Writing a summary of the library literature on metadata from 1995 to the present is a little like trying to drink from a fire hose. Library interest in this field has exploded in the past five years, leading to a tremendous amount of published material, some in print journals but largely on the Web. It is difficult to assimilate the variety of information that is out there, let alone to transform it into a useful summary. Yet in the library domain it is still a field in the very early stages of its development, as much of the publishing that has been done is of an introductory nature. Authors more often than not feel the need to begin their work with their own brand of a definition of metadata, indicating that the concept is not yet established enough within the library profession that we can say that we all know what we mean, like we do when we say “authority control” or “classification.” For 27 different definitions of metadata, see appendix 2 of the “Summary Report, June 1999,” by the Task Force on Metadata (ALA 1999).

I must note that the concept of metadata has been in use in computer science circles for decades, and that the (comparatively) recent emphasis on it in the library world indicates that librarians are collaborating with computer scientists to solve the problems of information discovery and retrieval in the digital age. The divide between librarians and computer scientists may not be as wide as Kathleen Burnett, Kwong Bor Ng and Soyeon Park (1999) seem to indicate in “A Comparison of the Two Traditions of Metadata Development.” According to the authors, librarians come from the bibliographic-control tradition, focusing on the description of individual objects with an emphasis on discovery by users, and computer scientists come from the data management tradition, encompassing the concerns of the bibliographic control tradition plus issues of security, data sharing, and data integrity. The present summary, however, in deference to the economy of space and to the limitations of its author, concentrates on library literature, and regrettably leaves even a lot of that out.

Various communities (e.g. the archival community, the museum community, the education community, the geographic and spatial data community, etc.) have been working to develop a plethora of metadata systems to enable the organization, searching, and use of their particular data. Comprehensive lists of metadata systems and metadata projects are easily found on the Web. International Federation of Library Associations and Institutions (IFLA) maintains an impressive list; the Committee on Institutional Cooperation (CIC) site “Links to Metadata Web Pages” (itself a useful list) describes the IFLA site as “the Metadata Gateway [which] should be the first stop for any search for information on digital libraries and metadata resources.” Also useful is the site maintained by UKOLN, “Metadata Resources.”

These long lists can be rather overwhelming, however, unless one has had an introduction to metadata and its uses. One of the most accessible introductions from the library perspective is Jessica Milstead and Susan Feldman’s “Metadata: Cataloging by Any Other Name…” (1999), which defines metadata and identifies the need for it in terms catalogers will understand: “If all documents carry the same fields, and also use the same controlled vocabularies, then we should be able to improve searching.” Milstead and Feldman’s companion article “Metadata Projects and Standards” (1999b) identifies some of the players on the metadata scene (International Organization for Standardization, World Wide Web Consortium, etc.) and gives short descriptions of a few dozen of the most famous metadata projects. Milstead and Feldman also refer to Dempsey, Russell, and Heery of the DESIRE project’s “A Review of Metadata: A Survey of Current Resource Description Formats” (1997), which classifies metadata projects into three groups: Band One, or simple formats employing mainly unstructured data automatically extracted and indexed, such as Yahoo, Lycos, etc.; Band Two, or structured formats supporting user discovery of resources and some human intervention in the task of indexing, such as Dublin Core; and Band Three, or richly structured formats, which require elaborate tagging, such as MARC or TEI. Judith Ahronheim’s article “Descriptive Metadata: Emerging Standards” (1998) is helpful to those trying to sort out the alphabet soup of metadata-speak. She gives short explanations of SGML, RDF, XML, TEI, FGDC and others, and provides a useful “Basic Resource List”, that includes the Web sites of the above-mentioned systems.
It is no accident that writing on metadata issues increased exponentially in 1995. That year saw the beginning of the Dublin Core (DC) Metadata Initiative, which over the past five years has emerged as the metadata system of broadest interest to libraries, as well as the one in which librarians have had the most influence in creating (aside from MARC, of course). The DC workshop reports available, from March 1995, April 1996, September 1996, March 1997, October 1997, and November 1998 (the seventh and latest DC workshop, in October 1999, did not have a report as of this writing) are the best sources for the theoretical discussions that have contributed to the DC. The system has developed from a 13-element set for describing primarily textual DLOs (document-like objects) to a 15-element set adaptable to images and other materials. At its heart, however, it has remained an attempt to enable users to discover huge numbers of worthwhile Internet materials, by using simple (in comparison to MARC format) embedded metadata creatable by authors and non-catalogers. No actual research exists as of yet measuring the cost savings resulting from using DC instead of MARC, or documenting how using DC vs. MARC affects retrieval, both of which issues are commended to future research by Clifford Lynch (1998). Another area for future research is into whether Web authors know about DC and use it and, if so, how well they apply it.

The DC has been used in many metadata projects around the world; an extensive list of these projects can be found on the DC Web site. One of the more famous projects is the Nordic Metadata Project, which involved an evaluation of different metadata formats and the selection of the DC, and the development of a "toolbox": a means of conversion from DC to MARC and vice versa, a DC syntax, user environment and user interaction, and a DC metadata-aware search service (Hakala et al. 1998). Another important project is CORC, an OCLC-sponsored project to create a database of both DC and MARC records for selected Internet resources. CORC stores records in XML and delivers them to users in their desired format (Medeiros 1999). As the initial phases of many projects draw to a close, we can expect more case studies and project reports to be published.

Harold Thiele's article "The Dublin Core and the Warwick Framework: A Review of the Literature, March 1995–September 1997" (1998) does a superb job of reviewing and categorizing DC literature during that time period. Thiele proposes several areas for future DC research, including user studies comparing the effectiveness of DC with other metadata systems in satisfying searcher needs and in improving precision ratios in retrieval. The effect of using DC on improving cache performance in the search process and reducing bandwidth problems is another area of research proposed by Thiele. Also needing investigation is whether DC favors centralized indexing search engines like AltaVista over non-centralized indexing engines like Harvest, and whether including DC metadata in some Web pages will separate Internet resources into the "academic" group, using metadata, and the "non-academic," not using metadata. Thiele's conclusion that "[m]ost of the literature up to this point has been of a descriptive nature" still largely holds true, save perhaps for the following debate.

At the fourth DC workshop, a debate emerged in the DC community between the "minimalists" and the "structuralists" (Weibel, Iannella, and Cathro 1997). The minimalists believe the primary value of the DC is its simplicity and suitability for authors and others untrained in cataloging, and so argue against extensions to the 15 elements. The structuralists believe richer metadata is necessary for it to be useful and support a variety of qualifiers (called the Canberra Qualifiers, in honor of the host city to the fourth workshop) to be added to the original 15 elements. Roger Clarke (1997) argues forcefully for the structuralist position in what he terms "a reaction against what [he] perceives as the dangerous simplicity of the Dublin Core." Clarke is particularly concerned with the paucity of rights-management information in DC records, which he feels will be problematic as more and more Web sites charge for access. It would be interesting for future research to track whether this prognostication is indeed true. The theme is picked up by Godfrey Rust in his article "Metadata: The Right Approach" (1998), where he alleges that "rights metadata will have to rewrite half of the Dublin Core or else ignore it entirely." He declares further: "Dublin Core has seemed to be the only metadata game in town, and that is precisely why it is dangerous."

But Rust exaggerates, of course, as DC is by no means the only metadata game in town. Marcia Lei Zeng (1999) describes a project to provide metadata for digital images of clothing for a historical fashion collection. While DC was one of the candidates evaluated by the project team, it was deemed still too text-centric and not useful in describing the visual information they needed to describe. The project ended
up using a modified VRA Core system with some additional elements. VRA Core was chosen particularly for its understanding of subject elements as capturing more “ofness” rather than “aboutness” as subjects are intended to do in DC (and in MARC, for that matter.) Bipin C. Desai (1997) argues for use of the Semantic Header rather than DC or other schemes because it provides more information so that users can find out more about a resource before they actually access it, which will be increasingly important to users if the Internet ceases to be largely free. Stuart Sutton (1999) describes the creation of GEM, metadata for instructional materials on the Internet, to be accessed by teachers, parents, and students. GEM begins with the 15 elements of DC, then adds 8 elements, derived from a study of the search habits of educators. In addition, a Java metadata-creation software application (GEMcat) was created to hide the metadata framework from the user, so the metadata creator has only to fill in the content of the surrogates. All of these articles are mainly descriptive, and follow-up research is necessary for these and the hundreds of other projects like them to assess their effectiveness for the user.

With the large variety of metadata systems being developed by specialists in various fields, the question of how to translate metadata from one system to another must arise. "Mapping" or the development of “crosswalks” is necessary in order to make metadata available to a wider audience, limit duplicate creation of metadata in various formats, and present the user with seamless access to a wide variety of resources. The National Information Standards Organization (NISO) White Paper “Issues in Crosswalking” (St Pierre and LaPlante 1998) outlines several steps to greater interoperability among metadata systems: harmonization, or the ensuring of consistency in terminology between metadata specifications; element to element mapping, in which slight differences in semantic interpretations of elements must be accounted for; and the development of a fully specified crosswalk, consisting of both a semantic mapping and a metadata conversion specification. The authors conclude by suggesting the development of a “metadata specification language” and a standard method for crosswalk writing. One crosswalk that has been developed is the “Dublin Core/MARC/GILS Crosswalk” by Library of Congress (LC) (1999); progress toward an FGDC/MARC21/DC crosswalk is also underway (Chandler, Foley, and Hafez 2000).

RDF functions as a sort of “super-crosswalk” in that it allows different metadata systems to work together to describe an object. As an XML application, RDF uses the namespace feature to define the metadata system from which each descriptive element is taken (as in: <DC.Creator>John Smith</DC.Creator>) so that the semantics of each element are perfectly clear (Miller 1998). Its first official specification published just in March 2000, RDF was developed largely out of PICS, another W3C project, which was originally conceived as a method whereby parents and educators could use a selective rating system to block certain Internet sites from access by children. PICS can also be used for any selection/de-selection purposes (Resnick and J. Miller 1996). Needless to say, PICS has sparked some controversy among those concerned about censorship. See, for example, Simson Garfinkel's article “Good Clean PICS” (1997).

There are those who believe that MARC format is still the best metadata system to describe Internet resources. Amanda Xu (1997) argues that metadata in other formats should be converted into MARC format in order to be integrated into a library's OPAC. This is important in providing the user with seamless access to a variety of resources which may use different metadata systems. Z39.50 gateways offer the possibility of seamless searching, but only a few metadata systems have Z39.50 profiles. For other systems, Xu proposes a scheme for converting data into MARC format. Others express the concern that other metadata systems do not provide complete enough information for effective searching. Vianne Sha (albeit in 1995, near the beginning of libraries' involvement in metadata) cites the established mechanisms for sharing MARC records, the limited capacity of OPACs to handle anything but MARC records, and the role of libraries in ensuring that the public will be able to search and access Internet resources through their catalogs, whether or not they can afford Internet access, as arguments in favor of using MARC format. Michael Gorman (1999) suggests that there should be different “levels” of cataloging, depending on the worth and permanence of the resource. Higher levels of worth would indicate MARC format is necessary; lower levels may be serviced by DC or simply by leaving them to the mercies of search engines. While she does not argue for MARC format only, Sherry Vellucci (2000) points out that though some metadata systems have the capacity to support authority control, most, except for TEI and EAD, do not even officially recommend the use of controlled access points and vocabularies.
Whether authority control is applied depends on the directors of each particular project, so the quality of the metadata is highly variable.

As can be seen from this short summary, research in the field of metadata systems in libraries has just begun, and possibilities for future research are vast. An important goal facing all members of the metadata community, according to Jennifer Younger (1997), is the creation of “a data registry delineating each [metadata] scheme and identifying common and unique elements between and among them” (483–484). Not only would this foster an awareness of existing systems, thus encouraging metadata creators to use an existing system and so increase the standardization of existing systems, it would also support conversions from one system to another. The DESIRE project has begun such a project with its recent development of the DESIRE Metadata Registry Framework (Heery et al. 2000).

**Acronyms Expanded**

CORC: Cooperative Online Resource Catalog

DC: Dublin Core

EAD: Encoded Archival Description

FGDC: Federal Geographic Data Committee

GEM: Gateway to Educational Materials

GILS: Government Information Locator Service

MARC: Machine Readable Cataloging

OPAC: Online Public Access Catalog

PICS: Platform for Internet Content Selection

RDF: Resource Description Framework

SGML: Standard Generalized Markup Language

TEI: Text Encoding Initiative

VRA: Visual Resources Association

XML: Extensible Markup Language

**Works Cited**


Prepared by Patricia M. Dragon, Special Projects and Collections, Monograph Cataloging Division, University of Michigan Library, [pdragon@umich.edu](mailto:pdragon@umich.edu).