RSS can benefit from the expertise of the Dublin Core community. One can continue to use the current RSS and provide only Title, Link, Description; but if you already have the metadata and want to make it available, then we wanted to create a standard way to do so. RSS file for O'Reilly Network contains items that are Dublin Core compliant.

In addition, to Title, Link, and Description, RSS has supplied the following fields: Creator, who in this case is the author of the article; a list of Subject keywords; the Type of item, in this case, a technical article; the Language in which it is written; the Date it was published; the file Format; and a statement about who owns the Rights to this article as well as the name of its Publisher.

5.1.3 UDDI - Universal Description, Discovery and Integration

The following information has been extracted from a document untitled UDDI Data Structures Reference on the web site: http://www.uddi.org/pubs/UDDI_XML_Structure_Reference.doc

The UDDI has based its approach on XML but has developed its own set of structure and schemas.

The programmatic interface provided for interacting with systems that follow the Universal Description, Discovery and Integration (UDDI) specifications make use of Extensible Markup Language (XML) and a related technology called Simple Object Access Protocol (SOAP), which is a specification for using XML in simple message-based exchanges.

The UDDI Programmer's API Specification defines approximately 30 SOAP messages that are used to perform inquiry and publishing functions against any UDDI-compliant Business Registry. This document outlines the details of each of the XML structures associated with these messages.

The purpose of UDDI-compliant registries is to provide a business discovery platform on the World Wide Web. Service discovery is related to being able to advertise and

locate information about different technical interfaces exposed by different parties. Services are interesting when you can discover them, determine their purpose, and then have software that is equipped for using a particular type of Web service complete a connection and derive benefit from a service.

A UDDI-compliant registry provides an information framework for describing services exposed by an entity or business. Using this framework the description of a service that is managed by a UDDI registry is information about the service itself. In order to promote cross platform service description that is suitable to a "black-box" Web environment, this description is rendered in cross-platform XML.

The information that makes up a registration consists of four data structure types. This division by information type provides simple partitions to assist in the rapid location and understanding of the different information that makes up a registration.

These four structure types make up the complete amount of information provided within the UDDI service description framework. Each of these XML structures contains a number of data fields (that is elements and attributes) that serve a business or technical descriptive purpose.

These structures are described in the UDDI API Programmer's Schema. The schema defines approximately 20 requests and 10 responses, each of which contain these structures, references to these structures, or summary versions of these structures.

5.1.4 OASIS - XML.org

Web site and source: <http://www.oasis-open.org/>

XML.ORG is an independent resource for news, education, and information about the application of XML in industrial and commercial settings. Hosted by OASIS and funded by organizations who are committed to product-independent data exchange, XML.ORG offers valuable tools, such as the XML.ORG Catalog, to help making critical decisions about whether and how to employ XML in the business.

As XML rapidly becomes the key data interchange standard for the Web, customers and developers are recognizing the need for effective administration of the XML schemas related to e-commerce, business-to-business transactions, and tools and application interoperability. As an established vendor-independent organization that has been hard at work at enabling interoperability for the past 8 years, OASIS has recognized this need. Through XML.ORG, OASIS will collect, manage, and distribute information about XML applications, including vocabularies, schemas, namespaces and DTDs.

An XML.ORG Registry has been built. People can submit their XML Schemas or can search for some. The Registry is a community resource for accessing the fast-growing body of XML specifications being developed for vertical industries and horizontal applications. The XML.ORG Registry offers a central clearinghouse for developers and standards bodies to publicly submit, publish and exchange XML schemas, vocabularies and related documents. It is a self-supporting resource created by and for the community at large.

Industry groups and other organizations that have developed XML schemas or vocabularies can freely register their work at the XML.ORG Registry. Today, the Registry is in its first phase of development. It is offered as a call for participation, an opportunity to experience the potential of an open XML registry and an invitation to the community to help shape its evolving functionality.

5.1.5 *BizTalk*

BizTalk is an industry initiative started by Microsoft and supported by a wide range of organizations, from technology vendors like SAP, CommerceOne, and Ariba to technology users like BASDA. BizTalk is not a standards body. Instead, it is a community of standards users, with the goal of driving the rapid, consistent adoption of XML to enable electronic commerce and application integration.

The BizTalk Framework is being currently defined. Its is a set of guidelines for how to publish schemas in XML and how to use XML messages to easily integrate software programs together in order to build rich new solutions. The design emphasis is to leverage what you have today - your existing data models, solutions, and application infrastructure - and adapt it for electronic commerce through the use of XML.

On the BizTalk web site, there's a library of XML schemas for you to review and download for use in your own applications. It is even encourage to publish your own schemas here for others to use.

The library web site is at: <http://www.biztalk.org/library/library.asp>

The library section of the web site presents card catalog and librarian features that provide BizTalk.Org members with the ability to locate schemas that others have registered and cataloged. Members are given the opportunity to register their organizations and establish publishing rights. They can also freely share their work and technical information describing how their organization defines their use of the XML standard.

5.2 **Publishing sector**

Correspondent: Laurie Causton, Clearbay Limited

5.2.1 *Current state of domain*

The last few months have seen progressive developments and announcements rather than milestones. The major area of activity remains electronic publishing, and in particular the various issues of rights within this.

Governmental - or perhaps quasi-governmental - interests in this area are discussed in the document from the Electronic Publishing, Books and Archives Project of the Council of Europe, "Draft guidelines on legislation and policy measures for book development and electronic publishing" (<http://culture.coe.fr/epba/eng/ecubookR.5.htm>).

A little progress in the news domain - the IPTC has ratified NewsML v1.0.

In September:

- CISAC held its General Assembly in Santiago, again with a focus on digital rights management, and a recognition of importance of identifiers.
- there was an announcement of a major revision of the Santa Fe convention of the Open Archives Initiative, resulting from experiences gained and a growing interest in the application of Santa Fe convention-concepts outside its original domain of e-prints.

October saw a number of announcements:

- the RIAA signalled its intention to develop a new identification system for sound recording, primarily to "facilitate and accelerate the growth of digital delivery of music on the Internet." More on this below.
- at the Frankfurt Book Fair the International DOI Foundation announced the formation of a new IDF Working Group - DOI-EB (DOI for E-books), to focus on the development of proofs-of-concept for the use of DOIs with e-books, taking into account related work such as the recommendations of the AAP/Anderson Consulting report on e-book numbering and activities in the Open e-Book Forum, the Electronic Book Exchange Working Group, ONIX International, as well as the IDF's own work on metadata standards and registration procedures.
- and at its quarterly Work Summit, the Open e-Book Forum itself released a document, "A Framework for the Epublishing Ecology" (<http://www.openebook.org/framework>) aimed at facilitating standards work in DRM and other e-publishing areas. The OeBF sees this document as a key component of two initiatives, announced at the same time - its requirements gathering and harmonizing initiative, with a focus on digital rights management (DRM); and a standards coordination initiative, discussed below.

5.2.2 *Overlaps and gaps identification*

The progress of the new RIAA identifier initiative will be interesting to watch - it is an identification system for music, with an emphasis on rights issues. It will aim to be fully compatible with the existing ISRC, but the music world already has a number of identifier- and rights-related activities - including the ISRC itself, the Secure Digital Music Initiative (SDMI), and the Interdeposit Digital Number (IDDN) (<http://www.iddn.org>) which also has music application.

Nevertheless, overlaps and parallel activity are perhaps inevitable, and maybe with this in mind, the OeBF inaugurated the new initiative in standards coordination noted above - this will involve regular technical coordination workshops to provide a forum for coordination and liaison between the various disparate, but related, standards efforts. There will be representation from a significant number of bodies including EBX, NISO, W3C, the International Publishers Association, MPEG, the American Association of Publishers, IDF, and Editeur. One to watch, perhaps.

5.2.3 *Trends*

The e-book industry continues to gain momentum, with estimates of \$12million sales in downloaded books in 1999 . Predictions for the future vary but still indicate an enormous growth. However digital rights management, copyright and copy protection, and anti-piracy continue to be the major issue for the industry bodies. Not surprisingly, this arises from understandable commercial concerns of revenue protection - as a Microsoft representative pointed out, calling for a focus on copy protection. "... without the ability to make a profit, we don't have a business ...".

The concept of Application Profiles is relatively new. Its "predecessor," the Warwick Framework, had a promising start, but here not much has happened since. Application Profiles are still largely under discussion in the academic and library domain, and as yet there is little related activity in this publishing world, but if they gain credence, then there may be merit in investigating their use in publishing. There are areas where APs might offer a useful solution, where the standard concept of a publication or book is no longer adequately addressed by established metadata approaches. An example

might be a reference book on music, published with an attached CD containing musical works, some perhaps copyrighted.

5.2.4 Main issues

There is a view that the major obstacle for publishers to successful e-book development and growth is the lack of a standard, secure method of exchanging e-book content. This seems to fall squarely within the remit of the Electronic Book Exchange Working Group (EBX). Even so, it is perhaps worth noting in this context that, while the OeBF enjoys a good representation from publishers (and booksellers), these are lacking in the membership of EBX, whose members are generally those companies with a vested interest in the development of the enabling technologies. Nonetheless, it has to be said that publishers and booksellers (typically the same ones that you find in the OeBF membership) participate in EBX activity.

In essence, of course, the issue here is rights, which will continue to predominate, and we will undoubtedly see much more of the separate but often overlapping activity in this area, particularly as the boundaries between different forms of content become less distinct. The OeBF's coordination initiative might play a useful role here.

5.2.5 Activity planned for the next period

The e-Book World conference is to take place in early November in New York, and is claimed to be the first event dedicated solely to the electronic book marketplace. It will focus on the technological aspects of e-book publishing, and the concerns facing traditional publishers.

The next DOI Workshop, also to be held in November, will look at, among other things, e-books and metadata.

The OeBF Working Group Summit takes place in December 2000, and will host a workshop of the coordination initiative.

5.3 Audio-visual sector

Correspondent: Annemieke de Jong, Nederlands Audiovisueel Archief

5.3.1 Current state of the domain

Compared to the networked 'bibliographical' domain, the development of metadata dictionaries and metadata schemes for audiovisual media is a relatively young area. Although certain specialized standards for parts of the audiovisual production process have already been developed and in some cases are indeed operational, not many of them have been given formal status. No detailed, official standards for interoperable 'audiovisual' metadata (that includes and integrates all digital production, distribution and archiving processes) yet exist and there are no commonly accepted models for the audiovisual production environment as a whole.

In order to define a metadata scheme for a media management system for the production environment two interrelated aspects have to be considered. In the first place the system has to be geared to the specific requirements of the audiovisual organisation it is used for; in the second place the necessary protocols and standards should be moulded as closely as possible to international schemes and standards as to enhance interoperability and facilitate exchange of audiovisual documents on a national and on an international scale. Metadata schemes that are developed locally have therefore to be synchronized with internationally accepted standards. Whenever possible local definitions should be translated to international standards from the start

on. For the production and distribution of digital television and radio material the SMPTE Metadata Dictionary and Engineering Guidelines function as an important reference. The same goes for the activities of the MPEG-7 group, that -although still in an early stage of development- are being closely monitored for developing schemes for multimedia content in the professional environment. For the category 'descriptive metadata' Dublin Core usually functions as an important reference. The Standard Media Exchange Framework of the BBC (SMEF) is sometimes used as a reference data model for broadcasting production environments and media management systems.

Within the audiovisual production, distribution and archiving domain interoperable schemes are developed on various levels:

1. Projectbased metadata schemes.
 - Local projects or initiatives for digital production, e.g. within one public or commercial broadcasting company
 - Projects with a broad, national scope e.g. The Digital Platform of the public broadcasters in the Netherlands, that prepares a common digital infrastructure covering a metadata dictionary and a datamodel. Another example is the Digitaal Erfgoed Nederland (DEN) initiative, that aims at establishing a web based interoperable interface that gives access to digital collections of various nature, including the audiovisual national heritage.
 - European projects within the Fifth Framework, that are aimed at establishing digital video libraries e.g. the European Chronicles Online (ECHO).
2. 'Industrial' metadata schemes.

Industrial players sometimes develop their own schemes and models, usually including (parts of) existing standards or standards-in-development and adding metadata and datamodel requirements coming from their customers. For media management systems these models are usually SMPTE compliant.

3. Interoperable schemes coming from the standardizing community.

e.g. P/Meta, one of the projectgroups of the European Broadcasting Union (EBU) aims at harmonizing the SMPTE Metadata Dictionary, Dublin Core and MPEG-7 with the requirements for exchange of audiovisual information between European broadcasters, archives and consumers.

5.3.2 Main issues

Generally speaking one of the main issues is the difficulty to synchronize the work that is being done within standardizing committees, projects, pilots and other initiatives on the level of tempo, goals en scope. Currently there is no centralized, systematic and well-timed exchange of information on the various local, international and national projects that concern the development of 'audiovisual' metadata schemes. Standardizing committees work slow and thoroughly and broadcasters and producers on the other hand - due to the rapid technological developments - are forced to act and cannot wait for yet another new version of the standard. Another problem is the still rather inadequate communication between the various contributing communities to schemes, both on an internal, company level and internationally. Communication problems may be caused by the differing professional (technical, archival, production) backgrounds of the contributors that need to integrate requirements to come to an interoperable scheme. The variation in the many audiovisual production models also leads to problems in establishing common metadata models, which makes

it difficult to standardize on a detailed level the various groups and classes of metadata that are required for digital media systems.

Issues more specific:

- Harmonizing the user requirements within the various audiovisual production domains
- Determination of the universe of discourses: to what level should information be exchanged?
- Standardizing audiovisual production, distribution and archiving process models and working procedures
- Standardization of data structures and data models, entities and attributes
- Standardizing type fields and field values
- The development of globally accepted, registered unique identifiers for audiovisual essence to link the stored audiovisual material with the related documentation through the audiovisual production process.
- Multilingual aspects in interoperable schemes
- Understanding and integrating of the various categories and types of metadata in the audiovisual environment including media processing metadata, technical metadata, production related metadata, descriptive categories
- The issue of integrating the legacy archives and legacy catalogues into the digital domain.

5.3.3 Trends, overlaps and gaps

Generally speaking the same kind of efforts are being undertaken in many 'similar' audiovisual environments concerning the development of metadata schemes. However, as for the more specific levels of the schemes and models this can hardly be avoided. Each audiovisual production environment has its own specific procedures and requirements and these need to be reflected in the company process schemes and data models. Besides this, the environments may differ in the extent to which exchange of information is restricted in view of copyrights and commercial exploitation. In other cases (parts) of internal schemes hold detailed information on internal business procedures and working processes and can therefore not be freely shared with other environments. It is for this reason that initiatives like that of EBU's P/Meta, that work on connecting the requirements, schemes and models of various broadcast environments to international standards, are of vital importance. Still, in order to enhance interoperability from the start on, it would be very useful if more professional platforms are established where broadcasters and other av-producers may acquire detailed, objective information on which work is already done and what schemes are available for reference.

Presently many pilot projects are running within the broadcast environment that include the development and use of metadata schemes for production, distribution and archiving or for any of these three processes. Forced by the rapid increase of digital production of audiovisual materials and the lack of usable common schemes and standards, many broadcasting companies and other producers are currently developing their 'own' metadata schemes for these pilots and projects. During this process they usually start with formulating local user requirements and defining local process models and dataflows. After this they might see to which extent their proprietary

models can be harmonized and mapped with existing standards or standards-in-progress for (parts of) the audiovisual domain, if any. New, commercial broadcasters that hold no legacy archives or catalogues currently tend to be in front of the developments with often a full working digital environment, using proprietary models and metadata dictionaries

The awareness amongst broadcasters, av-producers and others players in the audiovisual field that interoperability with other environments is an essential requirement, fortunately is growing rapidly. More and more it is realised that failure to adopt international standards in an early stage effectively means that local systems are inefficient and become more inefficient. Subsequent adaptation of local systems is extremely labour intensive and the gradual implementation of standards tends to be guided by technology rather than by user requirements. In this scenario the number of unique identifiers would slowly grow and the chances of interoperability would diminish. Broadcasting companies for instance, could be manouvered into a very unfavourable position compaired to audiovisual organisations that are more experienced in e-commerce. But although it is recognized that much can and needs to be integrated into international standards, there will allways remain elements of purely local importance which then may be taken on as extensions.

The contribution to the development of interoperable schemes for various areas at different stages of growth still stem from the respective professional qualifications and commercial interests of the participants. Although communication is still not sufficient, it can be observed that, especially during the last 12 months, the various developments are increasingly coordinated at European and international levels, whilst at the same time organisations that produce audiovisual materials have their own internal services and production services work much closer together than before in jointly establishing projects and pilots for the digital production environment. Broadcast engineering, information technology and the world of documentation and archiving are joining forces more and more to the benefit of quality levels all over.

In Appendix A, a number of application profiles for the audio-visual sector are described.

5.4 Cultural Heritage sector

Correspondent: Walter Koch, AIT

5.4.1 Current state of domain

Metadata in the field of Cultural Heritage relate to the description of ‘things’ as there are:

- Real/scientific objects,
- bibliographic (‘intellectual’) objects,
- (digital/analogue) surrogates of objects (image/film),
- objects instantiated by a (markup) language.

All these types of ‘things’ can be interrelated and appear in sub domains like: libraries, museums, and archives and are embedded in a space/time framework. Each sub domain has developed in the past separate (domain specific) meta data systems even there are similar ‘concepts’ which are described. Eg in the museum domain the term (meta data element) ‘creator’ (of a ‘thing’ called artwork) has similar meaning like ‘author’ in the library domain (creator of a ‘thing’ called book). Creator and

author are domain specific meta data and up to now it was common to enumerate all meta data which are needed to describe things, relations, and dimensions. This leads to exhaustive lists of meta data elements which turned out to be a great obstacle when a common meta data system for all sub domains have to be developed. The adventure of the internet has made it popular to look for 'resources' in a general way independently of sub domains. The pragmatic way consists in mapping meta data elements into a new set of elements like MARC based meta data elements in a few 'use' attributes of the Z39.50 environment. A more radical approach is provided by the Dublin Core which maps domain specific elements into 15 'core' elements. A more sophisticated approach might be to look 'behind the scenes'. If we take the term 'author' we can assume that this is a person or robot fulfilling a function (role) in relation to an object ('thing'). This brings us slowly to the term 'ontology' which is defined in different ways. The Knowledge System Laboratory, Stanford defines ontology as "specification of concepts to be used for expressing knowledge". Basic elements of this framework (system 'S-0') are: Types of entities, Attributes and properties, Relations and functions, constraints. In the library domain (system 'S-1') the meta data element 'author' can be considered as a composition of specifications of basic S-0-elements: person (*entity*), create (*function*) related to a 'thing' (*entity*). This is a very rough and not precise clarification of the environment which meta data is part of. In the practice there are first steps in the museum area to leave conventional paths: The consortium of interchange of museum information (CIMI) has introduced the meta data elements: who, what, when, where ('4w') which can be considered as a domain independent conceptualisation of four entities (person, thing, time, place). If we look into the IDEF5 (integrated definition) methodology which distinguishes different levels of ontologies we can consider the '4w' or Dublin Core elements belonging to a '*domain ontology*' for the cultural heritage sector, meta data used in the library sub domain as part of a '*practice ontology*' and metadata used within coin collections as part of a '*site ontology*'.

5.4.2 Main issues

Main issues in the CH-sector are still the harmonisation of meta data systems in subdomains. In the **library sub domain** the mapping of different MARC (MAchine Readable Cataloguing) standards into a unified system like UNIMARC reflects this effort. In the **museum sub domain** harmonisation efforts are much more difficult due to the heterogeneity of the objects (things) to be described. A coin needs quite different metadata elements compared to a painting. Specific attributes (eg. Genre in the arts environment) have been introduced to develop site specific domains. In the field of arts the AMICO (art museums image consortium) project has gained some importance; another system used in this area is called CDWA (categories for the description of works of art). The documentation group (CIDOC: International Committee for Documentation) of the International Council Of Museums (ICOM) has developed different system in the recent years: the ICOM 'information categories' which are based on (business) processes used in the museum world (this is related to some extent to the system (SPECTRUM) introduced by MDA (Museum Documentation Association). The CIDOC relational data model (based on collection management systems) defined also relations between basic 'entities' like person, event, and object. The latest development followed an object oriented approach and has led to the 'Conceptual Reference Model' (CRM). In the **archive's sub domain** a new issue evolves: the organisation of a collection. This is reflected especially by the system developed the International Council of Archives (ICA) and called ISAD(G), 'General International Standard Archival Description'. Another system, the Encoded

Archival Description (EAD) is quite frequently used to generate metadata based on the EAD-DTD. Special Professional Associations (eg the International Association of Sound and Audiovisual Archives, IASA) have developed 'site specific' "cataloguing rules" on different levels (fonds, collections, etc).

5.4.3 Trends

Having in mind the three levels of ontologies as outlined under 5.4.1 (domain O., practice O., site specific O.) one can see that there are efforts on all levels: The issue of cross (sub) domain searching is on a broad basis covered by the Dublin Core (simple) meta data system. Practice Ontologies contain meta data which are undertaken a harmonisation process in specific sub domains (eg library, archive), and Site specific metadata are quite popular in the museum world due to the heterogeneity of the objects to be described. Dublin Core qualified might be a step into the direction of developing a meta data system which can be applied to all levels of ontologies but these efforts are just at the beginning.

5.5 Educational sector

Correspondent: Erik Duval, Katholieke Universiteit Leuven

5.5.1 Review and update of activities

The four main standards-related initiatives in the domain of education, are still:

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

Most of the metadata related activities still take place in the two organisations listed under a and b above.

The IEEE LTSC LOM group held a meeting in Sedona in September, where it was decided to 'open up' the data elements for which vocabularies of appropriate values are defined in the document. The working document has been updated to reflect this change and numerous small edits and modifications. The new version, working document version 5, has been submitted to ballot in the LOM working group. At the same time, members were asked to approve (or not) the forwarding of this document by the IEEE LTSC sponsor and executive committee to the IEEE for full ballot. The next meeting will be held in Athens, Greece.

The Dublin Core Metadata Initiative met for its 8th workshop in Ottawa in October. It is unclear at this moment whether the DC-Education made any substantial progress at that occasion.

Most noteworthy is that there are new attempts to organize collaboration between LOM and DC-Education.

The CEN/CENELEC ISSS Learning Technologies Workshop launched the second phase of its existence at a meeting in October in Brussel. As mentioned in the previous report, a substantial part of the proposed work is concerned with metadata, including work on vocabularies and taxonomies, profiles, bindings and internationalization of metadata. It seems very likely that European funding will be available to support the workshop. The next meeting will be held in January in Brussel.

5.5.2 *Educational metadata application profiles*

The attached table (Appendix B) summarizes the elements sets, schemes and technology bindings by some of the main initiatives in this domain: ADL, ARIADNE, DC-Education, EDNA, EUN Schoolnet, GEM, IMS, NEEDS and SMETE. Most of these initiatives were covered in previous reports.

It is clear from this overview that different organisations make different choices to profile the same specification for their communities. With respect to LOM, ARIADNE is clearly the more ambitious profile, as it makes 23 elements mandatory. In comparison, IMS decided to include 20 elements in the 'core' group. ADL only requires 7 of the LOM elements to be provided.

For the DC-based initiatives, it is clear that they all needed to add elements to the basic DC element set. The DC-Education group decided to adopt 4 elements from LOM, and an additional data element. They also decided to add a 'ConformsTo' qualifier to the Relation element. EDNA, Schoolnet and GEM added between 5 and 9 data elements to the basic DC element set. Approver, User Level and Version are common to more than one of these initiatives.

ARIADNE, EDNA and GEM are the more active consortia with regards to the development of vocabularies for data elements. There is some adoption of (mainly GEM related) such vocabularies by other consortia.

The main base technology for implementation is XML.

It is striking to note that, at present, there is little activity on interoperability between independently developed systems for metadata management, like the ones presented above. Some demonstrator work has been done between the ARIADNE and GESTALT consortia, but larger scale developments are rare at this moment.

We believe that it is now time to start this interoperability development, as the specifications and implementations are maturing, and as several communities have developed their own profiles of the specifications involved. In principle, such profiles should not hinder interoperability. But this needs to be demonstrated in practice now.

Only then will we be able to amass a critical quantity of learning material, and will the standards indeed serve their ultimate goal.

As an addendum, note that many smaller scale projects, typically confined to one organisation, are also including metadata in their development efforts. Sometimes, the basic idea is to provide a useful service within the own organisation, and sometimes, the main focus is on further research that requires metadata support as basic functionality.

5.6 Academic sector

Correspondent: Michael Day, UKOLN

5.6.1 *Updates*

The European Commission funded NEDLIB (Networked European Deposit Library) project has recently published its specification of Metadata for long term preservation. The specification defines core metadata elements for the preservation management of digital documents. As with the Cedars project, the specification generally follows the arrangement of the taxonomy of information objects defined in the Reference Model for an Open Archival Information System (OAIS), a draft ISO standard published by the Consultative Committee on Space Data Systems.

5.6.2 *Application profiles*

Application profiles are schemas that consist of data elements drawn from one or more defined namespaces that, combined in particular ways by implementers, are optimised for particular local applications. In previous 'academic' domain reports, a wide variety of metadata initiatives have been described. Not all of these initiatives fit very well into the application profile discussion, so only a select few will be included here.

In the Internet information gateway area, application profiles can be seen in use within both gateways and the services that broker between them. Increasingly these application profiles tend to be extensions (or refinements) of the Dublin Core Metadata Element Set (DCMES). For example, the 'minimal set' of elements defined for use in the UK's Resource Discovery Network (RDN) is a small subset of the Dublin Core comprising six elements with recommended data entry guidelines. In the Renardus project, the element set is slightly larger (there are eight elements) but is also - with one exception (Country Code) - defined as a subset of Dublin Core elements.

Metadata initiatives concerned with the development of metadata schemas for digital preservation have had less concern with defining application profiles. However, the outline metadata specification developed by the Cedars project didn't define specific resource discovery elements but assumed that any future fully-developed system would need to consider resource discovery requirements in more detail. The project, therefore, included a subset of unqualified Dublin Core elements to form its preliminary 'Reference Information' for resource discovery.

The concept of 'application profiles' is relatively unknown in the various initiatives concerned with defining recordkeeping metadata, but there are some interesting parallels within the range of Australian initiatives. A basic component of the recordkeeping standards developed in Australia is not the DCMES but the Dublin Core-derived Australian Government Locator Service (AGLS) Metadata Standard. The AGLS standard has been designed to facilitate the resource discovery and retrieval of Australian government information and services and is maintained by the National Archives of Australia. The standard is itself a type of 'application profile' - based on Dublin Core with four additional defined elements (Audience, Availability, Function and Mandate) - that has been defined for government information and services of all types. Not all of these will be considered 'records' in a recordkeeping context. The AGLS standard is designed to be extensible, so that those with different (or more specific) requirements are able to add additional elements or qualifiers, assuming that the semantics of AGLS elements are not changed and that any mandatory elements in AGLS remain so. Australian recordkeeping metadata initiatives have made a point of maintaining links with the AGLS standard. For example, the Recordkeeping Metadata Standard for Commonwealth Agencies - also developed and maintained by the National Archives of Australia - inherits concepts from AGLS but contains additional elements specifically related to recordkeeping. While some elements in the Commonwealth recordkeeping metadata set are the same as those in AGLS, many of the other elements extend the AGLS in a significant way. In order to note these extensions, version 1.0 of the Recordkeeping Metadata Standard contains a mapping from the metadata elements defined in that document to the AGLS standard. Similar links with AGLS are also maintained in the Recordkeeping Metadata Schema (RKMS) developed by the Australian SPIRT Recordkeeping Metadata Project. The RKMS was specifically designed for compatibility with Dublin Core, the AGLS standard and the Recordkeeping Metadata Standard for

Commonwealth Agencies. The RKMS was defined at a more conceptual level than the Commonwealth recordkeeping metadata standard, but - in common with that standard - the AGLS metadata was viewed essentially as a subset of any metadata set specified for recordkeeping purposes.

In the biological information sector, the main example of application profile type developments is the definition of the Biological Metadata Profile of the Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata (CSDGM). The Biological Metadata Profile inherits all of the CSDGM elements, but includes additional elements that are able to document information about taxonomy and nomenclature. Another potential example, is the development of a Z39.50 profile for natural history collections and observation data sets known as the Darwin Core (DwC).

The Z39.50 Biology Implementers Group (ZBIG) have developed this profile, and it includes access points for a variety of scientific names and the institutions that hold biological specimens. The profile was developed in a Z39.50 context but there is a hope that the profile may have some wider applicability.

5.7 Geographical information sector

Correspondent: Elise Sfeir, PricewaterhouseCoopers

5.7.1 Introduction

This brief introduction highlights the main standards that are used by major GI projects that are identified in the present report. Actually, many GI initiatives look at these main standards and use some of them to create their metadata application profiles. It is important to note that in the GI world, the FGDC which is an American committee has a lot of impact on the standards used and on the developments of metadata. All the other initiatives have had to do with the FGDC. The exception to these standards seems to be Dublin Core. The GI world does not use it very much and the FGDC, for the moment, does not really recommend Dublin Core for GI metadata.

Therefore, many initiatives do not look at Dublin Core. They only consider one standard and follow it, or they make their own standard. Yet, some European projects, such as ETeMII do not fully follow the FGDC and look at the DCMI. ETeMII, which aims at organising a network of excellence, bringing together most of the stakeholders of the Territorial Management Information market, coming from research, industry and public sector, liaise with Dublin Core. ETeMII will liaise with various groups in order to build its own application profiles mainly by mapping Dublin Core and ISO/ TC 211 - 19115. The Madame project also recommends and follows this mapping.

5.7.2 Standards

ISO/TC 211 - 19115

In 1994 the International Standards Organisation created technical committee 211 (ISO/ TC 211 - 19115) with responsibility for Geoinformation/Geomatics. They are also preparing a family of standards; this process involves a working group, a committee draft, a draft international standard and finally the international standard. ISO have now released the committee draft of 'ISO 15046-15 - GI - Metadata'. CEN/ TC 287 has liaison status with ISO/ TC 211 - 19115 which means the results of the work in Europe will be taken into account when developing the global standards.

The ISO standards on geographic information ISO/ TC 211 - 19115 can be found on: <http://www.standardsinaction.org/gismetadata/>

CEN/TC 287 metadata standard

In 1992 the Comité Européen de Normalisation (CEN) created technical committee 287 with responsibility for geographic information standards. A family of European Prestandards have now been adopted including 'ENV (Euro-Norme Voluntaire) 12657 Geographic information - Data description - Metadata'.

The FGDC impact

In the USA the Federal Geographic Data Committee (FGDC) approved their Content Standard for Digital Geospatial Metadata in 1994. This is a national spatial metadata standard developed to support the development of the National Spatial Data Infrastructure. The standard has also been implemented outside of the USA, for example for the South African Spatial Data Discovery Facility.

The OpenGIS Consortium (OGC)

The OpenGIS Consortium (OGC) is an international membership organisation engaged in a cooperative effort to create open computing specifications in the area of geoprocessing. As part of its draft 'OpenGIS Abstract Specification' OGC has a topic on recording metadata for spatial data. OGC are working closely with FGDC and ISO/ TC 211 - 19115 to develop formal, global spatial metadata standards. At their plenary meeting in Vienna, Austria in March 1999, ISO/ TC 211 - 19115 welcomed the satisfactory completion of the co-operative agreement between the OpenGIS Consortium and ISO/TC 211 and endorsed the terms of reference for an ISO/ TC 211 - 19115 / OGC co-ordination group.

5.7.3 FGDC, the Federal Geographic Data Committee

The FGDC is a very important and international initiative. It has developed its own metadata standards which has actually a lot of impact on the other metadata standards developments.

The FGDC Metadata Standard is being harmonised with the International Organisation for Standardisation (ISO) Technical Committee (TC)211 Metadata Standard 19115. The June 8, 1994 FGDC Metadata Standard was used as the base document for International Organization for Standardization (ISO) 15046 Part 15.

The FGDC developed the Content Standard for Digital Geospatial Metadata (CSDGM) which objectives are to provide a common set of terminology and definitions for the documentation of digital geospatial data. The standard establishes the names of data elements and compound elements (groups of data elements) to be used for these purposes, the definitions of these compound elements and data elements, and information about the values that are to be provided for the data elements.

The development of this standard is related to the Spatial Data Transfer Standard (SDTS) that was developed to allow the transfer of digital spatial data sets between spatial data software.

The Content Standard for Digital Geospatial Metadata uses to the maximum extent possible, existing International or National Standards, as documented in Office of Management and Budget Circular A-119 "Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity assessment Activities." American National Standards referenced in the Content Standard for Digital

Geospatial Metadata include the American National Standards Institute, 1975, Representations of universal time, local time differentials, and United States time zone reference for information interchange (ANSI X3.51-1975): New York, American National Standards Institute; American National Standards Institute, 1986, Representation for calendar date and ordinal date for information interchange (ANSI X3.30-1985): New York, American National Standards Institute; American National Standards Institute, 1986, Representations of local time of day for information interchange (ANSI X3.43-1986): New York, American National Standards Institute.

The FGDC has 220 elements composed of compound elements and data elements.

The FGDC proposes some guidelines to create application profiles based on the FGDC metadata standards. The information below has been taken from a very useful document Guidelines for Creating a Profile for the Content Standard for Digital Geospatial Metadata, that can be downloaded from the following Web site: <http://www.fgdc.gov/metadata/csdgm/profile.html>

The current Content Standard for Digital Geospatial Metadata provides metadata collectors with formally defined elements known as standard elements. The metadata Standard attempts to standardize the content of metadata elements for a wide range of digital geospatial data. However, some users may determine that modifications to the Standard are needed to create meaningful metadata for their data sets. The Standard allows the user to create extended elements and profiles. Extended elements are user-defined elements outside the Standard needed by the metadata producer. A profile is a document that describes the application of the Standard to a specific user community.

Profiles may be formalized through the FGDC standards process or may be used informally by a user community. FGDC is the approval authority for profiles. To become recognized by the FGDC, a metadata profile must go through the FGDC standards review and approval process. FGDC approved profiles must specify a maintenance authority. While the FGDC is the designated maintenance authority for the Metadata Standard the organization or agency sponsoring a profile will be considered the maintenance authority for that profile.

5.7.4 ANZLIC, the Australia New Zealand Land Information Council

ANZLIC's mission is to provide leadership for effective management and use of land and geographic information to support economic growth, and the social and environmental interests of Australia and New Zealand. Key objectives under the headings: data; infrastructure; standards; access; industry development and organisational framework, are the focus of efforts to provide this leadership.

A Working Group was formed by the ANZLIC Advisory Committee in April 1995 to work on the following tasks to improve community access to data:

- produce a Metadata Framework paper which identifies and defines the mandatory metadata elements of a national land and geographic data directory system, discussing creation, maintenance and directory custodianship issues;
- using the Metadata Framework paper, promote the concept of a national data directory system to help determine the priorities and issues for implementing a national directory;
- develop an implementation plan for a national data directory system, including procedures for transfer of metadata between jurisdictions and the national directory system;

- develop and circulate for comment, a discussion paper on national guidelines for developing land and geographic data quality, in a form suitable for developing into an Australia/New Zealand standard.

The ANZLIC guidelines have been developed to promote a consistent standard of description for this small number of core metadata elements, that are generally common for all types of data and designed to indicate what data exists, its content, geographic extent, how useful it might be for other purposes and where more information about the data can be obtained. The purpose is to make information about all available data freely available so that existing data can be reused for other purposes if it is suitable, reducing the duplication of effort.

Standards on which the ANZLIC Approach is based

ANZLIC can be actually considered as a standard on its own because it can be compared to the FGDC work and mainly the ISO/ TC 211 - 19115 which has had extensive Australian input, particularly from interests associated with ANZLIC.

The US approach, developed by the Federal Geographic Data Committee (FGDC), specifies the structure and expected content of some 220 items (elements) which are intended to describe digital geospatial datasets adequately for all purposes. The ANZLIC approach is deliberately less ambitious than what has been attempted in the US. Arguments advanced in support of the more modest objective rely on experience to date with the creation of high-level directories in Australia.

While ANZLIC has not adopted the US approach, the Australia New Zealand framework is, as far as possible, consistent with the guidelines on Digital Geospatial Metadata produced by the US FGDC and with the Australia New Zealand Standard on Spatial Data Transfer AS/NZS 4270. The reasons for this are:

- many organisations are already using these standards for their data management activities;
- some vendors of software are providing templates and other support to the implementation of the standards; and,
- these standards are being implemented in some discipline or theme areas where there is international exchange of metadata.

However, until now, there has been no unifying set of metadata elements that could be used as the basis for the development of national metadata standards. ISO/ TC 211 - 19115 will provide this unifying set of metadata elements.

ANZLIC have based it on the FGDC work and then they have made their own core elements and categories without taken them from another developed standard. They did not proceed to any mapping and comparison with another namespace.

There are 32 ANZLIC core metadata elements that are grouped in 9 categories.

To assist with the implementation, ANZLIC has developed a run-time software tool to support the collection of metadata and to ensure consistent description of core metadata elements. This software tool, based on Microsoft Access, is available for use by dataset custodians throughout Australia and New Zealand.

The Data Entry tool may be used within organisations to manage the metadata database.

5.7.5 *The National Geospatial Data Framework (NGDF)*

All the following information has been taken from very interesting documents untitled Discovery Metadata Guidelines and Discovery Metadata Transfer Format and Communications Protocol Guidelines that can be downloaded on the following web site: <http://www.ngdf.org.uk/>

Introduction to the NGDF action

The NGDF has developed guidelines on metadata standards indicating their choices of namespaces and how they use various standards in order to built their application profiles.

This NGDF document represents the first stage in the development of metadata services for the discovery of data resources that have a geographic component. The aim is to provide a consistent and simple method of documenting any data resources that are referenced in some way to the earth's surface whether by coordinates or geographic identifiers (addresses, administrative area, postcode area). The consistent recording of metadata or information about data resources and their presentation in catalogues accessible to the user community via the Internet considerably facilitates the discovery of such data resources.

There are a number of metadata standards in existence or under development. None of those existing were found to meet with the requirements of being simple and applicable to the full range of data resources that are geospatially referenced. In developing these NGDF Guidelines metadata standards which are not specifically for geospatial data such as the "Dublin Core" have also been examined.

These metadata standards have been produced for use in the United Kingdom. However, with the evolution of global information infrastructures, developments outside the UK cannot be ignored. Therefore the Guidelines must be regarded as interim pending the development of standards at the international level such as those being developed by Technical Committee 211 of the International Organisation for Standardisation (ISO).

Standards on which the NGDF metadata are based

The principals standards that the NGDF looked at for their application profiles are:

- Draft ISO Standard 15046-15 Geographic Information - Metadata
- European Standard prEN 287009

They took also in consideration the work of the United States Federal Geographic Data Committee (FGDC) that has developed a standard known as the Content Standard for Digital Geospatial Metadata and that has strong parallels with the draft ISO standard. A further initiative driven by the Open GIS Consortium (OGC) was also looked at.

The NGDF Guidelines that are provided are an interim solution pending the final release of the ISO, FGDC and OGC standards. It is actually hoped that these standards will converge, indeed the ISO, FGDC and OGC standards already have a great deal in common.

The definitions of NGDF are closely allied to the draft ISO standard. It is anticipated by the Working Group that these Guidelines will remain unaltered until the publication of the ISO standard in 2000 at which time it will be modified to match the ISO standard or possibly become a profile of the ISO standard.

The NGDF guidelines are based upon the draft ISO metadata standard 15046-15. The intention is to make the guidelines a profile of the ISO standard in time. They were produced after extensive research of existing standards and guidelines and following a workshop in which data producers were encouraged to compile metadata relating to their own datasets.

Some 42 metadata elements have been identified as necessary for documentation at the discovery level of which 16 are mandatory and a further 7 are conditional depending on the context, the remainder are optional. These elements cover title, theme, date and spatial extent, access constraints, nature of the resource, how to obtain additional information and data supply.

Finally, the "NGDF Profile" is actually a subset of the GEO profile developed by the FGDC. Although only a subset of full Z39.50 functionality, the GEO profile is over large for NGDF Discovery Metadata requirements. There are a few NGDF metadata attributes (i.e. Status of Start Date of Capture and Status of End Date of Capture, Alternative Title and level of Spatial Detail) that cannot be supported as there is no appropriate mapping to the GEO attribute sets. Until an NGDF Profile is defined these optional attributes will be excluded from the protocol.

5.7.6 ESMI, the European Spatial Metadata Infrastructure

The following information have been derived from the ESMI Web site: <http://esmi.geodan.nl/uk/standards.html>

ESMI has established a core metadataset much like the Dublin Core but based on the existing standards CEN, ISO and FGDC. As ESMI is aiming to specify a minimum set of descriptions to meet the needs of professional GIS users they have limited the mandatory fields to a minimum. Therefore, the core consists of a set of elements that are required, however some of these do not have to contain a meaningful content (content optional) as opposed to the content mandatory fields. This set was a technical requirements because the query interface needs to be able to query the various connected database for this core set of metadata elements. It is the responsibility of the data provider to be more specific and provide the best possible meta information.

ESMI proposes a metadata implementation based on existing standards that enable compliant metadata services to be searched over the Internet. Users of spatial data will be able to easily locate spatial data resources. To be compliant metadata services will need to make certain metadata fields accessible to the ESMI searching mechanism. The minimum set of metadata fields is the ESMI core metadata.

ESMI applies the CEN/TC 287 metadata standard. This is the metadata standard used in the ESMI project because many of the existing metadata services in Europe are based on the work of CEN/TC 287 and the standard will not change in the next two years. In 1992 the Comité Européen de Normalisation (CEN) created technical committee 287 with responsibility for geographic information standards. A family of European Prestandards have now been adopted including 'ENV (Euro-Norme Voluntaire) 12657 Geographic information - Data description - Metadata'.

Furthermore, the on-going work of ISO/ TC 211 will be considered during the work of ESMI to ensure as much compatibility as possible.

The core metadata

The following information has been based on <http://esmi.geodan.nl/uk/mapping.html>, where a document deals with core metadata, search profile and procedures for

managing semantics. The ESMI core metadata and search profile together with the controlled lists and definitions of technical terms have been agreed upon during a technical meeting to be used in this form for the first ESMI prototype.

There are 9 main metadata elements, composed by sub-element, recommended by ESMI and based on the CEN/TC 287 standard, (i.e. metadata element: dataset identification has 2 sub-elements: dataset title and abbreviated title). They represent the minimum selection of metadata elements needed for a provider to obtain ESMI certification. Other providers may also be accepted with other conditions. The elements of the core metadata as well as the recommended further metadata should be provided fully multilingual.

The necessary metadata elements for ESMI are separated into 'content mandatory' (CM) and 'content optional' (CO). Content mandatory - metadata elements that have to be provided through the server software and which MUST contain information. Only a limited number of metadata elements is classified as CM. Content optional - metadata elements that also have to be provided through the server software but are NOT required to have a 'real' content ('not available' is allowed).

ESMI has also some controlled lists for the thesaurus. Language translations of the keywords of the controlled lists are provided as well. The controlled lists apply for the following elements:

- Language
- Overall positional accuracy
- Datum
- Ellipsoid
- Map projection
- Vertical Datum
- Term
- Restrictions on use
- Price information
- Organisation role
- Point of contact role
- Formats

Regarding the semantics, a next phase is going to start using thematics thesauri that data providers can map to their own thesaurus, in this way users looking for data can create a query using the correct terms that are mapped to the terms of the local implementation.

The search interface, the most important issue for managing semantics is managing the issue of having different administrative boundaries and hierarchies. In ESMI, this issue will be solved by using MEGRIN's SABE file (Seamless Administrative Boundaries of Europe). SABE contains high quality data from 26 different national producers, which have been made homogeneous and form a unique and consistent product. SABE presents several features which make it unique. It is homogeneous, enabling the user to work across borders without incurring the risk of semantic variations on either side. Although it is derived from source data protected by

copyright in each country of origin, agreements have been signed between MEGRIN and each NMA to allow the marketing of all or part of SABE on the basis of a single ©MEGRIN licence. This simplification greatly helps user access. SABE is also regularly updated. To address the semantic issue within SABE itself, MEGRIN uses for the countries of the European Union the official NUTS nomenclature (Nomenclature of Statistical Territorial Units) and SABE presents it up to the NUTS 5 level ("the smallest administrative unit with its own elected assembly", i.e. the "commune" in France, or "ward" in England, ...).

In Appendix C, the application profiles of FGDC, ANZLIC, NGDF and ESMI are described.

6 References

- [1] Home page for the Dublin Core Metadata Initiative working group on Registries. URL: <http://purl.org/dc/groups/registry.htm>
- [2] Rachel Heery and Manjula Patel: Application profiles: mixing and matching metadata schemas. *Ariadne*, Issue 25, September 2000. URL: <http://www.ariadne.ac.uk/issue25/app-profiles/>
- [3] Introduction to the second SCHEMAS Workshop: Publishing and sharing your metadata application profile. A two-day workshop, 23-24th November 2000, Gustav-Stresemann-Institut, Bonn, Germany. URL: <http://www.schemas-forum.org/workshops/ws2/schemas-ws2.html>

Appendix A. Application profiles for audio-visual sector

	Digital Platform http://www.digitaalplatform.nl	European Chronicles Online (ECHO) http://pc-erato2.iei.pi.cnr.it/echo/	Digitaal Erfgoed Nederland http://www.den.nl
<i>Elements</i>	Proprietary list of app.210 attributes : -identification area -production area -administrative area -descriptive area -technical area -language area To be mapped with SMPTE Metadatadictionary, DC, IFLA Model, SMEF model BBC and MPEG-7. Using EBU AV-Genre Classification Escort 2.4	Basic list of attributes based on a selection from the current catalogue practice of the 4 content providers (NAA, INA, Instituto Luce, Memoriav) plus metadata fields for automated indexing of speech, closed captions and and images.	Referencelist: DC Combined with selection of metadata used by participating collections, museums and archives
<i>Schemes</i>	Proprietary datamodel To be mapped with SMEF	IFLA data model, adjusted to ECHO environment/requirements	Not yet decided
<i>Technology binding</i>	RDBMS, web based	RDBMS, web based	XML

Appendix B. Application profiles for educational sector

	ARIADNE	DC- EDUCATION	EDNA	EUN Schoolnet	GEM	IMS	NEEDS, SMETE
<i>Elements</i>	LOM (version 5), with 23 mandatory elements	DC with additional elements: <ul style="list-style-type: none"> - audience (as in LOM) - educational standard - interactivity type (from LOM) - interactivity level (from LOM) - typical learning time (from LOM) and qualifiers - conformsTo for Relation 	DC with additional elements: <ul style="list-style-type: none"> - entered - approver - reassessment - user level - categories - conditions - indexing - review - version 	DC with additional elements: <ul style="list-style-type: none"> - rights - approver - release - user level - version 	DC with additional elements: <ul style="list-style-type: none"> - audience - cataloging - duration - essential resources - grade - pedagogy - quality - standards 	LOM (version 3.5), with some minor modifications that have been integrated in subsequent versions of LOM and 20 core elements	own set (for search): title, contributor, publisher, subject heading, affiliates keywords, platform, MIME type

<i>Schemes</i>	own list of disciplines, subdisciplines and concepts		<ul style="list-style-type: none"> - Dates: ISO8601 - own list of types - own list of user levels - own list (for Australian context) for coverage 		controlled vocabulary for: <ul style="list-style-type: none"> - audience - format - grade - language - pedagogy - relation - resource type - subject 	suggestions (often multiple, with their origin and suggested applicability) for 12 elements	list of MIME types, affiliates and platforms
<i>Technology binding</i>	XML, DBMS		HTML META tags			XML DTD	

Appendix C. Application profiles for geographical information sector

	FGDC	ANZLIC	NGDF	ESMI
<i>Elements</i>	Own set: 220 elements composed of compound and data elements	Own set: 32 elements grouped in 9 categories based on the FGDC work	Own set: 42 elements, inspired from the ISO Standard 15046-15	9 elements composed of sub-elements based on the CEN/TC 287 standard
<i>Schemes</i>	List of extended elements but no specific controlled lists	No specific controlled lists	Use of a thesaurus based on predefined NGDF keywords and the HASSET thesaurus (Humanities and Social Science Electronic Thesaurus)	Controlled list for the thesaurus for the following elements: <ul style="list-style-type: none"> - Language - Overall positional accuracy - Datum - Ellipsoid - Map projection - Vertical Datum - Term - Restrictions on use - Price information - Organisation role - Point of contact role - Formats