Metadata use in document management systems, which support business processes

Lena-Maria Öberg

INTRODUCTION

The issue of preserving information is nothing new. The problem is how to preserve information that will be interpretable and understandable over time. Technology development leads to new generations of systems, file formats, etc., but the information often has to be interpretable over long periods of time which will include several technical generations. The importance of metadata to describe the context, which is the key to interpretation of information, is well known. With the widespread use of computerized information systems, a need for conceptualization of metadata arises (Gilliland-Swetland 2000). Today there are several examples of developed metadata sets for recordkeeping and archival needs. VERS, RKMS and ISAD(G) are some examples. The issue regarding

when to include recordkeeping metadata, during the design phase of systems versus after records have been created and accumulated, has been discussed. This discussion includes questions about which metadata is possible and feasible to include at the creation of a document (for example) and which should be added after a document has been created. Hofman (1996) writes that a pro-active approach is essential to be able to preserve (at reasonable cost) accessible and interpretable information for the long-term. This is also supported by Shepherd and West (2003). Their article is based on the assumption that some metadata has to be created during the creation phase. But when an organization is implementing a document management system, it is an important matter to decide what kind of metadata they should capture and when it should be captured. According to Hedstrom (2001) the capturing of metadata at the creation stage is especially important when the information is electronic. The creation of information that has to be prepared could of course be both electronically and paper-based. This paper concerns electronically created information. Nowadays it is common that organizations use some kind of electronic document management system to handle their documents. There are a lot of examples of document management systems on the market (examples from Sweden are: Archs Docs, Platina and W3D3). What kind of documents are included in the electronic document management systems differs from organization to organization, but some of those documents are part of the information that should be preserved for long time. The reality in Sweden is that there is no system that could be called a fully developed record[1] system and therefore this study covers electronic

---

[1] For definition and discussion of record systems see the next section

document management systems. This study covers what influences the metadata capturing which is an essential part of preserve interpretable and understandable information over time. Description of those differences will give important knowledge in developing record systems in the future. The aim of this research is to study metadata use in document management systems and compare it with metadata that is recommended for recordkeeping.

The research questions in this article are:

- Which metadata elements are used in document management systems that support business processes?
- What are the differences between those metadata and existing sets of metadata, which have been developed for recordkeeping?

The contribution of this paper is a reflection of metadata use in two electronic document management systems that support business processes in Swedish organizations. The reflection includes comparisons of metadata use within two organizations with three metadata specifications.

THEORETICAL STANDPOINT AND RELATED RESEARCH

The Australian School on recordkeeping that is based upon the Records Continuum Model (Upward 2005) makes a distinction between record systems and electronic document management

systems. According to Upward (2005) a characteristic of the Continuum Model is the view of records[2] as unstable. There is no such thing as end products that should be delivered to an archival institution and from this arises a need for continuing addition of process metadata meanwhile they change during space-time. Upward (2000) writes that records can have multiple lives in space-time, and a record is never finished in its creation, it is continuously in change. This paper and this research are based on the same theoretical viewpoint as records continuum model. Thereby it is important that consider the long-term perspective already in a document management system if it is where the information is created and managed. The Records Continuum model thereby form an idealistic view of how information that has to be preserved should be managed and is one of two views that have been used to compare empirical data with existing standards. This study confirm what was showed by Öberg and Borglund (2006) in their study of four Swedish organizations. All those Swedish organizations have not yet implemented the records continuum model thought in their strategies for information and information management. Their strategies have greater similarities to the life-cycle model and paper-based information. This shows in for example a clear distinction between electronic document management system and long-term preservation. And when it comes to long-term preservation it is a problem that is owned and should be solved by archivists. As shown in the introduction a proactive approach is necessary to be able to take care of electronic information over time. While these differences

---

[2] For discussion about differences between record and information see section Definitions.

between the theoretical viewpoint and reality have been revealed there is a need for studying how this influences electronic documents management systems within Swedish organizations.

There is research in this field that focuses on development of metadata specifications (see for example McKemmish et al. 1999) and there are examples of comparison between different metadata sets (Cunningham, 2001; Duff, 2001; Shepherd and West, 2003). Päivärinta, Tyrväinen & Ylimäki (2002) have made an interesting study of organizational document metadata. Their paper includes a comparison of 19 metadata standards and specifications. The result from the comparison was later used as an input for the selection of metadata in an electronic document management system used by an energy corporation in Finland. Päivärinta, Tyrväinen & Ylimäki's (2002) study is relevant to this research but the selection of metadata standards does not include any metadata standards or specifications originating from recordkeeping. The Finnish study is however one example of case studies within this field. In 1998 Murphy writes about the lack of empirical understanding of the metadata use within organizations. Day (2003b) expresses the same thing in his article when writing that the development moving towards more practical focus but that there still is an urgent need for example advice to implementers. There are still only a few examples of empirical studies of metadata use in organizations. This paper will contribute to this lack of knowledge.

DEFINITIONS

Process

This paper is based on a study that is part of a research project called *Archives of the Future*. The author has in this project performed several studies that all focus on business processes and the traceability between process and information. The document management systems that have been the focus of the study that is presented here are both intended to support the business processes of their respective organizations. Both organizations are working through the process of describing and developing their business processes. As a part of this progress they have chosen to develop document management systems that will strengthen the connections between their documents and their business processes. A process is in this paper defined as follows: "Processes are differences or changes in states of space, time and form. It can produce information, matter-energy or both" (Eriksson 2004 p.490). A process consists of activities and activities consist of transactions. When the transaction needs verification the transaction will lead to the creation of information (in this study called a document, see below). A process can also be divided into sub processes, which can be necessary to give a reasonable level of detail. The process of describing and developing processes is of interest because it will offer the organization the possibility of tracing their information (in this case in the form of documents) to their business processes. This is an important part of preserving the context. Within archival practice it is essential that the preserved information can be used as evidence for organization transactions (Bantin, 2002; Duranti, 2001; Thomassen,

2001). Cox (2001) stated that the evidential value of a record could only exist if the content, structure and context are preserved.

Records vs. Document

According to ISO-15489-1 (2001) a record is defined as: "Information created, received, and maintained as evidence and information by an organization or person in pursuance of legal obligations or in transactions of business." A record is then a verification of business transactions as discussed in the section about process. A record can be a single item or a group of items (Reed, 2005). Reed (2005) further noted that the organization has a choice about whether they manage each record separately or if they manage several records as the record of one transaction. This means that a record (in this case a document) can include connections to several transactions. The definition of records includes many different formats of information. A document is one of those formats; a document is here defined as a coherent unit like a PDF- or a DOC-file. The document type that has been the focus of this study is text documents. In this paper the term document is used because this term is used by the studied organizations. None of the organizations within this study has chosen to manage one transaction per document; one document can be connected to several transactions.

Metadata

Metadata is a term that has been used since the middle of the 1970s. The most common definition within the context of databases and computer science is "data about data" (Sundgren and Steneskog,

2003). This definition gives one view of what metadata can be but since this paper regards use of metadata to preserve context the following definition is suitable: "the sum total of what one can say about any information object at any level of aggregation" (Gilliland-Swetland, 2000).

There exist several metadata initiatives in the field of preservation but with various purposes. There are many possible ways of classifying different types of initiatives or use of metadata. One way of classification is by making a classification based on type, for example administrative metadata and descriptive metadata, see for example Gilliland-Swetland (2000) and Lagoze (2006). Hofman (2004) and Day (2003a) on the other hand base their classification on where the metadata sets have their origin. They have groups like recordkeeping, preservation and national and research libraries. Hofman (2004) has classified the initiatives in two groups, record oriented, retrospective approach and business process oriented approach. His argument for this classification is based on the underlying view of records. The retrospective approach has its base in a paper environment and tends to focus on physical objects. This approach is built upon the life-cycle model and metadata thereby is an issue when for example describing and arranging records after they are transferred to an archive. The business oriented approach "takes the business process as a starting point for determining metadata requirements and emphasises the strong interrelationship between records, the business process in which they are created, and the management of records" (p. 6). This approach is based on the continuum model (Hofman, 2004).

224

METHOD

The method used for this study is a qualitative interpretative method. The chosen qualitative research method is the case study, a method well suited to information systems research where a phenomenon must be studied in reality and in its natural environment (Myers and Avison, 2002; Yin, 2003). The course of action consists of two main parts. The first part consisted of an empirical case study and the second analyzed differences between the empirical data and metadata set. The case study was carried out in two organizations and as recommended, different data collection techniques have been used in order to capture the complexity and uniqueness of the settings of interest (Eisenhardt, 1989; Yin, 2003). The organizations were chosen as suitable for the research project because they had an implemented computerized document management system, which supports business processes. The data collection methods were semi-structured interviews, the collection of screen shots from the systems and access to user manuals and system documentation. The interviews were held with persons that had been part of the analysis, design and implementation phases of the systems. The interviews were recorded and transcribed directly afterwards and are available on demand.

Organization A is a public authority in Sweden whose main business are national registration, taxes and administration. The organization has 13600 employees, who are located at 100 different locations in Sweden. Organization A implemented their system late in 2003. The EDMS mostly includes authority documents and management

documents (examples: decisions, minutes of meetings and policies). At the time of writing the system contains over 40000 documents and all employees are users of the system. Organization A has done some development of the organization using process modeling but the system is not used throughout the whole organization.

Organization B is a result-oriented section of a public company. Their main business activity is maintenance of infrastructure and they have the overall responsibility for railway transportation in Sweden. They have 7300 employees located at offices all over Sweden. Organization B implemented their system in the beginning of 2004. The EDMS only includes authority documents (examples: document management plan and manual) and management documents at the moment. About 3000 of the employees have access to this system. Organization B is certified by ISO 9000-1 and structures their work in a process-oriented way.

Analysis

The second part of this study is to compare the metadata found from part one with the recordkeeping metadata sets. The first part of the analysis was to choose sets for the comparison. As mentioned in the section about metadata there are a lot of examples of metadata specifications within the field of recordkeeping. This research does not claim to compare and evaluate different metadata specifications, so the selection is not covering the field in any sense. The three sets that were chosen are VERS (Public Record Office Victoria, 2006),

RKMS (Recordkeeping Metadata Project, 2006) and ISAD(G) (ICA, 1999). All of those sets are well established. VERS represent a metadata specification with its origin in the context of official records meanwhile RKMS is the result of a research project within the Records Continuum Research Group from Australia. ISAD(G) is an example of an internationally accepted archival standard by the International Council on Archives (ICA). They are all presented in more detail below. According to Hofman's classification VERS and RKMS are typically business oriented and ISAD(G) represents the retrospective group. The selection thereby represents different types of metadata specification. When the selection of metadata specification was made the analysis was done by comparison of the metadata elements from the empirical data and the metadata elements of the three sets with an aim of seeking for differences. Information about the three selected sets was gathered from manuals and other official documents. The result is presented in tables, which show differences and similarities between metadata elements from the empirical data and the metadata specifications.

Descriptions of the selected metadata sets

VERS (Public Record Office Victoria, 2006): the development of Victorian Electronic Records Strategy (VERS) started in 1996 in a project that the Public Record Office of Victoria was running. VERS is a framework of standards, guidance, training, consultancy and implementation projects, which is centered around the goal of reliably and authentically archiving electronic records. This study uses only the metadata standard from the strategy. VERS metadata set is divided into the following groups, encapsulated object metadata,

records VEOs, document, encoding, signature block and file VEO. VEO is an abbreviation for VERS encapsulated object. The Victorian government uses VERS and has implemented it in their agencies. VERS is, according to Day (2003a), in opposition to RKMS, a schema based on a practical base.

RKMS (Recordkeeping Metadata Project, 2006): The Australian Recordkeeping Metadata Schema (RKMS) is the main deliverable from the SPIRT Recordkeeping Metadata Research Project (McKemmish et. al, 1999). According to McKemmish et al. (1999) the project RKMS include a "standardized set of structured recordkeeping metadata elements and a framework for developing and specifying recordkeeping metadata standards and finally a framework for reading or mapping metadata sets in ways which can enable their semantic interoperability by establishing equivalences and correspondences that can provide the basis for semi-automated translation between metadata schemas". This study has used the standardized set of structured recordkeeping metadata elements. The scheme is divided in four classes, business, agents, records and business-recordkeeping. According to Day (2003a, 2003b) RKMS is conceptual in its nature and this include for example possibilities to include other metadata schemas in RKMS.

ISAD(G) (ICA, 1999): The International Standard on Archival Description, General (ISAD(G)) was developed by the International Council on Archives (ICA). The first edition of the schema builds on existing practice from Canada, United States and Great Britain (Duff, 2001). According to Duff (2001), ISAD(G) was developed for

archivists creating descriptions at the end of the life cycle and the focus was on paper based records. The metadata elements are divided into seven groups: identity statement area, context area, content and structure area, conditions of access and use area, allied materials area, notes area and description control area (ICA, 1999).

EMPIRICAL RESULT

The focus of the study has been two electronic document management systems (EDMS). Organization A implemented their system late in 2003. The EDMS mostly includes authority documents and management documents (examples: decisions, minutes of meetings and policies). At the time of writing the system contains over 40000 documents. A user can create and read documents within the EDMS while they are connected to a standard word processor. Before a document gets published the responsible chief has to approve it. The publication makes the document available to all users. There are 13600 employees who are all users of the EDMS. The main reason for developing this system was a desire to create a document just once, and save it in only one place, and then make it available depending on the situation. The metadata is captured when the user creates a document. The system does not allow the user to ignore the metadata window. Some of the metadata is also included in the letterhead in all documents, but is stored separately. Organization B implemented their system in the beginning of 2004. The EDMS only includes authority documents (examples: document management plan and manual) and management documents at the moment. The next types of documents that will be included are

forms. Before a document gets published in the EDMS it has to be examined and established. As in the other system under study a user can create and read documents within the EDMS while it is integrated with a standard word processor. Today the system contains several hundreds of documents. All employees have read access (3000) but only about 100 have access to create documents. The main reason for developing the system was a certification process, which set up new requirements on traceability of information. The metadata is captured when the user creates a document. The system does not allow the user to ignore the metadata window. Some of the metadata are also included in the letterhead in all documents but are stored separately.

Both systems have a web interface and the access is through an ordinary browser.

Table 1 below includes all metadata that are captured when a document is created. There is no distinction made between obligated metadata and optional ones. The metadata is sorted into subject groups to increase the readability. The metadata that are similar in both organizations are marked with light grey. In some cases the name of the metadata is not exactly the same, but they are defined in the same way by the organization and only named differently.

Table 1. Captured metadata

Organization A	Organization B
Date and time	
Date of filing	
Date of issue	Date
Date of publication	
Date of sorting out	
Displayed from year-year	
Displayed until further	
Valid from	Valid from
Valid until	Valid until
Valid until further	
Year	
Year of assessment	
Author	
Author	Administrator
Categorization	
	Document level
Document type	Document category
Information type	
Revision control	
Replaced by document	Replaced by document
Revision	Revision
Links and associations	
District	

Field of action	Process
Recipient	
Responsible section	
Sub district	
	Sub process
Working divisions	Aimed at
Others	
According to law	
Action required	
Publicized at extranet	
Registration number	
Secrecy	
SFS-number	
State	
Take action latest	
	Document number
	Environment
	Established by
	Examined by
	Quality
	Road safety
	Title
	Working environment
29	20

Organization A has chosen to capture 29 metadata elements when a user creates a document. The underlying reasons for the choice of those metadata elements are several. The organization's need was of course the main focus but they were also influenced by the Dublin Core metadata set. Dublin Core is a metadata set developed in information and library science and was originally made for mark up of web pages. In Organization A an archivist participated in the requirement phase and thereby had the opportunity to influence the metadata that were chosen. The metadata elements 'date of filing' and 'date of sorting out' is evidence of these requirements. There is an ongoing project in Organization A to capture documents to an electronic archive. The plan is that different information systems shall deliver documents or information in other forms to the electronic archive after a specified period of time, based on the date of filing. The employees in Organization A report that the system is difficult to use.

Organization B has chosen to capture 20 metadata elements. In Organization B it was business needs, an earlier system and a certification process that influenced which metadata to capture. Examples of metadata elements that originate from the certification process are: working environment, environment, road safety and quality. The users of the system reported satisfaction with the system and the most common opinion is that the system is intuitive and easy to use.

Matching of metadata

The mapping was done by comparison of metadata elements found in the empirical study and metadata in three selected metadata specifications. A match generated a 'yes' in an Excel sheet and no match was marked with a 'no'. In some cases it was not easy to decide whether there was a full match or not. First of all there was problem from a language point of view. The empirical study was conducted in Swedish and the metadata set is all presented in English. The table also includes the exact names of the corresponding metadata element.

Table two provides the completed metadata map. The metadata are listed in the same order as in table one (subject groups). The symbol x represents were there is a match between an empirically found metadata element and an element from one of the metadata specification sets.

Table 2 Matching of organizations metadata and metadata sets

Organization A	VERS	RKMS	ISAD(G)	Organization B	VERS	RKMS	ISAD(G)
Date and time				**Date and time**			
Date of filing	x	x	x				
Date of issue	x	x	x	Date	x	x	x
Date of publication	x	x	x				
Date of sorting out	x	x					
Displayed from year-year	x	x					
Displayed until further							
Valid from	x	x	x	Valid from	x	x	x
Valid until	x	x	x	Valid until	x	x	x
Valid until further							
Year	x	x					
Year of assessment							
Author				**Author**			
Author	x	x	x	Administrator	x	x	x
Categorization				**Categorization**			
				Document level			
Document type	x	x	x	Document category	x	x	x
Information type							
Revision control				**Revision control**			
Replaced by document	x			Replaced by document	x		

Revision				Revision			
Links and associations				**Links and associations**			
				District	x		
Field of action	x	x		Process	x	x	
Recipient	x	x					
Responsible section	x	x	x				
Sub district							
Working divisions	x	x		Sub process	x	x	
				Aimed at	x	x	
Others				**Others**			
According to law							
Action required	x						
Publicized at extranet							
Registration number							
Secrecy	x	x	x				
SFS-number							
State							
Take action latest							
				Document number	x	x	x
				Environment			
				Established by		x	x
				Examined by		x	x
				Quality			
				Road safety			

				Title	x	x	x
				Working environment			

Comments and remarks

Nine of the metadata from the empirical study are the same in both organizations. Of those nine, five have been matched against all selected sets within this study. Those five are: *date, valid from, valid until, author and document type*. Those metadata elements form a base set of metadata that is valid for this study. Concerning metadata that describes processes there is full match between the empirical data and VERS and RKMS. ISAD(G) does not include any metadata elements that describe the connections between information and process. Of the remaining metadata that is similar there is one that is not included in any of the studied metadata sets. This is about *revision*. The metadata *replaced by document* is only included in VERS. There are also examples of metadata elements that are not included in any set, for example, *according to law, registration number, state, environment and quality*. All of the sets are built to allow addition of own metadata so in a way it is possible to say that any metadata could be part of the set. That is, the metadata sets are extensible to meet the specialized needs of organizations.

More than half of the metadata elements that the organizations capture are suggested in VERS and RKMS. The matching against ISAD(G) shows that the similarities are less than half and in

Organization A only a third. So the conclusion is that VERS and RKMS, representing the business process approach, have a higher match rate than ISAD(G), which represents the retrospective approach.

DISCUSSION

Differences between organization A and B

This research shows some differences between organization A and organization B. An archivist participated in the metadata design in organization A. Metadata like *date of filing* and *date of sorting* out are signs of this. The Dublin Core metadata standard also influenced them during this phase. RKMS was developed to include other schemas and Dublin Core is one schema that it would be possible to include (Day, 2003a). So, the choice of Dublin Core does not exclude using a schema that has its origin within record keeping. These are two reasons that explain why organization A has chosen to capture a larger quantity of metadata. Organization B has however been influenced by a certification process. *Environment* and *quality* are examples of metadata that originate from this process. This research has shown that both organization A and B have been aware of how important metadata is and their choice is not made by coincidence. But there are a lot of examples of metadata specifications with roots in document management or record keeping but neither of the organizations has chosen to use them in their work. Those specifications arise from a lot of earlier experiences, which could have

been helpful. Organization A seems to have been satisfied with the fact that they looked at a specification (Dublin Core) and did not think that it was necessary to investigate further. Organization B seems to not have considered looking at any other specification. Their work with selection of metadata functioned well for their current needs and they did not see any reason to look at some standards. Another possible explanation is that both organizations have chosen to capture over 20 metadata elements each and maybe thought that it will be hard to motivate more metadata. At least this is the case when the organizations have chosen to work with manually capturing of the metadata. Both VERS and RKMS and research within this field do generally recommend automatically capturing of metadata. If the organizations have chosen to work with more automatically capturing of metadata they would not have had problems to motivate their employee to fill in metadata. Another risk with manually added metadata is of course that the employees interpret the metadata and how they should be filled in differently. Another risk is that users may fill in the metadata fields incorrectly because they think that it takes too much of their time. The Clever Recordkeeping Metadata project that is on going has pointed at this problem. They state that metadata should be captured only once and that it should be done automatically. According to the homepage of the project (McKemmish et. al, 2006) the problem is that in practice there are significant implementation problems. And the development of metatools for the automatic creation of metadata and the translation of metadata values has not kept pace with the theoretical advances and standards development. This is has also been shown in this study of two Swedish organizations which gives them problems with the reliability

of their metadata. The theoretical discussion about how metadata should be captured and managed is not the way it has been implemented in those organizations. This will give those organization problems with their information over time.

There are also examples of metadata elements that were not represented in any of the analyzed sets. There will always be a need for enterprise-specific metadata to extend standard metadata sets. If the organizations had chosen to use any of the sets that have been analyzed in this study this could have been solved easily. All of the sets have the opportunity to include an organization's own metadata elements. Something that has to be further investigated is the organizations' use of different words for the same thing, which could have an impact on the matching of metadata elements. With a development process striving for provision of electronic services it will be important that agencies can exchange information with each other. This increasing need for exchange of information will also require that organizations follow the same types of standards. Because of this there are reasons to observe standardization project focus on interoperability.

A problem that needs to get more focus is how researchers and practitioners within this field can reach out with their results. There are a lot of experiences and research behind metadata specifications sets like VERS, RKMS and ISAD(G) and both organizations would have been helped if they had used them. It is wastefulness with resources if every single organization should start all over again when designing which metadata elements they should use in an electronic

document management system. The issue is partly to reach the individuals that are part of such design processes. They could be system designers and system engineers for example, but also archivists and records managers. Both the VERS and RKMS are developed in Australia and this study shows that the research results unfortunately have not been taken up in this part of Europe.

Differences between the organizations' metadata use and the metadata specifications

VERS, RKMS and ISAD(G) all include a lot of metadata that neither of the organizations has chosen to capture. It is not possible to capture some of those metadata elements as long as the document is active. There will always be metadata that has to be captured when the document has reached the status of inactive. But as pointed out by for example the Records Continuum Research Group (Recordkeeping Metadata Project, 2006) it is important to create metadata once (early) and use many times. The Clever Recordkeeping Metadata project at Monash University is one example of an ongoing metadata project. According to their homepage the goal is to "develop a proof of concept prototype to demonstrate how standards-compliant metadata can be created once in particular application environments, then used many times to meet a range of business purposes." It is well known that it is expensive to capture and preserve metadata. It may be argued that the cost will not be lower if it is done in an early stage, merely that the cost will be spread out. But there might be a risk that if the metadata were not captured during the creation it would not be prioritized when it is time to do it. Further on might it

also be impossible to catch all needed metadata after creation. There are lot examples of archives that are in bad shape because of the fact that metadata design and capture were not prioritized by organizations.

A fact that I would like to give some attention is why are there so many metadata elements in the sets that none of the organization has chosen to use? One thinkable explanation could be that the standards are too extensive. VERS for example includes 151 metadata elements. Not all of the metadata set elements are obligated and it is possible for an organization to add metadata and use the ones they want. The standards are made to suit different kinds of organizations and this is one explanation why the standards are so extensive. The standards have also maybe been developed to be a goal to strive for. The intentions with the standards may not be that all organization should use all metadata that is included in the set. On the other hand as pointed out before the problem might be to reach out with research result and for example standards. Which implies that the extensive nature of the standards is not the problem. A reflection from this study is though that at the first sight, none of the sets included in this study is very easy to adopt and understand. Even if one of the organizations had searched for an existing standard there might have been a lot of start up problems especially with VERS and RKMS that are more extensive than ISAD(G). I believe there is a real need for case studies of implementations of different kinds of metadata specifications. The example of the Clever Recordkeeping Metadata project is one way of doing it. A prototype is something that could easily be shown to a broader public.

Organizations like to participate in different kinds of conferences to tell the world how they solved a problem, which of course is one way of reaching out. However, organizations frequently do not reflect upon their own solutions in an objective manner. This research has done this; a reflection on two organizations' choice of metadata compared to existing solutions.

Evaluation of results

The first issue that has to be evaluated is the sources of possible errors in this research project. Deciding whether or not metadata elements match was not always easy. VERS, RKMS and ISAD(G) each have their own terminology. This means that they use different words to describe the same thing, or to describe what can at least be interpreted as the same thing. This may have had an impact on the result of this research but only in specific cases and not on a general level. To avoid these types of "errors" the uncertain cases were marked with a special sign and examined and interpreted once again. The metadata elements of the analyzed sets were all documented in detail, which of course was an important condition to be able to accomplish a study like this. One way to validate the result would be to make consolidations with the designers of the standards. If a consolidation like this would show that some of the metadata elements from the sets have been misinterpreted it could be used as input in the development of the descriptions of the sets.

The aim of this research has been to investigate differences between metadata use in electronic document management systems with metadata sets for recordkeeping and archival purposes. An important matter to discuss is whether the selection of research sites and metadata sets is representative. VERS, RKMS and ISAD(G) are sets that are made to suit different kinds of records and different kind of organizations and are therefore general. They are thereby satisfactory examples from both from a business process approach (VERS and RKMS) and retrospective approach (ISAD(G)). The research thereby shows the importance of which approach you want to implement when an organization wants to preserve information over time. If the metadata shall show connections between business processes and the information it is necessary to use a metadata set that has a proactive approach. To describe those types of connections will be impossible to do if the organization chose a metadata set with a retrospective approach. The general conclusion, that there are larger similarities between the empirical data and business process metadata set, would probably have been the same with other representative sets. When it comes to the metadata used in document management systems, organizations A and B comprise only a very small sample. Nevertheless, the method used has been fruitful and it has pointed out several new research questions.

CONCLUSIONS

This research shows that there are differences between metadata used in EDMS and metadata in metadata standards and specifications for recordkeeping and preservation. Recordkeeping metadata sets

that are designed from a business process perspective (VERS and RKMS) have larger similarities with the metadata elements that the organizations have chosen to capture in their document management systems than sets with a retrospective perspective (ISAD(G)). The result also indicates that processes and links between records and processes are a part of business processes sets (VERS, RKMS) but not of the retrospective set (ISAD(G)).

A result of this research is an identified set of metadata that was the same for both organizations and the three standard sets. Those metadata elements are: *date, valid from, valid until, author and document type.* Those metadata were identified in both of the studied organizations and they also were included in all analyzed standard metadata sets.

This research indicates that there is a need for further research to investigate which factors influence the choice of metadata elements. Examples of such factors may be: type of organization, legislative regulations, and differences between countries, what professional expertise takes part in the development process and usage of standards. An interesting research question is also how to influence organizations to use existing standards and experiences.

There is a lot of research needed to be able to design solutions for this problem. The method used in this paper could serve as a basis for further similar investigations. The method could be used for

organizations to make self-reflective evaluation of their choice of metadata compared to existing standards.

References

Recordkeeping Metadata Project. Accessed 2006-06-26. http://www.sims.monash.edu.au/research/rcrg/research/spirt/

Bantin, P., "Electronic Records Management - A Review of the Work of a Decade and a Reflection on Future Directions", in *Encyclopedia of Library and Information Science*, vol. 71, 2002, pp 47-81.

Cox, R. J., *Managing records as evidence and information*. Westport, Conn. ; London: Quorum Books, 2001. ISBN: 1567202314

Cunningham, A., "Six Degrees of Separation: Australian Metadata Initiatives and Their Relationships with International Standards", in *Archival science*, vol. 1, 2001, pp 271-283.

Day, M., "Integrating metadata schema registries with digital preservation systems to support interoperability: a proposal.", presented at *2003 Dublin Core Conference: Supporting Communities of Discourse and Practice - Metadata Research and Applications, Seattle, Wa, USA*, 2003a.

Day, M., "Preservation Metadata Initiatives: Practicality, Sustainability, and Interoperability", presented at *Metadata in preservation: selected papers from an ERPANET Seminar at the Archives School Marburg 3-5 September 2003*, 2003b.

Duff, W., "Evaluating Metadata on a Metalevel", in *Archival Science*, vol. 1, 2001, pp 285-294.

Duranti, L., "Concepts, Principles, and Methods for the Management of Electronic Records", in *The Information Society*, vol. 17, 2001, pp 271-279.

Eisenhardt, K. M., "Building theories from Case Study Research", in *Academy of Management Review*, vol. 14, 1989, pp 532-550.

Eriksson, D. M., *Four Proposals for Enterprise Modeling, in Industrial Management and Economics*. Göteborg: Chalmers University of Technology, 2004, pp.490

Gilliland-Swetland, A. J., "Setting the stage", in *Introduction to Metadata: Pathways to Digital Information Version 2.0*, M. Baca, Ed.: Getty Research Institute, 2000, pp. 1-12. ISBN: 0-89236-533-1.

Hedstrom, M., "Recordkeeping Metadata. Presenting the Results of a Working Meeting.", in *Archival Science*, vol. 1, 2001, pp 243-251.

Hofman, H., "Lost in cyberspace : where is the record?", in *The concept of record : report from the Second Stockholm Conference on Archival Science and the Concept of Record, 30-31 May 1996*, K. Abukhanfusa, Ed. Stockholm: Swedish National Archives, 1998 ISBN: 91-88366-35-9.

Hofman, H., *Metadata and the management of Current Records in Digital Form*. Accessed 2004-04-29.www.ica.org/old/biblio/com/cer/metadata.htm

International Council on Archives, *ISAD(G) : General International Standard Archival Description : adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19-22 September 1999*, 2. ed. Madrid: Subdirección General de Archivos Estatales, 2000. ISBN: 84-369-3404-0

International Standards Organization, *Information and Documentation and Records Management Part I: General*. Geneva: International Standards Organization, 2001

Lagoze, C., Lynch, C. A., and Ron Daniel, J., *The Warwick Framework A container Architecture for Aggregating Sets of Metadata*. Accessed 2006-05-04.http://www.ifla.org.sg/documents/libraries/cataloging/metadata/tr961593.pdf

McKemmish, S., Acland, G., Cumming, K., Reed, B., and Ward, N., "Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema", in *Archivaria*, vol. 48, 1999

McKemmish, S., Gilliland-Swetland, A. J., and Cunningham, A., *Clever Recordkeeping Metadata Project*. Accessed 2006-08-04. http://www.sims.monash.edu.au/research/rcrg/research/crm/index.html

Murphy, L. D., "Digital Document Metadata in Organizations: Roles, Analytical Approaches, and Future Research Directions", presented at *31st Hawaii International Conference on Systems Sciences, Hawaii, 1998*.

Myers, M. D. and Avison, D. E., *Qualitative research in information systems: A reader*. London: Sage, 2002. ISBN: 0-7619-6632-3

Public Record Office Victoria, P., *Victorian Electronic Records Strategy - Forever Digital*. Accessed 2006-2006-06-26.

Päivärinta, T., Tyrväinen, P., and Ylimäki, T., "Defining Organizational Document Metadata : A Case Beyond Standards", presented at *Proceedings of the Tenth European Conference on Information Systems, Gdansk, Poland, 2002*.

Reed, B., "Records", in *Archives: Recordkeeping in Society*, S. McKemmish, M. Piggott, R. Barbara, and F. Upward, Eds.: Quick Print, 2005 ISBN: 1876938846.

Shepherd, E. and West, V., "Are ISO 15489-1:2001 and ISAD(G) compatible? Part 1", in *Records Management Journal*, vol. 13, 2003, pp 9-23.

Sundgren, B. and Steneskog, G., "Information Systems for Concerted Actions", in *Exploring Patterns in Information Management - Concepts and Perspectives for Understanding IT-Related Change*, B. Sundgren, P. Mårtensson, M. Mähring, and K. Nilsson, Eds. Stockholm: The Economic Research Institute Stockholm School of Economics, 2003, pp. 356. ISBN: 91-7258-63-1.

Thomassen, T., "A First Introduction to Archival Science", in *Archival Science*, vol. 1, 2001, pp 373-385.

Upward, F., "Modelling the continuum as paradigm shift in recordkeeping and archiving processes, and beyond - a personal reflection.", in *Records Management Journal*, vol. November, 2000

Upward, F., "The records continuum", in *Archives: Recordkeeping in Society*, S. McKemmish, M. Piggott, R. Barbara, and F. Upward, Eds.: Quick Print, 2005, pp. 348. ISBN: 1-876938-84-6

Yin, R. K., *Case study research: Design and methods* (3. ed). Thousands Oaks California: Sage, 2003. ISBN: 0-7619-2552-X

Öberg, L.-M. and Borglund, E., "What are the Characteristics of Records?", in *International Journal of Public Information Systems*, 2006, pp 55-76.