

---

**ARTICLES**

---

**D-Lib Magazine**  
**April 2002**

Volume 8 Number 4

ISSN 1082-9873

## Metadata Principles and Practicalities

Erik DuvalKatholieke Universiteit Leuven, Belgium  
<[Erik.Duval@cs.kuleuven.ac.be](mailto:Erik.Duval@cs.kuleuven.ac.be)>Wayne HodginsStrategic Futurist  
Autodesk  
<[wayne.hodgins@autodesk.com](mailto:wayne.hodgins@autodesk.com)>Stuart SuttonAssociate Professor, The Information School  
University of Washington  
<[sasutton@u.washington.edu](mailto:sasutton@u.washington.edu)>Stuart L. WeibelExecutive Director  
Dublin Core Metadata Initiative  
<[Weibel@oclc.org](mailto:Weibel@oclc.org)>

---

### I. Introduction

The rapid changes in the means of information access occasioned by the emergence of the World Wide Web have spawned an upheaval in the means of describing and managing information resources. Metadata is a primary tool in this work, and an important link in the value chain of knowledge economies. Yet there is much confusion about how metadata should be integrated into information systems. How is it to be created or extended? Who will manage it? How can it be used and exchanged? Whence comes its authority? Can different metadata standards be used together in a given environment? These and related questions motivate this paper.

The authors hope to make explicit the strong foundations of agreement shared by two prominent metadata Initiatives: the Dublin Core Metadata Initiative (DCMI) and the Institute for Electrical and Electronics Engineers (IEEE) Learning Object Metadata (LOM) Working Group. This agreement emerged from a joint metadata taskforce meeting in Ottawa in August, 2001. By elucidating shared principles and practicalities of metadata, we hope to raise the level of understanding among our respective (and shared) constituents, so that all stakeholders can move forward more decisively to address their respective problems.

The ideas in this paper are divided into two categories. *Principles* are those concepts judged to be common to all domains of metadata and which might inform the design of any metadata schema or application. *Practicalities* are the rules of thumb, constraints, and infrastructure issues that emerge from bringing theory into practice in the form of useful and sustainable systems.

### II. Principles

The paragraphs in the Principles section set out general truths the authors believe provide a guiding framework for the development of practical solutions for semantic and machine interoperability in any domain using any set of metadata standards.

#### A. Modularity

Metadata modularity is a key organizing principle for environments characterized by vastly diverse sources of content,

styles of content management, and approaches to resource description. It allows designers of metadata schemas to create new assemblies based on established metadata schemas and benefit from observed best practice, rather than reinventing elements anew.

In a modular metadata world, data elements from different schemas as well as vocabularies and other building blocks can be combined in a syntactically and semantically interoperable way. Thus, application designers should be able to benefit from significant re-usability as they gather existing modules of metadata and 'snap' them together much as individual Lego™ blocks can be assembled into larger structures. The appeal of the Lego™ metaphor has partly to do with the underlying engineering and design that sustains 'interoperability' across many years of evolution, and partly from the variety of 'semantics' reflected in the various themes of Lego™ sets.

Children think nothing of mixing cowboy themes and pirate themes and undersea exploration themes. While the 'semantics' of such combinations may not always be obvious to adults, children don't seem to be bothered by such incongruities. Similar flexibility should be achievable in the metadata architecture of the Web, allowing application designers to mix a variety of semantic modules within a common syntactic foundation, even though the designers of the modules might not have anticipated a given combination. For example, a discovery metadata module and an instructional management metadata module, expressed in a common syntactic idiom such as XML, should be able to be combined in a compound schema that embodies the functionality of each constituent. In this way, modular sets can be assembled to meet the specific requirements of a given application, meeting domain-specific and local requirements without unduly sacrificing cross-domain interoperability.

### **Namespaces and metadata modularity**

The notion of namespaces is a fundamental part of the infrastructure of the Web (and particularly XML [[NAMES](#)]), though the concept predates the Web and is familiar to most. Simply put, a namespace is a formal collection of terms managed according to a policy or algorithm. For example, the base protocol of the Web is HTTP, which is a namespace that guarantees that a given URI is globally unique. LCSH (Library of Congress Subject Headings) is a namespace managed by the U.S. Library of Congress according to rules governing the assignment of subject headings to intellectual artifacts. Any metadata element set is a namespace bounded by the rules and conventions determined by its maintenance agency.

The technicalities of declaring and managing namespaces in an XML environment are beyond the present discussion, but the idea is a critical part of the infrastructure necessary for deploying modular metadata systems on the Web. Namespace declarations allow the metadata schema designer to define the context for a particular term, thereby assuring that the term has a unique definition within the bounds of the declared namespace. Thus, the declaration of various namespaces within a block of metadata allows the elements within that metadata to be identified as belonging to one or another element set.

Expressed as natural language, such a declaration might read:

The Dublin Core metadata element set is defined at a Web location specified by a URI; all Dublin Core elements within the scope of this namespace declaration can be recognized by the prefix *dc*::

The IEEE-LOM metadata element set is defined at a Web location specified by a URI; all IEEE-LOM elements within the scope of this namespace declaration can be recognized by the prefix *lom*::.

Using this infrastructure, metadata system designers can select elements from suitable existing metadata element sets, taking advantage of the investment of existing communities of expertise, and thereby avoid reinventing well-established metadata sets for each new deployment domain.

### **B. Extensibility**

Metadata systems must allow for extensions so that particular needs of a given application can be accommodated. Some metadata elements are likely to be found in most metadata schemas (the concept of *creator* or *identifier* of an information resource, for example). Others will be specific to particular applications or domains (*degree of cloud cover*, for example, in remote sensing data).

Metadata architectures must easily accommodate the notion of a base schema with additional elements that tailor a given application to local needs or domain-specific needs without unduly compromising the interoperability provided by the base schema. Another application encountering such extensions should be able to ignore such extensions while making use of any elements understood by both.

## C. Refinement

Application domains will differ according to the degree of detail that is necessary or desirable. The design of metadata standards should allow schema designers to choose a level of detail appropriate to a given application. Populating databases with metadata is costly, so there are strong economic incentives to create metadata with sufficient detail to meet the functional requirements of an application, but not more.

There are two notions of refinement to consider. The first is the addition of qualifiers that refine or make more specific the meaning of an element. *Illustrator*, *author*, *composer*, or *sculptor* are all examples of particular types of the more general term, creator. *Date of creation*, *date of modification*, and *date of acceptance* are all narrower senses of a date attribute. Such refinements might be useful or even essential in a given metadata application, but for general interoperability purposes, the values of such elements can be thought of as subtypes of a broader element.

A second variety of refinement involves the specification of particular schemes or value sets that define the range of values for a given element. Thus, identifying that a metadata value has been selected from a controlled vocabulary or has been constructed according to a particular algorithm may make it much more useful, especially for automated processing. In this way, semantic interoperability across applications can be increased, by relying on a common value set.

The encoding of dates and times is an example of the use of an encoding standard to remove ambiguity from the expression of a metadata value. The string *03/06/02* is interpreted as *March 6, 2002* in North America and *June 3, 2002* in Europe and Australia. By using an encoding standard such as the W3C date and time format [[W3C-DTF](#)], a date can be encoded in an unambiguous manner (*2002-03-06*). Specifying the encoding format in the metadata allows unambiguous machine processing as well as improving human comprehension.

The use of controlled vocabularies is another important approach to refinement that improves the precision for descriptions and leverages the substantial intellectual investment made by many domains to improve subject access to resources. The Dewey Decimal Classification System, for example, affords a multilingual classification system long used in traditional library environments that can be applied to electronic resources as well. There are hundreds of domain-specific thesauri and classification systems, as well, that can be imported into the Web metadata architecture to support subject descriptions. Specifying the use of a particular vocabulary in a given collection of metadata will allow applications to provide more coherent search and browsing facilities. Even in cases where an application is not designed to take advantage of a classification scheme or thesaurus, users may still benefit from the inherent coherence that such a scheme affords.

## D. Multilingualism

It is essential to adopt metadata architectures that respect linguistic and cultural diversity. The Web as a global information system is important in that it affords unprecedented access to resources of global scope. However, unless such resources can be made available to users in their native languages, in appropriate character sets, and with metadata appropriate to management of the resources, the Web will fail to achieve its potential as a global information system.

Standards typically deal with these issues through the complementary processes of *internationalization* and *localization*: the former process relates to the creation of "neutral" standards, whereas the latter refers to the adaptation of such a neutral standard to a local context.

It is important to note that these two processes can sometimes work at cross-purposes. While global resource discovery is best served by internationalization (common conventions of practice, languages, and character sets), the needs of any given community may be better served by supporting local conventions. One of the challenges for a global metadata architecture is to assure that the underlying infrastructure can support either strategy equally well, or a mix of the two. Thus, a given application will reflect design choices based on an understanding of this balance and its implications.

A basic starting point in promoting a global metadata architecture is to translate relevant specification and standards documents into a variety of languages. DCMI maintains a list of translations of its basic documents. Likewise, the European workshop on Learning Technologies is maintaining translations of the LOM specification.

Another essential dimension is to include provisions in the metadata for the description of lingual and other cultural aspects of a resource. For example, metadata can describe the language and character set of the resource. The metadata may identify alternative versions of resources, in different languages, as well as the origin of the translations.

On a somewhat more technical level, it is important for global adoption of the standards that both the specifications and the ways these specifications are encoded are as "culturally neutral" as possible. As an example, it would be inappropriate to define the value space of a data element such as educational context in a way that is specific to one national system. Likewise, encodings will often be based on numerical representations of elements or their values, although there is wide practice to use some form of "pseudo-English" as well... (HTML tags are a typical example: the <LI> tag refers to the notion of a "List Item" and is thus somewhat biased linguistically.)

Multilingualism is one aspect of the broader issue of multiculturalism, which includes, for instance:

- The way in which dates are represented in different calendars,
- The direction in which text is displayed and read,
- Cultural connotations of certain icons and pictograms,
- Standards of practice (name order, collation standards, leading article standards).

Clearly, many of these aspects go beyond the immediate context of metadata. However, as mentioned above, it is important that metadata can describe the relevant characteristics, and that it can do so in ways that respect cultural and language differences.

### III. Practicalities

The metadata principles as set out above, lead, at a minimum, to the following practicalities. These practicalities represent aspects of the emerging ecology of metadata creation and management on the Internet.

#### A. Application Profiles

No single metadata element set will accommodate the functional requirements of all applications, and as the Web dissolves access boundaries, it becomes increasingly important to be able to also cross discovery boundaries. Application profiles will facilitate this by allowing designers to 'mix and match' schemas as appropriate.

An application profile is an assemblage of metadata elements selected from one or more metadata schemas and combined in a compound schema. Application profiles provide the means to express principles of modularity and extensibility. The purpose of an application profile is to adapt or combine existing schemas into a package that is tailored to the functional requirements of a particular application, while retaining interoperability with the original base schemas. Part of such an adaptation may include the elaboration of local metadata elements that have importance in a given community or organization, but which are not expected to be important in a wider context.

One of the benefits of this approach is that communities of practice are able to focus on standardizing community-specific metadata in ways that can be preserved in the larger metadata architectures of the Web. It will be possible to snap together such community-specific modules to form more complex metadata structures that will conform to the standards of the community while preserving cross-community interoperability.

Application Profiles achieve this modularity through a number of mechanisms:

1. **Cardinality enforcement:** Cardinality refers to constraints on the appearance of an element. Is it optional? Mandatory? Conditional? The status of some data elements can be made more stringent in a given context. For instance, an optional data element can be made mandatory in a particular application profile. A typical example would be an element that specifies the human language of a resource: such an element can be made mandatory in a multi-lingual community. Along the same lines, an application may make the status of an optional element conditional, or a conditional element mandatory. As an application profile must operate within the interoperability constraints defined by the standard, it cannot relax the status of data elements.
2. **Value Space Restriction:** For some data elements, the value space can be made more restrictive than in the standard.
  - a. This mechanism can apply when the standard is very loose about the values for a data element. A typical example is the restriction of values about people involved in the life cycle of a resource to references into a registry of people and organizations (e.g., as an LDAP service).
  - b. The same mechanism can also apply when the standard is already quite explicit about the value space, when the context of use allows for further restrictions. A typical example can restrict reference to the

human language of a resource (typically an extensive list, as defined in ISO 639) to those languages that are relevant in a particular community.

3. **Relationship and dependency specification:** An application profile can define interrelationships between data elements and their value spaces. For instance, the presence of one data element may impose the requirement that another element be present. Similarly, an application profile can restrict the value set of a data element, based on the value of another data element. A typical example would restrict the value space of the data type of a resource, based on its genre: for instance, a 'text' document cannot be of type MP3.
4. **Declaration of namespaces:** Application profiles support the use of multiple namespaces, such that designers may choose elements appropriate to their needs from various different element sets. Schema designers may also add local elements through the use of a locally defined namespace.

As described in an earlier section, namespace declarations are the XML infrastructure that allows the construction of mixed metadata sets within an application profile. A schema designer can invoke several such declarations to include elements from existing schemas that can be combined in a modular way to form a compound schema that meets the functional requirements of an application without destroying the possibility of interoperability with existing schemas that also use these elements in other combinations.

The main goal of application profiles is to increase the "semantic interoperability" of the resulting metadata instances within a community of practice, by going beyond the universal consensus of a single standard, without compromising the basic interoperability that the standard enables across the boundaries of these communities. [[SIGMOD](#)].

## B. Syntax and Semantics

Semantics is about meaning; syntax is about form. Agreements about both are necessary for two communities to share metadata. Two communities may agree about the meaning of the term title or creator or identifier, but until they have a shared convention for identifying and encoding values, they cannot easily exchange their metadata.

It is important, however, to keep syntax and semantics separate as far as possible. The rapid changes of the first decade of the Web illustrate this well. We have witnessed several versions of HTML, the emergence of XML, and the development of derivative technologies that include at this time both XML Schemas and RDF Schemas. The lack of stability in the structured markup realm emphasizes the necessity of maintaining independence between the semantics of metadata elements and their syntactic representation. However, as more information is 'born digital', one expects metadata facilities to be an intrinsic part of the creation and management of the resources, so issues of syntax cannot be ignored even though we are in general more concerned with the meaning of metadata statements rather than how they are exchanged.

At this writing it is not possible to predict which, if any, of the various metadata encoding schemes will prevail. A few observations are appropriate, however.

HTML-encoded metadata accounts for the majority of metadata embedded within Web resources (and hence available for harvesting). This approach has the great virtue of simplicity (no additional systems are necessary—Web infrastructure provides the system in the form of HTML markup and http protocols), but it limits the structural richness of the metadata assertions that can be made.

XML markup, while still a small part of the total markup on the Web, is the idiom of choice for the encoding and exchange of structured data. The XML namespace facility provides structural capabilities that HTML lacks, making it easier to achieve the principles of modularity and extensibility. The XML Schema specification defines a schema language that allows for the specification of application profiles that will increase the prospects for interoperability.

The Resource Description Framework (RDF) promises an architecture for Web metadata and has been advanced as the primary enabling infrastructure of the Semantic Web activity in the W3C. Designed to support the reuse and exchange of vocabularies, RDF is an additional layer on top of XML that is intended to simplify the reuse of vocabulary terms across namespaces. Most RDF deployment to date has been experimental, though there are significant applications emerging in the world of commerce (Adobe's deployment of their XMP standard which is based on RDF).

The IEEE Learning Object Metadata standard provides an example of how this critical need for independence between the semantics of metadata and their syntactical representation can be addressed. LOM will be what is known as a "multi-part standard" where the semantic data model is an independent standard and then each syntactical representation is an independent standard developed as a specific "binding" of the LOM Data Model standard. DCMI

also provides recommendations on encoding of Dublin Core metadata in alternative encoding idioms.

Finally, it should be noted that there is a third requirement beyond syntax and semantics for interoperability: content vocabularies. This may be as open and unconstrained as a shared natural language (English, Dutch, German...). The use of a specific controlled vocabulary or namespace will further narrow the scope and increase the precision of a description, as discussed elsewhere in this paper.

### C. Association Models

There are various ways to associate metadata with resources:

**Embedded metadata** resides within the markup of the resource. This implies that the metadata is created at the time that the resource is created, often by the author. Experts differ concerning whether author-created metadata is best or whether it is better to have trained practitioners evaluate and describe resources. As a practical matter, resource description expertise is a scarce and costly commodity, and thus any investment by authors in the description of their intellectual products is likely to be of value.

Embedded metadata can also be harvested, and the presumptive increase in visibility that might result is an incentive for creators to assign metadata. Early studies of the efficacy of such metadata are only recently becoming available [[GRE-01](#)].

**Associated metadata** is maintained in files tightly coupled to the resources they describe. Such metadata may or may not be harvestable. The advantage of associated metadata derives from the relative ease of managing the metadata without altering the content of the resource itself, but this benefit is purchased at the cost of simplicity, necessitating the co-management of resource files and metadata files.

**Third-Party metadata** is maintained in a separate repository by an organization that may or may not have direct control over or access to the content of the resource. Typically such metadata is maintained in a database that is not accessible to harvesters, though the emerging Open Archives Initiative Metadata Harvesting Protocol proposes a system that encourages the disclosure of metadata repositories among federated OAI servers [[OAI-02](#)].

Syntax issues and association models are often confused. Many assume HTML based metadata is equivalent to embedded metadata, and that other representations are necessarily other types. Any of these three syntactic idioms can easily be embedded within the markup of an electronic resource or managed as a separate entity.

A given information resource will often have multiple metadata records reflecting the various purposes and perspectives of the organizations that create and manage them. A resource may be created with embedded metadata supplied by the author. A separate record might be created by the issuing organization (an academic department or publisher, for example) and stored in a separate database. A third party (perhaps a library) might create yet another version of metadata, either from scratch or derived from a previous record. In most cases these records will not be managed in a coordinated way, and differences may arise among them that may cause ambiguity or confusion. This may be less than ideal, but must be expected in an environment where various organizations may choose to manage resource descriptions with different objectives.

### D. Identifying and Naming Metadata Elements: Tokens Versus Labels

The global scope of the Web URI namespace means that each data element in an element set can be represented by a globally addressable name (its URI). Invariant global identifiers make machine processing of metadata across languages and applications far easier, but may impose unnatural constraints in a given context.

Identifiers such as URIs are not convenient as labels to be read by people, especially when such labels are in a language or character set other than the natural language of a given application. People prefer to read simple strings that have meaning in their own language. Particular tools and applications can use different presentation labels within their systems to make the labels more understandable and useful in a given linguistic, cultural, or domain context.

### E. Metadata Registries

Metadata registries represent an important topic of digital library research at this time. As the number of metadata and application profile schemas designed to meet the needs of particular discourse and practice communities increases, the importance of the management and disclosure roles of registries will similarly increase. The expectation is that registries will provide the means to identify and refer to established schemas and application profiles, potentially including the

means for machine mapping among different schemas. In addition, it is expected that such registries will contain, or link to, important controlled vocabularies from which the values of metadata fields can be selected.

Such registries will assume the characteristics of an electronic dictionary, available for consultation by:

- Application designers, who will be able to consult registries to identify existing metadata schemas and schema components that might meet their needs or to identify extensions to those schema that other application designers have developed to meet a given local need.
- Creators and managers of metadata, who can consult a registry to ascertain the definition or usage statements concerning an element or the available or preferred candidate value sets to be used to populate particular elements.
- Applications, which can resolve URIs associated with a schema, an element, or a value set in order to compare or evaluate elements or their values in a set of metadata.
- End users, who might consult a registry to better understand definitions or context of metadata terms, and thereby improve their search or processing effectiveness.

Thus, registries will provide the means to manage and disclose metadata schema declarations, application profile declarations, and value space declarations. As any given metadata schema or application profile evolves, registries will maintain the relationships among that schema's various versions in order to promote semantic and machine interoperability over time [[HEE-00](#)].

The DCMI Registry Working Group is exploring some of these issues through the explication of functional requirements for a multilingual DCMI metadata registry and vocabulary management system. Initial prototypes for this system can be accessed at [[DC-REGISTRY](#)].

It is likely that registries will vary in the depth of their functionality with some being simple links to schema declarations while others may be richly functional databases. Some registries will be managed by namespace authorities and will hold the canonical copies of schema and value space declarations while other registries will harvest those declarations from such authoritative sources and thereby make them available in a more distributed manner [[HEE-00](#)].

## F. Completeness of Description

There is a strong inclination on the part of creators of metadata to 'fill in all the blanks.' If an element is available, people want to use it in a description. Applications should be designed to make evident that not every available element is necessarily appropriate for every resource type. Similarly, applications should provide assistance where possible in selection of an appropriate value for a particular element. To the extent that metadata creation facilities are built into content creation applications, the application can identify values for some elements more reliably than the user.

Ultimately, the richness of metadata descriptions will be determined by policies and best practices designated by the agency creating the metadata, and those policies and practices will be guided by the functional requirements of services or applications. Some of the tradeoffs for systems and searchers:

### Detailed metadata descriptions:

- may improve searching precision
- require higher investment in creation of metadata
- make it more difficult to promote consistency in creation of metadata

### Simple descriptions:

- are easier and less costly to generate
- may result in more false results, or more effort on the part of searchers to identify most relevant results
- improve probability of cross-disciplinary interoperability

## G. Mandatory Versus Optional Elements

Designing metadata standards for a global, cross-disciplinary information environment requires a high degree of flexibility. An element that is essential in one domain may not even be sensible in another, hence few, if any, elements in

a general metadata set should be thought of as mandatory.

On the other hand, it is entirely reasonable within a given application or even an application domain, to require particular elements. Thus, communities of practice should be encouraged to further specify standards of practice for a given metadata standard that will encourage uniformity of descriptions within a given domain. This can be done in the form of an application profile as described earlier, and shared with others within a community of practice in order to promote convergence and thereby increase interoperability.

## **H. Subjective and Objective Metadata**

Metadata is broadly defined as structured data about data. However, the process of creating metadata can involve both subjective and objective input. Some metadata is clearly objective: assertions of fact about authorship, date of creation, version, and other attributes are generally able to be determined in an objective way. This objective metadata can also be machine generated in most instances, such as the "properties" metadata generated when creating a file in a word processor or spreadsheet application.

Other metadata may be subjective, either because such elements are subject to differing points of view (assignment of keywords, summarization of content in an abstract), or because they are specifically intended to represent a subjective evaluation (a review of a book or a presentation). Even more formal metadata elements become subjective when used within a cultural or domain context that is subject to local interpretation. For example, a pedagogical characteristic that is dependent on a particular educational philosophy may be important within a given context, but will have no meaning outside that context. The requirement for metadata design is, as far as possible, to make that context explicit so that applications can more easily recognize when a given element is constrained by such context as opposed to being more broadly applicable.

## **I. Automated Generation of Metadata**

Most resource discovery metadata prior to the Web was created by humans in the labor-intensive activity of library cataloging. Cataloging metadata remains the most successful standard for resource discovery of books and periodicals, but it is costly to create and impractical for many materials available on the Internet.

Web search engines harvest and index a significant portion of the Internet and provide low cost index access to it, generally in an advertiser-supported model. Such indexing can be thought of as a kind of metadata, and for many information needs, it provides a surprisingly cost effective solution to resource discovery.

Between these two extremes lies a broad range of metadata creation that can be automated to some degree, and which can be expected to grow in importance as advances in such areas as natural language processing, data mining, profile and pattern recognition algorithms become more effective.

Content creation applications (word processors, electronic paper such as PDF, and Website creation tools) often have facilities for author-supplied attributes or automated capture of attributes that can simplify the creation of metadata. As these facilities grow more sophisticated, it will be easier and more natural to combine application-supplied metadata (e.g., creation dates, tagged structural elements, file formats and related information), creator-supplied metadata (keywords, authors, affiliations, for example) and inference-based metadata (classification metadata based on automated classification algorithms, for example). Combining attributes from these approaches will increase the quality and reduce the cost of metadata descriptions.

## **IV. Conclusions**

Metadata is a key part of the information infrastructure necessary to help create order in the chaos of the Web, infusing description, classification, and organization to help create more useful stores of information. Sources of metadata, like the sources of the resources themselves, will be of different quality and organized around different purposes to reflect the different objectives and business models of information providers. The social policies, organizational priorities, and market forces that shape the information spaces of the Web will undoubtedly create unforeseen opportunities and niches.

For these opportunities to be realized, some convergence of encoding formats and commonly agreed semantics will be necessary. This paper expresses some common understandings about metadata principles and practicalities that two metadata communities agree to be at the heart of their work. It is worthy of note that these commonalities did not emerge by design or intentional agreement, but rather are the expressions of years of independent work and the development of community practices. It has been encouraging to find the degree of convergence among our communities. The authors offer this distillation in hopes that not only our own, but other constituencies will find it useful

for enrichment of the intellectual Commons we share.

## V. Acknowledgements

The authors would like to acknowledge the critical attention of the following, whose suggestions and perspectives helped to shape this common vision:

Makx Dekkers	Managing Director, Dublin Core Metadata Initiative
Jon Mason	Assistant Director, IMS Australia
Neil Mclean	Director, IMS Australia
Robby Robson	Chair, IEEE Learning Technology Standards Committee
Ed Walker	CEO, IMS Global Learning Consortium, Inc.

## VI. References

### [DC-REGISTRY]

The Open Metadata Registry Prototype [Home Page]. Accessed February 2002. <<http://wip.dublincore.org:8080/registry/Registry>>

### [GRE-01]

Greenberg, Jane, M. Pattuelli, B. Parsia, W. Robertson, Author-generated Dublin Core Metadata for Web Resources: A Baseline Study in an Organization, *Journal of Digital Information*, volume 2 issue 2 (November 2001) <<http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Greenberg>>

### [HEE-00]

Heery, Rachel and Manjula Patel, Application Profiles: Mixing and Matching Metadata Schemas, *Ariadne*, Issue 25 (September 2000) <<http://www.ariadne.ac.uk/issue25/app-profiles/intro.html>>

### [NAMES]

*Namespace in XML*, World Wide Web Consortium, 14-January-1999, Editors: Tim Bray, Dave Hollander, and Andrew Layman, <<http://www.w3.org/TR/REC-xml-names>>

### [OAI-02]

*The Open Archives Initiative Protocol for Metadata Harvesting*, Protocol Version 1.1 of 2001-07-02, Document Version 2001-06-20, <<http://www.openarchives.org/OAI/openarchivesprotocol.htm>>

### [SIGMOD]

Special section on semantic interoperability in global information systems, *ACM SIGMOD Record*, Vol. 28, No. 1, March 1999. <<http://www.acm.org/sigmod/record/issues/9903>>

### [W3C-DTF]

*Date and Time Formats* [Technical Note] Misha Wolf and Charles Wicksteed 1998-08-27 <<http://www.w3.org/TR/NOTE-datetime>>

## VII. Further Reading

Baker, Thomas.

A Grammar for Dublin Core *D-Lib Magazine*, Vol. 6 No. 10 <<http://dlib.org/dlib/october00/baker/10baker.html>>

Dekkers, Makx and Stuart Weibel.

Dublin Core Metadata Initiative Progress Report and Workplan for 2002. *D-Lib Magazine*, Volume 8, #2 February 2002 <<http://www.dlib.org/dlib/february02/weibel/02weibel.html>>

Duval, Erik.

Metadata standards: What, who & why. *Journal of Universal Computer Science*, 7(7):591-601, July 2001. Special Issue: I-Know 01 - International Conference on Knowledge Management. <[http://www.jucs.org/jucs\\_7\\_7/metadata\\_standards\\_what\\_who](http://www.jucs.org/jucs_7_7/metadata_standards_what_who)>

Gilliland-Swetland, Anne et al.

*Introduction to Metadata: Pathways to Digital Information*. <<http://www.getty.edu/research/institute/standards/intrometadata/index.html>>

Lagoze, Carl.

The Warwick Framework: A Container Architecture for Diverse Sets of Metadata. *D-Lib Magazine*, (July, 1996).  
<<http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>>

Paepcke, Andreas, Chen-Chuan K. Chang, Terry Winograd, and Hector Garcia-Molina.

Interoperability for digital libraries worldwide. *Communications of the ACM*, 41(4):33-42, 1998. <<http://www.acm.org/pubs/citations/journals/cacm/1998-41-4/p33-paepcke/>>

Sutton, Stuart, and Jon Mason.

The Dublin Core and Metadata for Educational Resources. *Proceedings: International Conference on Dublin Core and Metadata Applications 2001*, pp 25-31. (2001) <<http://www.nii.ac.jp/dc2001/proceedings/product/paper-04.pdf>>

## VIII. Glossary

**Binding:** The association of a metadata assertion or statement with a particular syntactic encoding. A given metadata statement can be expressed in any of a variety of encodings. On the Web, these presently include HTML, XML, and RDF-XML, but other encodings or bindings may emerge over time.

**Cardinality:** Specification of how many times a metadata element can or must appear in a metadata description.

**Controlled vocabulary:** a formally maintained list of terms intended to provide values for metadata elements.

**Element:** a formally defined attribute or category of description in a metadata set. Often simply thought of in an attribute-value pair (element = "string-value"), but values may have additional structure (element = structured-value).

**Metadata architecture:** a coherent collection of enabling technologies, element sets, and standards of practice that collectively support the creation, management and exchange of interoperable metadata.

**Namespace:** a formally managed vocabulary with designated bounds.

**Namespace declaration:** a convention for declaring a namespace in XML syntax that includes the URI for the namespace and specifies a colon-delimited prefix token that is prepended to all terms from that namespace used within the scope of the declaration.

**Schema:** a formal grammar for a metadata element set expressed in a formal schema language (in the context of this paper, either a XML Schema or RDF Schema). Schemas may be simple (composed of elements drawn from a single namespace) or compound (composed of elements drawn from multiple namespaces).

**URI:** Uniform Resource Identifier: a globally unique identifier that identifies a Web resource (either a URL or a URN) constructed according to the HTTP namespace rules.

**Value set:** a controlled set of terms from which a value for a metadata element is selected.

Copyright 2002 Erik Duval, Wayne Hodges, Stuart Sutton, and Stuart L. Weibel

---

[Top](#) | [Contents](#)  
[Search](#) | [Author Index](#) | [Title Index](#) | [Back Issues](#)  
[Previous article](#) | [Next article](#)  
[Home](#) | [E-mail the Editor](#)

---

[D-Lib Magazine Access Terms and Conditions](#)

[DOI: 10.1045/april2002-weibel](#)