2002

Defining Organizational Document Metadata: A Case Beyond Standards

Tero Paivarinta
Agder University College, tero.paivarinta@hai.no

Pasi Tyrvainen
University of Jyvaskyla, pasi.tyrvainen@jyu.fi

Tanja Ylimaki
University of Jyvaskyla, tanja.limaki@titu.jyu.fi

DEFINING ORGANIZATIONAL DOCUMENT METADATA: A CASE BEYOND STANDARDS

Tero Päivärinta

Agder University College, Dept. of Information Systems,
Serviceboks 422, N-4604 Kristiansand, NORWAY
Tel: +47 3814 1662, Fax: +47 3814 1029
Email: Tero.Paivarinta@hia.no

Pasi Tyrväinen, Tanja Ylimäki

University of Jyväskylä, Faculty of Information Technology,
P.O. Box 35 (Agora), Fin-40351 Jyväskylä, FINLAND
Tel: +358 14 260 3093, +358 14 260 3275, Fax: +358 14 260 3068, +358 14 260 2544
Email: pasi.tyrvainen@jyu.fi, tanja.ylimaki@titu.jyu.fi

ABSTRACT

Despite of the vast interest on metadata research and the number of metadata standards proposed for digital documents, a dearth of research efforts on their practical application in organizations outside the public archives and information collections continues. Organizations thus lack practical means to identify, implement, and utilize the organizational metadata needed for effective document management in the enterprise. We report a case in which document metadata in an engineering company with 2600 employees were defined after a major merger. At the outset, a review on 19 metadata standards created a baseline for the definition. The subsequent effort resulted in an organizational metadata definition that included 48 metadata elements, which did not straightforwardly correspond to any standard reviewed. The study can also be considered as a blueprint for a method of defining organizational metadata for digital documents.

1. INTRODUCTION

Metadata describe "...the enterprise for the purposes of managing the information resource" [Kerschberg et al. 1983], for the development, use, and integration of that resource. Descriptions of a number of organization-related entities can be included in this broad concept, e.g., those of documents, employees, customers, projects, products and services, and information technology [Tozer 1999]. Since a major proportion of information nowadays resides in digital documents [Lyman, et al. 2000], a traditional issue of information resources management must be revisited to meet the challenges of the digital era: document metadata in the organization [Murphy 1998, 2001].

Thus far, the use of document metadata has been explicit most often in the fields of bibliographic control and data management, within the disciplines of library and information science (LIS) and computer science, respectively [Dempsey & Heery 1998, Burnett, et al. 1999]. In addition, inside the wide definition of the metadata concept above, a number of related ideas have appeared in the literature for decades under other topics, such as corporate information architectures [IBM 1984, Brancheau & Wetherbe 1986], data dictionaries [Martin 1990], and information infrastructures [King & Shaw 1999]. These fields of study, however, neglect the issue addressed here. LIS lacks attention to the organizational dimensions of metadata, despite of discussing document-based information,

whereas data management and corporate-level information architectures, dictionaries, and infrastructures have targeted their focus on structured and voluminous databases, lacking attention to a great proportion of the heterogeneous digital documents. Several institutions advancing bibliographic control in public administration, libraries, and the World Wide Web have proposed metadata standards for digital documents (Appendix 1). However, in her recent essay, Murphy [2001] denotes the rarity, even absence, of the empirical studies on organizational document metadata.

This paper contributes by presenting a case study in which an engineering company constructed an enterprise-wide metadata specification for its documents. The specification was based on a survey on those metadata elements suggested by the contemporary public standards and recommendations as well as on the organizations' traditional metadata elements originating in previous systems and paper-based practices for document management. Our study addresses the need for scrutinizing organizational metadata elements, rather than adopting the elements recommended by any single metadata standard as such.

The remainder of the paper is organized as follows. Section 2 describes a case study of defining organizational document metadata. Section 3 discusses the implications of the study, concluded in section 4 with suggestions for further research.

2. CASE OF DEFINING ORGANIZATIONAL DOCUMENT METADATA

2.1 Base-line and objectives

Fortum is an international energy corporation headquartered in Finland, formed through the merger of the IVO Group and the Neste Group in 1998, employing approximately 19 000 persons in c. 30 countries. The company operates throughout the energy chain: from oil and gas production through refining, distribution, and marketing, to the engineering, construction, operation, and maintenance of power and heating plants and related infrastructures. This study focuses on two engineering companies within the corporation: Fortum Engineering and Neste Engineering, which started to integrate their operations in 1998, involving approximately 2600 employees.

These previously separate companies, both with their own history, corporate culture, document management practices, and legacy systems, have been struggling to integrate their information infrastructures after the merger. Several development initiatives on managing data, documents, and knowledge have been conducted, including the introduction of a new electronic document management system (EDMS), ARKI®DM.

The companies pursued effective search and retrieval of documents for every employee. This was to be achieved with metadata related to every document instance and powerful full-text search tools. The project thus sought answers to the questions to integrate the organizations' document resources:

- What is regarded as document metadata in general?
- Which metadata elements should be attached to every document under the EDMS in the target organizations and why?

One researcher participated in the project in collaboration with the archivists and experts on information management of the two target organizations. Her role was to review contemporary metadata standards, to facilitate the definition process as an external stakeholder, and to report the results together with the intermediaries responsible for the project inside the organizations. The research data consist of the observations, interview notes, and organizational documents she gathered during in the project as well as the project documents themselves.

2.2 Defining document metadata in Fortum

The metadata definition project was conducted between August and December 1999. To create the definition, each document instance can be described with metadata elements [AGLS 1999], which

vary among the existing metadata standards and recommendations. The elements, in turn, can be described with a set of generic attributes for the purposes of their comprehension and implementation. The project comprised four steps: (1) review of metadata standards and specifications, (2) studying the organizations and defining the attributes to describe metadata elements for the purposes of this project, (3) gathering the elements, and (4) refining and grouping the elements.

In order to create the baseline for the definition of metadata elements, the researcher conducted a literature review, including a comparison of metadata standards. The 19 metadata standards and specifications in Appendix 1 were considered potentially relevant here, as they focused on digital document metadata in a form or another. These were reviewed further with respect to their metadata elements. The final number of the standards and specifications included in the subsequent comparison was 18, as UNIMARC was chosen as the representative of the family of MARC standards (Appendix 2). In an article published just after this review, Burnett et al. [1999] had chosen six metadata standards for their analytical comparison. Hence, we considered this review to cover the field rather satisfactorily for the purposes of the project in question.

The Dublin Core (DC) metadata specification was reviewed first, as the initial intention had been to form a generic recommendation for digital document metadata, resulting in a baseline list of candidate metadata elements. Additional elements mentioned in the standards analyzed were added to the list. After all the standards and specifications were reviewed, the process was repeated and the comparison was revised and summarized as a comparison table. Appendix 2 lists those metadata elements mentioned in at least five metadata standards and specifications. We considered the rest of the elements - mentioned in three or fewer specifications - to be too domain-specific to guide the general level specification of organizational document metadata sought in the first step.

Using a simple heuristics, we defined the "core" elements to be such ones mentioned in ten or more metadata standards and specifications and identified sixteen of them from the specifications. We observed also that the names and contents of the elements vary between the standards, which lead to some difficulties in their interpretation. Therefore, for example, the concepts availability, access rights, use constraints and retrieval were regarded as equal (Appendix 2). Table 1 presents the identified core elements, arranged in alphabetical order.

Table 1. The core elements of the 18 reviewed metadata standards

Element name	Description
Availability	Access rights, use constraints or any other information available for document retrieval (e.g. time period when the document is valid).
Creator	Creator, author, designer or other person responsible for the document content.
Date	Date of creation or publication of the document.
Description	Textual description of the document content (e.g. abstract, table of contents).
Format	File format of the document (e.g. doc, pdf, tiff, dwg).
Identifier	Unique identifier of the document (e.g. URL, ISBN, ISSN).
Keywords	Keywords describing the document content. Thesauri may be used.
Language	Language of the document.
Location	Location (physical or logical) of the document (e.g. URL).
Notes	Notes and comments about the document content, usage etc.
Organization	Organization of the creator or organization, which is responsible of the document content.
Publisher	Publisher of the document.
Relation	Sources to which the document is based on; references of the document; relations to other documents and objects (e.g. document is part of a collection).
Subject	Subject of the document.
Title	Title of the document.
Type of resource	Type of the document (e.g. invoice, report, or memo).

The core elements found by Burnett, et al. [1999] matched almost identically to this review (which was conducted independently before their article was published). Possibly due to different interpretations, minor differences exist between these two sets of "core elements" concerning the naming and contents of the elements. However, they remain to be scrutinized elsewhere being out of the focus of this study.

As the second step, the organizational factors affecting requirements for document metadata were examined and the attributes to describe the resulting metadata elements were defined. Information was

gathered by discussions with the representatives of the target organizations, as well as from the related documents including Fortum annual report, instructions for document management, the quality manual, and reports on development projects concerning document and knowledge management. By studying the domain the researcher was able to achieve an understanding of the target organizations and their metadata requirements. The external researcher was considered also useful in providing an outsider viewpoint in addition to the organizational stakeholders.

The attributes to describe the metadata elements for the metadata requirements analysis were identified based on the AGLS metadata specification, ISO 11179-3, DC (Appendix 1), and the organizational needs observed in this step (table 2). The attributes 'Identifier' and 'Data type' were considered relevant for the implementation phase of metadata specification within the EDMS, and the values of those attributes were not defined further in this project.

Table 2. Attributes to describe metadata elements

Attribute name	Description	Source
Name	Name of the metadata element in question.	DC, ISO 11179-3
Identifier	Unique identifier of a metadata element (e.g. Internal identifier in a document management system).	DC, ISO 11179-3
Definition	Short description of a metadata element; what is the content of the element.	DC, ISO 11179-3
Purpose and comments	Justification; why is this metadata element needed? How is it used? Other comments or instructions.	AGLS, DC, ISO 11179-3
Producer(s)	Organization/department/team/person/role, that produces the content of a metadata element and is responsible for it.	DC, ISO 11179-3
User(s)	Organization/department/team/person/role, that uses a metadata element e.g. for searching information.	Organizational need
Obligation	Obligation of a metadata element: mandatory (M), conditional (C) or optional (O).	DC, ISO 11179-3
Max. occurrence	Number of values assigned to a metadata element. The repeatability of the metadata element.	DC, ISO 11179-3
Value qualifier	Name of the set of values or list of values that can be assigned to a metadata element. There can be one or more sets of values.	AGLS
Default value	Default value of a metadata element.	DC, ISO 11179-3
Sub-elements	Sub-elements of a metadata element.	AGLS
Data type	Data type of a metadata element (e.g. character string).	DC, ISO 11179-3
Examples	Examples of the values assigned to a metadata element.	AGLS
Standard	Standard or specification, which defines the metadata element in question (name of standard and element).	Organizational need

During the third step, the candidates for metadata elements were gathered to a spreadsheet. Because the metadata standards include few elements related to the use or the lifecycle of a document, as addressed also by Murphy [2001], it seemed evident that no single standard or specification could be applied in the target organization without adjustments. Therefore the standards and specifications were exploited only as a justification for the organizational metadata elements discovered in this step. Also the names of the elements defined in the standards were preferred when naming elements found in the target organizations. After defining the final Fortum metadata elements, it will be possible to encode them by using a specific standard, such as DC, if necessary. The element candidates were gathered from several sources. In parallel, guidelines for document management were under development at the corporate level. These instructions, parts of quality manual, other reports related to document and knowledge management, legacy databases (the metadata elements defined in them), and discussions with the representatives of the target organizations were the main sources of the candidate elements, in addition to the literature review; resulting in a list of 79 candidates.

The criteria for selecting the final elements were twofold. First, the elements were to be related to documents. Hence the candidates related to other resources, such as customers and products, were excluded. Only a few aspects of workflow, such as the reviewer and acceptor (of a document) were included in the specification. Second, only the "essential" elements related to documents were included. However, the decisions on this essentiality relied largely on the tacit expertise of the domain experts, who represented experienced professionals from every business and administration area of the target organizations. They had knowledge about which elements (originally needed in the paper-based environment) were no longer accurate and which elements would be currently needed. Furthermore, they decided which elements would be needed if the EDMS were adopted throughout the whole company. The heuristics to include an element in the specification was that more than 50% of the

participating representatives considered it useful. The elements were also categorized (table 3) in order to improve the comprehension and management of the metadata specification. The categorization was elaborated collaboratively among the project participants.

2.3 Results

The revised Fortum metadata specification consists of 48 elements, divided into 5 categories (table 3).

Table 3. Categories of metadata elements

Category name	Description
Basic metadata 1	Metadata elements related to document content description and physical description; the minimum metadata set for document description.
Basic metadata 2	Other metadata elements related to document content and physical description.
Change history	Metadata elements related to changes in document content.
Use and handling	Other metadata elements related to document usage and life cycle.
Relations to other resources	Metadata elements that relate to document context in an organizational communication, i.e. elements describing document's relations to other resources (e.g. products, projects and customers).

The justification for this categorization was primarily practical: the same categorization was to be used as a basis for implementing metadata key-in forms. The first category includes the minimum set of metadata elements that are to be attached to every document managed by the EDMS (or by a manual system). Additional metadata elements or categories may be taken into use as necessary, e.g., in a particular project or an organizational unit. Table 4 outlines the resulting metadata specification including the attributes: Element name, Definition, Purpose, and Obligation (as defined in table 2; the other attributes are not declared here due to space limitations); grouped by the five categories.

The first category of the metadata specification (i.e., the minimum set of metadata elements; table 4) and the core elements found from the reviewed metadata standards (table 1) overlap to a large extent. The core elements 'Availability', 'Keywords', 'Language', 'Notes' and 'Subject' are not included in the minimum set of metadata elements in the Fortum specification, but they can be found in the other categories. However, the minimum set of Fortum metadata still contains three elements not straightforwardly corresponding to the core elements of the metadata standards: 'Revision' (number), 'Rights', and 'Security Level' (although the latter two do overlap with 'Organization' and 'Availability', they are still defined somewhat differently). Among the reviewed standards DC includes a major part of the elements in the first category (table 4), providing thus the best match to the organizational needs of Fortum.

Table 4. New Fortum metadata specification

Category 1: Basic Metadata 1

	Element name	Definition	Purpose	Obl.
1	Identifier	A unique document identifier.	Search attribute.	O
2	Document type	Name of the document indicating its type or purpose.	Search attribute.	M
3	Creator	Creator of the document or the handler of a document received.	Reveals the responsible person. Search attribute.	M
4	Date of creation	The date when document is created or handled.	Evaluation of the usefulness and relevance.	M
5	Title	Title of document.	Search attribute.	M
6	Description	Description of document content, e.g. an abstract, table of contents, unstructured description	Evaluation of usefulness and relevance.	O
7	Revision (number)	Revision number of a published document that indicates a certain state of the content.	Evaluation of usability.	C
8	Appendices	Identifiers of external appendixes of a document, or parts of a compound document.	For information.	C
9	Is Appendix in	Identifier of that document, the document in question is part or appendix of.	For information.	C
10	Security level	Security level of a document, e.g. public, confidential, or classified.	Rights management. Safety issues.	M
11	Location	Location of the original document, e.g. URL, path, archive.	Retrieval.	M
12	Format and file name	Format and file name of the original document.	For information.	M
13	Rights	Organization which owns the document (content).	For information.	M

Category 2: Basic Metadata 2

	Element name	Definition	Purpose	Obl.
1	Keywords	Keywords describing the document content.	Search attribute.	O
2	Media	Media of the content of the original document, e.g. textual document, technical drawing, photograph, video, or voice.	Search attribute.	M
3	References	Documents referenced in the document in question.	For information. Evaluation of usefulness.	C
4	Size of hard copy	Size of a printout of the original document, e.g. A0 A3, A4.	For information.	O
5	File size	File size of the electronic document.	For information.	M
6	Number of pages	Number of pages of the document or units the document consists of, e.g. a number of files handled as a document.	For information.	O
7	Language	Language of the original document.	Search attribute. For information.	O
8	System requirements	System requirements to modify the original document.	Reveals the system needed for modification.	O

Category 3: Use and Handling

	Element name	Definition	Purpose	Obl.
1	Date of receive	Date and time, when the document is received from the external party, like customer or supplier. Sender is located into the creator element.	Control and management.	C
2	Date of delivery	Date and time, when the document is delivered to external party.	Control and management. Search attribute.	C
3	Addressee	Party to which the document is delivered.	Control and management.	C
4	Date of publication	Date and time when the document is published (to be available to users).	Evaluation of usability and accessibility.	C
5	Publisher	Name of the publisher.	Reveals the responsible person.	C
6	Replaces	Identifier (or title) of the document which is replaced by the document in question.	Evaluation of accuracy and usability.	C
7	Is replaced by	Identifier (or title) of the document which replaces the document in question.	Evaluation of accuracy and usability.	C
8	Check-out by	Name or identifier of the person who has checked-out the document either from the EDMS or from paper-based archive.	Control and management. For information.	C
9	Date of check-out	Date of check-out.	Control and management. For information.	C
10	Back-up	Date & procedure of back-up, e.g. microfilming, scanning, CD-ROM.	For information.	C
11	Status	Phase of the document's life cycle, or state of design (in case of technical drawings).	Evaluation of accuracy and usability. Search attribute.	M
12	Version	Version number indicates the state of unfinished and unaccepted document content (see revision).	Evaluation of accuracy and usability. For information.	O
13	Validity	Start and end date of the period the document is valid. Status has to be final.	Evaluation of usability. Selecting documents for disposal.	C
14	Disposal	Date of disposal of the original document and name of the person in charge of the disposal. Metadata can be preserved.	Evaluation of accuracy and usability of a printout.	C

Category 4: Change History

	Element name	Definition	Purpose	Obl.
1	Reviewer	Reviewer of the document. Reviewer is responsible of the accuracy of the content.	Control and management.	C
2	Date of review	Date of review.	Control and management.	C
3	Acceptor	Acceptor of the reviewed document.	Control and management.	C
4	Date of acceptance	Date of acceptance.	Control and management.	C
5	Description of modification	Description of modification indicates what has been modified and how the modified part used to be.	For information.	C
6	Modifier	Modifier ID or name.	For information. Evaluation of accuracy.	C
7	Date of modification	Date of latest modification.	For information. Evaluation of accuracy.	C

Category 5: Relations to Other Resources

	Element name	Definition	Purpose	Obl.
1	Customer	Customer ID or name is given if the document is related to a customer in some other way than being received from or sent to customer.	Search attribute.	C
2	Plant or delivered system	Name or ID of the plant or system delivered.	Search attribute.	C
3	Project or work	Name or ID of a project or work.	Search attribute.	C
4	Supplier	Name or ID of the supplier is given if the document is related to a supplier in some other way than being received from or sent to supplier.	Search attribute.	C
5	Additional identifiers	Additional identifiers that are needed in respect of use and search of the document, e.g. device ID.	Control and management. Search attribute.	C
6	Additional information	Any additional information needed for search, retrieval and utilization of the document.	For information. Search attribute.	O

When comparing the Fortum metadata specification with the elements identified with the standards (Appendix 2), a few organization-specific elements excluded from the standards emerged, including 'Revision' (number), 'Size of a hard copy', 'Addressee', 'Description of modification'. Alike, the Fortum elements concerning the change history and life-cycle of a document ('Reviewer', 'Status', 'Acceptor') as well as other resources (‘Customer', 'Supplier', 'Plant', and 'Project or work') were rather absent in the reviewed standards. None of the reviewed standards and specifications included all the metadata elements defined in Fortum.

The target organizations' document resources had traditionally been strongly polarized into technical drawings originating mainly in computer-aided design (CAD) systems and "non-technical" documents originating in office applications. Thus the existing document (and metadata) databases were somewhat focused either on the one or the other. These legacy databases and their retrieval tools had also been rather isolated. Simultaneous retrieval of technical drawings and textual documents related to a project, a plant or some other topic was difficult. However, 42 metadata elements of the new specification could be found in one or more metadata specifications of the legacy databases. Hence, one contribution of the new specification was to merge the legacy specifications and to dissolve the distinction into drawings and textual documents. For example, one of the previous specifications defined the 'Title' element, while the other specification included the 'Description' element. Both of these are now included in the new specification. Alike, 'Security Level' was included in two old specifications and 'Location' was mentioned in a third one; now the both being a part of the new.

The Fortum specification also combined, renamed, or extended some of the previous metadata elements. A few examples follow. The 'Identifier' element now contains both the 'Document Identifier' and 'Sheet Code' of a multi-sheet technical drawing. The 'Microfilmed' element was renamed as the 'Back-up' element including now also information about scanning and other back-up procedures concerning a document. The 'Document Type' element combines the types of office documents and technical drawings and the 'Creator' element includes 'Internal creator', 'External creator' and the 'Person handling the document' from previous specifications. The 'Location' element contains a file path or the physical location (of a paper document) in an archive, and the 'Status' element describes the design phase of a technical drawing or the life-cycle stage of an office document.

A few previous elements were excluded from the new specification, as they were regarded as obsolete or not pertaining to documents as such. For example, the 'Scale' of a technical drawing was considered too specific. Furthermore, the elements 'Functional Group' and 'Document Group' were considered to relate to the folder structure of the EDMS, instead of documents per se. The 'Accumulation and Screening of Documents' (for disposal) pertained to more general archiving policy within a project or an organization and 'Approval Route' was regarded as a workflow issue irrelevant to all document instances. The 'Distribution' element originating from the paper-based era was removed, as there exists no longer a need for physical distribution in the new EDMS and the 'Distribution Information' was now replaced by the 'Date of Publication' and 'Publisher'.

The new specification contains also six elements additional to the previous ones: 'File Size', 'Date of Publication', 'Publisher', 'Validity', 'Customer', and 'Plant or Delivered System'. These originated from the corporate instructions for document management and the expertise of the participants. The 'File Size' is important in the digital environment for deciding whether to transfer a document through slow communication links. The 'Validity' element declares the temporal status of usability and expiration, e.g., how long the document can be used as a reference. The elements 'Customer' and 'Plant or Delivered System' link a digital document to the customer and product databases.

3. DISCUSSION: APPLYING METADATA STANDARDS IN ORGANIZATIONS

Three of Fortum's 13 core metadata elements were not directly corresponding to the core elements of the reviewed standards and a few of the standards' core elements were, in turn, excluded from the minimum requirements of Fortum. Hence, we suggest that the contemporary standards and specifications do not necessarily correspond to the contextual requirements for organizational

document metadata straightforwardly. Our case thus seems to support Murphy's [2001] speculation about the needs for scrutinizing organization-specific metadata elements and her critique of the mainly bibliographic metadata stressed by the contemporary standards to be adopted in organizations as such.

However, the review on the 19 contemporary metadata standards at the outset of the project appeared useful for facilitating the definition of new metadata elements. As DC provided the best fit among the reviewed standards with the core elements regarded as essential in Fortum, we consider it to be useful as a base-line also for other organizational initiatives defining document metadata. Anyhow, this case also clearly illustrated the importance of the knowledge embedded in organizations' existing information systems and tacit experience of the domain experts in providing important viewpoints to organizational document metadata beyond the contemporary standards. Hence, we suggest that one should not rely on the contemporary metadata standards only - knowledge of the business context will probably complement the results, with a terminology already familiar to the organization in question. The metadata definition process and results reported above can also be regarded as a blueprint towards a more generic method for defining organizational metadata, to be elaborated further in similar initiatives elsewhere. For the next research cycle elaborating this kind of method we would suggest the following shortcomings of this process to be scrutinized and improved.

The starting point of the definition process requires further investigation: whether to adopt a particular standard extending it with elements derived from organizational needs or whether to build an organizational specification first, then exploiting a standard for naming, refining the contents of, and categorizing those metadata elements that match to a particular standard. Yet another option is to mix and match parts of several standards while creating an organization-specific metadata definition. Our current pragmatic suggestion is to use the core metadata from DC and our review as the baseline.

Furthermore, the identification of numerous candidates for metadata elements was easy, whereas it was difficult to decide which elements actually do relate to documents (rather than to other resources). The selection criteria for "essential metadata" were neither thoroughly elaborated yet in this case. A future method could also elaborate quantitative measures to optimise the number of metadata elements of organizational documents. The main problem thus resides not in the definition of metadata elements in an organization *per se* rather than in optimizing the assembly of elements into a set large enough to be useful, but a small enough to be used. Convincing the members of the organization on the importance of metadata and metadata specification and motivating them to thoroughly fill in a metadata form for documents can be challenging, even if some of the metadata could be extracted automatically from the document contents (title, author) and from the workstation of the employee (e.g., the name of the project). As a 100% automated metadata generation remains unlikely, the means for filling in metadata by the document author ought to be easy-to-use in the first place.

According to Milstead and Feldman [1999] "the value of metadata elements is limited if there is no common agreement on what elements to use or what their content should be". In this project neither value qualifiers, which indicate how a value of a certain element should be interpreted [AGLS 1999], nor the components of the elements were fully formalized yet. In fact, these issues are being discussed in subsequent development initiatives in the target organizations and at the corporate level.

4. CONCLUSION AND DIRECTIONS FOR FURTHER RESEARCH

This report represents a case study on defining digital document metadata in an organization. The results suggest that organizations might want to scrutinize organizational specifications for their document metadata instead of adopting any single metadata standard as such. The review of 19 metadata standards conducted in a connection to the development process in this case can be regarded as a useful basis for facilitating the definition of document metadata in other organizations. Finally, the definition process above can be regarded as a blueprint for similar efforts elsewhere.

In the future, longitudinal research on the evolution of the Fortum metadata specification would provide insight into the dynamics of organizational document metadata and the evolution of EDMSs in

organizations. The granularity of digital document information in organizations will be ever more diverged in the near future, e.g., along with the diffusion of XML and related technologies for managing ever increasingly structured documents [cf. Chin 2001] and linking digital document entities with each other in several ways [cf. Yoo & Bieber 2001]. This trend will revolutionize the concept of document metadata further, providing challenging issues for researchers as well as practitioners within the field. The relevance of organizational metadata related to digital information at various levels of granularity will proliferate - will the research community be able to follow?

ACKNOWLEDGEMENTS

We wish to thank Kari Kotirinta, Tuire Paajanen and Ulla Sorsa for their participation and efforts. The Finnish Technology Development Centre has funded this research.

REFERENCES

AGLS (1999). *AGLS (Australian Government Locator Service) User Manual*. http://www.naa.gov.au/govserv/agls/user_manual [September, 1999].

Brancheau, J.C. and J.C. Wetherbe (1986). Information Architectures: Methods and Practice. *Information Processing and Management*, 22 (6), 453-463.

Burnett, K., K.B. Ng and S. Park (1999). A Comparison of the Two Traditions of Metadata Development. *Journal of the American Society for Information Science*, 50 (13), 1209-1217.

Chin, A.G. (2001). *Text Databases & Document Management: Theory & Practice*. Idea Group, Hershey PA.

Dempsey, L. and R. Heery (1998). Metadata: A Current Review of Practice and Issues. *Journal of Documentation*, 54 (2), 145-172.

IBM (1984). *Business Systems Planning: Information Systems Planning Guide*. IBM, Atlanta.

Kerschberg, L., D. Marchand and A. Sen (1983). Information System Integration: A Metadata Management Approach. In *Proceedings of the Fourth International Conference on Information Systems (ICIS)* (Ross, C.A. and E.B. Swanson Eds.), pp. 223-240, ICIS.

King, R.C. and M.J. Shaw (1999). Information Infrastructure for Enterprise Coordination and Integration. *Journal of Organizational Computing and Electronic Commerce*, 9 (1), 1-5.

Lyman, P., H.R. Varian, J. Dunn, A. Strygin and K. Swearingen (2000). *How Much Information?* School of Information Management and Systems, University of California at Berkeley. http://www.sims.berkeley.edu/how-much-info/ [November 10, 2000].

Martin, J. (1990). *Information Engineering, Book II*. Prentice Hall, Englewood Cliffs NJ.

Milstead, J. and S. Feldman (1999). Metadata: Cataloging by Any Other Name… *ONLINE*, 23 (1), 24-30.

Murphy, L.D. (1998). Digital Document Metadata in Organizations: Roles, Analytical Approaches, and Future Research Directions. In *Proceedings of the 31st Hawaii International Conference on System Sciences* (Sprague, R.H. Ed.), CD-ROM, IEEE Computer Society Press, Los Alamitos CA.

Murphy, L.D. (2001). Addressing the Metadata Gap: Ad Hoc Digital Documents in Organizations. In *Text Databases & Document Management: Theory & Practice* (Chin, A.G. Ed.), pp. 52-77, Idea Group, Hershey PA.

Tozer, G. (1999). *Metadata Management for Information Control and Business Success*. Artech House, Boston.

Yoo, J. and M. Bieber (2001). A Systematic Relationship Analysis for Modeling Information Domains. In *Information Modeling in the New Millennium* (Rossi, M. and K. Siau Eds.), pp. 485-501, Idea Group, Hershey PA.

APPENDIX 1. An overview of 19 metadata standards and specifications with references

Standard	Full name	Description & Reference
AGLS	Australian Government Locator Service.	Used to describe Australian government resources in the network to facilitate discovery of those resources. Based on Dublin Core. Extended with additional elements. http://www.naa.gov.au/govserv/agls/user_manual/cover.htm. [September, 1999].
BibTeX		Used in scientific and academic communities and in industry. Originally a program to create bibliographies in conjunction with the LaTeX Document Preparation System. http://www.coling.uni-freiburg.de/~neuhaus/manuals/btxdoc/btxdoc.html,. [September, 1999].
CDWA	Categories for the Description on Works of Art	Used in communities that provide and use art information (e.g. museums and archives). Provides guidelines for formulating the content of art databases. Consists of 30 categories. http://www.getty.edu/gri/standard/cdwa/HOMEPAGE.HTM. [September, 1999].
CIMI	Computer Interchange on Museum Information	Description of cultural heritage information in museums, archives and libraries. http://www.cimi.org. [September, 1999].
CIMS	Chesapeake Information Management System	Description of all kinds of information within the CIMS-system. Based on FGDC and NBII Metadata Standard. http://www.chesapeakebay.net/cimsindex.htm. [September, 1999].
DC	Dublin Core (Dublin Metadata Core Element Set)	Description of electronic resources on the Internet. Can be applied in diverse communities, e.g. museums, libraries, government agencies, and commercial organizations. Consists of 15 elements. http://purl.org/DC/. [September, 1999].
EEVL	The Edinburgh Engineering Virtual Library	A set of elements created in the EEVL project (in the UK) to describe engineering resources in the Internet. Consists of 22 elements. http://www.eevl.ac.uk/pub3.html. [September, 1999].
FGDC	Federal Geographic Data Committee: Content Standard for Digital Geospatial Metadata	A complex model to describe digital geospatial information. Includes 10 sections, some of which can be compound. Over 300 elements that can also be compound. http://www.fgdc.gov/metadata/contstan.html. [September, 1999].
GILS	Global Information Locator Service	Describes government resources in the network (in USA) to facilitate discovery. Based on DC, MARC and FGDC. http://www.gils.net/elements.html. [September, 1999].
IMS	Instructional Management Systems	IMS project aims to developing a standard for distributed learning environment including a metadata specification. Metadata comprises of four types of elements: categories, semantic elements, abstract data types and data types. http://imsproject.org/work_public/meta-data_did188.html. [September, 1999].
ISO/IEC 11179-3	Specification and standardization of data elements	Does not define (meta)data elements, but the attributes used in describing (meta)data elements. Attributes are classified into five categories, 10 general descriptors. http://www.sdct.itl.nist.gov/~ftp/l8/other/coalition/Coalition.htm. [September, 1999].
MARC	Machine Readable Catalogue Format	Originally designed for exchanging library catalogue records between libraries. Format has been developed by various organizations according to their own requirements, and many MARC standards have been evolved (e.g. USMARC, UKMARC, and FINMARC). Consists of numbered tags classified in 9 groups. http://www.loc.gov/marc/marc.html. [September, 1999].
ODMA	Open Document Management API	API that allows applications and document management systems to inter-operate. Contains a set of document attributes (i.e. metadata elements). http://www.aiim.org/odma/odma20.htm. [September, 1999].
PANDORA	Preserving and Accessing Networked Documentary Resources of Australia	PANDORA project aims e.g. at developing a proposal for a national approach to the long-term preservation of online publications. Metadata elements are classified into five categories. http://www.nla.gov.au/pandora/ldmv2.html. [September, 1999].
RFC 1807	A Format for Bibliographic Records by R. Lasher & D. Cohen	RFC is a memo, not a standard. It defines a format for emailing bibliographic records of technical reports. Used in US technical communities. http://ifla.inist.fr/documents/libraries/cataloging/metadata/rfc1807.txt. [September, 1999].
RLG	RLG Working Group on Preservation Issues of Metadata	Used in scientific communities. Consists of 16 descriptive data elements that are associated with digital master files that have preservation-based intent. (RLG = Research Libraries Group). http://www.rlg.org/preserv/presmeta.html. [September, 1999].
SOIF	Summary Object Interchange Format	SOIF is actually an internal record format of the Harvest and related systems. http://www.roads.lut.ac.uk/RADAR/soif-review.html. [September, 1999].
TEI	Text Encoding Initiative Independent Headers	Used in research communities and libraries. Defines a set of generic guidelines for the representation of textual materials in electronic form. Contains four parts: file, encoding, profile, and revision description. http://www.uic.edu/orgs/tei/.[September, 1999].
UNIMARC	The Universal MARC Format	International exchange of bibliographic data in machine-readable form between libraries (See MARC).

APPENDIX 2. Comparison of elements defined in the reviewed metadata standards

Due to the space limitations, the full table can be seen in the URL: http://www.jyu.fi/~pttyrvai/ecis/odm2002a2.pdf and reference information in odm2002.html