Metadata for Digital Libraries: a Research Agenda

EU-NSF Working Group on Metadata

15 March 1999

About This Paper

The EU-NSF Working Group on Metadata was one of five study groups funded jointly by the US National Science Foundation and the European Union (through ERCIM, the European Research Consortium for Informatics and Mathematics) on strategic issues in technology for digital libraries during 1997–98. The purpose of these groups was to identify digital library research areas that could profitably be addressed through international collaboration.

Introduction

"Metadata" is the Internet-age term for structured data about data. Typical examples are library catalog records, bibliographic headers in Web pages, "terms of use" statements, and ratings. Different user communities — from librarians and computer scientists to government agencies, cultural heritage organizations, publishers, businesses, and the legal community — scope and purpose metadata differently. International communities in areas such as biodiversity, the space sciences, and museums seek to refine the semantics of specialized metadata for the rapidly evolving needs of their fields. Likewise, publishers and other content providers are seeking agreements on standards to enable new forms of electronic commerce.

The creation and management of metadata is a sizeable and rapidly expanding industry. In the scholarly communities, libraries and abstracting and indexing services invest heavily in the creation of metadata to manage the published literature. Scholars are increasingly publishing their work in electronic journals or in less formal Web forums, and scientific datasets are proliferating. Metadata will be needed for these materials. As libraries and museums digitize cultural heritage information, they must create metadata to organize and manage it; indeed, the cost of creating this metadata is often comparable to the cost of digitization itself. In the consumer world, metadata in the form of ratings and reviews has long been important. Now it is becoming clear that good metadata is needed to allow consumers to find products for sale on the Web as well. As we look towards worldwide commerce in intellectual property over the Internet, metadata to support rights management will be an essential part of this new marketplace.

In all such contexts, metadata helps people find what they need, verify its authenticity, process it in an appropriate format, and perhaps to order or pay for it online. Some types of metadata will be used by humans; others will be processed automatically by new types of software tools and systems. No single type of metadata can suit every such application, every type of resource, and every community of users. How the diverse forms of metadata will co-exist and interoperate is a complex issue for research.

The EU-NSF Working Group on Metadata focused on the architectures, tools, and models needed for managing metadata in a distributed, networked environment and on aspects very broadly related to the problem of resource discovery. We did not cover specialized uses of metadata *within* information services, such as using thesauri to reformulate queries. Also outside our scope was metadata for knowledge representation in the broader sense of natural-language ontologies, classification schemes, domain-specific terminologies, and controlled vocabularies of element values.

1. Architectural Issues for Metadata

A number of high-level issues concern the management of metadata and the relation of metadata to the overall design of networked information systems.

1.1. Models for Metadata-Resource Association

There are several models for associating resources and metadata. For simple text documents on the Web, descriptive information is most commonly embedded in documents using the META tag of the Hypertext Markup Language (HTML). No additional infrastructure is needed for this other than the Web itself. The metadata can be created by authors and others who build networked resources, and it can be harvested automatically by indexing robots. Alternatively, metadata can be tightly coupled with a resource and transported together with that resource's contents in Web transactions using the Hypertext Transport Protocol (HTTP). This allows metadata and resources to be maintained separately, but at the cost of a somewhat more complex infrastructure. A third model finds metadata located physically apart from a resource in a separate database, managed in some cases by the owners of the information described, in other cases by third parties. Research is needed to understand the implications of these models for performance, cost, scalability, interoperability, implementation, maintenance, and systems design.

Trusted third-party resource description has long been provided by libraries, archives, and indexing services in a well-established metadata industry that predates the Internet, with well-established standards such as the MAchine-Readable Cataloging format (MARC). Since the rise of the Web, new entrepreneurial metadata services have emerged, including ratings bureaus, which may operate in conjunction with filtering systems in browsers, and full-text search engines. Search engines typically derive their metadata algorithmically, rather than intellectually, from a statistical and semantic analysis of full-text documents on the Web. This metadata is used only to resolve queries and is not made available directly to users.

As work advances on metadata in areas as diverse as resource discovery, electronic commerce, and rights management, it is becoming clear that interoperability will depend on a clearer understanding of data models for information objects. Efforts are underway to explore the relationships among the data models underlying the so-called Digital Object Identifier (DOI), the Dublin Core Element Set for resource description, and INDECS (INteroperability of Data in E-Commerce Systems), a metadata initiative for supporting global commerce in intellectual property. One common reference point for these various communities is the Functional Requirements for Bibliographic Records of the International Federation of Library Associations and Institutions (IFLA), which describes a range of possible "states" for information resources from the abstract *work* through the physical *item*. Defining a logical framework that subsumes or reconciles a variety of data models is a major research challenge with implications for the exchange and reuse of different types of metadata for a broad range of applications.

1.2. Metadata Creation and Management

Producing metadata has long been the task of professionals such as librarians and indexers. Today, metadata is also being produced by non-specialists (e.g., ordinary users), publishers, for-profit commercial agents, and even software systems. Creating metadata is often labour-intensive, though automated procedures and new types of tools for creating and managing metadata are evolving and will continue to become more sophisticated. Some word processors, editors, and format filters generate embedded metadata tags when a document is first created. Other tools automatically extract metadata values from the documents themselves.

Metadata may be generated or updated at various times in the life cycle of a resource, so workflows must be designed accordingly. Declarative rules or prescriptions can be used to explicitly describe how, when, and by whom a metadata element will be produced or updated. Research on metadata creation and management will enable systems to use document analysis to automate metadata capture, automatically extract subject classifications, trigger the update of metadata when a resource is modified, support the capture of temporal metadata, certify metadata at the time of publication, and remove metadata when a resource is obsolete. Such tools must become integrated into a variety of environments from Web site management utilities to databases, data warehouses, and legacy resources. If used on a wide scale, they could significantly improve the quality and cost-effectiveness of metadata in the networked environment.

1.3. Metadata for Repositories and Services

Proprietary digital resources will increasingly be stored in data repositories and may not be accessible to external Web harvesters. Publishers may want to use metadata to advertise their materials, which would then be available for a fee. Metadata will play a crucial role as a surrogate which makes the contents of such restricted resources and services visible to searchers. Some services, such as weather forecasts, may make their contents accessible only through dynamic interfaces that cannot be harvested. Metadata will not just serve as advertising for content; some kinds of metadata, such as reviews, will have markets in their own right.

Designing effective mechanisms for such repositories will enable large-scale commercial publishing on the Web. One repository model, the Kahn-Wilenski Digital Object Repository, supports formal specification of the services, policies, and transaction methods associated with a collection of resources. Research on repository architectures and their metadata will need to balance market forces and organizational needs to ensure that the technological options enable rather than constrain the expression of policy.

2. Simplicity, Complexity, and Interoperability

Current approaches to metadata, such as the Warwick Framework, assume that complex metadata needs are best met by a multiplicity of separate but functionally focused metadata schemes, relatively orthogonal and independently maintained by communities of expertise and practice, which can be mixed and matched as needed, rather than by just one comprehensive but monolithic set of elements. The metadata associated with a given object would be separated into a series of “packages” that would be marked as using these various metadata schemes.

The foundations for such modularity are already in place. Further research is needed now to scope and elaborate new schemes and to specify the interrelations among packages. There are also research issues involved in the management of multiple packages for an object under the Warwick Framework within various distributed computing environments, including questions of consistency, overlap, relationships among packages, and the linkage of packages to nested or complex objects.

2.1. Core Metadata Sets

Metadata schemes that meet fundamental functional needs of users in a wide range of fields and applications are called "core element sets." The most important work in this area to date has been in resource description, where the Dublin Core forms a central reference point of fifteen broad categories onto which more complex or specialized descriptive schemas can map. Such core element sets are needed now for functions complementary to description, such as structure and navigation, administration of digital objects, authentication, certification and provenance, terms and conditions, trust and quality, privacy, and longevity.

Particular communities will develop discrete (but possibly overlapping) packages of metadata for various purposes. For example, structural metadata for display and navigation is needed for specifying the range of file formats a library can deliver, the size of particular documents, or the parts of documents that are accessible individually, such as tables of contents and chapters. Administrative metadata can cover the creation of an object (such as date of capture or digitization technology used), identify an instantiation (version or edition), or specify the technology needed to view or use an object (e.g., storage or delivery file format, compression scheme, or location). Metadata on trust might include certifications or ratings of the quality of services or content offered by information providers. It is probably impossible, at this stage of development, to establish a comprehensive enumeration of core element sets; however, the need for a number of such sets is clear, and managing their evolution and scoping will be a challenge.

2.2. Metadata Pidgins and Complexification

In all forms of human communication there is a tension between the virtue of simplicity and the need for complexity. Historically, attempts to create simplified common languages such as Esperanto have often splintered between those who wanted to keep it simple to promote mass usage versus those who called for larger vocabularies and grammatical nuance. Pidgins — simple languages improvised by speakers of different languages who need to work together or conduct trade — also start with small vocabularies and simple grammars. When adopted by entire communities, however, pidgins typically evolve into more complex and expressive "creoles."

Such dynamics have been evident in efforts to define global standards in metadata. The effort to simplify and hybridize the resource description conventions of various communities has led to the pidgin-like (unqualified) Dublin Core Element Set. Such core schemas are extensible via additional elements or local refinements, but complexified adaptations can compromise or at least reduce interoperability. As users seek greater nuance of semantics and syntax, the natural follow-on to this pidginization is a re-complexification, or creolization, resulting in metadata that may be more precise and expressive, yet less compatible with the global standard.

This tension between simplicity (hence interoperability) and greater semantic richness is a natural feature of evolving metadata conventions. A search for balance has been evident in the evolution of the Dublin Core, particularly in the differences between qualified and unqualified approaches. Each new core metadata set will pose the same problem of balance as it is extended and specialized beyond its original core use.

2.3. Metadata for Complex Digital Resources

There are well-established conceptual models and practices for describing the contents and structure of traditional resources. For example, the consensus understanding of component parts of texts (titles, chapters, paragraphs) is so strong that it is often implied by layout and typography. This understanding can be formalized, along with higher-level structures, through constructs such as the Document Type Definitions (DTDs) of the Standard Generalized Markup Language (SGML) or word processor style templates. Other practices surrounding the use of text (e.g. citation and quotation) are also relatively well established, as are means of associating metadata with texts (e.g., independent MARC records, HTML META tags, and DTDs of the Text Encoding Initiative). A relatively limited number of encodings are in popular use for text documents — ASCII, word processor formats, PDF, HTML, SGML, and perhaps soon XML — and viewing software often allows users to convert among these formats.

For non-textual data the variety of encodings in use is much wider, and the problem of conversion is accordingly more complex. Many of these conversions will need to be controlled and automated by new types of metadata. Time-based media such as audio and video will need "mechanical" metadata for controlling processes such as synchronization (e.g., sampling rates and frame sizes), which will need to be linked to higher-level descriptive metadata through abstractions that are still very much at the research stage. For describing the intellectual contents and structure of resources that are more complex and less well understood than normal text documents — such as collections of documents, time-based media, and dynamically generated objects — there are few conventions in general use. A "Multimedia Content Description Interface" (MPEG-7) is currently under development within the Moving Pictures Experts Group (MPEG), a committee of the International Organisation for Standardisation that focuses on the encoding and processing of motion pictures, audio, and related multimedia formats. MPEG-7 is intended to include not just basic description, but metadata on terms and conditions, ratings, encoding formats, and "scenarios" of how multimedia components are combined in presentations.

More broadly, we have a relatively good understanding of traditional publication genres such as printed newspapers, magazines, journals, books, music CDs, and movie videos. Each of these genres is associated with a relatively clear, often long-standing analytical, intellectual, economic, and legal framework. However, we do not yet have such an understanding or consensus about evolving genres of digital publication. We may need to define new genres as the sum of multiple dimensions relevant for their management and handling, such as their encoding format, publication type, access controls, description, and terms and conditions of use. For all genres, including text, we need metadata conventions that support versioning.

Metadata is also needed for describing groups of resources as collections. Collections may be purely administrative entities, or they may be characterized by attributes such as medium, subject, or provenance. Collections may exist within other collections, such as a reference collection within a library. Distinctions between "whole" and "component" collections could facilitate navigation by helping users discover specific databases. Collection description could also provide at least some access where no item-level descriptions exist, such as in certain image databases and manuscript archives.

2.4. Metadata Diversity and Resource Discovery

While we have much experience with homogeneous descriptive metadata (i.e., the library catalog), very little is known about how to effectively repurpose and integrate the much broader range of metadata now being associated with digital objects into the discovery process. Moreover, the availability of various types of metadata will be highly variable in practice, and it is unclear how best to use types of metadata that may be only sparsely or spottily available. We face the challenge of breaking a vicious circle: without enough metadata of a given type, people will not build harvesters, but if that type of metadata is not widely used, there will be no motivation to create it in the first place. Solutions to this problem will advance metadata deployment, discovery systems, and networked services in general.

3. Infrastructure

Interoperability among applications benefits from common conventions on semantics and syntax. Semantic interoperability may be achieved either by agreeing on shared meanings or by using mediating software to create coherence across applications. Syntactic interoperability requires standard formats and protocols for expressing metadata structures.

3.1. Metadata and Information Architecture Standards

A number of standards that are emerging for the Internet and the Web will provide much of the basic architecture and context for digital libraries. For example, a working group of the World Wide Web Consortium (W3C) is developing the Resource Description Framework (RDF), a set of standards for supporting the exchange of metadata on the Web. This group has recently published the public specification of a general model for metadata. Its underlying encoding syntax is the eXtensible Markup Language (XML), which is expected to become a primary format for document encoding on the Web. RDF has been designed to support the encoding of metadata semantics in discrete functional packages that can be used by applications in the modular, plug-and-play manner called for by the Warwick Framework. The use of Dublin Core metadata in RDF will provide early deployment experience that should expose some of the strengths and weaknesses of this evolving infrastructure.

RDF uses metadata schemas to specify the structure and semantics of formal metadata standards in a way that is usable both by humans and by machines. Schemas will be extensible, either by referencing other schemas or by adding local refinements or specialized elements. Sharing a common set of elements across multiple metadata schemas will promote semantic interoperability by allowing the creation of new schemas that inherit and preserve (while perhaps specializing) the semantics of a parent schema. Languages, techniques, and systems for expressing such schemas will remain an important area for research. The refinement of a generalized metadata architecture based on RDF will require several iterations as it is deployed for digital libraries and in commercial applications.

Generalized formal models of metadata, such as RDF, provide a common language for defining the structure of the many types of metadata used in digital libraries. Within such overarching frameworks, more specific models may be needed to define specialized domains or facets of resources, such as intellectual property restrictions or geospatial coordinates. These formal models may exist in a hierarchy of description, with general models such as RDF pointing to more specific models. For example, the Federal Geographic Data Committee Content Standard might describe specialized aspects of geospatial data, while a more general model is used to associate that data with related non-geospatial resources. The combination of general and specific models is more practical and flexible than monolithic approaches and supports the integration of heterogeneous resources into coherent collections.

The metadata activity of W3C also includes work on standards for privacy and rating services. PICS (Platform for Internet Content Selection) and P3P (Platform for Privacy Preferences) are intended to provide a foundation for this work. Metadata for both will be expressible in RDF. These standards, however, have not yet been widely deployed. Research is needed both on technical and social aspects of their use and on their interaction with other metadata sets.

Other evolving standards involving substantial use of metadata will require continued refinement through deployment experience. In many cases, metadata-using standards have developed somewhat independently of the efforts of the metadata research community, and considerable work may be needed to understand how the conceptual models in these standards relate to work such as RDF and the Warwick Framework. Examples include the Common Object Request

Broker Architecture (CORBA) under development by the Object Management Group, which enables applications to manipulate distributed computing objects; the various proposals for electronic rights management and digital identifier systems (such as the DOI); and the Z39.50 protocol for information retrieval, which includes methods of providing metadata to characterize databases and servers. The standard ISO/IEC 11179 (Specification and Standardization of Data Elements) will need to address the varied requirements of metadata for digital libraries and electronic commerce. ISO's Basic Semantic Register will be relevant to the future standardization of metadata vocabularies and to the development of ontologies across schemas.

3.2. Crosswalks and Registries

A variety of constructs allow users to search and retrieve across different metadata schemas by translating the elements of one system or subject organization into the terms of a second. Mappings, whether based on tables or on formal definitions, represent relationships that are unambiguous; they support transparent searching across domains. Crosswalks are more complex frameworks that establish the relationship between schemas that have significantly different syntaxes or semantics. They can be based on thesauri or on more elaborate semantic frameworks, such as ontologies, which describe the metadata elements for each domain and map their internal relations. Crosswalk tools could use ontologies to help users understand semantic complexities across systems, but further work is needed on formalisms for expressing such linkages.

The metadata schemas available on the Web will form a global collection of namespaces that will effectively function as a distributed registry. These registries will need to be managed, coordinated, and ultimately connected. Registries will define the elements of metadata schemas in a machine-readable syntax (e.g., RDF) and offer authoritative listings of legal values, local extensions, mappings to other schemas, and guidelines for good usage. They will serve both humans, with readable text, and programs, with structured content that can automatically be parsed. Their role will be both to promote and to inform, thereby encouraging the use of standard formats.

Like good dictionaries, registries will need to describe actual usage while prescribing good practice. They will help providers of content in languages other than English both in keeping their translations of global standards up-to-date and in constructing specialized local schemas that are as compatible as possible with global standards. They will also assist in aligning element definitions among diverse existing and proposed schemas, thus reducing duplication of effort. At a more complex level, registries could offer knowledge bases that support the automatic translation of metadata into other languages and schemas, automatic validation, and the automatic updating of remote metadata to new versions.

The interaction between metadata registries and particular user communities will never be simply a technical problem. Like natural languages, metadata languages will evolve with use. Interoperability over time can be ensured only if registries are supported by a public forum with formal processes that allow user communities both to negotiate meanings collectively and to adapt these rules to local needs. Since the shared semantics of communities differ in complex ways, the definition and maintenance of crosswalks and registries imply support for complex social processes. Building such systems must draw on our understanding of how ontologies, whether of specialized vocabularies or of natural languages, can formally be related or linked among themselves. Constructions used in federations of thesauri, such as "interlinguas," could be adapted for linking ontologies of metadata.

Metadata element sets will be managed by agencies that are responsible for their development and maintenance. An ecology of registries will emerge that reflects a diversity of organizational motives, market forces, and user requirements. A distributed model for registries will reflect this diversity by supporting cooperation at multiple levels — global, regional, domain (by subject or resource type), and sector. The development of such infrastructures will provide an important focal point for international collaboration.

4. Policy and Management

The Web has substantially reduced barriers to information flow and access around the world and we now see increased interaction across sectors and disciplines on an international basis. Content and metadata are both moving rather casually between communities and across national boundaries. This radical change in accessibility has major implications for creating and managing both content and metadata.

4.1. Trans-Border Information Policy

Legislators and policy-makers are paying increased attention to the Internet environment and to harmonizing both intellectual property law and a wide range of information policies (e.g., privacy, content rating, commerce, and censorship) on a global basis. However, national differences in policies and legislation on such issues will inevitably become more prominent. The semantics of metadata, the operational characteristics of applications that use metadata, and the interaction of diverse forms of metadata will receive substantial scrutiny. Libraries, archives, museums, government agencies, and publishers — as well as their constituents — will need to take account of international implications as they establish metadata standards and infrastructures in support of their goals and interests.

Various classes of assertions embodied in metadata may well have legal status, at least in some usage contexts and legal regimes. We need to understand the implications of this within a complex diversity of national and international law and practice, especially when metadata crosses borders and jurisdictions. In some situations, it may be necessary to translate metadata values between languages to another or to map metadata packages from one legal regime to another. The assertions embodied in metadata and the sources of metadata will raise issues of liability and accountability; in order to cross borders, for example, some metadata may have to be digitally signed.

4.2. Metadata Research and the International Standards Process

Much of the value of metadata comes from its character as a shared community understanding. This means that progress in metadata rapidly moves into the arena of standardization. Work on metadata standards — and particularly the sort of cross-community standards discussed in much of this paper — is particularly challenging because it demands collaboration from historically autonomous and specialized communities of practice and their standards development organizations. This is creating significant difficulties as the various standards organizations struggle to define their roles and establish processes for coordination and collaboration with other groups.

It is becoming clear to many communities of practice that as their activities and information resources move into a highly distributed network-based environment they will need to evolve community metadata standards. Research is needed to develop faster, more effective and efficient ways for these communities to develop, achieve consensus upon, and document community practices, processes, and data models. Streamlining such efforts could have a high impact on a range of fields, from scientific disciplines to the cultural heritage community.

The development of standards for digital libraries and networked information is a very important process for establishing consensus about intellectual models and for translating research into practice. Standards development, particularly at the international level, is a slow and costly process that has in the recent past been dominated by commercial interests and government representatives with the resources to participate. It is essential that resources be made available to permit the metadata and digital library research communities to participate fully in these ongoing efforts.

4.3. Integrity, Accuracy, and Authenticity of Metadata

As metadata becomes more important both for discovering information and for expressing assertions about networked content, it will become increasingly necessary to verify its accuracy, integrity, and trustworthiness. One can already see situations in which content providers attempt to mislead Web indexing services with irrelevant index terms in order to make their content more visible to searchers. In the present environment, Web indexers can inspect the actual source content of Web pages to check for such subterfuge, and the major commercial services have developed proprietary algorithms that attempt to filter out indexing misinformation. As metadata use expands to capture more abstract assertions about objects (e.g., ratings or subject analysis), the design of such validation algorithms and heuristics becomes an extremely challenging and largely unexplored problem for research.

An alternative to verifying metadata is to consider the source of each metadata assertion associated with an object. This suggests research in at least two areas. The first is the design of systems to verify the source of metadata assertions, presumably in conjunction with digital signatures and a public key infrastructure, in a highly distributed international environment. The second is how to factor user-provided (and third-party-provided) constraints about the relative trustworthiness of metadata providers into processes such as resource discovery.

4.4. Evaluation and Metrics

The deployment of metadata through standard mechanisms such as RDF provides an opportunity to explore the evolving use of metadata within various communities. The sampling and analysis of metadata will provide a valuable window onto the Web as a medium for distributing the cultural, scientific, and technical products of a society and onto the practices used to describe them. Knowledge gained from tracking the diffusion of metadata use can be applied to the further development of metadata schemes and to the creation of dictionaries that track the transition from pidgin to creole. Usage patterns can serve as a guide for developers of schemas and feed back into the design of core metadata sets and functional packages.

A fundamental question about the use of metadata, particularly in areas such as resource discovery, is cost-effectiveness. Is intellectually derived descriptive metadata sufficiently better than algorithmically computed metadata to justify the cost of its creation? Benchmarks and experimental evaluation frameworks similar to the Text Retrieval Conference (TREC) for full-text databases will be needed. Tracking deployment will also help identify which data elements users consider to be important in processes such as resource discovery and show how these elements are being used. Such data will guide the further evolution of core metadata sets.

Conclusion

This paper identifies numerous topics for research on issues that will be central to fostering the growth of networked information resources, digital libraries, electronic commerce, and network-based publishing. In the near term, a commitment to work on registry infrastructures could provide an important focal point and source of cohesion for research and development in many of these areas, as well as being an important research project in its own right. The Dublin Core seems poised to provide a metadata system for resource discovery that is consistent across a wide range of applications and domains, usable by both experts and non-experts, interoperable with existing library catalogs and legacy databases, and coherent across many languages (over twenty to date). RDF is a deployment vehicle not only for Dublin Core but also for a wide range of new core and discipline-specific metadata sets that will be developed.

Effective research progress on metadata will need to involve intense collaboration between metadata specialists and communities trying to solve functional problems, such as rights management, resource discovery, archiving and preservation, or the organization and management of data specific to various disciplines. The definition and maintenance of metadata standards over time is a complex social process requiring negotiation, consensus-building, and iteration. Learning to manage such processes effectively and to coordinate the ever-growing activities of many disparate communities of interest is clearly a long-term research undertaking involving complex economic, technical, and social questions.

The world has never had information systems of this scale, nor has it ever tried to provide consistent organization and access to sets of resources and services as large and diverse as those appearing on the Internet today. We must seek not just to serve specific disciplines or communities of practice, nor must we focus just on facilitating the use of information across communities. Rather, we must seek to provide the emerging global information infrastructure with coherent methods of organization and access that transcend the historical boundaries of nations, languages, and cultures. Addressing these challenges will require a two-pronged research approach: engagement in near-term, applied questions involving specific metadata sets and architectural issues raised by developments such as RDF, and a continued investment in theoretical and foundational work that can abstract out broad principles and model the effects of scaling.

References

[1] Baker, Thomas, "Languages for Dublin Core," *D-Lib Magazine*, December 1998, http://www.dlib.org/dlib/december98/12baker.html.

[2] Bearman, David, Eric Miller, Godfrey Rust, Jennifer Trant, Stuart Weibel, "A Common Model to Support Interoperable Metadata," *D-Lib Magazine*, January 1999,
http://www.dlib.org/dlib/january99/bearman/01bearman.html.

[3] Berkeley National Laboratory (U.S. Department of Energy), "Joint Workshop on Metadata Registries," [workshop report], July 1997,
http://www.lbl.gov/ olken/EPA/Workshop/report.html.

[4] DOI, The Digital Object Identifier System, [home page], http://www.doi.org/.

[5] Dublin Core Metadata Initiative, [home page], http://purl.org/DC/.

[6] ERCIM News Online Edition, "Advanced Databases and Metadata," [special theme issue], October 1998, http://www.ercim.org/publication/Ercim_News/enw35/en35contents.html.

[7] INDECS (INteroperability of Data in E-Commerce Systems), [home page], http://www.indecs.org/index.htm.

[8] IFLA (International Federation of Library Associations and Institutions), "Digital Libraries: Metadata Issues," [evolving directory of resources], http://www.ifla.org/ifla/II/metadata.htm.

[9] IFLA Study Group on the Functional Requirements for Bibliographic Records, "Functional Requirements for Bibliographic Records," UBCIM Publications, New Series Vol. 19, 1998, http://www.ifla.org/VII/s13/frbr/frbr.pdf.

[10] Kahn, Robert, Robert Wilensky, "A Framework for Distributed Object Services," Corporation for National Research Initiatives, May 1995, http://www.cnri.reston.va.us/home/cstr/arch/k-w.html.

[11] Lagoze, Carl, Clifford A. Lynch, Ron Daniel, Jr. "The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata," Cornell Computer Science Technical Report TR96-1593, June 1996, http://cs-tr.cs.cornell.edu:80/Dienst/UI/1.0/Display/ncstrl.cornell/TR96-1593.

[12] Lynch, Clifford, "Identifiers and Their Role in Networked Information Applications," *ARL* 194, October 1997, http://www.arl.org/newsltr/194/identifier.html.

[13] Lynch, Clifford, Avra Michelson, Cecilia Preston, Craig A. Summerhill, "CNI White Paper on Networked Information Discovery and Retrieval," Coalition for Networked Information, http://www.cni.org/projects/nidr/www/toc.html.

[14] MPEG (Moving Picture Experts Group), [home page], National Institute of Standards and Technology, http://www.cselt.it/mpeg/.

[15] TREC (Text REtrieval Conference), [home page], http://trec.nist.gov/trec.

[16] World Wide Web Consortium, "Metadata Activity Statement," http://www.w3.org/Metadata/Activity.html.

EU-NSF Working Group on Metadata

- Clifford A. Lynch, CNI, Washington DC, co-leader
- Thomas Baker, GMD, Germany and Asian Institute of Technology, Thailand, co-leader
- Gene Alloway, University of Michigan, USA
- Howard Besser, University of California, Berkeley, USA
- José Borbinha, INESC, Portugal and National Library of Portugal
- Rachel Heery, UK Office for Library and Information Networking (UKOLN), UK
- Ole Husby, BIBSYS, Norway
- Renato Iannella, Distributed Systems Technology Centre, Australia
- Shigeo Sugimoto, University of Library and Information Science, Japan
- Anne-Marie Vercoustre, INRIA, France
- Stuart Weibel, Online Computer Library Center, USA