Digital Document Metadata in Organizations: Roles, Analytical Approaches, and Future Research Directions

Lisa D. Murphy • Indiana University • School of Business • Bloomington IN • limurph@indiana.edu

Abstract

Metadata characterize documents for discovery and use. In digital environments, conventional metadata are less likely to exist, and if they do, are unlikely to have been created by a trained cataloger. Within organizations, document metadata play roles beyond that of document discovery, for example, organizational memory, visibility, and network management.

An author-created metadata structure, the Dublin Core, maps incompletely to identified organizational metadata needs, particularly "publisher". Analytical approaches to metadata identify factors such as terseness vs. richness and logical and physical separability vs. integration from the object described.

In organizations, document discovery is often the start of a process of use and re-use of both formal and informal document types. Additional research is need to establish elements that would be important in organizations, the context and roles of metadata, and to explore approaches to metadata creation and use.

Introduction

Today, an increasing amount of information of importance to organizations is created and maintained as both formal and informal digital documents ([15]; [19]; [25]). Examples of documents created and used inside organizations include: policies and procedures, strategy papers, marketing reports, product specifications, customer service guides, and electronic maps. Researchers are currently investigating documents from several perspectives, including: genre ([16]; [31]; [32]), ability to transmit hard vs. soft information [22], social interactions [4], supporting tools [19], and searching (e.g., [10]; [13]; [20]). Much of this research has been under the auspices of digital libraries (e.g., [1]; [4]; [9]; [12]; [18]; [24]) while other investigations have looked at organizational issues (e.g., [15]; [19]; [16]; [22]; [25]; [32]).

For organizations to gain value from digital documents, potential users must be able to discover them, to assess their value for a particular information need, and to access and use them [15]. However, few organizations such as business corporations employ professional catalogers to create the descriptive information (also known as metadata, document descriptors, meta information, bibliographic records, and resource descriptors) that has typically been provided in libraries for searching and discovery. Similar to interactions with external databases, document users in organizations are believed to minimize their searching effort relative to their expected payoff and to make judgments about the likelihood that a document will be useful for their need (cf. [11]). Lacking a professional cataloger to create descriptors to aid in their evaluation, what do document users in organizations rely on?

Metadata for Digital Documents

The creation and provision of metadata have been core services of traditional libraries and information scientists. Smith [23] describes metadata as "the characterization of information objects for the purposes of locating, evaluating, and accessing appropriate sets of objects." These metadata have been "used principally in aiding users to access information objects of interest" ([24].

To many of us, the most familiar metadata are in a library card catalog. Today in many libraries, the collections remain significantly physical (i.e., books, journals, videos, etc.) while the card catalogs are likely to be digital. These metadata are created and stored separately from the information product. In digital libraries, both the metadata and the document are digital, allowing for alternative implementations. For example, metadata can be separate from the digital document in a traditional attribute-value listing in a searchable index or embedded within the document as hypertext mark-up codes readable by a specialized application such as a web browser or a "softbot" ([10]; [13]).

When metadata are conceived of as descriptors which may be internal or external to the information object, thousands of distinct metadata structures or systems can be identified. For example, metadata structures exist for music, images, historical texts, archeological records, medical records, maps, and software, as well as the traditional journals and books. Further, collections may implement the same metadata structure slightly differently to conform to their specific needs [9]. Many types of metadata require extensive training and specialized knowledge of the topic area [28].

As capabilities to access digital texts emerged, the potential for full text searching to substitute for document descriptors was recognized (e.g., [21]). However, full text

searching has limits for user-directed document discovery (e.g., [5]) such as finding spurious matches and the resulting information overload, inability to query on non-text elements such as illustrations and tables, and inability to query on contextual or supplemental information not encoded in the text. Thus, full text searching has not supplanted the role of metadata as document surrogates to support discovery [23].

In a digital world, metadata remain important. Dempsey suggests that "in an infinitely large resource space, effective management of networked information will increasingly rely on effective management of metadata" ([9], p. 2). As the number of documents in digital form increases, metadata become more important for their effective use. However, high value-added metadata are expensive to create and require resources that many document repositories do not have, that is, a trained cataloger. Alternatives to expert metadata creation include the World Wide Web (WWW) search robots that "harvest" metadata from the text of digital documents (e.g., [10]; [13]) and user-created metadata encoded in documents as HTML "meta" tags (e.g., [29]).

Metadata in Organizations

Metadata as "data about data" are most recognizable in today's organizations in the database element descriptors of database management systems and in system development and information management data dictionaries (e.g., [6]). Yet, metadata in organizations exist beyond that used to describe such highly structured forms of information. For example, they are captured as "properties" for word processed documents, exist in the header records of e-mail messages and directories of re-usable software objects, and function as indexes for digital image management and manual record retention systems. However, metadata in organizations generally lack the centralized or controlled aspect that both librarian-created and web crawler-created metadata have. Without an agent (living or otherwise) to create or find and then assemble the metadata, the usefulness of organizational metadata is likely less than optimal for the objectives of information sharing, organizational learning, or knowledge management.

Intranets may bring web crawler types of terse metadata into organizations, and with them, their noise and lack of quality ([9]; [10]; [13]). Natural language processing and information filtering advances are expected to improve the outputs of web indexing systems [10], but the lack of metadata not derivable from content remains a limitation.

Considering the case of organizational members creating documents to be used principally by other organization members, do we expect different information to be needed? Paepcke [19] offers insight into document creation and use inside a corporation from an extensive interview and observation study. Document creators were rabid plagiarists who repackaged existing material with new into what Paepcke calls "document compounds". The same material was often revised and re-presented to different audience, in a different format, for a different purpose.

Paepcke [19] found that identification of documents and interpreting their content were important in an environment where information discovery is the beginning of a process of re-use and recombination. Factors identified included who it was that captured or created it, their purpose in doing so, and other context information. Access to such metadata could facilitate the transmission of such information to the "next generation" of users in organizations.

Within organizations, Paepcke's [19] study provides a lone source for empirical results on the relationship between document discovery and document creation. He observes that models of information use from library environments are inadequate for within-organizational purposes [17]. While there is a shortage of empirical research related to organizations, Paepcke's results suggest that metadata containing information about the context of their creation and expected use available to prospective users would add value. However, it seems unlikely that organizations will hire specialized information professionals to create metadata for internal information resources (especially the informal and work-in-progress documents which are in majority) without evidence that the lack of it is a detriment to achieving organizational goals. While the exact nature of the organizational-specific elements is not yet known, some items to be considered include: department name or organizational affiliation; purpose of the document; work-in-progress status; approvals or authority levels of the authors; security or usage limits; expected date for final version.

Roles of Metadata in Organizations

Digital metadata has existed in organizations for many years playing the following major roles:

- supporting information search and retrieval (e.g., querying a database or retrieving an archived record);
- defining relationships between information elements and organizational entities (e.g., in database design or systems integration);
- serving as evidence (e.g., for legal or audit purposes) [3].

While data warehousing has increased attention to data-oriented metadata for the integration of structured information from diverse sources, this use does not introduce a new role beyond these three. On the other hand, the proliferation of both formal and informal genres of digital documents has increased the variety of business functions and users that may interact with metadata, and may therefore have introduced new roles. Users may rely on metadata for document interpretation and re-use (e.g., providing cues needed for accessing the content to create new documents), to describe costs, rights, responsibilities, and limitations of use (e.g., confidentiality, proprietary only, or regulation-required distribution), and as an element of organizational memory (e.g., reconstructing

organizational history from metadata without using document content). Document creators may use metadata to persuade (e.g., to promote a document as suitable for a certain purpose), as an aid to their own memory, or to increase visibility (e.g., to increase the likelihood of discovery by a search engine) while administration may use it to manage networks and allocate resources.

Table 1. Roles for document metadata in organizations.

Traditional Roles	New Roles
Discovery	Interpretation for re-use
Evidence	Individual memory aid
Use limitations and cost	Organizational memory
Retention and destruction	Persuasion or visibility
Managing resources	To describe accessibility

These new roles (shown in Table 1) reflect that digital document metadata in organizations are less formal, technical, and invisible, and more idiosyncratic, public, and contextual. In fact, access to metadata is considered crucial to support organizational information sharing through digital documents [15]. Table 2 compares some possible attributes of data-oriented and document-oriented metadata. While speculative at this time, the contents of Table 1 and Table 2 suggest that metadata will be increasingly author-created, responsible for conveying context, and pluralistic.

The Dublin Core Metadata Structure

Much of the development in digital metadata comes out of digital library projects in which access to networked resources through common interfaces (e.g., web browsers) is contemplated for a collection of digital materials, often of multiple formats (e.g., [2]; [9]; [23]). Recognition of the amount of information available through resources that do not have professional information provider involvement (i.e., on the Internet) has led to efforts such as the OCLC/NCSA Metadata workshop [30] and its products, the Dublin Core [27] and the Warwick Framework [8].

Table 2. Comparison of data-oriented and document-oriented metadata in organizations.

Attribute	Data-oriented	Document-oriented
Responsibility for creation	Technical professionals Third parties (other than content creators/users)	Author/creator Amateurs without professional or technical training
Meaning conveyed	Technical context Content Certainty	Social context Interpretation Equivocality
Form	Predefined Controlled Standardized Syntax-driven	Idiosyncratic Pluralistic, uncontrolled Localized Genre-driven
Motivation	"Push" – expressly defined, required, needed for technical performance	"Pull" – optional, available if created, visibility for author

As a metadata structure, the Dublin Core deserves attention because it is intended to be created by authors themselves without the aid of a cataloger, supports more than text-based information resources, and can be implemented in a variety of syntax. The fifteen elements of the Dublin Core (version 2) are described in Table 3. The elements are optional, repeatable, and expected to be modified and extended for local needs. [Author's note: Due to their large size, Tables 3 and 4 are at the rear following the references.]

Table 4 shows potential intra-organizational equivalents to the Dublin Core elements. The mappings between some elements seem quite straightforward (e.g., description, format, language, resource type, resource identifier, subject, source, title). However organizational equivalents for other elements show greater variation, such as: coverage, date, other contributors, publisher, and relation.

A limitation of the Dublin Core emerges when intra-organizational document genres diverge from conventional author-title-publisher model. Table 5 shows some potential additional metadata elements that may be appropriate in organizational environments but do not directly map to the Dublin Core elements. While not a comprehensive listing, Table 5 highlights that in organizations, the context of the document originator and its intended use may be as important as or more so than author name and title.

Table 5. Potential additional organizational metadata elements.

ELEMENT	EXAMPLES
AFFILIATION	Project, program, budget, sponsorship, department name, committee name
AUTHORIZATION	Policy, procedure, regulation, driver, initiative, approved by, directive

CONTACT	Mail Stop, phone number, email, address
EFFECTVITY	Product models, departments, job classifications affected
SECURITY LEVEL	Confidentiality, access level, destruction instructions, military security rating
STATUS	Work in progress, final, released, expired

A large gap between convention metadata and organizational metadata lies in the concept of "publisher." Certainly organizations use much material that is published in the sense of being available for public distribution and subject to relevant laws in that regard. However, for many organizations, the volume of internal use documents dwarfs the amount from outside sources. The information that the organization itself is the publisher, has limited "news" value for potential document users. Instead, we see a host of alternatives to the implied role: sponsor, approved by, authorizing party, and contact information. In a related vein, the state of being published appears to have several manifestations inside organizations. Organizational members (and some unscrupulous outsiders) may be most interested in works-in-progress, a state conventionally considered non- published. In other cases, a document may not be usable until it reaches a certain stage of authorization, regardless of the "publicness" of access to it.

While not exempt from issues of copyright and intellectual property, digital documents created and used internally may exist under a wide variety of controls as to who can use them and how. Beyond basic modes of company proprietary and confidential, internal documents range from restricted to attorney/client interaction (e.g., an environmental audit) to being required to be posted in a public location (e.g., annual OSHA injury and illness summary). Additional variations from retention and destruction policies.

Other approaches to metadata have come through the Internet as services such as Lycos, Alta Vista, and Yahoo, which offer users the ability to search massive indexes of electronic documents. However, these mechanisms provide both less and more information than desired, are confusing to novice users, categorize information differently than people do, and have difficulty dealing with certain types of documents (e.g., [10]; [13]; [17]; [20]). Such tools are now available for use on Intranets, however, the above problems remain, while their effectiveness in supporting non-HTML formatted information resources is uncertain.

Two Analytical Frameworks for Metadata

Dempsey [9] offers a useful analytical framework for considering metadata roles and characteristics for digital libraries (shown in Figure 1). Based on a continuum between terseness and richness, Dempsey describes three examples of metadata structures which he characterizes as performing roles from location through discovery to documentation. The three examples used suggest that the relative richness is related to variation in four areas: 1) amount of manual effort in creation, 2) amount of specialization, 3) level of description, and 4) external sources for metadata content.

This analytical framework helps explain the differences in usefulness and coverage we see in different document descriptor approaches. Not only are different data being captured, the cost of that capture tends to limit the potential coverage. Consequently, terse metadata such as that produced by web crawlers have a small range of information (i.e., that extractable from content) on a very large number of networked information products, whereas rich metadata structures such as specialized systems have large amount of data on fewer information products. Automated capture (e.g., Lycos) is less expensive than manual capture but precludes access to supplementary knowledge and information not explicitly contained within the document object itself.

Logical vs. Physical Separability

A reliance on the traditional library as analogue for the information environments of organizations has its limitations [17], including where and how much metadata exist. The library card catalog, whether automated or not, represents both logical and physical separation of metadata from the information product it represents. Library card catalogs typically include data that do not exist in the document (e.g., author's birth date, date entered into a collection, existence of other copies). Because it cannot be extracted from the resources, this information is logically separate from the document object itself. Digital card catalogs or abstract listings for non-digital information resources represent physical separation of the metadata from the object it describes. In contrast, digital metadata may be physically integrated due to being embedded in the document such as HTML meta tags or a word processor-generated summary file.

Logically integrated metadata include document components that are physically included within the document (e.g., table of contents, abstract) which, when reproduced elsewhere, demonstrate physical separability.

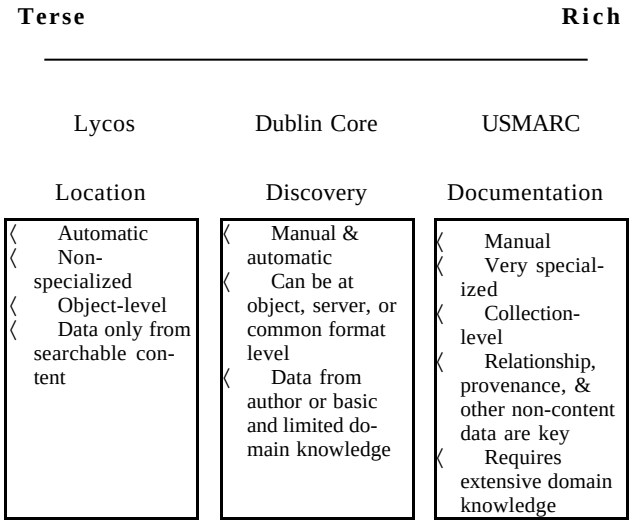

Figure 1. An analytical framework for metadata based on richness vs. terseness. (Adapted from [9].)

Formal definition of these two concepts are offered to promote the discussion about organizational metadata. To be logically separable, a metadata structure contains or can contain information not determinable from direct access to the full content of the document. To be physically separable, a metadata structure must be able to persist when the object it represents no longer exists. While in practice these dimensions may be more of a continuum, each is shown in Figure 2 as discrete and independent in a two by two matrix with examples.

In an organizational context, metadata would be logically separable when it provides context information that cannot be determined from the document itself. Examples include the purpose of the document (e.g., such as justifying a previous decision) or constraints placed on the authors (e.g., extreme time pressure or legal sanctions). Another type of metadata that is likely to be logically separate is related to disposition of the document (e.g., superceded, errata, approval status) and actions taken on the document after it leaves the author's control (e.g., restrictions on its distribution, translation to another language, or formatting changes).

Physical separability may be visible to the document creators and users (or may be transparent) and may exist on several levels. Users may not be aware that they are searching a separate index of metadata rather than actively searching document content. For example, consider the Internet search engines such as Lycos or Yahoo. Because it is an entity (an index) distinct from the original documents (HTML pages), Lycos itself has physical separation. However, what the users see may vary. While bookmarking the URL for a Lycos query allows it to be re-run in the future, the metadata so created is not identical to the original due to the continual update cycle of Lycos. Saving the result of the Lycos query to a file would result in metadata persisting beyond the life of the original document and thus would met the persistence criterion for physically separateness.

Metadata may include units of the document such as abstracts, tables of contents, footnotes, glossaries, or bibliographies when these are accessible as separable components. For example, the practice of posting the conference paper abstracts on a WWW site and publishing the entire papers on CD-ROM would allow such a web site to exist after the last CD-ROM became unavailable.

In organizations, the significance of separable metadata may vary with the situation. For example, storage limits or lack of network capacity may promote the use of metadata that is physically separated from the document repositories although this separation may not be visible to the user. Alternatively, the nature of the information products such as propensity for versioning, audit needs and legal status, and size or complexity, may promote a logical separation of metadata. Physical separation, in particular, implies the potential for a time lag in discovering that an information product exists or has been changed and the creation of the reference to it. In organizations, the persistence of metadata beyond the life of the document may be desired in some cases (e.g., organizational memory) and to be avoided in others (e.g., due to legal concerns). Logical separation implies that value may be enhanced by seeing the author and the information product in a larger context of organizational use and reuse.

Organizational Metadata Research Needs

The absence of empirical results about metadata existence, creation, and use in organizations lends a speculative air to this topic. Currently, visible metadata and document discovery efforts are underway in the digital library and WWW community, where the need for author-created metadata has been recognized (e.g., [30). However, no similar efforts were identified for organizational metadata. Nor have theoretical approaches to metadata in non-traditional settings emerged. Research directions have been grouped into two categories for the following discussion: the context and roles of metadata, and metadata creation.

	Logically Separable	Logically Integrated
Physically Separable	Traditional on-line card catalog elements not available in the document such as author's date of birth	Table of contents, abstracts, bibliographies, and other document structure reproduced in another location.
Physically Integrated	Word processed file with summary information element.	HMTL "meta" tags used to label elements within the document itself.

Figure 2. An analytical framework for metadata based on logical vs. physical separation or integration.

The Context and Roles of Metadata in Organizations

The factors that distinguish metadata in organizations from those in libraries or on the WWW need investigation. Some of the questions that arise include:

- What metadata are needed? Do library-originated metadata structures contain the right elements? Is there such a thing as too much metadata? What value do metadata elements convey to document users?
- Who should create it? When should professional catalogers be used, and when is the author the best source? What elements can be automatically extracted or generated? What factors influence metadata creation?
- Do the roles hypothesized for metadata occur in organizations? Are these roles useful in determining metadata needed, who should create them, and what should be captured?
- What should management's role be regarding metadata? When should it be mandated, standardized, automated, or prohibited? What incentives and disincentives to author-created metadata exist?
- What are appropriate measures of metadata quality? Should these measures vary by the role it plays?

Metadata Creation

The rapid proliferation of digital documents suggests that efforts to increase the range, amount, and quality of metadata in organizations short of bringing in specialists deserves investigation. Specifically, automated support for metadata capture, creation, and evaluation and mechanisms or facilities that encourage authors to create metadata, are likely fruitful directions.

Automation approaches might involve capturing metadata using "softbots" and web crawlers (e.g., [9] [10]; [13]) deployed on an organization's internal network as they are on the WWW. Because organizational information interests are less catholic than those of the general web browsing public, automating text-based creation could be enhanced by the use of information filters based on factors specific to organizations such as their structure, products, markets, industry-specific technologies, and genre. Efforts to ease the limitations of such post hoc approaches could address how to infer context information and avoid problems with data overload and spurious findings.

Additional areas for automation include: (1) ex ante creation of metadata at document inception and (2) applications that function as "interpretation engines" to aid in document understanding. Improving the information captured at the time of document creation could be accomplished working through the authoring application to extract information from document templates, software registration information, and user interactions with the document. For example, currently available word processing applications such as Microsoft's Word for Windows create a document descriptor from user interactions which can be supplemented by the author and searched by other components of the Microsoft Office 97 Suite [14].

Behavior-oriented approaches are needed that would increase the availability of metadata in organizations and the effectiveness of automation efforts. In most cases, document creators are in sole possession of key information and their cooperation in creating metadata must be pursued. Norms for information sharing [7] and subsidies or incentives for contributing to a common data store [26] are two areas with behavioral implications for metadata creation. Individual differences may also be important. In addition, the nature of the task and its context are factors to be considered in developing behavioral based approaches to the creation of metadata. For example, Paepcke [19] found people in the chemical analysis lab routinely annotated their tasks (e.g., running a chemical sample) while technical support personnel were often frustrated when it came time to categorize and document a reported problem and its solution, even when aided by a software application intended for this purpose.

Finally, document metadata could be enhanced by providing support for users during the evaluation process. One approach would be to create an "interpretation engine" that utilizes organizational information captured over time and infers context by tracing relationships and interactions. While many implementations are possible, an interpretation engine could have a component that routinely captured organizational information such as organizational charts, mission statements, annual capital budget priorities, external pressures such as regulatory deadlines, project and team composition and goals, and the like. To aid the prospective document user, perhaps utilizing data mining and inference techniques (cf. [10] regarding use of data mining techniques on the WWW for document discovery), the interpretation engine would provide context information about the document at the time of its creation or revision. For example, the mission, membership, and reporting structure of an ad hoc committee could provide crucial cues to appropriately situate project findings (cf. [19]). Thus, the interpretation engine represents a form of organizational memory, but one that provides context for interpretation (i.e., a form of "soft" or tacit knowledge [22]) rather than formalized knowledge about processes or products [25]. While meeting criteria for logical and physical separability, metadata used and created by an interpretation engine would not maintain an historical one-to-one relationship between metadata and its object.

Conclusion

Document metadata are no longer the sole domain of highly trained collections librarians. However, there is little guidance available to organizations about when and how to develop and provide metadata for their needs. In fact, we lack empirical evidence of how people in organizations interact with metadata, so we are uncertain about what the needs are. In this vacuum, an organization's Information Systems personnel may erroneously come to

believe that the more familiar data-oriented metadata with its greater formality is appropriate. Others may recognize differences but lack understanding of when richer description would return greater value or be confused by the wide variety of potential implementations (e.g., logical and physical separability vs. integration). When do factors such as persistence over time, who creates it, and how it is deployed on the infrastructure, become important for organizations to gain value from the information captured in digital documents?

The Dublin Core, a candidate intra-organizational digital document metadata structure, maps inconsistently to an organizational environment. While some elements appear to have organizational equivalents, others show greater variation, and many elements important for organizational needs do not appear at all. A greater conceptual and empirical understanding of the metadata element domain that is effective within organizations is needed.

Empirically- and conceptually-based discussions of digital document interactions in organizations suggest some directions for research on metadata in organizations. Situating metadata in organizational contexts can provided much-needed clarification of the roles played in addition to the familiar search and retrieval, element definition, and archiving. Means to improve metadata creation or interpretation could benefit from combining behavioral and automation approaches, and perhaps by breaking the conventional model of metadata as describing a single information object.

Paepcke reports that in organizations, "migration in and out of different information spaces is constant, as information is created, imported, and moved from one individual or group to another" ([19], p. 87). As recognized by others concerned with document interaction and use, metadata is a crucial element of digital information spaces [9]. Empirical and theoretical work on digital documents in organizations must no longer neglect metadata.

References

[1] Adam, N., Holowczak, R., Halem, M., Lal, N., & Yesha, Y. (1996) Digital Library Task Force. IEEE Digital Library Task Force Technical Activities Forum. URL: http://cimic.rutgers.edu/ieee_dltf.html

[2] Arms, W., Blanchi, C., Overly, E. (1997). An architecture for information in digital libraries. D-Lib Magazine [ISSN 1082-9873]. February. URL: http://www/dlib.org/dlib/februrary97/cnri/o2arms1.html.

[3] Bearman, D. & Sochats, K. (1996). Metadata requirement for evidence. URL: http://www.sis.pitt.edu/~nhprc/BACartic.html.

[4] Bishop, A. & Star, S. (1997). Social Informatics of Digital Library Use and Infrastructure. Annual Review of Information Science and Technology.

[5] Blair, D. & Maron, M. (1985). An evaluation of retrieval effectiveness for a full-text document retrieval system. Communications of the ACM. 28(3), pp. 289-299.

[6] Cammarata, S., Kameny, I., Lender, J., & Repologle, C. (1994). A Metadata Management System to support data interoperability, reuse and sharing. Journal of Database Management. 5(2), Spring, pp. 30-40.

[7] Constant, D., Kiesler, S., & Sproull, L. (1994). What's mine is ours, or is it? A study of attitudes about information sharing. Information Systems Research. 5(4), pp. 400-421.

[8] Dempsey, L. & Weibel, S. (1996). The Warwick metadata workshop: A framework for the deployment of resource description. D-Lib Magazine. July/August, URL: http://www.dlib.org/dlib/july96/07weibel.html.

[9] Dempsey, L. (1996). ROADS to Desire: some UK and other European metadata and resource discovery projects. D-Lib Magazine. July/August, URL: http://www.dlib.org/dlib/july96/07dempsey.html.

[10] Etzioni, O. (1996). The World Wide Web: Quagmire or Gold Mine? Communications of the ACM. 39(11), 65-68.

[11] Hart, P. & Rice, R. (1991). Using information from external databases: Contextual relationships of use, access method, task, database type, organizational differences and outcomes. Information Processing & Management. 27(5), 461-479.

[12] Hearst, M., Kopec, G., & Brotsky, D. (1996). Research in support of digital libraries at Xerox PARC; Part II: Papers and digital documents. D-Lib Magazine. June, URL: http://www.dlib.org/dlib/june96/hearst/06hearst.html.

[13] Lynch, C. (1997). Searching the Internet. Scientific American. [WWW article referenced.] URL: http://www.sciam.com/0397issue/0397lynch.html.

[14] Microsoft Corporation. (1997). Microsoft Word 97. [software]. Redmond, WA: Microsoft Corporation.

[15] Murphy, L. (1996). "Information Product Evaluation as Asynchronous Communication in Context: A Model for Organizational Research." Proceedings of ACM Digital Libraries '96, 134-142.

[16] Orlikowski, W. & Yates, J. (1994). Genres repertoire: the structuring of communicative practices in organizations. Administrative Science Quarterly. 39, 541-574.

[17] Paepcke, A. (1996a). Digital libraries: Searching is not enough; what we learned on site. D-Lib Magazine. May, URL: http://www.dlib.org/dlib/may96/stanford/05paepcke.html.

[18] Paepcke, A. (1996b). Summary of Stanford's Digital Library Testbed design and status. D-Lib Magazine. July/August. URL: http://www.dlib.org/ dlib/stanford/07paepcke.html.

[19] Paepcke, A. (1996c). Information needs in technical work settings and their implications for the design of computer tools. Computer Supported Cooperative Work: the Journal of Collaborative Computing. 5(), pp. 63-92.

[20] Pollock, A. & Hockley, A. (1997). What's wrong with Internet searching. D-Lib Magazine. March, URL: http://www.dlib.org/dlib/march97/03pollock.html.

[21] Salton, G. (1986). Another look at automatic text-retrieval systems. Communications of the ACM. 29(7), pp. 648-656.

[22] Schultze, U. & Boland, Jr., R. (1997). Hard & soft information genres: An analysis of two Notes databases. Proceedings of Thirtieth Annual Hawai'i International Conference on Systems Sciences. [CD-ROM version] New York, NY: IEEE Computing Society.

[23] Smith, T. R. (1996a) Call for Papers: International Journal of Digital Libraries Special Issue on Metadata and Digital Libraries. URL: http://www.llnl.gov /liv_comp/metadata/mail_dir/0027.html.

[24] Smith, T. R. (1996b). The meta-information environment of digital libraries. D-Lib Magazine. July/August. URL: http://www.dlib.org/ dlib/new/07smith.html.

[25] Stein, E. & Zwass, V. (1995). Actualizing organizational memory with information systems. Information Systems Research. 6(1), 85-117.

[26] Thorn, B. & Connolly, T. (1987). Discretionary databases: A theory and some experimental findings. Communications Research. 14(5), 512-528.

[27] Weibel, S. & Miller, E. (1997). Dublin Core Metadata Element Set: Reference Description. Document last modified January 15 1997. URL: http://purl.org/metdata/dublin_core_elements.

[28] Weibel, S. (1995). Metadata: The foundations of resource description. D-Lib Magazine. July. URL: http://www.dlib.org/dlib/july95/07weibel.html.

[29] Weibel, S., ed. (1996). A proposed convention for embedding metadata in HTML. [This is a report on the consensus of a break-out group from the W3C Distributed Indexing and Searching Workshop.] URL: http://www.w3.org/pub/WWW/Search/9605-Indexing-Workshop/ReportOutcomes/S6Group2.html

[30] Weibel, S., Godby, J., Miller, E. & Daniel, R. (1995). OCLC/NCSA Metadata Workshop Report. URL: http://www.oclc.org:5054/conferences/metadata/dublin_core_report.html.

[31] Yates, J. & Orlikowski, W. (1992) Genres of organizational communication: A structurational approach to studying communication and media. Academy of Management Review. 17(2), pp. 299-326.

[32] Yates, S. & Sumner, T. (1997) Digital genres and the new burden of fixity. Proceedings of the Thirtieth Annual Hawaii International Conference on Systems Sciences. [CD-ROM version.] New York, NY: IEEE Computing Society

Table 3. Dublin Core metadata elements (version 2). (From [27].)

ELEMENT	DESCRIPTION	EXAMPLES
COVERAGE	The spatial location and/or temporal duration characteristics of the resource	North America; 18th century; before 1922.
CREATOR	The person(s) or organizations primarily responsible for the intellectual content of the resource.	Authors; artists; photographers; illustrators.
DATE	The date the resource was made available in its present form.	December 3, 1996 (or 19961203).
DESCRIPTION	A textual description of the content of the resource.	Abstracts for document-like objects or content descriptions for visual resources.
FORMAT	The data representation of the resource.	Text/HMTL; ASCII; Postscript file; executable application; JPEG image; MIME type.
LANGUAGE	Language of the intellectual content of the resource	English; French; Japanese
OTHER CONTRIBUTORS	Person(s) or organizations(s) in addition to those specified in the CREATOR element who have made other significant intellectual contributions to the resource but whose contribution is secondary to the individuals or entities specified in the CREATOR element.	Editors; illustrators; translators; convenors.
PUBLISHER	The entity responsible for making the object available in its current form.	Publisher; university department; corporation.
RELATION	Relationship to other resources.	Chapters in a book; images in a document; items in a collection.
RESOURCE IDENTIFIER	String or number used to uniquely identify the resource.	URL; ISBN; any unique resource file name or key.
RESOURCE TYPE	The category of the resources. [Author's note: The original (1995) description called this the "genre" of the resource.]	Home page; novel; poem; working paper; preprint; technical report; essay; dictionary.
RIGHTS MANAGEMENT	A link to a source that provides information about terms and conditions for use and rights of that use.	Copyright notice; rights-management statement.
SOURCE	The work, either print or electronic, from which this object is derived, if applicable.	Paper version from which electronic source was transcribed; earlier version of same document.
SUBJECT	The topic of the resource, or keywords or phrases that describe the subject or content of the resource.	Selections from controlled vocabularies such as the Library of Congress Subject Headings.
TITLE	The name of the resource given by the CREATOR or PUBLISHER.	"The Elements of Style"; "Form 1040".

Table 4. Potential organizational equivalents to Dublin Core elements.

DUBLIN CORE ELEMENT	POTENTIAL ORGANIZATIONAL EQUIVALENT	INTRA-ORGANIZATIONAL EXAMPLES
COVERAGE	Duration, Extent	FY 1996; Western Division; Marketing Group
CREATOR	Author, Compiler, Manager, Producer, Team	CQI Team; EDI Steering Committee; Supervising Engineer; Webmaster
DATE	[same plus extensions]	Authorization date; Effective date; Release date
DESCRIPTION	[same]	Executive Summary; "Comments" from MS Word Summary Info box
FORMAT	[same]	[same]
LANGUAGE	[same]	[same]
OTHER CONTRIBUTORS	[same plus extensions]	Additional material supplied by; Reviewed by
PUBLISHER	[no direct equivalent; candidates included in examples]	Department of author, Sponsor, Authorizing party
RELATION	"Supersedes", Version, Phase identifier	Addendum; Initial Draft; Part 2 of 4
RESOURCE IDENTIFIER	Document, Procedure, or Report identifier	File name; URL; Policy AB10-2
RESOURCE TYPE	Genre, Document type; Template used	Customer Service Bulletin; ECR; Trouble report; Home page; Procedure; Action plan; Memo
RIGHTS MANAGEMENT	Copyright, Distribution limitations, Legal status	Company proprietary; Internal audit result; Authorized distribution list only
SOURCE	[same]	"Update to"; Transcription of interview, teleconference, or videoconference
SUBJECT	[same]	Key words; Acronym list; Project Names
TITLE	[same]	Report name; subject line from memo