CORES

Contract: IST-2001-35273

A forum on shared metadata vocabularies

Issues in cross-standard interoperability

D31

Document number:

CORES-PwC-WP1-D31-Final-20021008

General Information

Title	Issues in cross-standard interoperability
Creator	Makx Dekkers
Subject-Keywords	Interoperability, metadata
Description	The objective of this report is mainly to present and summarise the answers and opinions provided by the various respondents to the questionnaire whilst trying to identify major consensus and/or divergence areas. These issues shall then be used as a base for discussion at the CORES Interoperability Meeting
Publisher	PricewaterhouseCoopers
Contributor	Thomas Baker
Contributor	Christian Eilert
Date	8 October 2002
Type	Text Manuscript
Format	application/MSWord 2000
Identifier	CORES-PwC-WP3-D31-Final-20021008
Language	English
Rights	European Commission; public distribution on CORES Web site

Dublin Core metadata for this document

```
<?xml version="1.0"?>

<rdf:RDF xml:lang="en"

xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

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

<rdf:Description rdf:about =
"http://www.pwcextranet.lu/QuickPlace/cores/Main.nsf/h_Toc/2097BF6C4793768EC
1256BAB0050980A/?OpenDocument">

<dc:title> Issues in cross-standard interoperability </dc:title>

<dc:creator> Makx Dekkers</dc:creator>

<dc:subject> Interoperability </dc:subject>

<dc:subject> Metadata </dc:subject>

<dc:description> The objective of this report is mainly to present and
summarise the answers and opinions provided by the various respondents to
the questionnaire whilst trying to identify major consensus and/or
divergence areas. These issues shall then be used as a base for discussion
at the CORES Interoperability Meeting. </dc:description>

<dc:contributor> Thomas Baker </dc:contributor>

<dc:contributor> Christian Eilert </dc:contributor>

<dc:publisher> PricewaterhouseCoopers </dc:publisher>

<dc:date> 2002-10-08 </dc:date>

<dc:type> Text </dc:type>

<dc:format> application/MSWord </dc:format>

<dc:identifier> CORES-PwC-WP3-D31-Final-20021008</dc:identifier>

<dc:language> en </dc:language>

<dc:rights> European Commission; Public distribution on CORES Web site
</dc:rights>

</rdf:Description>
```

Table of Contents

1. Introduction

CORES is an Accompanying Measure under the Semantic Web action line aiming to serve the Community of metadata implementers, by both aiming to enhance interoperability between metadata solutions and to provide registration of metadata schemas.

One of the main objectives of the project is the organisation of a Metadata Interoperability Forum. In a meeting of a Metadata Interoperability Forum in November 2002, the CORES project will bring together key figures from major standardisation activities to discuss practicalities of cross-standard interoperability. The goal will be to discuss the feasibility of building semantic bridges between metadata standards with the objective of building a platform for future co-ordination and co-operation. This report summarises the responses submitted by potential participants in that forum to a series of open questions regarding the purposes and practicalities of cross-standard interoperability.

2. Scope

The discussion topics addressed at the Metadata Interoperability Forum will come from the various issues which were identified by the means of a survey and which are presented hereafter. The Forum shall bring together major standards activities from around the world, to discuss basic elements for cross-standard interoperability, on both strategic and technical levels, feeding the results into W3C's Semantic Web Activity and other international standards initiatives.

3. Objective

The objective of this report is mainly to present and summarise the answers and opinions provided by the various respondents to the questionnaire whilst trying to identify major consensus and/or divergence areas. These issues shall then be used as a base for discussion at the CORES Interoperability Meeting.

4. Methodology

Selection of participants

The selection of participants for the Interoperability Survey in CORES was driven by the intention to involve important metadata activities. Already during the proposal stage in 2001, we extended invitations to participate in the CORES project to a number of metadata activities with a global reach and concerned with description of electronic resources. We repeated this invitation in the first months after the start of the project in April 2002. In addition, we have also invited two major umbrella organisations that have an interest in metadata standardisation.

Approach

In order to collect opinions and views of the participating metadata standards activities, we have sent a questionnaire to key representatives of major standardisation activities – primarily from Standards Developing Organisations (SDOs) which develop and maintain metadata element sets from a semantic perspective but also from two key umbrella organisations for related standardisation work, the

Organization for the Advancement of Structured Information Standards (OASIS) and the World Wide Web Consortium (W3C).

The full text of the questionnaire is contained in Appendix A. Four areas are addressed in the questionnaire:

1. Opinions on diversity of standards and approaches to interoperability
2. Modelling issues and approaches underlying the various standards
3. Publication and maintenance issues
4. Registry issues

The evaluation of the responses to the questionnaire is not based on a statistical analysis. Such a statistical analysis would neither be possible with the number of respondents (10) nor useful for our purposes as this opinion-finding exercise uses mostly open questions to bring out discussion issues.

We also do not attribute opinions or statements to individual respondents or organisations, as we are interested in identifying issues for discussion in the CORES Interoperability Forum. The respondents were in broad agreement on many areas but with interesting points of disagreement; in our summary we try to do justice to both.

5. Results

Participants

The following initiatives have participated in the survey:

Acronym	***Full name***	***Description/objective***	***URL***
CERIF	Common European Research Information Format	Exchange of information about research programmes in Europe	http://www.cordis.lu/cerif/
DCMI	Dublin Core Metadata Initiative	Cross-domain resource discovery	http://dublincore.org/
DOI	Digital Object Identifier	Management of intellectual property in the digital environment	http://www.doi.org/
GILS	Government/Global Information Locator Service	Protocol and attributes for searching and retrieving metadata, initially for US government information	http://www.gils.net/
IEEE/LOM	Learning Object Metadata	Metadata for learning objects	http://ltsc.ieee.org/wg12/
MARC-21	Machine Readable Catalogue	Cataloguing format for materials held by libraries	http://www.loc.gov/marc/
MPEG-7	Moving Pictures Expert Group	Description of video, audio and multimedia content	http://mpeg.telecomitalialab.com/
OASIS	Organization for the Advancement of Structured Information Standards	Not-for-profit, global consortium driving the development, convergence and adoption of e-business standards	http://www.oasis-open.org/
ONIX		Exchange of book industry product	http://www.editeur.org/onix.html

		information in electronic form	
W3C	World Wide Web Consortium	Global forum leading and co-ordinating technical developments for the World Wide Web	http://www.w3.org/

The standards participating are described in a bit more detail in Appendix B.

Issue 1: Diversity of standards and approaches to interoperability

The existence of multiple metadata standards is a recognised fact of life. Different communities use different terms to describe their content, and do so from a practical need within the community, based on requirements of the scholarly or business processes that need to be supported. The respondent agreed that it is important that standards, once established, be adhered to so that metadata is reliable and consistent. There is a perceived need for a standard way of declaring metadata vocabularies to facilitate communication and achieve interoperability.

Although broad consensus on a single standard would provide a common basis for interoperability, it is not realistic to hope for a single standard that meets all needs. In specific communities, on the other hand, it can be beneficial to converge on a single standard in order to provide high-quality and consistent services. Interoperability with other communities and within other application contexts should then be achieved through mappings and transformations.

The respondents tend to see cross-standard interoperability not as a goal in itself, but as something that needs only to be considered where communities or services want to work together. Where there is a requirement to interoperate in specific cases, it would be helpful if there were well-accepted mechanisms, such as automated transformations, to support interoperability. The availability of machine-readable definitions of standards is a first requirement for enabling such automated processing.

Mixing and matching of different metadata standards is not a primary concern of most respondents, especially those with a strong domain perspective. Many work in environments where a single standard adequately meets the specific requirements of their domain, and mappings and transformations are used if and when exchange of information with others is needed. Mixing and matching only makes sense if the standards concerned can be reconciled through a process of semantic analysis or mapping. Some felt that standards can only be mixed sensibly if they are designed with this type of usage in mind, and can act as “building blocks” or modules in a larger framework.

As to the notion of ‘imperfect understanding’, so central to the Semantic Web vision, it is felt that there is a need to know which parts are understood and which are not. Broadly agreed semantics or higher-level ontologies could provide some help on this level, but some respondents point out that it can be dangerous for services to rely on and use information that is only partly understood – especially where transactions have legal or financial value or implications. This may be one reason why the concept of the Semantic Web has not proved very popular outside of the research domain. Generally, it was felt that the Semantic Web vision may not be realised for many years.

Issue 2: Modelling issues and approaches

Modelling is approached differently in the various standards communities. On one end of the scale, there may be a set of data types or elements defined pragmatically with regard to community culture or specific business needs; on the other end of the scale, there may be formally defined and logically consistent data models. The level of modelling seems to be related to the focus of the standard: where a standard aims at supporting descriptive and process requirements of a particular, clearly defined application, strict models and rules are put forward; where the primary concern is with information exchange or 'broad-brush' discovery, there is more emphasis on a flexible approach to allow data from different applications to be integrated into more loosely defined buckets.

All participating standards recognise the concept of 'data elements' in some form or other. Within a standard, these elements are identified with names or numbers, but not always with policies formalising these identities within an overarching namespace. In only a few cases namespace URIs have been assigned to elements for the purpose of unique identification within the Web context. Those who have not done so, however, acknowledge this as a possible first step towards automated mapping and translation mechanisms.

Many standards version both sets of elements as wholes and the individual elements within them.

Many of the standards surveyed have some notion of "application profile", or "profile", usually seen as defining a subset of the full standard for specific purposes, though some types of profiles can use elements or constructs from other standards.

Interoperability is seen as a problem to be addressed on several levels. To support specific applications or services, interoperability needs to be considered on the level of data models. On a broader level, there is a need to look at relations between individual elements.

Issue 3: Publication and maintenance

All of the standards are published as paper and Web documents. If standards are published in a machine-readable form, this publication is mostly in the form of XML DTDs or XML Schemas. RDF Schemas are only used in a very limited number of cases.

Metadata elements are themselves described with sets of attributes, but these attribute sets vary from standard to standard. Some follow ISO11179, while most follow their own internal schemas.

Issue 4: Registry services

In CORES terminology, a *metadata registry* is one of many emerging Web-based services that index the semantics of metadata standards for a wide range of potential uses, from search and look-up to automatic record conversion. Some registry-building initiatives focus on term-to-term metadata indexing, while others focus on providing full schemas for use in tool configuration or record verification.

Some of the CORES respondents look at registries primarily as information resources for the publication of a standard or information related to a standard (such as application profiles, policy disclosure, and guidelines).

In general, registries can be seen as tools to support the implementers of a standard, and standards developers themselves are not primarily concerned with them. Nevertheless, some of the standards developers involved in CORES can see the usefulness of registry services to be available to support the use of the standard over time and to ensure appropriate and consistent use. Also, standards developers themselves could use a metadata registry to find out about similar work in other domains and thereby reduce overlap and double work.

From the responses to the questionnaire, metadata registries may be used as services to store and find:

- The authoritative specification of a standard;
- Archived copies of its versions over time – supporting persistent understanding of legacy metadata, so that in ten years time it will still be possible to find what the semantics of a particular element were at a given point in time in the past;
- Translations and annotations of a standard in multiple languages;
- Application profiles, descriptions of specific usages of a standard in specific contexts;
- Mappings and crosswalks, including style sheets (such as XSLT) supporting transformations between standards
- Ontologies to be used as a tool to declare and relate semantics – this would reduce the risks of incorrect interpretation, misunderstanding and incorrect usage in transformations
- Various additional types of information to support the use of or be used in conjunction with a standard: thesauri, domain ontologies, controlled vocabularies such as code lists or pick lists and available software

The role of a metadata registry that would allow users to find appropriate standards for a particular use is not generally considered necessary or possible – it would require a sophisticated expert system that is difficult to build and maintain, and in practice most metadata implementers already have a fairly clear idea of the standard they want to use based on their domain knowledge.

6. Survey conclusions

Standards developers are interested in interoperability issues, as demonstrated by the fact that almost all standards activities contacted have agreed to participate in the CORES Interoperability Survey. Some standards serve specific communities or business sectors, while others are oriented to specific types of services or processes. However, all initiatives realise that their standard is applied in an environment where multiple standards exist, and that their standard does not necessarily satisfy all potential requirements.

Interoperability, however, is not a goal in itself. It is not surprising that standards developers are mostly concerned with satisfying the needs of their specific user communities and see interoperability with standards from other domains as a secondary issue. Many respondents would want to see what interoperability aims to

achieve in specific cases before they can consider mechanisms (mappings, transformations) and boundary conditions (how 'deep' understanding should be) to support interoperability. The 'Semantic Web vision' constitutes, as yet, insufficient rationale as a driver for this, until such time as practical applications can be demonstrated.

There is an emerging consensus that unique identification of the elements defined by a metadata standard, e.g. using URIs, would provide a basic mechanism supporting the expression of relationships between them, for example in mapping or transformation specifications and machine processing.

Metadata registries are seen as potentially helpful tools to support the users of metadata standards to find information (both machine- and human-readable) about or related to the application of a particular standard.

Appendix A. Questionnaire

```
-----------------------------------------------------------------

CORES Interoperability Forum Questionnaire

-----------------------------------------------------------------
Please take a stand on the following assertions -- either on
their inherent truth or on their desirability. Feel free to
reword or re-frame the assertion itself.
-----------------------------------------------------------------

1.  Just as lawyers, film directors, and chemists do not speak
    the same language, a diversity of metadata standards is
    inevitable and not entirely undesirable.

2.  Agreement on or normalization to one particular standard
    to the exclusion of others is not a worthy goal.  Rather,
    the point is to create Web services that aid in the
    (sometimes lossy) translation or transformation between
    different views of structured information.

3.  Well-defined conventions for declaring metadata
    vocabularies in standard, machine-processable formats (such
    as RDF schemas) would further the cause of interoperability.

4.  In practice, implementers tend to "mix and match" metadata
    standards pragmatically, drawing on multiple relevant
    standards, as needed, to construct a model for a particular
    project or application.

5.  The "Semantic Web" vision -- repurposing and merging data
    on the basis of "imperfect understanding" -- is the best we
    can hope for as a basis for interoperability on any really
    broad scale.

-----------------------------------------------------------------
Modelling issues
-----------------------------------------------------------------

1.  What is the grammar or data model underlying your standard?
    To what extent has this model been formally described?
    (please provide URL if available). How important in your
    model is logical consistency as opposed to pragmatic
    flexibility?

2.  Does your data model have "data elements" -- units of
    meaning analogous to "words" in natural language?  Do these
    data elements seem roughly comparable with data elements in
    other standards, either for the purpose of data merging or
    for the construction of cross-standard dictionaries or
    crosswalks?

3.  Do the elements of your standard have (or could they be
    assigned) unique identifiers in the Web context?  Could
    a common basis for such unique identification be provided
    by URIs?

4.  Is your standard versioned over time? If it is versioned, is
    this at the level of the entire vocabulary or at the level of
    individual terms? Do successive versions of a term share a
    common identity?

5.  Does your standard have a notion of Profile, or Application
    Profile?   Please describe the philosophy and purpose of such
    Profiles in a paragraph or two.  Are Profiles limited to
    subsets of your standard, or can they in principle include
    elements or constructs from other standards?

6.   From your perspective, should interoperability efforts
```

focus most on the level of individual elements (with element-to-element references and crosswalks); on the level of schemas (sets of elements usable as templates for instance metadata); or on the level of data models?

Publication and maintenance of standards declarations

---

1. In what forms do you publish your metadata terms and definitions (e.g. as Web pages, printed documents, XML schemas, RDF schemas; please provide URLs if available)? Are any of these forms more "authoritative" than others?

2. Have these forms of publication themselves been formalized? Do they follow particular data-description standards such as ISO 11179?

Metadata registries

---

1. Please comment on the following user-need scenarios for Web-based registry services, describing how you imagine such needs might be met and what this implies for element set providers. If a scenario does not seem realistic, please explain.

   a. A cataloger needs to find best practice in applying a standard to the description of a particular set of materials.

   b. The implementers of an information service need to find out what standards and elements are relevant to designing a schema for their metadata.

   c. An information provider needs to be able to translate local metadata into other formats in order to export records or provide particular views of its data.

   d. A software developer wants its metadata management utilities to configure themselves with the most up-to-date specifications of an element set.

   e. A quality controller wants to discover and eliminate errors, redundancy, and bad practice from a body of records.

   f. Ten years from now, an information service wants to know what a standard said in the year 2002 for the purposes of migrating legacy data.

   g. Any of the above want to use the element sets in their Japanese translation.

2. What other types of registry services do you foresee?

3. What role might metadata vocabulary providers have in the realisation of registry services?

Appendix B. Description of participating initiatives

CERIF

The Common European Research Information Format (http://www.cordis.lu/cerif/) provides a set of guidelines for data exchange of research information. The data model is defined in terms of entities, attributes and relationships. There are three primary object types: project, person and organisation. Data elements are defined for these entities, e.g. names and titles, dates, keywords etc.

DCMI

The Dublin Core Metadata Initiative (http://dublincore.org/) is a worldwide grouping of practitioners from various domains (libraries, education, government, business etc.) interested in providing a means to facilitate interoperability for resource discovery across domains using a set of 15 metadata elements plus a set of qualifiers for these elements. Core elements include for example: Title, Creator, Publisher, Subject, Identifier, Coverage etc.

DOI

The Digital Object Identifier (http://www.doi.org/) is a system for persistently identifying and managing intellectual property in the digital environment. It is developed and maintained by the IDF, the International DOI Foundation, a membership organisation aiming to provide a framework for managing intellectual content, for linking customers with content suppliers, for facilitating electronic commerce, and enabling automated copyright management for all types of media. Metadata is considered to be an essential component of the DOI System in order for it to be able to provide services relating to intellectual property in the network environment. The DOI specifies 10 basic attributes for resources (e.g. Identifier, Name, Agent, Category, Time and Place etc.).

GILS

The GILS initiative (which originally in the US stood for Government Information Locator Service, now in wider contexts for Global Information Locator Service, http://www.gils.net/) defines a mechanism based on ANSI Z39.50/ISO 23950 to search and retrieve metadata without prescribing the way that this metadata is physically held by the information provider, defining search attribute concepts and data attribute concepts that are used as an intermediary language to access indexes and transfer resulting metadata records. It is not technically a metadata initiative as it only defines an intermediary exchange mechanism between systems with their own metadata solutions.

IEEE/LOM

The purpose of the Standard for Learning Object Metadata (LOM) is to facilitate search, evaluation, acquisition, and use of learning objects, for instance by learners or instructors. The work has its origins in both the Ariadne and IMS projects, and also builds on metadata work done by the Dublin Core group. Data elements describe a learning object and are grouped into categories. The Base Scheme consists of nine such categories: General; Lifecycle; Meta-metadata; Technical; Educational; Rights; Relation; Annotation; Classification.

MARC21

The MARC (MAchine Readable Catalogue) format was developed in the 1960s to provide a standard for the creation of machine-readable catalogue records for books and journals. It now also supports a multitude of formats of materials held in libraries. The description uses numerically tagged fields with subfields.

MPEG-7

This is the standard for description and search of audio and visual content defined by the Moving Picture Experts Group (http://mpeg.telecomitalialab.com/). It contains a large number of descriptive elements for video, audio and multimedia content.

OASIS

OASIS (http://www.oasis-open.org/) is a not-for-profit, global consortium that drives the development, convergence and adoption of e-business standards. Members themselves set the OASIS technical agenda, using a lightweight, open process expressly designed to promote industry consensus and unite disparate efforts. OASIS produces worldwide standards for security, Web services, XML conformance, business transactions, electronic publishing, topic maps and interoperability within and between marketplaces.

OASIS has more than 500 corporate and individual members in 100 countries around the world. OASIS and the United Nations jointly sponsor ebXML, a global framework for e-business data exchange. OASIS operates XML.org, a community clearinghouse for XML application schemas, vocabularies and related documents. OASIS hosts The XML Cover Pages, an online reference collection for interoperable markup language standards. The OASIS Network includes UDDI, CGM Open and LegalXML.

ONIX

ONIX (http://www.editeur.org/onix.html) is an international standard for representing and communicating book industry product information in electronic form. It provides a format for exchange of information between systems that may, internally, have their own metadata formats. Data elements are defined for message headers, reference and product numbers, title, authorship, extent, subject, publisher, etc.

W3C

The World Wide Web Consortium (W3C) is a forum for information, commerce, communication, and collective understanding related to the World Wide Web, developing common protocols that promote its evolution and ensure its interoperability. This membership organisation has currently over 500 members from academia and industry. The Semantic Web is an initiative of W3C, based on a vision that the World Wide Web should evolve to a universally accessible platform that allows data to be shared and processed by automated tools as well as by people.