From Paper to Digital Documents

Challenging and improving the SGML approach

by

Tone Irene Sandahl

Preface

This Dr. Scient. thesis has been submitted to the Faculty of Mathematics and Natural Science, Department of Informatics, University of Oslo. My work was conducted within the Document Information Systems (DIS) project at the System Development Group. The DIS project consisted of nine Masters students, a doctoral student (myself) and two associated professors, Kristin Braa and Pål Sørgaard. The DIS project is part of the Internet Project, which is an international research network of more than twenty researchers from computing, informatics, and ethnology. The ambition of the project is to give a rich overview of the main aspects of Internet use (see http://internet.informatik.gu.se/).

During my doctoral study a number of people have been of invaluable help, by commenting on and critiquing each of the papers, but also by being available for inspiring discussions. I would like to express my gratitude to my present supervisor Kristin Braa for comments and discussions at all stages of this work. Kristin has been indispensable, and her enthusiasm and "turbo-brain" have made it very inspiring to do research with her. I do also appreciate Kristin's supportive and encouraging spirit. I would like to thank Riitta Hellman, who helped me initiate the study, and Pål Sørgaard, who has shared ideas and been very encouraging along the way. Riitta and Pål were my first and second supervisors respectively.

Warm thanks to the co-authors of the papers, Kristin Braa, Astrid Jenssen, Nina Lundberg and Pål Sørgaard. I am grateful to Eric Monteiro, Ole Hanseth, Tone Bratteteig, Nina Lundberg, Astrid Jenssen and Erlend Øverby for their comments on the 'final contribution'. Thanks are also due to the students participating in the document project (DIS), especially Bjørn Tore Kristiansen, Espen Berger and Arnstein Schei. I have appreciated the discussions and comments from my colleagues in the Internet Project: Magnus Bergquist, Staffan Björk, Kristin Braa, Claudio Ciborra, Antonio Cordella, Per Dahlberg, Bo Dahlbom, Lars Bo Eriksen, Henrik Fagrell, Jennica Falk, Leif Grönqvist, Ole Hanseth, Lars Erik Holmquist, Ann Johansson, Steinar Kristoffersen, Jan Ljungberg, Fredik Ljungberg, Peter Ljungstrand, Nina Lundberg, Eric Monteiro, Carsten Sørensen, Kai Simon, Ulrika Snis, Tomas Stegberg, Dagny Stuedahl, Lars Svensson and Pål Sørgaard.

My thanks also go to Joan Greenbaum and Frieder Nake for their inspiring discussions when visiting our group. My colleagues Eevi Beck, Annita Fjuk, Ole Smørdal and Dagny Stuedahl have been helpful in many ways. I would like to thank the management group at the University Center for Information Technology for their patience in waiting for me to finish this thesis. Special thanks to Astrid Jenssen, an SGML expert who is always available for questions and discussions; also to Nina Lundberg for all the help and encouragement she gave me in telephone conversations. Margaret Forbes deserves special thanks for her efforts with my English.

In addition to my colleagues, there is a group of people to whom I owe a big **THANK YOU**, and they know why: my dearest Bjart, my parents, the rest of my family, and Astrid.

Summary

This research has been initiated on the basis of practical experiences in developing a relatively large SGML system at the University of Oslo. This thesis contributes to the field of information systems, with a particular focus on document systems. The aim of this work is to inform the design of document systems by considering the transformation from paper to digital documents in organizations. The Standard Generalized Markup Language (SGML, ISO 8879) approach is emphasized. The SGML approach takes the documents' structure and content as the starting point in design, and regards the document as a collection of structured information. This approach is challenged and tentatively improved by empirical studies of documents in use and theoretical considerations of artifacts at work.

The research approach has been an Action Case, as defined by Vidgen and Braa (1997). The interpretation of the transformation process from paper to digital documents is based mainly on an in-depth case study that was conducted at a Norwegian news agency from January 1996 to March 1998. The empirical findings are discussed according to theoretical concepts that emphasize the significance of artifacts at work to illuminate the various roles of documents at work.

Concepts from the Actor Network Theory (ANT) (for example, see Callon, 1986; Latour, 1987; Law, 1986) are applied to emphasize the interrelations of humans and artifacts, as well as the importance of artifacts' properties in these relations. The concepts of 'boundary object' (Star and Griesemer; 1989) and 'borderline issues' (Brown and Duguid, 1994) are applied to get various perspectives on the actor-network.

The study illustrates that it is challenging to substitute paper documents with SGML documents. Firstly, two different types of technology, with different properties and features, are exchanged. By removing paper documents, we also remove resources that go beyond the canonical meaning of the artifact. These resources are related to paper as a technology. Secondly, the document perspective in SGML is too restricted in relation to the various perspectives on documents in practical use. The emphasis on structure complicates the production of documents. Thirdly, the application of shared document models across work practices turns the various heterogeneous actor-networks into one network, which requires a common objective among the actors involved. The dilemma of "who does the job and who gets the benefits" (Grudin, 1989; 1994) arises as well.

The study indicates that an investigation of the actor-networks that include documents provides an insight into the more hidden aspects of work. By regarding documents' central, peripheral, local and shared properties, one can gain an understanding of how documents are embedded in work, including the importance of documents and related artifacts to aspects such as awareness, articulation and coordination of work. The properties determine how things become interrelated into heterogeneous networks. The research shows how a document's properties or inscriptions are essential to its production and application in use. Insight into these prerequisites helps us to understand how the computer system can fit into work practices, even if we do have no guarantees that it will be used in the way that we expect. According to design, work practices are improved by changing the technical properties or the technical fundamentals, by adding various inscriptions into the system. This thesis describes how an existing system was improved by the use of 'gateways'. In the design of the gateways, the idea has been to keep the technical possibilities that SGML provides, and at the same time take into account our knowledge about the paperwork.

Background

This research has been initiated on the basis of practical experiences in developing relatively large SGML system at the University of Oslo. At the end of 1992, a project was initiated at the University Center for Information Technology Services (USIT) to determine what type of electronic infrastructure could deal successfully with electronic documents at the university. The infrastructure had to address the whole life cycle of a document, i.e. production, updating, filing, administration, distribution, presentation and reuse. In addition, to derive benefits from documents in electronic form, the intention was to develop new functionality such as advanced search and retrieval features, as well as services like 'document on demand'. The use of Standard Generalized Markup Language (SGML, ISO 8879) in the project emerged as a possible key tool for describing the documents and their content.

In March 1993, to gain experience in the practical use of SGML for some parts of the information produced at the university, USIT established a pilot project that involved developing an open and flexible solution for the production, exchange and distribution of the university's course catalog. The catalog is divided into sections for each faculty, describing courses offered by that faculty, and different sections for other kinds of courses such as distance education, information about student services, and collaborating institutions. The information is written within the different units, and coordinated and presented in a printed catalog twice a year. The pilot project was initiated to develop a technical infrastructure and administrative routines for dealing with the catalog as well as providing "new functionality" for the students and the staff. In this project I participated as the project leader and as one of the system developers.

The project had a technical focus. A critical aim was to get a structured and well-defined catalog in order to develop new services based on the information in the catalog. We wanted to develop scripts that could manipulate the catalog and its information elements depending on existing situations and products. The intention was to implement services such as room allocation, in which a writer allocating a room could get suggestions for a place and time based on existing information in the catalog, as well as advanced search in the catalog information and customized publishing. A basic principle was that the writers should have to enter information only once. As a result, updating and maintenance of information across presentations and products should be easier, and redundancy could be avoided. If the information was to be presented in different settings and applications, the computer would do this automatically. This presupposes that the method of describing the catalog in a structured way is standardized across the work practices involved.

The pilot project did not turn out as expected. During the pilot project we did not develop any new services like the ones mentioned above. The only innovation was the catalog available on the Web. The printed catalog was still the product; however, it was regarded as a "better" document because of its improved structure and layout. How was it that after applying two years of hard work to improve the production process and functionality of the catalog, all we achieved was a catalog on the Web in addition to the printed catalog that had been produced perfectly well without the use of SGML throughout the years?

The university course catalog was standardized in a way that compelled all the contributors in the production process to follow the same predefined document model, which represented a very detailed document structure. From a technical point of view, this standardization and high level of structure and specification has its benefits. The standardization is necessary for communication over networks, and for automation of processes, such as merging pieces of information into a catalog, or for various kinds of 'virtual' documents, and their presentation on paper or an electronic device. The

standardization and the level of specification are important for the functionality as well. If it should be possible to search in the catalog for all courses that start on January 22, the date must be specified in the text in an unambiguous way across the various departments at the university. From a technical point of view the standardization and rich specifications support a high degree of technical flexibility.

The project followed a typical SGML approach, meaning that the document's content and structure formed the foundation for the design. The aim was to standardize the document on a shared document model that represented the catalog, to allow automation of document handling and use of the documents' information elements in different settings and applications. The document model was essentially based on document analysis of the paper catalog.

The experiences from this project indicated that standardization of documents across work practices is problematic. It was problematic for the writers who had to mark up their text based on a predefined document model. It was also problematic for the application of documents at work. SGML presupposes that documents are only structured sources of information. This is one interpretation of the documents in use. Previously, the writers at the departments and faculties could present and use the catalog in their own way. With a standardized catalog, this freedom was restricted.

This research is based on these experiences. It was interesting to focus on why it is so hard to get an SGML system into use; to challenge the SGML approach to digital documents and to improve the approach based on experiences of SGML systems in practical use.

Table of Contents

From Paper to Digital Documents

Challenging and improving the SGML approach

1. Introduction

This thesis contributes to the field of information systems, with a particular focus on document systems. The aim of this thesis is to inform the design of document systems by considering the transformation from paper to digital documents in organizations. The approach emphasized is the Standard Generalized Markup Language (SGML, ISO 8879) approach. The SGML approach takes the documents' structure and content as the starting point in design. This approach is challenged and tentatively improved by empirical studies of documents in use and theoretical considerations of artifacts at work.

At least 80% of electronic information in organizations is in the form of documents, as opposed to database records (Reinhardt, 1994). These documents usually belong to only one user and passed from one person to another in printed form. During recent years, there has been growing interest in digital document applications. There is a tendency to make the documents available and accessible across boundaries, and to integrate document technology into larger systems in order to create, handle and present documents in new ways in addition to improving their functionality (Reinhardt, 1994; Denning and Rous, 1995). The enormous diffusion of the World Wide Web (Web) makes it easy for anyone to publish, exchange and distribute documents within and across organizations. A large number of document management systems and document databases has been introduced into the market as well (for example, see BYTE May 1997 for an overview), and standards, e.g. ODMA[1], OLE[2] and OpenDoc, have been developed to improve interconnectivity and interoperability between various kinds of document technology. There is ongoing interest in making documents digital, and this thesis is a contribution to this process.

SGML is one approach to digital documents that is attracting growing interest. SGML is intended for use in the publishing field (Smith, 1992), but its focus on 'structured documents' makes SGML relevant for document systems in general (Goldfarb et al., 1998; Travis and Waldt, 1995; Goldfarb, 1990; Herwijnen 1990; André et al., 1989).

This thesis also addresses the tendency to integrate information and communication technology into large interorganizational or even worldwide systems. The term 'information infrastructure' has been widely used to describe these interconnections. In infrastructures, standards are the technical backbones (Hanseth et. al., 1996). To allow globalization, computer networks have to be connected in a way that makes communication and information exchange possible. Also, there are at least two different ways of achieving information interchangeability between systems and applications: standardization on applications, so that the applications can work on each other's data, or standardization of the data itself, so that it can be processed by any application. For instance, a standard such as OLE supports the former way, while SGML supports the latter. HTML (HyperText Markup Language) is an SGML application, and provides information interchangeability by standardizing the data

---

[1] Open Document Management API
[2] Object Linking and Embedding

through encoding the document. The new XML (eXtensible Markup Language) initiative is another example of this (Light, 1997; Goldfarb & Prescot, 1998). XML has an SGML profile, which means that it can support an unlimited range of applications as opposed to HTML where the aim is to display Web pages and provide hyperlinking (Ibid.).

This thesis is not about information infrastructures as such, but it focuses on the changes in interconnected work practices against the background of infrastructure visions. This research takes the work into the discussion of infrastructures, and how the infrastructures' need for standardization may impact that. In this way this research contributes to the infrastructure research as well.

The 'final contribution' is structured as follows: The next subsections present the research questions and a short summary of the contributions included in this thesis. Section 2 presents the research approach and section 3 gives an empirical description of a case study. The discussion appears in section 4, while the interventions are covered in section 5. There are some concluding remarks in section 6.

1.1 Research questions

The aim of this thesis is to inform the design of document systems by regarding the transformation from paper to digital documents with a particular focus on the SGML approach. The research questions are based on the challenges experienced during the catalog project at the University of Oslo, and the motivation to improve the SGML approach.

The research questions are as follows:

- How does the transformation from paper to digital documents impact work practices?

- How does the SGML applications impact the transformation and work practices?

- How can one improve the SGML approach?

The results from discussion of the two first questions are foundations for improvements of the SGML approach. The following subsections, 1.1.1 and 1.1.2, motivate the research questions. The section 1.1.2 also gives a definition of the SGML approach.

1.1.1 Documents

In general, documents arise from the basic human need for stable, external, communicative resources (Levy, 1994; 1988). What is said is transient and insubstantial, while what is written down is fixed and external, and therefore potentially available to people across time and space (Ibid.). In this way documents are 'immutable mobiles', which makes them a powerful resource in an ever-changing world (Latour, 1986). Because of the "sharedness" of documents, they have been important in many aspects of social evolvement. They have been central in the development of organizational management (Yates, 1989), in democracy and individual freedom (Ong, 1982), as well as in the creation of independent nations (Anderson, referred in Brown and Duguid, 1995). Because of their "massively public" nature, documents have a 'social life' as much as they have a directly informational life

(Brown and Duguid, 1995). For instance, documents have a role in linking people, in political linkage, and as a means of negotiation. Scientific journals bind researchers together by enabling them to read the same papers, keeping each individual conscious of being a member and aware of what others are working on. Newspapers may reflect culture and create identity, as well confirming or supporting it. Documents may provide a shared context for constructing meaning, and may be the beginning of the process of negotiations (Ibid.).

Documents are parts of our everyday work and life. Documents have been essential in organizational work since the mid-nineteenth century (Yates, 1989), even if their applications and genres have changed through the years (Yates and Orlikowski, 1992). As Hughes and King (1993, p. 153) state: "Much work in a variety of settings consists of working with and through documents in such a way that it could be said that the work consists of producing, managing and preserving documents". Documents have both context-independent and context-dependent properties (Ibid.). The context-independent properties point to documents as "massively public", "formatted artifacts", "traces", with "socially defined purposes" and "normatively and pragmatically regulated" (Ibid.). Accordingly, documents in work practices have properties that depend on their use, and that cannot be generally defined. The context-independent properties do no more than point towards the detailed examination of appropriate domains of use. The study of an airline operation room, conducted by Suchman and Trigg (1991), illustrates that documents ("complex-sheets") have a variety of context-dependent properties, which the community of practice relies on. A complex sheet is a document whose intention is to map incoming to outgoing aircraft and to include information about people and baggage required while the aircraft are on ground. However, in practice, these sheets have other roles as well. For instance, they are central in making available to employees the cooperative structure of their work (Ibid.).

Paper documents have properties that are very well-suited to cooperative work (Luff et al, 1992; Harper and Sellen, 1995). The fact that paper is tangible, flexible, and light has implications for the ease with which it can be physically transported, manipulated and laid out in space (Ibid.). Paper is 'tailorable' and 'ecologically flexible' (Luff et al., 1992). People tailor their documents in order to differentiate and highlight particular items. As illustrated by Luff et. al (1992), doctors may underline or mark text with colored pens in records to make colleagues aware of irregularities in treatments, and architects sketch in and ring changes to their plans. Regarding documents as ecologically flexible highlights the adaptability of documents' properties to a range of situations and contingencies. A doctor can examine a patient and glance at the record at the same time (Ibid.). In his studies of applications of medical records, Berg (1996) points out the importance of carrying the records around into different work situations, in talks with a patient or with other medical personnel. Even if the main aim of a medical journal is to document information about the medical status and treatment of a patient, this is only one of several purposes that the records serve in practical medical work.

This research points out the challenges of replacing paper documents, with all their context-independent and context-dependent properties, by well-structured SGML documents.

1.1.2 The SGML approach

SGML - Standard Generalized Markup Language (ISO 8879) - is the internationally recognized standard for identifying text, enabling the different parts of the text's structure to be identified and described so that each part can be handled and

accessed as appropriate (Bryan, 1989). It is intended for use in the publishing field (Smith, 1992). SGML's origins date back to the period when the publishing industry was dissatisfied with the way that computers handled text. The original intention was to rationalize the publishing process in a way that made it easy to transfer data between electronic typesetting systems (Alschuler, 1995). A number of references provide a broad historical presentation of SGML (for example, Smith, 1992; Alschuler, 1995; Travis and Waldt, 1995).

Like information systems in general, SGML systems are based on an abstraction, a description or a model of 'the reality'. In an SGML system the model is a document model, called the 'document type definition' (DTD) (Goldfarb, 1990). The SGML language follows a BNF grammar, and the model is described by the use of 'regular expressions'. A DTD specifies the information elements that make up a document type, and how they are structured in relation to each other. The DTD forms the foundation for every SGML system by rigorously recording and enforcing the requirements for document structure. For instance, the DTD controls the text editors that insert and keep track of the structure, allowing some authoring functions to be automated. The DTD controls the systems that manage whole and partial documents, and it provides information about documents to the software that formats them, indexes them for retrieval, and forms the foundation for all kinds of functionality (Maler and Andaloussi, 1996).

It is important not to confuse an SGML application with the SGML language itself. As mentioned, the SGML language can be used to define different DTDs according to the documents at hand. Adopting SGML does not mean that the documents will automatically "fit into" any other SGML system. The documents have to be encoded based on the same DTD (or documents have to be converted between the different DTDs) to support the potential for consistent processing within and between documents of the same type, which is the foundation for various kinds of new functionality and services. In this respect, the "shared" DTD represents the standardization of documents across work practices.

Within the 'SGML community' there are some 'rules' or assumptions about what the foundation of DTD design should be. This is defined as the SGML approach. The idea is to analyze the potential contents of each document type to determine what the structure will be. The DTD modeling concerns structuring information elements (Travis and Waldt, 1995, Maler and Andaloussi, 1996), and in practical DTD design this is very often done by taking the existing documents as the point of departure. Hanseth and Monteiro (1997) have experienced corresponding assumptions in the community of EDIFACT.

In a way the SGML approach is an example of data modeling, where the entities in the document model represent information elements in the real document. The intention is to share this information between work practices, and to create functionality and services on top of it. Data modeling has been criticized from several perspectives. However, the essence of the criticism is that traditional data modeling is blind to the dependencies of what data mean for people at work (social settings). In data modeling the data are separated from the contexts of which they form part (for example, see the references).

In this thesis, the SGML approach is criticized because it does not pay sufficient attention to the context. However, the challenges are based not only on the shortcomings in considering the information (the documents' content) according to its context, but also the shortcomings in looking at the document (the artifact itself) as a part of work practices. As stated above, the paper documents are significant artifacts

4

at work, and their application goes beyond the aspects of information sharing. This perspective is further spelled out in section 2.1.

As experienced in the catalog project at the university and as pointed out by Monteiro and Hanseth (1995), standards are not neutral or transparent, but impact the environments they become parts of. The SGML itself, by its focus on structure and standardization, impacts the transformation from paper to digital documents in addition to the fact that documents become electronic. This research is about the consequences of SGML in applications of documents in work practices.

1.1.3 Existing research fields that concern digital documents

Various research fields discuss different aspects of digital documents. For instance, the field of 'hypertext' concerns changes in (re)presenting and reading text in hypertext in addition to stating and discussing technical challenges (for example, see the proceedings from Hypertext'96[3]). The field of 'electronic publishing' is a more technical field for the publishing industry (for example, see Hersch et. al., 1998). The field of digital libraries (DL) also has a fairly technical focus, with a strong emphasis on search and retrieval facilities as well as on scanning and presentation technology (see the special issues of the Communications of the ACM, April 1995 and April 1998). However, there are very good examples of exceptions representing non-technical contributions (such as Levy and Marshall, 1995; Kling and Elliot, 1994).

In addition, various literatures from the field of information systems (IS), including infrastructure-specific literature, and the field of computer supported cooperative work (CSCW) have been applied to shed light on various aspects of the issues discussed. It has been important to get some understanding of documents in work practice and how structured digital documents may impact this. Where the literature is published - within the fields of IS, CSCW, DL or elsewhere - is less significant.

1.2 The Contributions

This thesis is based on a collection of papers. This final contribution discusses the findings presented in the papers according to the overall research questions by using theoretical concepts presented in section 2.1. The aim of the final contribution is to consider my findings, which are documented in the papers, from the perspective of informing the design of document systems, and especially SGML systems.

The first paper in the list below reports from the initial study. Based on the initial study the further research took two separate but not mutually exclusive paths. Firstly, investigation of paper documents and the transformation to electronic documents was conducted in order to understand the application of documents in work practices, and the impact that the structured document might have on this. Section 1.2.2 presents the papers (no. 4, 5, 6 in the list below), while section 4 discusses the transformation from paper to structured digital documents.

Secondly, some research was conducted in order to improve the writers' work situation by investigating other approaches to standardization on SGML. The papers (2, 3) are presented in section 1.2.3. This research has been significant in order to improve the SGML system studied.

The papers are:

---

[3] http://www.cs.unc.edu/~barman/HT96/index.html

1) Sandahl, T. I. and Jenssen, A. (1997): The First Step in Designing an SGML-based Infrastructure for Document Handling: A Discussion of Challenges that arose. In *Scandinavian Journal of Information Systems*, 1997, Vol 9, No. 2.

2) Sørgaard, P. and Sandahl, T. I. (1997). Problems with Styles in Word Processing: A Weak Foundation for Electronic Publishing with SGML. In *Proceedings of the 30th HICSS*, January 7-10, 1997, Wailea, Hawaii.

3) Braa, K. and Sandahl, T. I. (1998) Approaches to standardization of documents. In Wakayama, T., Kannapan, S., Khoong, C. M., Navathe, S., and Yates, J. (eds.) *Information and process integration in enterprises: Rethinking documents,* Kluwer Academic Publishers, Cambridge, Massachusetts, USA.

4) Braa, K and, Sandahl, T. I. (1998). From Paperwork to Network. In *Proceedings from the Third International Conference on the Design of Cooperative Systems (Coop '98)*, Cannes, France, May 26-29 1998.

5) Braa, K and Sandahl, T. I. (1998). Documents in Infrastructure. Challenges for Design. In *Proceedings from the European Conference of Information Systems (ECIS),* Aix-en-Provence France, June 4-6 1998.

6) Lundberg, N. and Sandahl, T. I. (1998). Artifacts in Work Practice. Submitted.

In paper 3, 4, 5 and 6 the authors are presented in alphabetic order only.

1.2.1 Initial study

The first paper, *"The First Step in Designing an SGML-based Infrastructure for Document Handling",* presents and discusses experiences from a relatively large SGML project at the University of Oslo. This project is referred to as the initial study, because it is the foundation for the issues discussed in this thesis. The initial goal was to improve the "functionality" of documents in order to access, update, search, use and reuse, retrieve, present, exchange and distribute documents independently of time and place and without loss of information. The problem statement bears evidence of a rather technical perspective: how could SGML be applied in order to reach these goals? However, during the project period the focus changed to more non-technical issues, as these were the most challenging ones in the development of the SGML system.

The paper outlines a few concluding remarks, which have been the point of departure for the research conducted. The initial study concludes that a prerequisite for benefits from SGML-based documents is producing the SGML documents in the first place, which requires a satisfactory writing environment. The SGML standard is non-neutral, and the writers were caught within the walls of standardization. Selection of the editor or word processor to be applied is essential, and has to be taken into account in DTD design. Secondly, documents are integrated into work practices; they have roles, and conventions grow around them. An SGML implementation may (differentially) change these roles and conventions and thereby impact work practices.

1.2.2 Documents in work practices

Three papers present and discuss empirical findings that help to answer the question of how the SGML approach impacts the transformation from paper to digital documents. The paper *"From Paperwork to Network - A Field Study"* emphasizes the

importance of understanding multiple roles of document at work in order to inform design of document systems. The difficulties one faces in the transformation are discussed and illustrated by a case study at a Norwegian news agency and its production of TV schedules.

We applied the concept of 'borderline issues' to identify and discuss borderline resources of documents that are established through conventions at work. These resources lie beyond what is regarded as the canonical meaning of the artifact, and therefore may disappear when various kinds of artifacts are 'replaced' by information technology. We state that understanding the artifacts' unnoticed resource embedded in the work practice could be a starting point for design of document-based systems. A prototype is developed to 'keep some of these resources alive'.

However, the complexity increases when one is dealing with inter-organizational systems or infrastructures, since the borderline resources vary in different communities of practice.

In the paper *"Documents in Infrastructure - Challenges for Design"* the role of documents in infrastructures is highlighted, since much of the communication and coordination among organizations is based on documents of some kind. In the paper, we describe a particular transformation process in a work chain producing TV schedules to illustrate the complexity of the transition from paperwork to network and discuss the interorganizational challenges in that respect. The work chain investigated consisted of various TV channels and a news agency. The concept of 'boundary object' has been applied for analytical purposes.

The SGML-based solution designed to support the document-based infrastructure was never adopted as planned. The essential reasons why it failed are, as we see it, the lack of focus in the design on existing practice within and between the involved organizations, the unawareness of the role of documents in keeping the chain together, their idea of a common world across the organization involved - there was no room for flexibility, as well as the fact that success depended on people outside the news agency – the TV channels – adopting the system, with the result that "those doing the work are not those getting the benefits".

However, because of the interdependencies in infrastructures, breakdown in one practice has an impact on practices in other organizations. This is clearly shown in our study. To avoid a breakdown in the whole chain of work, we propose to introduce gateways, which can function as a transformer between different practices and/or different technological solutions (Hanseth et al., 1996). In this way the practice and technology in one organization can change without forcing breakdown in other organizations. We introduced two such attempts to solve some of the problems discussed in this paper.

In the paper *Artifacts in work practice*, the artifacts' roles and meanings are further investigated by focusing on *how* artifacts in general become resources in work practices. We took our examples from investigation documents and related artifacts in practical use. The aim of this paper is to be more specific about the artifacts' properties and, based on that, how they become resources in the communities of practices. The paper combines concepts from the Actor Network Theory and the borderline issues.

See section 2.1 for further presentation of the theoretical concepts.

1.2.3 Approaches to standardization on SGML

The aim of the paper: *"Approaches to Standardization of Documents"* was to study standardization to SGML based on different approaches: "soft", guided, and enforced. The approaches are defined from the writers' perspective. The "soft" approach describes a situation where the writers produce text as before and the text is structured by others or partly converted to SGML documents by programs. The guided approach describes the quasi-standardization where the writer follows predefined style sheets, and other people or programs are used to convert the text to suitable structured and standardized documents. In terms of SGML, the producer deals with styles to "mark" text elements instead of using tags. The last one, the enforced approach, describes a situation where the writer must work according to the standard defined. In terms of SGML the producer must mark up the text using tags defined in the DTD.

An analytical framework was developed as a means to reveal and categorize limitations, problems and new possibilities seen from three different perspectives associated with the different approaches to standardization. The framework is based on the quality framework developed by Braa (1995).

The framework is as follows:

	"Soft" Standardization	Guided standardization	Enforced standardization
Use perspectives			
Organizational perspectives			
Technical perspectives			

Figure 1.1: Analytical framework: Approaches to standardization on SGML

Applying this framework may reveal different, problematic, and possible defective aspects of the approaches. It is then possible to get an overview of the resources needed to improve the process. The approach to choose will depend on a variety of factors, i.e. staff competence, organizational effort, technical support and technical infrastructure - or it could be dealt with as an organizational choice: where to implement changes in the organization.

The aim of the study, reported in the paper: *"Problems with Styles in Word Processing: A Weak Foundation for Electronic Publishing with SGML"* was to investigate the use of paragraph styles, in order to discuss whether current word processors are suitable applications for the capture of text in SGML. Based on our experiences we identified four categories of problems associated with the use of paragraph styles: overlooking styles, weak functionality, semantic problems and structural problems.

Our main findings were i) that there are great variations in the use of paragraph styles, ii) that in general paragraph styles are seldom used, and iii) that there are several problems associated with the use of paragraph styles.

The study showed that text from word processors is a weak starting point for automatic conversion to SGML and thus capture of important data. However, our

conclusion is more modest: one cannot simply assume that comprehensive use of word processing represents a simple way to capture text in SGML.

On a general level the "paper metaphor" of current WYSIWYG word processors can be used to explain the findings, but there are also problems connected to the way styles have been implemented in some word processors.

2. Research Approach

This research has emphasized an understanding of the transformation from paper to structured digital documents. In addition, some small-scale interventions have been conducted to test practical improvements as well as theoretical issues. The approach has been an Action Case, as defined by Vidgen and Braa (1997).

Research methods can be separated broadly into two categories: positivist and interpretivist. The positivist approach assumes that phenomena can be observed objectively and rigorously; good research is legitimated with reference to the virtues of repeatability, reductionism, and refutability (Easterby-Smith et al., 1991). In contrast, the interpretivist approach considers the methods of natural science to be inappropriate where human beings are concerned, mainly because different people (including researchers) will interpret a situation in different ways (Ibid.). In these terms, this research is typically interpretative.

However, Vidgen and Braa (1999) argue that in both positivist and interpretivist approaches the researcher is making an intervention, and there will be unexpected outcomes. In action research the aim is to gain learning and knowledge through making deliberate interventions to achieve some desirable change in the organizational setting. And, as they say, with any research initiative in in-context information systems - regardless of the research tradition and methods adopted - there will be unintended consequences and a degree of uncertainty about how the research project will play itself out. They argue, based on the philosophy of Habermas philosophy, that change can always be perceived to contain a critical element.

The three categories of research - positivism, interpretivism, and intervention - form the basis of the research framework for information systems (Vidgen and Braa, 1999.). Vidgen and Braa (ibid.) state, based on McGrath (1982), that it is impossible to have a research method that maximizes the three aspects of research. Thus, it is not possible for a researcher to make interventions as though she/he were entirely and indistinguishably part of the organization, while also being an observer who can stand back from the situation and make interpretations, and at the same time produce rigorous results in the positivist tradition.

The Action Case is an approach that focuses on understanding and change (interpretivism and interventions). However, in the Action Case only small interventions are conducted, compared to Action Research.

The theoretical perspectives grounded for both the interpretation and the interventions are presented in section 2.1. The empirical approaches are presented in section 2.2, while the change-oriented approaches are presented in section 2.3. The interpretations and the interventions are discussed in sections 4 and 5.

The empirical studies have been conducted partly in cooperation with Kristin Braa and three of our Masters students: Arnstein Schei, Bjørn Tore Kristiansen and Espen Berger. Bjørn Tore and Espen have programmed the prototype presented in section 5.

2.1 Theoretical perspectives

The purpose of information system design is to design a computer artifact, and to change the range of possibilities for action in the social organization that will apply the artifact (Ehn, 1988; Stolterman, 1991). The purpose of systems design is to formulate a vision of these changes, and the result of a design process is a set of specifications of computer systems and work processes (Andersen et al., 1986).

There is a fairly widespread belief that information technology simultaneously enables and constrains the restructuring of work organizations (Mathiassen, 1981; Orlikowski, 1991), However, it is difficult to foresee how the technology will impact the work practices. A principal reason for this difficulty is the contingent, interwoven and dynamic nature of the relationships, which also create challenges for conceptualization. Structuration theory (Orlikowski and Robey, 1991; Orlikowsi, 1992), phenomenology (Boland and Greenberg, 1992), hermeneutics (Klein and Lyytinen, 1992) and Habermas's theory of communicative action (Gustavsen and Engelstad, 1990) are examples of theoretical approaches applied to devote this problem. However, in terms of design, I find it necessary to know more about how information technology shapes, enables and constraints organizational changes, and to grasp the interplay between information technology and work in more detail. The beliefs that an interplay exists do not take us very far (Monteiro and Hanseth, 1995). To be specific about the technology, which is an aim in systems design, we need to incorporate a more thorough description and understanding of the technical properties and how these are translated into non-technical ones (Monteiro and Hanseth, 1995). As stated by Suchman and Trigg (1991), there is an intimate relation between work and technology. The development of the artifacts with which people work and the development of their work practice go hand in hand.

The point of departure in systems design and development is to get an understanding of the work practice where the future system will form a part (Ehn, 1988; Greenbaum and Kyng, 1991). Based on that, systems designers have to make some representation, which is challenging (Kyng, 1995; Suchman, 1995). A representation, as a DTD is in an SGML system, is the foundation for the system's properties, functionality and features. In an SGML approach, the design of a DTD is based on the content and structure of the documents, and how that structured information is applied. The information is separated from the artifacts itself. In contrast to the SGML approach, in this research it has been important to get an understanding of how documents are applied in work practices, and to get some indications of what will happen when these documents become electronic SGML documents as a part of an SGML system. According to the arguments above, it has been important to get an understanding of the interplay between human, documents and related artifacts in order to be specific about the documents' importance in work practices.

The theoretical concepts have been applied as part of an iterative process of data collection and analysis as well as being a foundation for interventions. The concepts are chosen based on the need to understand applications of documents in work practices. The concepts of 'heterogeneous actor networks' and 'inscriptions' are borrowed from the Actor Network Theory (ANT) (for example, see Callon, 1986; Latour, 1987; Law, 1986) to emphasize the interrelations of humans and artifacts, as well as the importance of artifacts' properties in these relations. The concept of 'boundary objects' (Star and Griesemer, 1989) emphasizes that an actor-network consists of several networks that are 'linked' by boundary objects. The concept of borderline issues (Brown and Duguid, 1994) searches for properties that lie beyond

the canonical meaning of the artifact, and in this respect make more 'hidden' networks visible, which improves the understanding of work practice.

Hopefully, the concepts can help researchers as well as SGML practitioners to reframe their own experiences of documents in use and to make progress in their understanding of design of document systems.

2.1.1 Heterogeneous actor network and inscriptions

In this research, work practice is regarded as a heterogeneous actor network, meaning that the work practice is simplified and represented in a way that highlights how humans and artifacts (non-humans) are intertwined in order to reach initial goals. In an actor network both humans and artifacts are *a priori* seen as equal parts of the social world. This approach argues that the society would not exist if it were simply social. In these terms, work practices are nothing other than patterned networks of heterogeneous materials (Law, 1992).

The artifacts' inscriptions define how humans and non-humans are interconnected in the heterogeneous network. The notion of inscription refers to the way artifacts embody patterns of use (Ackrich, 1992). Inscriptions may be properties, i.e., features, characteristics, possibilities and allowances inscribed in the artifact and that are of vital importance in the 'relations' between humans and artifacts. The inscriptions shape the social link between different actors and therefore influence the actors' performance. Thus, the artifacts' inscriptions will constitute the form and the substance of the actor's interactions.

By emphasizing the inscription, the properties of the technology become important. According to SGML systems, the properties of SGML are emphasized and regarded in relation to their context. In this way, we can analyze how standardized documents impact work practices, and be able to distinguish between the challenges forced by the properties of structure and standardization and those due to the fact that they become electronic. Using the concept of inscription makes it possible to be more specific about the technology and its impact. For instance, by evaluating the use of an SGML editor, one considers the editor's inscription - its properties - in the light of this use; one does not regard it merely as an editor. The SGML editor focuses on structure in order to achieve standardized documents, and this focus on structure impacts work. One can evaluate this impact by observing how these inscriptions influence the network (work situation) of which they form part.

By regarding work practice as an actor network, the interrelations and the dependencies of the involved actors or components are emphasized as well as how they are interrelated. Simply expressed, this refers to the changes that one work practice may force in another. Or, in more detail, how, for instance, technical artifacts are intertwined into larger technical arrangements. For example, at the news agency fax sheets are intertwined with shelves or binders. Fax sheets can be placed on shelves, because papers have inscriptions such as tangibility and lightness, which makes them easy to lay out in space. This interrelation or interconnection is visible, for example when parts are removed from the network. For instance, when the paper sheets are gone, there is no need for shelves, fax machines, binders, etc. How human and artifacts are intertwined in the news agency is further presented in section 3 and discussed in section 4.

The application of the concepts of heterogeneity, network and inscription make it understandable that how things are intertwined is not accidental, but the socio-technical arrangement is established over time to fulfil some intentions or aims.

2.1.2 Boundary objects

Documents are boundary objects. In general, a 'boundary object', as defined by Star and Griesemer (1989), is something that arises over time from ongoing cooperation between communities of practice, such as a way of resolving different definitions of things, situations, or problems. It is a complex set of arrangements between two or more cooperating communities, which satisfies their different information needs. Boundary objects are thus objects that are plastic enough to adapt to the local needs and constraints of the different communities, yet robust enough to maintain a common identity across sites. They are weakly structured in common use, and become strongly structured in individual-site use. The objects may be abstract or concrete. They have two important qualities: ambiguity of meaning across sites or communities; and durability in time as a set of working arrangements (ibid., Star, 1989). In this way a boundary object may hinder conflicts and create coalitions and alliances between various actors.

The concept of boundary objects is an improvement of the network thinking in the Actor Network Theory. The concept keeps in mind that, for instance, a work practice is part of several networks and has several intentions or aims to fulfill.

A boundary object can be interpreted differently in different work practices, but still be a part of the same overall aim or intentions of the work. The concept of a boundary object opens up the view of a network as a collection of networks. In the cases presented in this thesis, the humans and artifacts are intertwined in an actor network to produce a course catalog or TV schedules. In addition, the different work practices are separate networks. They have their own aims and intentions that are not shared by the other work practices involved. However, the catalog or the TV schedule is part of all these networks. The documents become boundary objects that on the one hand have local roles or meanings, and on the other are "shared" or external in a way that links these work practices together.

The concept of boundary objects is applied in this research to focus on the local and the shared aspects of the TV schedules and how they link work practices together. In order to standardize on document types, the concept of boundary objects helps us to be aware that documents can be applied differently in various work practices even if they are "shared" material.

2.1.3 Borderline issue

The concept of the borderline issue (Brown and Duguid, 1994) is applied to discuss the importance of the additional or unnoticed resources of artifacts beyond what is usually recognized as the canonical meaning. In general, artifacts have both central and more peripheral properties, and there is a border between them (Ibid.). What is recognized as a central or peripheral property varies in different communities of practice. Some users see properties as a central and natural part of the artifact, while other users of the same artifact regard the same properties as uninteresting and peripheral.

There are no clear boundaries indicating what one should see as a central part of the artifact and what is peripheral. How the artifact, e.g. the document, is regarded depends on the work practice. The border is defined as the dividing line between the central and the peripheral properties. Borderline issues are those resources that are shared and constitute a social meaning for a group of people. These unnoticed resources are developed over time as artifacts are integrated into ongoing practice

and social conventions are developed. The border resources are maintained by communities of practices, and are resources that workers often rely on (Ibid.).

Borderline issues develop over time based on material continuity and communities of common practice. Material continuity is needed in order to recognize the properties, and community of common practice is necessary for members to share, recognize and reformulate conventions (Ibid.).

The concept of borderline issues emphasizes the importance of looking for artifacts' peripheral properties and their significance in work practice conventions. By removing or substituting artifacts, these properties (inscriptions) may disappear and impact the work practice in unwanted ways. The concept of borderline issues makes us focus on properties that go beyond the canonical meaning of the artifact. The concept of inscription will explain why they become resources, since the inscriptions 'decide' the interactions with the other actors (human and artifacts). For instance the fax sheets have peripheral properties (and inscriptions) of being tangible and light. The central property is to share information between the TV channels and the news agency. However, because they are tangible and light they can be placed on shelves. Over time, the location of faxes takes on meanings beyond the canonical meaning of sorting information from TV channels. The employees use the fax sheets on shelves to keep track of the status of work: for instance, they see which fax sheets have arrived and which are outstanding. Convention to deal with the status of work is developed over time. And when fax sheets are removed, the shelves disappear as well as the convention that has been developed.

The focus on peripheral properties might find "other" and more "uncovered" actor networks, which improve our understanding of work practice. The description of a network only tells one side of the story. Applying the concepts of boundary objects and borderline issues enriches this perspective of actor networks. The boundary object focuses on local and shared properties, while the borderline issues focus on central and peripheral properties.

All the concepts presented give serious consideration to the artifact when work practices are analyzed. In this research, where the documents in work practice are the main concern, this is appropriate. The aim is to understand the transformation from paper to digital documents, and this necessitates concepts that take documents (artifacts) seriously and at the same time regard their properties as something essential for their use.

2.2 Empirical approaches

The study is based on the belief that the primary laboratory for information system research is organizations, where the development and use of information technology can be studied in-context (Ehn et al., 1995). In-depth case study is an approach to interpretative studies (Yin, 1994), which is the main approach in this research. The case studies are used to provide an understanding of work practices both before and to some degree after the introduction of SGML systems in organizations, which is necessary since the transformation itself is of interest.

2.2.1 Initial study at the University of Oslo

The initial study was located at the University of Oslo. A project was initiated at the University Center for Information Technology Services (USIT) to determine what type of electronic infrastructure could deal successfully with electronic documents and other forms of information at the university. In March 1993, to gain experience in the

practical use of SGML for some parts of the information produced at the university, USIT established a pilot project that involved developing an open and flexible solution for the production, exchange and distribution of the university's course catalog. The pilot project ended in December 1994.

The development process was evaluated continuously. During the project we interviewed 22 different people involved in the project, representing catalog writers across the departments and faculties as well as management people at USIT and the university's central administration unit. Some were interviewed several times. We had access to 393 email messages sent to a distribution list where writers could ask all kind of questions. We reviewed a number of questions and problems from writers' telephone calls and direct mail as well. We examined minutes of 13 meetings with the writers and more than 20 internal meetings dealing with more technical problems. In addition, we had access to 3 project reports written by the project members (not only the project leader). After finishing the project, we reflected on and recorded what we had experienced during the process, and documented it in a paper (see Sandahl and Jenssen, 1997). The interpretations based on this study have served as a point of departure for further studies as well as being influenced by them.

In this project I participated as a project leader and system developer. This made me familiar with the problems and challenges discussed, but also made it more difficult for me to keep a distance and remain objective about the subjects at hand. This was not necessary a bigger problem for me than for others in projects with an action-research orientation. However, experiences from this study may impact the interpretation of the introduction of the SGML system at the news agency, by influencing what we looked for.

2.2.2 Case study at a Norwegian news agency

This research is mainly based on an in-depth case study that was conducted at a Norwegian news agency from January 1996 to March 1998. The news agency is an organization providing news services to the press — newspapers, magazines, radio and television. The need for a new editorial system that supports electronic text exchange had been recognized in the news agency for several years. The existing system was based on old-fashioned mainframe technology, which is difficult for the IT department to maintain. The focus on a new editorial system intensified from 1993 where the International Press Telecommunication Council (IPTC) introduced the SGML-based "New Industry Text Format" (NITF) to increase electronic text interchange between news agencies and their customers / suppliers. IPTC recommended that all news agencies adopt the standard in their editorial and communication systems.

The news agency wanted to build competence within the organization on a minor project before continuing with an implementation for the whole organization. In 1995 the IT department, together with external consultants, took an initiative to apply SGML in the production of one of news agency's services: the TV schedules. The production of TV schedules involves a chain of work tasks. Starting with the production of program information in the individual TV channels, this chain involves merging, quality control and coordination in the Media Department at the news agency, ending in a variety of weekly products to be delivered to different newspapers and magazines.

There were two reasons for our interest in the news agency. Firstly, the news agency was going to introduce an SGML system into their production of TV schedules (at that time) in the near future. The design process was ongoing, but the system had not yet

been introduced. This meant that we had time to do investigations before the SGML-based system was introduced as well as after the introduction. However, we did not have the opportunity to impact the design process at that point. Secondly, the IT people as well as the employees at the Media Department, which is the department that works with the schedules in the news agency, were really forthcoming and interested in our research questions. We found this important in order to get access to the people involved.

In total, 23 interviews were conducted at the news agency, while 8 were conducted in the TV channels. The interviews lasted from 30 minutes to two hours. We interviewed writers, desktop publishers (those who use the desktop program to create the TV schedules) and managers at the Media Department, and writers and program planners at the TV channels. The purpose of the interviews was to create a broader and deeper knowledge of the users' work practice as well as their way of experiencing and understanding their own work and roles. The interviews can primarily be characterized as open-ended qualitative interviews (Yin, 1994). However, some focused interviews were also conducted to clear up points that came up during the open-ended interviews. No tape recorder was used. There were almost always two of us present at the interviews and both took notes. Each of us transcribed our own notes into a description immediately after the interview. The different descriptions were compared and contrasted. Questions and comments on the descriptions as well as inconsistency between them formed the basis for the more focused interviews later on.

About 70 hours of observations were conducted at the news agency. From February 1996 to December 1997, 10 days of about 4 hours of observation were conducted at intervals. In March 1998 a week of approximately 30 hours of observation ended the empirical study (so far) at the Media Department.

Observations were carried out to illustrate the complexity and multiplicity of performing the job and to show how work is related in time and space. The observations were made by following documents around in the work practice, as well as observing the employees in their work in order to understand what they are doing and why. We asked questions about their work along the way. We took notes, which were 'transcribed' into emails when the investigator was back home. The emails were sent to the participants in our research group.

When errors occur in the system or the writers have questions about it, they usually call the IT personnel. But some emails were sent as well. We went through these emails. In addition, we had access to the project plans and reports. The reports, in particular, were used to get information about technical problems that occurred. In addition, a workshop was carried out in the Media Department at the news agency to get an overview of their work practice, and to increase the understanding among IT staff, managers, writers and researchers of the tasks writers perform that include technology use.

The case description is based on all the presented sources of empirical data, and appears in section 4.

2.2.3 Small-scale case studies

As mentioned, one of the results from the initial study was that the writers had problems dealing with SGML and the SGML-editor. We found it interesting to evaluate different approaches to SGML. Kristin Braa and I carried out interviews at three various sites that have different approaches to SGML. The university case

represented one of the approaches. Inspired by previous research by Braa (1995), we interviewed writers, managers and technicians in order to evaluate the approaches from a use, organizational and technical perspective. The other two sites were a major public-sector organization and an administrative unit at a large educational institution. See Braa and Sandahl (1998) for further details.

Pål Sørgaard and I investigated the writers' use of style sheets or templates, and evaluated this use in terms of conversion to SGML. This approach is one of the three approaches studied by Braa and Sandahl (1998) as mentioned above. However, in this study we went into more detail regarding the use of styles from a technical perspective. We used our established contacts with three different organizations to get access to documents and to conduct interviews with users. We analyzed the use of paragraph styles in a set of documents. Document analysis was primarily performed manually by opening documents with the word processor. When going through the documents, we looked for use and non-use of styles, redefinitions of styles, manual formatting, and all the aspects of paragraph styles we could think of. For some of the documents we conducted follow-up interviews with authors or other people involved. See (Sørgaard and Sandahl, 1997) for further details. Both these studies are important for the interventions presented in section 5.

2.3 Change-oriented approaches

Based on the interpretations, some improvements of the SGML approach are suggested to provide better support for the transformation from paper to SGML documents in work practices. These improvements form the basis for the interventions presented in section 5.

The theoretical perspective presented in section 2.1 is the foundation for the interventions as well. Simply put, it has been important to make improvements that are based on the application of documents at work. The documents and related artifacts have properties that are essential to the work process. When we change these properties, as we do when documents become electronic in a computer system, the work will change. However, how it will change depends on the properties inscribed in the new system even if there is no guarantee that the system will be used as desired. Designers have to strive for system inscriptions that fit in with existing practice and the objectives of changes.

When we separate the content from the document artifact itself, and make the necessary abstraction and formalizations based on that, several meanings and roles of documents in practical work get lost. The foundation for intervention is to regard the sociotechnical network and the significance of artifacts in that respect.

Experiences from the studies of style sheets (Sørgaard and Sandahl, 1997; Braa and Sandahl, 1998) are significant to the design of the prototype. The improvements have been evaluated according to use, organizational and technical perspective as recommended by Braa (1995).

3. Empirical description of the production of TV schedules

In 1995 the news agency's IT department, together with external consultants, took an initiative to apply SGML in the production of one of the news agency's services: the TV schedules. These schedules, produced for newspapers and magazines, were

regarded as the most structured and well-defined product they produced, and thus assumed to be easy to standardize in SGML. In addition, the Media Department was not satisfied with the way the schedules were produced at that time and had stressed the need for changes for a long time.

The production of TV schedules involves a chain of work tasks. Starting with the production of program information in the individual TV channels, this chain involves merging, quality control and coordination in the Media Department, ending in a variety of weekly products to be delivered to different newspapers and magazines. There are approximately 40 TV channels, and the number varies as TV channels close down and start up.

From the news agency's point of view, there was an urgent need to facilitate the document exchange process between the TV channels and the news agency. Almost all schedules from TV channels arrived as faxes. However, some were transferred by mail or electronic devices as well. The information was retyped and encoded into an old mainframe system, where the schedules were edited before being converted for import into a desktop program to create pages ready for printing, or transmitted as data to the newspapers and magazines. The electronic exchange of TV program schedules from TV channels to the news agency was expected to result in enormous time savings, because the people in the Media Department would then not have to retype the text, but only to improve the quality.

The project started by including the Norwegian TV channels, except the major national TV channel. The Nordic Channels also deliver their information by fax, but they were not included in the pilot project. Nor were the other international channels. Information from the international TV channels is received from a firm called Intermedia.

The intention of the project was to make a faster and better production process within the news agency, but also to make the production process more effective by standardizing the information from the TV channels. Standardization makes it possible to automate some work tasks in the Media Department.

The goals of the project were: i) to make it possible to write the information only once; this meant that the text written by the TV channels should be directly applied in the Media Department as well, ii) to enable some automatic validation by defining the elements in the TV schedules (for instance, check whether a channel has more than one program that starts at the same time at the same day), iii) to enable using the definition of the text elements in a way that made some manual procedures unnecessary. For instance, one could automate the conversion into QuarkXpress files, and make sure that the text length fits the available space in the style sheet. iv) to create new services. SGML makes it possible to present the information in new media - for instance on the Web, where people can search the TV information.

3.1 The work practice at the TV channels

The different TV channels are organized in various ways. However, they all send out information about their TV programs to the TV viewers in one way or another. They all produce TV schedules; however, the schedules may vary in content and form. In the TV channels the production of schedules is most often done in two steps. The first step is to decide what kind of programs are to be broadcast at what time. The second step is to add a variety of program information to the schedules. At a particular time, usually 2-3 weeks before the programs are broadcast, the schedules are regarded as ready and sent out to different recipients, including the news agency.

The number of recipients varies from 30 - 200 depending on the channel's size. In fact, many newspapers and magazines get the TV information directly from the TV channels in order to prepare reports related to forthcoming TV programs. The same newspapers and magazines buy schedules from the news agency some time later. The TV channels also send schedules to people who are interested. They want the schedules to be attractively formatted. Most of the channels create their schedules in Microsoft Word.

Each TV channel has a particular profile with respect to content and the presentation of the channel's identity. There are differences in the information that the channels want to emphasize, e.g. some channels want to emphasize the actors starring in a film, while others emphasize the film itself. In practice, one channel presents the actors before the film title in the schedule, and another presents the title first, or one emphasizes the title in bold, others emphasize the actors. Some channels use a great deal of text to introduce a special sporting event; others just mention the participating teams. They all have a profile indicating the type of channel they represent. In addition, the way that they present programs varies. For instance, if there is a new series, they usually want to use more space than if the series has gone on for years. And the programs that are broadcast in the middle of the day or night are more superficially presented than those broadcast in prime time.

Some channels broadcast almost the same programs every week, while others have more varied schedules. They seldom write the information more than once; they copy and paste between documents. For instance, the series are usually broadcast at the same time every week. In this case, only the series number and the name of the episode need to be changed. The information about the news programs usually does not change at all.

3.2 The work practice at the Media Department

The Media Department at the news agency improves the TV information further. Even though the TV channels spend a great deal of time on the production of the schedules, they include all kinds of errors, inconsistencies and poorly written text. The employees at the department go through the information and correct spelling and grammatical errors, edit long sentences, and polish the text. The previous week is used as the basis for data entry in the mainframe system. There are always some recurring programs, so "cut and paste" functionality is almost always used as the starting point. They state that more than two-thirds of the work is done by copying from the previous week.

The Media Department produces three different kinds of products based on the TV information. Firstly, customers can get the TV information more or less directly from the mainframe system, called ATEX. Secondly, the Media Department offers converted text to formatting languages that some customers use to produce their own pages. And thirdly, they produce print-ready pages in the desktop program QuarkXpress. The print-ready TV schedules are offered in various formats as well. Six different style sheets are created to represent the different schedules. Some style sheets cover a whole page in the newspaper, others only half a page. Style sheets also represent differences in how much space each channel has available. For instance, some style sheets give the Norwegian channel more space than the other channels.

It is important for the Media Department to deliver schedules of high quality to the customers. This means that the information is updated and correct as well as that the

pages look good and are easy to read. The TV channels often change their schedules just before the deadline of some magazines. The Media Department wants to make the updates as late as possible. Another central issue within the department is to make sure that the 'ShowViews' are correct. ShowViews are applied to program home video players. Every program has own ShowView[4].

As mentioned, the TV channels send their TV schedules to the news agency almost entirely by fax. One channel has its own solution, called Data-TV. All the faxes are sorted by channel and placed on a shelf. There are three different types of shelves: the in-shelf, week-shelf and out-shelf. The in-shelf has a space for each TV channel. When a schedule is received, the fax is placed on the shelf where it belongs. The in-shelf provides an overview of what was received and what is missing. When the faxes contain information about TV schedules for the week the employees are working with, the fax is placed directly on to the week-shelf. Otherwise, the faxes are moved from the in-shelf to the week-shelf when it is time to start working with them.

The week-shelf is not separated into channels, but only into 'to write', 'to proofread', 'corrections' and 'next week'[5]. For every week, regardless of when it starts and ends, information about all programs every day for each channel has to be retyped. When the work is done, about 6-9 different people within the Media Department may have been involved.

When the week is finished, the faxes are moved to the out-shelf. If there is time, a controller takes the faxes out of the out-shelf and checks them against the text in the mainframe system. From the out-shelf, the information goes into the trash. When the faxes are in the out-shelf, the information is considered ready for preparing print-ready files in QuarkXpress.

Figure 3.1: Parts of the in-shelf and the week-shelf

The TV channels send schedules for the whole week at once. However, the definition of a week is not uniform. What denotes a week varies among the different stakeholders. Most of the TV channels use a week that starts on Monday and ends on Sunday. The Media Department operates with a week that starts on Saturday and ends on Friday. The week in the newspapers and magazines depends on which day TV guides are published. The TV guides consist of information about radio and TV programs for "all" channels for a week. The "different" concepts of a week force the Media Department to operate with a week consisting of 11 days.

The huge number of faxes, the different time frames for newspapers and the magazines' different publications, as well as the various types of schedules, make the

---

[4] For more information, see: http://www.gemstar.co.uk/en/showview/home.html
[5] There are some other labels as well, but the spaces are not in use.

coordination of work within the agency quite complex. A coordination form is therefore applied to formalize the complexity in a manageable way.

Figure 3.2: The coordination form

The coordination form is marked when the TV information is retyped into the mainframe system and also when it is proofread. The 'done by' and 'corrections by' columns are signed respectively. The last column, 'comments', is used intensively. The information for a week is not always retyped at once, and in this right-hand column the employees mark what they have done on which days. If the proofreading is not finished, they write an 'O' in the forms' left-hand column. When the work for a channel's week is finished, the employees delete all the comments in the 'comments' column using correction fluid. They register the date when the last work was done, as seen in figure 3.2 above.

The huge number of faxes, the different concepts of weeks and the number of employees require coordination of many intertwined and interdependent activities to make it possible to create the schedules in time. The various actors in the Media Department coordinate work activities that take place in a different time and space - they articulate the activity affecting the other actors in producing the TV schedule for a week without being in the same time and space. In practice, moving faxes from one shelf to another, and signing the coordination form accomplishes this.

The shelf in which the faxes are placed, assimilable at a glance, represents the state of work. When the faxes are on the in-shelf, no writing has started. However, the staff notices whether all TV schedules have arrived or not. When the faxes are in the out-shelf, the registration is complete. In this way the faxes on shelves indicate progress and trace the states of the work process. In addition, the work process is visible to all the workers at the Media Department.

This visibility is important in the division of labor and job rotation at the department. The workers can see who is working with what, and they are aware of how much a person has typed in the last few hours. Since they do not want to assign all the typing to one person alone, they change when they sense it is their turn. The community of practice relies on this implicit information to carry out the work.

The employees at the Media Department are organized in the Graphical Union, and they do not want the entire typing workload to be allocated to one person, so that job rotation is important. The coordination form and the shelves simplify this job rotation.

There is only one ATEX file for each TV schedule, and this file gets converted according to the different needs. Every TV schedule is represented in at least five different ways within the Media Department. The reason is that the customers want different products and that the time for delivery depends on the customers. This makes the updating a challenge, if the changes from the TV channels come after the conversions are done.

3.3 The SGML-based Document System

Early on, it was decided that the document exchange system should be based on SGML, because of the IPTC's recommendations. The modeling of the document – the construction of a Document Type Definition (DTD) – was done by consultants, working together with the IT manager to some extent. No writers were included in the DTD-modeling process, even though the DTD is the core of an SGML-based

document system, and affects the system's design and use. The level of structure was determined on the basis of discussions with the IT manager about future possible services, e.g. on the Internet. The TV schedules, presented in newspapers and magazines, were the starting point for the DTD design.

The aim was to ensure that all program schedules consisted of the same information and were structured in the same way. After validation of the document according to the DTD syntax, scripts can be converted for desktop programs or other applications automatically, and provide facilities for advanced search such as "on demand" documents.

Since the DTD specifies a model of the document, the text producers both in the TV channels and in the Media Department have to produce the document according to the predefined structure, normally using an SGML editor. An SGML editor "interprets" the SGML standard and has mechanisms to help the writers to encode the information in relation to the predefined DTD. In this case the text producers use Author Editor. SGML editors were installed in the TV channels, being the beginning of the work chain. The news agency provided the SGML software and training, to motivate the TV channels to use the system.

Figure 3.3: The user interface in the Author Editor

The files were to be saved in a particular file structure on a server: /channel/week/day/file name. The channel, week number, day and some status information form the file name. The idea was that computer programs would pick up the right files and convert them into various style sheets for a desktop program for making print-ready pages.

Later, the intention was to have an information system based on the information in the program schedules to offer the press the opportunity to retrieve the information they wanted directly from a database. The information could also be presented on the WWW. This could provide "on demand" services for the public, e.g. selecting movies in the thriller genre with a certain actor. The newspapers and magazines should not notice any differences.

The SGML-based system was never adopted in the way it was supposed to be. Many channels withdrew from the project, and the people in the Media Department were not very happy with it either. However, some channels tried using it. Even though it failed, and the system was not used as expected, we had the opportunity to make some observations and conducted some interviews with people who had tried the system. Our data from those observations and interviews are presented in the discussion below.

4. From paper to SGML documents

The SGML approach to system development is motivated by a desire to represent exact and explicit information in an structured and objective way, more or less independently of its context of use. This is a rather mechanistic view of system design (Dahlbom and Mathiassen, 1993). This approach to modeling is in line with approaches such as data modeling and use of flow diagrams. The DTD is an abstract model of a document structure, which separates the information elements from its context as well as its physical appearance. To design document systems, this separation is necessary, and there is a need for some structure and formalism in order to benefit from computer capacity.

This section discusses the transformation from paper to SGML documents and indicates some problems with the SGML approach in this respect. The automation is a central goal for introducing SGML. From this perspective, the aim is to replace documents or entire document-production processes with computer-based representations to improve the production, storing and utilization of documents, and hence to make the organization more effective and efficient. This was the aim within the catalog and the TV schedule projects as well. However, removing paper documents just like that is not necessarily trivial, because people may rely on them in ways that are unexpected. This is further discussed in section 4.1. It emphasizes that the information element structure of documents is not necessarily the only important aspect to consider in the design of document models. The experiences from the empirical studies are discussed according to the theoretical concepts presented in section 2.1.

This illustrates that a paper document is technology with several and diverse inscriptions that are essential for people carrying out their work. The central property is to share information between people independently of time and space, and documents consist of more or less structured information in this respect. However, they also have the properties of being tangible, light, etc, which make them adaptable - as well as giving them different meanings and roles in different contexts.

When one separates the information (data) from the artifact (paper), many of the paper documents' properties disappear, which forces changes in the work practices. In addition to becoming digital, the SGML documents are structured according to a shared model, which also standardizes the document across work practices. The inscription of structure impacts the work practices in several ways, which go beyond the fact that they become digital. The writers find it challenging to follow a DTD since the model does not necessarily fit their perception of the document. This is further spelled out in section 4.2. And the standardized document model allows only one representation or perspective of the document, which restricts the documents' boundary properties. This impacts the interdependencies of work, which is discussed in section 4.3.

4.1 From paper documents to digital documents

According to Actor Network Theory, humans and artifacts are interrelated at work to fulfill some aims or intentions (Law, 1992). How artifacts, such as documents, are embedded into work practice depends on their inscriptions. All artifacts have inscriptions such as intentions, properties and features, and many of them are more or less generally agreed upon. For instance, documents are produced and used for the need of sharing information. However, as we have observed, documents may have more peripheral properties as well. These peripheral properties may become important common resources that the communities of practice rely on (Brown and Duguid, 1994; Braa and Sandahl, 1998b; Lundberg and Sandahl, 1998). When documents become electronic some of these resources disappear.

In the news agency, artifacts were introduced to organize work in such a way that the employees can produce the schedules in time. Fax machines were introduced to make it easier and faster for TV channels to deliver their TV information to the Media Department. Shelves were introduced because fax sheets had to be separated by channel and by week. The coordination form was introduced to help coordinate a complex process.

22

However, over time these artifacts acquired roles other than mere organizers (Lundberg and Sandahl, 1998). Computers, shelves, fax sheets and forms are visible and present in the rooms where people are working; they are also essential in the individual's understanding of the activities of the others, which provides a context for their own activities (as defined as awareness by (Dourish and Bellotti, 1992)). For instance, the process of moving faxes between shelves makes the personnel aware of the status of their work as well as who is working on what. This awareness helps to keep the articulation of work on a relatively low level (Braa and Sandahl, 1998b) and becomes a resource that the community of practice relies on.

Arrangements of artifacts, such as documents on shelves, are important for the awareness of human activity. When documents, as well as other artifacts, are represented in computer systems in one way or another, there is no need for shelves, tables or fax machines at the workplace. The visible paper documents are gone; the process of carrying them, sorting them or using them becomes invisible as well. Implicit information necessary for peoples' awareness is gone. Instead of getting an overview of the progress by simply glancing at the shelves, the employees have to browse through 350 files for every week to see which files have arrived and what is still missing.

The document as an artifact may have other resources as well. For instance, a document can be a coordinating artifact. At the Media Department, the coordination of work is indicated by 'who is holding the fax sheet'. The paper acts as a token and the shelf in which the documents are placed represents the state of work. The coordinated role of the document supports the progress of work, and it becomes essential in the division of labor.

The significance of paper in the Media Department is illustrated by the way that the department managed the corrections from TV channels that arrived as emails. The emails were sent to one particular PC at the department, and it is logical to think that each email could be forwarded to a particular person for updating. However, the email was printed out and placed in the correction shelf or on someone's desk, depending on when in the production process the corrections arrived. The person who picked up the mail on paper, or received it on her/his desk, was responsible for the corrections. The email was printed out to maintain the coordination property, in the same way that the property was embedded in the faxes.

The articulation work became more complicated, and additional coordination mechanisms became necessary when these things changed. To organize the work, the coordination form was extended and became more essential. The personnel were totally dependent on the coordination form to keep track of the work process, and started using the form in new ways. Previously, they signed only after the week was finished, but after the introduction of the SGML-based system they signed for every work task performed. The coordination form was a way to get explicit information on what was happening, which made it an important artifact in work cooperation. Similar is experienced by Carstensen and Sørensen (1997) and Schmidt and Simone, (1996) as well. As illustrated in the case studies, this explicit coordination was not enough to keep track of the work. The more implicit coordination from the faxes on shelves was very important as well.

In a way, documents are active actors within a community of practice. They are not only "passive" information sources. The fact that documents are tangible, light and ecologically flexible has implications for the ease with which they can be physically transported within the communities and laid out in particular spaces (Luff et al., 1992; Harper and Sellen, 1995). When and if documents are placed in specific locations,

23

they represent signals. What happens on the fax machine, the number of faxes received, and the shelves in which they are placed, are factors of vital importance to how the working day will develop. Replacing paper documents with electronic SGML documents gave the documents other properties, which did not fit in with previous use and existing technical arrangements such as shelves and fax machines (Lundberg and Sandahl, 1998).

The one-sided focus on structured information to automate and improve search and retrieval facilities "black boxes" the production process into a simple system: "information is input, it is merged and presented or utilized in an appropriate way". An SGML approach overlooks the fact that the work process consists of a complex network of heterogeneous actors that have evolved over time and where conventions have grown around the use of artifacts. The lack of paper impacts the flexibility of documents in use. Documents only took on the significance of the 'structured information source'. We have to remember that there are other properties of documents that are essential for the conventions that grow around them as well as how they become resources (Lundberg and Sandahl, 1998). When we overlook them and regard documents as sources of information only, some of these resources disappear. And work practices may break down.

This subsection illustrates the importance of regarding artifacts as embedded parts of work practices to gain a better understanding of their use. It also illustrates the danger of separating the content from the document, and basing the modeling on that alone. A paper document is a carrier of information that goes beyond the text written on it, which is hard to recognize if we do not look at its use in real work settings. Paper documents on shelves support awareness of colleagues' work, support the articulation of work and provide for division of labor. When documents become electronic they lose properties that seem to be important in work practice. How these properties become important has to be taken into account in order to understand the use of documents in real work. The next section discusses impacts in work practices based on the inscription of a digital SGML document.

4.2 Impacts of inscribed structure in digital documents

Digital documents acquire new properties that will impact the work practices in new ways. First of all, electronic media differ significantly from paper ones. The medium itself has properties that impact document production and use (Hendry, 1995; Bolter, 1991; Ong, 1982). Secondly, the designer gives the documents new inscriptions by defining document models or in other ways preparing for the production and application of documents. According to SGML, the DTD specifies the properties of SGML documents' properties, and the documents' structure is essential in this respect. The DTD also specifies the concepts applied in describing the document's content.

This subsection discusses the challenges of following a predefined structure and adopting predefined concepts in the production of documents. The challenges are based on the experience that a model does not necessarily fit the writers' perception of the document. For instance, there is a common understanding across work practices of how TV schedules should look. However, for producers there might be small, yet significant differences in the structure and content within the particular document instances. These differences are grounded in the practice producing and using them.

The TV channels structure their TV information in different ways, even if the differences are small. The TV channels use the schedule to profile their channel. A film channel is interested in presenting the film and the actors in detail, while other channels only mention them. If the film is well known, the channels might find it important to describe the film before the actors are presented. In situations where the actors are well known, but the film has a lower rating, they find it important to emphasize on the actors involved and not on the film itself. With the TV schedule DTD, the first description is possible, but the latter is not. How the TV channel presents their information is important to them; they use it as promotion (Braa and Sandahl, 1998c). Similarly, there are also several different ways to represent the same information in a course catalog. At the university, there was a great deal of disagreement among the writers on how to represent information about the day, time and place for lectures. Some wanted to present it in a table, while others wanted prose (Sandahl and Jenssen, 1997). Since there was only one catalog DTD and one TV schedule DTD, the differences in structure were not supported.

The concepts used in describing the TV schedules vary between the TV channels as well between the TV channels and the news agency. For instance, the title of a program was termed a 'title' in some channels and 'title line' in others. The same applied to concepts such as 'date', which was also termed 'day', and 'actor', which was termed 'participant' as well. There were 18 attributes to choose from in identifying the genre of a program, and the practice of classifying differed among the work practices involved. For instance, a program for 'youth' may also be a 'documentary', and a program of 'music' may also be 'entertainment' and 'culture'. The classification of information across work practices and organizations is also discussed by (Bowker et. al.; 1995; Hanseth, 1991). The numbers of attributes and elements needed differ between the various channels. For instance, a film channel needs only elements that help to identify 'title', 'actors', etc. This is hardly enough for an all-round channel. The differences in meanings and the number of concepts are not supported when there is only one common document model.

As illustrated, the writers find it hard to follow a predefined model because it does not reflect their perception of the documents they are supposed to produce. The use of a native SGML editor forces the writers to follow the predefined DTD. An SGML editor provides correct and consistent documents according to the DTD defined. The Author/Editor was applied at both the university and the news agency. Use of an SGML editor to approach standardization is mentioned as an 'enforced approach' by Braa and Sandahl (1998).

The writers hesitated to start using the SGML editor. First of all, the SGML editor compelled the writers to follow a document model that was more or less unfamiliar in terms of their own understanding and aim of the document in use. Secondly, an SGML editor emphasizes the structure and inscribes a writing behavior that does not fit into the writers' way of working. They had to do a great deal of extra work without getting any benefits for doing so, which leads to lack of motivation (Grudin, 1989; 1994).

The use of an SGML editor requires discipline in the way that text is written. The writers have to follow the DTD 'all the time' when they are writing. This means that the documents have to be consistent according to the DTD at all stages of the writing process, and not only in the final version of the document. This is experienced as challenging, since people do not produce text in a linear way (for example, see Haas, 1990).

The focus on correct and consistent structure also impacts the editor's copy and paste facilities. The copy and paste facilities are restricted, and are not allowed if the predefined structure gets damaged. It is possible to turn off the "rules checker". However, this is not the practice. As stated in section 3, the copy and paste facilities are very important for the writers, since they take previous schedules as the point of departure for producing new ones.

The use of an SGML editor forces the writers to relate to all the elements and attributes defined in the DTD. Even if the SGML editor is context-sensitive, meaning that it keeps track of which elements the writers can apply according to where they are in the text, there are many elements that are visible to the writers, but that they never use. The TV schedule DTD consists of elements that cover all that the channels need (and more) for describing their TV information. But a film channel has no need for elements or attributes that are used in describing news, and a sport channel does not need concepts to describe films. The writers drown in elements and attributes, and they find it hard to know which to use.

In addition, the focus on detailed structure documents and the introduction of an SGML editor turns the text processing into a kind of 'data punching'. The SGML editor resembles a user interface to an SGML system rather than an authoring tool in a writing process. The interface of an SGML editor is seen in figure 4.3. In this case, the writers have to fill in information about ?M?. When there is standardization at a very detailed level, as at both the university and the news agency, the writing becomes more or less a process of filling in a database schema as opposed to using a word processor where the writing involves building sentences.

The DTD represents a shared model for various practices and it represents only one perspective of the document, which is forced on the writers by the application of an SGML editor. This is challenging. The SGML document has lost its boundary aspects. The model represents the shared perspective, while the local or individual perspectives disappear. When documents are regarded as a collection of structured information only, the heterogeneity of work practices is overlooked and it is assumed that documents have common 'meanings' and applications across work practices, which they do not necessarily have. In this case, there is a lack of agreement on what TV schedules 'are'. The previously allowed differences that create the agreement are gone.

The next section discusses how the transformation from paper to structured digital documents impacts the interdependencies of the work practices involved.

4.3 Impacts in the interdependencies of work

According to ANT, changes in artifacts' inscriptions impact the network they belong to. SGML documents have properties different from paper documents, and substituting paper documents has consequences for the work processes they are parts of. The previous subsections illustrated that there are challenges in switching from paper to digital documents, because conventions have grown around paper documents over the years and they are embedded into work practices in several ways. And the emphasis on information structure, which the SGML approach stresses, makes it hard for the writers to perform the work in the way that they want, which decreases their motivation for starting to use it. This subsection illustrates that the transformation from paper to SGML documents also impacts the interrelations between the work practices involved.

Starting at the TV channels, which represent the beginning of the work chain for producing TV schedules, the introduction of SGML separated the production of TV schedules sent to the news agency and schedules sent to other recipients. The same document existed in two versions, depending on the technology applied to produce them. One work task became two, and this was regarded as extra work. This is another example of the lack of boundary properties in SGML documents. The designers of the SGML system did not uncover the interrelations of work tasks that the writers had to perform, and how these tasks were connected by the document as a central actor.

Through the introduction of SGML documents, the writers at the TV channels became parts of a relatively large information system. It was not only the technical devices that were integrated, but the people, work practices and organizations as well. The writers had to produce their TV information in SGML to support a more or less automatic production of TV schedules in the news agency, which they hesitated to do (Braa and Sandahl, 1998c). As stated above, writing of SGML documents according to a predefined structure and content is challenging. And it is even more challenging to motivate writers to produce SGML documents in one organization if it is other organizations that reap the benefits (Grudin, 1989; 1994). This was the situation at the TV channels.

This interconnectivity has consequences for the employees at the Media Department as well. Before the introduction of SGML the employees had to concern themselves only about the TV schedules' content. With SGML, the structure of the TV information became important as well. If the TV channels delivered documents that had errors according to the predefined structure in the DTD, the Media Department had to correct these as well before they could start to improve the text and produce the schedules in the desktop program. In addition, as discussed in section 4.1, the lack of fax sheets impacted the awareness of work, the articulation of work and the division of labor at the Media Department.

The introduction of SGML documents improves the technical scope for making new and advanced technical solutions by defining document models and by integrating technology into large sociotechnical arrangements. Use of an SGML system presupposes that the production and application of documents are similar across work practices. The boundary objects that allow the work practices to have their own perspectives and perceptions are gone. The collection of the various networks becomes one network. This is challenging and even unacceptable for the various work practices involved.

To benefit from introducing SGML, someone has to create the SGML documents. At the news agency and the university much effort was spent on getting the writers to produce highly structured SGML documents - so much effort that there was little time to implement new functionality, features and services. At the university, no new services were implemented during the project period of almost two years. Working hours were devoted to improving the DTD so that writers could deal with it when writing, to improving printing facilities, to provide training and support and to organize a new work flow (Sandahl and Jenssen, 1997). The same happened at the news agency. From the time that we entered the project in early 1996 until this thesis was written up, no new services or functionality were developed in order to apply the TV information in new ways. The effort was spent on producing the schedules, which they had done perfectly well before SGML was introduced.

It is challenging to provide some flexibility in the production and application of documents. This is necessary to allow the production of TV schedules that include

several networks and do not limit it to only one. It is also challenging to ensure that the writers get some of the benefits of using the system, to motivate the production of SGML documents, which is the foundation of the functionality and features of the system.

4.4 Challenges of the transformation from paper to SGML documents

According to the discussions above, the challenges of substituting paper documents with SGML documents are founded in three different, but related aspects. Firstly, we substitute two different types of technology with different properties and features. By removing paper documents, we also remove resources that go beyond the canonical meaning of the artifact. These are resources that the communities of practices rely on, and that have to be taken into account in the design of document systems. The document model represents one perspective of the document, which is hard to combine with the various applications of documents in use within and between work practices.

Secondly, the document model represents a collection of structured information, based on an abstraction from the content presented in the document, which is too restricted according to the various perspectives of documents in practical use. The emphasis on structure complicates the production of documents.

Thirdly, the application of shared document models across work practices turns the various heterogeneous actor-networks into one network, which requires a common objective among the actors involved. The dilemma of "who does the job and who gets the benefits" (Grudin, 1989; 1994) emerges. The dilemma impacts the users' motivation for starting to the system. It is a challenge to make sure that users can benefit from it right from the first day of use.

The challenges mentioned relate to the work practices needed to provide some flexibility in the production and application of documents at work, which conflicts with the technological need for structure, formalism and standardization. The lack of flexibility may have the result that the systems are not used. People continue working as before, as we have experienced in our case studies, or they find various other workarounds to avoid using the system (Gasser, 1986).

Previous research in system design has illustrated that deficiency in the understanding of work practice often contributes to computer system failure (Greenbaum and Kyng, 1991; Ehn 1988). The main complaint against the systems design process at both the university and the news agency was that the design was based on inadequate understanding of the work. However, it is not trivial to define cooperative work and even more complicated to understand what it is. Many aspects of the work are even 'invisible' (Suchman, 1995). Various research studies emphasize the importance of taking into account aspects such as tacit knowledge (see e.g. Suchman, 1995; Sachs, 1995), articulation work (Schmidt and Bannon, 1992) and implicit awareness (Heath and Luff, 1992; Hughes et al., 1992) in the design of computer systems. The Scandinavian tradition approaches this problem by involving future users in cooperative design processes (see e.g. Ehn, 1993). In this way there is mutual learning between the designers and the users, and the users can set requirements for the system based on their knowledge and experiences of their own work practice. I believe, as illustrated in this section, that studies of socio-technical arrangements help to improve the understanding of work practices as well, including their more hidden aspects. For instance, by regarding the central, peripheral, local and shared properties of documents, it is possible to achieve an understanding of how

documents are embedded in work, including the importance of documents and related artifacts to awareness, articulation and coordination of work. This observation contrasts with the SGML approach to documents, where the document content is separated from the artifact itself. Because of this, there are several roles or meanings of documents that remain invisible.

The next section discusses how to improve the SGML approach by using 'gateways' to provide more flexibility in the production and application of documents at work. The gateway design is based on a consideration of how properties of documents and related artifacts may impact their surroundings.

5. Improving the SGML approach by using 'gateways'

The transformation from paper to SGML documents is challenging. As stated in section 4, the SGML approach focuses on the application of the documents' content and overlooks the application of the artifact as such. The SGML systems designed at the university and the news agency are designed based on that principle. It is emphasized that it is necessary to have an understanding of the heterogeneous network of which documents form part. In this way it is easier to consider what the future system should look like in order to transform the work practice in the desired direction.

Hanseth et al. (1996) introduce the concept of 'gateways' to deal with the tension between standardization and flexibility. The aim of gateways is to allow different communities of practices to apply different technical solutions that serve various work organizations, but at the same time enable them to communicate and interact independently of the differences. In terms of this research, the local flexibility in the production and application of documents is improved by allowing documents to have peripheral and local properties in addition to the central and shared ones.

By focusing on the existing sociotechnical arrangements, we get an idea of the prerequisites for the work being done. The properties of artifacts are given serious consideration, and regarded in the context of their surroundings. The properties are vital for determining how things are interrelated into heterogeneous networks. The previous section illustrates how a document's inscriptions are essential for its production and application in use. An understanding of the prerequisites helps us to understand how the computer system can fit into work practices, even if we have no guarantee that it will be used in the way that we expect. According to design, the work practices are improved by changing the technical properties or the technical fundamentals, by adding various inscriptions to the system.

This way of thinking emphasizes the properties of the product itself and how these can impact the existing work. The designer's main task is to enhance technology with inscriptions that hopefully make it easier to embed the system into ongoing work to transform the practice into a "better one". In this way the designers become specific about the technology and its impact. The design of gateways is an attempt to change the SGML system's inscription in a way that makes the system easier to use by decreasing the level of structure, increasing the local flexibility and making the borderline resources explicit.

Since the DTD is the core of an SGML-system, a work-oriented DTD design is probably preferable. The DTD should not only represent the structured information, but also the application of documents at work. Making a DTD more flexible, by untying the structure and simplifying it, impacts the technical properties and thus the work itself. How to make DTDs more work-oriented is an issue for further research. An

improvement of the SGML language itself might strengthen the document model - for instance, by making it more like object-oriented languages (e.g. including mechanisms of generalization and letting the element declarations include actions as well). However, this has not been an issue for this research.

Below, three different gateways are presented: style sheets, sub-DTDs and interfaces. These gateways can beneficially be combined into well-functioning hybrids, as exemplified in section 5.1. However, the various gateways focus on improving different aspects. The use of style sheets and the sub-DTDs enhance flexibility in writing by allowing well-known concepts in the specifications as well as avoiding all concepts that are not in use. This is an attempt to deal with the fact that the production and application of documents varies between work practices. By introducing style sheets the writers can use their own editor, and provide structured documents using a method that is familiar to some extent. The organizational differences in the promotion and profile of the channels and the fact that they produce TV schedules for many recipients are taken into account. In many respects, the style sheets become boundary objects: by allowing differences, they make it possible to keep the boundary aspects of the documents. The interfaces take paper's important properties and make the resources they provided explicit for the employees at the Media Department. The resources that are provided by the documents' peripheral properties are given serious consideration in the design process. These gateways are only examples of how to change the inscription of the existing system.

The three gateways are evaluated and discussed according to Braa's (1995) framework and the theoretical concepts presented in 2.1 to investigate their potential in approaching the challenges stated in the previous section. The style sheets are presented in section 5.2, where experiences from previous research on the use of style sheets (Sørgaard and Sandahl, 1997; Braa and Sandahl, 1998) are essential to the discussion. Sections 5.2 and 5.3 present and discuss the use of sub-DTDs and interfaces respectively.

5.1 Prototype

The SGML system designed for the news agency did not work as intended. The TV channels had problems in delivering SGML-encoded TV information. The Media Department also experienced problems in their division of labor as well as in the awareness and coordination of work. They had to browse 350 files to see whether all the files had been received; they had no control over who was working with what, and they had lost a source of information on the progress of the work. The stable work chain of producing the TV schedules was broken after the introduction of the SGML system. Reasons for this are discussed in section 4. An experiment introducing gateways was conducted to improve the local flexibility and at the same time keep the standardization of documents in order to provide new functionality and features.

To support the TV channels in their production of the SGML-encoded TV information, sub-DTDs and style sheets were regarded as appropriate gateways. The style sheets reflected a new, less structured DTD made for each channel. The original DTD had 63 elements. For a particular channel, the DTD as well as the style sheet consisted of 16 styles, reflecting the elements in use. The 18 attributes defined in the DTD became 6 in the sub-DTD and in the style sheet.

To satisfy the channels' profiling needs, for instance, by making a film title bold or formatting actors' names in italics within a paragraph of plain text, macros were

applied. For more detailed descriptions of the implementation of style sheets, see (Kristiansen, 1998).

The sub-DTDs were designed as a subset of the original "shared" DTD, by selecting the most important elements for a particular practice. There are scripts that translate from a simple DTD into the more structured DTD. The subset DTD can accommodate different word processors (and macros). The aim is to reduce complexity and increase flexibility by allowing the channels to develop their own profiles on the schedule. It also facilitates the production of schedules as part of various other work chains in the TV channels. In this way, properties connected to boundary objects are implemented in the gateway as style sheets that can be flexible for each channel and as sub-DTDs that reflect a subset of shared elements (see Figure 5.1).
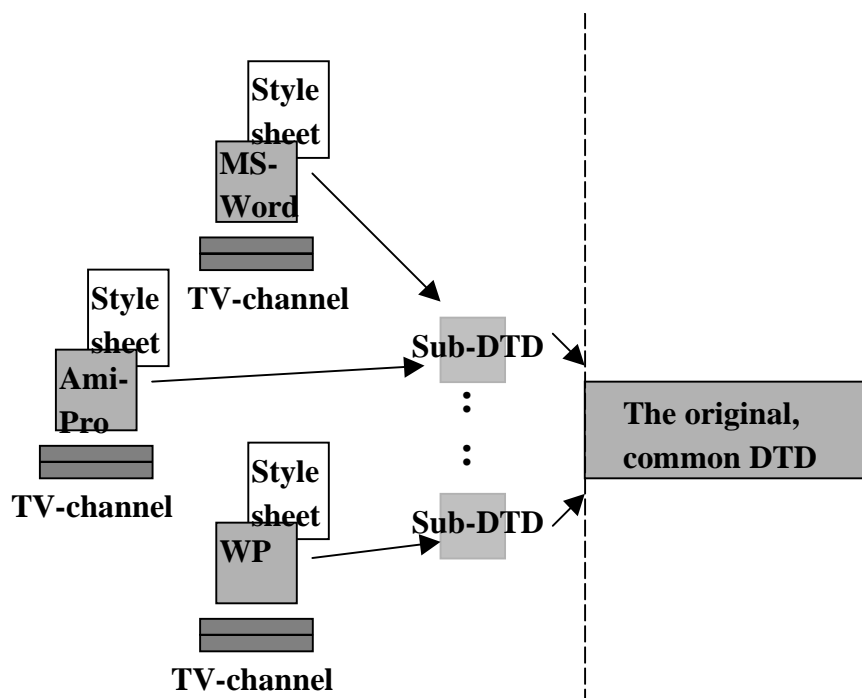
Figure 5.1: Style sheets and sub-DTDs as gateways

The use of style sheets has its challenges as well (Sørgaard and Sandahl 1997; Braa and Sandahl, 1998). This is further discussed in section 5.2. To improve the use of styles, training programs were established and conducted at the TV channels. The editor's user interface (Microsoft Word in this example) was improved to ease the access to the styles. The styles appear in a permanent window, instead of in a menu (for technical details, see Kristiansen, 1998).

Figure 5.2: The user interface of Microsoft Word with the styles easily available

A prototype was developed to make it possible to experiment with implementing an interface that represents some of the resources identified at the Media Department, and presented in section 4.1. In addition, the prototype supports the notion of different weeks. What denotes a week varies among the different actors involved (see section

3). This makes it necessary for the Media Department to operate with a week consisting of 11 days.

The prototype generates an overview showing which channel has sent information for which day, and the status of the production process. The user interfaces, which are represented as calendars, represent both the TV channels' weeks and the Media Department week, and 11 days in total are displayed in the interface. A user interface consists of 12-14 channels. Figure 5.3 shows the Norwegian channels. The status is represented as different color codes for each date. The faxes may have the following status: not yet arrived (black), arrived (blue), in process (yellow), ready for proofreading (light blue), and finished (green). Since the TV channels often send corrections in the same way as the original, the corrections (pink) are implemented as well. If a schedule is not recognized, it is represented as unknown (red). The colors represent the faxes' progression through the work process. They provide a way to represent the visibility of faxes in the various shelves.

When the user selects a channel (for instance, nrktv1) and double-clicks on a certain date the editor starts, and opens the corresponding file. The prototype takes into account the role of the documents as coordinating artifacts, represented by color. The color also helps the employees to be aware of the status of the work, and in this way it also supports the division of labor. If the date is yellow it indicates that someone is working with it, and that it should not be touched. If it is blue, it is ready to be worked on.
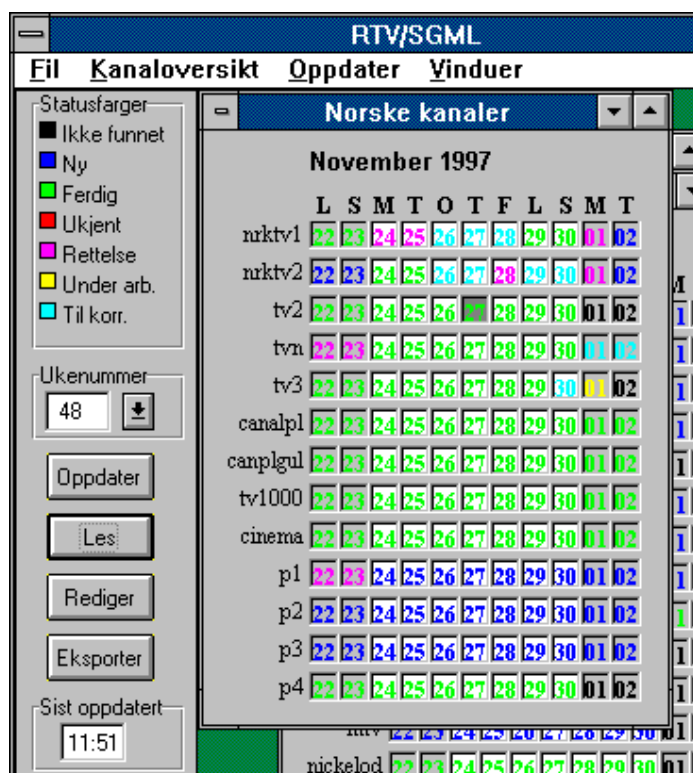
Figure 5.3: The prototype's user interface

The style sheets have been implemented for a couple of TV channels, and the Media Department is using the prototype to keep track of its work. This hybrid of various gateways functions as desired, and the users apply them as intended.

The next subsections evaluate and discuss the respective gateways.

5.2 Use of style sheets

An approach that would give writers more flexibility would be to offer them other tools, free them from doing the coding, and instead absorb the cost of conversion into SGML. An obvious approach would be to use a WYSIWYG word processor and appropriate templates or style sheets (Herwijnen, 1990). WYSIWYG word processors such as Microsoft Word, Word Perfect or AmiPro are familiar tools for the writers we studied. The word processors are embedded into all their work situations and applied for writing all kind of documents. These kinds of text processors do have a one-sided focus on structure, but offer a variety of functionality that is intended to support the writing process.

Use of style sheets, which is a guided approach to standardization (Braa and Sandahl, 1998) can improve the writers' situation in the production of TV schedules and still get the schedules into SGML in order to improve functionality and create new services. Use of paragraph styles has potential benefits for conversions to SGML (Sørgaard and Sandahl, 1997). They help in capturing text elements such as headings, footnotes, and lists, as well as more content-oriented information. Styles such as 'episode', 'actor', and 'date' can be defined as well. This mechanism is a foundation for conversions to SGML, and, for example, for improving the search and retrieval functionality within and across documents and making new services available.

Almost all writers at the TV channels applied Microsoft Word in their production of the TV schedules. However, they were unfamiliar with the use of styles (Kristiansen, 1998). With style sheets the writers can write the text more or less as before. In addition to writing the text, they have to style it by choosing the right style from a menu. Using style sheets, the TV channels can still copy and paste between schedules, and they can get the layout they want and even improve it by being more consistent (for instance, all headings look alike and the text consist of only one font type). There is also room for some individuality among the different TV channels. For instance, the channels can use style names that reflect their own use of concepts. One channel may use styles that are named 'actor' and 'day', while another channel uses the concepts of 'participant' and' date'. However, the logical meaning is the same. The numbers of styles may vary as well. In these cases, the various 'local' perspectives and perceptions are taken into account, which improves the boundary aspects of documents in use.

With styles there is no way to indicate the relationships between the elements or to verify that they follow a certain pattern, which from a technical perspective is problematic. It is possible to define style sheets in a way that indicates which style should come next. For instance, after a 'title' the next style is 'actor'. However, the writers can easily change this by selecting another style. To benefit from the use of styles to create structured documents, the styles have to be applied in a correct and consistent way without any support from the authoring tool.

The main problem in using style sheets is, as we observed in our analysis of press releases, that style sheets are not applied even if they exist (Sørgaard and Sandahl, 1997). Instead, the format of a paragraph is manually modified, for example by

selecting the font size and "bold" from the toolbars to achieve the same effect. Other problems, such as incorrect or incidental use of paragraph styles, are observed as well. These as well as other problems are further discussed in (Sørgaard and Sandahl, 1997)

We absolutely do not blame the writers for the lack of style use. The word processors have inscriptions that do not necessarily encourage the writers to apply style sheets. Until a few years ago the only use of word processors was to produce printed documents (Warnock, 1992). The functionality is designed to produce attractive documents on paper. From this perspective, it does not matter whether the text element is formatted by using styles or by selecting the font size, alignment and bold face manually from the toolbar. Nowadays, however, many documents are published on paper as well as electronically. With these trends, the use of word processors has shifted from producing printed documents to producing documents that can be reproduced or communicated in several different ways. The word processors' focus on paper presentations does not necessarily keep up with these changes in use. For instance, the press releases were all written according to an agreed layout, although they did not use a document template. They were manually formatted. In the agreed layout, several different "fields" of information, for example the date and organizational unit, appear on the same line. Since paragraph styles are associated with complete paragraphs, they always result in a line break. This makes it difficult to combine the agreed layouts with meaningful paragraph styles, and to identify different text elements for purposes such as improving search facilities. See (Sørgaard and Sandahl, 1997) for more details.

There are limitations on the structure of styled documents, because of the style sheet's restriction in defining styles within styles. If a new style is selected within a styled paragraph, only the layout changes, not the 'meaning' of the text element itself. For this reason, complex and large hierarchical structures cannot be styled. This implies that the DTD has to be simpler (or sub-DTDs have to be developed) in order to apply word processors and style sheets as authoring tools, which in itself improves the flexibility in authoring. The emphasis on detailed structure decreases. For instance, the course catalog DTD and the TV schedule's DTD are much more rigid and rich than the DTD for the administrative handbooks. The administrative handbooks were studied as a guided approach to standardization (see Braa and Sandahl, 1998).

From a technical perspective, the use of style sheets increases the technical challenges. When the styled documents are to be converted to SGML, they have to follow a predefined structure in one way or another. The conversion scripts capture the text elements. Chaos may occur if the structure in the DTD is different from the structure in the styled document. The conversion script will mix the text elements to fit into the structure defined in the DTD, and that is not always appropriate.

As pointed out above, the use of word processors and style sheets is not necessarily an appropriate approach for producing SGML documents. However, they have some benefits compared with an SGML editor, and some style sheets were created to take into account some local needs in the production of TV schedules. It is an advantage that style sheets can be tailored to some degree according to the practical situations in which they are used. In addition, since word processors are familiar tools and embedded in work practice, there is reason to believe that even production of text according to a given structure is easier with style sheets than with a native SGML editor.

A word processor such as Microsoft Word is applied in the production of all kinds of documents, and in various chains of work. For this reason, the writers may not regard the structuring as imposing a heavy additional workload. They have to do some work to apply the style sheets correctly, but they may get some benefits in improved layout. Also, the same document can be sent to all the recipients. The use of word processor does not separate the production processes, and the word processor provides functionality such as copy and paste, which is an important facility for those who produce documents.

From a technical point of view, it is a disadvantage that no mechanisms verify that the writers follow the defined patterns. If they do not, it is challenging to identify the text elements and their structure, which is necessary to convert to an SGML DTD. The technical challenges increase with an approach like this. If the style sheets are appropriately applied, they support some structuring and identifying of text elements. In a way, the word processor itself becomes a boundary object between the writers and the technicians' requirements.

To support the writers in using styles correctly, the accessibility to the styles was improved. The styles became visible in the Word interface, making them as readily available as the toolbars. A course was offered as well, to teach the writers about the style mechanisms and to tell them about the benefits of using them.

Each style sheet represents a kind of document model, which means that within the TV channels and news agency several models of the same document may exist. These various models may provide some flexibility in the production of TV schedules. By avoiding a one-sided focus on structure and standardization, and taking into account that there are different ways of describing the same document, the flexibility in writing is improved. In this way, the use of word processors and style sheets support some of the boundary aspects needed.

To convert to an SGML DTD, the structure in the DTD cannot be too complex, since the mechanism of styles within styles is not supported. The introduction of sub-DTDs is helpful in this respect.

5.3 Use of sub-DTDs

A way to improve the local flexibility is to make more flexible and/or smaller DTDs that are tailored to suit each of the TV channels, departments or faculties. To make the DTD more flexible, strong expressions like 'the information about *date* has to come before the *title*, which must be followed by *actor*' can be replaced by expressions like 'the information about *date*, *title* and *actor* must be present, but the sequence is optional'. This is also an approach to decrease the focus on structure, and to make models that are easier for writers to relate to, depending on their context of use.

At the news agency and the university, none of the writers use all the elements defined in the DTDs. All of them use only a subset. DTDs tailored for each TV channel, department or faculty can be developed. In this case, all the elements that are not in use are stripped, and not visible to the writers at all. In addition, concepts such as 'actor' can be used in a sub-DTD even it is converted into 'participant' in the "shared" DTD. The sub-DTDs are compatible through a "shared", larger and richer DTD that the functionality of the system is based upon.

However, these solutions have to be weighed against technical requirements and the effort of maintenance. DTDs are subject to change, and it is obvious that changing one DTD is simpler than changing 10 or 20. For a programmer, it is easier to develop

new functionality, search and retrieval routines as well as various presentations on paper and the Web, if the DTD is strict and represents standardization across the work practices. With a highly flexible DTD or multiple DTDs, the programming is far more complex, because of all the alternatives that have to be taken into account. From a user perspective, the DTD should be as flexible as possible. On the other hand, from a technical perspective, the DTD should be as strict as possible.

Sub-DTDs can be used together with style sheets, but also with the use of native SGML editors. Sub-DTDs make advanced technical solutions possible, since the technical facilities are based on the "shared" DTD. The writers still have to follow a predefined structure, but, even if an SGML-editor is applied, the concepts are known, and only elements in use are defined and visible in the editor.

With sub-DTDs, several models of documents are defined. However, in these situations, as in the SGML approach, the models do not represent the use of documents at work. But boundary aspects can be supported by tailoring the DTD to various work practices.

5.4 Making implicit resources explicit in user interfaces

When paper documents become electronic, the benefits of paper are lost. This study, as well as related studies, has illustrated that the properties of paper are essential for people to carry out their work. This do not mean that we have to keep using paper forever, but it does mean that we cannot focus only on the document's content in the design of document systems. It may be appropriate to make the essential conventions that have developed around documents in work practices explicit in the system in order to avoid breakdowns in work practices. One way to do this is to develop an interface of the system that represents not only the content of the document, but also how documents are applied in practical work. In this way the implicit resources that are essential to the division of labor, awareness and articulation of work become explicit. In other words, an interface as such will keep the resources that go beyond the document text.

5.5 Improvements to the SGML approach

This section has illustrated how the SGML approach can be improved by regarding the documents as artifacts with various properties that go beyond the property of structured information. Consideration of the documents' properties and how they are interrelated with humans and other artifacts provides insight into both explicit and more hidden uses of documents at work. An understanding of the sociotechnical network leads to an understanding of the prerequisites of the work taking place. When these prerequisites change, the work practices will change as well. This is a very good argument for not separating the content from the document, and for using only the content as the foundation for the document modeling. As stated, the understanding of work is the foundation for systems design.

The SGML approach focuses on structure and standardization, and this is represented in a document model. Section 4 illustrates that the switch from paper to electronic documents is challenging in itself, and the focus on structure and standardization impacts the production of application of documents in a way that users do not accept. By changing these inscriptions and adding new ones, one can improve the system in a way that fits in better with the work. The conventions are made explicit and represented in the system, and the one-sided focus on structure and standardization is untied by increasing the local flexibility. The gateways

36

presented are designed after taking the documents' central, peripheral, local and shared properties into account, which improves the local flexibility and still fulfils the technical need for structure and standardization.

Taking the central, peripheral, local and shared properties into account is a result of not separating the content from the documents itself, based on the assumption that the document properties are essential to the way the work is processed. It is difficult to represent 'everything' in a DTD itself, but we have to make sure that the system takes these aspects into account.

By changing the properties of the technology, we are changing the work by changing the technological prerequisites. This is possible only if we know the properties beforehand, which we can accomplish only if we study the sociotechnical network, or how the properties are critical to the work.

To study the documents' central, peripheral, shared and local properties we gained an understanding of the work, which is necessary in designing document systems. These concepts of documents at work can help SGML practitioners and other designers to reframe their own perspectives of documents in use and enhance their understanding of the design of document systems.

6. Concluding remarks

The aim of this thesis has been to inform the design of document systems by regarding the transformation from paper to digital documents in organizations. The SGML approach is emphasized. The SGML approach takes the documents' structure and content as the starting point in design. This approach is challenged and improved by empirical studies of documents in use and theoretical considerations of artifacts at work.

This research has illustrated that the transformation from paper to digital documents is challenging because paper documents and related artifacts are highly integrated into work practices, and they have intentions or roles that go beyond the artifacts' canonical meaning. This is in contrast to the SGML perspective on documents as a "collection of structured information". The SGML approach emphasizes the documents' content and structure, which is an obstacle to discovering the variety of documents' application at work. In addition, SGML systems have inscriptions that emphasize structure and standardization, which overlook the heterogeneity of work practices and presuppose that documents have common 'meanings' and applications across practices, which they do not necessarily have. As a consequence, the various work practices are regarded as one. The work performed at the TV channels is seen as an input to the system, and not as independent work performed with independent aims.

To summarize, in the transformation from paper to structured digital documents, the SGML approach has to be improved by a better understanding of documents at work, decreasing the level of structure and increasing the local flexibility.

The study illustrates that an investigation of the actor-networks of which documents are parts promotes an understanding of the more hidden aspects of work. By regarding documents' central, peripheral, local and shared properties, it is possible to achieve an understanding of how documents are embedded in work, including how important documents and related artifacts are to aspects such as awareness, articulation and coordination of work.

The concept of inscription (properties), which provides an insight into how humans and artifacts are intertwined, is helpful in order to understand documents at work as well as specifying how new technology impacts work. In design, it is necessary to be specific about the technology by being aware of various aspects and features of it, and the concept of inscription provides for this. It does not mean that the new technology should be exactly like the previous one. A digital document is an artifact in its own right, distinct from paper documents. Allowing users to draw on their experience with paper documents when using a digital documents could be a good way to get them started, but only that. Digital documents lack some essential features that paper documents have. On the other hand digital documents create numerous new possibilities, such as search, cross-referencing or linking. However, to understand the documents in use, we have to take the paper documents as a point of departure. An awareness of how paper documents are integrated into work can make it easier to predict changes by changing the prerequisites for such use in a way that does not make breakdowns inevitable.

In the design of the gateways, the idea has been to keep the technical possibilities that SGML provides, and at the same time to take into account our knowledge about the paperwork. From a technical point of view, SGML provides new services and functionality. However, it requires encoding of the documents in SGML in the first place. Without any SGML documents, there is no foundation for new services. The technological conditions have to be arranged to accommodate this. In addition, it is necessary to provide some benefits from the start of the implementation, and provide for these benefits to increase in line with the implementation of the system. The tension of "who does the work and who gets the benefits", as stated by Grudin (1989; 1994), is present in situations where various heterogeneous actors become parts of the same actor-network.

I find the concepts of heterogeneous actor-network, inscription, borderline issues and boundary objects fruitful in order to understand documents at work as a foundation for design of document systems. I believe that these concepts can reframe the SGML practitioners' as well as other designers' perspective of documents at work, and in this respect promote progress in the understanding of design of document systems.

References

Akrich, M., (1992). The de-scription of technical objects. In Bijker, W. E. and Law, J. (eds.) *Shaping technology/building society*. MIT Press.

Alschuler, L. (1995). *ABCD...SGML. A User's Guide to Structured Information*. International Thomson Computer Press

Andersen, N. E., Kensing, F., Lassen, M., Lunding, J., Mathiassen, L., Munk-Madsen, A. and Sørgaard, P. (1986). Professionel systemudvikling: erfaringer, muligheder og handling (in Danish). Teknisk Forlag.

André, J., Furuta, R. and Quint, V. (1989). *Structured Documents*. Cambridge University Press.

Berg, M. (1996). Practices of Reading and Writing. The Constitutive Role of the Patient Record in Medical Work. *Sociology of Health and Illness* 18: 499-524

Boland, R. J. and Greenberg, R. H. (1992). Method and metaphor in organizational analysis. *Accounting, Management and Information Technology*, 2(2), pp. 117-141.

Bolter, J. D. (1991) *Writing Space. The Computer, Hypertext, and the History of Writing.* Lawrence Erlbaum Associates Publishers.

Bowker, G. C., Timmermans, S., Star, S.L. (1995) Infrastructure and Organizational Transformation: Classifying Nurses' Work. In Olikowsky, W. J., Walsham, G., Jones, M. R. and DeGrose, J. I (eds) *Information Technology and Changes in Organizational Work. Proceedings of the IFIP WG8.2 working conference on information technology and changes in organizational work, December 1995.* Chapman & Hall.

Braa, K. (1995) Beyond Formal Quality of Information Systems Design, Dr. Scient Thesis, June 1995, Department of Informatics, University of Oslo.

Braa, K. and Sandahl, T. I. (1998) Approaches to standardization of documents. In Wakayama, T., Kannapan, S., Khoong, C. M., Navathe, S., and Yates, J. (eds.) *Information and process integration in enterprises: rethinking documents,* Kluwer Academic Publishers, Cambridge, Massachusetts, USA, pp. 125-143.

Braa, K and, Sandahl, T. I. (1998b). From Paperwork to Network. In *Proceedings from the Third International Conference on the Design of Cooperative Systems (Coop '98)*, Cannes, France, 26-29 May 1998.

Braa, K and Sandahl, T. I. (1998c). Documents in Infrastructure. Challenges for Design. In *Proceedings from the European Conference of Information Systems (ECIS),* Aix-en-Provence, France, 4-6 June 1998.

Brown, J. S. and Duguid, P. (1995). The Social Life of Documents. At the Web site http://www.parc.xerox.com/ops/members/brown/papers/sociallife.html (This paper was first published in Release 1.0, October 1995 in the first issue of First Monday 1(1))

Brown, John S and Duguid, P (1994). Borderline Issues: Social and Material Aspects of Design. In *Human-Computer Interaction*, 1994, Volume 9, pp. 3-36.

Bryan, M (1989). *SGML. An Author's Guide to the Standard Generalized Markup Language.* Addison-Wesley Publishing Company.

Callon, M (1986). Some Elements of Sociology of Translation: Domestication of the Scallops and the Fishermen. In J. Law (Ed.), *Power, Action and Belief: A New Sociology of Knowledge* London: Routledge & Kegan Paul.

Carstensen, P.H., Sørensen, C (1997). From the Social to the Systematic - Mechanisms Supporting Coordination in Design. In Computer Supported Cooperative Work - An International Journal, pp. 1-26.

Dahlbom, B. and Mathiassen, L. (1993). *Computers in Context.* NCC Blackwell.

Denning, P. J. and Rous, B. (1995). The ACM electronic publishing plan. *Communications of the ACM*, 38(4): pp. 97--103, April 1995.

Dourish, P. and Bellotti, V. (1992). Awareness and coordination in shared workspaces. In *Proceedings of the Conference on Computer-Supported Cooperative Work*. October 31 to November 4, 1992. ACM Press.

Easterby-Smith, M., Thorpe, R. and Lowe, A. (1991). *Management Research: An introduction. Sage Publications* Ltd.

Ehn, P. (1993). Scandinavian design: on participation and skill. In Schuler, D. and Namioka, A. (eds). *Participatory design : principles and practices*, pp. 41--78. Lawrence Erlbaum Ltd.

Ehn, P. (1988). *Work-Oriented Design of Computer Artifacts*, Falköping: Arbetslivscentrum.

Ehn, P., Meggerle, T., Steen, O. and Svedmar, M. (1995). What Kind of Car is this Sales Support System? On Styles, artifacts and quality-in-use. In Kyng, M. and Mathiassen, L. (eds). *Proceedings of the Third Decennial Conference on Computers in Context: Joining Forces in Design*, August 14-18, Aarhus University, Denmark.

Gasser, L. (1986). The integration of computing and routine work. *Acm Transactions on office information systems*. Vol 3, No. 4., pp 203-225.

Goldfarb, C. F. (1990). *The SGML handbook*. Clarendon Press, 1990.

Goldfarb, C. F., Pepper, S. and Ensign, C. (1998) *SGML Buyer's Guide : A Unique Guide to Determining Your Requirements and Choosing the Right SGML and XML Products and Services.* Goldfarb Series on Open Information Management, Prentice Hall.

Goldfarb, C. F. and Prescot, P. (1998). *The XML Handbook.* Charles F. Goldfarb Series on Open Information Management, Prentice Hall.

Greenbaum, J. and Kyng, M. (1991) *Design at Work: Cooperative Design of Computer Systems*. Hillsday: Lawrence Erlbaum Associates.

Grudin, J. (1994) Groupware and Social Dynamics: Eight Challenges for Developers. In *Communications of the ACM*, Vol. 37, No 1.

Grudin, J. (1989). Why groupware applications fail: Problems in design and evaluation (1989). *Office: Technology and People*, 4(3): pp. 245--264, June 1989.

Gustavsen, B. and Engelstad, P. H. (1990). Creating systems by dialogue – an emergent trend in organization development projects, In Bjerknes, G. and dahlbom, B. (eds). *Organizational competence and system development*. University of Lund.

Haas, C. (1990). Composing in Technological Contexts. In *Written Communication*. Vol 7 No 4. Sage Publication.

Hanseth, O. (1991). Intergarting information systems: The importance of context. In Stamper, R. K., Kerola, P., Lee, R., and Lyytinen K. J. (eds). *Collaborative Work, Social Communication and Information Systems*. North-Holland, 1991, pp. 133-156.

Hanseth, O. and Monteiro, E. (1997). Inscribing behavior in information infrastructure standards. Accounting, Management & Information Technology. Vol. 7, No. 4, pp. 183-211.

Hanseth, O., Monteiro, E. and Hatling, M. (1996). Developing information infrastructure: the tension between standardization and flexibility. In *Science, Technology & Human Values*, Vol 21, No. 4, pp. 407-426, Sage Periodicals Press (1996).

Harper, R. and Sellen, A. (1995). Paper-Supported Collaborative Work. *Technical Report EPC-1995-109*. Rank Xerox Ltd.

Heath, C. and Luff, P. (1992). Collaboration and Control. Crisis Management and Multimedia Technology in London Underground Line Control Rooms. *Computer Supported Cooperative Work (CSCW)* 1, pp. 69-94.

Hendry, D. G. (1995). Breakdowns in writing intentions when simultaneously deploying SGML-marked text in hard and electronic copy.. In *Behaviour & Information Technology*, Vol. 14, No. 2

Hersch, R. D. André, J. and Brown, H. (Eds.) (1998): Electronic Publishing, Artistic Imaging, and Digital Typography. Proceedings from the 7th International

Conference on Electronic Publishing, EP'98. Held Jointly with the 4th International Conference on Raster Imaging and Digital Typography, RIDT'98, St. Malo, France, March/April 1998.

Herwijnen, Eric van (1990). *Practical SGML*. Kluwer, Dordrecht, 1990.

Hughes, J. and King, B (1993). Paperwork. In Benford, S. and Mariani, J. (eds*). Requirements and Metaphors of Shared Interaction. Part II: Shared Artifacts*. COMIC Deliverable 4.1, pp. 153-325. Lancaster, England, 1993

Hughes, J, Randall, D., and Shapiro, D. (1992). Faltering from ethnography to design. In *Proceedings of Computer-Supported Cooperative Work*, 1992, Toronto, Canada

Klein, H. and Lyytinen, K. J. (1992). Towards a new understanding of datamodelling. In Floyd, C. et. al (eds.) *Conference on software development and reality construction*, pp. 203-220. Springer Verlag, 1992.

Kling, R. and Elliot, M. (1994). Digital library Design for usability. In *Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries*. June 19-21, 1994 - College Station, Texas, USA

Kristiansen, B. T. (1998). *Standardization and support for local flexibility in an inter-organizational setting. A case study and an experiment*. (In Norwegian). Master Thesis. Department of Informatics, University of Oslo.

Kyng, M (1995) Making Representation Work. In *Communications of the ACM*, September 1995, Vol. 38, No. 9

Latour, B (1996a). *Aramis or the Love of Technology*. Cambridge, Massachusetts: Harvard University Press.

Latour, B (1996b). Social theory and the study of the computerized work sites. In Orlikowski, W. J and Walsham, G and Jones, M. R and DeGross, J. I (eds). *Information Technology and Changes in Organizational Work*, Chapman & Hall.

Latour, B (1987). *Science in action*. Open University press, 1997

Latour (1986) Visualization and Cognition: Thinking with eyes and hands. In *Knowledge and Society: Studies in the Sociology of Culture Past and Present*. Vol. 6.

Law, J. (1992). Notes on the theory of the Actor-Network: Ordering, strategy, and heterogeneity. In *Systems Practice*, Vol. 5. No. 4, 1992. pp. 379-393

Law, J. (1986) On the methods of long-distance control: vessels, navigation and the Portuguese route to India. In Law, J. (ed) *Power, action and belief*. Routledge & Kegan Paul Inc, 1986.

Levy, D (1994). Fixed or Fluid? Document Stability and New Media. In *European Conference on Hypermedia Technology 1994*. Conference Proceedings. New York: ACM Press, 1994

Levy, D. (1988). Topics in Document research. In *Proceedings from the ACM Conference on Document processing systems*, pp. 187-193. Santa Fe, New Mexico, December 5-9, 1988

Levy, D. and Marshall, C. C. (1995). Going Digital: A Look at Assumptions Underlying Digital Libraries. *Communication of the ACM,* Vol. 38. No. 4. April 1995

Light, R. (1997). *Presenting XML*. Sams net.

Lundberg, N. and Sandahl, T. I (1998). Artifacts in Work Practice. Submitted.

Luff, P, Heath, C. and Greatbatch, D. (1992). Tasks-in-interaction: paper and screen based documentation in collaboration activity. In *Proceedings of Computer-Supported Cooperative Work, 1992*, Toronto, Canada

Maler, E and Andaloussi, J. (1996*). Developing SGML DTDs. From text to model to markup.* Prentice Hall, PTR.

Mathiassen, L (1981). *Systemutvikling og systemutviklingsmetode.* Doktoravhandling (In Danish). Datologisk afdeling, Aarhus Universitet.

McGrath, J. (1982). Dilemmatics: The study of research choices and dilemmas. In McGrath, J., Martin, J., and Kulka, R. (eds) *Judgements Calls in Research*, pp. 69 – 102. Sage, Beverly Hills.

Monteiro, E. and Hanseth, O. (1995). Social shaping of information infrastructure: on being specific about technology. In Orlikowski, W.J, Walsham, G and Jones, M. R. and DeGross, J. I (eds). *Information technology and changes in organisational work.* Chapman & Hall, 1995

Ong, W. J. (1982). Orality & Literacy. The Technologizing of the Word, T. J. Press Ltd.

Orlikowski, W. J. (1992). The duality of technology. Rethinking the concept of technology in organizations. *Organization Science.* Vol 3, No. 3, pp 398-427.

Orlikowski, W. J. (1991). Integrated Information Environment or matrix of control? The contradictory implications of information technology. *Accounting, Management and Information Technology* 1(1), pp. 9 - 42.

Orlikowski, W. J. and Robey, D. (1991). Information technology and the structuring of organizations. *Information systems research.* 2(2), pp. 143-169.

Reinhardt, A. (1994) Managing the New Document, in *Byte*, Vol 19, No. 8, August 1994.

Sachs, P. (1995). Transforming work: Collaboration, Learning, and Design. In *The Communication of The ACM*, Vol.38, No. 9, September 1995.

Sandahl, T. I. and Jenssen, A. (1997). The First Step in Designing an SGML-Based Infrastructure for Document Handling in *Scandinavian Journal of Information Systems,* Vol 9, No. 2, pp. 25-44.

Schmidt, K and Bannon, L (1992). Taking the CSCW Seriously: Supporting Articulation Work. In *Computer Supported Cooperative Work (CSCW). An International Journal.* Vol. 1, pp. 7-40.

Schmidt, K and Simone, C (1996). Coordination Mechanisms: Towards a Conceptual Foundation of CSCW Systems Design. *In Computer Supported Cooperative Work: The Journal of Collaborative Computing* 5: pp. 155-200. Kluwer Academic Publishers.

Smith, Joan M (1992) SGML and Related Standards. Document description and processing language. Ellis Horwood Limited.

Star, S. L. (1989) The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving. In Gasser, L. and Huhns, M. (eds) *Distributed Artificial Intelligence*, Vol. 2. Pitman, London.

Star S., L. and Griesemer J., R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *In Social Studies of Science*, August, 1989.

Stolterman, E. (1991). The hidden rationale of design work–a study in the methodology and practice of system development (in Swedish). Department of Informatics, University of Umeå.

Suchman, L. (1995). Making work visible. In The Communication of The ACM, Vol.38, No. 9, September 1995.

Suchman, L. and Trigg, R. (1991). Understanding Practice: Video as a Medium for Reflection and Design. In Greenbaum, J and Kyng, M (eds). *Design at Work - Cooperative Design of Computer Systems*, Lawrence Erlbaum, 1991.

Sørgaard, P. and Sandahl, T. I. (1997). Problems with Styles in Word Processing: A Weak Foundation for Electronic Publishing with SGML. In *Proceedings of the 30th HICSS*, January 7-10, 1997, Wailea, Hawaii.

Travis, B. and Waldt, D. (1995). *The SGML Implementation Guide. A Blueprint for SGML Migration*. Springer.

Vidgen, R & Braa, K (1999). Interpretation, intervention, and reduction in the organizational laboratory: a framework for in-context information system research. In *Accounting, Management & I*T (forthcoming).

Vidgen, R & Braa, K (1997). Balancing interpretation and intervention in information system research: the "action case" approach, in *Proceedings of IFIP 8.6: Information Systems and Qualitative Research*, Philadelphia, 31 May - 3 June, 1997.

Warnock, John E. (1992). The new age of documents. *BYTE*, 17(6):257--260, June 1992.

Yates, J (1989). *Control through Communication. The Rise of System in American Management.* The Johns Hopkins University Press, Baltimore and London.

Yates, J. and Orlikowski, W. (1992) Genres of organizational communication: A structurational approach to studying communication and media. In *Academy of Management Review*, Vol 17, No. 2, pp. 299-326.

Yin, R. K (1994). *Case Study Research*. Sage Publications Ltd.

The First Steps in Designing an SGML-Based Infrastructure for Document Handling

Tone Irene Sandahl and Astrid E. Jenssen

Abstract

In this paper we present and discuss efforts to design an SGML-based infrastructure in an institutionalized work setting. The initial goal was to improve the "functionality" of documents in order to access, update, search, use and reuse, retrieve, present, exchange and distribute them independently of time and place and without loss of information. From a technical point of view, standardization is a prerequisite for success. In the pilot project presented in this paper, Standard Generalized Markup Language (SGML, ISO 8879) is applied. However, the study has shown that writers experience standardization as a restricting factor in their work, and local flexibility is essential. In relation to SGML-based infrastructures, this has consequences for DTD design and for the selection of authoring tools. The study has also indicated that documents are artifacts that are integrated into practice, and this has to be taken into account in design of document technology.

Keywords: *SGML-based infrastructure, DTD-design, documents in work*

1. Introduction

Since early on in the computer age, there has been a need to classify, compute, combine and recombine, count, sort, and manipulate information. From the late 60s, this has been done by cutting the information into little pieces (data) and putting it into databases. Databases are fine for discrete, predictable pieces of information, but they do not work very well for information like stanzas, scientific descriptions, or maintenance procedures (Alschuler, 1995). A database system is basically a computerized *record*-keeping system (Date, 86). As stated by Reinhard (1994), at least 80% of electronic information in organizations is in the form of documents, as opposed to database records. Traditionally, documents have been static, represented as files on disks. Until PCs were networked, these files usually belonged to only one user and passed from one person to another in printed form. It is a challenge to make these documents, or files, more dynamic, in order to be able to access, search, use and reuse, retrieve, present, exchange and distribute them without loss of information (Reinhard, 1994).

This paper presents and discusses the efforts involved in the first steps of designing an SGML-based infrastructure, and considers the complexity of this. The concept of an infrastructure denotes all the documents and practice required to support people adequately in carrying out their work (Jewett & Kling, 1991; Star & Ruhleder, 1994). The concept of an SGML-based infrastructure indicates that the technical solutions are based on SGML. From a technical perspective, application of SGML is an

appropriate approach for three main reasons: i) Standardization is necessary for infrastructures to exist; it is the technical backbone (Hanseth et al., 1996). ii) SGML makes it possible to use structured concepts in text in general, which lessens the difference between documents and databases, as well as improving search and retrieval in texts. iii) SGML is independent of software, systems and presentations, and it supports "open systems".

At the end of 1992, a project was initiated at the University Center for Information Technology Services (USIT) to determine what type of electronic infrastructure could deal successfully with electronic documents and other forms of information at the university. The infrastructure had to address the whole life cycle of a document, i.e. production, updating, filing, administration, distribution, presentation and reuse. The use of SGML in the project emerged as a possible key tool for describing the documents and their content.

In 1993, to gain experience in the practical use of SGML for some parts of the information produced at the university, USIT established a pilot project that involved developing an open and flexible solution for the production, exchange and distribution of the university's course catalog. The pilot project was initiated to develop a technical infrastructure and administrative routines for dealing with the catalog, which contains dynamic information and "new functionality" for the students and the staff. A detailed presentation of the development process itself can be seen in Jenssen & Sandahl (1996).

The paper is structured as follows: In Section 2, a short introduction to SGML is presented, followed by the research approach in Section 3. Section 4 presents the practice of catalog work, while Section 5 briefly presents the design and implementation. The evaluation of the pilot project is in Section 6, the discussion in Section 7, and finally the conclusion in Section 8.

2. Standard Generalized Markup Language (SGML)

There are at least two different ways to achieve information interchangeability between systems: standardization on applications, so that the applications can work on each other's information, or standardization of the information itself, so that it can be processed by any application. SGML supports the latter solution.

SGML is designed to enable text interchange, and it is intended for use in the publishing field (Smith, 1992). Since then, it has been increasingly adopted as the international standard for data and document interchange in open system environments, including the automotive, defense, commercial aerospace, pharmaceutical, electronics, and telecommunications industries (Van Herwijnen, 1993).

A basic design goal of SGML is to ensure that information encoded according to its provisions should be portable from one hardware and software environment to another without loss of information (Goldfarb, 1992). The basic idea is very simple: the text is described in terms of its structural components (descriptive markup) rather than its presentation (procedural markup) in a single medium.

A descriptive markup system uses markup codes, which provide names to categorize parts of a document. Markup codes such as <course> identify a portion of a document and assert of it that "the following item is a course". All the text is coded as plain text. SGML thus enables the interchange of text across platforms, because there is no need for "translation" to met hardware requirements. The same document can readily

be processed by different types of software, each of which can apply different processing instructions to those parts of it which are considered relevant. In addition, different sorts of processing instructions can be associated with the same parts of the file. Since only the structure and/or content of a document are marked, a given viewer of that document can decide what the "look" and "use" will be. The markup of the document never changes - only the way it is interpreted.

SGML supports the notion of a document type, and hence a "Document Type Definition" (DTD). An SGML document always has an associated DTD that specifies the rules of the model of the document; for example a DTD for a course catalog might specify that the document type (catalog) must have information about one or more courses. Furthermore, each course must have a code, a title, an optional description (descrip), followed by zero or more combinations of day, time and place. Information about lecturer(s) is required, and so forth. The type of a document is formally defined by its parts (course, day, time, place, ...) and their structure in the DTD. Here is an example of the syntax of part of a DTD:

```
:
<!ELEMENT catalog - - (course+)>
<!ELEMENT course - - (code, title, descrip?, (day, time, place)*, lecturer+)>
<!ELEMENT teacher - - (fname, sname, email?, phone*, fax?)>
:
```

Figure 1. Part of a Catalog DTD.

The users of the standard design the DTD; hence, the DTD is not predefined in the standard itself.

3. Research Approach

The research approach was pragmatic, as described below.

First, we decided to use the university study catalog as a pilot document. The goal of the pilot project was stated before the project start, namely to make the catalog electronic and structured in order to improve its production, exchange, distribution and application. SGML is intended to support this.

We entered the problem environment, analyzed the work situation, and identified roles related to the production, exchange, distribution, and application of the catalog. We took part in the design process and evaluated it continuously. After finishing the pilot project, we reflected on and recorded what we had experienced during the process, which is documented in this paper.

3.1 The Catalog as a Pilot Project

The catalog was chosen as a pilot project for several reasons. The university catalog is well known to the students and staff at the university, and many people use it when carrying out their jobs. The production process of the catalog is attractive in relation to other crucial documents; the information is supplied from different units (faculties, departments and central administration) at the university and a number of writers are responsible for updating different parts of the catalog.

The electronic version of the catalog facilitates advanced search, reuse of information and presentation on different media (printed, electronically on screen), as well as possibilities for links to other relevant information such as university/faculty regulations, syllabi, curricula, and so on. From a more technical perspective, the information was seen as structured, so it was well suited for building SGML applications.

The main goals of the pilot project were to produce a better (and in the long run, cheaper) catalog, to make it easier to update and maintain, and to gain practical experience introducing SGML at the university. To produce a better catalog involved developing a more readable catalog, making it electronically accessible and adding some new functionality.

3.2 The Idea of a Structured Catalog

As we saw it, a critical aim was to get a structured and well-defined catalog in order to develop new services based on the information in the catalog. Later, we wanted to develop scripts that could manipulate the catalog and its information elements depending on existing situations and products. Services like room allocation, in which a writer allocating a room could get suggestions for a place and time based on existing information in the catalog, as well as advanced search in the catalog information and customize publishing, were meant to be implemented.

A basic principle was that the writers should have to enter information only once. As a result, updating and maintenance of information across presentations and products should be easier, and redundancy could be avoided. If the information was presented in a different setting, the computer would do this automatically. The writers would still have to collect information from lecturers and others, and enter it in the catalog.

3.3 The People Involved

The pilot project was organized as a project group responsible to a steering committee. Both groups consisted mainly of staff from USIT. The project group consisted of 3–5 IT people with 1–2 working full-time, the others working part-time on the pilot. The group was responsible for the system development process. The central administration was responsible for editing the catalog and some of the writing. Each semester, there were about 40 writers from the different university faculties and departments, and several of them worked with the catalog for more than one semester. The writers were located in all the administrations and secretariats at the faculties, the central administration and the study department. In the faculties and the study department, the secretaries or administrative consultants are the writers, where the advisers do the writing at the department level. However, the production of the catalog is only a small part of their daily work.

3.4 Evaluation Methods

During the project we interviewed 22 different people involved in the project, representing writers and management people at USIT and the central administration unit. Some were interviewed several times. We analyzed 393 email messages sent to a distribution list. We reviewed a number of questions and problems from telephone calls and direct mail. We analyzed minutes of 13 meetings with the writers and more than 20 internal meetings dealing with more technical problems. In addition, we analyzed 3 project reports.

4. The Practice of Catalog Work

The work with, and the use of, the catalog were manifested in the ongoing work procedures. The writers' work, the coordination, and the merging of the catalog had been done in the same way for many years. In this section we describe how the work was done before the pilot project.

The catalog is separated into sections for each faculty, describing courses offered by that faculty, and different sections for other kinds of courses such as distance education, information about student services, and collaborating institutions. The writers at the central administration maintain information about all sections except the information from the faculties. They have to coordinate with other units at the university to collect information to be presented in the catalog. The writers at the faculty level maintain information common to a faculty, and the writers in the subordinate departments maintain information mainly about courses offered by the department. In addition to writing the text for the catalog, the writers at the faculty level are responsible for making sure that all the text from each underlying department "fits together", e.g. that registration deadlines for classes are correct in relation to the dates set by the faculty, and that there is no overbooking of lecture rooms.

However, all the departments and units are responsible for their parts of the catalog, for collecting information and for distributing updated information to the students and staff. For instance, the writers at department level are in contact with different lecturers to collect information about which courses are to be offered each semester. They have to coordinate the allocation of lecture rooms, taking into account the lecturers' preferences for the day, time and place and, in cooperation with the writers at faculty level, to avoid overbooking. They do the updates and distribute the results to the different lecturers for proofreading. Then they update the information again as often as necessary. If students need to be informed about important changes after the catalog is sent to the printing office, e.g. changes in class time scheduling, the writers make an A4 document that they display on boards outside their office, students' lecture rooms, and workplaces.

Different departments and faculties presented the same information in the catalog (e.g. information about courses, time and the places for classes) somewhat differently. For instance, the mathematics presented course information in a tabular form, while the law faculty presented the same information as free text.

The writers used word processors such as MS Word or WordPerfect. They were told which font and font size to use in headings, paragraphs and so on, but no style sheets were available. They saved the documents on their own computers, and paper copies were exchanged or distributed during the production process by mail. Much of the updating work was done on paper versions.

When the departments and faculties were finished with their work, they all sent the files or documents on diskettes to the central administration unit. At the central unit, the final proofreading was done, and references and indexes were created manually. The administration unit sent the catalog on diskettes to the print office. Even before the catalog came back from the printing office, some of the information was outdated due to last-minute changes in course schedules, etc. Hence, there was a need to update information until just before the publication date, and a continuous need for updating of further changes. The writers or the central administration contacted the printing office directly to have them incorporate the changes in the final printed version.

We observed that some information elements in the catalog were also part of other important handbooks, brochures and catalogs at the university. There were three or four original versions of the information elements, and it was hard to avoid inconsistency. The catalog was published only on paper. It was about 450 pages long, and 50,000 copies were printed twice a year. These were available to students and administrative staff at the university as well as other educational institutions. It was fairly expensive. Electronic services and services such as customizing publishing and publishing on demand could save both paper and money.

5. The Design and Implementation

In the pilot project period, the following were emphasized in design: the DTD design, the editing environment, the printing environment, WWW presentations, an environment that supports the interdependence in work to some extent, and finally training programs and support. A summary of the process is presented in the table below. However, a more detailed presentation of the implementation can be seen in Jenssen & Sandahl (1996).

Work period	Work done	By whom
Catalog 1 93: March–June	Document analysis	USIT/writers
	Developing the first version of the DTD	USIT/writers
	Manually encoding the information for the autumn catalog	USIT
	Developing the printed version	USIT
	Developing scripts for conversion to HTML	USIT
	Setting up the catalogs, files and access in the Unix file system	USIT
	Organizing the work flow	USIT
	User training	USIT
	User support	USIT/writers
	Evaluation	
Catalog 2 93: July–December	Developing the second version of the DTD	USIT/writers
	Merging already encoded information into the new version of the DTD	USIT
	Encoding the information using templates in word processors	Writers
		USIT
	Conversion of new information to SGML	USIT
	Improving the printed version	USIT
	Printout possibilities for the writers through Unix	USIT/writers
	Organizing the Unix file system	USIT
	Organizing the work flow	USIT
	User training	USIT/writers
	User support	
	Evaluation	
Catalog 3 94: January–June	Improving the second version of the DTD	USIT/writers
	Improving the electronic version	USIT
	Using the SGML editor for updating the information	Writers
	Developing schemes to be used with the SGML editor	USIT
	Improving the conversion to HTML	USIT
	Organizing the Unix file system	USIT
	Organizing the work flow	USIT/writers
	User training	USIT
	User support	USIT
	Evaluation	USIT/writers

Catalog 4 94: July–December	Further improvements of the second version of the DTD	USIT/writers
	Printout possibilities through World Wide Web	USIT
	Using SGML editor for updating the information	Writers
	Improving the style sheet used by the SGML editor	USIT
	Organizing the Unix file system	USIT
	Organizing the work flow	USIT/writers
	User training	USIT
	User support	USIT
	Evaluation (interviewing)	USIT/Dep. of Informatics/writers
95: January->	The system in ordinary use	USIT/writers
	Evaluation	USIT/Dep. of Informatics/writers

Table 1. The work done at the different stages of the system development process

5.1 DTD Design

The first version of the DTD was developed through document analysis led by the system developers involving the writers, managers and people from the central administrative unit. The intention was to make the catalogs' structure rich enough to allow retrieval of information directly from databases, links to other information and functions for presenting different views of some parts of the information in the catalog. As a result of the way that the conversion routines were programmed for the printouts, some elements were also included in the DTD to ensure that the layout in the paper version was correct and attractive. For example, for information about courses, there were different elements to be used depending on the day information, such as <day> for one day, <dayint> for day interval and <dayoppr> for specifying a selection of days: <day>Monday</day>, <dayint>Monday-Thursday</dayint> and <dayoppr>Monday, Tuesday, Friday</dayoppr>. Later, the DTD and the conversion routines for making printouts were changed to use only one element for information about days, and still ensure an appropriate layout.

The focus on technical solutions and products such as printed and electronic versions led to a DTD that contained a large number of elements for the users to deal with in the writing process. A DTD that seems appropriate for the technological solution may not be appropriate for the writers. On the basis of practical experience, the DTD was reorganized and improved during the project. It became less detailed for the whole catalog, removing some elements and generalizing others, but was still rich and strict.

5.2 Editing

A goal specified for the project was to have the writers produce SGML documents according to the relevant DTD. Because of the importance of correct input, using a native SGML editor was seen as the appropriate solution for doing this. An SGML editor is context-sensitive. It knows the predefined structures defined in the DTD. It may incorporate a validating parser that makes it possible to avoid markup errors and guarantee that the document is structurally correct. The editor Author/Editor was applied, one of the reasons being that it was one of the few tools available for the PC, Macintosh and Unix platforms at that time. The editor is a tool for focusing on the document's content and structure. It has some functionality for adding different layouts according to the structure, but this is not adequate to fulfill WYSIWYG layout requirements according to the printed version of the catalog.

50

Use of the editor was integrated step-wise. For the first catalog, the information from the writers was manually encoded using SGML editors and other tools by a group of people at USIT. The second catalog was produced by the writers using word processors and style information, and then further structured and converted to SGML by USIT. For the third and subsequent catalogs, all the writers have used the SGML editor.

5.3 Printouts

The university had previously developed a print spooling system (PRISS), making it possible to print any file from any computer (Macintosh, PC, Unix Workstation) to any printer on the network. PRISS was applied to get printouts of the catalog on the writer's (or other) local printer with the same layout as the final catalog. The layout was an improvement on earlier versions created before the pilot project; for example, it was consistent throughout the catalog, it included a table of contents for each faculty, and introduced symbols in the margins to highlight important information.

TeX/LaTeX ((Knuth, 1984), (Leslie, 1986)) was used as the tool for generating postscript files in order to typeset the catalog on paper with an appropriate layout. Scripts were made to convert the SGML DTD to TeX/LaTeX.

The writers order printouts by using WWW interfaces prepared for them, making it possible to print different parts of the catalog to their local printer.

5.4 World Wide Web Presentations

To make the catalog available through the WWW, scripts for conversion from the SGML DTD to HTML were developed. The requirements considered for the implementation included presenting all the information in the printed catalog, making it possible to do dynamic updates, and making the result available for users of different WWW clients, e.g. for blind and visually impaired people.

5.5 Interdependence in Work

The coordination between the writers at the departmental level and the faculty became more "electronic". All writers have read and write access from their Macintosh or PC to their own catalog file(s), and read access to the other files. Server/Client technology is applied. From their desktop computers, they establish a connection to their catalog on the common Unix server, which is used to manage the different files and the access to them.

The writers leave their files on the servers. When the deadline expires, scripts merge the files to create one common catalog. This is sent to each writer's printer to produce a printout for final review and approval before an electronic version is created and the catalog is sent to the printing office.

5.6 Training and Support

Training and support was emphasized. Writers needed to learn about the structure of the DTD, and the Author/Editor. Twice a year the writers were invited to a two-day course covering both the structure and the editor. An email list and an "SGML phone number" were established for ongoing questions and comments from all the involved participants. The developers also visited the writers at their offices when needed.

6. The Evaluation of the Pilot Project

This section presents our empirical data from the evaluation of the pilot project. As shown in Table 1, the evaluation was a continuous process during the pilot project. The empirical data from interviews, emails, reports, telephone calls, direct mails and meetings form the basis for this section. The categories presented are based on the main issues raised by writers during the evaluation. The quotations below are from the interviews.

6.1 The Writing of the Catalog Production

During the pilot project, the number of writers grew and their work with the catalog production changed. The catalog was still a product, but use of the new technology changed the process. Before the pilot, several of the writers were primarily concerned with collecting information, and the typing was done at the faculty level. This changed to include the responsibility for direct data entry into the system as well.

During the pilot project, the process of producing the catalog was more time-consuming for the writers than before the introduction of the new infrastructure. Some of them had a lighter workload, and ended up as "experts" on the catalog. Others had a longer working day because of the additional demands. Still others received new work tasks, concerned with production of different kinds of information for distribution.

> I do another job now. I have been on courses, and spent a lot of time to become qualified to do my job. In fact, I should get paid more now (laughs).

The use of SGML requires discipline in the way text is written. Structuring the information according to a given definition of the document type creates constraints on dealing with it. Usually, people can present their information in their own way by using the tools they prefer. With the SGML, this freedom is restricted.

> It is problematic with SGML, because you have to be so damned correct, otherwise you get problems with your printouts. A few "typos", and then chaos. This is no problem in other word processors that I know. OK, you see the misprint on the paper, but you can read it, and use it!

Some writers pointed out that the freedom to use a well-known word processor, and to present the information in your own way, was gone. One said that he had the feeling of going back 10 years in time, dealing with text markup in editors like RUNOFF. The SGML editor used does not have the same functionality as word processors such as MS Word and WordPerfect. An SGML editor is an assistant for the writer doing markup according to the predefined document definition. It may incorporate a validating parser that makes it possible to avoid markup errors and guarantee that the document is structurally correct. Some writers stated that when working with the SGML editor they had to concentrate more on the technology and the structure than on the text itself. For most of the text, however, they needed only to fill in information in the right places.

> When using Word you almost forget that you are using a computer; it is just there – a tool, which is incorporated in my work. When using the SGML editor I have to think about how to use it – how to include which element, and so on. But I believe I will get used to it (laughs).

The writers were confused by the difference between the logical structure (represented by markup in the text) and the physical structure (or layout, presenting the catalog on paper). We received many questions related to the use of the logical structure. At the beginning, almost every writer related the logical markup directly to the printed catalog. Knowing how a specific markup in a context would look on paper, they used this markup for layout rather than for its logical meaning. For example, they wanted to use the element <emphasize> to mark up a title instead of using the element <title> in the appropriate context for this purpose. Problems related to the differences between logical and physical structure were felt strongly in the beginning of the project. The email concerned with these questions diminished during the pilot project.

6.2 The Need for Local Flexibility

There was a great deal of disagreement among the writers on how to structure the catalog, and what information elements should be required. The first design of the DTD was meant to be based on requests and proposals by the writers, managers and designers. However, the writers themselves did not have the same requirements for the DTD. On the contrary, they had conflicting requests and the proposals differed widely among departments, and between departments and faculties and the central administrational unit.

> No, I do not agree with anyone! In a way I am happy to have this opportunity to change the catalog. It should have been done years ago. And I see the potential of getting the catalog into SGML; you know .... save money on printing, and you know .... Web and all this stuff. However, we will never totally agree on a common structure for the catalog. The faculties and the departments are too different for that. I can spend hours telling you the structure and content I want, but it will, for sure, be in conflict with what she (a writer from another department) wants.
>
> How can anybody expect us to want to structure our information as a table? We are not the Department of Mathematics. We like to write and read prose. (Laughs). No, I do not have any rational reasons for that. (Laughs).

We experienced unwillingness among some of the writers to change their way of structuring and presenting the information. The writers had strong opinions on a detailed level about their own information. For instance, some wanted to have the general information about a course presented before other information such as day, time, place and lecturer. Others felt that it was important to have the general information at the end of the information about the course.

6.3 Some organizational aspects

There was a heavy workload for all the people involved in the process of introducing SGML at the university. It was time-consuming to develop both technical solutions and administrative routines. The project report documented over 1200 hours overtime for

the technical staff related only to the first edition. It decreased to less than 700 hours for the second edition, and decreased further with later editions.

As mentioned previously, the role of several of the writers extended from pure information-gathering to include updating the system and proofreading as well. In addition, the administration of much of the catalog production process was shifted from the central administration unit to USIT. The technical solution resulted in new deadlines for updates, approval of the catalog and delivery to the printing office. This led to a shift where USIT set the agenda for the writers by dealing with deadlines and organizing courses. USIT assumed responsibility for the graphic representation of the layout for both the printed and the electronic version of the catalog. The editorship was and still is the responsibility of the central administration unit.

6.4 Training and Support

All of the writers had to participate in training programs, and they all needed time to understand the underlying structure (DTD), to learn the SGML editor, and to get an understanding of how the integration of SGML might influence their work situation. The writers stated that they needed to learn and understand the SGML in order to work with it.

> I see the SGML people as a kind of a doctor for my information. They say that I have to mark it up to gain some new functionality. Of course, I will do that if I know why I have to do it. Comparing it to medicine – I take my medicine if my doctor tells me why I have to do so. I do not take medicine if the doctor cannot give me an appropriate reason. Obvious!

They also wanted to know the benefits of using SGML. They emphasized the need to know the main structure of the DTD, and the where and how of adding new information to the document. Knowing the structure of the DTD requires some understanding about what a logical structure is, and the ability to distinguish between the logical and physical structure. This took time to achieve, but it gradually evolved.

Despite the scheduled training, the writers needed access to some form of help all the time. They needed help to solve technical problems and to figure out what to do with the different parts of the information, how to code and where to put the markup. The interviews, email, and minutes from meetings show that the writers saw the training and support as highly important and necessary, and the many questions from the writers related to the process show that there were obvious reasons for emphasizing support.

6.5 The Catalog as a Product

The WWW version of the catalog became very much an electronic presentation of the printed version with the same sequence of the main structure elements, adding some new functionality for searching and navigation through listings of parts of the information. At the early stage of the pilot, very few of the writers were familiar with using the WWW, and they were mainly concerned about the printed version of the catalog. They primarily used the printed version as a tool in their work with student services.

After the deadline for the printed version, only a few writers took advantage of the possibilities to update the WWW version continuously. Some writers as well as some

managers were concerned about which presentation of the catalog should be used as the reference: the paper version or the WWW version. There was no overall agreement from the organization on this subject.

Writers emphasized that the catalog had for years been a kind of contract between the departments and the students. The departments demand that the students read (parts of) the catalog and that they follow the information provided there. On the other hand, the students use the catalog as documentation for what they need to know.

> Before, the catalog was a kind of a contract between the students and us, and we wanted it to be like this. How will this be when the catalog changes all the time?

They also mentioned the fact that students could bypass information by not clicking on links to it. Some writers feared that information they saw as important would be less visible on the WWW than in the paper version, and some feared that others would update the information. The writers are very much aware of their role in giving the students the right information. If the catalog does not contain enough information, or the students do not find the information they need, they ask the writers and others in the administrative units. In the printed catalog this information is represented in different "visible" chapters. In WWW they saw this information as "hidden" behind links.

7. Discussion

In this section we point out different aspects that demand serious consideration during the design of SGML-based infrastructures. The discussion is based on empirical data presented in previous sections, and seen from the perspective of system development. In relation to the main goals for the pilot stated in Section 3.1, the discussion focuses on the writing process and the catalog in use, since these are the main issues that the participants emphasized during the evaluation.

7.1 Focus on Structure Impacts Flexibility in Writing

There is a conflict between the requirements of a strict DTD and flexibility when authoring. As stated in the sections on empirical work, the writers felt that their freedom to write was restricted because of the editor and the underlying strict DTD. When writers use an SGML editor, they have to be aware of the predefined structure, and they have to deal directly with it when authoring. On the other hand, when using an SGML editor it is fairly easy to avoid ambiguity in markup at input. The editor displays only those elements that are valid in a particular context, which helps the writer to choose the right element. The SGML-encoded documents are ready for further use, handling and management without any conversion or other forms of adaptation. A more flexible DTD may offer the writers a more flexible writing process, but further use of the SGML-encoded document may be restricted, depending on the degree of functionality, (re)usability and the presentation of (components of) the documents required (Maler & Andaloussi, 1995).

Each department presents some general information to the students before the listing of the different courses offered by the department. This might be information about important dates, student services, services for disabled students, and so forth. This should be structured and presented in the same way for all the departments, making it possible to develop services based on the structure. In the DTD, the required

structure and content are specified. To fill in the information, the writers must use the correct markup for the different kinds of information and be aware of the predefined structure.

As stated in Section 6.3, the writers have reasons for rejecting the common structure. They could not see the point of making the structure common to all departments and faculties.

There is a tension between the concepts of local flexibility and "global" standardization. The standardization is necessary for communication over networks, and for automation of processes, like merging pieces of information into a catalog and publishing it on paper or the WWW, and advanced search through the catalog information. The tension between standardization and flexibility is also observed by Hanseth et al. (1996).

A solution to the conflict could be to offer writers other tools, free them from doing the coding, and instead absorb the cost of conversion into SGML. An obvious approach would be to use a WYSIWYG word processor and appropriate templates (Van Herwijnen, 1993). An argument for the use of WYSIWYG word processors and templates from the writers' point of view is that they can use the word processor they know well, and they deal with layout instead of logical markup. The change in their work situation may be smaller when a WYSIWYG word processor is applied, since the WYSIWYG word processor is already integrated into their work practice. However, we see some drawbacks as well. Studies have shown that users of WYSIWYG word processors do not necessarily apply the templates available, or template styles may be used incorrectly (Sørgaard & Sandahl, 1997). Templates that consist of many different styles are difficult to manage. The style list gets long, and which style to choose next is not necessarily obvious. These kinds of problems may force a lack of standardization that *may* introduce other problems - in interactions, merging, and presentation of the information.

As we see it, the choice of an SGML editor or a WYSIWYG word processor should be based on the complexity of the DTD at hand. If the DTD is small with few and understandable elements, a WYSIWYG word processor is preferable. Where the DTD is relatively complex, the effort of learning an SGML editor can be worthwhile, since the WYSIWYG word processor has no mechanisms for managing large sets of styles. In Braa and Sandahl (1998), different approaches to standardization of documents are discussed further.

Another way to solve this problem of restrictions on writing is to make more flexible and/or smaller DTDs that are tailored to suit each of the departments or faculties. To make the DTD more flexible, strong expressions like 'the information about *day* has to come before the *time*, which must be followed by *place*' can be replaced by expressions like 'the information about *day*, *time* and *place* must be present, but the sequence is optional'.

At the University, none of the writers use all the elements defined in the DTD. All of them use only a subset. For greater clarity, a DTD tailored for each department or faculty can be developed. In this case, all the elements that are not in use are stripped, and not visible to the writers at all.

However, these solutions have to be weighed against technical requirements and the effort of maintenance. DTDs are subject to change, and it is obvious that changing one DTD is simpler than changing 10 or 20. For a programmer, it is easier to make presentations on paper and the Web, or to prepare solutions for advanced search, if

the DTD is strict. With a highly flexible DTD the programming is far more complex, because of all the alternatives that have to be taken into account. From a user perspective, the DTD should be as flexible as possible. On the other hand, from a technical perspective, the DTD should be as strict as possible. Where the line is drawn depends on the situation. However, this decision is of vital importance for the success of the SGML system.

7.2 The Catalog in Use

From a rational point of view the case for an electronic catalog is more or less obvious: it keeps information more accurate, complete and up-to-date, improves portability and makes information less complex and less disorienting (Ventura, 1988). However, paper documents carry an aura of authenticity and legality that is difficult to dispel from peoples' minds (Berry and Goulde, 1994).

Some writers were unwilling to accept electronic documents as a substitute for "the real thing", especially at the beginning before they saw them for the first time. There were several reasons for this. People feared that information they considered important would become less visible in WWW than in the paper version. For instance, the information on the first pages in the printed catalog is regarded as important. The main concern of the central administration unit is the information to new students. To reduce the number of inquiries, they want students to read the information before they contact the department. The central administration unit was afraid that the student would click directly into the information about courses, and not see the other information. However, it is a challenge to present information on the WWW in a way that takes this into account.

The catalog is a contract, which is a genre of organizational communication between the students and the staff at the university (Yates & Orlikowski, 1992). As pointed out in the empirical data, the departments demand that the students read (parts of) the catalog and follow the information given there. On the other hand, students use the catalog as documentation for what they need to know. The catalog links the staff and the students together; it is necessary for the functioning of the whole university system. The writers experience a contradiction in the meaning between the catalog on paper and the catalog on WWW, because of the change from static to more dynamic information. Contracts are meant to be stable, but updating will of course occur. The writers were afraid that the technology would force an evolution from catalog-as-contract to catalog-as-encyclopedia. In this way the meaning of the catalog would change, and it would no longer be a contract, but more a source of information. However, updates occurred as notices on boards. The students are committed to paying attention to the boards. The catalog, like documents in general, has both fixed and fluid properties, independent of which media they are based on (Levy, 1994).

We regard it as important to organize updates in the catalog in such a way that the catalog is still regarded as stable and at the same time a source of relevant information. This is not necessarily a contradiction. Information about deadlines, intake, rules, and so on, is relatively stable, and it has to be. Information about course times and locations is more dynamic, and the students know that this information may change during the semester. To make it possible to rely on the electronic catalog, conventions such as 'keep an eye on the notice board' in the 'paper world' have to be developed and introduced in the 'electronic world' as well. Probably, these conventions will be developed over time (Brown & Duguid, 1994, Yates & Orlikowski, 1992). However, we propose to develop new ways to take this convention into account, and make them explicit for the organization.

We observed that the catalog had underlying and "hidden" intentions, which are developed over time and integrated into ongoing practice. The catalog is more than a medium to communicate information from the administration to students; it is an artifact that also coordinates work practice, e.g. the writers' coordination of the updates and the students' organization of the semester.

To produce the catalog for the next semester, the writers take the previous year's catalog as a starting point and update dates, delete old courses, and so on before they print out their pages. However, there are blank spaces and question marks in the text, since this is just a draft. The printouts are sent to the lecturers and others that have opinions and the right to make changes. The lecturers make their updates on the paper, and return it to the writer. The writer coordinates the rooms and types the new and updated text into the catalog. The 'new' catalog is distributed in the same way for proofreading until it is finalized.

The main goal of the catalog is to be a tool for students to plan and organize their semester. They use the catalog to get to know about the studies at the University, to decide which courses to take part in, what time they want to have the group lectures in order to avoid conflict with other courses or part-time jobs, and so on.

These two examples show that the catalog is integrated into ongoing practice, and not something 'added on'. There is no limit on how the catalog can be used. We have given only two examples here. The standardization of the catalog impacts the catalog in use, and it is important that the local need for flexibility of use is regarded in design. We have to be aware that the catalog, or other documents, is embedded in practice, and has roles there. This is further discussed in Braa and Sandahl (1998b). As an example, at least the printing facility, as discussed in the section above, has to function properly to support the need for drafts. In addition, the electronic version of the catalog should help students to plan their semester, because that is one of the key objectives of the catalog.

8. Concluding Remarks

This paper describes the first steps of designing an SGML-based infrastructure. The goals of the pilot project presented were to produce a better catalog through a better structure and layout, to make it easier to update and maintain, and to gain practical experience in SGML. We also wanted to develop services like room allocation, advanced search and customized publishing based on the structured catalog. However, in the period of the pilot project we did not achieve all these goals. The design and development had a technical focus, and significant problems related to use occurred as a consequence.

Based on our experiences in this pilot project, we have the following concluding remarks:

i) In order to benefit from SGML-based documents, they have to be produced in the first place, which requires a satisfactory writing environment. Selection of the editor or word processor to be applied is essential, and has to be taken into account in DTD design.

ii) Documents are integrated into work practices; they have roles, and conventions grow around them. An SGML implementation may (differentially) change these roles and conventions and thereby impact work practices.

These points are discussed in more detail below.

8.1 DTD Design

The DTD has to be designed in a way that supports the functionality desired. From a technical perspective a strict and rich DTD is preferable. From a use perspective the DTD should be flexible and as small as possible. We have indicated three different approaches to the problem. First, the DTD can be made more flexible by reducing elements and changing required sequences to optional sequences where appropriate. This is the easiest way to reduce strictness, and the most obvious. Second, the DTD should be made smaller by reducing the number of elements. This can affect the functionality, because some definitions disappear. However, in our experience there were far too many elements from the start, and the reduction of elements did not necessarily affect the functionality desired. Third, 'writers' DTDs' can be introduced as a subset of the full DTD. These DTDs consist only of elements that a particular writer uses. The effort of maintenance has to be taken into account before implementing a solution like this. The same goal is reached if each writer sees only the elements (s)he needs. The DTD is the same, but the editor 'hides' the unnecessary elements. Using predefined forms for each writer is an equivalent solution. In short, the DTD design impacts both the writing and the presentation or functionality, but in opposite ways. What one should emphasize depends on the situation at hand (competence of writers, degree of functionality, and so on). We state that SGML editors are better tools for producing text based on a complex DTD than WYSIWYG word processors are. On the other hand, if the DTD is simple and small, a WYSIWYG word processor should be applied. The fact is that the WYSIWYG word processors are highly integrated into work practice. The writers know them and use them daily. However, if there is a need for continuous conversions to and from SGML to provide continuous updates, the use of templates can be a difficult approach, because of the errors that often occur in conversions (Maler & Andaloussi, 1995).

8.2 Catalog in use

Because of the paper documents' ecological flexibility, they easily fit into different situations for staff and students. We have observed conventions related to updates of the paper catalog, conventions among coordinating the updates, and students' use in the organizing of their semester. The catalog is essential in these situations; however, it plays different roles. To be able to achieve a critical mass of use (Grudin, 1994) the electronic catalog has to be flexible enough to be adapted to different situations. Some conventions related to the paper disappear, and new ones have to be introduced to get the electronic documents adopted into the organization.

8.3 Goals That Have Been Accomplished

In a survey carried out by a master's degree student, students reported that they were satisfied with the 'new' printed catalog. It was experienced as more structured than the earlier ones, and easy to navigate in (Markussen, 1998). The accounting department reports that the catalog costs 200,000 less than the earlier versions not based on SGML. The reason is that the number of pages has decreased (Ibid.).

8.4 The Pilot Projects' Implications for Further Development

Further development based on this experience has been carried out to achieve more of the overall goals. In brief, we can mention:

8.4.1 Catalog as One of Several Documents

The organization has gained a whole new understanding of the content of the information produced, and there are several initiatives from different units at the University to deal with different document types according to the solutions for the catalog. Several document types in addition to the catalog now constitute an infrastructure of information, making it possible to reuse information components across different types of documents. Future work will address further integration with other information systems, e.g. database solutions for student systems for all universities and colleges in Norway. This has the potential to improve students' possibilities for planning the semester and signing up for exams.

8.4.2 DTD Design

Modular DTDs have been developed to tailor DTDs to specific document types and to reuse general structures of elements between the DTDs.

8.4.3 Functionality for the Writer

Administration routines are available through the WWW, including basic functionality such as publishing a document (making it available among the official WWW information), archiving and restoring data, ordering printouts and conversion to the local WWW (to look at the result before publishing). Functionality is tailored to suit different document types, e.g. extracts of the information on lecturers and lecture rooms in the catalog for support of the planning work.

8.4.4 The WWW Presentation

The conversion routines have been developed further to provide integrated solutions for including the university's visual profile on every HTML page, including relevant meta information such as "last updated by", who is responsible and contact points. In addition, there are better navigation tools to show the relevant context of the HTML page at all times, and improved search facilities. It is still a challenge to tailor the WWW presentation to different user groups, e.g. new students, researchers, or administrative staff.

8.4.5 The printing

Much work has been done to improve the printing facilities. Today, the SGML files can be printed regardless of their status.

8.4.6 Continuous updates

The organization has developed strategies and routines for dealing with continuous updates for some of the document types, e.g. the catalog. The aim of these strategies is to make it clear how late the writers can make updates in relation to deadlines for printing, and how and where students can catch up on updates in the electronic version. The latter has been designed, but not yet implemented at all levels at the university.

Acknowledgements

References

Alschuler, Liora (1995) *ABCD ... SGML A user's guide to structured information*, International Thompson Computer Press.

Berry, Margaret D and Goulde, Michael A (1994). *A New View of Documents. Integrated Information Management in the ´90s*. In Workgroup Computing Report Vol. 17, No 8.

Braa, K and Sandahl, T. I. (1998). Approaches to Standardization of Documents. In Toshiro Wakayama, Srikanth Kannapan, Chan Meng Khoong, Shamkant Navathe, JoAnne Yates (Eds.) *Information and Process Integration in Enterprises. Rethinking Documents*. Kluwer Academic Publishers.

Braa, K and, Sandahl, T. I. (1998b). From Paperwork to Network. Submitted to *Third International Conference on the Design of Cooperative Systems (Coop '98)*, Cannes, France, 26.-29. May 1998.

Brown, John Seely and Duguid, Paul (1994) Borderline issues: Social and Material Aspects of Design. In *Human Computer Interaction*, 1994. Volume 9, pp 3-36.

Date, C. J. (1986) *An introduction to database systems*. Addison-Wesley Publishing Company.

Goldfarb, Charles, F (1992). *The SGML Handbook*. Clarendon Press, Oxford.

Grudin, Jonathan (1994). Groupware and Social Dynamics: Eight Challenges for Developers. In *Communication of the ACM*, Vol. 37, No 1, pp 93-105.

Hanseth, Ole, Monteiro, Eric and Hatling, Morten (1996). Developing information infrastructure: the tension between standardisation and flexibility. In *Science, Technology and Human Values*, Vol. 21 No. 4, pp. 407-426, Sage Periodicals Press 1996

Herwijnen, Eric Van (1993) *Practical SGML*. Kluwer Academic Publishers.

Jenssen, Astrid E. and Sandahl, Tone Irene (1996). Conflicts between the possibilities and the reality in the field of structured electronic documents: Experiences from a large-scale SGML-project. In Bo Dahlbom et al., editors, Proceedings of the 19th *Information Systems Research Seminar In Scandinavia*, pages 935-955, Lökeberg (Göteborg), 10-13 August 1996.

Jewett, Tom and King, Rob (1991). The Dynamics of Computerization in a Social Science Research Team: A Case study of Infrastructures, Strategies and Skills. *In Social Science Computer Review* 9:2 (1991). Duke University Press.

Knuth, Donald (1984): *The TeX-book*, Addison-Wesley, Reading, Massachusetts and Lamport.

Leslie (1986): *LaTeX, user's guide and reference manual*, Addison-Wesley Publishing Company, Reading, Massachusetts.

Levy, D. M. (1994) *Fixed or fluid? Document stability and new media*. In proceedings of ECHT´94. ACM (pp 24-31), New York.

Maler, Eve and Andaloussi, Jeanne El (1995). *Developing SGML DTDs: From text to model to markup*. Prentice Hall.

Markussen, Atle Holmer (1998). *Fra produkt til tjeneste i dokumentbaserte IS. Perspektiver og utfordringer*. Master Thesis (in Norwegian). Department of Informatics, University of Oslo.

Reinhardt, Andy (1994) *Managing the New Documents* in Byte, Vol. 19, No. 8, August 1994.

Smith, Joan M (1992) *SGML and related Standards. Document description and processing language.* Ellis Horwood Limited.

Star, Susan Leigh and Ruhleder, Karen (1994). Steps Towards an Ecology of Infrastructure: Complex Problems in Design and Access for Large-Scale Collaborative Systems. *In Proceedings of CSCW 1994*, ACM. Chapel Hill, NC, USA, pp 253-264.

Sørgaard, Pål and Sandahl, Tone Irene (1997). Problems with Styles in Word Processing: A Weak Foundation for Electronic Publishing with SGML. In *Proceedings from 30th HICSS*, Wailea, Hawaii, Jan 7-10.

Ventura, Cheryl A. (1988). Why Switch from Paper to Electronic Manuals? In Proceedings from Santa Fe, New Mexico, December 5-9 1988.

Yates, JoAnne and Orlikowski, Wanda (1992). Genres of Organizational Communication: A Structurational Approach to Studying Communication and Media. In *Academy of Management Review* 1992. Vol. 17. No. 2, pp 299-326.

Problems with Styles in Word Processing: A Weak Foundation for Electronic Publishing with SGML

In Proceedings of the 30th HICSS, January 7--10, 1997, Wailea, Hawaii

Pål Sørgaard and Tone Irene Sandahl

Abstract

Most word processors have facilities for styles and document templates. These mechanisms may help maintain typographic consistency while editing, and they may support document exchange and conversion. Document conversion is of special importance to electronic publishing. Our own experience suggests that there {\em are} problems with the use of styles and templates. A sample of documents drawn from three case organisations confirms this suggestion, and indicates that the use of these mechanisms is infrequent and riddled with difficulties. An initial classification suggests that most of the problems with the use of paragraph styles can be described as ignoring or overlooking these mechanisms. These problems have consequences, and one should be careful in assuming that material written with current word processors easily can be converted to formats for electronic publishing. Careful organisational implementation of word processing may be critical. Our interpretation of the problems encountered also indicates that they may be related to the paper metaphor communicated by the principle of WYSIWYG. We claim that the paper metaphor does not communicate any understanding of the structure beneath the surface of a digital document.

1. Introduction

The recent massive diffusion of the World Wide Web (Berners-Lee et al., 94) has turned electronic publishing into a practical reality for many providers and consumers of documents. As a result, the aim of electronic word processing has expanded. Where traditionally the emphasis has very much been on the production of printed documents (see, for example (Preece et al., 94 p. 18,) and (Barker et al, 92), today the purpose of word processing for many users is extended to the production of digital documents.

There are several important opportunities with digital documents in areas such as search, retrieval and reuse of text. Several organisations plan to use these possibilities. For an example, see ACM's Electronic Publishing Plan (Denning & Rous, 95).

Several of the opportunities of digital documents can only be fulfilled if the documents are available in some standardised format, as specified by, for example, SGML (Standard Generalized Markup Language, see section 2) (Goldfarb & Rubinsky, 90). In fact, the World Wide Web relies on this standard, as the Web's language for document exchange, HTML (Hyper Text Markup Language), is defined as a so-called SGML DTD (Document Type Definition).

Hence, to provide a Web-service, and to benefit from digital documents, it is highly relevant to look at how to make text available in SGML (or some other standard). Eric van Herwijnen identifies four approaches for capturing text in SGML (Herwijnen, 90, ch. 5):

1. Typing SGML as plain text,

2. OCR (optical character recognition) reading of existing documents combined with an intelligent markup system,

3. converting text written in conventional word-processors, and

4. use of specialised SGML editors.

Today, the third approach could be extended to include word processors with built-in converters to HTML/SGML. van Herwijnen argues that the success of the third approach will depend on the use of so-called paragraph styles (see section 2). People working in Web-services often spend a lot of time cleaning up errors in automatically converted HTML-files. Moreover, conversion to HTML can be much simpler than conversion to "richer" Document Type Definitions. Such rich DTDs are needed to get the full benefit of digital documents, i.e. search and retrieval based on the logical structure of the documents (see figure 1).

Paragraph styles provide a mechanism which helps such conversions, and paragraph styles can conceivably be used for conversions to very rich DTDs. Little is known, however, about the actual use of paragraph styles in word processing. In this paper we report from the first of a series of studies of use of word processing, electronic publishing, and organisation of Web-services. Here we study the use of paragraph styles, in order to discuss whether current word processors are suitable applications for the capture of text in SGML. Our findings are not restricted to SGML, but apply to several contexts where text is to be converted and made available in more than one medium.

In the next section paragraph styles, SGML and related issues will be discussed. In section 3 our research method is described. In section 4 we present a classification of the problems we have observed with the use of paragraph styles. Our findings are discussed in section 5, and finally some conclusion are drawn in section 6.

2. Styles, Templates, Markup and SGML

Styles or *style rules* are widespread mechanisms in word processing. Styles have been available in commercial WYSIWYG (what you see is what you get) word processors since the end of the 80s, and were available in several research prototypes a decade earlier (Johnson & Beach, 88). Styles are named formats, enabling consistent formatting of all parts (typically paragraphs) of a document referring to the same style.

Styles can be applied to several property domains of a document, i.e. characters, paragraphs, tables, headings, etc. (Johnson & Beach, 88). Typical paragraph styles are 'Heading1,' 'Body,' 'Author,' etc. A word processor will typically have a dialogue box for the definition of properties of a style.

It is common to define *style sheets* or *document templates* which contain a coherent set of styles to be used for the composition of specific document types like letter, article, memo and book. Often efforts are made to make a set of company specific

templates. Making good document templates requires good knowledge of the word processor as well as a fair sense of typographic quality.

Johnson and Beach distinguish between *static functionality* and *dynamic functionality* of a word processor (Johnson & Beach, 88): Styles do not add to the power of a word processor in terms of what kind of paper documents it can be used to produce (the static functionality is the same). The dynamic functionality, however, is increased through the use of styles, typically in terms of making consistent updates, document-wide format changes, typographic consistency, and reuse of text more easy.

In classical, markup-based, word processors (like LaTeX (Lamport, 86)) there is a distinction between *visual markup* and *logical markup* (Flynn, 95). Visual markup indicates how the text should be set (size, width, etc.), while logical markup classifies or "tags" the text so that it can be appropriately set using the definitions of the word processor (or of the user, if the definitions have been modified). LaTeX is in fact a macro package implementing logical markup on top of TeX's (Knuth, 84) commands for visual markup.

SGML (ISO-8879) is a standard for how to describe and provide logical structure in documents. Originally SGML was introduced for technical documentation, but has later been used in several other areas. Logical structure is defined in Document Type Definitions (DTDs), for an example see figure 1. DTDs essentially define context-free grammars (Aho & Ullmann, 72) for the markup of one type of documents. The markup itself (see an example in figure 2) follows specific conventions for syntax: <EMNE>...</EMNE> is a typical example of an element. Moreover there is notation for *attributes* (i.e. DEL="in260"), and for *entities* (e.g., å used to denote `å'). A DTD may or may not be a part of the document itself. There are also several standard DTDs, some of which defined by ISO 12083. A wide range of SGML-tools like parsers, editors, converters and browsers is available.

Figure 1: Part of a rich DTD for describing courses

```
<!element emne - - (kode?, emnenavn?, emneinnl?, emneund*) >
<!element emneund - - (delemne+ | (periode+ | seksjon+ | (urekke | underv)*)) >
<!element delemne - - (kode?, emnenavn?, emneinnl?, emneund*) >
<!element ikkeund - - (tittel, (kode, emnenavn)+ ) >
<!element kode - - (#PCDATA) >
<!element emnenavn - - (#PCDATA | stikkord) * >
<!element emneinnl - - (vektall?, avsn*, (ukap | oriente)*, eksamen?) >
<!element vektall - - (#PCDATA) >
<!attlist vektall synlig (synlig | usynlig) usynlig>
<!element tid - - (#PCDATA) >
<!element underv - - ((nr | tittel)?, avsn*, ((dag | dagint | dagoor | dato | fulldato), tid*, sted?,
navn*, start?, varighet?)*, unfelles?) >
<!attlist underv type (forel | gruppe | seminar | kollok | ovelse | orakel | kurs | lab) #REQUIRED
          oblig (obl | ikke-obl) ikke-obl
          aapen (aapen | lukket) aapen >
<!element nr - - (#PCDATA) >
<!element dag - - (#PCDATA) >
```

```
<EMNE DEL="in260">
<KODE>IN260</KODE>
<EMNENAVN>Systembeskrivelse og spr&aring;k,
5 vt</EMNENAVN>
<EMNEINNL><VEKTTALL>5</VEKTTALL></EMNEINNL>
<EMNEUND>
<UNDERV TYPE="FOREL"><TITTEL>Forelesninger</TITTEL>
<DAG>Mandag</DAG>
<TID>12.15 - 13</TID>
<STED>Lille auditorium</STED>
<DAG>Onsdag</DAG>
<TID>14.15 - 16</TID>
<STED>Lille auditorium</STED>
<NAVN><FNAVN>P&aring;l</FNAVN>
<ENAVN>S&oslash;rgaard</ENAVN></NAVN>
<NAVN><FNAVN>Tone Irene</FNAVN>
<ENAVN>Sandahl</ENAVN></NAVN>
<START>22. januar</START></UNDERV>
<UREKKE><TITTEL>Gruppe&oslash;velser</TITTEL>
<UNDERV TYPE="GRUPPE"><NR>Gr. 1</NR>
<DAG>Mandag</DAG>
<TID>15.15-18</TID>
<STED>Sem 3A(+3C,3D)</STED></UNDERV>
<UNDERV TYPE="GRUPPE"><NR>Gr. 2</NR>
<DAG>Torsdag</DAG>
<TID>16.15-19</TID>
<STED>Sem 3A(+3C,3D)</STED></UNDERV>
<UNDERV TYPE="GRUPPE"><NR>Gr. 3</NR>
<DAG>Fredag</DAG>
<TID>9.15-12</TID>
<STED>Sem 3A(+3C,3D)</STED></UNDERV>
<URFELLES><START>22. januar</START></URFELLES>
</UREKKE></EMNEUND></EMNE>
```

Figure 2: Example of a SGML encoded document

SGML is built on the idea of a strong separation between markup and lay-out (formatting *can* be specified, but this is not encouraged). SGML-files are assumed to be processed by suitable output programs, which will process the SGML-tags and make lay-out according to instructions from some other source. Hendry argues that it may be very hard to write a text independently of how it will be processed (Hendry, 95), nevertheless the potential is attractive, and there are working examples. At the University of Oslo, the complete catalogue of all teaching is written in SGML, with a DTD defined for the purpose (Jenssen & Sandahl, 96). The printed catalogue is produced through a conversion from SGML to TeX, while an online version is available in WWW. Figures 1 and 2 originate from this example. Course information is typed in and modified using a syntax directed editor: the editor only allows input as defined by the DTD, tags need not be typed, but can nevertheless be displayed. The system takes input and displays output in ISO-Latin-1, reducing the need for SGML-entities like å.

Use of paragraph styles have several potential benefits, ((Barker et al., 92), (Bryan, 94), (Johnson & Beach, 88), (Miastkowsky, 93)). For the purposes of this paper, we focus on benefits related to electronic publishing and SGML, three of which are identified below, named and labelled with letters to ease reference to them later in the paper.

M: meaning

Paragraph styles provide a mechanism for classifying text. Different kinds of paragraphs can be explicitly named, using paragraph names like 'Example,' 'Definition,' etc. This is useful in large, technical documents like software manuals. This may be used to attach meaning to the text, supporting the authors in discussions about the document, its content and its lay-out.

C: conversion

In several settings there is a need to convert the files for a document. Conversions may be needed to transfer a document between different word processors or to formats like SGML. Today, authors often need a lot of assistance to get documents on the World Wide Web. Sometimes separate organisational units are set up for this purpose. Electronic publishing efforts are simplified if the authors use a format which is easily converted to HTML.

S: search

Indexing, search, and retrieval are improved when text is equipped with logical markup. Although free-text search also has its advantages, it is certainly desirable to be able to distinguish between occurrences of a name as an author name and as any other textual occurrence of the same name. ACM's Electronic Publishing Plan (Denning & Rous, 95) explicitly calls for digital documents which are logically structured for search and retrieval.

3. Research Method

We used a pragmatic approach for analysing use of styles in word processing. We used our established contacts with three different organisations to get access to documents and to conduct interviews with users.

The idea behind our approach was to analyse use of paragraph styles in a set of "normal" documents. Since we found no other studies or attempts to systematise the use of paragraph styles, we developed our own classification of the problems we found.

3.1 Initial Classification of Problems

Before we started our empirical work we identified four categories of problems associated with the use of paragraph styles: overlooking styles, weak functionality, semantic problems and structural problems. This classification was based on personal experiences with word processing and on experiences from development projects and consulting. Our understanding and naming of these problems changed during the study. Still, the initial categories survived, and have been used to structure the findings in this paper (subsection 4.1-4.4).

3.2 Case Organisations

We collected documents from three different organisations. Here we call them AA, BB, and CC, respectively.

AA

has 100 employees, most of whom are engineers, mathematicians or computer scientists doing contract research. AA is highly computerised, each employee has at least one computer. Several different kinds of computers as well as different word processing packages are being used. There is little systematic education in word processing. Document templates are available, their use being mildly encouraged, but not followed up. Only limited resources have been put into developing templates, although efforts have increased during the last year. Facilities for converting documents are not always available.

BB

has 4.500 employees, most of whom use Microsoft Word or Word Perfect. Within the administration, Word is the dominating word processor. Every year BB offers courses in word processing for both Microsoft Word and WordPerfect. In 1995 about 500 persons participated in internal word processing courses. Availability and enforcement of document templates vary heavily. Within BB we have had contacts with the *administrative* staff at different levels. Some of these have attended courses in word processing, others have not. For the administration, a document template for letters is available from servers in the network.

CC

is a major public bureaucracy, consisting of several, large, and very independent units. Most employees use PCs, using AmiPro, Microsoft Word and WordPerfect for word processing. From 1993 CC introduced email for most employees, and in 1995 CC launched a common, externally oriented WWW-service. This service is considered politically important. The WWW-service is operated by a newly established unit (CCW) within the department for common services. CCW receives documents for publishing via electronic mail, on floppy disks, and sometimes on paper. Documents are normally received from the authors, but for large publications the text is often obtained from a print shop.

Use and availability of document templates vary heavily among the units in CC. CCW has no authority over authors, and makes documents available on an "as is" basis, taking whatever steps which may be needed (in some cases retyping the text). Thus the burden of authors not using appropriate templates falls on CCW. As often as possible, the personnel in CCW perform automatic conversions. Within the half-year period we have followed CCW there has been an increasing emphasis on encouraging the use of paragraph styles among authors. New tools for document analysis and conversion are brought into use as they become available.

3.3 Selection Procedure

We had different strategies for the selection of documents in the three organisations, but the general emphasis was on selecting documents where use of paragraph styles would make sense.

In AA we searched for a project where several authors were involved. We obtained access to a more than 300 page software manual (referred to as AA-01) written together with people from an external customer. The documented software was of a very technical nature and was written by domain specialists, not by computer scientists. To avoid bias we considered avoidance of computer scientists important, and this made it hard to find many more documents.

In BB we randomly selected 18 persons in the administration, and asked them to send us the last three documents they had written. Not all responded. We received 18 documents from nine different authors.

In CC we asked CCW to file all documents they received for publishing on the Web during a week. The result was a floppy disk containing 27 documents.

Document collection was supplemented by interviews in AA and BB. Some interviews were conducted using the telephone, and others were face-to-face meetings. We did not draw up a structured questionnaire, but asked questions related to the concrete use of styles in the documents we received.

Given the difficulties in obtaining a representative sample of documents, we restrict our interpretations to state the existence of problems. We cannot report reliable figures about "problem frequencies," but we believe the classification reported here can be used to compare the use of word processing in two or more different organisations.

3.4 Document Analysis

Document analysis was primarily performed manually. We addressed only the use of paragraph styles. This delimitation means that there are other problems in the way word processors are used which will not be covered. It is only when paragraph styles have a role that we identify the problem.

Examples of problems falling outside our categorisation are the practices of adding empty paragraphs to align a new heading on top of a page and of "simulating" an indented list by manually breaking the text in single lines and indenting these lines with tabs or blanks (this is also reported by Trigg and Bødker (94)). Of course, such practices create problems when the document is being changed or converted. To us, however, this is not related to paragraph styles. It is more an issue of other mechanisms in the word processors not being used as intended.

Manual document analysis was performed by opening documents with the word processor. When going through the documents, we looked for use and non-use of styles, redefinitions of styles, manual formatting, and all aspects of paragraph styles we could think of. For some of the documents we conducted follow-up interviews with authors or other people involved. The findings were classified in discussions in our research group. The problems presented below were developed bottom up from our findings, but we choose to present them within the four predefined categories, as there were not really any new categories of problems encountered.

In an attempt to do automatic document analysis we developed a system based on DynaTag (from Electronic Book Technologies) and Balise (from Advanced Information Systems/Berger-Levrault). With this system we could automatically compare the use of paragraph styles (and indeed use of character formats) with the structure defined by a "normal" Document Type Definition provided by us. Automatic document analysis turned out to be of little use. It could only be applied to Microsoft Rich Text Format (RTF), and thus hard to apply to all documents. Moreover, the automatic analysis was geared to look for structural problems, a category of problems which was not very common.

4. A classification of problems

In this classification section we present a classification of the problems we found. In most of the categories we have identified several subcategories. There were large frequency differences among the problems. For each problem we also try to identify which of the benefits of styles (identified in section 2) which may not be achieved when the problem is present. Other consequences, like loss of dynamics in editing, are not discussed. The findings are summarised in table 1.

4.1 "Overlooking Styles:" Not Using Style Capabilities

We define "overlooking styles" as the manual formatting of text which could have been formatted using paragraph styles. Since the use of paragraph styles was not very extensive, this category is by far the most common category. Analysis of the documents reveals a series of different problems in this category.

I-a: Style exists but is not used.

It is very common to find documents where the document template contains styles which could have done the required formatting, but where these styles are not being used. Instead the format of a paragraph is modified to achieve the same effect. We have seen this in practically all the documents we analysed. In many documents styles are not used at all. It is also quite common with documents that only use the most elementary styles, typically the styles for headings at the higher levels. Some interviewees explained that these styles were useful, as they helped constructing the table of contents. In a large technical document (AA-01), where styles were used heavily, we found spurious examples of "manually" created section titles at low levels.

Not using existing styles may lead to problems when the file is to be *converted* to some other format. Moreover, depending on which styles are not being used, *search* (for, e.g., 'Author') and *meaning* (of, e.g.,'Example') may also suffer.

I-b: New type of paragraph, but no style.

In developing a text, new "types" of paragraphs often come into being. In a strategy document we analysed, there were some important paragraphs in each section, and these paragraphs were printed in bold. This consistent formatting was done manually, rather than defining a new paragraph style. Clearly, consistent formatting of a new kind of paragraph is more easily achieved with a specific style.

This problem may have consequences when the file is *converted*. In addition, the possible benefits of attaching *meaning* to paragraphs are lost.

I-c: Examples of a style consistently reformatted, while style is unchanged.

In early versions of one of the documents we studied, all examples of one specific heading style were reformatted. This was not done, however, through a modification of the style for that heading, but by manually changing the format for each individual example of the kind of heading. In the final version of the document this had been changed to a corresponding modification of the style.

Although inconvenient, this problem does not affect any of the mentioned benefits of meaning, conversion and search.

I-d: Incidental use of style.

In some of the documents from BB, only one paragraph style was applied, for example the style 'Date'. This came as a surprise to us, since we had expected that if a document should have been written in only one paragraph style, it would be 'Normal.' It turned out that many documents were written using a letter template, where 'Date' was the first style encountered. As a side effect, we saw an example where a quote which clearly was pasted in from another document, was carrying the style 'Date'. We have personally experienced this problem in other contexts, where the choice of template, however, was more appropriate.

This problem causes difficulties for any conceivable benefit from the use of paragraph styles. *Meaning*, *conversion* and *search* are all affected. As an example, one needs to be prepared for documents where there are many examples of the 'Date' style, and where the contents of such paragraphs do not have any resemblance with a date.

4.2 "Weak Functionality:'" Style Use Made Difficult

Style mechanisms often have unsatisfactory designs. This can be observed by analysing documents, and is also brought forwards in interviews with users. It is our impression that these problems make style use difficult or even directly discouraging. What we report here are problems with the use of styles which we have observed and which we believe are related to the design of style mechanisms. To try to catch the full range of problems with styles, a more detailed analysis of different implementations of style mechanisms is needed.

II-a: Context-free mechanisms cause many styles.

In a large technical document (AA-01) we were struck by the number of paragraph styles associated with various kinds and levels of indented lists. The lists occurred at three levels, and in three kinds: bulleted, numbered and equipped with descriptions. To achieve good lay-out, therewere separate styles for beginnings and ends of lists. As a result, 24 out of a total of 45 paragraph styles were involved with lists. Even so, not all alternatives were available at level three, and there were no definitions beyond that level.

In the word processors we looked at, it is not possible todefine that a style should be indented 1 cm relative to the preceding or surrounding paragraph. There is no notion of *context-sensitive* style definitions (see Johnson and Beach (88)) for ananalysis of various design alternatives).Hence the style for an itemised list must have different names for each level of indentation.This problem does not occur in markup based systems like LaTeX or with DTDs like HTML: both feature general definitions of three kinds of lists: unnumbered, numbered and labelled.

This problem clearly makes it difficult to work with styles. There is an unnecessary conflict between simplicity (few styles) and generality(many styles) for those who make document templates. Document *conversion* is made difficult, although not impossible.

II-b: Copy and paste resulting in enormous style catalogues.

In some word processors, it often happens that when text from one document is pasted into another, the style catalogue of the target document is updated with the styles of the pasted text. We have found documents with style catalogues containing styles which obviously do not belong there (see also problem I-d above). We and our

colleagues have experienced shared authoring or editing processes where the number of styles became simply overwhelming. We have experienced as much as 200 styles in a catalogue, many of which were "synonyms." Mixing different language versions of the same word processor tends to generate this problem.

It is clearly hard to work with large style catalogues. Selecting the appropriate style requires scrolling in the style catalogue, and it often happens that one document mixes different styles for the same purpose. It is not surprising that those who have been confronted with such examples find paragraph styles hard to use and understand. Large style catalogues also make *conversion* more difficult, since large tables of style names need to be maintained.

II-c: Association with line-oriented formatting.

We analysed several press releases from CC. They were all written according to an agreed lay-out, although they were not written using any document template. In the agreed lay-out, several different "fields" of information, for example date and organisational unit, are placed on the same line. Since paragraph styles are associated with complete paragraphs, they always result in a line break, making it impossible to combine the agreed lay-outs with paragraph styles carrying meaning[6].

This problem has several consequences. Use of paragraph styles to attach *meaning* to elements of the document is discouraged. The example with heading lines containing several important pieces of information is typical. Some of the fields found in such headings could also be useful when *searching* for documents. Depending on how users circumvent the problem, problems may also arise with *conversion*.

4.3 Semantic Problems

We define "semantic problems" to be situations where styles with specific meaning are used without regard to this meaning.

III-a: Wrong document template.

In BB we saw several examples of documents written with a template for letters although the documents themselves were not letters. Clearly this resulted in strange use of styles as well (see problems I-d and III-b). An obvious reason for the problem to turn up was the lack of other document templates.

This problem does, like problem I-d, cause difficulties for all mentioned benefits of styles.

III-b: Wrong paragraph style.

When a paragraph style is used only for its lay-out properties, but contrary to the meaning associated with the name of the paragraph style, we would classify it as use of the "wrong paragraph style". The 'Author' style, for example, should not be used as a convenient way to centre text in a heavy type face. We have encountered this problem personally, but we did not really observe it in the set of documents we analysed. Those examples we found were hard to interpret as using the wrong

---

[6] Clever use (abuse?) of table mechanisms or columns may be used to circumvent this problem. Alternatively character styles could be used, but in most cases it is natural to use paragraph styles to classify text, like with `Author,' `Abstract,' etc

paragraph style. Instead we classified them as incidental use of styles (I-d), related to use of the wrong document template (III-a).

Again, all mentioned benefits of using paragraph styles are affected.

Table 1: Overview of problems and Consequences, M = meaning, C = conversion, S = search

Problem	Where found	"Frequence"	Consequences
I-a style exists but not used	All	very high	C(MS)
I-b new type of paragraph, but no style	BB	high	MC
I-c examples of a style consistently reformatted	BB-01	observed	-
I-d incidental use of style	BB	high	MCS
II-a context-free mechanisms cause many styles	AA	observed	C
II-b copy and paste resulting in enormous style catalgoues	BB	observed	C
II-c associate with line-oriented formatting	CC	observed	MCS
III-a wrong document template	BB	high	MCS
III-b wrong paragraph style	-	see I-d	MCS
III-c logical style not applied	AA-01	low	MS
IV structural problems	-	not observed	C

III-c: Logical style not applied.

Some documents have a well developed set of logical paragraph styles, styles which are of specific relevance to the document or the kind of document in question. In document AA-01 there were special styles for 'Code,' 'Note,' and 'Example,' etc. Sometimes, such logical styles are defined, but not used when appropriate. There is clearly a heavy overlap between this problem and problem I-a. When we propose this as a separate problem, we want to highlight the potential for expressing meaning with the paragraph style mechanism.

The main consequence of this problem is the lack of representation of *meaning*. In addition, *search* for documents with specific content is made more difficult. As some word processors (e.g., Word and FrameMaker) allow search for examples of specific styles, even search within documents may suffer.

4.4 Structural Problems

SGML Document Type Definitions normally impose restrictions on where different elements may appear in a document. Examples of such restrictions may be that 'Author' should come after `Title,' that a heading should not be more than one level below the previous heading, or that all letters should have a 'Title.' None of these restrictions apply with the use of styles. Simple-minded translation of style examples

to SGML elements may therefore result in text which cannot be parsed according to the DTD. Clever conversion programs may fix several such "errors," but practical experience with automatic conversion shows that some "errors" always tend to slip through.

Initially we defined this as structural problems. We thought this was an important category, since it addresses a problem which arises when text is to be *converted* to some DTD, say HTML. In the set of documents we analysed, we found no clear examples of structural problems. Still, when documents are to serve as input to information systems, the structure of the documents must conform to what we expect.

4.5 Summary of Findings

Our main findings are (1) that there are great variations in the use of paragraph styles, (2) that generally paragraph styles are seldom used, and (3) that there are several problems associated with the use of paragraph styles. The problems with the use of styles, and their consequences (affected benefits), are summarised in table 1.

5. Discussion

In discussing our findings we will first go back to our question regarding approaches for electronic publishing and capturing input to SGML-based systems. Thereafter we will analyse our cases and findings with respect to organisational implementation of word processing. Finally we will try to establish a connection between the problems we have observed and the dominating paradigm of modern WYSIWYG word processors, which we here will call the paper metaphor.

5.1 Word processors as input to SGML

Compared to other ways of obtaining input to SGML-based systems there are obvious advantages with the approach of converting documents written in a word processor to SGML. Other approaches involve disadvantages as (Braa & Sandahl, 96):

- Setting up special units to manually tag plain text files. This implies no burden on text producers, but may require considerable effort in the new unit.

- Asking end users to work with special SGML-editors enforcing conformance with the DTD. Such editors look pretty much like other WYSIWYG-editors, but they are easily perceived as strict.

Our study shows that the apparently attractive alternative of taking SGML input from word processors may not be that attractive either. Text from word processors, *as written today*, is a very weak starting point for automatic conversion to SGML and thus capture of important data. For some DTDs, like HTML, it will always be possible to make a conversion which looks reasonable (converting paragraphs with big letters to <h5>...</h5>, etc.), but this is not a way to capture the structure of the documents.

The press releases of CC may serve as a good example, since they contain a series of "fields" of information which would be relevant to capture for further processing, indexing, search, etc. Examples of such fields are date, organisational unit, topic, and contact person. Today, press releases are made such that they *look* reasonably similar, but they are not standardised with respect to use of styles. Technically it is of course possible to attack this problem with AI-inspired techniques for heuristic classification (see, i.e., Clancey (85)). The documents could be analysed with respect

74

to words, formatting, etc. In such an approach it would most likely be possible to convert a large proportion of the press releases to a suitable DTD. Still, there would be uncertainty with respect to the correctness of the classification. It is not our purpose in this paper, however, to enter into a debate on the relevance of heuristic classification and other AI-based techniques. To us, it appears as attractive to let those who write a text declare that the title indeed is a title, and we are looking for ways of making this convenient.

Our findings imply that other approaches to capture text in SGML, after all *may* turn out to be more attractive than the conversion of word processor files. Such an inference cannot, however, be made in general. We have not tried to change practice with respect to use of paragraph styles in our case organisations. The efforts needed in order to change word processing practice in a more "SGML-friendly" fashion may vary. We discuss these efforts below. Our conclusion is therefore more modest: one cannot simply assume that comprehensive use of word processing represents a simple way to capture text in SGML.

5.2 Organisational implementation

Common to the three case organisations we have worked with is the lack of what we would call organisational implementation of the use of word processing and use of styles.

In AA there have been some efforts, but in the project where the software manual (AA-01) was written, there were no discussions about the specific template used for this document (it was not a template from AA). The person we interviewed had done her best in using styles as she saw they were used in the document, but some features of the template were unknown to her.

In BB only one template, the one for letters, had been developed and made available.

In CC there is no common policy on word processing. It is left to the different units, and the people in the Web-service do what they can to cope with what they get. The Web-service is an add-on to the organisation, and has no authority to instruct the units regarding their word processing practices.

Although some organisations claim to succeed with the implementation of word processing, it is clear from our cases and other examples we know of, that often too little effort is spent on developing templates and on training in how to access and use document templates. As a result, the interpersonal, long-term benefits of using word processing are not realised. Of course, it may always be difficult to realise such benefits, since they are inherently associated with Grudin's dilemma of "who does the job and who gets the benefit" (Grudin, 89), (Grudin, 94), but in our cases one has hardly attempted to pave the way for these benefits, much less made an effort to make potential benefits visible. Benefits visible to the users are effective, as demonstrated by the use of styles which support the construction of a table of contents.

To be effective, we believe the implementation of word processing should include the following tasks in order to prepare for electronic publishing and conversion of documents to SGML:

- Design, distribution, and maintenance of a comprehensive set of document templates covering most of the needs of those who write documents. Paragraph styles ought to be consistently named across different templates. Some word

processors are shipped with a set of templates which may serve as a good starting point.

■ Training in use of the templates. The training ought to include examples which illustrate the benefits of sticking to the templates, also to other people "down the line". The training should also address problems of cluttering up the style catalogues.

■ Making it beneficial to use styles, e.g. through provision of conversion tools between the templates and other useful formats. There should, for example, be documented ways to receive documents written in other word processors. There should also be ways of delivering documents to (and receiving them from), for example, external print shops. Ways to convert to HTML should of course be defined. Ideally there should be an SGML DTD for each template. The SGML extensions of some word processors may support this job.

■ Preparations for electronic archiving. Although not mentioned in this paper, there are issues related to archiving. DTDs corresponding to defined templates represent a mechanism for application independent storing of editable and searchable versions of the documents.

5.3 The paper metaphor

The use of paragraph styles does not change the looks of a printed document. It is only a property of the digital, unprinted document residing in the computer's storage. Currently WYSIWYG — what you see is what you get — and direct manipulation are the dominant principles for user interface design (Buxton, 93), especially for off the shelf software like word processing. Unfortunately, it appears that WYSIWYG communicates the idea that word processing is just another way of writing on paper. As an example Preece et al. (p 18) write:

> Alongside developments in interactive graphics, interactive text processing systems were also evolving at a rapid rate. Following in the footsteps of line and display editors was the development of systems that allowed users to create and edit documents that were represented fully on the screen. The underlying philosophy of these systems is captured by the term WYSIWYG. In other words, the documents were displayed on the screen exactly as they would look in printed form. This was in stark contrast to earlier document editors, where commands were embedded in the text and it was impossible to see what a document would look like without printing it.

Implicitly or not, Preece et al. place WYSIWYG and "how a document looks like in *printed* form" on an equal footing. A related opinion is expressed by Barker et al. (p.306):

> Although non-WYSIWYG word processors can produce a handsome final product, you won't see what you'll get until you get it. That's fine for people who like surprises. But if you work in a busy environment, where the deadline is always yesterday, you need an editor that gives you a clear idea of how your document will look before it comes off the printer. You need a product that does word styling as well as word processing.

The main problem with the idea that word processing is for processing printed paper - — the paper metaphor — is that invisible differences have no importance. This may be appropriate for documents that will only be viewed in one format and one medium. Documents which look exactly the same on paper may, however, have vastly different properties with respect to conversion, reformatting, search and exchange.

Warnock claims that until a few years ago the only use of word processors was to produce printed documents (Warnock, 92). Nowadays, however, many documents are published on paper as well as electronically (e.g., in WWW), and exchanging documents through email is also quite common. With these trends, the use of word processors has shifted from producing printed documents to producing documents which can be reproduced or communicated in several different ways[7].The paper metaphor does not keep up with these changes in use of word processing.

Users clearly focus on paper. The person working with the large manual (AA-01) used styles a lot, the motivation being a consistent layout, not a digital document. In BB, after a course in the use of styles, a user stated "I get the same results without [styles]," even asking back "Do you see the point in using them?" Users also refer to the use of simple icons in the toolbar to change the alignment of text. One user referred to the use of shortcuts: she knew all the standard formatting commands by heart. There were also cases where users were unaware of planned reuse of the text in WWW. When perceived as beneficial, i.e. for the construction of table of contents, some styles were indeed used.

Based on our findings and interviews we claim that the paper metaphor communicated by WYSIWYG is a part of the problem. Within the paper metaphor the use of styles makes little sense. Many of the problems reported from our empirical material can be explained by the users' focus on printed results. The paper metaphor provides no room for invisible structures in the document, while inviting non-style based editing of paragraph formats. Thus, we feel that the paper metaphor is an overly simple notion of what is going on. In the era of electronic documents people should not be misled to believe that they write on paper.

6. Conclusions

We have developed a classification of problems with paragraph styles. The classifications appears to cover the problems we have identified, but one category, "Overlooking styles," dominates quantitatively.

In a sample of documents, paragraph styles appear to be little used. Even in documents where use of paragraph styles would make sense (for example, because of planned publishing in WWW), there is little use. These practices will make it difficult to benefit from the opportunities of digital documents, standard exchange formats, etc. One cannot realistically assume that current practices in use of word processing provide a good basis for electronic publishing.

In our cases, little effort has so far been made to encourage or support the use of paragraph styles. We have provided some recommendations about how this could be done, but further studies are needed to interpret and understand differences in style use in different organisations.

---

[7] For a discussion of the difficulties in writing text for multiple kinds of presentation, see (Hendry, 95).

On a general level the "paper-metaphor" of current WYSIWYG word processors can be used to explain the findings, but there are also problems connected to the way styles have been implemented in some word processors. Comparative evaluations and studies of use of different word processors under comparable conditions could shed light on these issues.

Acknowledgements

References

Aho, Alfred V. and Ullman, Jeffrey D (1972). *The Theory of Parsing, Translation, and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, 1972.

Barker, D., Edvards, David L. and Wszola, Stan (1992). Writing in style. *BYTE*, 17(6):306--315, June 1992.

Berners-Lee, Tim, Cailliau, Robert, Luotonen, Ari, Nielsen Henrik Frystyk, and Secret, Arthur (1994). The World Wide Web. *Communications of the ACM*, 37(8):76--82, August 1994.

Braa, Kristin and Sandahl, Tone (1996). Standardization and flexibility in the distibution and exchange of documents. *International Working Conference on Integration of Enterprise Information and Processes ``Rethinking Documents'' (IPIC'96)*, Boston, November 1996.

Bryan, Martin (1994). Document markup for open information exchange. In *IEE Colloquium on "Adding Value to Documents with Markup Languages"*, London, 6 June 1994. The Institution of Electrical Engineers. IEE colloquium digest no. 1994/142.

Buxton, Bill (1993). HCI and the inadequacies of direct manipulation systems. *SIGCHI Bulletin*, 25(1):21--22, January 1993.

Clancey, William J (1985). Heuristic classification. *Artificial Intelligence*, 27:289--350, 1985.

Denning, Peter J. and Rous, Bernard (1995). The ACM electronic publishing plan. *Communications of the ACM*, 38(4):97--103, April 1995.

Flynn, Peter (1995). *The World Wide Web handbook*. Thomson Computer Press, London, 1995.

Goldfarb, Charles F. and Rubinsky,Yuri (1990). *The SGML handbook*. Clarendon Press, 1990.

Grudin, Jonathan. Why groupware applications fail: Problems in design and evaluation (1989). *Office: Technology and People*, 4(3):245--264, June 1989.

Grudin, Jonathan (1994). Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM*, 37(1):92--105, January 1994.

Hendry, D. G. Breakdowns in writing intentions when simultaneously deploying SGML-marked texts in hard copy and electronic copy (1995). *Behaviour and Information Technology*, 14(2):80--92, March--April 1995.

Herwijnen, Eric van (1990). *Practical SGML*. Kluwer, Dordrecht, 1990.

Jenssen, Astrid E. and Sandahl, Tone (1996). Conflicts between the possibilities and the reality in the field of structured electronic documents: Experiences from a large-scale SGML -project. In Bo Dahlbom et al., editors*, Proceedings of the 19th Information systems Research seminar In Scandinavia*, pages 935--955, Løkeberg (Gøteborg), 10--13 August 1996.

Johnson, Jeff and Beach, Richard J (1988). Styles in document editing systems. *Computer*, 21(1), January 1988.

Knuth, Donald E (1984). *The TeX book*. Addison-Wesley, Reading, Mass., 1984.

Lamport, Leslie (1986). *LaTeX: A Document Preparation System*. Addison-Wesley, Reading, Mass., 1986.

Miastkowsky, Stan (1993). Beyond word processing. *BYTE*, 18(5):85--86, 88, 90, Spring (Special issue) 1993.

Preece, Jenny et al. (1994). *Human-Computer Interaction*. Addison-Wesley, Wokingham, England, 1994.

Trigg, Randall and Bødker, Susanne (1994). From implementation to design: Tailoring and the emergence of systematization in CSCW. In Richard Furuta and Christine Neuwirth, editors, *ACM 1994 Conference on Computer Supported Cooperative Work — CSCW'94*, pages 45--54, Chapel Hill, October 22--26, 1994. ACM Order Number 612940.

Warnock, John E. (1992). The new age of documents. *BYTE*, 17(6):257--260, June 1992.

Approaches to Standardization of Documents

In Information and process integration in enterprises: Rethinking documents, Kluwer Academic Publishers

Kristin Braa and Tone Irene Sandahl

Abstract

This paper identifies and compares different approaches to standardization of documents in order to enable electronic document exchange based on SGML. Case studies applying the different approaches were conducted. The dilemma between standardization and flexibility is emphasized. An analytical framework has been developed as a means to reveal and categorize limitations, problems and new possibilities associated with the different standardization approaches.

1. Introduction

This paper focuses on the design and implementation of electronic documents in organizations. Although at least 80 % of corporate electronic information in organizations is in the form of documents, as opposed to database records (Reinhardt, 94), current systems have little access to this information. Traditionally, documents have been static, represented as files on disks. Until PCs were networked, these files usually belonged to only one user and passed from one person to another in printed form. Along with the adoption and diffusion of Internet communication technology, there is a growing interest in exchanging and distributing documents through data networks. Document distribution is necessary not only as a means of exchanging information between people but also as a means to integrate systems.

To enable exchange and distribution of documents through networks, some standardization is necessary (Hanseth et al., 96). In the case of the World Wide Web (WWW), the text is marked up in HTML (de facto standard). Thus, all WWW browsers[8] can read and present text (documents) on the Internet in a suitable way, whichever hardware platform is used.

This paper evaluates and compares different approaches to standardization of documents in order to enable electronic document exchange based on SGML. These are approaches to building the infrastructure for document exchange within organizations, and approaches to the division of labour for accomplishing the process. Using SGML is not the only way to achieve standardization. RTF, PDF, and ODA are other examples. However, the case studies presented are concerned with document exchange systems based on SGML.

We have identified three main approaches to standardization: One approach, **"soft" standardization**, is to let the producers of a text write freely and then hire people to go through the text and mark it up. Another approach, which we call **guided standardization**, is to guide text producers to use styles in a word-processor and

---

[8] Vendors are making their own version of HTML, this may over time change the compability between the browsers

then convert the documents to SGML. With the third approach, **enforced standardization**, the producers of a text produce the SGML-coded text themselves.

The aim is to study standardization processes, based on different approaches. This paper is based on a comparison of examples of the different approaches, studied in three different organizational settings. An analytical framework is developed as a means to reveal and categorize limitations, problems and new possibilities associated with the different standardization approaches.

The paper is organized as follows: First, the research approach and the theoretical framework are outlined, followed by a brief presentation of SGML. The application of the framework in the organization is then considered. There is a section discussing our findings, which are then summarized in the conclusion.

2. Research Approach

In this section we will state the background for the selection of cases, and show how we came up with the theoretical framework for analyzing the approaches to electronic document exchange systems found in the cases.

2.1 Empirical Work

Comparative case studies were conducted in organizations where different approaches to standardization of document exchange were applied, with special emphasis on the organizational setting and application domain. The selection of cases reflects the ongoing discussion within the SGML community about different approaches to converting non-SGML documents to SGML. Van Herwijnen (90) presents five different approaches: i) Tagging documents with a non-SGML editor, ii) using editors that give formatting feedback (e.g. LEXX and DECwrite), iii) adding the tags with a program, iv) imposing structure with style sheets, and v) using native SGML input systems. Organizations that apply at least one of the approaches above were selected. We were, however, not able to find organizations in Norway applying approach i) or ii). The selected cases are: Case 1, Organization of Web service; Case 2, Production and exchange of administrative handbooks; and Case 3, Production of a university catalog, representing approaches iii), iv) and v), respectively. However, we reframed and renamed the approaches (see section 2.2). The theoretical framework presented in the following section was applied as an analytical tool in the comparative studies.

The chain of work tasks in the standardization process involves people in several roles from different parts of the organization. Qualitative interviews were conducted including all those involved in the document exchange process - text producers, text providers, editors, converters, managers, and technical staff - representing use perspectives, organizational perspectives and technical perspectives. In addition, electronic documents that were written by text producers were analyzed with respect to the use of styles (Van Herwijnen´s approach iv). Technical reports documenting the process of standardization were studied, and emails sent to internal distribution lists, reflecting text producers' experiences with the standardization process, were analyzed.

2.2 Approaches to standardization of documents

There are several ways to standardize documents for electronic distribution and exchange (e.g. Van Herwijnen, 90). These approaches are, however, viewed from a

technical perspective that does not include the use situation or the organizational context. We found that a use perspective on standardization approaches was more applicable and richer, since Van Herwijnen's technical approaches imply preconditions on use practices that are not dealt with. We have, more generally, characterized approaches to standardized document exchange systems as:

1) **"Soft" standardization**: the producer may produce text as before, but the text will be standardized/structured by others or partly converted by programs (related to iii).

2) **Guided standardization**; quasi-standardization where the producer has to follow predefined style sheets and where other people or programs are used to convert the text to suitable code. In terms of SGML, the producer deals with styles to "mark" text elements instead of using tags (related to iv).

3) **Enforced standardization**; the producer of text must work according to the standard. In terms of SGML the producer must mark up the text with use of tags (related to v).

Within the **"soft" standardization** approach the producers of a text use traditional word processing products, like Microsoft Word, AmiPro or WordPerfect. There are no rules or standards that must be applied in the text processing process. Authors usually format the document by choosing the font, font size, margins, and so on, from a toolbar. They are not required to define the different text elements (title, heading, paragraph), nor to define the document structure (title before heading, and heading before paragraph). With respect to understanding the content and the structure in a document, this approach satisfies human user needs, but for the computer it is more or less incomprehensible. Programs may do some conversion, but manual structuring is necessary to standardize such documents in SGML. In this case, the text producers (with their word processor) do not create documents that are "ready" for the SGML-based document system, and manual editing has to be added.

In the **guided standardization** approach, the producers of a text use style sheets in word processors like Microsoft Word, AmiPro and WordPerfect. In terms of SGML, the producer deals with styles to "mark" text elements instead of using tags. There are possibilities for representing structure with the use of style sheets, but the possibilities are restricted (Johnson & Beach, 88; Woods, 91). With the correct use of style sheets, a conversion to SGML can be done automatically using conversion programs (Van Herwijnen, 90).

With **enforced standardization** the producers of a text use an SGML editor, and encode the text in SGML directly. An SGML editor "reads" and "understands" the definition of the document type (DTD). The editor knows the content and the structure, and therefore helps the authors follow the "standardized structure" in the text processing process. The author marks up the text using tags. In SGML the marked title may look like this: <title>This is a title</title>. When one uses an SGML editor, the text is marked up directly. If the coding is done correctly, no further conversion is necessary.

The "soft" standardization approach focuses mainly on the use perspective, while the enforced standardization approach mainly views the standardization process from a technical perspective. The three different approaches to producing SGML-encoded documents are viewed as ideal types, meaning that in reality one may find that different approaches are applied at different stages of the process.

2. 3 Quality perspectives

To be able to evaluate and compare different approaches to standardization in an organizational context, we apply perspectives in Braa's (95) IS quality framework. The IS quality framework is an analytical tool for understanding and evaluating information systems in an organizational context. The perspectives are the use quality perspective, organizational quality perspective, and technical quality perspective. These perspectives provide a useful contribution to evaluating document exchange systems as well, since in the end designing quality systems is what designing computer-based document exchange systems is all about.

Use quality is seen from a subjective perspective: the end users' actual experience of using the system. Different users and user groups will have different experiences and ideas about what good quality is. In applying a use perspective, the consequences of using SGML for different kind of users (authors, editors, readers etc.) are emphasized.

Information systems are operative in an organizational context and social practice, so that organizational perspectives on quality are also relevant for judging quality.

Quality seen from a technical perspective refers to the software system's structure and performance. The technical quality of a software system is the basis of its functionality (the computer must perform according to expected operations), thus often regarded as the most important.

Braa (95) concludes that neither of the perspectives is sufficient when the aim is to evaluate and understand the quality of information systems in an organizational context. The relationships between the perspectives are closely intertwined, e.g. a robust technical artifact is needed if the system is to function effectively and the artifact is used in an organizational setting, for an organizational purpose. If an IS is not used, it does not have any qualities. How to evaluate the use quality will depend on the perspective of the user or group from which it is regarded.

The approaches to electronic exchange systems, "soft" standardization, guided standardization and enforced standardization, combined with the IS quality perspectives from Braa (95), constitute the collective framework for analyzing and comparing the different cases.

2.4 Scope of the research

The study focuses on different approaches to standardization of documents based on SGML. The aim is to study the standardization process, not to do an evaluation of SGML itself, nor to discuss different ways of processing the SGML documents into further applications.

The choice of SGML is based on the fact that it is an international standard (ISO 8879) for describing documents independently of specific processing systems, operating systems, and document models, designed for data longevity and portability, and the fact that there is a fast growing interest in SGML within the private as well as the public sector in Norway.

3. Standard Generalized Markup Language

There are at least two different ways to achieve document interchangeability between systems: standardization of applications, so that the applications can work on each

other's documents, or standardization of the document itself, so that it can be processed by any application. SGML supports the latter solution.

SGML is used to describe the structure and the content of a document (descriptive markup), not its appearance (procedural markup). A descriptive markup system uses markup codes, which provide names to categorize parts of a document. Markup codes such as <descrip> identify a portion of a document and assert of it that "the following item is a description". All the text is coded as plain text. SGML thus enables the interchange of text across platforms, because there is no need for "translation" to any machine-dependencies. The same document can readily be processed by many different pieces of software, each of which can apply different processing instructions to those parts of it that are considered relevant. In addition, different sorts of processing instructions can be associated with the same parts of the file. Since only the structure and/or content of a document is marked, a given viewer of that document can decide what the "look" will be. The markup of the document never changes; only the way that it is interpreted does.

SGML supports the notion of a document type, and hence a document type definition (DTD). An SGML document always has an associated DTD that specifies the rules of the model of the document; for example, a DTD might specify that the document must have a description, and cannot have any warnings that are not immediately followed by one or more specifications. Its parts (description, warning, specification, etc.) and their structure in the DTD formally define the type of a document.

A basic design goal of SGML is to ensure that documents encoded according to its provisions should be portable from one hardware and software environment to another without loss of information.

4. Application of the framework in the cases

In this section we describe the approaches applied in three different organizational settings. We found that each case mainly applies to one approach. Hence, the approaches are described by our findings from the cases. The description is viewed from the three perspectives: use, organizational, and technical perspective.

4.1 "Soft" standardization

Case 1: **Organization of Web service.** The organization is a major public bureaucracy, consisting of several large and independent departments. A common externally oriented Web service was launched in 1995, where governmental information, press releases and publications are published. The aim is to inform the public, thus the service is considered politically important.

A newly established unit herein referred to as the technical editorial board operates the Web service. The technical editors receive documents for publishing via email, on floppy disks, and sometimes on paper. The documents come from a variety of sources. Most employees in the organization use PCs, using AmiPro, Microsoft Word or WordPerfect for word processing. Within the departments individuals decide which documents will be published on the Web service. Documents are normally received directly from the authors, but large publications such as governmental committee reports are often obtained from print shops. These latter documents are already coded in SGML when the technical editors receive them.

The organizational strategy seems to minimize changes in work in the text producers' use situation. The "soft" approach can be explained by the role of the technical editorial board of the Web service. The technical editors have no authority to instruct authors, and make documents available on an "as is" basis, taking whatever steps may be needed (in some cases retyping the text). The staff of the technical editorial board has to manually go through the documents, identifying and understanding the structure. Further, they must either add styles for subsequent conversion to HTML[9] or code the document directly in HTML in order to get the document into a standardized format. Often hybrid documents are received, where some of the styles are simulated in WYSIWYG editors. New tools for conversion are brought into use as they become available. Thus experimenting with tools has been undertaken in order to build competence.

4.2 Guided standardization

Case 2: **Production and exchange of administrative handbooks.** In the spring of 1994, it was decided to collect information about existing and related procedures of formal work routines applied in administrative work in a large educational institution. The information was collected into two books: the regulations handbook and the administration handbook. The first step was to make printed documents, standardized by means of styles in Microsoft Word. The next step was to convert the books into SGML and HTML for presentation of the electronic books in the Web viewer. The styles were designed based on negotiation in a working group.

Working with style sheets was perceived as time-consuming by the text producers. They were unfamiliar with the use of styles, and spent much more time than expected in applying styles. The analysis of the styled documents showed that they did not use the style sheets completely. About 25% of the styles in the style sheets were not applied. Instead they used, for example, the style "normal" even if it was logically a heading or a quote. Consequently, manual structural proofreading was necessary before automatic conversion could be accomplished. Thus, the technical situation became more like that in the "soft" standardization approach, with a need for recognition of the structure and adding styles by the technical staff.

A training manual was written, in which the styles were presented and explained. All the text producers received this manual, but very few used it. No training was provided. The result was that the text produced was inconsistent and difficult to use as standardized documents, and was no more standardized than in the situation of "soft" standardization; further use is dependent on manual reconstruction. The results from our study of the use of style sheets show that text producers, as well as technical editors, spent more effort on creating a handbook than before. The assumptions about automatic conversion failed, because the text producers did not use styles consistently. It was necessary to establish support personnel for completing the styles in documents before conversion, to make up for the lack of consistent use.

4.3 Enforced standardization

Case 3: **The production of a university catalog[10]**. Case 3 is a pilot project developing a solution for the production, exchange and distribution of a university's course catalog by means of SGML. The catalog was previously only published on paper. There are about 40 text producers, distributed among about 20 units, involved

---

[9]According to SGML terminology, HTML is a DTD.
[10] A more detailed presentation of this case can be seen in (Jenssen & Sandahl, 96)

in maintaining the catalog. The text producers use an SGML editor. The SGML editor incorporates a validating parser that makes it possible to avoid markup errors and invalid structure during editing.

Responsibility for the production of different parts of the catalog is divided between different units at the university. The text producers in the central administration maintain information about all parts except the information from the faculties. They coordinate contributions from other units at the university that will be presented in the catalog. At a lower level, each unit (department) is responsible for its part of the catalog, collecting information and distributing updated information to the students and staff.

The technical staff developed the DTD, which represents the model of the structure and the content of the catalog, with some participation from text producers. The intention was to make the DTD rich enough to be able to retrieve information from databases, to link to other information and to have functions for presenting different views of (some of) the information in the catalog. Programs have been developed for conversion both to HTML, so that the catalog can be presented on the Web, and to PostScript through TeX/LaTeX, for presentation of the catalog on paper. When the information is ready in SGML, the technical staff merges the files to produce the catalog. The catalog as a whole is parsed, and conversion to Web and paper can be done whenever needed.

Users found work with the document exchange system hard and frustrating, especially in the beginning. Use of SGML requires discipline as to the way text is written. Structuring the information according to a DTD creates limitations on work styles. Usually, people can present information their own way using their preferred tools. Dealing with SGML, they felt that flexibility was limited. Others pointed out that work with the catalog had changed. The catalog is still the product; however, using SGML has changed the process. They spent more time writing and editing the catalog than before, because of the new tool and the structuring. Because of the time-consuming process, some got fewer work tasks, and ended up as "experts" on the catalog. In other words, introduction of SGML created specialized work tasks, more like typing pools.

From a more organization-oriented perspective, the introduction of SGML required much effort in training and support. Even with planned training, the text producers needed to have access to some form of constantly available user support. They needed help to solve technical problems, to figure out what to do with the different parts of the information, and how to code and where to put the markup. User support created a heavy workload for the technical staff.

As a benefit of introducing SGML, the organization wanted to offer some new services to customers, such as publication on the Web, integration between the catalog and relevant databases, and providing facilities for a matrix on room allocation. From a technical perspective, this is easy to implement because the capabilities are embedded in the structure of the information itself. However, one year after implementation the database integration and the room allocation service were still not implemented.

There has been a change in administrative routines and distribution of responsibility since the SGML implementation. Before the introduction of SGML, the central administration unit was responsible for the production process - deadlines, proofreading after the deadline, new kinds of information in the catalog, etc. Some

time after the first production of a computer-based catalog, the technical department assumed responsibility for the production process, without any clear agreement.

From a technical perspective, the document exchange system imports and processes the SGML-encoded document automatically using predefined programs. However, the technical staff detects numerous syntax inconsistencies during the conversion process. Errors in the documents are possible because when users experience too many failure messages on syntax, they switch the validation checker off. The technical staff then has to manually check and mark up the document.

5. Discussion

The three different cases described above mainly represent three different approaches to standardization of electronic documents. The approaches to standardization are presented in relation to three perspectives: use, organizational, and technical perspectives. This constitutes the framework applied in the discussion below. An outline of the results is presented in Table 1.

Table 1: Summary of the results, applying the framework to the cases.

	"Soft" Standardization	Guided standardization	Enforced standardization
Use perspectives	• Little change in work content • Little change in work organization • Little change in roles	• Some structure in work • Requires common understanding of styles • "Workaround"	• High structure in work • Change in work content • Change in work organization • New roles emerge • "Workaround"
Organizational perspectives	• Allocation of parallel organizational unit • Allocation of resources to technical staff	• Common agreement on styles • New roles emerge	• Allocation of - training support - user support • common agreement on DTD
Technical perspectives	• Manual conversion • Reconstructing the structure of document • Low compatibility • "Workaround" • Need for creative conversion tools	• Medium automatic conversion • Medium compatibility • Need for tools (text processors) facilitating styles	• Automatic conversion • High compatibility • Need for development of advanced editors

5.1 Use perspectives

From a use perspective, the "soft" standardization approach differs significantly from the other two approaches. In the "soft" approach the text producers have little, if any, change in their work. They use their word processors the same way as before the introduction of standardization. They format their text manually using a toolbar within the word processor, and the paper copy of the text is seen as the final product. The work organization of the text producers is as before. They do not deal with the standardization directly. There are, however, some restrictions on the layout in some of the documents they produce. The organization wants to have a common profile on some of its information. In this respect, the text producers "fake" the profile using manual formatting. The appearance of the text on paper complies with the profile, but the electronic versions of the text are quite different, because of the use of spaces, tabs, fonts and font sizes. The text producers do find themselves in a flexible work

situation when it comes to production of text, and they do not experience any limitation on the process.

Within the guided approach, the text producers have to use predefined styles. Observations in the second case study show that text producers have problems using styles correctly and consistently. They are not familiar with the use of styles and think of them as "difficult to use". Some text producers stated that the style mechanisms have an unsatisfactory design, and do not necessarily support the idea of structuring. When they did not know how to use a style, or which style to use, they just dropped it and used "normal" instead and then simulated the layout. Further, they might use the style "quote" on an ordinary paragraph, because they knew that the "quote style" was formatted as italic text. Gasser (86) defines "workaround" as the development of new ways of using computer systems, which were not intended by the original design. Working around the predefined style sheets by "faking" the styles is easily done, and many text producers did. Thus, the situation of text processing was not experienced as very standardized. It seems that motivation for using styles as a part of the standardization process was lacking. An emphasis in user training on understanding how use of styles is part of the standardization process, and why consistent use is crucial for achieving that, might help.

In the enforced standardization approach, both the work routines and tasks were changed because of SGML and the SGML tools. The text producers' work situation was structured by the predefined DTD, and was experienced as restricted with respect to the order of work tasks, as well as the work itself. However, when they experienced the structure as too annoying they switched the validation checker off. This "workaround" occurs and creates a breakdown in the automatic conversion routines.

From the use perspective, the "soft" approach is the most flexible, and few changes are observed in work situations. The standardization process will not affect the text producers. In the guided, and even more in the enforced approach, the users feel that the work is strongly affected by the standard.

5.2 Organizational perspectives

From an organizational point of view, the three approaches have different effects. Within the "soft" approach there was a need for establishing an additional unit within the existing organization - just for converting the documents into formats that are more appropriate in relation to standardization. Since the text producers are left out of the standardization itself, others had to do the job. In case 1, the staff developed competence by experimenting with the conversion tools. This resulted in delays in publications on the Web.

In the cases of "soft" standardization and guided standardization the technical staff is part of the work chain of standardization, and will easily be seen as a bottleneck between the text producers and the final product. The text has to go through a standardization process in another part of the organization before it is ready for further application.

Within the guided standardization approach the organizational effort is on at least three different levels. First of all, it is necessary to negotiate common style sheets (templates) for different document types. These style sheets are standardized for the organization. Second, the text producers have to learn to use the style sheets, and how to style documents. As observed in case 2, production of administrative university manuals, and also reported in Sørgaard and Sandahl (97), this is not necessarily a

trivial task. The assumptions about automatic conversions failed, because the text producers did not use styles consistently. Third, a parallel organizational unit was established and given resources to make up for the lack of standardized documents received from the users. Our experience is that an understanding of the rationale for using styles consistently is necessary; the role of styles in the standardization process should be emphasized.

Within the enforced standardization approach, the workload is placed on the shoulders of the text producers. They have to work within a structured environment enforcing standardized work tasks, and to learn the principles of SGML as well as a new tool. There has to be an agreement on a DTD, which is an organizational challenge (Jenssen & Sandahl, 96). Modeling the DTD in a way that supports the work processes. not only modeling the document as a product, needs to be an organizational effort. Much user training and user support is needed, and the organization has to allocate resources to do this. In addition, SGML expertise is necessary, which takes time to achieve. Technical competence is important in this approach.

Table 2: "Workaround" creates breakdowns and leads to a technical situation as in "soft" standardization.

	Soft Standardization	Guided standardization	Enforced standardization
Use perspectives		Workaround ("faking" the style use)	Workaround (switching off the checker)
Organizational perspectives			
Technical perspectives	- Manual standardization - Reconstructing the structure of document - Low comparability - Need for creative conversion tools		

In the case of the catalog production we observed a change in responsibility for the production process, which was transferred from the department (or faculty) level before SGML implementation to the technical department coordinating the whole production process. As far as we know this was not due to a conscious decision.

From a technical perspective we see that the potential for document exchange systems can be achieved at the organizational level, but this has so far not been implemented. Managers have little involvement in the implementation of document exchange systems. This is often regarded mainly as a technical task. Organizational perspectives need to be addressed explicitly; this implies challenges for further research.

5.3 Technical perspectives

From a technical viewpoint, the "soft" approach necessitates a great deal of manual conversion. Software tools that identify layout as structure can do some conversion; otherwise people have to standardize the text manually by using styles or directly into SGML. Since the text has a restricted structure, and no defined content, the text has to be treated before further use is possible.

The guided approach could provide for automatic conversion, if the styles were used in a correct and consistent manner. Since the use of styles in a way defines the structure and to some extent the content of the document, conversion to other formats is easier to provide (Sørgaard & Sandahl, 97).

In the enforced approach, the technical condition for further processing is directly present. This, however, requires correct and consistent code.

"Workarounds" in the text production create breakdowns of the standardization process in both guided and enforced standardization, and must be compensated for by manual conversion efforts like those in "soft" standardization (Table 2). Observations in the cases reported on in this paper show that breakdowns occur easily, lead to bottlenecks in the standardization process and cause delayed or unpublished documents. To avoid breakdowns and gain the benefits of standardization, adequate support is needed, with regard to both better tools and training.

5.4 Standardization and Flexibility

The application of the three perspectives shows that there are differences in who gets the benefits and who performs the work, as discussed in Grudin (94), depending on which approach is applied. Working according to standards will doubtless limit the flexibility of work styles but may create new flexibility in terms of new services or products. The relationship between standardization on the one hand and flexibility on the other is illustrated in Table 3.

Table 3: Standardization and Flexibility

	"Soft" Standardization	Guided standardization	Enforced standardization
Use perspectives	• high degree of flexibility in use • low degree of standardization of work routines	• some flexibility in use • some standardization in work routines	• low degree of flexibility in use • high degree of standardization in work routines
Organizational perspectives	• low degree of flexibility regarding new applications • low degree of control of product / work	• some flexibility regarding new applications • some control over products	• high degree of flexibility regarding new possibilities of applications • high control over products and work routines
Technical perspectives	• little degree of flexibility regarding (developing) new technical solutions • low degree of standardization of conversion routines	• some flexibility regarding (developing) new technical solutions • some standardization of conversion routines	• high degree of flexibility regarding (developing) new technical solutions • high degree of standardization of conversion routines

The relationship between standardization and flexibility appears different depending on which perspective and which approach is applied. In the "soft" standardization approach the use perspective is considered dominant and flexibility in the use situation is very much present. The conversion routines in the same approach provide a low degree of standardization, thus implying manual reconstructing of documents and manual conversion.

In the enforced standardization approach, the technical perspective is dominant and flexibility with respect to developing new technical solutions will be supported, while the work processes are conceived of as highly standardized. The DTDs enforce a high level of detail for encoding the information. To support the design of new functionality and provision of new services based on information in the document, the DTD has to be well defined. On the other hand, if the DTD is very structured, more user problems will occur. Technically, it is easier to go from a highly structured DTD towards a less structured one, than to add structure. Consequently, the future functionality and services must be envisioned from the very beginning of the DTD modeling. This represents a challenge in DTD modeling.

The DTD modeling is seen as a model of the document, thus a technical matter, not as a model of the work with documents in an organizational context. A DTD that seems appropriate as a technical solution may not be appropriate for the text producers. IT departments are the driving force in DTD-modeling, and their understanding of the work in organizations is limited. In the end, the standardization effort done in the organization might not give the expected flexibility, because the DTD is not satisfactorily designed and can be difficult to change.

The guided standardization approach is a hybrid attempting to provide some flexibility in use situations and some standardization in conversion routines. However, our observations in the cases discussed in the present paper show that consistent use is critical in order to avoid breakdowns in conversion routines, which create a technical situation similar to the "soft" approach (Table 2).

In the organizational perspective, the tension between standardization and flexibility appears as the trade-off between the costs of dealing with standardization (e.g. user support, training) and the benefits of flexibility with respect to new services and

products. To create flexibility from an organizational point of view, such as providing new services and products, standardization is required.

5.5 The relationship between the approaches

The boundary between two of the categories, "soft" standardization and guided standardization, is blurred in the case of the organization of a Web service. As an initial step, the "soft" standardization approach was chosen. The costs were that a heavy workload was imposed on the technical editorial board (reducing the number of documents being published). After a new conversion tool became available, in which the technical staff saw the potential of using styles, they subsequently begun to emphasize the use of templates (style sheets). Products and tools change fast in the early stage of Web technology, and, as observed, they affect work practices. However, use and availability of document templates or style sheets vary widely between the units. We saw several examples of difficulties in applying styles consistently; consequently, manual conversion work was still needed.

The organization's longer-term aim turns out to be to proceed with an enforced standardization approach. The plan is to install SGML editors at the front desks in the departments, and have the coding done by the receptionists in addition to their customary work tasks. At this stage this has not been implemented. Time will show whether this is a realistic strategy or not. However, it seems that different approaches were applied step by step due to increasing maturity in use, motivation, technology and competence. The strategy seems to be to start small with "soft" standardization in order to gain motivation, knowledge and experience, and then to proceed with the next approach.

Going from the "soft" and guided standardization approaches to the enforced standardization approach involves changing the text producer's tools. In the "soft" and the guided standardization approach, well-known WYSIWYG editors are usually applied. SGML editors are not necessarily WYSIWYG and are more loosely connected to the end product. Traditionally, with the use of WYSIWYG editors, the text producers control the whole process: from editing through manipulation, formatting and presenting the documents. When an SGML editor is applied, this is not necessarily the case. In addition, the DTD sets restrictions on how to edit, because of the predefined structure, and to some extent, content. Thus, the transformation process from word processors to SGML editors can be hard for the text producers.

6 Conclusion and further research

The aim of this paper has been to discuss different approaches to the standardization of documents, and to categorize limitations, problems and possibilities within the different approaches according to use perspectives, organizational perspectives and technical perspectives. The framework has been applied in three organizations to analyze the experience and to present the results. Although the borders between the identified approaches, "soft" standardization, guided standardization and enforced standardization are blurred in some situations, the characteristics of the approaches proved to be useful. We found that the framework supports a more holistic evaluation of the standardization process and of the possibilities of achieving success, by explicitly regarding each perspective. By applying this framework, different, problematic, and possible defective aspects of the approaches may be revealed. An overview of the resources needed to improve the process may then be achieved.

Which approach to choose will depend on a variety of factors, i.e. staff competence, organizational effort, technical support and technical infrastructure - or it could be dealt with as an organizational choice: where to implement changes in the organization.

For each approach, we have found contrasts in who will get the workload and who will get the benefit. Due to the chain of work tasks involved in document exchange, it is not obvious that those doing the work will be those experiencing the benefit. This situation creates a motivation problem and is a major obstacle to the accomplishment of successful document exchange.

In the "soft" standardization approach the technical staff got the extra workload, due to the effort to convert the text into SGML, and the producers of the text got the benefit in the sense that no changes in work situations were added. In the enforced standardization approach the producers of the text had to apply a new authoring tool, and, to some degree, make use of SGML. Consequently, the text producers definitely experienced an extra workload, especially in the beginning. The technical staff benefited because there was little or no correction in the SGML files. However, they were busy providing user support and solving user problems. The user problems of working with the standard have to be solved in order to more easily achieve the text producers' acceptance of the document standardization. In the guided standardization approach, success depends on whether, or how, the styles are applied. If the style mechanisms are consistently used, it is proportionately easy to convert automatically to SGML. We have found that "workaround" in use situations in both guided and enforced standardization approaches imply a technical situation like that of "soft" standardization.

The issue of standardization and flexibility will be different depending on the perspective and the approach being applied. Flexibility in use is experienced within the "soft" standardization approach, while the technological solutions within the same approach are not flexible at all. While the work processes are conceived as highly standardized and difficult to apply, flexibility in technological solutions is supported within an "enforced standardization" approach, although not automatically adopted. Organizational effort and acceptance are necessary. Seen from the organizational perspective the tension between standardization and flexibility results in the costs of dealing with standardization (e.g., user support, training) while the benefits of flexibility appear as new services, products, etc.

Independent of the strategy applied, we have observed that the organizational potential of SGML has not been realized so far. The situation is rather that the documents are implemented as an electronic version of paper documents, and new functionality is seldom added. This can partly be explained by the fact that the work related to text production, user support, corrections and conversions is so time-consuming that the further development of new functionality has little priority. Another important factor is that organizational and managerial involvement and commitment are lacking. We have indicated the necessity of considering organizational needs as early as the DTD modeling stage. So far it seems that the organizations are only paying the costs of the standardization, and not yet obtaining the advantages. Since the potential is not realized, there may also be reason to ask questions about the usefulness of SGML itself. Challenges on how to take advantage of the potential of the new medium, as well as firm evaluation of SGML itself, need to be addressed in further research.

Time and product development are significant variables in this area of early adoption of electronic document technology, making it necessary to carry out longitudinal studies. A challenge for further research will be to develop tools and techniques to

reduce the bottlenecks; tools for supporting the use of styles; tools for "embedding" the standardization into the document processing; as well as techniques for including users in the standardization process - in addition to putting organizational needs on the agenda from the very beginning of the design of the document exchange systems.

There are, however, some problems in comparing the cases for various reasons: The products are different. They exist in different domains, and the document exchange systems have been introduced for different reasons. However, this study "paints a picture" of typical application domains and of a chaotic situation of immature technology as well as lack of competence and experience.

Acknowledgments

References

Alschuler, L. (1995) *ABCD ... SGML A users guide to structured information*, International Thompson Computer Press, London

Braa, K. (1995) *Beyond Formal Quality of Information Systems Design*, Ph.D Thesis, Department of Informatics, University of Oslo.

Checkland, P. & Scholes, J. (1990) *Soft Systems Methodology in Action,* John Wiley & Sons, New York.

Derose, S. J., Durand D. G. (1994) *Making Hypermedia Work. A User´s Guide to HyTime*. Kluwer Academic Publishers.

Gasser, L. (1986) The Integration of Computing and Routine Work, in *ACM Transactions on Office Information Systems*, Vol 4, No. 3, p 205–225.

Grudin, J. (1994) Eight challenges for developers, Communications of the ACM, Volume 37, Number 1, p. 93–105.

Goldfarb, C., F (1990). *The SGML Handbook*, Claredon Press, Oxford.

Gruber, T., R., Vemuri, Sunil, Rice, J. (1995) Virtual documents that explain How Things Work: Dynamically generated question-answering documents, at http://www-ksl.standford.edu/people/gruber/virtual

Hanseth, O., Monteiro, E. and Hatling, M. (1996) Developing information infrastructure standards: the tension between standardization and flexibility, *Science Technology and Human Values*, 1996 (To be published??).

Kim, K. K. (1989) User Satisfaction: A Synthesis of Three Different Perspectives, in *Journal of Information Systems.*

Kitchenham (1989) Software quality assurance, in *Microprocessors and Microcomputers,* Vol. 13 (6), pp 373-381.

Jenssen, A. E, Sandahl, T. I. (1996) Conflicts between the possibilities and the reality in the field of structured electronic documents. Experiences from a large-scale SGML-project. *In Proceedings of the 19th Information Systems Research Seminar in Scandinavia, 1996*

Johnson, J., and Beach, R. J. (1988). Styles in document editing systems. *Computer*, 21(1), January 1988.

Reinhardt, A. (1994) Managing the New Document, in *Byte*, Vol 19, No. 8, August 1994.

Smith, J., M. (1992). *SGML and related standards. Document description and processing language*. Ellis Horwood Limited.

Sørgaard, P, and Sandahl, T.I (1997) Problems with styles in word processing: a weak foundation for electronic publishing with SGML. In *30th Hawaii International Conference on System Sciences (HICSS),* Wailea, Hawaii, Jan 7-10, 1997

Travis, B. E. and Waldt, D. C. (1996) *The SGML implementation Guide*, Springer, 1996

van Herwijen, E. (1990) *Practical SGML*. Kluwer Academic Publishers, Dordrecht, Boston, London.

Woods, M., J. (1991) Techniques for textual mark-up using Microsoft Word. *University Computing*, 1991, 13, 171-179.

From Paperwork to Network - a field study

In Proceedings from the Third International Conference on the Design of Cooperative Systems, Cannes, France, May 26-29 1998

Kristin Braa and Tone Irene Sandahl

Abstract

The importance of understanding multiple roles and resources of documents in work practices for designing document systems is emphasized in this paper. These resources often vanish when documents become digital. The difficulties one faces in the transition from paperwork to digital work are discussed based on a case study. Prototypes are described to illustrate ways of capturing these resources.

Keywords: Document systems, paperwork, work practices, articulation work.

1. Introduction

Increasingly, companies are considering digital documents as the only way to successfully manage the flow of information throughout the enterprise (Berry & Goulde, 1994). The industry of workflow systems and document management is growing (Ibid.). Visionaries state that documents will no longer be an electronic analogue to paper, but rather dynamic, modular, multimedia entities (Reinhardt, 1994). However, studies show difficulties in exploiting the potential of electronic documents in providing new services (Braa & Sandahl, 1998b, Braa & Sørgaard, 1997, Eriksen, 1997).

The roles of documents in work practices are emphasized, to improve design of digital document systems. We regard documents not as isolated artifacts, but intertwined in a heterogeneous network of humans and non-humans. Brown and Duguid's (1994) concept of borderline issues is applied to identify border resources of documents developed through practice and conventions over time, and to use them as input for the design. We also adopt the analytical distinction between cooperative work and articulation work (Schmidt and Bannon, 1992, Schmidt and Simone 1996), where cooperative work consists of the interdependence of multiple actors who interact through changing the state of a common field of work. To deal with the source of confusion and disorder, individual and interdependent activities must be coordinated, scheduled, aligned, meshed, and integrated. The orderly accomplishment of cooperative work requires what has been termed articulation work (Strauss, 1985). In the empirical work presented in this paper, we have found that the paper documents, and other artifacts adapted to the paper medium as a part of the production process, play an essential role in coordinating work. This implies that the paper-based production process reduces the need for articulation work, and thereby the need for additional coordinating artifacts. We have seen that when documents become digital, these resources vanish and the need for articulation work increases. It is, however, possible to identify these properties and implement them into document systems. We describe paperwork at a New Agency (NA), and the process of transforming paperwork into a digital document system in an inter-organizational context. The attempt was a failure. One of the reasons is that the resources that the users rely on

to do their work were not regarded as central and were removed in design. The paperwork described is the production of TV schedules in TV channels, coordinated and merged in a news agency and sent to magazines and newspapers. We present an alternative prototype, which takes some of the identified border resources into account. The prototype has been implemented and adopted at the news agency.

The study can be characterized as an action case (Vidgen & Braa, 1997). An action case is a research method, which deals with the dilemma of intervention and interpretation in in-context research. The study has lasted for approximately two years. We have done qualitative studies of the actors involved, such as consultants and TV channels, as well as the Media Department and the IT department at the news agency (NA). These studies include interviews, observation and document analyses. We have also conducted small-scale interventions such as workshops, prototypes of alternative solutions, and training in order to evaluate our findings and hypothesis.

In section 2, related studies of documents in use are presented. In section 3, documents as unnoticed resources in the concept of borderline issue are described. The case is described in section 4, and the SGML-based system in section 5. The transition from "paperwork to network" is discussed in section 6, and some design challenges and prototype suggestions are presented in section 7. Finally, some concluding remarks are presented in section 8.

2. Related Studies of Documents in Use

Levy (1994) stresses that documents have both fixed and fluid properties, independent of the media they are based on. "Documents arise from the basic human need for stable, external, communicative resources. What is said in the hallways is transient and insubstantial, while what is written down is fixed and external and therefore potentially available to participants across time and space (Levy, 1988, pp. 187)".

Luff et al (1992) state in their study of paperwork in architectural and medical practice that paper documents have special properties suitable in synchronous and asynchronous collaboration. They emphasize the potential for tailoring paper documents and their ecological flexibility. People tailor their documents to differentiate and highlight particular items. Doctors may underline or mark text with color pens in records to alert colleagues to irregularities in treatments, and architects sketch in and ring changes to their plans. It is relatively easy to add information to documents that is essential for their application in work situations. The term "ecological flexibility" refers to the capability of documents to be adapted to a range of situations and contingencies. A doctor can examine a patient with an eye on the record at the same time. Doctors can compare records to evaluate different treatments when they brief other doctors.

Hughes and King (1993) have observed in their study that paperwork and technical knowledge about accountancy are integrated with the socially organized work. The documents have different roles in the work settings. First, the documents are objects that consist of relevant information, and are artifacts that carry information around in the organization and between organizations. Second, the documents are "displays" of the methods used to produce them. What readers could find in documents would depend on their interpretation of the displays. A reader has to know what to look for, which means knowing what the "object" represents and its implications, as well as how to read the "object" as displays of the procedures used to produce them in their organization.

Hughes and King (Ibid.) studied the invoices through the different work processes in the organization. They observed that production work leaves its mark on it, as comments, memos, queries and questions stuck to it. Since the sequence of stages and the activities are standardized, a glance at the invoice is enough to be able to tell what has happened, where things are up to, and what problems might be. The fact that the documents circulate as an element in the division of labor provides not only the means of achieving the various transformations of the materials, but also an indication of how the organization is performing.

Brown and Duguid (1995) state that documents are also a powerful resource for constructing and negotiating social space. They point out that the social rather than directly informational life of documents is prominent within a historical and geographical perspective. For instance, the roles of documents include linking people, political linkage, and a means of negotiation (Ibid.). Scientific journals link researchers who read the same papers, keeping each researcher conscious of being a member and aware of what others are doing. Newspapers may reflect culture and create identity, as well confirming or supporting it. Documents may provide a shared context for constructing meaning, and may be the beginning of the process of negotiations.

Documents are instruments of control and power within and between organizations (Yates, 1993). Yates (Ibid.) has found in her study that formal communication emerged as an important control mechanism in organization in the beginning of the nineteen century. Procedures, rules, and financial and organizational information were documented at all levels to enable upper management to monitor and evaluate processes and individuals at lower levels. A century has passed since the establishment of 'formal management'; however, the documents are even more important in management today.

Documents' roles depend on genres. Different kinds of documents may represent different genres, which relate to what people expect of the document. Expectations for prose are different from those for a law book; expectations for a formal letter are different from those for a memo. According to the argument of Yates & Orlikowski (1992), new communication genres in general evolve over time as a result of changes in existing genres based on new or different individual needs. The letter became a memo because of the need for less formalized written communication within organizations (Yates, 1993; Yates and Orlikowski, 1992). In other words, documents that are produced and in use exist because they are essential for people carrying out their work (Yates & Orlikowski, 1992).

3. Document's unnoticed Resources

In more general terms, we apply the concept of the borderline issue (Brown & Duguid, 1994) to discuss the importance of documents' additional or unnoticed resources, beyond what is usually recognized as the canonical meaning of the document. In general, artifacts have both *central* and more *peripheral* properties, and there is a *border* between them (Ibid.). What is recognized as a central or peripheral property varies within different communities of practice. The border is distinguishable from the other peripheral properties if it plays a socially recognized role in the artifact's use. For most of us, an invoice is a piece of paper that tells us how much we have to pay or how much others owe us. For us, the financial transaction itself is the central property of the invoice. As reported by Hughes and King (1993), accountants may regard the comments, memos, queries and questions attached to the invoice as the central property, in order to be able to tell what has happened, what the status is, and what problems have to be solved. Some users see properties as a central and natural part

of the documents, while other users of the same document regard the same properties as uninteresting and peripheral.

There are no clear boundaries indicating what is seen as a central part of the artifact and what is peripheral. How the artifact, or more specifically the document, is regarded depends on the community of practice. The border is defined as the dividing line between the central and the peripheral properties. Borderline issues are those resources that are shared and constitute a social meaning for a group of people. These unnoticed resources are developed over time as artifacts are integrated into ongoing practice and social conventions are developed. The border resources are maintained by communities of practices, and are resources that workers often rely on (Brown & Duguid, 1994). For instance, the lack of a mark in an invoice may be an unnoticed resource, indicating that a colleague is not at work, and that someone else is doing her job (Hughes & King, 1993).

The noise of a keyboard is another example, providing people with the useful information that a colleague is at work in the room next door. A central property of the keyboard is its capability of making text. A peripheral property is the noise it makes when a person uses it. The border, as an unnoticed resource, is the information that a colleague is at work in the room next door.

Borderline issues develop over time based on material continuity and communities of common practice. Material continuity is needed in order to recognize the properties, and community of common practice is necessary for members to share, recognize and reformulate conventions (Brown & Duguid, 1994).

Because the border resources often lie beyond what is regarded as part of the artifact, designers often remove these when developing new systems. We apply these concepts to identify and discuss resources of documents that are established through conventions in paperwork, and to analyze the design of a document system that replaces the paperwork, in order to investigate whether the border resources are replaced or disappear. Further, a prototype has been designed in order to capture some of these resources.

4. The Case

NA is a news agency providing news services to the press — newspapers, magazines, radio and television. The need for a new editorial system that supports electronic text exchange has been recognized in NA for several years. The existing system is based on old-fashioned mainframe technology. The focus on a new system intensified in 1993, when the International Press Telecommunication Council (IPTC) introduced the SGML-based "New Industry Text Format" (NITF) to increase electronic text interchange between news agencies and their customers / suppliers. IPTC recommended that all news agencies adopt the standard in their editorial and communication systems.

NA wanted to build competence within the organization on a minor project before continuing with an implementation for the whole organization. In 1995 the IT department, together with external consultants, took an initiative to apply SGML in the production of one of NA's services: the TV schedules. TV schedules produced for newspapers and magazines were regarded as the most structured and well-defined product they produced; it was thus assumed that they would be easy to standardize in SGML.

4.1 The production of TV schedules

The production of TV schedules involves a chain of work tasks. Starting with the production of program information in the individual TV channels, this chain involves merging, quality control and coordination in the Media Department in NA, ending in a variety of weekly products to be delivered to various newspapers and magazines. Currently NA is typesetting TV schedules of about 40 TV channels, and the number fluctuates as TV channels close down and start up. The documents' role in this infrastructure is further discussed in Braa and Sandahl (1998).

From the NA's point of view, there was an urgent need to facilitate the document exchange process between the TV channels and NA. Almost all schedules from TV channels arrive as faxes. The information is retyped, integrated with editorial material, and encoded into an old mainframe system, where the schedules are edited before being converted for import into a desktop program to make pages ready for printing, or transmitted as data to the newspapers and magazines. The electronic exchange of TV program schedules from TV channels to NA was expected to result in enormous time savings, because the people in NA would not need to retype the text. Instead, they could concentrate on improving the quality of the products to be sent to newspapers.

4.2 The community of practice: Media Department

The TV channels send their TV schedules to the Media Department in NA almost entirely by fax. Only a few use mail, and one channel has its own solution, called Data-TV. All the faxes are sorted by channel and placed on a shelf. There are three different types of shelves: the in-shelf, week-shelf and out-shelf. The in-shelf consists of space for each TV channel. When a schedule is received, the fax is placed on the shelf where it belongs.

The in-shelf gives an overview of the channels that have been received and those that are missing. A glance at the shelf reveals outstanding channels.

The TV channels do not necessarily send schedules for the whole week at once. However, the definition of a week is not uniform. What denotes a week varies among the different actors in the network. For most of the TV channels, a week starts on Monday and ends on Sunday. The Media Department operates with a week that starts on Saturday and ends on Friday. The week in the newspapers and magazines is dependent on which day TV guides are published. The TV guides consist of information about radio and TV programs for "all" channels for a week. The "different" concepts of a week force the Media Department to divide up the information from the TV channels and to define the week according to the recipient's wishes. When the TV schedules do not send the TV information for the whole week at once, a coordination form is used to register which days have been received and which have not.

When all the schedules from TV channels have been received and placed on the in-shelf (in which each channel has one shelf), all the faxes containing information about the specific week get moved to the right space on the week-shelf. However, in busy periods the faxes are placed directly into the week shelves. The week-shelf is not separated into channels, but only into weeks. For every week, regardless of when it starts and ends, information about all programs every day for each channel has to be retyped. To maintain an orderly system, the personnel sign the weekly coordination form when a particular TV channel is finished. When the coordination form is completed, the week is done. By then, 6 different people within the Media Department may have been involved. In coordinating the different definitions of weeks, the faxes, shelves and forms play important roles.
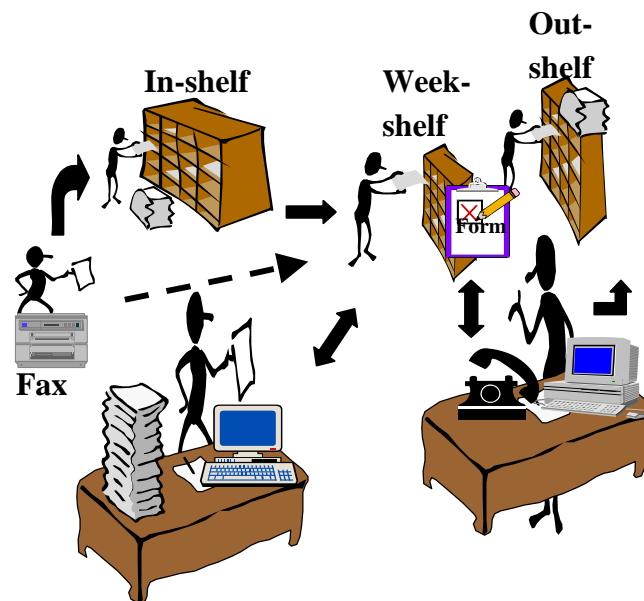
Figure 1: The work practice at the Media Department

When the week is finished, the faxes are moved to the out-shelf. If there is time, a controller takes the faxes out of the out-shelf and checks them against the text in the mainframe system. When this has done, the coordination form is signed in the correction cell. From the out-shelf, the information goes into the trash. When the faxes are in the out-shelf, the information is considered ready for preparing files in QuarkXpress that are ready for printing. The text converted into QuarkXpress has to be justified in order to fit into predefined spaces. From time to time, TV channels change their schedules and fax corrections the Media Department. The faxes are placed in a correction shelf or on somebody's desk. In this case, the coordination form is not in use.

The previous week is used as the basis for data entry in the mainframe system. There are always some recurring programs, so "cut and paste" functionality is almost always used as the starting point.

The employees at the Media Department are organized in the Graphical Union, and they do not want to allocate typing to any particular person, feeling that job rotation is important. The visibility and physicality of the faxes, the coordination form and the shelves simplify this job rotation and division of labor.

5. The SGML-based document system

The first step in the SGML project at the NA was the design of the Document Type Definition (DTD) that represents the structure of TV schedules. An SGML document always has an associated DTD that specifies the model of the document. The DTD defines the document's structure and concepts. The more a DTD is structured, the more functionality can be provided (Maler & Andaloussi. 1996). The DTD is the core of an SGML-based document system, and affects the system's design and use.

The consultants did the DTD modeling, working together with the IT manager to some extent. The level of structure was determined on the basis of discussions with the IT manager about future possible services, e.g. on the Internet. It is easier to go from a highly structured DTD to a less structured one than vice versa. Thus highly structured DTDs are often chosen in order to cope with possible future applications and services.

Since the DTD specifies a model of the document, the text producers both in the TV channels and in the media department have to produce the document according to the predefined structure, normally using an SGML editor. An SGML editor "interprets" the SGML standard and has mechanisms to help the writers to encode the information in relation to the predefined DTD. In this case the text producers use Author Editor. SGML editors were installed in the TV channels, being the beginning of the work chain. NA provided the SGML software and training, in order to motivate the TV channels to use the system.

The files were to be saved in a particular file structure on a server: /channel/week/day/file-name. Channel, week number, day and some status information give the file-name. The idea was that computer programs would pick up the right files and convert them for a desktop program for making pages ready to print.

Those at the end of this document exchange chain, the newspapers and magazines, should not notice any differences. The newspapers and magazines can choose between seven different TV schedules. These products differ in size, the selection of channels, each channel's space, and so on. The products are represented as different style sheets on the producers' computers.

There were suggestions that later it could be possible to make an information system based on the information in the program schedules, and to offer the press the opportunity to retrieve the information they wanted directly from a database. The information could also be presented on the WWW. This could provide "on demand" services for the public, e.g. selecting movies in the thriller genre, with a certain actor.

The SGML-based system was never adopted as such; only parts of it have been adopted. The TV channels have more or less refused to use the system, because the SGML editor was inflexible and they lost their channel profile and identity. However, the problems related to the interdependencies between the organizations are further discussed in (Braa & Sandahl, 1998).

6. From paperwork to network

In this section, we will discuss how document's properties change or even vanish during the transition from paper to digital documents. These are often resources that users rely on in order to get the work done. Implementation of document systems creates breakdowns in present work practices, resulting in resistance to using the system, as in this case. We present and discuss some *central* and *peripheral*

properties of documents, and how they are *border resources* in the cooperative work setting at the news agency that we have observed.

In this discussion we focus on the digital work situations at the Media Department. However, the paperwork related to channels that were never required to deliver electronic information is as before. The fax, shelves and forms exist as before in order to cope with this.

6.1 Coordinating artifacts

The huge number of faxes, the different concepts of weeks and the number of employees require coordination of many intertwined and interdependent activities to make it possible to create the schedules in time. The various actors (humans as well as artifacts) in the Media Department coordinate work activities that take place in a different time and space, articulating the activity affecting the other actors in producing the TV schedule for a week without being in the same time and space. In practice, moving faxes from one shelf to another, and signing into the coordination form does this. In this sense, the coordinating property of the faxes is the central one, and the faxes as coordinating artifacts are resources that the work community relies on.

Since the papers are ecological flexible and easily accessible through shelves, the division of labor is indicated by 'who is holding the fax'. The paper acts as a token. The state of work is represented by the shelf in which the faxes are placed, assimilable at a glance. When the faxes are in the in-shelf, no writing has started. However, staff note whether all TV schedules have arrived or not. When the faxes are in the out-shelf, the registration is complete. In this way the faxes in shelves indicate progress and trace the states of the work process. In addition, the work process is visible to all the workers at the Media Department.

This visibility is important in the division of labor and job rotation at the department. The workers can see who is working with what, and they are aware of how much a person has typed in the last few hours. Since they do not want to allocate typing to any particular person, they change when they sense it is their turn. This understanding of the activity of others, which provides a context for one's own activity (defined as awareness in Dourish and Bellotti (1992)), is in this case implicit. The community of practice relies on this implicit information in order to carry out the work.

However, the IT-designers regard the central property of the faxes as the content (structure, with elements). They were not aware of the additional resources, and focused on how to represent the structure and content. In the new system the faxes, shelves and forms were replaced (see figure 2). One might say that files on servers represented faxes on shelves. But an employee cannot get an overview simply by glancing at the shelf; (s)he has to browse through 350 files for every week. The file name reflects the status, instead of the forms and faxes in different shelves. Again, one has to browse through the 350 files to find out what has been done, and what is outstanding.

There is no overview; the border resources such as the public properties of the faxes and the trace properties have been removed, and nothing has replaced the coordination capacities. The foundation for implicit awareness broke down. It is difficult to see who is working with what and what remains to be done. The personnel were totally dependent on the coordination form to keep track of the work process.

This invisibility has serious consequences in the process of typing and proofreading, because writers are able to update the same file. The files are on a server (not a database) and there is no technical locking mechanism to prevent this.

In the paperwork situation, they saw people holding faxes, and recognized that faxes were moved from shelf to shelf, which indicated status and progress, as well as generating action. In short, the TV schedules on paper were essential in the coordinating mechanisms within the department. In the digital document system these mechanisms become invisible and broke down. The border resources developed through conventions over time were removed and not replaced.

6.2 Articulation work

The coordination form is an example of what Schmidt and Simone (1996) call "specialized artifacts, which, in the context of conventions and procedures, are instrumental in reducing the complexity of articulation work". We have, however, observed that not only "specialized artifacts" are essential in this respect. At the Media Department, faxes, shelves *and* forms are all artifacts that have roles within their community of practice and are resources that "minimize" the articulation work. We have seen that the need for articulation work increases when the coordinating properties of the faxes and shelves disappear and hinder the workspace awareness. The personnel needed to be aware of the activities of others in using 'shared artifacts' to help them coordinate and accommodate them into the ongoing work practice. This awareness could be explicit, as reported by Rogers (1993), or more implicit, as reported by Heath and Luff (1992) and Hughes et al. (1992).

The coordination form is a way to get explicit information on what is happening. We have observed that this information is not enough, and the implicit information from the faxes in shelves is important as well.

The need for articulation work also increased as a consequence of changes in some procedures and conventions that stipulate the articulation of writing the TV schedules. Since the newspapers and magazines require TV schedules for a week at a time, the pragmatic approach was to work on schedules for a whole week. In practice, the previous week was copied and acted as a point of departure for the current week production. However, the DTD is based on a day. According to the DTD, a document consists of information for one day, which means that when the day is ready the file must be saved, and a new file created for the next day. Using the SGML-based system, they were no longer allowed to work with a week at the time; the DTD forces the writers to produce TV information for a day and not a week. Dividing up weeks into days increases the complexity of the production process and the coordination form becomes even more important. However, the coordination form represents a week as well. In addition, signing the form each day creates difficulties in terms of space available and complicates status monitoring.

Previously, the faxes acted as coordinating artifacts, and the production of TV schedules was based on the concept of a week. The articulation work became more complicated, and additional coordination mechanisms become necessary when these things changed. In order to organize the work, the coordination form was extended and became more essential. Or as one person at the Media Department put it:

104

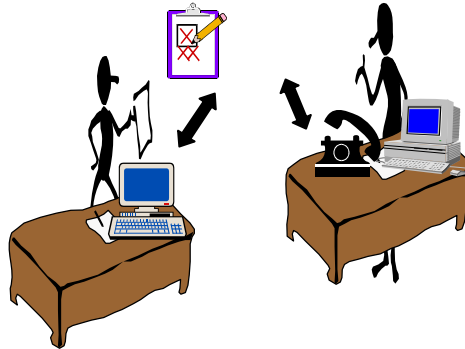"Now, we have to put everything on the form, or else it is just chaos."

Figure 2: Digital work.

If we compare the paperwork illustration in figure 1 with the digital work illustration in figure 2, the differences are striking. The paperwork situation is a heterogeneous production system including humans, faxes, forms, shelves and computers. The faxes were only not objects to be worked on, but were also acting as subjects in the production process, informing the operators what to do. When these were replaced without supporting these resources the need for articulation work, with additional coordinating artifacts, became essential. The complexity of the work has increased even though it is invisible. We have seen that the artifacts border resources is essential in keeping the articulation work on a low level in a cooperative work arrangements.

6.3 From network to paperwork

We have observed that digital work requires new and additional paperwork methods in order to keep the practice together. The TV channels were supposed to deliver the corrections in SGML to the NA, but no one does so. Most of the corrections from the TV channels come as email to one particular PC in the Media Department. Some corrections come as faxes. Corrections that come by email are in electronic form, and the email could be forwarded to a suitable person for updating. However, the email is printed out and placed in the correction shelf or on someone's desk, depending on when in the production process the corrections arrive. The person who picks up the email on paper, or receives it on her/his desk, is responsible for the corrections. The email is printed out in order to maintain the coordination property, in the same way that the property is embedded in the faxes.

When one of the personnel opens up one of the predefined style sheets in QuarkXpress, all the different files from each channel are automatically converted into the desktop program and placed in a predefined space on the QuarkXpress page. Unfortunately, the text is always too long and it is necessary to delete and rewrite the text to fit it into the space available. However, the next time a style sheet is opened, the conversion into QuarkXpress is based on the same SGML files, which still are too long. If the space available for a channel is exactly the same size within different style sheets, the personnel copy and paste between them. Usually, they print out the first TV schedule that is ready, and use it to see how they shortened words and phrases so that they can use the same formulation in other style sheets instead of producing new ones. This automation presupposes that the text is in an appropriate length, and

the goal is to present the information in TV schedules without human intervention. However, the text length is a central property of the document for the producers of the schedules, and much time is used to adjust text to fit into predefined spaces. The SGML system does not support this work practice.

7. Design challenges and prototype suggestions

Documents are, as discussed above, integrated into social activities, and cannot be separated from the practice in which they are incorporated. In NA the forms display activities and regulate the division of labor. Faxes have unnoticed resources that intertwine conventions and practice such as: "I am holding the fax, so I do the job". When all faxes have been moved from the in-shelf to the week-shelf, this indicates that data entry should start. These are examples of resources that are essential for people to carry out their work. A challenge for design is to identify the conventions that "keep the practice together", in order to avoid breakdowns. This will probably vary in different situations, since the resources of the paper document differ. The additional resources of documents should be studied and assessed when designing digital document systems.

According to the argument of Brown and Duguid (1994), these resources have to be taken into account in design for at least two reasons. First, because the border resources are not transformed automatically into new designs; (in the frames of references) they get removed, e.g. the visibility of faxes in week shelves disappeared. Second, because of the continuous proliferation of new technologies, workers have less time to develop new borderline resources.

In relation to SGML-based solutions, the DTD design is important. It is challenging to design DTDs that are based on practice instead of end products, in order to avoid the breakdown of essential conventions as happened in the Media Department when the concept of a week was split into days.
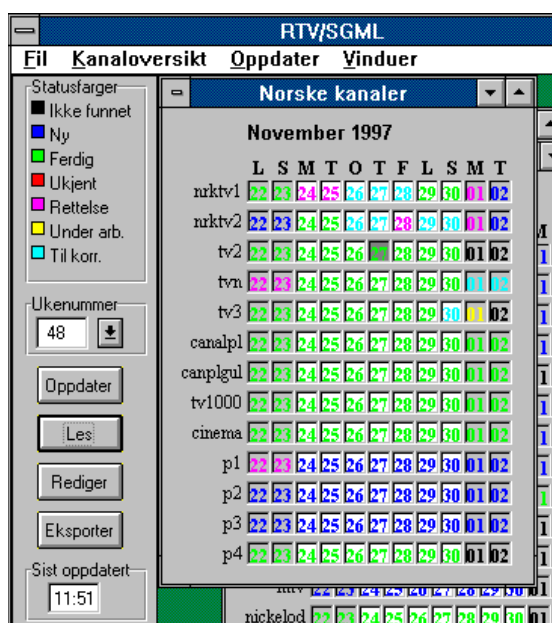

Figure 3: The prototype's user interface

A prototype was developed to make it possible to experiment with implementing some of the identified border resources in the system.

The prototype generates an overview showing which channel has sent information for which day, and the status of the production process. The user interfaces, which are calendars, represent both the TV channels' weeks and the Media Department week, and 11 days in total are displayed in the interface. A user interface consists of 12-14 channels. Figure 3 shows the Norwegian channels. The status is represented as different colors on the dates. The faxes may have the following status: not yet arrived (black), arrived (blue), in process (yellow), ready for proofreading (light blue), and finished (green). Since the TV channels often send corrections in the same way as the original, the corrections (pink) are implemented as well. If a schedule is not recognized, it is represented as unknown (red). The colors represent the faxes' progression through the work process. They provide a way to represent the visibility of faxes in the various shelves. In addition, the different notions of a week are supported.

When the user selects a channel (for instance, nrktv1) and double-clicks on a certain date the editor starts, and opens the corresponding file.

The prototype takes into account the role of the documents as coordinating artifacts, represented by color, and the conventions related to the concepts of a week. The news agency is using the prototype in production. However, so far it is only the typist who uses it. The people who work with the page layout in QuarkXpress are still highly dependent on the form.

Another prototype was developed to allow the TV channels to reduce the complexity of the DTD in general, as well as to support the different needs of the various TV-channels by designing sub DTDs. These DTDs are linked with style sheets applied in the various text processors. The conversion into SGML is based on these style sheets. The prototype has been successfully tried out, and is now in production for this particular TV channel. How this solution act as a gateway between the different communities of practices is further discussed in Braa and Sandahl (1998).

8. Conclusion

In this paper we address the power of documents as artifacts with both social and material aspects. The objective has been to reveal the document's central and peripheral properties as well as to describe the unnoticed resources of the documents that are shared by a community of practice. The study shows that documents and document-related artifacts are powerful in "keeping the work practice together", and they are essential to handle the articulation work. In our case, when the documents and shelves are visibly present, the articulation work is minimal. However, when the artifacts disappear, the need for articulation work increases, because of the increased complexity in work organization.

We state that understanding the unnoticed resource of the documents and document-related artifacts embedded in the work practice could be a starting point for design of document-based systems. The system developers for the NA and the TV channels have focused only on what they saw as the TV schedules' central property. The border resources were never taken into account. Removing these resources without designing a substitute caused the cooperative work to collapse. However, it is challenging to identify and design for these resources. In the prototypes we have tried to implement some of these resources into the design. They have been adopted in practice.

In relation to SGML-based solutions, the DTD design is important. The DTD design impacts work. In this case, the process and practice related to the production were never taken into account in the design; only the structure and content of the TV schedule itself.

However, the complexity increases in inter-organizational systems or infrastructures, since the border resources vary in different communities of practice. This aspect is further discussed in (Braa & Sandahl, 1998).

Acknowledgement

References

Berry, M and Goulde, M.A. (1994) A New View of Documents. In *Workgroup Computing Report* Vol. 17. No 8.

Brown, John, S and Duguid, P (1995). The Social Life of Documents. *In First Monday, peer-reviewed journal on the Internet*.

Brown, John S and Duguid, P (1994). Borderline Issues: Social and Material Aspects of Design. In *Human-Computer Interaction*, 1994, Volume 9, pp. 3-36.

Braa, K and Sandahl, T. I. (1998). Documents in Infrastructure. Challenges for Design. To be presented at the European Conference of Information Systems (ECIS), Aix-en-Provence France, June 4-5-6 1998

Braa, K and Sandahl, T. I. (1998b). Approaches to Standardization of Documents. In Toshiro Wakayama, Srikanth Kannapan, Chan Meng Khoong, Shamkant Navathe, JoAnne Yates (Eds.) *Information and Process Integration in Enterprises. Rethinking Documents*. Kluwer Academic Publishers.

Braa, K. & Sørgaard, P. (1997) Stages and diversity in the implementation of World Wide Web and document technology. In 2nd *IFIP 8.6 International Working Conference on Diffusion, Transfer & Implementation of Information Technology*, 25th-27th June, 1997, Ambleside, Cumbria, UK.

Dourish, P. and Bellotti, V. (1992) Awareness and Coordination in Shared Workspaces. *In Proceedings of the Conference on Computer-Supported Cooperative Work*. October 31 to November 4. 1992. ACM Press.

Eriksen, L.B. (1997) Digital Documents, Work and Technology, three cases of Internet News Publishing, In *Proceedings from 30th HICSS*, Wailea, Hawaii, Jan 7-10.

Heath, C., and Luff, P. (1992). Collaboration and Control: Crisis management and multimedia technology in London Underground Line Control Rooms. In *Computer Supported Cooperative Work (CSCW), An International Journal*, Vol. 1, pp. 69-94.

Hughes, J and King, V (1993). Paperwork. In Steve Benford and John Mariani (eds) *Requirements and Metaphors of Shared Interaction*. Part II: Shared Artifacts. COMIC Deliverable 4.1, pp 153-325. Lancaster, England. 1993

Hughes, J, Randall, D., and Shapiro, D. (1992). Faltering from ethnography to design. In *Proceedings of Computer-Supported Cooperative Work*, 1992, Toronto, Canada

Levy, D (1994). Fixed or Fluid? Document Stability and New Media. In *European Conference on Hypermedia Technology 1994*. Conference Proceedings. New York: ACM Press, 1994

Levy, D (1988). Topics in Document Research. In *Proceedings from ACM conference on Document processing systems*, pp. 187 - 193. Santa Fe, New Mexico, December 5-9, 1988

Luff, P, Heath, C. and Greatbatch, D. (1992). Tasks-in-interaction: paper and screen based documentation in collaboration activity. In *Proceedings of Computer-Supported Cooperative Work*, 1992, Toronto, Canada

Maler, E and Andaloussi, J. (1996*). Developing SGML DTDs. From text to model to markup*. Prentice Hall, PTR.

Reinhardt, A (1994). Managing the New Document. In *BYTE,* Vol 19, No. 8, august 1994

Rogers, Y. (1993) Coordinating Computer-Mediated Work. In *Computer Supported Cooperative Work (CSCW), An International Journal*, vol 1, pp. 295-315.

Sandahl, T.I. and Jenssen, A. (1997) The First Steps in Designing an SGML-Based Infrastructure for Document Handling. Submitted to Scandinavian Journal of Information Systems.

Schmidt, K and Bannon, L (1992). Taking the CSCW Seriously: Supporting Articulation Work. In *Computer Supported Cooperative Work (CSCW). An International Journal.* Vol 1, pp. 7-40.

Schmidt, K and Simone, C (1996). Coordination Mechanisms: Towards a Conceptual Foundation of CSCW Systems Design. *In Computer Supported Cooperative Work: The Journal of Collaborative Computing* 5: pp. 155-200. Kluwer Academic Publishers.

Straus, A (1985). Work and Division of Labor. In *The Sociological Quarterly*, Vol 26, No 1.

Sørgaard, P. and Sandahl, T.I (1997). Problems with Styles in Word Processing: A Weak Foundation for Electronic Publishing with SGML. In *Proceedings from 30th HICSS*, Wailea, Hawaii, Jan 7-10.

Vidgen, R & Braa, K (1997) Balancing interpretation and intervention in information system research: the "action case" approach, in *Proceedings of IFIP 8.6: Information Systems and Qualitative Research*, Philadelphia, 31 May - 3 June, 1997.

Yates, J (1993). *Control through Communication. The Rise of System in American Management.* The Johns Hopkins University Press, Baltimore and London.

Yates, J. and Orlikowski, W. (1992) Genres of organizational communication: A structurational approach to studying communication and media. In *Academy of Management Review*, Vol 17, No. 2, pp. 299-326.

Documents in Infrastructure - Challenges for Design

In Proceedings from the European Conference of Information Systems, Aix-en-Provence France, June 4-6 1998

Kristin Braa and Tone Irene Sandahl

Abstract

The objective of this paper is to reveal the heterogeneous complexity of document infrastructures in work practices. The case shows that documents in infrastructures have both physical and social properties, as well as roles in keeping the web of work practice together. Some design challenges relating to addressing and coping with this infrastructure are discussed.

I. Introduction

Digital technology offers immense possibilities both for effective production of documents and for providing new services and products. However, studies have identified difficulties in exploiting the possibilities of electronic documents as well as in providing new services (Braa & Sandahl, 1998a, Braa & Sørgaard, 1997, Eriksen, 1997).

Work practices need to share more than raw data in databases. They need to share interpretations and meanings of data in the context of its use, that reflects, supports or enables daily work activities within an organization (Wakayama et. al., 1998). This sharing of interpretations takes place in implicit as well as explicit ways, and is often manifested in documents (ibid.). Hughes and King (1993, p.153) go further and state that "[m]uch work in a variety of settings consists of working with and through documents in such a way that it could be said that the work consists of producing, managing and preserving documents".

Paper documents can be tailored and are ecologically flexible in use, and therefore suitable in synchronous and asynchronous collaboration (Luff et. al., 1992). People tailor their documents in order to differentiate and highlight particular items. Some may underline or mark text with color pens to alert colleagues to irregularities and importance. The ecological flexibility of documents makes it possible to adapt them to a range of situations and contingencies. Even if the document's content and structure can be interpreted differently depending on the social setting (Hughes & King, 1993), the paper document as an artifact can be adopted and geared to other work practices without any additional features (Luff et. al., 1992).

The role of documents in infrastructures is highlighted to emphasize that much of the communication and coordination among people is based on documents of some kind, and often represent shared cross-organization.

We emphasize the importance of paper documents' ecological flexibility in cases where they act as links or "bindings" between different work practices, as well as how

this flexibility decreases when documents become electronic. In order to design document systems, an understanding of the heterogeneous complexity of work practice where documents play a central role is essential.

An infrastructure is often characterized as something relatively big and complex, which is developed over a long time and consists of a relatively large number of independent actors involved in design as well as use (Hanseth, 1997). Often infrastructures are compared with substructures or substrates that something else "runs" or "operates" on, and have a rather technical connotation. Berg's (1997) notion of work practice, where work practice is viewed as a configuration of materially heterogeneous elements that in its circulation and interrelation is geared towards the realization of some output, has very much the same emphasis, but with a bias towards the role of work practice.

We want to regard infrastructure as a heterogeneous actor network (Callon, 1991) consisting of humans and non-humans in a socio-technical network, where work practice as well as documents play an important role. This actor network perspective can be scaled up and down for analytical purposes. We will emphasize the role of documents in infrastructures, the way documents tie work practices together, and the way documents mediate work. In another paper, we emphasize the documents' role within a particular work practice (Braa & Sandahl, 1998b).

We describe a particular transformation process from paper to electronic documents taking place in a news agency with the aim of enrolling a variety of TV channels into the network, and discuss the challenges for designing such document infrastructures. Various autonomous and independent actors are involved in this infrastructure, making it difficult to change.

The attempt was a failure, and there may be several reasons for this. The document design was treated as a technical task implying that the news agency could enforce a particular new design, neglecting the way that the documents act as a bridge between the different actors. Changes in these documents involve internal changes in the various actor networks. The various TV channels have different content and interpretation of these documents. However, they share something. We use the concept of boundary objects (Star & Griesmar, 1991) to explain the role of these shared documents in the infrastructure.

The study can be characterized as an action case (Vidgen & Braa, 1997). Action case is a research approach, which deals with the dilemma of intervention and interpretation in in-context research. The case study has lasted for two years, consisting of observations and interviews of different actors involved in the process: text producers in the TV-channels and news agency, line managers, IT personnel and IT consultants. In addition, we conducted small-scale interventions such as workshops, prototypes of alternative solutions, and training are developed and introduced into the work situations.

In the next section we will broaden the perspective on documents in infrastructures by discussing documents as boundary objects. The case is then described and discussed. The paper concludes with lessons learned from the failure that imply challenges for design, in addition giving examples of some alternative design solutions.

2. Documents as Boundary Objects

The concept of a boundary object is applied for analytical purposes (Star & Griesmar, 1991) in order to understand the role of documents in infrastructures. A boundary object is defined as something that arises over time from ongoing cooperation between communities of practice, such as a way of resolving different definitions of things, situations, problems, etc. It is a complex set of arrangements between two or more co-operating communities, which satisfies their different information needs. Boundary objects are thus objects that are plastic enough to adopt to the local needs and constraints of the several communities, yet robust enough to maintain a common identity across sites. They are weakly structured in common use, and become strongly structured in individual-site use. The objects may be abstract or concrete. They have two important qualities: ambiguity of meaning across sites or communities; and durability in time as a set of working arrangements (ibid., Star, 1989).

Boundary objects are essential in infrastructures because they link different work practices together. They are the objects that create and maintain the dependencies between work practices. Documents are often boundary objects. On the one hand, "[d]ocuments arise from the basic human need for stable, external, communicative resources. What is said in the hallways is transient and insubstantial, while what is written down is fixed and external and therefor potentially available to participants across time and space (Levy, 1988, pp. 187)". On the other hand, the documents have different roles in work settings. Documents are "displays" of the methods used to produce them. What could be found in documents would depend on the interpretation of the displays. A reader has to know what to look for, which means knowing what the document represents and its implications, as well as how to read it as a display of the procedures used to produce them in their organization (Hughes & King, 1993). In addition, documents as artifacts can have additional, unnoticed resources that are essential within one work practice and absent in another. In one work practice, a lack of a mark in an invoice is an unnoticed resource that tells that a colleague is not at work, and that someone else has to do her job (Hughes & King, 1993). In another work practice, this lack of a mark has no significance at all. These unnoticed resources are developed over time as artifacts are integrated into ongoing practice, and they are essential to people in carrying out their work (Brown & Duguid, 1994). In Braa & Sandahl (1998b) we examine the document's role within a particular work practice and emphasize the additional resources observed in that context.

In other words, documents can be stable and external, as well as open to flexible interpretation and internal, at the same time.

In the next section we will illustrate how a transformation process from paper to electronic documents was planned and conducted in a News Agency (NA) and failed to be diffused to the TV channels.

3. The Case: The News Agency

NA is a news agency providing news services to the press — newspapers, magazines, radio and television. The need for a new editorial system that supports electronic text exchange has been recognized in NA for several years. The existing system is based on old-fashioned mainframe technology. The focus on a new system intensified from 1993 where the International Press Telecommunication Council (IPTC) introduced the SGML-based "New Industry Text Format" (NITF) to increase electronic text interchange between news agencies and their customers / suppliers.

IPTC recommended that all news agencies adopt the standard in their editorial and communication systems.

NA wanted to build competence within the organization on a minor project before continuing with an implementation for the whole organization. In 1995 the IT department, together with external consultants, took an initiative to apply SGML in the production of one of NA's services: the TV schedules. TV schedules, produced for newspapers and magazines, were regarded as the most structured and well-defined product they produced, and thus assumed to be easy to standardize in SGML.

3.1 The production of TV schedules

The production of TV schedules involves a chain of work tasks. Starting with the production of program information in the individual TV channels, this chain involves merging, quality control and coordination in the media department, ending in a variety of weekly products to be delivered to different newspapers and magazines. There are approximately 40 TV channels, and the number varies as TV channels close down and start up.

From the NA's point of view, there was an urgent need to facilitate the document exchange process between the TV channels and NA. Almost all schedules from TV channels arrive as faxes. The information is retyped and encoded into an old mainframe system, where the schedules are edited before being converted for import into a desktop program to create pages ready for printing, or transmitted as data to the newspapers and magazines. The electronic exchange of TV program schedules from TV channels to NA was expected to result in enormous time savings, because the people in NA would then not have to retype the text, but only to improve the quality.
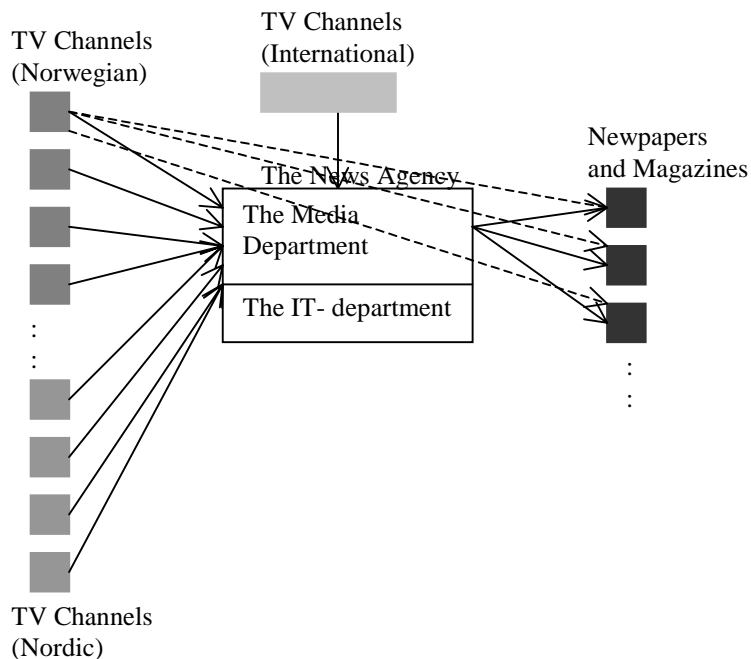

Figure 1. The organizations involved in the production of TV schedules

They started by including the Norwegian TV channels, except the major national TV channel. The Nordic Channels also deliver their information by fax, but they were not

included in the pilot project. Nor were the other international channels. Information from the international TV channels is received from a firm called Intermedia. NA is waiting for Intermedia to deliver TV information in SGML format. Negotiations on both the price and the solution are still taking place.

3.2 The SGML-based Document System

Early on, it was decided that the document exchange system should be based on SGML. From a technical perspective, application of SGML is an appropriate approach in such design for three main reasons: i) Standardization is essential for infrastructure to exist; it is the technical backbone (Hanseth et al., 1996). ii) SGML makes it possible to use structured concepts in text in general, which lessens the difference between documents and databases, as well as improving search and retrieval in texts. iii) SGML is independent of software, systems and presentations, and it supports "open systems". Expectations for SGML as a technology and a platform for the solution were thus immense.

The modeling of the document – the construction of a Document Type Definition (DTD) – was done by consultants from a print shop. An SGML document always has an associated DTD that specifies the model of the document. The DTD defines the document's structure and concepts. The more a DTD is structured, the more search functionality can be provided (Maler & Andaloussi. 1996). The DTD is the core of an SGML-based document system, and affects the system's design and use.

The aim is to ensure that all program schedules consist of the same information and are structured in the same way. After validation of the document according to the DTD syntax, scripts can be converted for desktop programs or other applications automatically, and provide facilities for advanced search such as "on demand" documents.

Since the DTD specifies a model of the document, the text producers both in the TV channels and in the media department have to produce the document according to the predefined structure, normally using an SGML editor. An SGML editor "interprets" the SGML standard and has mechanisms to help the writers to encode the information in relation to the predefined DTD. In this case the text producers use Author Editor. SGML editors were installed in the TV channels, being the beginning of the work chain. NA provided the SGML software and training, in order to motivate the TV channels to use the system.

The files were saved in a particular file structure: /channel/week/day/file-name. Channel, week number, day and some status information give the file name. The idea was that computer programs would pick up the right files and convert them for a desktop program for creating pages ready to print.

Those at the end of this document exchange chain, the newspapers and magazines, should not notice any differences. However, there were suggestions that it could be later be possible to make an information system based on the information in the program schedules, and offer the press the opportunity to retrieve the information they wanted directly from a database. The information could also be presented on the WWW. This could provide "on demand" services for the public, e.g. selecting movies in the thriller genre, with a certain actor.

4. Why did it fail?

The SGML-based system as described above was never adopted. The project plan states that the system was to be implemented by March 1, 1996. However, it was only in the middle of October 1996 that the first SGML-encoded schedules arrived from a channel. Since then the same channel has sent three SGML-encoded schedules - the only channel to do so. The Media Department was supposed to start working with the SGML-based system on January 1, 1997. The system was never used in practice.

> "The only thing we want is to receive the text electronically in order to get away from all the extra work of retyping, not to take part in the idea that 'SGML should gather everybody to be a part of the same world" (Manager, Media department).

The essential reasons why it failed are, as we see it, the lack of focus in the design on existing practice within and between the involved organizations, the unawareness of the role of documents as a boundary object in keeping the chain together, and the fact that success depended on people outside the control of the news agency – the TV channels – adopting the system. In this section, we will discuss some explanations for the failure of these attempts.

4.1 Work practice at the media department

The TV channels send their TV schedules to NA almost entirely by fax. Only a few use mail, and one channel has its own solution, called Data-TV. All the faxes are sorted by channel and placed on a shelf. There are three different types of shelves: the in-shelf, week-shelf and out-shelf. The in-shelf consists of space for each TV channel. When a schedule is received, the fax is placed on the shelf where it belongs. The in-shelf gives an overview of what was received and what is missing. A glance at the shelf reveals the channels that are outstanding.

The TV channels send schedules for the whole week at once. However, the definition of a week is not uniform. What denotes a week varies among the different stakeholders. Most of the TV channel's week starts on Monday and ends on Sunday. The Media Department operates with a week that starts on Saturday and ends on Friday. The week in the newspapers and magazines depends on which day TV guides are published. The TV guides consist of information about radio and TV programs for "all" channels for a week. The "different" concepts of a week force the Media Department to split the information from the TV channels and to define the week according to the recipient's wishes.

In coordinating the different definitions of a week, the faxes, shelves and forms play important roles. When all the schedules from TV channels are received and placed on the in-shelf, all the faxes that containing information about the specific week get moved to the right space on the week-shelf. The week-shelf is not separated into channels, only into weeks. For every week, regardless of when it starts and ends, information about all programs every day for each channel has to be retyped. To maintain an orderly system, the personnel sign the weekly coordination form when a particular TV channel is finished. When the coordination form is completed, the week is done. By then, 10 different people within the Media Department may have been involved.

The employees at the Media Department are organized in the Graphical Union, and they do not want to allocate typing to any particular person, so that job rotation is important. The coordination form and the shelves simplify this job rotation.

When the week is finished, the faxes are moved to the out-shelf. If there is time, a controller takes the faxes out of the out-shelf and checks them against the text in the mainframe system. From the out-shelf, the information goes into the trash. When the faxes are in the out-shelf, the information is considered ready for preparing files in QuarkXpress that are ready for printing.

The previous week is used as the basis for data entry in the mainframe system. There are always some recurring programs, so "cut and paste" functionality is almost always used as the starting point.

4.2 The work practice at the TV channels

The different TV channels have a special profile for their channel, with respect to content and the presentation of the channel's identity. There are differences in the information that they want to emphasize, e.g. some channels want to emphasize the actors starring in a film, while others emphasize the film itself. In practice, the first channel presents the actors before the film title in the schedule, and the latter presents the title first. Some channels use a great deal of text to introduce a special sporting event; others just mention the participating teams. They all have a profile indicating the type of channel they represent.

Some channels broadcast almost the same programs every week; others have more varied schedules. They seldom write the information more than once; they copy and paste between documents.

When a schedule is finished it is printed out and faxed to NA and others. In fact, many newspapers and magazines get the schedules directly from the TV channels in order to prepare reports related to forthcoming TV programs. The same newspapers and magazines buy schedules from the NA some days (weeks) later.

4.3 The DTD modeling process: The Idea of a Common World

The first step in the SGML project at the NA was the design of the DTD model. The consultants did the DTD modeling, working together with the IT manager to some extent. The point of departure was a program schedule from a newspaper that was analyzed according to its structure and concepts. The level of structure was determined on the basis of discussions with the IT manager about future possible services, e.g. on the Internet. It is easier to go from a highly structured DTD to a less structured one than vice versa. Thus high structured DTDs are often chosen in order to cope with possible future applications and services.

The heterogeneous work practices in the media department and TV channels were never studied nor a matter of concern. The differences in practices were thus not recognized, and one common DTD was designed for both text producers in the TV channels and text producers in the media department. This required a common understanding and use of domain-oriented concepts in all the TV channels and in the Media Department. In practice, there are some major differences in the use of concepts within the different organizations. For instance, the concept of a tableau has different meanings in the Media Department and the TV channels. In the Media Department it denotes a program description only. In some of the TV channels it denotes a program description including some text about the story or event. Synopsis

was defined as a term to include the additional sentences. However, the Media Department never understood what it was, and the TV channels have never heard of it.

Since the DTD defines and standardizes the information structure in the schedules, some channels complained that they lost their profile. Following the structure in the DTD, channels are no longer allowed to present the actors before the movie title.

The DTD model was presented for the reference group. The reference group had representatives from some of the Norwegian TV channels and the manager of the Media Department. There were three meetings in the reference group. Changes to the DTD model suggested in the meetings were never implemented.

The print shop treated the work task of designing a DTD model as a product to be developed, not as a production process involving variety actors. The idea of one common DTD failed because the concepts and profile did not suit all the participants in the work chain.

4.4 Enforced Approach to Standardization

Three main approaches to standardizing documents in SGML have been identified: "soft", guided and enforced (Braa & Sandahl, 1998a). These approaches are viewed from a text producer's perspective. The *"soft"* approach is to let the text producers write freely and instead hire people to go through the text and mark it up in SGML. This approach upholds the existing practice. However, it requires extra work to prepare the documents for exchange related to the predefined DTD. The *guided* approach to standardization is to guide text producers to use styles as a tool for structuring in a word processor and then convert the documents to SGML. With *enforced* standardization, the producers of a text produce the SGML-coded text themselves. They also have to make sure that the text is written in relation to the predefined DTD.

In this case an enforced approach to standardization was applied, which requires the text producers to follow the rigid structure when writing; forced to follow a certain practice.

Previous studies (Ibid.) show that writing a text directly in an SGML editor requires a high level of training and user support. Since the DTD has to be followed, for instance, the cut and paste facilities are restricted. The SGML editor will not allow cut and paste in documents if there is a risk of destroying the structure. So, on top of the increased inflexibility in use, the practice of cut and paste was not allowed. The result was that it took substantially longer to produce the text, both for people at the TV channels and at the Media Department.

There is high personnel turnover in TV channels. Work with TV schedules, in particular, is not regarded as having a high status, and people leave or get other jobs. The enforced approach to standardization requires time for training, but text producers seldom stay in their jobs for very long. Frequently, new persons have to be trained.

4.5 Those doing the work are not those getting the benefits

The text producers at TV channels were responsible for delivering a document to the media department, which was 100 % correct according to the DTD.

117

Normally, the TV channels use Microsoft Word for this task, which is only a part of their daily work. Another complication is that the program text in turn forms part of another chain of work tasks within the TV channels, e.g. the channels distribute it to about 100 other recipients. Today the TV channels produce two types of schedules, one that NA requires and one that the other sources want.

The result is that so far the TV channels have refused to use the system. The suggested SGML solution relies on leaving the job to the TV channels. Why should they do it, when they do not gain any benefits? This was the most visible of this project. The success of implementing this system relied on people outside the host organization.

4.6 Coordinating artifacts

The huge number of faxes, the differences in weeks and the number of employees require coordination of many intertwined and interdependent activities to make it possible to create the schedules in time. In such situations there is a need for structuring and control of how the interrelated activities should be meshed (Carstensen & Sørensen, 1997). The shelf and forms are artifacts that play such roles in the Media Department. These artifacts coordinate and organize the interactions, and are essential in the division of labor.

The various actors in the Media Department coordinate work activities that take place in a different time and space, articulating the activity affecting the other actors in producing the TV schedule for a week without being in the same time and space. The coordination form also accumulates the activities carried out: one can read the form to get a brief history, assimilable at a glance, of the work that has taken place (for that particular week).

The faxes, shelves and forms were replaced. One might say that faxes on shelves were represented by files on servers. But an employee cannot get an overview simply by glancing at the shelf; (s)he has to browse through 350 files for every week. The file name represents the status instead of the forms and faxes in different shelves. Again, one has to browse through the 350 files to find out what has been done, and what is outstanding. There is no overview, and there is nothing to replace the coordination capacities of the form. It is difficult to see who is working with what and what remains to be done.

In the paperwork situation, people in the media department saw people holding faxes, and recognized that faxes were moved from shelf to shelf, which indicated progress. In short, the coordination mechanisms broke down, because of the narrow view on documents.

5. Challenges for Design

Documents are integrated into social activities, and cannot be separated from the practice in which they are incorporated. In NA the forms display activities and regulate the division of labor. Faxes have unnoticed resources that intertwine conventions and practice. Resources such as: I am holding the fax, therefore I do the job. When all faxes have been moved, this indicates that data entry is complete. These are conventions that create practice and confirm it. A challenge for design is to identify the conventions that "keep the practice together", in order to avoid breakdowns.

What are the aspects of paperwork that are necessary to keep in order to "keep the practice together"? This will probably vary in different situations, while the meanings of the paper document also differ. The additional resources of documents should be studied and assessed when designing such infrastructures. This is further discussed in the Braa & Sandahl (1998b).

In this case, schedules delivered by faxes act as boundary objects between the TV channels and the News Agency. The TV schedules produced in the different channels are used for different purposes; for instance, they all send the schedule directly to the newspapers and the TV schedules are frequently reused in ongoing production.

The TV schedules (faxes) as a boundary object "link" the practices together in the sense that they are at the same time stable and durable, and "plastic" enough to be adopted into local practices. For instance, in the Media Department the faxes are placed on shelves, which have coordinating properties and are essential artifact in the division of labor.

In order to avoid breakdown in the network we propose to introduce gateways, which can function as a transformer between different practices and different technological solutions, as a mechanism of implementing the properties of boundary objects.

Gateways may link together incompatible networks, meaning networks that are internally aligned and stabilized but not aligned to each other. (Hanseth & Monteiro, 1998). In this way the practice and technology in one organization can change without forcing breakdown in other organizations. Below, we will describe two such attempts to solve some of the problems discussed in this paper.

However, in relation to SGML-based solutions the DTD design is important. It is challenging to design DTDs that are based on practice instead of on products, and whose elements and attributes reflect the document's role in social settings as well as physical properties.

In this case, the heterogeneity of the chain of work tasks was not supported. However, this is not impossible to implement in a document system if one is aware of them. In collaboration with two of our master students we have experimented on prototypes illustrating how to solve some of the problems discussed in this paper. The major problem, discussed in this paper, is to create satisfactory flexibility for the TV channels and the same time creating necessary standardization so that the chain of work tasks is not broken. We have made interconnecting gateways in order to align different work practice into the SGML system at the NA.

Gateways: A gateway was designed by using a style sheet that reflects a new, less structured DTD made for each channel. This is a guided approach to standardization. The style sheet consists of 12 styles, which reflect the most frequently used text elements. In order to satisfy each channel's profile, macros have been implemented. The macros go through the text and make sure that the formatting is right. The sub DTD for a particular application area (TV channels) was designed as a subset of the original "common" DTD, by selecting the most important elements for that particular TV-channel. There are scripts that translate from a simple DTD towards the more structured DTD. For instance, while the "common" DTD has 63 elements, the subset DTD has 12, which are all known at the TV channel. The subset DTD can facilitate different macros (and word processors). The aim is to reduce complexity and increase flexibility by allowing the channels to develop their own profiles on the schedule, and to be able to use the same product to deliver to other sources. It also facilitates the production of schedules as part of other different work chains in the TV

channels. In this way properties connected to boundary objects are implemented in the gateway as style sheets that can be flexible for each channel and as sub DTDs that reflect a subset of shared elements (see Figure 2). All this effort reflects how flexible and easy to tailor the paper-based work practice based on faxes is.
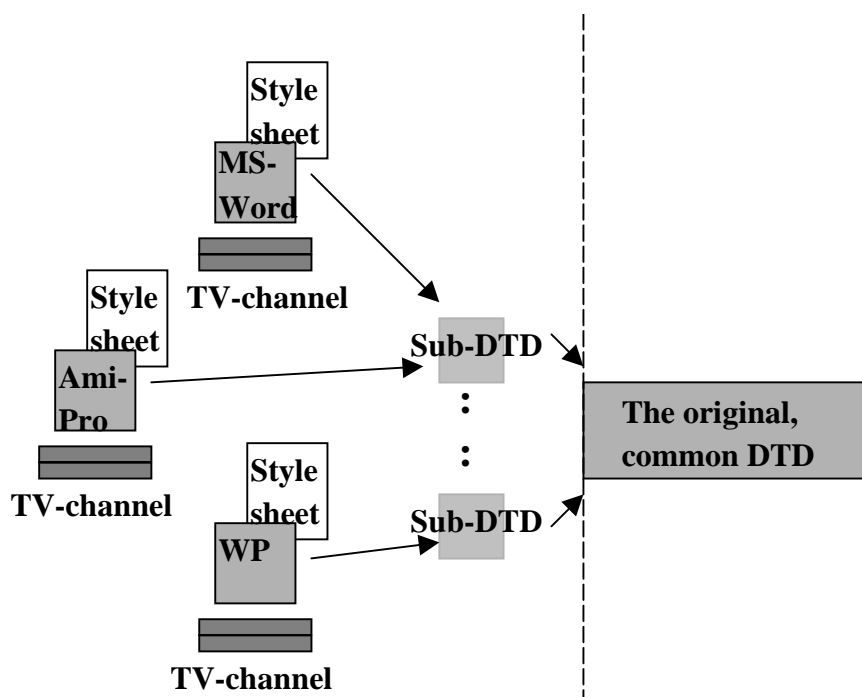

Figure 2: The Style sheets and sub-DTDs as Gateways

Previous studies of the use of styles (Sørgaard & Sandahl 1997; Braa and Sandahl, 1998a) have shown that people hardly use styles. In order to improve the use of styles, training programs were established and the user interface was improved. For instance, the styles appear in a permanent window, instead of as elements in a menu.

Another problem superficial discussed in this paper, is that the coordinating artifacts were not recognized as a part of the chain of work tasks and were not implemented as such.

Coordinating properties: A user interface of the SGML system was implemented in Visual Basic. When the 350 files are read by the SGML system, the interface generates an overview showing which channel has sent information for which day. In addition, the status of the information is shown. When the user selects a certain date an overview will appear, communicating with the SGML system, presenting all files that have been received for that particular day. This is further spelled out in Braa and Sandahl (1998b).

The prototypes have been successfully tried out in one production chain from one TV channel delivering the text into the news agency's SGML-based system, and are now in production for this particular TV channel. NA is considering implementing such solution for the remaining TV channels.

7. Conclusion

The objective of this paper has been to reveal the heterogeneous complexity of socio-technical networks, where documents play an essential role. The action case shows that documents in infrastructures have both physical and social properties, as well as roles in keeping the web of work practice together. Understanding of this practice should inform design. Some design challenges relating to addressing and coping with this practice are discussed. In infrastructures where organizations and people are related together in heterogeneous networks, gateways are suggested as an attempt to connect different work practices. The concept of boundary objects is introduced in order to identify such document properties to be implemented by gateways. This has been illustrated by the prototypes described, aiming at creating a flexible solution reflecting local conventions and work practice as well as contributing to the common production process. When designing document infrastructure, the interdependencies in work should be considered; that is, each group should profit by using the system, unlike the news agency case where the TV channels were supposed to do the job without gaining any benefit. In addition, an enforced approach to standardization is a difficult approach because the actors in the infrastructure are independent and autonomous and will hardly change the work practice in the same pace.

Acknowledgement

References

Berg, M (1997) Accumulating and Coordinating: Occasions for Reading and Writing Artefacts in Medical Work, paper presented in Department of Informatics, University of Oslo, October 20. 1997.

Brown, John S and Duguid, P (1994). Borderline Issues: Social and Material Aspects of Design. In *Human-Computer Interaction*, 1994, Volume 9, pp. 3-36.

Braa, K and Sandahl, T. I. (1998a) Approaches to standardization of documents, in *Information and process integration in enterprises: rethinking documents,* Kluwer Academic Publishers, Cambridge, Massachusetts, USA, pp 125-143.

Braa, K and, Sandahl, T. I. (1998b). From Paperwork to Network. Submitted to *Third International Conference on the Design of Cooperative Systems (Coop '98),* Cannes, France, 26.-29. May 1998.

Braa, K. and Sørgaard, P. (1997) Stages and diversity in the implementation of World Wide Web and document technology, in 2nd IFIP 8.6 International Working Conference on Diffusion, Transfer & Implementation of Information Technology, 25th-27th June, 1997, Ambleside, Cumbria, UK.

Callon (1991) Techno-economic networks and irreversibility. In Law, J. (ed) *A sociology of monsters. Essays on power, technology and domination*, pp. 132-161. Routledge.

Carstensen, P.H., Sørensen, C (1997). From the Social to the Systematic - Mechanisms Supporting Coordination in Design. In Computer Supported Cooperative Work - An International Journal, pp. 1-26.

Eriksen, L.B. (1997) Digital Documents, Work and Technology, three cases of Internet News Publishing, in HICSS'30, Hawaii.

Hanseth, O. (1997). Information Technology as Infrastructures. Doctoral dissertation. Department of Informatics. Göteborg University.

Hanseth, O, Monteiro, E and Hatling, M (1996). Developing information infrastructure: the tension between standardisation and flexibility. In *Science, Technology and Human Values*, Vol 21 No. 4, pp. 407-426, Sage Periodicals Press.

Hanseth, O. and Monteiro, E. (1998) Changing irreversible networks: Institutionalisation and infrastructures, in ECIS'98 Proceedings of the sixth European Conference on Information Systems, Aix-en-Provence, France, June 4-6.

Hughes, J and King, V (1993). Paperwork. In Steve Benford and John Mariani (eds) *Requirements and Metaphors of Shared Interaction*. Part II: Shared Artifacts. COMIC Deliverable 4.1, pp 153-325. Lancaster, England. 1993

Jewett and Kling (1991). The dynamics of computerization in a social science research team: A case study of infrastructures, strategies and skills. In *Social Science Computer Review*, no 9, pp. 246-275.

Levy, D (1994). Fixed or Fluid? Document Stability and New Media. In *European Conference on Hypermedia Technology 1994*. Conference Proceedings. New York: ACM Press, 1994

Levy, D (1988). Topics in Document Research. In proceedings from ACM conference on Document processing systems, pp. 187 - 193. Santa Fe, New Mexico, December 5-9, 1988

Luff, P, Heath, C. and Greatbatch, D. (1992). *Tasks-in-interaction: paper and screen based documentation in collaboration activity*. In Proceedings of Computer-Supported Cooperative Work, 1992, Toronto, Canada

Maler, E and Andaloussi, J. (1996*). Developing SGML DTDs. From text to model to markup*. Prentice Hall, PTR.

Star, S. L. (1989) The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving. In Gasser, L. and Huhns, M. (eds) *Distributed Artificial Intelligence*, Vol. 2. Pitman, London.

Star S. L. and Griesmer J., R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkley's Museum of Vertebrate Zoology, 1907-39. *In Social Studies of Science*, August, 1989.

Sørgaard, P. and Sandahl, T.I (1997). *Problems with Styles in Word Processing: A Weak Foundation for Electronic Publishing with SGML*. In Proceedings from 30th HICSS, Wailea, Hawaii, Jan 7-10.

Vidgen, R & Braa, K (1997) Balancing interpretation and intervention in information system research: the "action case" approach, in *Proceedings of IFIP 8.6: Information Systems and Qualitative Research*, Philadelphia, 31 May - 3 June, 1997.

Wakayama, T., Kannapan, S., Khoong, C. M., Navathe, S., and Yates, Y. (1998). Documents, Processes and Metaprocesses in *Information and Process Intergration. In Enterprises*. Rethinking Documents. Wakayama, T., Kannapan, S., Khoong, C. M., Navathe, S., and Yates, Y (Eds.). Kluwer Academic Publishers

Artifacts in Work Practice

Submitted

Nina Lundberg and Tone Irene Sandahl

Abstract

The emphasis in this paper is to improve the understanding of work practice by focusing on artifacts' properties in work. This is done by analyzing and identifying how artifacts peripheral properties become common resources within a community of practice. Concepts from 'Actor Network Theory' and 'Borderline Issues' are applied for analytical purposes. This paper illustrates how artifacts are active elements in the relationships of people and between people and their environments. This does not only mean that they are active in the sense that they are necessary, but also, active in a way that coordinate, and even trigger or initiate, work.

Keywords: Artifacts, Artifacts' Properties, Documents, Work Practice, Actor Network Theory, Borderline Issues.

1. Introduction

"If we took away my computer, my colleagues, my office, my books, my desk and my telephone, I would not be a sociologist writing papers, delivering lectures, and producing "knowledge". I'd be something quite other - and the same is true for all of us" (Law, 1992). No doubt artifacts play an important role in our lives and in our work practice. They are initially brought into our organizations for a specific purpose, however, their use may over time become manifold and not necessarily just serving the initial purpose (Latour, 1987; Law, 1992; Akrich 1992). The artifacts in use are resources that the communities of practices rely on, however, some of these resources may be 'invisible' for designers or other outsiders. Rules and conventions develop around the use of artifact over time, this gives them their meaning within a community of practice (Brown & Duguid, 1994). In the process of design, these resources and conventions are easily overlooked due to their highly local establishment. Since, the design and development of artifacts go hand in hand with the development of work practices (Suchman & Trigg, 1991), effective design involves a co-evaluation of artifacts with practice. By focusing on artifacts and our interactions with them resources are made visible, seen and talked about in order to inform design.

The aim of this paper is to analyze and illustrate *how* artifacts peripheral properties become important resources within communities of practices. This is done by focusing on artifacts' peripheral properties, emphasizing on the correlation between peripheral properties and use of artifacts in work. We try to reveal the meaning and importance of these resources by analyzing artifacts and their peripheral properties from the point of view of those who use them. And, as Suchman (1995) pointed out the goal of making work visible for systems design is to develop more appropriate technologies from the point of view of those who will be using them. Previous systems design research has illustrated that deficiency in the understanding of work practice often contributes to computer system failure (Greenbaum & Kyng, 1991).

Two case studies are presented. The first case is from a news agency in Norway and the second from a health care institution in Sweden. Our research approach is built upon in-depth case studies as well as ethnographic studies. Interviews and observations have been applied in both studies. Both cases illustrate how artifacts, for instance fax machines, fax sheets, coordination forms, examination requests, different types of shelves and tables, as well as computers - interrelate and interact with people and how this establishes recourses over time that the communities of practice rely on in the production of the desired results, i.e. TV schedules and radiological diagnoses.

We have observed that artifacts, such as documents have similar meanings and roles in spite of organizational differences. The news agency being concerned with TV information, faxes and schedules, whereas the health care institution deals with patients, medical records and image production technology. This contributes to the significance of our joint findings as well as to the richness of our stories to be told in this paper.

2. Related Research

In systems design, it is common to consider artifacts as tools or media for human activity. When artifacts in general are viewed from the perspective of their use, they can both support communicative and instrumental activities, and they can mediate our activity towards other humans or towards 'objects' (Ehn, 1988). An artifact can augment and even replace individual or cooperative human activities (ibid.). Ehn (1988) has an Heideggerian perspective on artifacts, which means that an artifact, e.g. a hammer, belongs to the 'background', it is ready-at-hand without reflection in the carpenter's world. It must primarily be understood as a practical artifact that she uses in her everyday life, not as a thing or an object external to her. This perspective of artifacts is applied in research projects within the system design field where the issues are cooperative design of computer artifacts (see e.g. Bødker et al., 1991). The users are included into the design process based on the motivation that artifacts as such have no meaning; they are given meaning only through their incorporation into social practice. When we design artifacts they can serve to mediate the activity of the users, and at the same time become meaningful tools in the users' practice (Ehn, 1988).

Investigations into work practices in operations and control rooms have uncovered the meanings of artifacts in some respects. For instance, in their study of an airport operation room Suchman and Trigg (1991) point out the important role of the paper sheets they use, and stress the difficulties involved in replacing this paper sheet with a computer based representation. Electronic representation has some benefits. However, computerized forms have their own problems; for example, they don't allow the same ease of document transfer (Ibid.).

In control rooms such as air traffic control rooms (Hughes et al., 1994), the London Underground (Heath & Luff, 1992) or the Paris Metro (Fillipi & Therureau, 1993), the co-location of workers allows them to observe each other and to monitor the work in progress. By looking at each other's radar screens as well as listening to colleagues' conversations, they improve their understanding of what is going on, and this is necessary for the workers to carry out their own work. These studies have shown that artifacts are important in the "understanding of the activity of the others, which provides a context for your own activity", defined as awareness by Dourish and Bellotti (1992). We stress this point by focusing on how different artifacts are introduced, arranged and co-developed over time within the work itself.

Brown and Duguid (1994) stress the artifacts' social and material aspects in the framework of their 'Borderline Issues', and Latour (1987), Callon (1986), Akrich (1992) and Law (1992) point to the artifacts' properties and features in socio-technical networks. These issues are elaborated further in the following section.

3. Conceptual Framework

In order to improve our understanding of artifacts in work practice, we apply some concepts from Actor Network Theory (ANT) (see e.g., Callon, 1986; Latour, 1987; Akrich, 1992 and Law, 1992) and 'Borderline Issues' (Brown & Duguid, 1994). It has been important for us to find concepts that take artifacts seriously into account and make them explicit in social arrangements.

ANT recognizes that establishing and changing a social order relies on a tight interplay between social and technical means. ANT argues that society would not exist if it were simply social. Humans and non-humans are regarded as 'equal', and are treated in the same way, they are in ANT termed 'actants'. The concept of actants blots out the distinctions between humans and artifacts as components in the actor-network. In other words, an actor-network consists of actants that are linked together. ANT says that there is no reason to assume, a priori, that either artifacts or people in general determine the character of social change or stability (Law, 1992).

According to ANT, social settings or work practices are nothing but patterned networks of heterogeneous materials (Law, 1992). This heterogeneity emphasizes the significance of artifacts in work practices. The concept of network focuses on interconnections and relationships between humans and artifacts. The ANT's notion of inscription refers to the way artifacts embody patterns of use (Akrich, 1992). Inscriptions can be properties, i.e., features, characteristics, and possibilities inscribed in artifacts as well as meetings, institutional arrangements, skills, etc. How these properties are perceived depends upon the interpreter and her context. And, they shape the connections between different actors and therefore influence the actors' performance.

According to Brown and Duguid (1994) artifacts have both central and more peripheral properties. What is recognized as a central or peripheral property varies within different communities of practice. What Brown and Duguid (1994) define as 'Borderline issues' are shared resources that constitute a social meaning for a group of people. They are based on continuously presence of the artifacts in a community of practice. Continuity is needed in order to recognize the artifacts' properties, and community of practice is necessary for members to share, recognize and reformulate conventions (Ibid.). These resources are developed over time as artifacts are integrated into current practice and social conventions are developed. Communities of practices maintain the resources, and workers often rely on them (Ibid.).

We find the combination of these concepts fruitful. 'Borderline Issues' focuses on both central and peripheral properties of artifacts, while ANT support us in identifying the properties as well as understanding how they become resources in the socio-technical work practice.

4. Research Approach and Work Settings

The studies were conducted at a news agency in Norway, a company providing news services to the media - such as newspapers, magazines, radio and television; and in

the radiology department of a hospital in Sweden. The radiology department is a diagnostic center at the hospital, giving service to all other departments within the hospital.

Our research approach is built upon in-depth case studies at the news agency and on ethnographic studies at the radiology department. Both studies included interviews and observations of work practices. The interviews can primarily be characterized as open-ended qualitative interviews. Totally, 23 interviews and 70 hours of observations were conducted at the news agency, while approximately 30 interviews and 40 hours of observations were conducted at the radiology department. Each of the interviews lasted from 30 minutes to 2 hours.

At the news agency, the production of TV schedules was studied. A TV schedule consists of information about TV programs and when they start. The production of TV schedules starts with the production of TV program information from each TV channel. The information is sent from the channels to the news agency on faxes. At the news agency the process continues with quality control, merging and coordination of the information, as well as typesetting of eight different types of schedules. It ends in a variety of weekly products to be delivered to various newspapers and magazines.

At the news agency there are 6-9 people working with the production of TV schedules the entire week. The information on the fax-sheets is entered into a mainframe system that keeps track of the program information for each channel every day of a particular week. The schedules are formatted in a desktop program on Apple computers and delivered to magazines and newspapers by use of ftp. All of the employees perform the entering of information into the system as well as the formatting of schedules.

Once a week newspapers and magazines have TV supplements for the following week. However, which days this week covers vary from newspaper to newspaper and from magazine to magazine. Some offer a week lasting from Sunday to Saturday, while others present it from Monday to Sunday, according to the day the supplement is published. The huge amount of faxes, the different time frames for different publications, as well as the various types of schedules, make the coordination of work quite complex within the agency. A coordination form is applied in order to deal with this complexity.

The radiology department involves the administrative staff, which is the link between the radiology department and the outside world; radiographers, who are the specialist in image production; radiologists, who are the interventional and diagnostic specialists; and computer technicians, who support all the other actors with regard to computer systems. Radiological work involves distributed actors that carry out activities occasioned by a high degree of unexpected events.

In the radiology department various computer systems are used, such as PACS (picture archiving and communication system), RIS (radiology information system), and HIS (hospital information system). PACS supports the electronic storage, retrieval, distribution, communication, display, and processing of image data. In combination with HIS and RIS it provides a means for managing work associated with radiological examinations.

The medical staff at the clinical wards writes examination requests on paper. Each request includes data such as the patient's name, the date requesting the examination, the name of the clinician, the type of examination required (e.g. computer tomography, magnetic resonance, angiography, chest examination,

ultrasound, mammography, etc.) and the patients symptoms and signs as well as the clinician's preliminary diagnosis.

5. Cases: News Agency and Health Care Institution

5.1 The news agency

It is a very busy day in the Media Department of the news agency. A lot of faxes keep arriving from the TV channels in Scandinavia, as well as from various other European countries. The department's largest customer, a weekly magazine, is supposed to get its TV schedules later this day. In addition, the newspapers must have their daily delivery. However, TV information for more than half of the channels is still missing.

The fax machine is ticking. An employee is already busy entering TV information from faxes into a mainframe system. She looks around, and sees that all of her colleagues are quite busy as well. She knows that it is her turn to pick up the faxes now since the others have already done this several times today.

She walks over to the fax machine, takes the two faxes and puts them in shelves. Both faxes are from Channel 8. One of them contains information that is supposed to be included in the schedules that are to be delivered today. She puts this fax directly into the specific space, labeled "to write" in the week-shelf. The other fax contains information concerning future TV programs. She puts that fax in the in-shelf. It will later be moved to the week-shelf when it is time to process it. She goes back to her desk and continues entering TV information from the Sports Channel that she was working on before the last faxes arrived.

One of her colleagues sees that there is a new fax in the week-shelf. One less missing channel, he thinks. He has just finished entering data from Channel Z into the system, and he is looking for more work to do. He takes the Channel Z fax with him and places it in the out-shelf. He marks the coordination form to show that the information for this particular channel is registered in the mainframe system. Then he picks up the fax from Channel 8 in the week-shelf, and returns to his desk and starts entering that information into the system as well. The last faxes he has processed all contained errors in one way or another. He has called three different channels earlier today, and he hopes that this time the faxes are error-free. All these telephone calls take time, and they bore him.

A third colleague is working with a desktop program to make the TV schedule pages ready for printing. She realizes that there is a new fax in the out-shelf, which indicates that some new information is electronically available in the system. She checks the coordination form and confirms that Channel Z has been finalized. If any information from a channel is missing, this is registered on the form as well. The customer's style sheet is open on her computer. She goes into the mainframe system and copies the information into the right place on the style sheet. She does some proofreading during the formatting, and she deletes some words, rewrites some sentences, etc., in order to fit the text into the space available for that particular channel. When she is through, she marks the coordination form.

One of the writers got tired of all the errors he had handled during the day, and went out to get something to drink. While he is gone, his boss comes by and places a piece of paper on his desk. The paper contains some corrections from the Sports Channel. A representative from this channel had called the boss to ask for some last minute changes in their program. When the writer returns after his short break, he sees the

paper and starts working on the new changes at once, before he continues the work he was doing before the break. He knows that when corrections come directly to his desk, it is important. He has to make sure that the corrections are implemented.

When all the TV information has arrived, has been entered into the system, and formatted in the desktop program, the schedules are ready for delivery to the customers.

5.2 The radiology department

An employee from the medical department's administrative staff is entering the hallway in the radiology department. There are shelves on the wall to the right, and he places a document in one of the shelves. When an employee from the radiology department's administrative staff passes the shelves 10 minutes later, she glances at the shelf where the document - an examination request form - is visible. She brings the request to the administrative area, and paper in hand she enters an appointment for the patient into the RIS. She places the examination request form in another predefined special shelf visible to the radiographer out in the hallway between the administrative and image-production areas. After a glance at the shelf, the radiographer is triggered to initiate the preparation of the patient, the equipment, and the room for the examination. She carries out the X-ray examination written in the medical request. She adds information into the medical request, to inform other medical staff of her actions. Then she places the light medical request on a table visible to the administrative staff. When an administrative staff sees it, she distributes it to another table in the diagnostic area.

A radiologist in the diagnostic area has just finished diagnosing a chest examination, he looks at the table where the new examination requests are placed and realizes there are more patients to diagnose. He walks over to the table, takes the new examination requests and brings them over to the diagnostic work stations. This patient has old X-ray films from an earlier examination; these need to be compared with the images in the PACS. The radiologist stands up and walks over to the lightboard and positions the X-ray films in a row. When diagnosis is accomplished the radiologist dictates the report on a tape. He walks over to the predefined shelves for secretaries, takes the examination request and the tape and puts them in a shelf. A secretary sees that there is a new report to transcribe. She brings the tape and medical request to the administrative area and adds an examination report to the medical request. Thereafter, she returns to the diagnostic area and places the examination request in the shelf predefined for the radiologist who diagnosed the images. The radiologist sees that there is a transcribed report in his shelf. He takes the document and checks it thoroughly and signs off. Then he places it in a predefined shelf for the next day's multidisciplinary conferences, and goes back to the diagnostic workplace by the computer screens. After the images have been diagnosed a second time at the conference, they are put in an 'out-shelf', finally, the administrative staff distributes the report including the diagnosis to the medical department.

6. Artifacts in Work Practice

We have observed that humans and artifacts are interrelated in work in order to fulfill some aims or intentions. Artifacts are developed and brought into our world to support us in our work. And, we are more or less dependent on them in that respect. If we look around in our work places we see documents, binders, computers, telephones, fax machines, shelves, pens, cups, etc that we use and interact with every day. At other work places hammers, nails and screwdriver or shots, stethoscope and tweezers are

present. However, these artifacts are, in some respect, 'parts of us' (Ehn, 1988), and we cannot do our job without them.

All these artifacts have intentions, properties and features that are more or less generally accepted. For instance, the aim of shelves is to support the organization of work by sorting documents, or other artifacts, in them. The hammer is used to nail. And, documents are produced and used for the need of sharing information. However, artifacts may have more peripheral properties as well, and these peripheral properties may become important common resources that the communities of practice rely on.

In the news agency artifacts are introduced in order to organize work in such a way that the employees produce the schedules in time. Fax machines were introduced in order to make it easier and faster for TV channels to deliver their TV information to the Media department. Shelves were introduced because fax sheets needs to be separated according to channels and weeks. Shelves have predefined places with names on them, and it is a simple technology to use. The coordination form was introduced to help coordinating a complex process. Carstensen and Sørensen (1996) report that introducing some kind of formalism is a way of dealing with complex work situations.

Artifacts do not only have roles as organizers. Computers, shelves, fax-sheets and forms are visible and present in the rooms where people are working, they are also essential in the individual's understanding of the activities of the others, which provides a context for their own activities. For instance, the process of moving faxes between shelves makes the personnel aware of the status of their work as well as who is working on what. Previous research has illustrated that this awareness keeps the articulation of work on a relatively low level (Braa & Sandahl, 1998). The similar phenomenon is observed in the radiology department.

We have seen that arrangements of artifacts, as documents in shelves, are important for the awareness of human activity. And, when documents become electronic, there is no need for shelves, tables or fax machines. The visible paper documents are gone, the process of carrying them, sorting them or using them becomes invisible as well. Implicit information necessary for peoples' awareness is gone. The implicit information may preferably be represented in some other technology in order to keep the work practice together.

Dourish and Bellotti (1992) emphasized the importance of awareness in work, and we illustrate how this awareness may depend on the arrangement of artifacts in work practices. Latour (1987) and Akrich (1992) state that artifacts have politics, in the sense that "they constitute active elements in the organization of the relationships of people to each other and with their environment" (ibid. pp.1). This does not only mean that they are active in the sense that they are necessary, but also, as we have seen, active in a way that they coordinate, and even trigger or initiate, work.

For instance, documents may be coordinating artifacts. The aim of the examination request and the fax sheets are to share vital information between different communities of practices and being a carrier of information. We have seen that the documents may coordinate activities in two senses. First, documents' material and visible presence, in a shelf or on a table according to their structured trajectory in the radiology department and news agency afford the linking of actions and events over different sites and times without personal interaction between staff. When the medical request or fax sheets are distributed to a particular place, the responsibility of work is handed over to a particular community or person. In both places, the coordination of

work is indicated by 'who is holding the document'. The paper acts as a token and the shelf in which the documents are placed represents the state of work. Secondly, in the radiology department the examination request is formatted in ways that trace work, which enables various communities of practice to coordinate particular activities between themselves. This is done in a way where one actor adds information to the medical request that is both supervising and required by the next actor in order to take action. For instance the clinician demands a 'particular' examination to be carried out at the radiology department according to his/her findings of the patient. Medical records integrates distributed and isolated entries into an assembly which itself carries meaning (Berg, 1997).

According to Schmidt and Simone (1996) particular artifacts are introduced in order to manage the coordination in work. This does not only mean that particular coordination artifacts support coordination, but also, as we have seen it, some artifacts support coordination in itself. The coordinated role of the document is what keeps up the progress of work in the news agency and at the radiology departments, it is essential by means of 'keeping the work practice together'.

In a way, artifacts may be active actors as well. For instance, the administrative staff at the radiology department receives an examination request from a clinician, the administrative staff places the examination request in a special shelf visible to the radiographer. After a glance at the shelf, the radiographer fetches the request and reads the information. She then carries out the X-ray examination requested in the document, etc. We have observed that, in these situations, artifacts, such as documents on shelves, trigger human action. The fact that these documents are tangible, ecological flexible and light have implications for the ease of which they can be physically transported within the communities and laid out in particular spaces (Luff et al., 1992; Harper & Sellen, 1995). When and if documents are placed in specific locations, they represent signals. The same is observed at the news agency. What happens on the fax machine, the number of faxes received, and the shelves in which they are placed, are of vital importance to how the working day will develop.

The perspective of saying that artifacts as well as humans can be actors, in the sense that they put other humans or artifacts into actions is an analytical stance, and not an ethical position (Law, 1992). We, as well as Law (1992), do not mean that we have to, or should, treat people as machines. We do not want machines to have rights, duties or responsibilities that we usually accord to people. An artifact cannot take control over humans literary speaking. The point we want to make, however, is that artifacts may have peripheral properties that enable various use within a community of practice, and that conventions in work 'allow' artifacts to have the role as actors.

In systems design, it is common to consider artifacts as tools or media for human activity. When artifacts in general are viewed from the perspective of their use, they can both support communicative and instrumental activities, and they can mediate our activity towards other humans or towards 'objects' (Ehn, 1988). In addition, we have seen that within particular communities of practices, both fax-sheets and examination request forms are actors that trigger work; they make things happen and are in a way 'subjects' that people rely on. Therefore, in order to understand work practices, we must see artifacts as actors as well as tools and media.

We found the combination of concepts from ANT and Borderline issues useful. The combinations of the concepts from the two theories are fruitful. Borderline Issues focuses on central and peripheral properties of artifacts, while ANT contribute with a conceptual framework that both identifies these peripheral properties and illustrate how they become recourses within the practice.

7. Conclusions

We have in this paper described *how* artifacts' peripheral properties become common resources that over time the communities of practices rely on. These resources, lying as they do 'beyond the object', seem to us important to understand in the process of understanding work.

In work, artifacts and their resources have over time established a relatively stable and pragmatic work process. However, when new computer technology is introduced the old technology as well as its accompanying resources may disappear, hence the stability in work may be challenged.

We have in our cases illustrated examples of *how* particular artifacts' have translated into unnoticed common resources within a community of practice. For instance, fax sheets and examination requests do not only transfer information but also simultaneously indicate the status and progress in work, this is enabled through its peripheral properties of being tangible, ecological flexible, tailorable, light etc. The visibility of artifacts, as well as their arrangements, is important for people's awareness and coordination in work. We have seen that artifacts have been actors that triggered human activity. This implies that to consider artifacts exclusively as tools and mediums is too restricted if we wish to understand work practices. To address problems of design and use of artifacts in these changing conditions, the common resources and its role in work practice needs to be better understood – even if only to explain what life will be like without them.

Acknowledgements

References

Akrich, M. (1992). The de-scription of technical objects. In Bijker, W. E. and Law, L. (eds) *Shaping technology/building society*. MIT Press.

Bellotti, V. and Bly, S. (1996). Walking Away from the Desktop Computer: Distributed Collaboration and Mobility in a Product Design Team. In *Proceedings of the Conference on Computer-Supported Cooperative Work*. ACM Press.

Berg, M. (1997). Accumulating and Coordinating: Occassions for reading and Writing Artefacts in Medical Work. *Workshop in Department in Informatics, University of Oslo, Norway.*

Bjerknes, G., Ehn, P., and Kyng, M. (1987). *Computers and Democracy – a Scandinavian Challenge*. Aldershot.

Braa, K and Sandahl, T. I (1998) From Paperwork to Network - A Field Study. In *Proceedings from the Third International Conference on the Design of Cooperative Systems (COOP)*, Cannes, France 26-29 may 1998.

Brown, J. S. and Duguid, P. (1994). Borderline Issues: Social and material Aspects of Design. In *Human-Computer Interaction*, Vol. 9, pp. 3-36.

Bødker, S., Greenbaum, J. and Kyng, M. (1991). Setting the Stage for Design as Action. *Design at Work: Cooperative Design of Computer Systems*, Hillsday: Lauwrence Erlbaum Associates.

Callon, M (1986). Some Elements of A Sociology of Translation: Domistication of the Scallops and the Fishermen. In J.Law (Editor), *Power, Action and Belief: A New Sociology of Knowledge* London: Routledge & Kegan Paul.

Carstensen, P. H. and Sørensen, C. (1996). From the Social to the Systematic - Mechanisms Supporting Coordination in Design. In *Computer Supported Cooperative Work - An International Journal*. Vol 5., No. 4. pp. 387-413. Kluwer Academic Publishers.

Dahlbom, B. (1993). En vetenskap om artefakter. *VEST: Tidskrift för vetenskapsstudier*, vol. 6, no 4.

Dourish, P. and Bellotti, V. (1992). Awareness and coordination in shared workspeces. In *Proceedings of the Conference on Computer-Supported Cooperative Work*. October 31 to November 4, 1992. ACM Press.

Ehn, P. (1988). *Work-Oriented Design of Computer Artifacts*, Falköping: Arbetslivscentrum.

Ehn, P. and Kyng, M. (1987). The collective resource approach to system design. In Bjerknes, G., Ehn, P., and Kyng, M. (eds.), *Computers and Democracy – a Scandinavian Challenge*. Aldershot.

Fillipi, G. and Theureau, J (1993). Analysing Cooperative Work in an Urbam Traffic Control Room for Design of a Coordination Support System. In *Proceedings from the European conference of Computer Supported Cooperative Work (ECSCW)*. Kluwer.

Greenbaum, J. and Kyng, M. (1991) *Design at Work: Cooperative Design of Computer Systems*, Hillsday: Lauwrence Erlbaum Associates.

Harper, R. and Sellen, A. (1995). Paper-Supported Collaborative Work. *Technical Report EPC-1995-109.* Rank Xerox Ltd.

Heath, C. and Luff, P. (1992). Collaboration and Control. Crisis Management and Multimedia Technology in London Underground Line Control Rooms. *Computer Supported Cooperative Work (CSCW)* 1, pp. 69-94.

Hughes, J., King, V., Rodden, T., and Anderssen, H. (1994). Moving Out from the Control Room: Ethnography in System Design. *Computer Supported Cooperative Work (CSCW)* 1, pp. 429-439.

Kyng, M. (1995). Making Representation Work. In *Communication of the ACM*, Vol. 38, No. 9.

Latour, B (1987). *Science in action*. Open University press, 1997

Law, J. (1992). Notes on the theory of the Actor-Network: Ordering, strategy, and heterogeneity. In *Systems Practice*, Vol. 5. No. 4, 1992. pp. 379-393

Luff, P., Heath, C., and Greatbatch, D. (1992). Task-in-interaction: paper and screen based documentation in collaboration activity. *Proceedings of the Conference on Computer-Supported Cooperative Work*. Toronto. ACM Press.

Suchman, L. and Trigg, R. (1991). Understanding Practice: Video as a Medium for Reflection and Design. In Greenbaum, J and Kyng, M (eds). *Design at Work - Cooperative Design of Computer Systems*, Lawrence Erlbaum, 1991.

Suchman, L. (1995). Making work visible. *In The Communication of The ACM*, Vol.38, No. 9, September 1995.