

Traceable Information Systems

– Factors That Improve Traceability Between
Information and Processes Over Time

Lena-Maria Öberg



Mittuniversitetet

MID SWEDEN UNIVERSITY

Department of Information Technology and Media

Mid Sweden University

Lena-Maria Öberg

Department of Information Technology and Media

Mid Sweden University

E-MAIL: lena-maria.oberg@miun.se

Licenciate Thesis 27, 2007

ISBN: 978-91-85317-68-4

ISSN: 1652-8948

Printed by Mid Sweden University, Sundsvall, Sweden, 2007.

Abstract

Preservation of information is not a new issue but preservation of digital information has a relatively short history. Since the 60's when computers began to be used within administration, digital information that has had to be preserved over time. The problem addressed in this research is how to preserve understandable information over time. Information is context dependent, which means that without context it is not possible to use the information. Process is one part of the context. And an important issue when preserving information is then to be able to trace an information object to the process where in it has been created and managed. Associating information to a particular process creates the possibility of relating information objects to each other and also to the context in which the information has been created and used. The aim of this thesis is to identify and structure factors that can improve the traceability between information and processes over time. A set of factors based on case studies and a set of analytical methods are presented that can improve the traceability over time. These factors have been identified and structured by the use of the Synergy-4 model. They have been identified within four different spheres namely: competence, management, organization/procedure and technology. The factors have further been structured in three different time states namely: creation time, short and middle term and long-term. The research concludes that there are a lot of factors influencing ability to preserve information. Preservation issues include selection of metadata standards, organizational culture, lack of understanding from management and formalization of documents. The conclusion is that if an organization wants to succeed in preserving traceable information they have to build strategies that cover the issues from a range of different angles. This thesis suggests that crucial angles are competence, management, organization/procedure and technology. Furthermore, the strategies must be in place at the stage of creation of the information objects.

Keywords: Traceability, Preservation of Information, Processes, Context, Synergy-4

Acknowledgement

The writing of this thesis would never have been possible without funding from the European Union 'Objective 1 programme for the Södra Skogsläns region' (Funding the research projects Future of the Archives and Development of Archival and Information Science III) and Vinnova (Funding the research project The Home Building Guide).

First, I would like to thank all my supervisors at Mid Sweden University. I would like to thank Viveca Asproth for open and fruitful discussions, which also has included helping me seeing my strength and weak sides as Ph D Student. To Anita Håkansson I would like to say that you have a fabulous ability to see the best in things, which sometimes makes me feel like I am flying from our meetings. Patrik Wallin, your competence within information and archival field has been important.

I feel like the environment is of great importance to a Ph. D student and in my case I have had the pleasure to colleagues both in Östersund and Härnösand. The person that has been my nearest colleague during the work with this thesis is undoubtedly Erik Borglund. Without all his support, telephone calls, e-mails, meetings and co-authoring of papers, this thesis would have been so much harder to finish! Thanks also to all my colleagues at different campuses at Mid Sweden University for fruitful discussion and all the joy during for examples joint travels!

A special thanks to Stig C. Holmberg and Christina Amcoff Nyström for valuable comments during my pre-seminar.

I am especially grateful to Karen Anderson who has helped me with both the supervision and proofreading of three of my research papers and also the first summary section.

Erik Öberg, besides being the most wonderful husband, has also helped me with the design and layout of this thesis! Thank you!

In the Archives of the Future I was participating in reference meetings with a number of organizations and an academic advisory board. Those meetings have been important in the development of my academic communication capability. Here I would like to mention Tom Sahlén at the municipality of Sundsvall, who has a talent for putting tricky and fruitful questions.

Special thanks go to all the organizations that have participated in the empirical studies.

Finally, Maja and Folke you have not helped me writing this thesis, but you sure have helped me relax from it! This is not a simple task with a sometimes scatter-brained and stressed mom!

Frösön October 2006

Contents

- 1. Introduction** 1
 - Objective and Scope of the Thesis 2
 - Assumptions 2
 - Thesis structure 3
 - Included papers 3
 - Research Perspective 4
 - Research process 5
 - Descriptions of the studies 6
 - Study 1 6
 - Study 2 7
 - Study 3 7
 - Re-analysis model 8
- 2. Conceptual Framework** 9
 - Information - Records - Documents 9
 - Process 13
 - Traceability..... 14
- 3. Related Research** 16
- 4. The papers** 21
 - Paper 1 23
 - Problem definition in the paper 23
 - Purpose of the paper 23
 - Summary of the results 23
 - The paper’s contribution to the overall aim of this thesis 23
 - Paper 2 24
 - Problem definition in the paper 24
 - Purpose of the paper 24
 - Summary of the results 24
 - The paper’s contribution to the overall aim of this thesis 25

Paper 3	25
Problem definition in the paper	25
Purpose of the paper	25
Summary of the results	26
The paper's contribution to the overall aim of this thesis	26
Paper 4	27
Problem definition in the paper	27
Purpose of the paper	27
Summary of the results	28
The paper's contribution to the overall aim of this thesis	28
5. Results: The factors	29
Detailed Presentation of the Factors	32
Time state: Design/Creation	32
Time state: Short and Middle Term	35
Time state: Long Term	38
Discussion of the factors	39
6. Concluding remarks	42
Lessons learned	44
Methodological	44
Personal	45
7. References	48

The Papers

What are the Characteristics of Records?	53
Metadata use in document management systems	
which support business processes	83
To Capture and Preserve Documents in Processes	103
Needed Components to reconstruct processes	110

1. Introduction

Preservation of information is not a new issue but preservation of digital information has a relatively short history. Since the 60's when computers began to be used within administration, digital information that has had to be preserved over time. Our society gives a lot of example of information that is of interest over time. There are a lot of examples of information's value to society over time, both from public and private organizations. In the manufacturing industry for example it could be expensive to loose information like in drawings if the organization wants to rebuild part of their buildings. The problem addressed in this research is how to preserve understandable information over time. It is not necessary to preserve all information in an organization, the important thing is to preserve valuable information. Examples are information that serves as evidence or that is seen as important for business. Before describing the problem in more detail it is important to remark that this problem is tightly connected to physical access. Without physical access such as questions about file formats and hardware it would be impossible to use the information again. However this research focuses on other issues. An important issue in this problem is what happens to information over time. At the time when information is created in an organization in order to carry out their out their business, a lot of things are known. An e-mail with an attached file from someone in one's own division regarding a current task is easy to understand and use in the present time. But hard disk crashes, system failures, and switches of program versions are examples of technical occasions that cause information losses over time.

Besides losses because of technical problems, information is also context dependent, which means that without context it is not possible to use the information. Information always exists in a context. The context could be divided into a lot of parts. Examples include the author, dates for creation, date when the information was archived, whom it concerns, present legislation, organizational structure, other information that it is associated with it and so on. Documenting the context will ensure the possibility of interpreting preserved information. The importance of the context when preserving information over time is also stated by for example, (Guercio, 1997,; McKemmish, 2005,; Thomassen, 2001,; Upward & McKemmish, 2001). Guercio(1997) notes that an amount of information that should be preserved is always associated with another amount of information and all those associations gives us contextual information. The contextual information includes information about relationships between different parts of the information and also to what could be seen as that which has been surrounded the information. The contextual information gives possibilities to place a certain amount of information in a coherent whole. As described it is important to associate different information objects to each other

but also to the surrounding “world”. An important issue when preserving information is then to be able to trace an information object to the process (Bearman, 1994; Duranti, 1997; McKemmish, 2005; Reed, 2005) wherein it has been created and managed. Associating information to a process will create the possibility not only to associate information objects to each other but also make associations to the context within which the information has been created and managed within.

Objective and Scope of the Thesis

The aim of this thesis is to identify and structure factors that can improve the traceability between information and processes over time¹.

The expected result of this thesis is a structured set of factors that can improve traceability, which would support design and development of computer based information systems that manage information that is valuable over time. The result could also be used to identify weaknesses in existing solutions. The set will also be valuable in building awareness of factors in an information system environment that are important to consider in order to maintain traceability over time.

Assumptions

This research is based on the assumption that to be able to use in the future information created today an organization will have to make up strategies and solutions for that – today. This will henceforth be referred to as the pro-active approach. Computerized information systems that include information that has to be preserved for a long time has to be designed for that purpose. This assumption has a strong connection with how the concept time is used in this thesis. Time has been divided into three time states namely: creation time; short and middle term; and long-term. The creation period is seen as important in order to be able to preserve information in the long run. The short and middle term is chosen to give broader spectra between the two extreme states.

¹ Time has in this thesis been divided into three time states namely: creation time; short and middle term; and long-term.

Thesis structure

The thesis is a collection of four research papers where each paper contributes to the synthesized result presented in this first part of the thesis. The purpose of the first part of this thesis is to give the background to and introduce the main purpose of this thesis. The first part also includes a re-analysis of all empirical material.

The outline of the first part of this thesis is as follows. In this first section the research perspective is described which includes a description of the research process as well as description of studies conducted. In the next section the conceptual framework is presented. It includes discussions about information, records, documents, processes and finally traceability. The third section include a presentation of related research and in the fourth section the four research papers are described briefly. In the fifth section the result from the re-analysis is presented which means all the identified factors the section is ended with a discussion of the result. The sixth section includes concluding remarks and reflections of the methodological and the personal process.

The first research paper has been co-authored together with Erik Borglund² in which both authors have contributed equally.

Included papers

The following research papers are included in the thesis and are found in part 2.

1. Öberg, L-M. & Borglund, E. (2006). What are the Characteristics of Records? *International Journal of Public Information Systems*, 2006(1), 55-76.
2. Öberg, L-M (2007). Metadata use in document management systems, which support business processes. *Accepted for publication in Archives and Social Studies. A journal of interdisciplinary research.*
3. Öberg, L-M (2005). To Capture and Preserve Documents in Processes. In K.S. Soliman (Ed.), *Information Management in modern enterprise: Issues & solutions. Proceedings of the 2005 international business information management conference July 5-7, 2005, Lisbon, Portugal* (pp 181-183.): IBIMA.
4. Öberg, L-M (2006). Needed components to reconstruct processes. *In proceedings of Archiving 2006 conference May 23-26, 2006, Ottawa, Canada* (pp. 165-168): IS&T

² A doctoral candidate

Research Perspective

The underlying methodological starting point for this research is adopted from systems science. System science history goes back to the 1940s and 1950s and started as a reaction to reductionism. Over the years the theories and models within systems science has evolved. Nowadays researchers like Jackson(2003) try to categorize systems theory into three main theories, namely hard system thinking, soft system thinking and critical system thinking. This research has most in common with the two latter, which is shown in the argumentation below. Those arguments also give a view over why systems science has been chosen as the underlying starting point. The reasons are that it support the following:

- Within systems science it is pointed out that it is important to study a phenomenon within its environment. In this thesis this means that preservation of information is of no interest outside the organization that has created it and that the organization emphasizes the importance of its value. It can be argued that this makes the unit of analysis too complex but systems science has its foundation in studies of complexity(Jackson, 2003).
- Modern systems science (soft system science and critical system science) encourages studies on the same problem from different angles. In this thesis this is most visible in the use of the Synergy-4 model as a re-analysis tool. The system science approach do not limit themselves to defined borders of disciplines, meanwhile real-life problem seldom are limited to one particular discipline (Eriksson, 1998).
- The third and last point is that it is emphasized to use different method based on the type of problem that aims to be solved (Flood & Jackson, 1991). This has strongly influenced the three empirical studies that have been undertaken during the work with this thesis leading to the use of a number of different data collection techniques and analysis methods.

Research process

The results presented in this thesis are from a three-year research process. As described in the last section the underlying starting point is from the systems science. During a research process the researcher has to make a lot of decisions that regard how to fulfil the aim of the research. This section is about all those decisions. This section will guide the reader through decisions that have been made during the three-year process and why those decisions have been made. The decisions that are described have not been explained in the procedure as they turned out in reality. The description is simplified to be easier to read. Interdependencies and influences between different parts of this thesis (the papers and the studies) are described in the section about The Papers. Although the reader might get the impression that

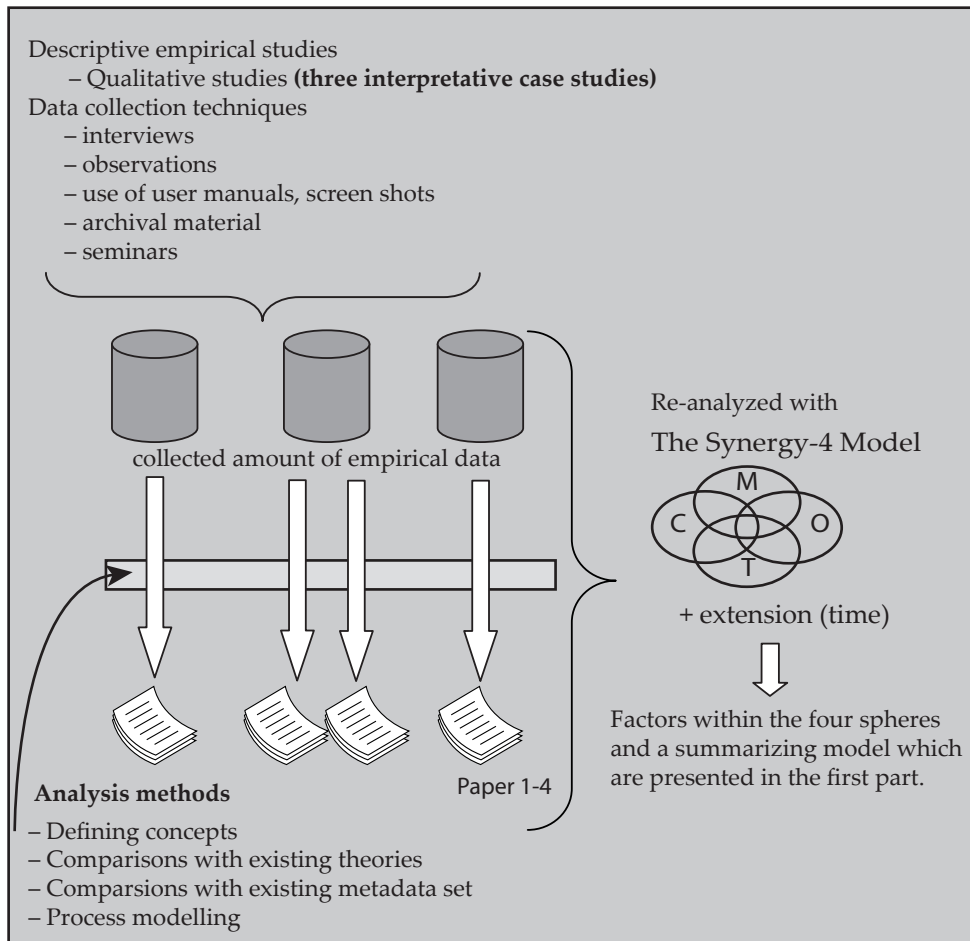


Figure 1: Overview of research process

all decisions about how to proceed with research had been decided from the very beginning, that is not the case.

The thesis includes four research papers and this first section. The research papers and the first section are based on three empirical studies. Early in the research process it was clear that it would involve descriptive empirical studies. The reason for this is that there were few examples of empirical studies within the information and archival science area, especially in Swedish settings. When it comes to the choice of qualitative studies and interpretative case studies it has been argued by Hartman [1998] that qualitative research can support such needs and also increase the knowledge of phenomena in their natural environment. Case study method is well suited for information systems research where it is necessary to study a phenomenon in reality and in its natural environment [Myers & Avison, 2002; Yin, 2003]. As recommended, different data collection techniques have been used in this research in order to capture the complexity and uniqueness of the settings of interest [Eisenhardt, 1989; Yin, 2003]. The empirical data has then been analyzed by use of different techniques such as defining concepts and comparisons with existing theories and process modelling. Those analysis techniques are described in detail in the research paper in which it has been used. The three empirical studies have been presented in four research papers.

Descriptions of the studies

Study I

The background for the study was a literature review within the archival science field, which generated a quantity of research questions. One of those questions was to describe the characteristics of records in real-life settings, and to answer this question a study was conducted in four organizations. The mode of procedure was a qualitative case study and we used interviews, observations and a seminar to collect data. The analysis aimed to identify potential characteristics of the records, a methodological step similar to what Orlikowski (1993) describes as defining concepts. When the identification of records characteristics was finished we compared the empirical based findings with archival theory. The results and the details about the method of the study are presented in research paper 1.

Study 2

The background to the second study was similar to the first one, namely that there were few examples available of empirical studies that could give information about the reality. But in the second study it was also possible to concentrate the study from records to a narrower focus on traceability. The second study was conducted in two organizations and aimed to investigate how those organizations used metadata in information systems (document management systems) to keep traceability between information and processes. The mode of procedure was a qualitative case study and the data collection methods used were interviews and access to screen shots, project documentation and manuals. The empirical data obtained (the metadata elements) were sorted and after that compared with three established metadata sets. The result of the comparison was presented in paper two. A second analysis of the collected data was then made and presented in paper three. This analysis was done by a discussion of similarities and dissimilarities between how the organizations have described their processes and how processes are defined theoretically. Details about the method are presented in research papers 2 and 3.

Study 3

The background of the third study was based on an insight of that reconstruction of processes could give valuable knowledge about how to build information systems that will make reconstructions possible in the future. The third study was conducted in two Swedish municipalities. The study aimed to investigate what kinds of components are needed to be able to follow the traces between information and processes when a case is finished. The mode of procedure was a qualitative case study and the data collection methods used were examination of archived material and seminars. By reading and interpretation of archived material reconstructions of the handling process were made. Both the process and the documents that were part of the process were analysed using the common process modelling technique. The process models were then validated in expert groups seminars, which aimed to fill in missing and/ or misinterpreted parts. The details of the methodology and the results from this study have been presented in research paper 4.

Re-analysis model

At the end of the research process, when summing up, it became clear that the results needed to be studied from a higher level of abstraction. Another level of abstraction would allow identification of the contribution from each research paper and also make the results less dependent of the participating organizations. The use of the analysis model has also been seen as an important tool for achieving the stated aim of this thesis, that is to be able to identify factors by a re-analysis of all the empirical data. It must be noted that the decision to use an analysis model was not made in the beginning of the research process. Rather, it was thought that it would be a helpful tool to force the required level of abstraction. The model was chosen from an own set of requirements of the model, which are described below.

The requirements for the re-analysis model were that:

- It encouraged reflections from different **perspectives**
- It was already **developed**.
- It was possible to adapt it for use in different time states.
- It should be **easy to learn** and thereby give results that can be easily understood and visualized by others.

The requirements that regarding the different perspectives was the most important in the choice of an analysis model. It could be said that this requirement does not limit the selection. But it was important that the perspectives together form a whole. There are a lot of important factors if an organization is to succeed in preserving information over time. From a research perspective there are several fields that must be involved in order to design durable solutions. The thinking concerning perspectives and the need to make studies and analysis from a holistic approach is based in systems science and therefore the model is taken from that field. The requirement that it should be an existing model and that it should be easy to learn was connected to the need of readability and clear communication of the results. The adaptability to be able to combine the chosen model with different time state was very important and a direct consequence of the aim of this thesis.

The model that best corresponded to the listed requirements was the Synergy-4 model which was developed by Holmberg (Holmberg, 2001). According to the Syn-

ergy-4 model a company or organization can be seen as a cooperation or synergy between four spheres. The four spheres are: competence, management, organization and procedures and technical infrastructure. The Synergy-4 model is described by Holmberg (2001) described as “approach and a methodology which aims at letting the employees develop the knowledge, skills and attitudes necessary for a balanced and coordinated development of their enterprises.” Development and improvement of organizations has often concentrated on just one of those spheres at a time. The interdependencies between the spheres have been neglected or excluded. The model thus strives to avoid these types of concentrations.

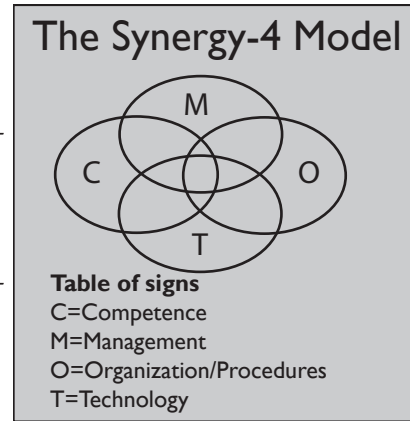


Figure 2: The Synergy-4 Model

2. Conceptual Framework

Information - Records - Documents

Throughout the work with this thesis it has become clear that information, that is one of this thesis' core concepts needs to be explained and clarified. The concepts of documents and records have also been used in this thesis. In this chapter, information, documents and records will be discussed and related to each other.

The kind of information that is of interest in this research is here defined as: information that belongs to an organizations is an asset that has a value over time. Information is an interpreted amount of data with a structure and with relationship to its context. Value to an organization can be defined differently depending on the organization; it could for example be about evidential or business advantage. Documents and records are here seen as subsets of information. This definition of information will be explained in the discussion below.

For someone that has the Scandinavian School of Informatics as background the most natural definition of information and data is that by Langefors (Langefors,

1995). Data is symbols that can be used to represent information and information is thereby interpreted data. Sundgren (Sundgren & Steneskog, 2003) who is one of Langefors students writes that it is important to realize that information can only be stored in human brains. Data can on the other hand be stored on external media. While this thesis concerns digital information, the term to use according to the above, should be data. The author agrees upon the reasoning about the distinction between data and information but in this coherence the term information gives a more significant view of the unit of the analysis in this thesis. When information is stored in a computer it is always cut into bits but is interpreted for users by different kinds of software application. This thesis has its focus at a logical level where for example, the interpretation of bits to a pdf-file is not of interest. To users and therefore the organization in which they are working, it is the information that they can read, use and understand. They are not aware of the “bits”.

In this thesis it is not any daily information flow that is interesting however. This thesis is concerned with information that is seen as important because it has value for an organization, over time. Within the emerging research field Enterprise Content Management (ECM) McKeen and Smith (2003) write about information assets and point out that organizations ought to manage those assets over their life-cycle. The formats of those assets include documents, forms, data, reports and web-pages. In this thesis different concepts has been used, in some cases document is the term that is used because that is what has been in focus in a certain paper and a certain study. In another paper information is the most appropriate word. But they should be seen as examples of the same thing, namely information that belongs to an organizations information asset that has a value over time. The ECM- related definition includes the term life-cycle so it appears that the definition used in this thesis and in the ECM-field is identical in the sense that they both include the variable time.

As mentioned earlier, information could be defined differently and used differently within different academic fields. Record is frequently used within the information and archival science and first of all, it should not be mixed up with the term used in database theory, where record is a tuple in a table (e.g(Connolly & Begg, 2005). The concept that is used in this thesis has its origin within archival science and recordkeeping. How to define ‘records’ is an essential issue within the archival science and there exists of course several parallel definitions. The Australian school of recordkeeping has been adopted in this thesis and therefore the definition from the ISO 15489 is the one that has been used in this thesis.

A record is “Information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in transaction of business.” (ISO 15489-1. International Standards Organization, 2001,p.3) An electronic record can be stored in one file as for example a word-file, but it could also be aggregated from multiple sources. ‘Record’ is here seen as a subset of information, a unique subset with unique characteristics. One example of this is that a record never may be changed. A record is always connected to a network of documentation, people and activities (Reed, 2005). Records are by definition context dependent and have a strong connection to transactions; indeed they are evidence of transactions. The transaction could may be a result of an activity (not all activities within an organization lead to information being created) and the activity is a part of a process³. By describing the connections between information and the process where it has been created and managed central part of the context is described. Bearman (1994,p. 54) described the importance of the process like this: “Without understanding the record system in relation to processes and activities of the organization ... it is not possible to identify what data in the system constitute evidence of an activity”.

In Sweden a record is often understood as the same thing as a public document, but the definition above shows that it could exist in all types of organizations not only public ones. According to McKemmish (2001,p.336) well-managed records can be “meaningful records for as long as they are of value to people, organizations and societies – whether that is for a nanosecond or millennia”. The idea that the value of the information is important, is well in line with the definition of information above. The boundary of one record may seem confusing. What is one record and what can be seen as several records? A record could even consist of several records. The general definition of a record is not enough in a practical sense and therefore it is up to each organization to clearly define when a new record should be created and which record should be part of other records. To make the concept understandable in a tangible way some examples from empirical data of records is presented below.

3 Process, activity and transaction are discussed more in detail later in this section.

Example 1; An application form from a citizen regarding the formation of a company.

Example 2; A building permit which includes the application form, the decision regarding the building permit, the examination report and letters to the agency.

Example 3: a technical manual.

The last concept to be discussed is the document. This term is a widely used concept to describe for example word or pdf-files. Probably this has its origin in the desk metaphor that is used within most common graphical user interfaces. Theoretically 'document' is a more complex term compared to the daily use of the concept as discussed above. According to Pédaque (2003) the view of document could be seen from three different perspectives. The three perspectives are form, sign and medium. Pédaque (2003,p. 9) defines from the first perspective (form) document as: "An electronic document is a data set organized in a stable structure associated with formatting rules to allow it to be read both by its designers and its readers." A document is thereby seen as a communication object. This view of a document has large similarities to the common view of document from the information system field. The tradition within the archival and information science is also a view of a document as a "thing" a material (see Buckland(1991) for a discussion about information as a thing) . But the view of a document as a sign includes the view of organizational aspects (Buckland, 1991) of documents. The organizational aspects emphasize that a document has to be arranged in a meaningful relation to other "evidence". This has strong connections to reasoning about context and the importance of relating information to the context in which it has been created. So, the view of a document is the same as that of many other researchers within information system science, "a data set organized in a stable structure" but to be valuable over time it is important to relate the information to its context. 'Document' and 'information' are thus used to describe the same thing, but in some coherences (like when dealing with document management systems) document is the common concept and to avoid misunderstanding the most common concept in that community has been used. And document is here seen as a subset of information.

To summarize the discussion above and as a reminder to the reader this thesis builds upon the following adoption of information.

The kind of information that is of interest in this research is part of an information that belongs to an organization's information asset because it has a value over time. Information is an interpreted amount of data with a structure and with relationship to its context. Value to an organization can be defined differently depending on the organization; it could be about evidence or business advantage. Documents and records are seen here as subsets of information

Process

The adopted definition of processes that is used in this thesis is originally from Le Moigne (Eriksson's translations is used here because the original source is only available in French). The definition of processes is tightly connected to the view of a general system, which is "something (a structure of components) that functions and transforms towards a goal in an environment" (Eriksson, 2004, p.280). A process is defined as: "a difference or change in states of space, time and form" (Eriksson, 2004, p.281). The most important difference (for this thesis) is that the general process theory (discussed in detail below) does not include the view of that implementation of new processes will lead to higher satisfaction or higher efficiency. The process view is used by Eriksson (2004) as one perspective to describe an organization. In this thesis the definition of processes by Le Moigne is adopted because it has a wider approach to processes. This is that processes exist in all organizations that consist of differences in states of space, time and form. Processes could thereby be compared to what could be explained as a procedure of how to solve a certain task. The procedure should be described and known by the person that is working with this certain task. Another difference to general process theory it is often emphasized that the processes should cross functional boundaries (see for example (Sundberg & Sandberg, 2006)). LeMoigne does not exclude this from his definition, but an underlying view within the systems science is a holistic view, which will also account for the view of processes. Within this thesis there will be empirical examples of both processes that do not cross organizational boundaries and those that do, but the adoption of processes is still the same.

The main idea within the research and practical use of process-oriented tools is called Business Process Reengineering (BPR). It has been a popular management tool since the early 1990s, and it has also been widely discussed and investigated within different research communities, see for example (Grant, 2002; Grover, Jeong, Kettinger, &

Teng, 1995). Katzenstein & Lerch (2000) describe the core in BPR as “organizational redesign that strives to improve coordination among people and other process entities to overall process goals more efficiently and effectively. “ A BPR project will by definition lead to a changed organization, the organization will become process oriented to some extent. Benefits of the use of BPR include higher productivity and greater customer satisfaction(Cao, Clarke, & Lehaney, 2001). Within the scope of BPR, processes should be seen as horizontal rather than the earlier hierarchical of functional view of organization. But it is also possible for an organization to use just parts of the BPR-concept, namely the modelling part. During the period of time that BPR has been used as a tool to change organization several methodologies have evolved. Vakola & Rezgui(2000) has criticized existing BPR methodologies and has also developed a generic model for BPR. The model is called CONDOR and consists of eight steps in an ongoing process. Another example of a model that makes suggestions for how to obtain comprehensive understanding of an organization is Eriksson(2004). The understanding of existing processes will give a good view of how an organization has accomplished some of their working tasks, which in turn give contextual information.

Concepts used in process modelling

Irrespective of which method is used to model processes there are some core concepts that are used within the process modelling. Activities and transactions are examples of such concepts which are also of relevance for this thesis. A process consists of activities and activities consist of transactions. When the transaction needs verification the transaction will lead to the creation of information. An activity is defined by Eriksson (Eriksson, 2004,p.283) as: “discrete actions that are conducted by one or several actors or processors within the system”. Within archival theory, ‘transaction’ is an important concept, and a record is sometimes described as transactional information (Bantin, 2002, Duranti, 2001, Thomassen, 2001) According to theory a record should be connected to a transaction to be able to serve as evidence of transactions.

Traceability

The two main concepts (information and process) of this thesis have now been discussed and the remaining part is to discuss the relationship between those two concepts. Throughout the process of writing this thesis different concepts have been used to describe this relationship. The final decision is to use traceability, which

like many other concepts is used in different contexts. Traceability here means that by associating a unit of information to the process where it has been created, the organization can reach information traceability. This creates the possibility of placing information in its context (Guercio, 1997;; McKemmish, 2005;; Thomassen, 2001;; Upward & McKemmish, 2001). As described in the section about processes, a process consists of activities (see for example(Eriksson, 2004;; Vakola & Rezgui, 2000) which in turn can consist of transactions. An example of where traceability is used

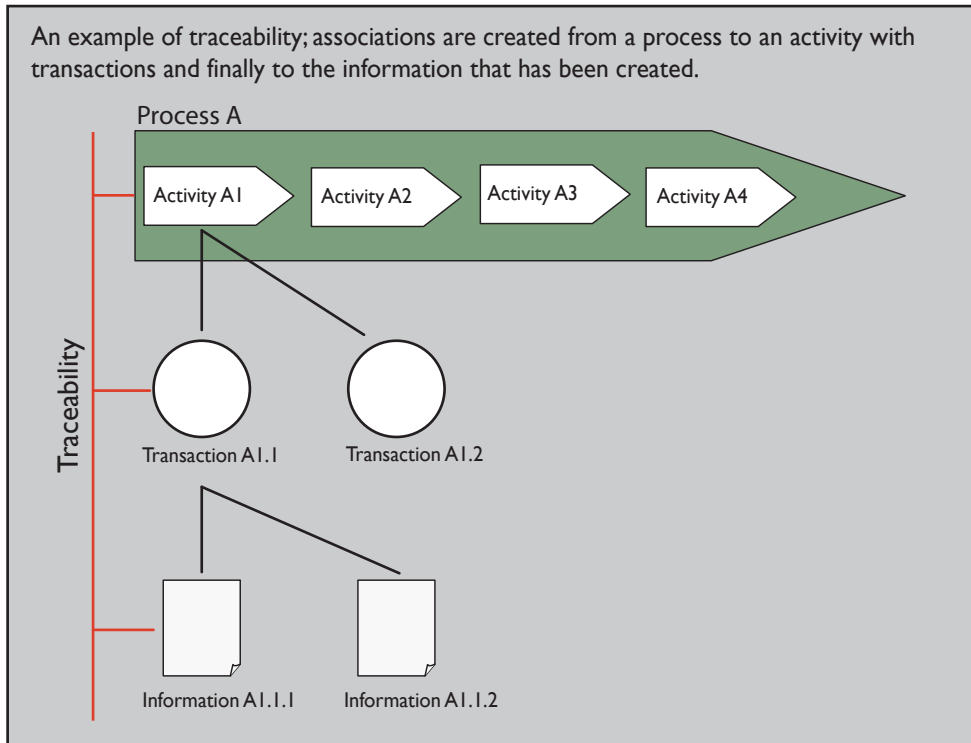


Figure 3: Example of Traceability

with the same meaning is an area of research that focuses on how to create traceability links between program source code and system documentation (see for example (Antoniol, Canfora, Casazza, De Lucia & Merlo, 2002; Marcus & Maletic, 2003). The difference here is between which parts the links or associations are made.

To reach traceability on the highest level the transaction then has to be associated with the information that has been created within that transaction. This is an ideal design of traceability on the highest level. Throughout the studies examples from

the participating organizations have showed that the associations between information and processes can be implemented in different manners and on different levels (Öberg, 2005).

The figure above shows the relationship between process, activity, transaction and information, which give traceability. The relationship could be a two-way association or single-way association, dependent on the need of the organization and the technical solution.

So far the discussion about traceability has been theoretical and a gives an idealistic view of implementation. Practical implication is here seen as important to show how and why traceability is important when an organization wants to preserve information over time. The first example is a fictive example from authority x that received a huge amount of appeals against their decisions every year. It is of great importance to their internal lawyers to be able to follow the handling process of the public servant after the event and identify which information has been used to make the decision. Another fictive example is a citizen who can follow his building permits process through a Webb interface. The citizen may need to use this kind of application while the process is ongoing but also to get the information again several years later. Both the examples require that the organization has control over its information and the processes it belongs to. Another example is collected from Landgren (2007) who states the importance of digital traces within the community of emergency response. According to his conclusions digital traces (from for an example a computer) can give valuable information about ongoing work for example in a rescue action.

3. Related Research

The overall problem that is in focus in this thesis is how to be able to preserve electronic information over time. This problem includes problems in several domains for example technological, organizational and legal ones. Asproth(2005) has made a survey of ongoing research projects within long-term preservation. The author argues that even though there are a lot of examples of problems that are of non-technological character most research concentrates on technical issues. Gilliland and McKemmish (2004) have made an identification of major and emergent areas of archival research between 1995 and 2005 and the list includes 19 areas. This list also includes both technical areas but also educational and organizational ones. According to Gilliland and McKemmish(2004) the development of archival research in the past 15

years is unique. This is obvious looking at the list of areas presented by the authors. They also argue that in most of the areas there are examples of approaches on three levels, building, evaluation and reflecting. The Gilliland and McKemmish(2004) article itself show signs of a maturing process where two researchers try to categorize both areas within the field and what kind of methods are used to accomplish research. This is an example of a reflecting activity.

This section about related research will henceforth concentrate on research that in some way is related to the research aim in this thesis. The exposition includes references from researchers that belong to both archival and information science and to information system research. The criteria that have been used in this literature search are; context, metadata, document management system, processes, document and record.

The Records Continuum Research Group (RCRG) at Monash University in Australia is an active research group within recordkeeping and archiving. According to their homepage⁴ they are: “looking at methods of analysis, which enable records to be controlled at different points in time throughout their lifespan”. They had accomplished several research projects (two of them are presented later in this section) since the start 1997. One important outcome of their research is the continuum model. Frank Upward has developed the model (Upward, 2000,; 2005). The continuum model consists of set of theories around records and recordkeeping and all will not be discussed in detail here. But according to Upward (2005) the characteristic of the model is the view of records as unstable. The records continuum model also emphasizes the pro-active approach which means that to be able to use information created today in the long run an organization has to design strategies and solutions for that – today. The context is seen as an important instrument to reach usable preserved information. McKemmish (2001,p.335) writes that the theory building and modeling has included development of conceptual models for describing records in their “societal, business, and documentary contexts. Another foundation concept in the continuum model is the view that “a record exists in a network of documentation, people and business activities” (Reed, 2005,p.111) Business activities, which in practical means forms processes is of course relevant to this research.

InterPARES (International research on Permanent Authentic Records in Electronic

4 <http://www.sims.monash.edu.au/research/rcrg/>

Systems 1 and InterPARES 2) is the most extensive project within this area⁵. The projects are extensive both in their aims and the project group (there are researchers from Asia, Australia, Europe, Canada and America). InterPARES 1 ended in 2001 and the main findings are published in a book and in several research papers. According to the conclusions in the book one part of the project developed a conceptual framework for establishing the requirements for preserving authentic electronic records. The project strove to create a single comprehensive typology of authenticity requirements for electronic records but did not succeed. InterPARES 1 focused on authenticity of records produced and/or maintained in databases and document management systems. InterPARES 1 has made delimitations in focusing on assessing and maintaining the authenticity of electronic records once they become inactive and are selected for permanent preservation. (Duranti (red), 2002;; Duranti & Thibodeau, 2006;; InterPares I and II, 2007). InterPARES 2 was initiated in 2002 and concluded in 2006. According to the project webpage the aim of the project is: "to develop and articulate the concepts, principles, criteria and methods that can ensure the creation and maintenance of accurate and reliable records and the long-term preservation of authentic records...". InterPARES 2 has focused on records produced in complex digital environments such as artistic, scientific and e-government activities. The concept of electronic records that was adopted by InterPARES 1 (included characteristics, parts, formal elements, attributes and digital components) has, according to Duranti and Thibodeau (2006), worked quite well with databases and document management systems, but will be problematic when dealing with records that have more fluidity. Both the InterPARES project and this thesis focus on preservation for a long time span, which includes information from the time it is created to the preservation. The focus within InterPARES 2 studying records in a complex digital environment is interesting but this kind of information has not been included in this thesis.

The overall aims of the InterPARES project are general (developing concepts, principles etc) with delimitation that has been discussed above. This thesis deals with context and within this area there are examples of more specialized research. The reason for this is that the widespread use of computerized information systems creates a need for conceptualization of metadata (Gilliland-Swetland, 2000). 'Metadata' is here used in the way it is defined within the field of archival and information science and should be seen as something wider than what kind of data is stored in a table in a specific database. Gilliland-Swetland(2000,p.1) defines it as: "metadata is the sum total of what one can say about any information object at any level of ag-

5 www.interpares.org

gregation”.

A group of researcher at Monash University have several ongoing projects within this area. The Recordkeeping metadata project resulted in a metadata schema (RKMS – the Australian Recordkeeping Metadata Schema)(McKemmish, Acland, Cumming, Reed, & Ward, 1999). The Clever Recordkeeping Metadata project is another example and it addresses problems around automating of metadata creation and sharing of metadata between business systems, current recordkeeping systems and archival systems (Evans, McKemmish, & Bhoday, 2005). There are also other examples of metadata schemas. See for example the metadata schema that is part of VERS (2006) (Victorian Electronic Records Strategy). Evans et al.(2005) argue that there is now a need for integration and translation between different metadata schemas. In paper two three existing metadata schemas have been used to analyze two Swedish organizations’ use of metadata. Paper two also includes details about the different metadata schemas. There are also other examples of comparison between different metadata sets. See for example (Cunningham, 2001,; Duff, 2001,; Shepherd & West, 2003). The research regarding metadata shows a natural progression of maturity when this kind of reflection upon existing research findings is made. Gilliland and McKemmish(2004) make the same conclusion in their paper. They point out that evaluation and comparison of conflicting conceptual models and descriptive schemas is an emergent area of research within the field. Research within the field of metadata also includes other issues. First of all there is a time perspective, when should it be done? The opinions are diverse in this question but my basic view is common with that of Hofman who argues that a pro-active approach is essential to be able to preserve (at reasonable cost) accessible and interpretable information for the long-term (Hofman, 1998) ,which means that metadata should be captured as early as possible. Reed(2005) notes that layers of metadata could be added at various times because they cannot be defined at the point of capturing. A second issue concerns how to add metadata, which can be done manually or automatically. There is research focusing on solutions to capture metadata automatically (Evans et al., 2005) but there are few examples of organizations that have come to such implementations.

So far, this presentation has only included research that has its origin within the academic field of archival science, but there are also examples of relevant research within information systems science.

At the University of Jyväskylä there is a research program called Document Management Research. Within this theme several research projects have been conducted from 1997 onwards. The general aim of the program is "to develop methods and techniques for the management of digital documents in the networked multimedia environments of enterprises."⁶ The research within this program includes several standardisation projects using SGLM and XML as techniques. In the first project called, RASKE, a Finnish acronym, the research settings were the Finnish Parliament and Government. The studies are interesting partly because of the chosen research sites. This kind of organization creates a lot of documents that have a value over time, and it is therefore important to design information systems that can handle this situation. Salminen, Lyytikäinen & Tiitinen (2000,p.623) shows awareness of this in arguing that "document standardization does not concern documents only. It concerns workers, their work, business partners and future systems as well". Besides dealing with techniques for standardisation the research project has also dealt with document analysis. Their document analysis method includes two tools for describing the work context of documents namely process modelling and life cycle modelling. Their work has shown that process-modelling techniques (in their case amongst other techniques) is useful when an organization wants to describe the context of documents. In this specific case the models are used as a base for standardization of documents. The research is however concentrated on standardization as a method for usable documents both when it comes to daily use and long-term use. Salminen et al.(2000) have not studied how to describe the processes themselves or how to preserve the traceability between the documents and processes. Päiväranta, Tyrväinen & Ylimäki (2002) have, on the other hand studied how to define 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. Nevertheless their study does not include any connections to processes and does not discuss how to manage metadata over time.

All the presented research has made delimitations in different manners. The delimitation is sometimes the unit of study (as for example with Interpares), in other cases the limitations are made based on the time horizon (for example long-term preservation). The varying limitations make it difficult to compare different research projects with each other, but they must be put in relation to the work of this thesis. There is one example of research that is based on the same concept about time as this thesis is. This concept is that the design and creation phase is of great importance for the

⁶ <http://www.cs.jyu.fi/~airi/index.html>.

ability to preserve information in the long run. The research made by the Records Continuum Research Group (RCRG) is in this sense close to the research in this thesis. Although the RCRG has not had the same aim as this research, when it comes to metadata the relationships between this thesis and their work are strong. The details of these relationships are discussed in research paper two. However when it comes to InterPares 1 and 2 projects they have another view when it comes to time. They focus on the preservation part without any consideration of the design and creation phase. The Document Management Research group from Finland has focused in their research on how to design document management systems where long-term preservation certainly is mentioned but not dealt with. This thesis thereby fulfil a need for research that has this wide perspective of how time will influences the possibilities for preserving information and also a need for research focusing on traceability with this perspective.

4. The papers

As mentioned before, this thesis consists of one first introduction part followed by four research papers. The research papers have of course all contributed to the final result and in this section it is discussed how the papers are related to each other and how they have contributed to the final result. The contribution can be seen as a two-stage rocket. The first stage of the contribution is the results from each research paper and the second stage is the result from the re-analysis. In the end of this section all papers are described in detail, the figure below gives an overview of the contribution and the relationships between the four research papers. As shown in the figure the papers have contributed to each other with a deeper understanding of central concepts.

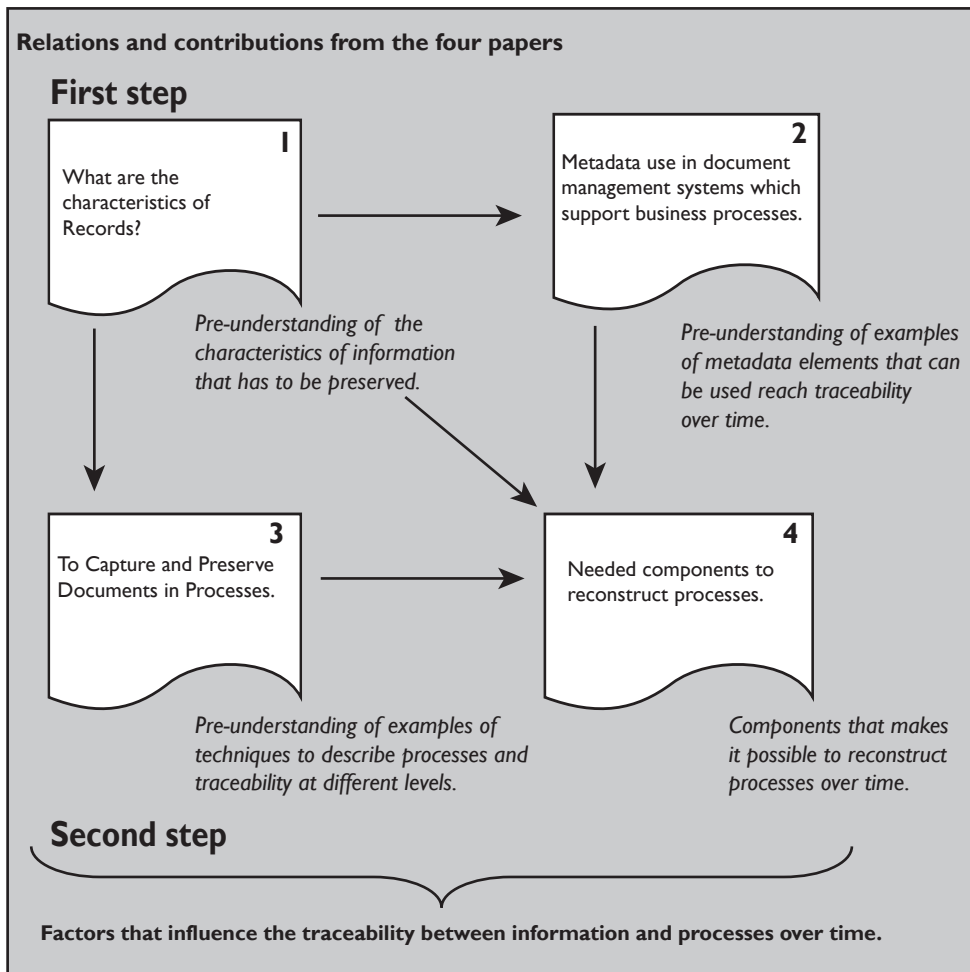


Figure 4: Relations and contributions from the four research papers

Each paper is summarized below and described under the following headings.

- Problem definition in the paper
- Purpose of the paper
- Summary of the results
- The paper's contribution to the overall aim of this thesis

Paper I

What are the characteristics of records? (Öberg & Borglund, 2006)

Problem definition in the paper

In theory, records are a subset of information and the definitions of records available are rather abstract and difficult to use as help in identification of records. The problem is that the theory does not to an acceptable extent explain what is seen as records in real-life settings. What kind of information is seen as records in an organization and what are the characteristics of those records? Identification of what are the characteristics of records is necessary if one aims to formalize records, which is necessary to enable automatic processing in a computer-based information system.

Purpose of the paper

The aim of the paper was to describe the characteristics of records based on both archival theory and empirical data. The research questions within the paper were:

- What is defined as records in organizations?
- What are the characteristics of these records?
- Are there any differences between the empirically grounded characteristics of a record and the characteristics described by recordkeeping and archival theory?

Summary of the results

The paper presents a list of several different characteristics found in four organizations. The paper also shows that there is a large variety in what organizations define as records. It also shows some dissimilarity between theories and empirically identified records. Legislation, other organization-dependent regulations, archival tradition, and different recordkeeping traditions are possible reasons for those dissimilarities. The paper presents examples of a large variation in recordkeeping and information management maturity between organizations.

The paper's contribution to the overall aim of this thesis

The paper has contributed in pre-understanding of the characteristics of information that have to be preserved. This knowledge has been essential to the following studies. Besides this it has also contributed a set of factors that could improve the traceability of information over time. Those factors are presented later in this section.

Paper 2

Metadata use in document management systems, which support business processes (Öberg, Accepted for publication 2007)

Problem definition in the paper

The importance of metadata for describing 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. 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 is an important issue. A pro-active approach is essential to be able to preserve (at reasonable cost) accessible and interpretable information for the long-term. The pro-active approach taken here is that metadata ought to be captured during the creation phase. However the standard procedure in Sweden is that documents are stored and managed in document management system as long as the document is in active use. When the document is no longer of immediate interest it is archived. The problem is that the integration of metadata use in document management systems and record keeping systems is not high enough.

Purpose of the paper

The aim of the paper was to study metadata use in document management systems and compare it with metadata that is recommended for recordkeeping in order to reveal differences.

The research questions in this article were:

- 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?

Summary of the results

The research shows that there are differences between metadata used in Electronic Document Management Systems 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 simi-

larities 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 associations between records and processes are a part of business processes sets (VERS, RKMS) but not of the retrospective set (ISAD(G)).

The paper's contribution to the overall aim of this thesis

The paper has contributed in deeper insights about the way choices of metadata and metadata standards influence how traceability can be reached. The paper has also contributed an understanding of the factors that influence choices of metadata, like organizational culture and the participants in the development projects. Finally, the second paper has contributed with a subset of the factors that are presented later in this section.

Paper 3

To Capture and Preserve Documents in Processes. (Öberg, 2005)

Problem definition in the paper

An organization undergoing transforming into a process-oriented organization may also need to change their management of information. Some of the computerised information systems that are used need to support the processes in order to give the organization continuity in their use of processes. By associating documents to the processes in which it has been created and changed an organization will have the ability to follow the information flows afterwards. The starting point for this paper is that a pro-active approach is essential when an organization wants to preserve information. This means that the information has to be prepared and captured for long-term preservation at the creation stage, otherwise it will be costly and difficult. The problem is however to know **how to** associate documents to a process and also how to describe this relationship so it will be understandable over time.

Purpose of the paper

The focus in paper three is what must to be captured in description of the processes and the associations between documents and information to reach traceability.

Summary of the results

The main result from paper three is that the organizations within the study showed two different examples of implementation of associations between documents and processes. It has been done at completely different levels (Organization A has associated documents to the head process that it belongs to and Organization B has associated the documents both to the head process and to the subprocess) but none of the organizations has reached traceability between documents and activities. One organization has described the processes. In the first empirical study the focus was on characteristics on preserved information and result indicates fragmented trace of the process. It is worth noting that if the process does not follow the normal procedure there are more traces to follow. In such cases codes are noted and copies of letters that describe what a customer needs to do to are preserved. Those organizations that capture this type of metadata have a chance of succeeding. But they have to supplement this with descriptions of the “normal” process. This could be implemented by making instances of a process only if the process does not following the normal case.

The paper’s contribution to the overall aim of this thesis

Besides contributing a number of factors the paper also provides examples of techniques to describe processes and traceability at different levels.

Paper 4

Needed Components to reconstruct processes. (Öberg, 2006)

Problem definition in the paper

The Freedom of the Press Act is part of the Swedish constitution. This gives every citizen rights of access to public documents. In Sweden all documents that are created and received in a public authority are public unless a classification of 'secret' is applied. Swedish systems also include openness; which makes traceability an important matter for the public authorities. Traceability makes it possible to follow processes such as a decision making process. Nowadays with a 24/7 e-service the intention is that citizens should be able to follow their on-going processes and access public documents by themselves. In Sweden there is an on-going research project called "Home Building Guide" (Bygga Villa in Swedish) the goal of which is to build an integrated e-service. This integrated e-service will cross agency borders in order to make it easier for a citizen to seek information, for example about a property unit purchase or to follow the different processes that are needed to build a house. Today a lot of public documents are electronic and set in different kinds of computer-based information systems. Examples of such systems are: document management systems; record keeping systems and workflow systems or combinations of these. The problem is, however, that the requirements for designing such an information system that takes account of traceability of processes and general requirements for record keeping over time under Swedish conditions are not known. However in the archival community the importance of tracking information processes has been known long before the process orientation era. Duranti (1997) has described records as associations in a chain, and Bearman (1994) even states that it is impossible to understand records without relation to processes and activities. There are several examples of standardization processes that include traceability between processes and documents/records. Examples are; ISO 9001 and ISO 23081-1(ISO 23081-1. International Standards Organization, 2006). Another example is provided by McDonald (1997) whose prototype made it possible to connect and track records in processes. Nevertheless to be able to build an integrated e-service like the "Home Building Guide" there is a need for studies that describes what kind of components there are used in Swedish archives of today and what kind of needed components are missing.

Purpose of the paper

The aim of paper four was to describe what kinds of components are currently in

archives that make it possible to reconstruct processes.

Summary of the results

This study has presented three different categories of components that make it possible to follow processes namely, context, documents and date. The study has shown different ways of implementing those components and there are of course further alternatives.

A high degree of openness is important for successful e-services and the study revealed some problems. The process that is preserved in the archive shows the municipalities' internal processes. The "Home Building Guide" and similar integrated e-services will place the citizen inside the process. This study does not reveal whether a citizen thinks that the preserved process gives a fair picture of what has happened. There are for example few components of dialogue with the applicant during the process of a building permit. With an e-service communication with the citizen can change into new ways. The communication will often be recorded by e-mail or by forms. This raises a need for discussion about appraisal of what has to be preserved or not. There are some audit trails that connect to other agencies' work (for example, consultations with the environmental health committee) but to build integrated, effective e-services the archives must also be more integrated with each other. This will include both co operators within the Municipalities and external actors if the "Home Building Guide" vision should become reality. How this should be solved is anything but straightforward. Questions that have to be solved are for example; where processes cross agency boundaries, who is responsible for the documents; who captures and preserves the documents and how the context of the documents is maintained and how documents should be exchanged. Another problem that has been revealed during this study concerns the process borders. When the building permit process is finished maybe a new sub process starts? Or some sub processes may have preceded the building permit process, or be contained within it. How and where should those components of traceability be preserved?

The paper's contribution to the overall aim of this thesis

The fourth paper has contributed knowledge about what components are needed to be able to interpret a process afterwards, by use of the information that is preserved. In addition the fourth paper has contributed by identifying factors that influence traceability over time.

5. Results: The factors

The analysis model has been presented in an earlier session but the practical work has not yet been described. The first step was to read through all empirical data and all research papers carefully. The papers aimed to identify factors that could influence traceability over time, and the analysis model managed this process in the sense, that it set out the type of factors (the four spheres) that the process aimed to identify. The factor identification part was done iteratively to be able to identify factors in all time states. Many of the factors have been clearly stated in the research papers but some were tacit. There were also some examples of factors that were not presented in any of the four papers but nevertheless were identified in one of the three empirical studies. This identification process was done parallel with the structuring part. An identified factor was immediately sorted into one of the four spheres (competence, management, organization/procedures and technology). Some of the factors have also been identified within several of the time states. The structuring process was also iterative in the sense that some factors shifted sphere after consideration. Some of the factors have also been considered as borderline which means that they might fit into a combination of two spheres. But after further consideration all factors were located in one sphere at this stage. A possibility for development of the model is to focus on these borderline factors more closely.

The process of identifying and structuring the factors included validation. The validation took place in an expert group where all factors were presented and discussed with the participators. The expert group was composed of seven archivists, one security manager, one Ph D student in political science and one document manager. The validation discussion focused on the structuring of factors and the analysis model. The discussion about the factors resulted in rearrangement of the factors and the comments about the analysis model is presented in a later section in this thesis (Lessons Learned in Concluding Remarks).

c Lack of competence ●○○○	t Traceability on normal process respectively alternative process ○○●○
c Similar handling process ○○○●	t Version Management ●●●●
m Clear borders ○○○●	t Metadata about the processes ○○●○
m Lack of understanding from management ○○○○	t Selection of metadata standards ○●○○○
o Organizational Culture ●●○○●	
o Participators in system development projects ○●○○○	
o Lack of procedures for document management ○○○○	
o ISO 9001 certification ○●○○○	
t Level of traceability ●○○●○	
t Selection of metadata	
t Formalization ●○○●●	
t Document-based preservation ○○○●	

Design and Creation

Table of signs:

- c - competence
- m - management
- o - organization/procedures
- t - technology

○○○● = marked dot indicates which paper the factor has been presented in. In this case the fourth research paper.

<div> <div>c</div> <div>Staff turnover</div> <div>○○○○</div> </div>	<div> <div>m</div> <div>Lack of understanding from management</div> <div>○○○○</div> </div>
<div> <div>m</div> <div>Lack of understanding from management</div> <div>○○○○</div> </div>	<div> <div>o</div> <div>Organizational changes</div> <div>○○○●</div> </div>
<div> <div>o</div> <div>Organizational changes</div> <div>○○○●</div> </div>	<div> <div>o</div> <div>Lack of procedures for document management</div> <div>○○○●</div> </div>
<div> <div>o</div> <div>Organizational Culture</div> <div>○○○●</div> </div>	<div> <div>t</div> <div>Document-based preservation</div> <div>○○○●</div> </div>
<div> <div>o</div> <div>Missing documents that have never come to the authority</div> <div>○○○●</div> </div>	
<div> <div>o</div> <div>What is defined as a process? Where does a process end?</div> <div>○○○●</div> </div>	
<div> <div>o</div> <div>Lack of procedures for document management</div> <div>○○○●</div> </div>	
<div> <div>t</div> <div>Metadata addition</div> <div>○●●○</div> </div>	
<div> <div>t</div> <div>Document-based preservation</div> <div>○○○●</div> </div>	
<div> <div>t</div> <div>Formalization</div> <div>○○○●</div> </div>	
<div> <div>t</div> <div>Connecting components</div> <div>○○○●</div> </div>	
<div> <div>t</div> <div>Dating</div> <div>○●○●</div> </div>	
<div>Short and Middle term</div>	<div>Long term</div>
<div> <div>Table of signs:</div> <div> <div>c - competence</div> <div>m - management</div> <div>o - organization/procedures</div> <div>t - technology</div> </div> <div> <div>○○○○ = No marked dot indicates that the factor has been identified in empirical data that has not been published</div> </div> </div>	

Detailed Presentation of the Factors

In this section all the factors are explained in detail. The factor itself is in bold type and is followed by an explanation. In some cases, where it has been judged necessary, examples from the empirical material are presented. The factors are structured in the same order as in the analysis model. The headings are divided into two levels, where the design and creation factors come first and thereafter the factors that have been categorized according to whether they are short and middle term and or long term factors. The spheres are listed in alphabetic order. The factors are not ordered within each sphere. In the cases where a factor has been identified within several time states, the factor has been explained several times to increase the readability.

Time state: Design/Creation

Competence

- **Lack of competence.** The lack of competence is caused by the fact that the group that creates documents and records has grown. There are examples of employees who do not understand why record keeping is important which influences for example how they fill in metadata, or which document they think should be preserved. This will also lead to difficulties when it comes to trying to make the employees use the similar handling process.
- **Similar handling process.** The more the process is standardised the easier it will be to follow what has happened. This is valid both from the internal view when an employee has to use information from a process for which someone else has been responsible and from an external point of view when a citizen wants to follow a process electronically in an e-service.

Management

- **Clear borders.** This factor has to do with where a process ends and if whether process should start at the same moment. This matter cannot be solved in a delimited part of an organization; it has to be coordinated with the whole organization.
- **Lack of understanding from management.** This research has shown several examples where management does not give sufficient priority and status to record keeping and preservation.

Organization/Procedures

- **Organizational Culture.** The culture of an organization influences a lot of things that concerning record keeping and thereby the ability to ensure a high level of traceability between information and processes. Examples of this is include the kinds of documents the employees think of as important to preserve; the employees' attitude to record keeping; and where in the organizational structure record managers and archivists are suited.
- **Participators in system development projects.** The kind of professionals engaged in a system development project influences the degree of traceability between information and processes. This is particularly true for systems that include information that should be preserved, an example being whether or not an archivist/records manager is engaged in designing the metadata requirements.
- **Lack of procedures for document management.** In all the studied organizations the widespread use of information technology has led to an increase in the number of employees that have opportunities and obligations for creating information. Instead of using a secretary to formulate a letter to a client the employee writes them themselves, for example by using e-mail. Without written and known procedures that regulate for example how the document should be structured, where the document should be stored and how to keep track of versions it will be difficult to achieve traceability between information and processes. It could even lead to difficulties to retrieving the information at all.
- **ISO 9001 certification.** The ISO 9001 does not force organizations to work in a process oriented way, but the organization is required to describe their processes and to make a document plan. This research has shown that a certification process like 9001 could thereby lead to an increased awareness of the importance of document flow in the organization. This awareness also leads to better control over processes and the documents that are created within those processes.

Technology

- **Level of traceability.** A process consists of activities. An activity consists of transactions. The connection between for example a document and the process that created it could be on the process level, sub process level, activity level or transaction level. Which level the organization chooses is governed by its own internal choice. An organization could choose to associate a document to the process level and then describe which eventual sub-processes and activities belong to that process. Transactions may be managed separately or together, as Reed (2005) describes. Associations set on the lowest level will give the highest

traceability but also most associations to manage. In study one, records were found that had no connection to a transaction and thereby no connection to process. This is an undesirable situation resulting in very low traceability.

- **Selection of metadata.** Metadata is the one of the keys to preservation of context, and the context is the key to preserving interpretable information over time. Deciding which metadata to capture is a very crucial issue.
- **Formalization.** Formalization includes designing and formalizing of the handling processes and use of templates. This factor is influenced by and thus connected to the factor about similar handling. Several organizations have ambitions to increase the level of formalization. Formalization has several positive effects on traceability. It is, for example, easier to follow the handling process afterwards if the organizations have consistently used templates and the same kinds of forms. The forms themselves include a lot of metadata, which makes them easier to interpret. Among the negative effects is that forms are stereotypes and could lead omission of important information about a process because it does not fit in the template.
- **Document-based preservation.** This research has shown examples of document-based preservation, meaning that the information that have been preserved is the documents that have been created and managed within a process. An example of problems with this strategy is that information saved in a database and only presented to the user on the screen will not be preserved. Possible alternatives to this strategy would be either to create documents of screenshots or to keep the whole system intact.
- **Traceability on normal process respectively alternative process.** The research has shown several instances of higher level of traceability when the process is not normal an example being an appeal process which is often very well documented. But when everything is done according to a normal process there are sometimes few signs what process looked like. An example of this is where the only part of a process found in the file is an application form and it is implied that it has been approved if there is no evidence of the opposite.
- **Version management.** Version control was of great importance to all the organizations in all of the studies but it may be seen as a matter of importance only in the present and not relevant for preservation. The strategy varies among organizations but in some it is clear that the only version that should be preserved is the final valid one. The first study however showed example of an organization that preserved all versions sent in to the authority, even those that included er-

rors which been corrected in later versions. As long as the document is in active use organizations usually strive to reach full traceability of all versions of a document.

- **Metadata about the processes.** The metadata that has been studied in this research is added to a document. The document is described by its metadata, for example about author, date of creation and to which process it belongs. Metadata about the process is necessary in order to be able to interpret the process itself. None of the studies have shown any examples of such metadata.
- **Selection of metadata standards.** Metadata standards have been developed in several academic and professional fields. There are many examples of metadata for geographical, statistical or preservation needs. The choice of metadata standard will influence how and from which angle the context is described. Preservation over time requires special metadata: some metadata element have to be added when the information is created while others have to be added afterwards. Cooperation between different professional and interest groups is particularly useful identifying the metadata concerning creation.

Time state: Short and Middle Term

Competence

- **Staff turnover.** Staff turnover is a natural and essential part of an evolving organization. But even though it is a normal part of any organization it causes some problems for preservation of information. First of all many years' experience is of important when using information that has to been interpreted because the employee is familiar with the context. The second identified problem is that some organizations do not have routines to ensure employees pass on information when leaving the organization. For example resigning employees could leave their computer without any information about its folder structures. Another example was a municipality that had to call in a retiree to describe and document how the archive was structured when new employees needed that information.

Management

- **Lack of understanding from management.** This research has shown several examples where management does not give sufficient priority and status to record keeping and preservation.

Organization/Procedures

- **Organizational changes.** Reorganizations are a common tool for improving an organization in one way or another. New demands on the market and a transformed commission could also lead to changes within an organization. Those changes often result in new divisions, new managers and transformed routines. Organizational changes will also affect how the organizations manage their information and in some cases also what kind of information they have to manage.
- **Organizational Culture.** The culture of an organization influences a lot of things that concerning record keeping and thereby the ability to ensure a high level of traceability between information and processes. Examples of this is include the kinds of documents the employees think of as important to preserve; the employees' attitude to record keeping; and where in the organizational structure record managers and archivists are suited.
- **Missing documents that have never come to the authority.** Current Swedish legislation has made the commissioner of a building responsible for certain types of documents. Once the building is finished the commissioner of the building is required to send in a certificate to validate that the building has been built according to the building permit. This is usually forgotten, and thus making it impossible to follow the process (the building permit) to the end.
- **What is defined as a process? Where does a process end?** It is important that a process is well defined in order to ensure traceability between information and processes. This includes where and how a process starts and how it ends. The study of two Swedish municipalities shows that it is important to discuss where the process ends. During the seminars discussion centred on whether the process should end when the building is finished, but as mentioned earlier it is the responsibility of the commissioner of the building to report that the building is finished. If the certificate is never sent in to the municipality it will never be possible to close the process. Other suggestions for ending the process were when the municipality has completed the administration of the building permit.
- **Lack of procedures for document management.** In all the studied organizations the widespread use of information technology has led to an increase in the number of employees that have opportunities and obligations for creating information. Instead of using a secretary to formulate a letter to a client the employee writes them themselves, for example by using e-mail. Without written and known procedures that regulate for example how the document should

be structured, where the document should be stored and how to keep track of versions it will be difficult to achieve traceability between information and processes. It could even lead to difficulties to retrieving the information at all. When time has passed, use and interpretation of the information would be facilitated if the procedures were preserved together with the information. Another type of problem that arises over time are missing descriptions of codes. Codes are used by several organization to describe what kind of error or what kind of action that has been done.

Technology

- **Metadata addition.** With the pro-active approach of the continuum model metadata should be captured as early as possible, but there will always be examples of metadata must be added afterwards, for example in which information system this information has been created and appropriate classification codes. Even so it is important to create routines for this already at the design phase so it will facilitate long-term management.
- **Formalization.** Formalization includes designing and formalizing of the handling processes and use of templates. This factor is influenced by and thus connected to the factor about similar handling. Several organizations have ambitions to increase the level of formalization. Formalization has several positive effects on traceability. It is, for example, easier to follow the handling process afterwards if the organizations have consistently used templates and the same kinds of forms. The forms themselves include a lot of metadata, which makes them easier to interpret. Among the negative effects is that forms are stereotypes and could lead omission of important information about a process because it does not fit in the template.
- **Connecting components.** The research has shown examples of different kinds of components that connect different documents and their context. One type is some sort of unique key, which will have to be dealt with carefully if the unique key can change over time, which is the case with property unit designation. Another important component necessary for the preservation of context is use of the diary. There, in the diary, it is possible to follow different audit trail for the process, such as which documents were part of the process and who was the applicant. The third identified type is something called a cover page. The page is part of the workflow and document management system and includes all the information that the diary contains together with other central information.
- **Document-based preservation.** This research has shown examples of document-

based preservation, meaning that the information that have been preserved is the documents that have been created and managed within a process. An example of problems with this strategy is that information saved in a database and only presented to the user on the screen will not be preserved. Possible alternatives to this strategy would be either to create documents of screenshots or to keep the whole system intact.

- **Dating.** Dating has turned out to be an essential part of preservation of information over time. Traditionally (as in paper archives) this is always fulfilled but it is important to bring this competence to designing computerized information systems. The component date has been shown to be very important to verifying the order of activities within a process.

Time state: Long Term

Competence

Unidentified factor

Management

- **Lack of understanding from management.** This research has shown several examples where management does not give sufficient priority and status to record keeping and preservation.

Organization/Procedures

- **Organizational changes.** Organizational changes are, as mentioned earlier a fact in the short and middle term and of course a natural part of the long term life of an organization. The effects of change will be even harder to overcome in the long term because of staff turnover and therefore loss of organizational memory.
- **Lack of procedures for document management.** In all the studied organizations the widespread use of information technology has led to an increase in the number of employees that have opportunities and obligations for creating information. Instead of using a secretary to formulate a letter to a client the employee writes them themselves, for example by using e-mail. Without written and known procedures that regulate for example how the document should be structured, where the document should be stored and how to keep track of versions it will be difficult to achieve traceability between information and processes. It could even lead to difficulties to retrieving the information at all. When time has passed, use and interpretation of the information would be facilitated

if the procedures were preserved together with the information. Another type of problem that arises over time are missing descriptions of codes. Codes are used by several organization to describe what kind of error or what kind of action that has been done.

Technology

- Document-based preservation.** This research has shown examples of document-based preservation, meaning that the information that have been preserved is the documents that have been created and managed within a process. An example of problems with this strategy is that information saved in a database and only presented to the user on the screen will not be preserved. Possible alternatives to this strategy would be either to create documents of screenshots or to keep the whole system intact.

Discussion of the factors

The result of the re-analysis has identified factors within four spheres, namely: competence, management, organization/procedures and technology. These factors have been presented in detail in the above section. To sum up and give an overview of the result, the factors have been graphically presented in the figure below. The figure visualizes the quantity of factors in each sphere and one Synergy-4 model for each

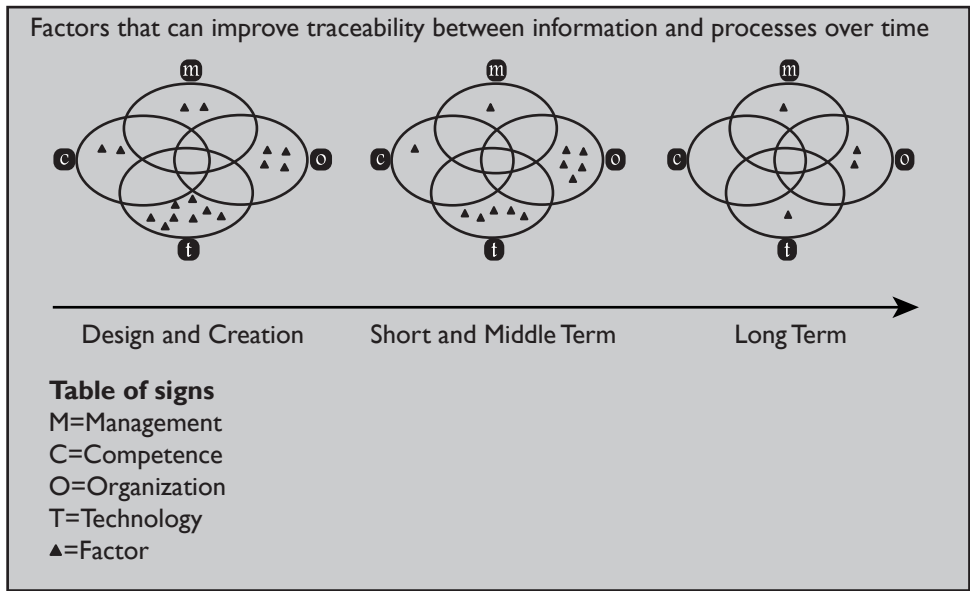


Figure 6 Summing up the factors

time state. The description of the signs is presented in the figure.

The majority of the factors have been categorized as technological ones. What is the explanation for this? This could imply that technical issues are of greater importance than other aspects of preservation of information over time. However, the more likely explanation is that this research has had a technical focus. This especially applies to papers two and three. Paper two and three discuss metadata capture within computerized information systems whereas paper one and four studied information independent of whether it is electronic.

An interesting matter for discussion is that there is one example of a sphere that does not include any factors at all. This is the sphere of competence in the long-term time state. There are however few competence-factors across all time-state. But the competence field is interesting in two different ways. First of all competence is an issue that is important to organizations generally: it is not confined information management and recordkeeping. The second reason is that the information asset itself is part of sources for competence in an organization. This research has revealed an example the necessity of asking a pensioner to return to the office to explain how the archive was organized. The head of the archived material had obviously not archived metadata of how to use the archived material. Staff turnover is one competence factor that was identified over short and middle term. In the long run it is not possible to avoid staff turnover, it has to be dealt with one way or another and process modelling techniques can be useful here. The fourth paper shows that it is possible to reconstruct processes afterwards, which could be used for the transfer of procedures/routines over time. The development of ISO 9001 also forces the use of process modelling as a tool to make the processes established and known. A final remark regarding competence is that a possible overlap between this research and the field of knowledge management and learning organizations has been shown. This relationship could be an interesting direction for further research.

When it comes to the factors within the organizational sphere it is apparent that organizational culture influences things such as metadata capture and participants in system development projects. Traceability between information and processes is of very high importance inside an organisation: the study in two municipalities (paper four) showed that it was their internal processes were visible in the archives. Nowadays upcoming e-services generate a problem both from an organization perspective

and from a citizen perspective. Organizations often want to slim their organization when they develop e-services and one idea in the project the Home Building Guide⁷ is that municipalities could help each other with building permits. This requires higher conformity of processes and information management across cooperating organizations. But when organizational culture influences the way procedures for information management and recordkeeping are followed it is a problem, even if it is only an internal problem because it engenders problems with traceability over time.

The last of the factors for discussion concerns management. As the case with competence there are few factors that have been categorized within the management sphere. One thing to observe is that the same factor, namely: *Lack of understanding from management* has been identified in all three time states. Traceability is in this sense part of a larger issue, namely information management over time. The big challenge here is to convince the management that to be able to use information that is created today in lets say, 50 years ,the strategy must be built today. The first step in the strategy is to include archival matters in a strategy of information management because archival issues cannot be solved later on.

Looking at figure 6 from a time perspective, most of the factors have been identified at the first time state (Design and Creation). The question is what does this indicate? First of all the pro-active approach and the continuum model provide the standpoint that the design and creation phase has a crucial role for information in electronic form. The studies are focused on electronic information, therefore it is no surprise that most of the factors are identified in this time state. Second, it is only the fourth paper that concern the long run, so the quantity of empirical data is lower in the two latter time states.

No difficulties were experienced in identifying which time state should be applied to which factor. But there is one important issue discuss namely that the borders between the three different time states are blurred. The view is that the time states used in this thesis are independent of differences in use over time. This has been discussed in Borglund and Öberg (2006)where the view is that the use of records has

⁷ "Home Building Guide" is an on-going research project (Bygga Villa in Swedish) the aim of which is to build an integrated e-service. This integrated e-service will cross agency borders in order to make it easier for a citizen to seek information, for example about a property unit purchase or to follow the different processes that are needed to build a house.

varies from active and semi-active independent the age of the information.

6. Concluding remarks

This research has shown that there are a lot of different factors that could improve the traceability between information and processes. The factors have been identified within four different spheres (competence, management, organization/procedures and technology) where the technology sphere has been the one with most identified factors. The factors have been divided into three time states where the creation phase has generated most factors. The conclusion here is that the creation phase is important to preservation of information. This may be seen as a consequence of the pro-active approach, which is an underlying assumption of this thesis. But the examples of factors like formalization and level of traceability show that it would not be possible to complement an information system with this type of functionality when it already been built and implemented. Information system that includes information that has value over time has to be designed with this type of functionality from the start.

An important remark from this research is that to be able attain traceability an organization has to implement a strategy that covers a range of different aspects. Those aspects everything from metadata capturing, choices of preservation strategies, training of staff, engagement and resources from management and common and shared procedures. The used re-analysis model used(Synergy-4) has been helpful in the identification and structuring of factors while it look at an organization from four different angles (Competence, management, organization/procedures and technology).

Another conclusion from the studies is that the widespread use of computers on a daily basis has caused some problems that influence traceability, for example a larger quantity of users that creating documents and problems with formats. But, on the other hand is there is also an enormous potential in the technology. Preserved information can be accessed more easily than before and there are greater possibilities for combination of information through use of different graphical presentations without destroying the original. Some organizations make use of this potential, but there are also examples where instead of using new ideas the traditional way of dealing with information is transferred without change. So, the digital solutions are copies of the manual ones with the same limitations.

An interesting area for further research is that of boundaries. To ensure traceability there is a need for definitions of the borders. An example identified in this research concerns where the borders of a process begin and end. Another example of an issue concerning borders is metadata. To be able to use metadata there will have to be descriptions of the metadata that will have to be described by metadata that will have to be described... Where should this spiral end?

Another fruitful area for further research would be to use the Synergy-4 model for comparative studies of factors that could improve traceability in other research settings for example in other countries or in private organizations. If the Synergy-4 model is used as a lens in the data collection phase it would be possible to fill in gaps like those reported in the two latter time states (the short and middle term and the long term).

As a final remark this research has produced s valuable knowledge both for building strategies for preserving electronic information over time and as detailed information of what is important in designing and building information systems that maintain traceability over time.

Lessons learned

Methodological

An important matter to discuss is whether it is possible to generalize the results of this research. This research is based on empirical grounds using a qualitative methodological approach. The studies have covered at most four organizations (in study one) and in the other cases two organizations (study two and three). The empirical data from those studies do not give a general picture, and that was not the aim of the studies. This research has been explorative due to lack of empirical studies in an emerging research field. The explorative nature of the research has resulted in studies that also fulfill the need of for identifying problematic areas. But, by use of different analytical tools it was possible to compare the empirical data from a small set of organizations to existing theory and models, which in turn provide possibilities for arguments for a higher level of recognition in other settings. The intention has been to reach high transferability by providing rich descriptions of both the research settings and the research process. Unfortunately in research paper three this was not possible on the limitations on acceptable length for the paper. However, the second paper is based on the same study and includes more details about the settings and the process.

The most important reflection on the way the chosen methodology has worked is that the three studies have been focused. As mentioned before this research has been explorative but in a structured manner. The studies were planned early (see also the next section) and there was small deviation from the “original plan”. Each study was of course planned in detail just before the study was carried through but the original idea was set out early in the research process. Thus the studies have not influenced each other when it comes to focus and research questions. The studies have of course increased the earlier understanding of the research domain and by that means influenced the analysis. An alternative approach would have been that the first study could have led to new questions that were answered in the second study and so on. If the latter approach had been used the result would not have been the same, but it is hard to say what the outcomes might have been, because it is impossible to predict how the chain of questions would have developed. The studies have each focused on parts of the aim of this thesis and the analysis model (Synergy- 4) was a necessary step to be able to study the results from another and wider perspective. A different analysis model would have led to an analysis that focused on other details in the existing empirical data and the presentation would

then have been different. The possible alternatives to the Synergy-4 model were traditional qualitative analysis methods. However, the choice was made based on requirements were presented in the Research Perspective section.

The model used has shown several benefits. First of all it has contributed to the wider view of the collected data and the contribution of each research paper. Secondly, the model has enabled visualization of the results in different ways and thirdly, the model was helpful in taking an outsider's perspective (see also the next section). During a validation seminar one attendee raised a question about factors that can influence the traceability that is outside the organization. One example of such factor would be legislation. The suggestion was to add a sphere to the model that could be called the surrounding world. This change could lead to interesting development of the model.

Personal

Working on this thesis or research in general is often described as a journey, so I will here use that metaphor to describe the progress of this thesis and my own learning as researcher. The journey started in the spring of 2004 when I was employed as a Ph D student in a project called Archives of the Future. My colleague and I were in the same situation; we were entering what was, for us, an unknown scientific field. We both had our master degree in informatics and the research project involved collaboration between informatics and archival science. Early in the project we were both requested to present the aim and direction of our thesis work. We had a couple of weeks to make a decision. I started to read about problems about preservation and my interest fell on processes and how information was often seen as chains in associations. To use the journey metaphor, it felt like jumping on a train that seemed be going somewhere interesting but I had actually no idea exactly where it would take me. I got fast feedback that processes were of high interest to the archival community and there were several existing projects on different levels.

In order to take the chosen railway-carriage my colleague and I made our first empirical study. It was both enjoyable and useful to study the real world and try to understand if and how what we saw relates to all the theoretical concepts we were reading about. The study was wide and open and could be compared with looking at the schedule to check what kind of journey alternatives there are. The study also revealed a lot of problems those different organizations meet when trying to manage

their information over time. The second study was carried out shortly after the first one was finished and in that study I focused on metadata in document management systems, which supports the business processes within the organization. After this study I ran in to some problems concerning a really important issue of a researcher's job, namely publishing. The journey takes you to different places, here described as railway carriages. During the trip and when it has ended it is important to document what you have observed. Part of the documentation is to be able to save data for later purposes and be able to analyze them. The other part has in my case been to write research papers. The second study resulted in two papers viewing different perspectives of the study (papers two and three in this thesis). The second paper was sent to a journal and after several months rejected. I have worked with revisions in several iterations before the paper was submitted again. I think that the hard part for me has been that the writing of a research paper requires that you take a step back and think of what the result really shows. There is a Swedish proverb that describes the problems I have well; It is hard to see the forest because of all the trees. During the empirical studies and also after I have finished them I had concentrated so much on interesting details within the study that it had been hard to describe the results on another level. To revisit the same metaphor again, in order to write the research paper you need to take a plane and look at the railway carriages from above. There it is possible to describe how the railway carriage I have been studying could be attached to carriages that already have been studied by other researchers. The higher perspective is also necessary to describe what kind of carriage you jumped into. This step would not have been possible without help from my tutors and I had not realized the problem fully until recently.

The second study focused on digital solutions and early in the process I felt that to be able to study the chosen problem I also had to study older material. How does one make journeys through time, is there any time machine? The archives make it possible to move backwards in time. The third study was a journey that took me to how building permits was handled before the introduction of computer-based information systems. Experiences from this study of an earlier period gave me a lot of new knowledge about what is needed in order to use information that has been created and used in another context. When I had made all those trips in different railway carriages and also described them using different perspectives in four different research papers the journey came to a new phase. The issue was then to describe my results from all the journeys and the relationships between those journeys. I also wanted this description to be done so that it could be used to improve the possibilities for preserving information. This step has been the hardest and it has puzzled

me a lot. First I concentrated my thoughts on how to switch from empirical studies to some kind of design activity. I could see that this was a necessary step and I tried to force my thoughts to switch to this design mode. I did not know if this would be possible. In my case the push in the right direction came in a seminar where I was supposed to present my thoughts about my final results and thereby this first section of the thesis. A colleague of mine said simply – maybe you need an analysis model? That gave me the input I needed to know in what kind of direction I should travel. My search for a suitable model resulted in use of the Synergy-4 model by Holmberg(2001). This model helped me look at what I have done from above; using the journey metaphor it has been maybe to use a space rocket so that even the journeys by plane could be studied from the above.

The journey is now nearly ended, and maybe I lied when I said that switching from an empirical mode to a design mode was the hardest, I think that writing this reflective section of my thesis was even harder. And now I wonder, what kind of journey metaphor could be used to reflect what the space rocket journey was like?

To end this section I would like to tell you a story about a PhD student I know well. At her first meeting with the reference group for the project in which she was working, she gladly said to all present

– I think I have got the best job in the world.

I know from a well-informed source that she still thinks she has the best job in the world and that she hopes that the journey will continue!

7. References

- Antoniol, G., Canfora, G., Casazza, G., De Lucia, A., & Merlo, E. (2002). Recovering Traceability Links between Code and Documentation. *IEEE Transactions on Software Engineering*, 28(10), 970-983.
- Asproth, V. (2005). Information technology challenges for long-term preservation of electronic information. *International Journal of Public Information Systems*, 2005(1), 27-37.
- Bantin, P. (2002). Electronic Records Management - A Review of the Work of a Decade and a Reflection on Future Directions. *Encyclopedia of Library and Information Science*, 71(Supplement 34), 47-81.
- Bearman, D. (1994). *Electronic evidence: strategies for managing records in contemporary organizations*. Pittsburgh: Archives and Museum Informatics.
- Borglund, E., & Öberg, L.-M. (2006). *Operational Use of Records*. Paper presented at the 29th Information Systems Research Seminar, Helsingoer.
- Buckland, M. K. (1991). Information as Thing. *Journal of the American Society for Information Science*, 42(5), 351-360.
- Cao, G., Clarke, S., & Lehaney, B. (2001). A critique of BPR from a holistic perspective. *Business Process Management Journal*, 7(4), 332-339.
- Connolly, T. M., & Begg, C. E. (2005). *Database systems: A practical approach to design, implementation and management (4th, rev.ed)*: Harlow Addison-Wesley.
- Cunningham, A. (2001). Six Degrees of Seperation: Australian Metadata Initiatives and Their Relationships with International Standards. *Archival Science*, 1, 271-283.
- Duff, W. (2001). Evaluating Metadata on a Metalevel. *Archival Science*, 1, 285-294.
- Duranti L. (2002). (red) *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*.
- Duranti, L. (1997). The archival bond. *Archives and Museum Informatics*, 11(3-4).
- Duranti, L. (2001). Concepts, Principles, and Methods for the Management of Electronic Records. *The Information Society*, 17, 271-279.
- Duranti, L., & Thibodeau, K. (2006). The concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES. *Archival Science*,

2006(6), 13-68.

- Eriksson, D. M. (1998). *Managing Problems of Postmodernity: some Heuristics for Evaluation of Systems Approaches*. Laxenburg: International Institute for Applied Systems Analysis.
- Eriksson, D. M. (2004). *Four Proposals for Enterprise Modeling*. Ph. D Thesis, Chalmers University of Technology, Göteborg.
- Evans, J., McKemmish, S., & Bhoday, K. (2005). Create Once, Use Many Times: The Clever Use of Recordkeeping Metadata for Multiple Archival Purposes. *Archival Science*, 5, 17-42.
- Flood, R. L., & Jackson, M. C. (1991). *Creative Problem Solving: Total Systems Intervention*. Chichester: John Wiley & Sons.
- Gilliland, A., & McKemmsish, S. (2004). Building an Infrastructure for Archival Research. *Archival Science*, 4, 149-197.
- Gilliland-Swetland, A. J. (2000). Setting the stage. In M. Baca (Ed.), *Introduction to Metadata: Pathways to Digital Information Version 2.0* (pp. 1-12): Getty Research Institute.
- Grant, D. (2002). A Wider View of Business Process Reengineering. *Communications of the ACM*, 45(2), 85-90.
- Grover, V., Jeong, S. R., Kettinger, W. J., & Teng, J. T. C. (1995). The Implementation of Business Process Reengineering. *Journal of Management Information Systems*, 12(1), 109-144.
- Guercio, M. (1997). Definitions of Electronic Records, the European Perspective. *Archives and Museum Informatics*, 11(3-4), 219-222.
- Hofman, H. (1998). Lost in cyberspace : where is the record? In K. Abukhanfusa (Ed.), *The concept of record : report from the Second Stockholm Conference on Archival Science and the Concept of Record, 30-31 May 1996*. Stockholm: Swedish National Archives.
- Holmberg, S. C. (2001). Taking Syntegrity-4 from Assumption Mode to Reflection Mode. *Systems Research and Behavioral Science*, 18, 127-135.
- International Standards Organization,(2001). *Information and Documentation and Records Management Part I: General*. Geneva: International Standards Organization.
- International Standards Organization,(2006). *Information and documentation -- Re-*

- cords management processes -- Metadata for records -- Part 1: Principles. www.iso.org.
- InterPares I and II. (2007). *Homepage of Research project InterPares II*. Retrieved 13 April, 2007, from www.interpares.org
- Jackson, M. C. (2003). *Systems thinking Creative Holism for Managers*. West Sussex: John Wiley & Sons Ltd.
- Katzenstein, G., & Lerch F. J. (2000). Beneath the Surface of Organizational Processes: A social Representation Framework for Business Process Redesign. *ACM Transactions on Information Systems*, 18(4), 383-422.
- Landgren, J. (2007). *Designing information technology for emergency response*. Göteborg: IT-University of Göteborg.
- Langefors, B. (1995). *Essays on Infology*. Lund: Studentlitteratur.
- Marcus, A., & Maletic, J. I. (2003). *Recovering Documentation-to-Source-Code Traceability Links using Latent Semativ Indexing*. Paper presented at the 25th International Conference on Software Engineering.
- McDonald, J. (1997). Towards Automated Record Keeping, Interfaces for the Capture of Records of Business Processes. *Archives and Museum Informatics*, 11, 27-285.
- McKemmish, S. (2001). Placing Records Continuum Theory and Practice. *Archival Science*, 1, 333-359.
- McKemmish, S. (2005). Traces: Document, record, archive, archives. In S. McKemmish, M. Piggott, B. Reed & F. Upward (Eds.) *Archives: Recordkeeping in Society* (pp. 1-20). Wagga Wagga: Center for Information Studies.
- McKemmish, S., Acland, G., Cumming, K., Reed, B., & Ward, N. (1999). Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema. *Archivaria*, 48.
- Orlikowski, W. (1993). CASE Tools as Organizational Change: Investigating Incremental and Radical Changes in Systems Development. In M. D. Myers & D. Avison (Eds.), *Qualitative Research in Information Systems*. London: Sage Publications.
- Pédaque, R. T. (2003). Document: Form, Sign and Medium, As Reformulated for Electronic Documents (3d ed., pp. 1-25). http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/05/94/sic_00000594.pdf.

- Public Record Office Victoria, P. (2006). *Victorian Electronic Records Strategy - Forever Digital*. Retrieved 2006-06-2006
- Päivärinta, T., Tyrväinen, P., & Ylimäki, T. (2002). *Defining Organizational Document Metadata : A Case Beyond Standards*. Paper presented at the Proceedings of the Tenth European Conference on Information Systems, Gdansk, Poland.
- Reed, B. (2005). Records. In S. McKemmish, M. Piggott, R. Barbara & F. Upward (Eds.), *Archives: Recordkeeping in Society*: Quick Print.
- Salminen, A., Lyytikäinen, V., & Tiitinen, P. (2000). Putting documents into their work context in document analysis. *Information Processing and Management*, 36, 623-641.
- Shepherd, E., & West, V. (2003). Are ISO 15489-1:2001 and ISAD(G) compatible? Part 1. *Records Management Journal*, 13(1), 9-23.
- Smith, H. A., & McKeen, J. D. (2003). Developments in practice viii: Enterprise content management. *Communications of the Associations for Informations Systems*, 11, 647-659.
- Sundberg, H. P., & Sandberg, K. W. (2006). Towards e-government: A survey of problems in organisational processes. *Business Process Management Journal*, 12(2), 149-161.
- Sundgren, B., & Steneskog, G. (2003). Information Systems for Concerted Actions. In B. Sundgren, P. Mårtensson, M. Mähring & K. Nilsson (Eds.), *Exploring Patterns in Information Management - Concepts and Perspectives for Understanding IT-Related Change* (pp. 356). Stockholm: The Economic Research Institute Stockholm School of Economics.
- Thomassen, T. (2001). A First Introduction to Archival Science. *Archival Science*, 1, 373-385.
- Upward, F. (2000). Modelling the continuum as paradigm shift in recordkeeping and archiving processes, and beyond - a personal reflection. *Records Management Journal*, 108(3), 115-139.
- Upward, F. (2005). The records continuum. In S. McKemmish, M. Piggott, R. Barbara & F. Upward (Eds.), *Archives: Recordkeeping in Society* (pp. 348): Quick Print.
- Upward, F., & McKemmish, S. (2001). In Search of the Lost Tiger by Way of Saint-Beuve:Reconstructing the Possibilities in 'Evidence of Me...' *Archives and Manuscripts*, 29(1), 22-43.
- Vakola, M., & Rezgui, Y. (2000). Critique of existing business process re-engineering

methodologies The development and implementation of a new methodology. *Business Process Managment Journal*, 6(3), 238-250.

Öberg, L.-M. (2005). *To Capture and Preserve Document in Processes*. Paper presented at the Information Management in modern enterprise: Issues & solutions., Lisbon Portugal.

Paper I



What are the Characteristics of Records?

Öberg, L.-M. & Borglund, E. (2006). What are the Characteristics of Records? *International Journal of Public Information Systems*, 2006(1), 55-76.

Abstract

Project documentation, minutes of meetings, emergency plans, and inspection reports are all examples of records, recorded information. There is a changeover from paper-based records to electronic records in many organizations and these organizations often implement information systems managing electronic records that only take archival requirements into consideration to some extent, if at all. This is a problem that makes preservation of electronic records difficult. The aim of this paper is to describe the characteristics of records, based on a study of empirical data and archival theory. It is essential to identify and understand the characteristics of a record in order to manage and preserve records in computerised systems. Knowledge of what characteristics a record has is one way to make it possible to formalize records. Formalization at different levels is needed for computerized management of records. This paper is based on a qualitative case study performed at four different organizations in Sweden. Empirical data was collected from multiple sources within each organization and resulted in five groups of characteristics with several sublevels. The empirical data were compared with recordkeeping and archival theory. Based on this study the essential characteristics of records were identified as context, form, organization, structure and version/copy.

Keywords: Archival theory, Electronic records, Formalization, Information systems, Record characteristics.

I Introduction

Many organizations are changing from paper-based records to electronic records. This is a natural process, when more and more information and documents are created by the usage of computer-based information systems (henceforth only the term 'information system' will be used). The problem is that many organizations implement information systems for managing electronic records, which take recordkeeping requirements into consideration only to some extent. A record is more than just

information, it is supposed to be trustworthy: reliable and authentic, able to serve as evidence, and to support accountability. The length of time for which the records need to be preserved could vary from months up to hundreds of years (Duranti, 2001a; Thomassen, 2001). For example, the Swedish police have implemented an information system that does not correspond to recordkeeping requirements. They have a police management information system from which all records should be preserved forever, in accordance with regulations from the Swedish National Archives. However, after 13 months all records are written to a flat ASCII-file (which is a format that has qualities which are suitable for the technical aspect of preservation), which makes it impossible to reuse the records e.g. to interpret and understand the records. Another example is from a public Swedish Organization (named organization 1 in this paper) which has implemented a digital archive system. In the digital archive system there are no connections between some of the records and the transactions, which is a key principle for the requirement of a record serving as evidence of an organization's actions. One example is when the organization 1 receives an application that is incorrectly filled in, a letter of need for change is written and sent back to the person who made the application. These letters are auto generated based on error codes. The system logs that a letter has been sent out, related to error codes. If error codes change, which happens occasionally, it may not be possible to recreate those transactions. Iacovino (2004) describes an Australian national electronic health record system, a system not following the recordkeeping requirements such as reliability and authenticity. The electronic health records system cannot create trustworthy records and lacks critical and necessary recordkeeping functions. For example the system does not manage to capture and preserve process records related to a more objectified record, yet Grimson (2001, p. 121) lists six different research issues given less attention, but which need to be solved in order to deliver electronic healthcare records of the 21st century, including "techniques for characterizing data quality and provenance, preservation of access to the record over time" are all research issues related to recordkeeping requirements. Holgersson (2001) reports that some of the information systems managing electronic records in the Swedish police, filter information which negatively influences possibilities to trust records and make correct decisions. Bearman (1994), makes a clear distinction between a recordkeeping system and an information system. The difference is that a recordkeeping system captures, manages, and provides access to records through time and meets records requirements. In this paper computer-based information systems are of interest, not because they are different from recordkeeping systems, but because record requirements must be fulfilled even in them. Records are created electronically in information systems, and it is important that the requirements are

captured even at creation to be able to maintain them through preservation. In design and development of information systems for management of electronic records, we propose a proactive approach. Within a proactive approach requirements and needs for trustworthy management of electronic records must be designed and implemented when the system is developed, a proposal implicitly supported by both the standard ISO 15489 (International Standards Organization, 2001a, 2001b) and by the European Commissions Model Requirements for the Management of Electronic Records, MoReq (European Commission, 2002). ISO 15489 recommends a careful analysis, and identification process of both requirements and strategies for electronic record management before the actual information system is developed. Even as early as 1992 Charles Dollar discusses why archivists should have knowledge of how object oriented computing works. An archivist may not be able to deliver archival and record requirements during information system design and development projects, without knowledge of how such a process works, and how system developers work (Dollar, 1992). Two years later David Bearman suggested a proactive approach, which he described as planning in risk management (Bearman, 1994).

Standard procedure when new information systems are designed and developed is to perform an analysis of the business activities and to capture requirements from businesses, legal aspects, users, customers and others involved. A component in this procedure is to define and make information intelligible, information of different types that the future information system is supposed to manage or process. This work is described as the formalization of information. According to Shipman & McCall (1994, 1999) information must be formalized so that it can be managed by information systems. In order to perform formalization, knowledge of the characteristics of the information is needed according to Shipman & McCall. In other words knowledge about the possible information types, properties, and interrelationships is needed. It is possible to grade the amount of formalization from less formal to most formal when more functionality is required and a higher level of formalization is needed (Shipman & McCall, 1999). A high level of formalization of records could for example enable automatic identification and detection of records within information systems.

‘The record’ as a concept in this paper should not be mixed up with the record concept used in database theory, where a record is a tuple¹ in a relation (e.g. Teorey, 1999). ISO 15489 defines records as: “Information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in transaction of business” (International Standards Organization, 2001a, p. 3), a definition used throughout the rest of this paper. Examples of records are; healthcare records, project documentation, minutes of meetings and inspection re-

1 Simplified a row in a table

ports.

Records are physical or virtual, have a content, a structure/form, are created in a context (Bearman, 1994; Hofman, 1998; McKemmish *et al.*, 2005), and are process bound information (Thomassen, 2001). But these features do not make records unique in relation to other forms or sorts of information. According to Thomassen (2001) a record has several criteria, which makes the record unique in relation to other types of information.

1. Records are evidence of actions and transactions;
2. Records should support accountability, which is tightly connected to evidence but which allows accountability to be traced;
3. Records are related to processes, i.e. “information that is generated by and linked to work processes” (Thomassen, 2001, p 374);
4. Records must be preserved, some for very short time and some permanently.

These four characteristics of records make them different from other types of information. In an organization records are part of the organizational memory and are used to support organizational management. Cox (2001) has stated that the evidential value of a record can only exist if the content, structure and context are preserved. The context is the link between different records that belong together, and also to the process where the record was created. The above criteria are defining what a record must achieve, but the criteria do not explain and exemplify what characteristics are needed in order for records to be formalized in an information system.

The aim of this paper is to describe the characteristics of records based on empirical data and archival theory. Accordingly the research questions are:

- a. What is defined as a record in organizations?
- b. What are the characteristics of these records?
- c. Are there any differences between the empirically grounded characteristics of records and the characteristics described by recordkeeping and archival theory?

This paper is addressed to both practitioners and scientists. Researchers within information systems, archival science, and records management as well as information systems designers, information systems architects, records managers, archivists, and information managers can be described as the target group of our contribution, which can be summarized as:

Identifying empirically grounded characteristics of records, which are compared with theoretical findings from archival theory and recordkeeping literature.

The application area is mainly development of new information systems managing electronic records.

2 Research method and research sites

This paper is based on a qualitative study, which has an underlying interpretative approach. In this research understanding and knowledge of the nature of records was our interest. According to Hartman (1998) qualitative research can support such needs and also increase the knowledge of phenomena in their natural environment. The chosen qualitative research method is the case study, a method well suited for information systems research where a phenomenon is supposed to be studied in reality and in its natural environment (Myers & Avison, 2002; Yin, 2003). The research can be described as interpretative, where an iterative interpretation process takes place during the analysis phase (Walsham, 2002). As recommended, different data collection techniques have been used in this research in order to capture the complexity and uniqueness of the settings of interest (Eisenhardt, 1989; Yin, 2003).

Qualitative research methodologies have been criticized for producing results with low quality, when the results are seldom generalizable (Lee, 1989). However, according to Guba & Lincoln (1989) results from qualitative research can be transferable to other organizations and settings and therefore be of high quality and usefulness. The aim of this paper is to present a description of the concept of records and their adherent characteristics, which is a first step towards a more transferable knowledge.

The research was carried out in four different organizations, which were chosen by three different criteria:

1. The organization must have a widespread use of records;
2. The organization must have information systems managing electronic records;
3. The organization must have an archive, for preservation of records.

The selection can be described as an adapted selection process (Hartman, 1998). An adapted selection process is suitable if the selection of research sites is assessed to answer the stated research questions (Hartman, 1998). The choice to select only four organizations was governed by the time available to perform this research. The se-

lection process also aimed to get a selection that was a range of representative organization types, with only four organizations. The organizations were:

– Org 1.

A Swedish public company, under the control of the Swedish government. The company has a large physical archive with more than 20 000 running meters of records. It has a large case flow of approximately 600 000 cases per year, which produce many records. Since 2004, it manages all its records in electronic form.

– Org 2.

A multinational enterprise. The enterprise is producing goods, and the production facility is vulnerable for unforeseen stoppage of production, which force the enterprise to manage their records with good structure.

– Org 3.

A Swedish municipality. A municipality has a variety of different types of records, dependent on the variety of tasks obligated for Swedish municipal government organizations.

– Org 4.

A Police authority in Sweden². The Swedish police both create and receives many records. The majority of the records are also going to serve as evidence in the courts of law. The police have several information systems managing records, which should be preserved forever.

2.2 Course of action

The research was designed to be performed in two stages at each research site, and was carried out during January and February in 2005. The first step was to study the archives, where records were stored and preserved. The empirical data were collected by physical reading and study of paper based records, electronic records, and official documents. During this data collection, interviews with archivists, records managers and even other employees working with records on a regular basis, were carried out. The second stage was to study records management in general, which includes both manual records management and use of electronic records and document management systems. This stage also contained studies of information sys-

² There are 21 police authorities in Sweden, which together with the National Police Board and National Laboratory of Forensic Science, form the National Police Service.

tems within the four organizations that managed information in a stipulated way.

The empirical material was analyzed in two steps. The analysis of the empirical material has similarities with the hermeneutical analysis described by Svensson & Starin (1994) where the analysis went from a holistic view to a detailed perspective iteratively. The analysis started by identifying detected record types. Each detected record was named after its content, or as the organization had named them.

The next step of the analysis process was to identify potential characteristics of the records, a methodological step similar to what Orlikowski (2002) and Strauss & Corbin (1998) describe as defining concepts. The potential characteristics were given descriptive names, and the names are not grounded in theory. A set of five characteristics was then further analyzed and compared with characteristics described within archival and recordkeeping theory.

3 Presentation and discussion of the results of the empirical and literature study

This section aims to answer the three research questions. First a summary of what organizations define as records is presented, followed by a presentation of the characteristics of these identified records. This section ends with a comparison between archival and recordkeeping theory-based characteristics and the empirically grounded characteristics in this research.

In Sweden no distinction is made between recordkeeping and archives in public organizations. A record is part of the archive at the moment it is created or received into an organization (Hörfeldt, 1998). Therefore most of the records that are presented in this section have been found in archives. It is also noticeable that Sweden also has a wider interpretation of public records. A public record in Sweden is all documents and records³ made or received in a public organization. The public have free access to public records, which is declared in the freedom of press act⁴. In private organizations and companies the term 'record' is not used often, they mainly use the term 'documents' instead, even if they have a physical archive. The wider interpretation of records within Swedish public organizations adds some bias to the empirical material that the reader must be aware of. This broader interpretation of what constitutes a record makes it possible that records that are not clearly bound to transactions are preserved and treated like all other records.

3.1 Records identified

³ The Swedish word "handling" is here translated to record instead of document.

⁴ TF 2kap (SFS 1949:105)

Table 1 presents a complete list of the records identified in the four different organizations.

Table 1.A complete list of all records identified

Organization	1	2	3	4	Organization	1	2	3	4
Staff information	x	x	x	x	Minutes of meeting	x	x	x	x
Account	x	x	x	x	Offer	x	x	x	x
Agreement	x	x	x	x	Orders	x	x	x	x
Archival description	x	x	x	x	Outer environment	x	x	x	x
Complaint	x	x	x	x	Policies	x	x	x	x
Contract	x	x	x	x	Postal giro and bank forms	x	x	x	x
Copies of receipt	x	x	x	x	Production	x	x	x	x
Correspondence	x	x	x	x	Propositions from the trade-union	x	x	x	x
Decisions	x	x	x	x	Purchases	x	x	x	x
Document plan	x	x	x	x	Regulations	x	x	x	x
Documentation/Archiving	x	x	x	x	Reports	x	x	x	x
Forms	x	x	x	x	Results/account documents	x	x	x	x
Goals and work of Improvements	x	x	x	x	Routines	x	x	x	x
Instruction/Order	x	x	x	x	Specification of traveling expenses	x	x	x	x
Invoices	x	x	x	x	Standards	x	x	x	x
IT/Computers/Telephone	x	x	x	x	Staff matters	x	x	x	x
Joint instructions	x	x	x	x	Statement of account	x	x	x	x
Legal information	x	x	x	x	Store	x	x	x	x
Lists of diary	x	x	x	x	Tenders	x	x	x	x
Lists of employee	x	x	x	x	Trade-union agreement	x	x	x	x
Lists of staff	x	x	x	x	Verification	x	x	x	x
Maintenance	x	x	x	x	Work of proposals	x	x	x	x
Minutes	x	x	x	x	Written report	x	x	x	x

Accommodations, decoration, and equipment	x			Laboratory reply		x		
Affirmation of grants			x	Legal cases from Supreme Administrative Court	x			
Application of allowance			x	Liquidation	x			
Ban on carrying on a business	x			Lubricant instructions		x		
Blueprints		x	x	Machine follow-ups		x		
Business cover paper	x			Main Ledger	x	x		
Care journals			x	Maps		x	x	
Certificate	x		x	Nature conservation		x		
Civil servant proposal			x	Work/performance plan		x		
Checklist	x	?	x	Notice of attending...			x	
Comments on a proposal circulated for consideration	x		x x	Official notes	x		x	
Commercial advertising	x			Outgoing request of...	x			
Commissions about exercise of public authority	x			Plans in detail			x	
Education material		x		Press cutting		x		
Emergency plan	x			Price lists				
Environment analysis		x		Process descriptions		x		
Environment management instructions		x		Process map/Master plan		x		
Environment reports		x	x	Project	x	x	x	
Excursions		x		Project documentation		x		
Exemption	x			Quality regulations documents		x		
Forest inventory reports		x		Random inspection compilations		x		
Forest rejuvenation		x		Refer to...for consideration	x		x	x
Forestry documentation		x		Register extracts	x			x
Geo analysis		x		Security and security management instructions/handbook	x			
Grant/permission application			x	Settlement of accounts	x			
Information campaign	x			Site Plans			x	

Information/answers from Authority	x			Staff accommodation		x		
Inquiry	x			Test results/analysis		x	x	
Inspection reports			x	Treaty revisions		x		
International cooperation	x			Utterance/statement to authorities and organizations	x		x	x
Inventory lists		x		Web structure	x			
Lists of revision	x							

The fact that the number of detected general records is less than the number of detected organization-specific records, may not be remarkable, but it shows a large variation and complexity of records. Table one visualizes and answers the first research question in this research; what is defined as a record in organizations?

3.2 Potential Characteristics

During the analysis phase aimed at describing the characteristics of records the transcribed material was read and all possible potential characteristics were noted, then grouped with similar characteristics on different levels. Some were seen as potential characteristics of records, some were seen as values of characteristics. There were also examples of characteristics with subcharacteristics. Therefore it is important to give details about all values identified in this phase of the research.. The potential characteristics, subcharacteristics, and their values are listed below in alphabetical order. In the following subsections all potential characteristics and their sub characteristics are presented. In the following subsections we answer the second research question: what are the characteristics of the records that organizations defined as records?

3.2.1 Cause

In this research different reasons for why records are preserved have been identified. This is sometimes explicit but often implicit. The identified causes are not mutually exclusive, and they are:

- Legislation;
- Business needs;
- Cultural and historical needs.

First all records preserved in public organizations are preserved because of legal demands. As mentioned above, citizens in Sweden have the right to access public

records. We also found records which are preserved because of other legislation, for example legislation for finance management, and the environment. Records have also been found to be preserved because of their business value.

Both the municipal and the multinational enterprise have examples of records which they think are unnecessary because they are never searched for again after their operational use has ended. Those records are saved because of legislation but obviously have low or no business value. In the municipality there were records about employees (Lists of employees), which record how many hours each employee has worked per month. The multinational enterprise declared that almost all long term preserved records were never used. For example, invoices were preserved due to legislation but some of the staff had difficulty seeing the usefulness of those records, when there was a copy of the invoice in the financial information system. This raises the issue of identifying and keeping the original as the archival record, but that is out of the scope for this paper. Even if evidence is one main cause for recordkeeping in archival and recordkeeping theory (e.g. Reed, 2005), this was not a stated reason within the organizations in this research.

3.2.2 Context

Context arises from the connection and coherence between the record and its creation, as well as connecting transaction to other related records. Context can be described with metadata.

The maturity of metadata use differs between the research sites. There were examples of no use of metadata, to fully automatic labeling of metadata. Every record contains of metadata. For example when filling in a formula every description of each field is metadata, i.e. data about the data (information) you are filling in. This leads to the conclusion that 'no use of metadata' is really no use of 'additional metadata'.

The 'context' and its sub characteristics can also be found as metadata elements in metadata standards e.g. VERS (Public Record Office Victoria, 2003), which also includes many other metadata elements, for example language, preservation history, and coverage.

The subcharacteristics within 'context' are:

- administrative process;
- transaction;
- general;

- what;
- when;
- where.

The 'administrative process' includes registration number, actions and case numbers.

The idea of the transaction is in recordkeeping and archival theory (International Council on Archives, 2000; McKemmish, 2002; McKemmish et al., 2005; Thomassen, 2001) almost a requirement for records. In this research both examples of a clear connection to transaction and almost no connection whatsoever have been found. For example, in police records, connections to transactions are visible, due to demands from the legal system. Both at the public company and at the multinational enterprise records were found with no connection to specific transactions. The subcharacteristic 'general' indicates the extent of metadata implementation. 'General' can take the values between the extremes 'no metadata use' to 'fully automated metadata use'. The name 'general' was chosen when metadata has been used as a tool to preserve the content at a general level. The subcharacteristic 'when' is related to time and had also been used for purposes other than recording when some action happened. There are examples of records including many 'when' values, for example when the record was created, when a person is born, when a specific situation occurred, etc. The subcharacteristic 'where' is a spatial value, and 'what' often contains descriptions of content. For example in the police, many records are based on standardized forms, where fields can be filled in or marked to tell the reader what the record is about. The 'what' characteristics can be used in some automated identification of the record's type.

3.2.3 Type of content

This potential characteristic is about the type of the content in records. The records that were found in this research had a variety of content types. In this research three values of 'type of content' have been found:

- original (raw);
- ordered;
- analyzed.

'Original' is when the information is saved as it appeared when it was created. Examples are test results from an analysis of chemicals or a hearing. In some cases the organization has 'ordered' the information for example an alphabetically ordered

list of employees. There were examples of ‘analyzed’ information such as statistical analysis presented as tables and graphs but with no connection to the original data. In the cases where the records are in raw form the connection to a transaction often is clear. A record with ‘original’ content is a record that is derived from a business transaction and the information has not been refined in any way. When information is ‘ordered’ or ‘analyzed’ there has been some kind of refinement involved. Those records are often composed from many information/data sources and in some case also several records. Those records have not always a clear connection to their main sources. For example the multinational enterprise can decide to either preserve the data from specimens taken, or to preserve the analysis. There is also an option to preserve both the raw data and the analysis of the data. If there has been an action of refinement there should also be a description of how, why, and by whom it has been done. As a potential characteristic, type of content is about differences in the origin of the contents, something that might not be specifically remarkable but can be important when, for example, the need for connection to the original source is needed.

3.2.4 Frequency of use

This potential characteristic consists of the following values:

- never used;
- annually used;
- used often;
- used daily;
- used very often.

All the studied organizations had separated their preservation of records physically. For example, in many organizations it was more or less a standard procedure that paper records created during the current year are kept within easy access. This construct was also used in the electronic system, where for example financial records follow the account period.

Records that are not used so often can be separated from one information system and kept in another information system (for example an e-archive). There will always be example of records that won’t be used again. When looking at similar types of records their usage may also differ. For example, not all environment analyses may be used again, but a few might. This depends on the situation and it is impossible to know exactly which record will be used again on that level. But this research shows examples of groups of records that are never used again, an example being the lists of employees in the municipality. The use discussed above is a record’s

primary use, i.e. to support some business activity. Whether or not a record can be used by future researchers (secondary use) has not been taken into account here. A relevant question here is why organizations keep records that are never used, and in some cases when no person knows why they are preserved. In some cases it is because of legal regulations, and by regulations from the Swedish National Archives, which look after future researchers' needs. When people in organizations are unaware of why records are preserved it highlights a lack of understanding of the archival value of the records by the employees concerned. In the multinational enterprise they reported that they feel insecure about whether or not they have a future need for those records, which results in more records than necessary being kept, which makes it harder to find records that they want to find.

3.2.5 Organizing

The organizing (i.e. intellectual control) of records has been a central concept for archival science. Here, 'organizing' is how records are organized in preservation both in the short term, middle term, and long term. The reason for this is that organizing has been the key to accessibility. In Sweden the use of the general archival register plan is widespread (e.g. Nilsson, 1983). Examples from the plan are that minutes of meetings are labeled with A, and economics labeled with G. Two problems that can often be detected in category F (documents ordered after business tasks) are that this category can be extensive and that it could include records from other categories. This category covers all records that are organized by case. All the organizations in the study have examples of records organized as case files. For example, in the municipality a case consisting of an official report regarding a school playground could include financial records, minutes of meetings, maps etc. In designing information systems it would be possible to organize the records in different ways depending on the situation. By using different interfaces the same record could be accessed both by content and by case.

The research has shown that organizations use different variants of organizational criteria for managing their archives. The Swedish National Archives' register plan is not mandatory for public organizations, it is only a recommendation. The differences in organizing are found both within the archives and in middle term preservation outside physical archives. The reasons for the choices sometimes seem to be random. In some divisions in the multinational enterprise there were individuals that kept and organized records on their own personal hard drives or in systems that are not formally part of the recordkeeping systems of the organization. This is seen as a problem by the organization because it limits access to records. Organizing has the following subcharacteristics:

- By cause;
- By types;
- By content;
- Personal;
- Multilevel.

The use of the register plan is placed within the subcharacteristic 'By content'.

3.2.6 Searchability

This characteristic consists of following values:

- No searchability (no search criteria);
- Low searchability (few search criteria);
- Good searchability (several search criteria);
- Group-based searchability;
- Indirect searchability.

This potential characteristic is dependent on computer-based information systems, either as an indexing system for paper based records, or as the system managing the electronic records.

There were no records found with unlimited search criteria as in a Google-search. A Google perspective influences expectation of searchability. In several cases the only possibility to access a record was by its unique key, which gives low searchability. In Sweden, for example, every organization has a unique id for their organization, and every citizen has an id-number which uniquely identifies every person. The experience of searchability is dependent on the users' experiences of search tools and archives. The recordkeeping staff may know exactly how to find a specific record but an external user may have serious trouble. In several of the studied archives searchability is indirect, which means that the end user has no access to archival records without help from archival staff. The values added to the characteristic have nothing to do whether full text searchability or search by controlled vocabulary is preferable. This debate is not within the content and aim of this paper.

Records are usually organized in groups. An example is that all minutes of meetings from a certain year are grouped together, sometimes with no possibility to search for a specific minutes of meetings.

3.2.7 Structure

Within this characteristic two extreme values have been found:

- full flexibility;
- standardized structure.

Structure can be seen as the level of formalization, and is about the structure of the records themselves, not of the system managing the records. For example the information system within the police service the system managing police reports compels the user to write different types of information in different fields, for example the system does not accept a string of text in a field which should contain a telephone number. This makes all records of this type (police reports) in the system conform to this highly formalised structure. Within the police systems these constraints also existed when police reports were written on paper. The other value can be found in the municipality, which has no mandate to influence how received records are structured. For example, a citizen can handwrite a complaint about some matter as well as use some formalized templates. Both these types of records structures have to be dealt with, even in an electronic information system.

In this research those extremes were also found in electronic records. But the majority of the electronic records had a more or less standardized structure. In the organizations within this research the majority of electronic records were in the form of electronic documents and formula.

3.2.8 Version and copy

Some records are copied and are in that case labeled with a clear indication of this. Copies of records were for example found when one record served as the basis for decision. Added to the newly created record about the decision taken, a copy was laid, to increase the understanding of the decision. Normally this is a conscious action, e.g. when the organization wants to distribute multiple copies of a record. Creation of multiple copies is not necessary when the record is electronic. It is possible to access the original record from many places, and limitations are only dependent on the information system. But if the system is to meet recordkeeping requirements, the system design must ensure that contextual connections are clearly maintained. Regarding version control, the present study has identified three different values in the research sites:

- Complete traceability of version;
- Partial traceability of version;

- Non-traceability of version.

Complete traceability includes all versions of records, even if there are records that are not correct. For example, if the minutes of a meeting fail to correspond to what was decided in the meeting, both the original and the revised versions of the record are preserved. In the organizations where no traceability was found, only the valid and latest version of the record was preserved.

3.3 Analysis of potential characteristics

From the empirical material eight different potential characteristics were identified, which are all presented above. This presentation of potential characteristics also answers the two first research questions in this research: What is defined as a records in organizations? and what are the characteristics of these records? The characteristics are based on analysis of the empirical data. After a careful analysis of each of those characteristics it was found that it was only possible to name five of the above as general characteristics of records. Those were:

- Context
- Form
- Structure
- Organizing
- Version/Copy

These characteristics can help to identify, tell apart and make records recognizable.

The other three, 'searchability', 'frequency of use', and 'cause' were not possible to define as characteristics. 'Searchability' and 'frequency of use' are both affected by how well the above five listed characteristics are implemented. 'Searchability' for example is both dependent on how records are organized, to what degree metadata is used and how structured a record is. 'Frequency of use' is dependent on the content and users' interest in using the record, and can hardly be defined as a characteristic. 'Cause' is possible both to interpret as embedded within the characteristic 'context', but can be interpreted as implicit metadata.

As mentioned earlier, one of the tasks for a record is to serve as evidence about means of actions or transactions. The evidential value of a record is a function or effect of the presented characteristics. Evidentiality can be better described as a meta-characteristic of records. Evidence is a product of good recordkeeping.

3.4 Comparison of findings with archival theory

Which of the empirically identified characteristics can also be found in archival theory? And are there any characteristics found in this research that are not to be found in the literature of archival theory? This is the last research question in this paper. Table 2 visualizes this comparison. Only the first and second level characteristics are presented in the table, even if the comparison was made with all levels.

Table 2. A summary of empirically based findings compared with theoretically identified characteristics of records

First level	Second level	Reference
Context		Thomassen (2001), Hoffman (1998), Shepherd and Yeo (2003), Duranti (1997, 2001b), Guercio (1997), Bearman & Trant (1997), Bearman (1997), Duff & Harris (2002), Upward, McKemmish (2001), Hartland et al (2005), Gilliland-Swetland (2000)
	Transaction	Thomassen (2001), ICA (2000), ISO 15489 (2001a), Duranti (1997, 2001b), Guercio (1997), Bearman & Trant (1997), McKemmish & Upward (1990, 2001), Dollar (1992), Gilliland-Swetland (2000), Reed (2005), Menne-Haritz (2001)
	General	Bearman & Trant (1997), Bearman (1997)
	Administrative process	Duranti (1997, 2001b), Guercio (1997), Bearman (1993), Reed (2005), ISO 15489(2001a), Gilliland-Swetland (2000)
	What	
	When	Duranti (1997)
	Where	
	Who	Guercio (1997), Duranti (1997)
Type of content		Gränsström et al(2000)
	Original	
	Ordered	
	Analyzed	
Structure		Duranti (2001b), Bearman & Trant (1997), Bearman (1993), Bearman (1997), Gilliland-Swetland (2000, 2005)
	Standardized	
	Full flexibility	
Organizing		Reed (2005), Bearman (1993), Gilliland-Swetland (2000)

	By Cause	
	By Types	
	By Content	
	Personal	
	Multilevel	
Version/ Copy		Duranti (1997)
	Complete Traceability of version	
	Partial traceability of version	
	Non-traceability of version	Reed (2005)

Conclusions

4.1 Naming the characteristics

The empirically grounded characteristics were given descriptive names without any reference to theory or standards. When comparing these identified characteristics with characteristics discussed in the theoretical literature, it became clear that there are discrepancies in the labeling of characteristics of records. In this section those differences will be discussed. It is important to note that, irrespective of differences in labels, the comparison is based on the underlying meaning and content of a characteristic and not merely on the literal labels. Discrepancies in naming have been identified with regard to structure, organizing and type of content.

Reed (2005) uses different labels to describe what we have called organizing. She uses the terms 'grouping', 'classification arrangement' and 'sequencing'. However the difference, if any, between those concepts is not clearly stated. Reed gives examples of sequencing which includes serial, sequence, and dossier. Further she states that which type of grouping/sequence/arrangement to use depends on the situation and different types have their benefits and weaknesses. Classification, arrangement and grouping are also the concepts that are used by Gilliland-Swetland (2000). She gives an example of classification by subject content, which has also been identified

in our study in the characteristic ‘organizing’ with subcharacteristic ‘by content’. Gilliland-Swetland (2000) also writes about hierarchies to build structures to provide access to material. To sum up this discussion, two main reasons can be identified as to why this concept of organizing is important within recordkeeping. The first is to make it possible to retrieve records, which requires some sort of index or classification. The other reason is managing organized records that can be demonstrated to be trustworthy enables interpretation of the records authenticity and reliability.

Regarding the characteristic structure, Bearman (1997, p.272) also used this term in the following citation: “Records are evidence when they are bound to appropriate metadata about their content, structure and context”. This is a rather common view within archival science. The question is what is meant by ‘structure’ in this context. Structure is a concept that deals with appearance of a record, which includes the material of a document, character sets and how the record is arranged with headings and paragraphs. Our view of structure is a little bit narrower than this more general description of structure. Gilliland-Swetland (2000) writes about structure in the same sense that we do. She writes that XML has been used to develop structures that are more predictable, which could be helpful for example with web resources. Her discussion also includes a description of how diplomatics uses structure to investigate the authenticity of records. Some records have been found to consist of the same type of elements and are built up the same way. But she also writes that structure could be both intellectual and physical, particularly in the sense of hierarchies with series and subseries. The latter is not included in our definition of structure but rather connected to our concept of organizing.

4.2 Comparison of characteristics

The empirically grounded characteristic ‘context’, with its sub levels ‘transaction’ and ‘administrative process’, together with the empirically grounded characteristic ‘structure’ have been found to correspond with recordkeeping and archival theory. This correspondence was not surprising since records are often described as consisting of context, content and structure and these characteristics are parts of both the ISO15489 and ICA definition of records (International Council on Archives, 2000; International Standards Organization, 2001a).

‘Context’ has been in focus in several research and standardization projects. This has resulted in several existing metadata schemas that cover different aspects of context (for example what, when and who). The present study has not made any attempt to compare different metadata schemas to find out which one most closely corresponds to our empirical result. The theoretical viewpoint regarding ‘structure’

is that computer-based information systems require a higher level of formalization of the structure than paper-based solutions. But some theoreticians emphasize the need for preservation of the original structure see e.g. (Bearman, 1994, 1997, 1999; Bearman & Trant, 1997; Duranti, 1997, 2001a, 2001b). Some organizations are obligated to preserve records even if they are unstructured and miss important metadata such as author, and date. One example of this is a municipal organization that has to preserve an anonymous letter from a citizen. If the municipality has implemented a digital recordkeeping system which requires some metadata that is not present in this kind of record (in this case the author, since it is an anonymous letter), should the organization decide not to preserve this record? The authors argue that the formalization process should not allow this kind of development. It must be possible to manage even records that are unstructured in a digital environment. The system should not force the organization to make changes to the record. An important question is; if the original structure is affected by formalization that is required by a digital system, will it influence the trustworthiness of the record? This issue is one part of the debate concerning the preserving of records with their original look/appearance. Is it good enough to preserve the content of the record or must the appearance remain exactly the same as the original? Formalization has both positive and negative aspects. The positive aspect is that it forces organizations to work with standardization of records and even standardization of working procedure. This situation is feasible in organizations striving to handle all citizens or customers the same way. Formalization is also a key in situations where organizations want to exchange information with each other. One negative aspect is that the formalization process may influence decisions about destruction and preservation. Maybe this could force organizations to base decisions about destruction on what kind of records can be part of an electronic document and workflow system, instead of be based on traditional appraisal methods. It would then lead to earlier decisions about destruction, which according to ISO 15489 and confirmed by the authors observations, has to be done early in system development. Another negative aspect is that formalization may lead to a situation in which content that earlier was part of the records, such as hand written notes, will not be preserved in the future. Regarding the characteristics 'organizing', our reflection is similar to that which McKemmish *et.al* (2005) argue. That is, within a computer-based information system it is possible to use several simultaneous methods of organization. By adding metadata (preferably automatically) it could be possible to design a system that lists all records by content or by types depending on what the user wants to do. Gilliland-Swetland (2000) writes about a logical method of organizing which replaces the physical way of organizing records.

Concerning the characteristic 'type of content' there are few examples about this phenomenon in the literature. This is however not surprising, because recordkeeping is about recording transactions in an organization, giving different content. If the record is a presentation of a statistical analysis without reference to the raw data upon which the analysis was based, the record is evidence only that the organization has made this analysis. If the record includes the raw data and the methodology by which the data upon which this fictive analysis were obtained, the record provides evidence of the transaction in which the organization has made some sort of investigation. A possibility is of course also that both those records are preserved based on the decision that the organization wants to preserve both. The recordkeeping system is in this sense caring about how the record looked originally, not only about the way it has been refined. But, information systems very often include different combinations of data, extractions of data, various types of presentation of data and so on. In Sweden there is currently an ongoing debate that relates to this characteristic and it is about potential records. Digitization has thus revealed a problem that needs to be discussed during appraisal. Gränström *et al* (2000) have a discussion about potential records in electronic environments in their book. They emphasize the need for decisions about preservation in the design phase of information system.

The present study shows that several organizations are concerned about problems with version and copy control. In many organizations there is a big proportion of the staff that have rights to create records. The multinational enterprise was a good example of this. About ten years ago there were a lot of records that were classified as correspondence. Most often the letters were written by a secretary and on letterhead papers with signatures. Today most of the employees write those types of letters themselves, often by e-mail. Version and copy control is something that is everybody's responsibility. It is important that any system that includes records is designed to ensure control over versions and copies.

Reed (2005) argues that whether an organization chooses to preserve drafts, final versions and copies is a decision has to be made based upon the organization's needs. There is nothing within the archival theory that clearly states which one is the right way to do it. Duranti (2001a) argues that authenticity partly depends on the state of transmission (i.e. draft, original or copy), which has to be clear.

The results of this study show that the four organizations have reached different levels of implementation of recordkeeping requirements. The recordkeeping requirements have, as shown, a theoretical base, sometimes founded upon a viewpoint derived from paper-based records. As discussed above there are reasons to question some of those theoretical requirements, but there are also reasons to point out weak-

nesses in the organizations studied. There are several examples of unwanted record-keeping practices in the studied organizations. One example is personal organizing of records that was found in the multinational enterprises and something that is not recommended and not even mentioned theoretically. Other similar unwanted recordkeeping practices are; examples of no added metadata, no separated metadata, no connection to transaction and lack of full traceability of versions.

A general conclusion is that the empirical data are based on more detail than is presented in recordkeeping and archival theory. Table 2 clearly visualizes that 'context' stands out as the most described and scientifically penetrated characteristic in recordkeeping and archival theory.

Table 1 is a summary of the different records that were detected in the four different organizations that were studied. According to e.g. Thomassen (2001) archives consist of records and, implicitly, only records should be found in archives. Many of the records listed in table 1 have been found in archives. A record is supposed to be created in a work process and should have a relation to transaction of business (International Council on Archives, 2000; McKemmish et al., 2005; Thomassen, 2001), see Table 2 for more references. In this research some records which can be questioned if one literally uses the definition of records have been found in archives.

For example:

- Educational material in covers, materials from courses in which some employees have participated.
- Postal giro and bank forms.
- Lists of staff, a summary of all employees at a specific time.
- Information campaign material. Material produced by an advertising agency. Web structure, documentation over the internal web structure. Press cuttings, where the archive owner were mentioned.
- Legal information, produced by the ministry of Justice
- DAFA Announcements. DAFA is a Swedish company supplying and selling public information from different public databases.
- Copies of legal cases from the Supreme Administrative Court.
- IT/Computer/Telephone equipment descriptions, for example information about which equipment is installed within the organization and where.

Based on previous presented criteria, and record definitions it is not fully obvious

that the list above can be defined as records. They were found with no visible connection to any transaction or work process. Within the studied organizations such connection might exist, but the context was not clear and visible by just studying the record. For example the educational material could have been the result of a procurement performed within the organization, but this example had no such contextual connection, which made such interpretations impossible. Another example is the information concerning IT/Computers/Telephone. Those records should have their justified existence in the archive if they had any connection or relation to for example offer, order, decision, and the invoice. By following archival definitions from e.g. ISO15489 (International Standards Organization, 2001a) or ICA (International Council on Archives, 2000) the relationship to a transaction is one of the components that separates records from documents and information.

The identified difference between practice and theory has several explanations. One thing that influences this is the discrepancy between what theories would define as a record and what is preserved in Swedish archives. The Swedish archival register plan has in some cases lead to organizing of archival material that does not keep links between records and transactions. There is, however, a changeover to an enterprise and process-oriented view of record keeping in Sweden that has similarities to the development in for example Australia. The Swedish National Archives⁵ is working on a new archival description, which will strengthen the contextual cohesion between a record and the processes wherein the record has been created. This is dependent on legislation in Sweden. Another possible explanation is that we have studied organizations, which do not meet best practice standards in recordkeeping. Bearman and Trant (1997) write about better/worse record, which is dependent on complete/incomplete metadata. Further, they write that a solution might be to make those responsible for record creation aware of the consequences of bad records. Some organizations in this research have mentioned problems with employees that do not understand the importance of good record keeping. They think that it is a matter that takes too much of their time. Which of these explanations is nearest to the truth is irrelevant to this discussion. The scenario is rather that both the explanations given above contribute to making the discrepancy between theory and empirical data clear. One thing that we want to emphasize is that organizations that plan to make parts of their recordkeeping digital should take into consideration those differences between theory and the practices observed in this case study. Standards or other tools can require that recordkeeping systems are designed in a certain way, which may not be the case.

5 <http://www.ra.se/ra/index.html>

4.3 Characteristics and the proactive approach

The characteristics of records presented above are implicitly also characteristics that make a record unique in relation to other types of information, in other words recordkeeping and archival demands concerning records that must be fulfilled. This statement argues for, and supports, the necessity of working proactively when dealing with electronic records. In a paper-based environment it is possible to recreate some of these characteristics afterwards. But when records are electronic and are managed within information systems any such strategy is likely to be impossible. Even if such activities are possible, the benefits and positive effects of using information systems can be lost if administrative and manual tasks have to be performed just to fulfill archival demands. Therefore the proactive approach, compliant with ISO15489 and suggested earlier in this paper, is necessary. In a proactive approach these characteristics are possible to interpret as basic requirements which need to be fulfilled at the design stage of information systems that manage electronic records. The ISO15489 Standard is influenced by the records continuum model: a model that sees records in a continuum of space time without a linear time perspective, contrary to the life cycle model where time is linear. The continuum model implicitly proposes a proactive approach in which record appraisal must be performed during the design phase (McKemmish, 2002; Upward, 2004)

4.4 Demarcation

An important question to discuss here is how the selection of organizations has influenced the result. Organizations within the public sector have to follow a more extensive regulation than private organizations when it comes to archives. This study has had no intention of showing differences between private and public organization but this would be an interesting future research question. The results give a richer picture of public organizations than private organizations. The private organization that has been studied is ISO 9001 certified and it is impossible to say how this has affected the result. ISO 9001 includes routines for document control, which sometimes is in accordance with archival requirement but is sometimes contradictory.

4.5 Potential application

Within archival science this research might open up a creative discussion about the need for a more detailed investigation and description of the record and its characteristics. To be able to build information systems that manage electronic records the record itself must be formalized to some extent. The formalization is necessary for automatic computer based management and processing. When designing informa-

tion systems that aim to take the full extent of archival requirements into consideration, knowledge about the characteristics of records could be helpful to achieve the objectives of records; to be reliable and authentic, be able to serve as evidence, and to support accountability. Table 1 also indicates the importance of carefully identifying general and organizationally specific records before a design and development process starts. By this discussion a potential application area could be the design and development processes of information systems that are going to manage electronic records.

4.6 Concluding remarks

The characteristics of records, based on this study are; context, type of content, organizing, structure and version/copy. The comparison with theory has shown that all characteristics except type of content are discussed within the literature. Although the characteristics have been identified both empirically and theoretically this research has shown some discrepancies when penetrating the characteristics. Those differences can be explained by different causes that have been discussed in detail in our paper. However, some of the main reasons are dissimilarities in legislation and recordkeeping traditions between countries and immaturity in implementation of record keeping. Those explanations are also valid for differences in which kind of records that are preserved in the four organizations.

Influencing organizations to adopt archival requirements is an important mission and there is a need for further research to suggest solutions based on organizations' individual situations. Both the list of identified records (section 3.1) and the list of identified characteristics (section 3.3) indicate that recordkeeping is implemented differently in the four organizations. Our research shows both examples of recordkeeping that have high correspondence to recommendations in theory and standards but also examples that do not follow those recommendations. An important remark is that if an organization plans to develop a new recordkeeping system, it must not be based on an existing but poor recordkeeping system that does not comply with accepted standards. These kinds of discrepancies will, if not taken into consideration, make the system development difficult. The step from manual handling of records to digital handling, which includes new routines, can be a big step to take, especially if the organization is unaware of this problem. Situations like this could then lead to a more extensive need of education than planned.

This study has shown examples of organizations that work with digital information systems with automated capturing of metadata and other kinds of automated

functionality. But there are also examples within these organizations' information systems that include records that do not meet recordkeeping requirements as well as manual systems with low maturity. The conclusion here is that in some senses, some of the organizations in this study have a lot to do before reaching good record keeping.

Acknowledgments

We would like to acknowledge our appreciation to Dr. Karen Anderson for her support in finishing this paper. Karen Anderson has guided us deeply in archival and record theories as well in correcting the language.

References

- Bearman, D. (1994). *Electronic evidence: Strategies for managing records in contemporary organizations*. Pittsburgh: Archives and Museum Informatics.
- Bearman, D. (1997). Capturing records' metadata: Unresolved questions and proposals for research. *Archives and Museum Informatics*, 11(3-4), 271-276.
- Bearman, D. (1999). Reality and chimeras in the preservation of electronic records, *D-Lib Magazine* (Vol. 5).
- Bearman, D., & Trant, J. (1997). Research: Towards testable specificity. *Archives and Museum Informatics*, 11(3-4), 309-321.
- Cox, R. J. (2001). *Managing records as evidence and information*. Westport, Conn.; London: Quorum Books.
- Dollar, C. M. (1992). *Archival theory and information technologies: The impact of information technologies in archival principles and methods*. Macerata: Univ. of Macerata.
- Duff, W. M., & Harris, V. (2002). Stories and names: Archival description as narrating records and constructing meanings. *Archival Science*, 2(3-4), 263-285.
- Duranti, L. (1997). The archival bond. *Archives and Museum Informatics*, 11(3-4), 213-218.
- Duranti, L. (2001a). Concepts, principles, and methods for the management of electronic records. *The Information Society*, 17, 271-279.
- Duranti, L. (2001b). The impact of digital technology on archival science. *Archival Science*, 1(1), 39-55.
- Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532-550.

- European Commission. (2002). *Model requirements for the management of electronic records*. Luxembourg: Office for Official Publications of the European Communities.
- Gilliland-Swetland, A. J. (2000). *Enduring paradigm, new opportunities: The value of the archival perspective in the digital environment*. Washington, D.C.: Council on Library and Information Resources.
- Grimson, J. (2001). Delivering the electronic healthcare record for the 21st century. *International Journal of Medical Informatics*, 64(2-3), 111-127.
- Gränström, C., Lundquist, L., & Fredriksson, K. (2000). *Arkivlagen: Bakgrund och kommentarer* (2., omarb. uppl. ed.). Stockholm: Norstedts juridik.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, Calif: Sage.
- Guercio, M. (1997). Definitions of electronic records, the european perspective. *Archives and Museum Informatics*, 11(3-4), 219-222.
- Hartland, R., McKemmish, S., & Upward, F. (2005). Documents. In S. McKemmish, M. Piggott, B. Reed & F. Upward (Eds.), *Archives: Recordkeeping in society* (pp. 348p). Wagga Wagga: Charles Sturt University, Centre for Information Studies.
- Hartman, J. (1998). *Vetenskapligt tänkande: Från kunskapsteori till metodteori*. Lund: Studentlitteratur.
- Hofman, H. (1998). Lost in cyberspace: Where is the record? In K. Abukhanfusa (Ed.), *The concept of record: Report from the second stockholm conference on archival science and the concept of record, 30-31 may 1996* (pp. 115-129). Stockholm: Swedish National Archives.
- Holgersson, S. (2001). *It-system och filtrering av verksamhetskunskap: Kvalitetsproblem vid analyser och beslutsfattande som bygger på uppgifter hämtade från polisens it-system*. Linköping: Univ., Institutionen för datavetenskap.
- Hörnfeldt, T. (1998). The concept of record: On being digital. In K. Abukhanfusa (Ed.), *The concept of record: Report from the second stockholm conference on archival science and the concept of record, 30-31 may 1996* (pp. 67-75). Stockholm: Swedish National Archives.
- Iacovino, L. (2004). Trustworthy shared electronic health records: Recordkeeping requirements and healtconnect. *Journal of Law and Medicine*, 12(1), 40-59.
- International Council on Archives. (2000). *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.
- International Standards Organization. (2001a). *Iso 15489-1. Information and documentation and records management part 1: General*. Geneva: International Standards Organization.

- International Standards Organization. (2001b). *Iso 15489-2. Information and documentation — records management — part 2: Guidelines*. Geneva: International Standards Organization.
- Lee, A. S. (1989). A scientific methodology for mis case studies. *MIS Quarterly*, 13(1), 33-50.
- McKemmish, S. (2002). Placing records continuum theory and practice. *Archival Science*, 1(4), 333-359.
- McKemmish, S., Piggott, M., Reed, B., & Upward, F. (Eds.). (2005). *Archives: Record-keeping in society*. Wagga Wagga: Charles Sturt University, Centre for Information Studies.
- Menne-Haritz, A. (2001). Access-the reformulation of an archival paradigm. *Archival Science*, 1, 57-82.
- Myers, M. D., & Avison, D. E. (Eds.). (2002). *Qualitative research in information systems: A reader*. London: Sage.
- Nilsson, N. (1983). *Arkiv i förvandling: Studier i arkivens teori och tillgänglighet*. Solna: Nordiska bokh. (distr.).
- Orlikowski, W. J. (2002). Case tools as organizational change: Investigating incremental and radical changes in systems development. In M. D. Myers & D. E. Avison (Eds.), *Qualitative research in information systems: A reader* (pp. 181-221). London: SAGE.
- Public Record Office Victoria. (2003). Vers metadata scheme pros 99/007 (version 2) specification 2. Retrieved 16 Mars, 2005, from http://www.prov.vic.gov.au/vers/standard/pdf/99-7-2_Std_ver2-0.pdf
- Reed, B. (2005). Records. In S. McKemmish, M. Piggott, B. Reed & F. Upward (Eds.), *Archives: Recordkeeping in society*. Wagga Wagga: Charles Sturt University, Centre for Information Studies.
- Shepherd, E., & Yeo, G. (2003). *Managing records a handbook of principles and practice*. London: Facet Publishing.
- Shipman, F. M., & McCall, R. (1994). *Supporting knowledge-base evolution with incremental formalization*. Paper presented at the CHI 94, Boston, Massachusetts.
- Shipman, F. M., & McCall, R. (1999). Incremental formalization with the hyper object substrate. *ACM Transactions on Information Systems*, 17(2), 199-227.
- Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2. ed.). Thousand Oaks, Calif.: SAGE.
- Svensson, P.-G., & Starrin, B. (Eds.). (1994). *Kvalitativ metod och vetenskapsteori*. Lund: Studentlitteratur.
- Teorey, T. J. (1999). *Database modeling & design* (3. ed.). San Francisco, CA: Morgan Kaufmann.

- Thomassen, T. (2001). A first introduction to archival science. *Archival Science*, 1, 373-385.
- Upward, F. (2004). The records continuum and the concept of an end product. *Archives and Manuscripts*, 32(1), 40-62.
- Walsham, G. (2002). Interpretative case studies in is research: Nature and method. In M. D. Myers & D. E. Avison (Eds.), *Qualitative research in information systems: A reader* (pp. 101-113). London: SAGE.
- Yin, R. K. (2003). *Case study research: Design and methods* (3. ed.). Thousand Oaks, Calif.: Sage Publications.

Paper II



Metadata use in document management systems which support business processes

Öberg, L-M (2007). Metadata use in document management systems which support business processes. Accepted for publication in Archives and Social Studies. A journal of interdisciplinary research.

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 arises¹. 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 writes that a pro-active approach is essential to be able to preserve (at reasonable cost) accessible and interpretable information for the long-term.² This is also supported by Shepherd and West³ 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 ⁴, the capturing of metadata at the cre-

1 Gilliland-Swetland, Anne. J. "Setting the stage". In: M. Baca (Ed.), Introduction to Metadata: Pathways to Digital Information. Getty Research Institute, 2000 pp 1-12. ISBN 0-89236-533-1

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

3 Shepherd, Elisabeth., & West, Victoria. "Are ISO 15489-1:2001 and ISAD(G) compatible? Part 1 In: Records Management Journal 2003, 13(1), pp.9-23.

4 Hedstrom, Margret. "Recordkeeping Metadata. Presenting the Results of a Working Meeting". In:

ation 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 that 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 system⁵ and therefore this study covers electronic 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.

Archival Science 2001, 1, pp.243-251.

⁵ For definition and discussion of record systems see the next section.

Theoretical standpoint and related research

The Australian School on recordkeeping that is based upon the Record Continuum model⁶ makes a distinction between record systems and electronic document management systems. According to Upward⁷ a characteristic of the Continuum Model is the view of records⁸ as unstable. There is no such thing as end products that should be delivered to an archival institution and this arises a need for continuing addition of process metadata meanwhile they change during space-time. Upward⁹ writes that records can have multiple lives in space-time, and a record are 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 is it important that consider the long-term perspective already in a document management system if it is where the information is created and managed. The Record 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 Borglund¹⁰ in their study of four Swedish organizations. All those Swedish organizations have not yet implemented the record 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 this differences between the theoretical viewpoint and reality has been revealed there is a need of studying how this influence electronic documents management systems within Swedish organizations. As mentioned the Australian school makes a distinction between electronic document management systems and records system.

There is research in this field that focuses on development of metadata specifications

6 Upward, Frank. (2005). "The records continuum". In McKemmish, Sue; Piggott, Michael; Reed, Barbara; Upward, Frank. (Eds.), *Archives: Recordkeeping in Society* Quick Print. 2005 pp. 348. ISBN: 1-876938-84-6

7 Ibid

8 For discussion about differences between record and information see section Definitions.

9 Upward, Frank.. "Modelling the continuum as paradigm shift in recordkeeping and archiving processes, and beyond - a personal reflection". In: *Records Management Journal* 2000, 10(3), pp. 115-139

10 Öberg, Lena.-Maria.; Borglund, Erik. "What are the Characteristics of Records? In: *International Journal of Public Information Systems* 2006 (1), pp. 55-76.

(see for example ¹¹ and there are examples of comparison between different metadata sets ^{12, 13 14}. Päivärinta, Tyrväinen & Ylimäki¹⁵ 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 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 Murphy¹⁶ writes about the lack of empirical understanding of the metadata use within organizations. Day¹⁷ express 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.

11 McKemmish, Sue; Acland, Glenda.; Cumming, K.; Reed, Barbara.; Ward, Nigel. "Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema". *Archivaria*, 1999, 48.

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

13 Duff, Wendy. "Evaluating Metadata on a Metalevel". *Archival Science*, 2001, 1, pp.285-294.

14 Shepherd, Elisabeth., & West, Victoria. "Are ISO 15489-1:2001 and ISAD(G) compatible? Part 1". *Records Management Journal*, 2003, 13(1), pp.9-23.

15 Päivärinta, Tero., Tyrväinen, Pasi.; Ylimäki, Tanja. (2002). "Defining Organizational Document Metadata : A Case Beyond Standards" In *European Conference on Information Systems*, Gdansk, Poland, 2002, pp.1154-1163.

16 Murphy, Lisa. D. (1998). *Digital "Document Metadata in Organizations: Roles, Analytical Approaches, and Future Research Directions"*. In *Hawaii International Conference on Systems Sciences*, Hawaii, 1998, pp. 267-276.

17 Day, Michael. "Preservation Metadata Initiatives: Practicality, Sustainability, and Interoperability". In: *ERPANET Seminar at the Archives School Marburg* 3-5 September 2003, pp.1-16.

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" ¹⁸. 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^{19 20 21}.Cox²² stated that the evidential value of a record could only exist if the content, structure and context are preserved.

18 Eriksson, Darek. M. (2004). Four Proposals for Enterprise Modeling. Ph. D Thesis, Göteborg,,Chalmers University of Technology,2004. ISSN 0346-718X:2112

19 Bantin, Philippe. "Electronic Records Management - A Review of the Work of a Decade and a Reflection on Future Directions". In: Encyclopedia of Library and Information Science, 2002, 71(Supplement 34), pp.47-81.

20 Duranti, Luciana. (2001). "Concepts, Principles, and Methods for the Management of Electronic Records." In: The Information Society,2001, 17, pp.271-279.

21 Thomassen, Theo. "A First Introduction to Archival Science". Archival Science, 2001,1, pp.373-385

22 Cox, Richard. J. Managing records as evidence and information. Westport, Conn. ; London: Quorum Books,2001. Electronic source.ar

Records vs. Document

According to ISO-15489-1²³ 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 items²⁴. Reed 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”²⁵. 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”²⁶.

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 ex-

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

24 Reed, Barbara. (2005). “Records”. In McKemmish, Sue; Piggott, Michael; Reed, Barbara; Upward, Frank. (Eds.), Archives: Recordkeeping in Society Quick Print. 2005 pp. 348. ISBN: 1-876938-84-6

25 Sundgren, Bo., & Steneskog, Gösta. “Information Systems for Concerted Actions”. In Sundgren, Bo; Mårtensson, Pär; Mähring , Magnus; Nilsson, Kristina. (Eds.), Exploring Patterns in Information Management - Concepts and Perspectives for Understanding IT-Related Change. Stockholm: The Economic Research Institute Stockholm School of Economics, 2003. ISBN 91-7258-63-1

26 Gilliland-Swetland, Anne. J. “Setting the stage”. In: M. Baca (Ed.), Introduction to Metadata: Pathways to Digital Information. Getty Research Institute, 2000 pp 1-12. ISBN 0-89236-533-1

ample²⁷ and²⁸. Hofman²⁹ and Day³⁰ 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. Hofman³¹ 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 model³².

Method

The method used for this study is a qualitative interpretative method. The course of action consists of two main parts. The first part consisted of empirical case studies and the second analyses differences between the empirical data and metadata set.

Data Collection

The empirical study was carried out in two organizations. 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 author-

27 Ibid

28 Lagoze, Carl; Lynch, Clifford. A.; Ron Daniel, J.” The Warwick Framework A container Architecture for Aggregating Sets of Metadata”. Accessed: 04-05-2006. <http://www.ifla.org.sg/documents/libraries/cataloging/metadata/tr961593.pdf>

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

30 Day, Michael. “Integrating metadata schema registries with digital preservation systems to support interoperability: a proposal”. In; Dublin Core Conference: Supporting Communities of Discourse and Practice - Metadata Research and Applications, Seattle, Wa, USA, 2003, pp.1-8.

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

32 Ibid.

ity in Sweden whose main business is administration and legislation. Organization B is a result-oriented section of a public company. Their main business activity is maintenance of infrastructure.

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 VERS³³, RKMS³⁴ and ISAD(G)³⁵. 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 is typically business oriented and ISAD(G) represent 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

VERS³⁶: 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

33 Public Record Office Victoria, P. Victorian Electronic Records Strategy - Forever Digital. Accessed 26-06-2006. www.prov.vic.gov.au/vers/vers/default.htm

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

35 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.

36 Public Record Office Victoria, P. Victorian Electronic Records Strategy - Forever Digital. Accessed 26-06-2006. www.prov.vic.gov.au/vers/vers/default.htm

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 Day³⁷, in opposition to RKMS, a schema based on a practical base.

RKMS³⁸: The Australian Recordkeeping Metadata Schema (RKMS) is the main deliverable from the SPIRT Recordkeeping Metadata Research Project³⁹. According to McKemmish et al.⁴⁰ 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^{41 42} RKMS is conceptual in its nature and this includes for example possibilities to include other metadata schemas in RKMS.

ISAD(G)⁴³: 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 Britain⁴⁴. According to Duff 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

37 D Day, Michael. “Integrating metadata schema registries with digital preservation systems to support interoperability: a proposal”. In: Dublin Core Conference: Supporting Communities of Discourse and Practice - Metadata Research and Applications, Seattle, Wa, USA, 2003, pp.1-8.

38 Recordkeeping Metadata Project (2006). Retrieved 20060626, 2006, from Recordkeeping Metadata Project Accessed: 26-06-2006 .<http://www.sims.monash.edu.au/research/rcrg/research/spirt/>

39 McKemmish, Sue., Acland, Glenda., Cumming, K., Reed, Barbara., & Ward, Nigel. “Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema.” *Archivaria*, 1999, 48.

40 Ibid.

41 D Day, Michael. “Integrating metadata schema registries with digital preservation systems to support interoperability: a proposal”. In: Dublin Core Conference: Supporting Communities of Discourse and Practice - Metadata Research and Applications, Seattle, Wa, USA, 2003, pp.1-8.

42 Day, Michael. “Preservation Metadata Initiatives: Practicality, Sustainability, and Interoperability”. In: ERPANET Seminar at the Archives School Marburg 3-5 September 2003, pp.1-16.

43 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.

44 Duff, Wendy. (2001). Evaluating Metadata on a Metalevel. *Archival Science*, 2001, 1, pp.285-294.

elements are divided into seven group; identity statement area, context area, content and structure area, conditions of access and use area, allied materials area, notes area and description control area ⁴⁵.

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 13 600 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.

45 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.

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 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 represent 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 include⁴⁶. 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

46 Day, Michael. "Integrating metadata schema registries with digital preservation systems to support interoperability: a proposal". In; Dublin Core Conference: Supporting Communities of Discourse and Practice - Metadata Research and Applications, Seattle, Wa, USA, 2003, pp.1-8.

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 project⁴⁷ 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

47 McKemmish, Sue., Gilliland-Swetland, Anne. J., & Cunningham, Adrian. (2006). "*Clever Recordkeeping Metadata Project*." Accessed 04-08-2006. <http://www.sims.monash.edu.au/research/rcrg/research/crm/index.html>

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 Group⁴⁸ 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 ex-

48 Recordkeeping Metadata Project (2006). Retrieved 20060626, 2006, from Recordkeeping Metadata Project Accessed: 26-06-2006 .<http://www.sims.monash.edu.au/research/rcrg/research/spirt/>

planation 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. There might be misinterpretation in some matching of metadata but the overall result is not affected. 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 re-

search 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. The sample is however satisfactory. They are however 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 an 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.

Paper III



To Capture and Preserve Documents in Processes

Öberg, L-M (2006). To Capture and Preserve Documents i Processes. In K.S. Soliman (Ed.), Information Managment in modern enterprise: Issues & solutions. Proceedings of the 2005 international business information management conference July 5-7, 2005, Lisbon, Portugal (pp.181-183): IBIMA

Abstract

A key criteria for information that should be preserved are that it has to be trustworthy over time. To reach this, feasible state, a proactive approach is important. Which means that information have to be prepared for long-term preservation already in the creation phase. The preparations include management of the context of information. By linking documents to the processes where it has been created and changed an organization will have the ability to follow the document flows afterwards. Two qualitative empirical studies constitute the base of this paper. In the two studied document management systems; there were no links between documents and activities but there were links between documents and processes and documents and sub processes. If an organization implements links on activity level it will give preservation of context and future interpretation of organization documents.

1. Introduction

A transforming into a process-oriented way of working may influence an organization management of information. This also includes information with long-term preservation need. By linking documents¹ to the processes² where it has been created and changed, an organization will have the ability to follow the information flows afterwards. Interpretation of information depends on context[1]. The usual implementation of capturing and preservation of context is, use of metadata (in this paper defined as the total sum of what one can say about any document[2]). Capturing of connections between information and the processes will lead to trust worthier information, because of the fact that there it serve as evidence of organizations

1 Document is a form of “implemented” information in a computerized information system. A document is a coherent unit such as a word-file or a PDF-file.

2 “Processes are differences or change in states of space, time and form. It can produce information, matter-energy or both.” [5] A process consists of activities. The first activity in a process is in this paper called a transaction.

actions. Trust worthy information is a key criteria for information that have to be preserved for long time [3, 4]. Which information that need to be preserved, depends partly on legislation and partly of an organization needs. One example is Swedish health care records that are preserved during a patient's lifetime and additional ten years. E-archive is becoming a reality and thereby arises a need of new strategies for the preservation and management of information.

2. Problem domain

The focus of this research is to identify which information that is needed to interpret and use processes, documents and the links in-between those two. The aim of this paper is to study metadata capturing in process supportive document management systems, in order to understand strengths and weaknesses in existing solutions.

The starting point for this research is that a proactive approach is essential when an organization want to preserve information. This means that the information has to be prepared and captured for long-term preservation at the creation stage, otherwise it will be costly and difficult. It is necessary to develop systems that mark and capture the information with metadata that is suitable for the information flow in current business processes [6]. Several metadata sets exist today all created for the long-term preservation of information.

3. Description of method and research sites

Two empirical studies have been accomplished. The first study focused on key characteristics of preserved information. The study covered four Scandinavian organizations: one public authority, one multinational enterprise, one municipal and one police department. The result from this study will be presented at the IRIS28 conference in august. In the second study, metadata capturing in two computerised document management systems was in focus. In those document management system there are links between documents and the processes of the organization. The empirical data (captured metadata) was in a qualitative manner matched to metadata sets for long-term preservation. The method used in both studies was interpretative case studies [7]. The data collection methods used was: studies of system documentation, screen shots and qualitative interviews. The second study took place in one Swedish public authority with main business in administration and legislation. This organization will in this paper be labelled with an A. Organization A implemented their document management system in 2003 and it includes documents such as: policies, regulations and decisions. Organization B is a result-oriented section in a public

company. Their main business is maintenance of infrastructure. They implemented their document management system 2004 and it includes document such as: check-list for dialogue with member of staff, manuals and templates.

4. Empirical results

The first empirical study showed that documents and transactions were not always linked together. In some cases it was tacit. This is a contradiction to one of the theoretical key concepts for preserved trustworthy information. Preserved information shall fulfil as evidence. This gives that the link between the transactions and in this case the document, has to be captured and preserved. The second study showed that this has been implemented different in the document management systems. This is visualized in figure 1 and figure 2.

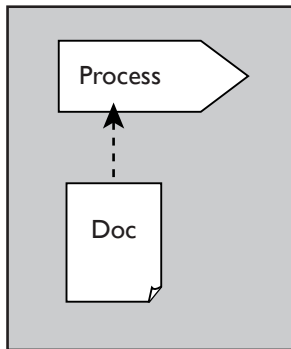


Figure 1 Visualization of implementation of links in organization A.

Organization A has one link to connect the process and document together. Organization B has implemented two links between the process and the document. There is one link between the process and the actual document and further links between the sub process and or sub processes and the document. Example from organization A: Regulations regarding VAT will be connected to the process impose taxes and bringing payment. When users want to search for the document it possible to get context information about which process it belongs to.

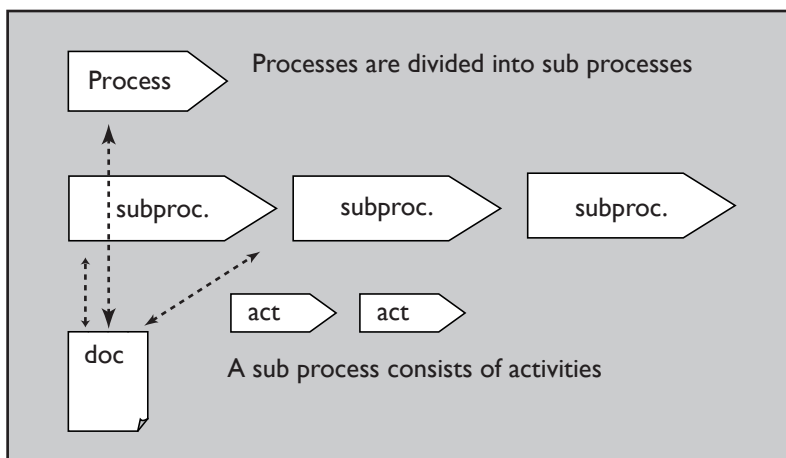


Figure 2 Visualization of implementation of links in organization B.

Example from organization B: A document that includes a checklist for plan of commission will be connected to the process accomplishes commission and the sub process Plan and manage commission. The links between documents, processes and sub processes, are two-way associated.

There were no links between documents and activities in none of the studied organizations.

A reflection from both studies is that none of the organizations has chosen to capture metadata that concern the processes in their document management systems. Organization B has chosen to capture and preserve metadata about processes in a separate information system. This pattern is also indicated in the sets (e.g. VERS[8], RKMS[9], ISAD(G)[10]) that are analyzed. The sets include a number of metadata for the document, and metadata to link documents and processes. There are no examples of metadata for description of processes. The first empirical study focused on characteristics on preserved information. The result indicates that some organizations only have fragmented trace of processes. The result shows that there are more traces of the process if it did not follow the normal course of event. Case codes are noticed, and copies of letters that describe what needs to be done to make a correct version are preserved.

One of the identified characteristics from the first empirical study is, full version management. Organization A has implemented this in their system. In their information system a user is not allowed to make updates of documents. They have to make a new document if changes are needed. If a document has been published the system guarantee that there will be no changes in the original version. The implementation in organization B has the same functionality but the same document can exist in several versions. Full version management is a key criterion for the ability to follow the document flow.

5. Concluding remarks and future research

To reach full traceability on processes and documents there has to be links on activity level. An example of feasible design is visualized in figure 3.

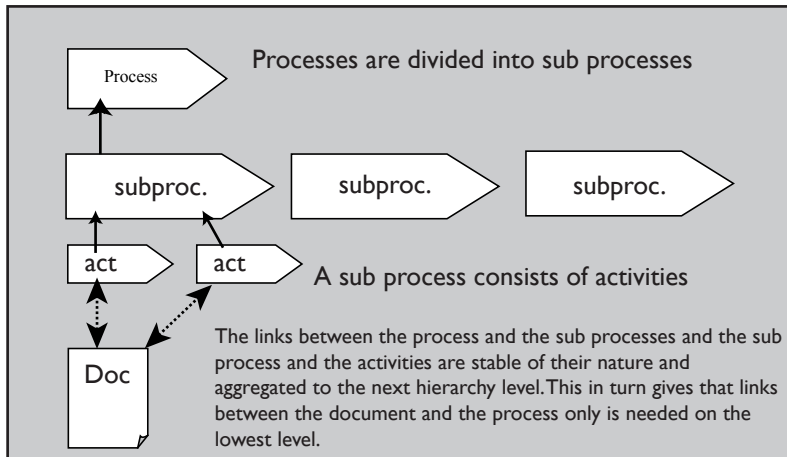


Figure 3 Example of feasible design of links

This kind of implementation will require metadata to catch those links and full version management. Only with full version management it will be possible to follow every change that a document will go through during a process. Future research is necessary to describe what kind of metadata that is needed to describe document flows within processes. There is also need of future research, to investigate differences between different types of organizations. The second study shows that there are a lot of differences in metadata capturing but the reasons for this are not known.

In some of the studied organization the process were more visible when the normal course of event were not followed. Organizations that capture metadata that shows this kind of trace have chance of succeeding. But they have to supplement this with descriptions of the normal process. This could be implemented by making instances of a process only if the processes not follow the normal course. Only organization B have captured about processes. They have built a special application for management of the processes where they also capture metadata about the process. Example of such metadata are; description of activity and description of sub process. Further research is necessary to compare this system to other similar system. A possible solution is of course to combine the process supportive system with the document management system.

The metadata capturing has to be as automated as possible and the solution for this

is a proactive approach. Some metadata has to be captured when the document is created and some later during the process. There are also metadata that have to be captured when the document is inactive. There is a need for further research to design metadata capturing during a life cycle of a document. Long-term preparation in early stages has to be on the agenda for system developers.

8. References

- [1] Sundgren, B. and G. Steneskog, "Information Systems for Concerted Actions", in *Exploring Patterns in Information Management - Concepts and Perspectives for Understanding IT-Related Change*, B. Sundgren, et al., Editors. 2003, The Economic Research Institute Stockholm School of Economics: Stockholm. p. 356.
- [2] Gilliland-Swetland, A.J., "Setting the stage", in *Introduction to Metadata: Pathways to Digital Information* Version 2.0, M. Baca, Editor. 2000, Getty Research Institute. p. 1-12.
- [3] Duranti, L., "Concepts, Principles, and Methods for the Management of Electronic Records." *The Information Society*, 2001. 17: p. 271-279.
- [4] Thomassen, T., "A First Introduction to Archival Science". *Archival Science*, 2001. 1: p. 373-385.
- [5] Eriksson, D. M. Four Proposals for Enterprise Modeling. Unpublished Ph. D Thesis, Chalmers University of Technology, Göteborg, 2004.
- [6] Hofman, H. Metadata and the management of Current Records in Digital Form. Retrieved April 29 ,2004, from www.ica.org/old/biblio/com/cer/metadata.htm
- [7] Walsham, G. "Interpretative case studies in IS:research nature and method." *European Journal of Information Systems*, 4(2), 74-81.
- [8] VERS: Victorian Electronic Records Strategy. Retrieved May 1,2005 from, http://www.prov.vic.gov.au/vers/standard/spec_02/default.htm.

[9] RKMS: Australian Recordkeeping Metadata Schema. Retrieved May 1, 2005, from, <http://www.sims.monash.edu.au/research/rcrg/index.html>.

[10] ISAD(G): General International Standard Archival Description. Retrieved May 1,2005 from, http://www.ica.org/biblio/cds/isad_g_2e.pdf

Paper IV



Needed Components to reconstruct processes

Öberg, L-M (2006). Needed components to reconstruct processes. *In proceedings of Archiving 2006 conference May 23-26, 2006, Ottawa, Canada 8pp. 165-168*: IS&T

Abstract

Access to public documents is a Swedish citizen's right. Decision processes should be traceable, which requires that the information systems that contain this type of information must fulfill requirements for traceability between information and processes. With the 24/7 concepts there are new challenges to be able to reach traceability. This study has focused on components that are needed to reconstruct processes. Two Swedish municipalities have been part of this study and it shows that the archived material gives a rather good picture of the internal process and the components that has showed to be important are; contexts, documents and dates. The study has shown that the handling and thereby the archive is limited to a process that include only one actor. Nowadays when authorities strive to build e-services that cross agency borders the record keeping systems have to be designed to support new kind of processes with for example several actors.

Introduction

The Freedom of the Press Act is part of the Swedish constitution. This gives every citizen rights of access to public documents. In Sweden all documents that are created and received in a public authority are public unless a classification of secret is applied. The Swedish systems also include openness; which makes traceability an important matter for the public authorities. The traceability makes it possible to follow processes such as a decision making process. Nowadays with a 24/7 e-service the intention is that citizens should be able to follow their on-going processes and access public documents by themselves. In Sweden there is an on-going research project called "Home Building Guide" (Bygga Villa in Swedish) the goal of which is to build an integrated e-service. The integrated e-service will cross agency borders in order to make it easier for a citizen to seek information, for example about a property unit purchase or to follow the different processes that are needed to build a house. (The research reported here is part of this project.) One example of such a process is applying for a building permit, during which several authorities could be

involved. If those authorities could cooperate in an effective manner the process of getting a building permit would be easier and more efficient.

Today a lot of the public documents are electronic and included in different kinds of computer-based information system. Examples of such systems are: document management systems; record keeping systems and workflow systems or combinations of these. The problem is however, that the requirements for designing such an information system that take account of traceability of processes and general requirements for record keeping over time under Swedish conditions are not known. However in the archival community the importance of tracking information processes has been known long before the process orientation era. Duranti [1] has described records as links in a chain, and Bearman [2] even states that it is impossible to understand records without relation to processes and activities. There are several examples of standardization processes that include traceability between processes and documents/records. Examples are; ISO 9001 and ISO 23081-1[3]. Another example is provided by McDonald [4] whose prototype made it possible to connect and track records in processes. Nevertheless to be able to build an integrated e-service like the “Home Building Guide” there is a need for studies that describes what kind of components there are used in Swedish archives of today and what kind of needed components are missing. Therefore the aim of this paper is to describe what kinds of components that make it possible reconstruct processes, are currently in archives.

Research Approach

Methodologically, this study uses a qualitative approach both for data gathering and analysis. The study has been conducted in two Swedish municipalities categorized [5] as larger cities. The reason for choosing the municipalities for the purposes of this study was that they have traditions of traceability and openness. It was necessary to include two municipalities, since the data gathering is extensive, and has to be an in-depth study because of its aim. The process that has been the focus of this study is building permits. This is because this process is an essential part of the “Home Building Guide”. The course of action includes three steps.

1. Process modelling by interpretation of public documents from different time periods. The process orientation has been a part of enterprise modelling since the early 1980s. The modelling technique in this paper follows the main-stream within this area. The first step includes studies of differences between time periods. The different time periods include processes from years with manual handling of building permits compared to processes from years with implemented supporting computer-

based information systems (such as document management systems or work flow systems). Some of the components may be kept in systems, which will influence the reconstruction of the processes. The time periods chosen are: 1960-1965 and 1999-2000.

2. A Seminar in which the process models were discussed with archival staff and local government officials with experience of building permits. The outcome of the seminar was revised process models with identified weaknesses and benefits of existing components.
3. Analyses of which components make it possible or impossible to follow the process. This was made by comparing the different findings against each other and analyzing the findings into categories of components that influence traceability.

Description of the research sites

The municipalities have chosen different types of solutions for their archive but they have some things in common. The link that connects documents together is the property unit designation. In Municipality A all microfilms, documents and digital documents are indexed by the property unit designation, and it is also the only search criteria. Municipality B has indexed all drawings and maps by the property unit designation whereas documents like the application form and certificates are ordered by diary number. In 1999-2000 both municipalities implemented electronic workflow and document management systems. Municipality A also implemented a digital archive so all documents from a building permit process are saved as TIFF-files. When a process is closed they make deliveries to the archive. The deliveries include all documents that should be preserved. There is no delivery of the workflow and document management systems or parts of it. The documents are in this sense separated from the system that they have been managed in. Municipality B also works with deliveries to the archive when a process is closed up. Drawings and site plans and such documents are sent to microfilming and indexed by the property unit designation. The rest of the documents are preserved on paper ordered by diary number.

Presentation of the Components

In this section the empirical data from the study is presented. The section includes one example of a process model and tables; the latter describes differences and similarities between both time periods in both municipalities. The process model in

figure 1 is one example of a model that has been used in one of the seminars. The notation used in the process models is taken from Microsoft Visio, which was used to make the models. The notation components should be interpreted as follows:

Light grey rectangle: Activity: White rectangle: Activity, which does not always have a link to a document and is sometimes indicated by other documents or activities; Black line: Relationship; Document symbol: Document that is part of the archived material; Asterisk: Used where a document and/or activity does not exist in every instance of the process.

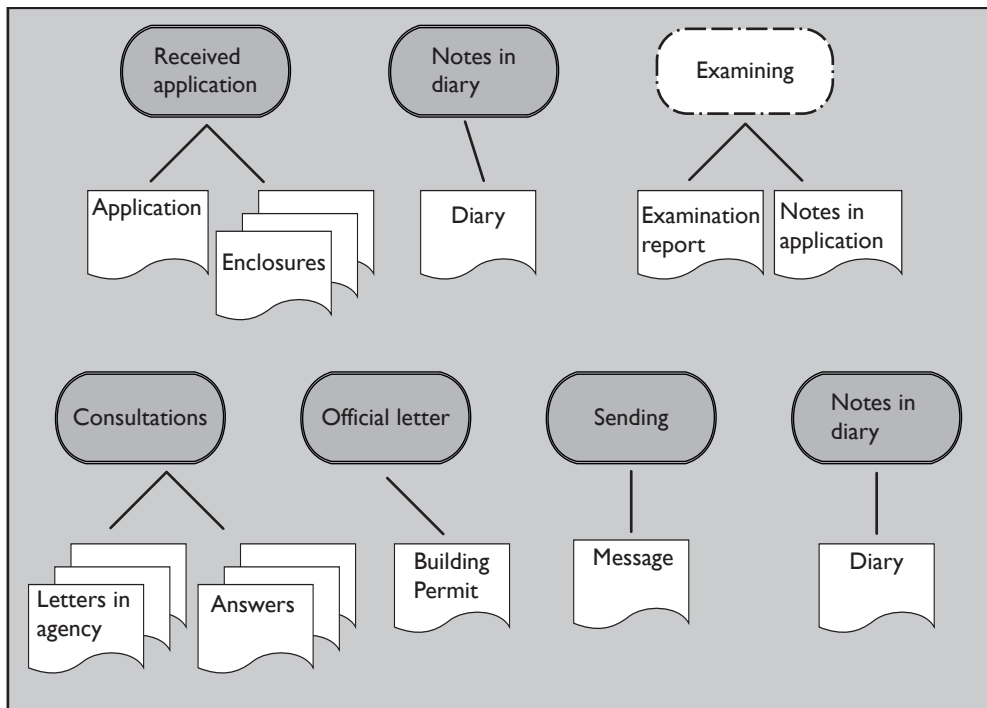


Figure 1: Example of a process model. Municipality A, 1960-1965.

The tables are divided into two main parts, identified components and missing components. Most of components that have been found in the archives are documents. They often have embedded metadata (author, dates etc), which makes it impossible to describe all metadata. The tables below therefore include the identified components on what could be called the document level and sometimes diary level. The first table represents Municipality A and the second table represents Municipality B. Components marked with an asterisk mean that this component did not appear in every instance of the process.

Table 1 Empirical results from Municipality A

Components Identified	
1960-1965	1999-2000
Application	Application
Enclosures	Enclosures
Diary	Cover Page
Consultations	Consultations
	Certifications
	Minutes from consultations
Minutes	Links to minutes regarding decisions*
	Inspection reports*
	Examination report*
Building Permit with motivations and conditions	Building Permit with motivations and conditions
	Checklists for qualification certificate
Message to the applicant	Message to the applicant
Components missing	
1960-1965	1999-2000
Examination Reports	Examination Reports*
	Decisions pertaining to the type of process

Table 2 Empirical results from Municipality B

Components Identified	
1960-1965	1999-2000
Application	Application
Enclosures	Enclosures
Diary	Printout from diary*
Examination report	Examination report
Certifications	Certifications
Building Permit with motivations and conditions noted on application form	Minutes from decision (includes the building Permit with motivations and conditions)
Flying inspection documentation	Flying inspection documentation*
	Minutes from consultation*

Delivered documents according to conditions in the permit	
Official letter by civil servant	
Consultations	
	Checklists for qualification certificate*
	Civil servant comments
Inspection Reports*	Inspection report*
Message to applicant	Message to the applicant
Components Missing	
1960-1965	1999-2000
Inspection Reports*	Printout from diary*
	Certifications*
	Minutes from consultation*
	Checklists for qualification certificate*
	Decision pertaining to the type of process
	Flying inspection documentation*

To the main differences between the municipalities are components concerning inspections and examinations. Municipality B has chosen to preserve different types of documents that describe those kinds of activities. From the mid 60's there are examples of handwritten notes made by a civil servant after flying inspections. During the 1999-2000 the municipality has preserved all examination reports, which are forms that are filled in by hand. In Municipality A there are no examples of inspection reports from the first period. This was pointed out during the seminar and the participants were surprised by this fact. They could think of two possible explanations. The municipality accomplished a migration from paper to microfilm during the 70's. The municipality then had to decide which documents should be preserved and the decision was influenced by the cost of the migration. The other possible reason was that the inspection report has never been part of the material that was delivered to the archive. The civil servants kept this kind of documentation on their files and they were never incorporated with the rest of the documentation. The interesting point here is that even during the later period of time there are few examples of inspection reports in Municipality A. The discussion about this fact was that only inspection reports that note divergences from the norm are preserved. To confirm this one ought to check within the workflow and document management system, which has not been part of the data in this study. However those differences

of culture and decisions between the two municipalities influence the components and thereby the possibilities for reconstruction of processes.

The processes of the two time periods differ a lot. The main reason for those differences is that the law that regulates building and site plans in Sweden changed in 1987 (Plan och Bygglagen 1987(10)). The new law has for example influenced what kind of certifications are needed. The most important change is that much of the responsibility moved from the authority to the applicant as the person responsible for quality. A lot of components are missing because of the fact that they are never sent in to the authority. The levels of formalization also differ between the time periods: it is higher during the later period. This could be explained by several things. First of all both municipalities implemented electronic workflow and document management systems in the late 90's. Computerization in itself leads to formalization of processes and documents. Both municipalities started a process-oriented transformation which partly aims to make the handling of building permits uniform. There is also a general trend of streamlining of public administration, which has led to the introduction of forms and templates. The missing Components were revealed during the seminars.

Analysis of Components

The components have been categorized into three categories and they are analyzed and discussed in the sections below.

Context

As mentioned earlier the main component that connects different documents and their context is most often the property unit designation. The other important component necessary for the preservation of context is the diary. Both Municipalities have used the diary as the center of context in the 1960's. There, in the diary, it is possible to follow the different audit trails of the process such as which documents have been part of the process and who was the applicant. Municipality A has during the later time period (1999-2000) replaced the functionality of the diary with a cover page. The cover page is part of the workflow and document management system and a digital print out is delivered to the archive together with all the other documents. Municipality B has the same kind of cover page in their system but they do not deliver it to the archive, which means that important context information is lost. Municipality B has not fully implemented an electronic handling process but when they do they ought to reconsider how the context information from the systems

should be preserved. To make use of an e-service as much as possible it is necessary to take the step to a higher degree of electronic automation and this will make this issue even more important to Municipality B.

Documents

Both the municipalities included in this study have chosen to preserve building permit information as documents, regardless of whether the system is paper-based or digital. The alternative nowadays is to use electronic forms and split up the information in several tables in a database, thus requiring preservation of the databases and systems within which they are held. The outcome of this document-oriented way of working is that the document gives information about context and activities in the process. This document-oriented way of working has in this sense a positive effect but there is also a risk by separating the documents from the system that they have been part of. There may be information about the process that is not within a document and the organization must in those cases take care of those components in some way. The argument here is not that the document-oriented way of working is the only alternative. But if an e-service will include for example applications in XML-format the organization has to take care of that situation. An important thing to deal with could be to preserve not only the content but also the metadata that are part of the application. The municipality's document-oriented way of working does not include separated metadata. They both have separated metadata in their workflow and document management systems but they are not delivered to the archive. This will influence both the possibilities of interpretation and the search ability especially if citizens will have direct access to the documents.

As noted earlier, the introduction of template-based documents helped to standardize processes across both municipalities, thus making it easier to follow the document trail afterwards. However, care must be taken in designing the template to ensure important components are not missed. A positive effect of the standardization is that this will make it easier when organization needs to exchange information. The need of exchanging will increase with integrated e-services.

Date

The discussion of context concerns how to identify which item belongs to a particular instance of a process. This sets the external border for the instance. When it comes verifying the order of activities the component date has been shown to be very important. There are very few examples of documents that have not been la-

beled with a date of creation or date when received which of course is good.

Conclusions

This study has presented three different categories of components that make it possible to follow processes namely, context, documents and date. The study has shown different ways of implementing those components and there are of course alternatives.

A high degree of openness is important for successful e-services. The municipalities in this study appear to be in a good position in preserving processes, which is a key factor for openness, traceability and in the future effective and useful e-services, but this study has revealed some problems. The process that is preserved in the archive shows the municipalities' internal processes. The "Home Building Guide" and similar integrated e-services will place the citizen inside the process. This study does not show if a citizen thinks that the preserved process gives a fair picture of what has happened. There are for example few components of dialogue with the applicant during the process of a building permit. During the seminars it was revealed that dialogue with the applicant takes a lot of time and includes help with drawings and help with filling in the application. With an e-service the communication with the citizen can change into new ways. The communication will often be recorded by e-mail or by forms. This raises a need for discussion about appraisal of what has to be preserved and not.

There are some audit trails that connect to other agencies' work (for example, consultations with the environmental health committee) but to build integrated, effective e-services the archives must also be more integrated with each other. This will include both cooperators within the Municipalities and external actors if the "Home Building Guide" vision should become reality. How this should be solved is anything but straightforward. Questions that have to be solved are for example; where processes cross agency boundaries, who is responsible for the documents; who captures and preserves the documents and how the context of the documents is maintained and how documents should be exchanged.

Another problem that has been revealed during this study concerns the process borders. When the building permit process is finished maybe a new sub process starts? Or some sub processes may have preceded the building permit process, or be contained within it. How and where should those components of traceability be preserved?

This study has made a contribution about requirements for designing information

systems that include public documents. In particular, it has looked at requirements that regard traceability and required components to be able to reconstruct processes. This study has also revealed new challenges to recordkeeping in a new environment, namely integrated e-services that cross agency borders. The paper has identified areas of need for future research before integrated e-government services

References

- [1] L. Duranti, The archival bond. Archives and Museum Informatics., 11 (1997).
- [2] D. Bearman, Electronic evidence: strategies for managing records in contemporary organizations. (Archives and Museum Informatics :Pittsburgh, 1994).
- [3] ISO 9001 & 23081-1 www.iso.org, 2006.
- [4] J. McDonald, Towards Automated Record Keeping, Interfaces for the Capture of Records of Business Processes Archives and Museum Informatics 11 (1997).
- [5] SKL, Sveriges kommuner och landstings hemsida www.skl.se 2006.

Author Biography

Lena-Maria Öberg received her Master in Informatics from the Mid Sweden University (2001). Since 2004 she has worked as a Ph D Student in the project Archives of the Future at Mid Sweden University.