

## Genre-Based Metadata for Enterprise Document Management

Anne Karjalainen  
 Faculty of  
 Information Technology  
 University of Jyväskylä  
 Finland  
[ankarjal@cc.jyu.fi](mailto:ankarjal@cc.jyu.fi)

Tero Päivärinta  
 Information Technology  
 Research Institute  
 University of Jyväskylä  
 Finland  
[tero.paivarinta@titu.jyu.fi](mailto:tero.paivarinta@titu.jyu.fi)

Pasi Tyrväinen  
 Faculty of  
 Information Technology  
 University of Jyväskylä  
 Finland  
[pasi.tyrvainen@jyu.fi](mailto:pasi.tyrvainen@jyu.fi)

Jari Rajala  
 CSC – Center for  
 Scientific Computing  
 Finland  
[jari.rajala@csc.fi](mailto:jari.rajala@csc.fi)

### Abstract

*Contemporary challenges for enterprise document management (EDM) include managing a mixture of technologies, recognizing the needs of several user roles and groups, and pursuing effective processes utilizing documents in digital form. Responding to these challenges means gathering and scrutinizing organizational metadata describing the organization's information resources. Despite the volume of metadata research in general, there continues to be a dearth of studies on metadata in the field of digital documents in the organizational context.*

*We suggest that an analysis of organizational metadata for EDM can be based on scrutinizing genres of organizational communication. This paper reports an action research project, in which this approach has its origin. Many of the genres that were identified in the target organization were implicit and implemented by heterogeneous means, i.e. "soft". In general, an organization should be aware of the continuous evolution of its document genres and should not allow an EDM system to freeze future evolution. An explicit organizational genre repertoire including metadata about the genres, is a solid basis for the design of an EDM system. As such, the action research cycle reported here represents the first attempt towards a practice-oriented method for gathering and scrutinizing genre-based metadata for EDM development.*

### 1. Introduction

The concept of organizational metadata has been used to cover all descriptions of an enterprise pursuing information resources management [10]. Organizational metadata can be used for several purposes of development, for human use and for the technological integration of the information resource [10, 7, 6, 14]. However, outside the fields of integration of structured

databases [e.g. 7, 6] and high-level corporate information infrastructures and architectures [8, 4, 12, 11], research on organizational metadata remains in its infancy [14].

The major proportion of organizational information resides in digital documents, utilized in various processes, used by several types of users in various roles for different purposes and processed in more and more technologically diverging applications [18]. Digital documents also enable increasing functionality [21]. Hence organizational metadata need to be systematically identified and scrutinized for development, human use and technological integration of enterprise document management systems (EDMS). However, most of the metadata research concerning digital documents so far has taken place in the digital library and WWW communities. These studies emphasize the use and creation of global standards for metadata [14]. Evidently, the field of organizational metadata for enterprise document management (EDM) deserves more rigorously grounded theoretical approaches and systematic development of practical methods for assisting the organizations to cope with information overload and various technological options for EDM.

This paper reports a piece of action research in which metadata related to genres of organizational communication were identified and analyzed. The metadata were primarily aimed at assisting the development of enterprise document management systems (EDMS). Hence this study focuses on document genres [15], i.e. typified, enacted and shared purposes and forms of documents occurring in recurrent situations (see also [27]) in the organization under analysis. This paper contributes by discussing a genre-based approach to capturing organizational metadata for EDM. The action research cycle reported here can be regarded *per se* as a starting point for method engineering, i.e. systematic development of methods and tools [5], towards a genre-based, methodical development of EDM.

**Table 1. The stages of this study from the viewpoints of it's practical and research goals and actions**

<b>Stage</b>	<b>Practical goals / actions</b>	<b>Research goals / actions</b>
	Main goal 1: Increased understanding of contemporary EDM. Main goal 2: A requirements definition for an EDMS.	Main goal: A method for collecting metadata about organizational document genres and experience of using a genre-based approach to EDM.
<b>Diagnosing</b>	Formulating practical goals. Gathering information about the problem setting.	Considering whether the problem situation is interesting enough from the viewpoint of research.
<b>Action planning</b>	Comprehending and agreeing on the method and techniques to be used in the project. Agreeing on the schedule. Assigning responsibilities for carrying out the project in the target organization.	Explicating the method and techniques to be tried out in the target organization. Deciding on research techniques for evaluating the developed method.
<b>Action taking</b>	Using the method: - identifying who should participate - identifying genres - gathering metadata about the genres - analyzing the metadata and formulating a requirements definition for the future EDMS	Facilitating and observing the use of the method.
<b>Evaluating</b>	Evaluating the results produced by the method with respect to the practical problem setting. Performing a follow-up study of the implications of the project and discussing the results within the organization.	Evaluating the use of the method and techniques: - a follow-up study through dialogue with the representatives of the organization - comparing the observations made during the action taking stage with the intended use of the method
<b>Specifying learning</b>	(Discussing the practical results of the project - performed on the evaluating stage.)	Structuring the experience gained for further research purposes. Elaborating the method.

Action research has been recognized as a relevant, *a posteriori*, research approach for method engineering. In this way, in-depth experience of a method can be gathered for rigorous development of the method [1, 22, 24]. A typical action research cycle consists of five stages: **diagnosing, action planning, action taking, evaluating, and specifying learning** [23]. In method engineering, every stage inherently includes the aspect of developing the method in question and testing it in a target organization. Table 1 summarizes the stages of this research from the viewpoints of research and practice.

The rest of the paper is organized according to the five stages (table 1). Diagnosing is discussed in section 2 (problem setting in the target organization), action planning is discussed in section 3 and action taking in section 4. Section 5 discusses the evaluating stage and section 6 summarizes what was learned from the viewpoint of research. Finally, we suggest some areas for further research.

## 2. Problem setting in the target organization

The CSC (Center for Scientific Computing) is a non-profit organization consisting of 80 specialists providing services in several areas of expertise related to scientific computing and communication [30]. For instance, the CSC provides the services of FUNET, the national university and research network in Finland. The CSC is organized into nine expertise groups consisting of seven

groups for different areas of technological expertise, a group for communications and a group for administration.

An effort to improve EDM was initiated in January 1998. The CSC established an EDM team aiming at shared archiving practices and improved utilization of organizational databases. Another objective was to improve collaboration and project practices among the groups and individual specialists.

The CSC's specialists operate with several kinds of personal computers, workstations, supercomputers, Internet servers and operating systems. The practices of document processing, delivery and reuse were neither effective nor easy to standardize. The employees did not have a clear understanding of the information produced by the other employees despite of large number of digital information storages within the organization. Locating and accessing documents from multiple servers incorporating complex directory structures was problematic as the amount of recorded information had exploded. Many employees had unofficial, personal versions of documents on which they relied in their jobs. Documents were stored in more than twenty file formats, preventing effective (re)use. Guidelines for document management - such as using common templates, systematic backup routines and unified versioning procedures - were incomplete, covering only a small proportion of all documents.

The EDM team decided to seek assistance in clarifying the problem and to obtain an external opinion about the development of EDM. Hence the CSC contacted a group of university researchers for help. Since the researchers

considered the problem setting in the CSC a good opportunity to further experiment with and evaluate a genre-based approach to EDM [for earlier studies see e.g. 25, 17], a cooperative project started in May 1998. The EDM team set their practical aims as increased understanding of the contemporary state of EDM and as obtaining a requirements definition for a future EDMS implementation. The action research project was carried out between September 1998 and May 1999. The CSC regarded the project as a starting point towards the creation of a systematic EDM to be continued after the project by the CSC EDM team.

### 3. Action planning

#### 3.1. Explicating the method

The researchers elaborated a genre-based method for gathering and analyzing organizational metadata for EDM. Experiences from earlier studies [25, 17] guided the construction of the method to be piloted in this project. The researchers planned the action together with the EDM team from the CSC. Figure 1 illustrates the process of the planned method by utilizing Yourdon's data flow diagram [29].

#### 3.2. Step I: Identifying units producing and using information

The first step includes the identification of organizational units that produce or use information.

These units serve as the starting point for the identification of the genres in the next step. In addition, external stakeholders (e.g. customers, subcontractors, and municipalities), with whom the units exchange information, are identified. The step results in a list of the units, their contact persons, and external stakeholders.

#### 3.3. Step II: Identifying the genres

The second step aims at identifying and naming the genres relevant to EDM, i.e. defining the *genre repertoire* [16] of documents in the organization. It has been suggested that a genre can be altered, versioned or merged with other genres along with organizational changes or incorporation of new technologies [e.g. 26, 28]. Hence there may exist explicitly defined and widely comprehended, "hard", genres - as well as more or less implicit or "soft" genres of documents and other information for which different people (even those expected to share that genre) have different connotations [20, 28, 3, 25]. Hard document genres can be characterized by rather uniformly identified metadata values whereas soft genres are more implicit. The more people participate in defining the rules and norms applicable to a genre, the "harder" it gets [20]. In this case, obtaining metadata about hitherto soft genres of documents and other information in the CSC was considered relevant by the researchers. If more than one employee agreed on a name for a piece of information [13] it was considered as identified genre.

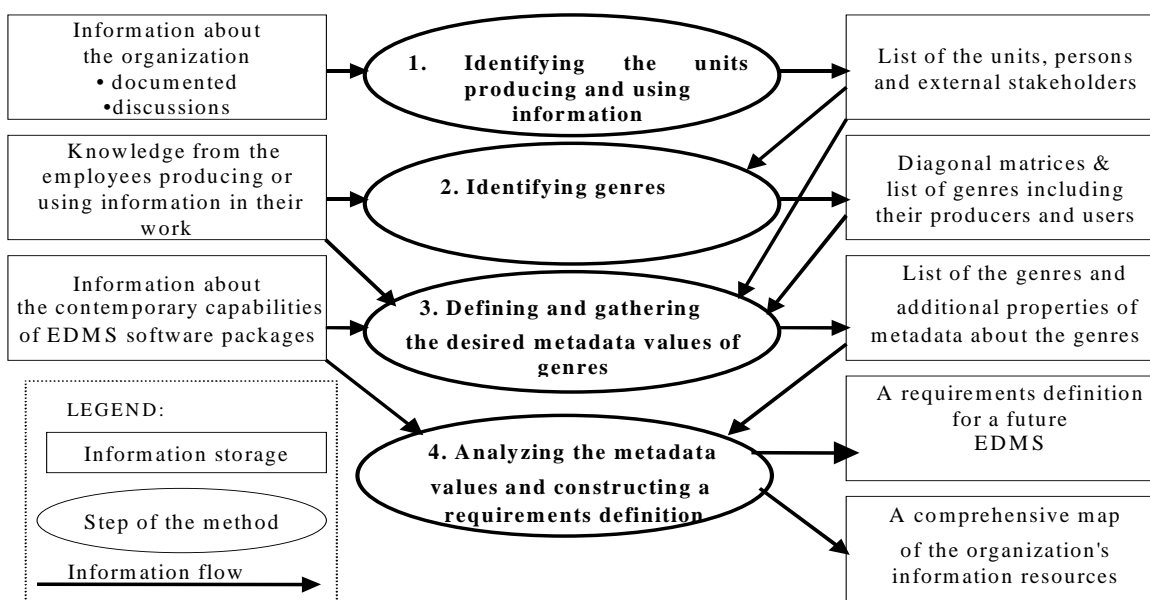


Figure 1. An overview of the preliminary method

The researchers concluded that the method should avoid the use of complicated terminology, but should nonetheless be able to capture the soft document genres as well. The means for obtaining the anticipated information is the formulation of questions. One should ask question like “What information do you produce for others?” instead of “What documents do you produce?”. In this method the document genres are to be identified later on from the set of named information elements produced by the participants.

The genres are identified in collaborative sessions with the representatives of the organizational units. In each session a diagonal matrix model is collaboratively built up on a wall [19] (Appendix 1 illustrates the idea). The model consists of the subunits of the organization and the information elements produced or used in the units identified in the first step. If the organization is small the subunits are replaced by the main tasks of the units. The hard and soft genres are identified and named in connection with the producers and users of the genres. The genres are then transferred from the resulting diagrams to a spreadsheet for the next step.

### **3.4. Step III: Defining and gathering the desired metadata about the genres**

In the third step, additional metadata about each genre are defined and transferred to the spreadsheet (Appendix 2). Within this project, we did not attempt to find universal metadata categories for all the imaginable purposes. Instead we focused on collecting information about the current state of EDM and on formulating a requirements definition for a future EDMS. Thus the metadata to be collected about each genre should contribute to reaching these practical goals.

Three main classes of metadata were considered relevant to the practical goals stated:

- 1) metadata describing the users, producers and the use frequency of the genres (to be used as a basis for organizing documents in the future EDMS, for access rights definitions and for determining those using or producing most instances of a particular document genre);
- 2) metadata describing the current technological implementations of a genre (to be used in defining the technological architecture of the future EDMS and the needs for the interoperability of specific applications); and
- 3) metadata about processing needs of a genre (to be used in defining the functionality performed by the future EDMS, such as workflow management, archiving or versioning procedures).

The metadata is collected by a form to be filled in by all the employees willing to participate in the project (Appendix 2). Instructions on filling in the form are given to the participants. The form is tailored for each organizational unit by prefilling it in with the information gathered on the previous step of the method. In the project the forms were printed on paper sheets to be filled in.

For the project another questionnaire is created for gathering opinions from the organization’s employees about the functionality of a future EDMS. This questionnaire also contains open questions on document archiving procedures and on responsibilities and policies for security and EDMS maintenance. These general questions are intended to complement the genre-based analysis and to foster discussion on the decisions that would need to be taken before implementing an EDMS.

### **3.5. Step IV: Analyzing the metadata and constructing a requirements definition**

The fourth step aims at analyzing the genre-based metadata in the light of the project’s goals. The spreadsheet serves as a means for doing this by documenting the genres and the metadata describing them. The genres can now be easily classified and clustered according to their metadata values. For instance, the spreadsheet can be sorted according to the producers of the genres to see what genres are produced by a certain unit, or according to the applications to see which genres are affected if certain applications are replaced. As it is, the spreadsheet offers a holistic genre-based map about the information resource.

In the diagnosing stage general requirements for EDMS functionality are preliminarily discussed. The analysis thus tries to identify the most frequently used document genres and the functionality emerging or needed for these genres. The analysis should also find needs to manage genres that are for some other reason critical for the organization, such as the documents supplying evidence in legal matters. Furthermore, the analysis should identify the most important soft genres to be elaborated and standardized towards harder document genres to be managed in an EDMS. The requirements definition for the future EDMS should be balanced according to the metadata values obtained for the document genres.

## **4. Action taking**

### **4.1. Step I: Identifying the units producing and using information**

Most of the information needed for this step had already been gained during the diagnosing stage. The nine expertise groups of the CSC were chosen to the organizational units to be analyzed in the project. The genres were grouped and analyzed primarily according to the expertise groups in organization and secondarily according to the general communicative purpose or content of a genre. More detailed subunits, i.e. tasks for each group and external stakeholders were identified in collaborative sessions during the next step.

#### **4.2. Step II: Identifying the genres**

The collaborative sessions were carried through as planned, one in each expertise group. Two researchers assisted each session. Two to six representatives of the expertise group in question identified and named the main tasks of the group and the genres produced and used as a result of those tasks, and the genres that are shared with other expertise groups or external stakeholders. The sessions took two to four hours. After all the sessions the researchers transferred the 850 genres identified to the spreadsheet with information about the producers and users of each genre.

#### **4.3. Step III: Defining and gathering desired metadata about the genres**

The researchers tailored the prefilled forms to each expertise group to gather metadata about the genres. In seven groups the forms were filled in individually by four to six employees, whereas two groups chose to conduct a collaborative session for this purpose. Altogether, 40 employees (one half of the personnel) participated in this step. The filled in forms and feedback were returned in closed envelope to a CSC intermediary, who collected and returned the envelopes to the researchers. The researchers coded the data from the forms onto the spreadsheet. Filtering the spreadsheet according to the metadata values formed groupings of similar genres. Altogether the spreadsheet consisted of 850 genres each with 19 metadata values. The high number of genres named was a surprise for both the researchers and the EDM team. Thus this step required more effort than expected on the part both of the participants in completing the questionnaires and of the researchers in encoding the metadata values into digital form.

#### **4.4. Step IV: Analyzing the metadata and constructing a requirements definition**

A significant part of the identified genres had not been recognized as "documents" by every group or employee.

For instance, a '*customer consultation*' was recorded and delivered as a html document in the Web by specialists, but sometimes sent and archived by an electronic mail system, and sometimes (if the customer contacted the specialist for consultation by phone) simply communicated orally on the phone with neither the question nor the answers recorded anywhere.

The researchers formed various groupings of the genres identified on the spreadsheet according to the metadata values. For example, the EDM team could now analyze which genres required versioning or workflow functionality, which genres had the greatest frequency on production or use, which genres were recognized and shared among the most expertise groups etc. These groupings and listings were used in creating the balanced requirements definition for the future EDMS together with the results from the questionnaire about general functionality.

The researchers wrote a report about the current state of EDM in the organization. The report included declaration of user roles, processes and technologies (operating systems and software packages) related to the document genres identified. The report also listed development needs for EDM.

The researchers also gave a recommendation for a future EDMS. The report discussed the identified development needs with regard to the common capabilities of commercial EDMS software packages. Also, the needs for the functionality of an EDMS were described. The balanced requirements definition, based on the genres grouped and listed as described above, formed the basis for this recommendation. The report also included a comparison of seven EDMS software packages with regard to the requirements definition.

One of the researchers categorized the genres further, primarily according to the expertise groups and secondarily according to their content. This categorization was adopted by the EDM team as the basis for architecture to logically organize documents in a future EDMS. It was considered important to elaborate the guidelines for processing and using the most important genres. For instance, the guidelines should include instructions for the templates to be used and for the groupwork procedures related the genres. Both automated support provided by the future EDMS and manual processing of the most important genres were to be considered in the guidelines.

### **5. Evaluating**

#### **5.1. Collecting the follow-up data**

Follow-up data from the CSC was collected in four ways:

- 1) during the action taking stage with feedback forms filled in by the participants;
- 2) in a feedback discussion by the EDM team and one of the researchers after the action taking stage;
- 3) by a separate follow-up session with the leader of the CSC EDM team (also a co-author of this paper); and
- 4) by interviewing the managing director of CSC after the project [9].

The researchers also made observations and notes about the collaborative sessions that they assisted in. The follow-up aimed at discussing what was learned, and obtaining information about the actual use and usefulness of the obtained results.

The evaluation was made with both practical and research objectives in mind. Both of these are discussed in the following subchapters.

## 5.2. Practical results

Regarding the first practical goal of gaining increased understanding about contemporary EDM in the CSC, the analysis of metadata values revealed that many of the challenges of EDM were not primarily technological. Instead, a great number of the problems identified could be solved by recognizing and enacting the most important document genres and by agreeing on common practices for the production, storage, processing and use of the genres - even in fact by using technologies already present in the organization. This finding was warmly welcomed in the CSC, since it is a policy of the organization that technological standards should not override the aspects of organizational development and specialized needs of the experts' [9].

The list of the 850 genres identified together with their producers and users now provided concrete answers to such questions as which information should be shared and among whom. The list *per se* illustrates what information exists in the CSC and whom one could ask about it. The list was also regarded as an informative basis for architecture to organize and categorize documents in the future EDMS. The first version of this architecture, even if only quickly formulated by one researcher, was regarded as a good starting point. After the project, the EDM team evaluated this architecture further. The identification and analysis of the most complex (soft or heterogeneous by their properties) genres raised the question of how much and what kind of variations in document processing should be allowed, and who should decide on the technologies constraining non-standardized practices for document production and use. The list thus enabled the EDM team to recognize many aspects that hindered shared and effective utilization of documents at the organizational level. The EDM team designed instructions for the most complex document genres after the project.

As a result, understanding of EDM was increased throughout the organization.

The second practical goal was to create a requirements definition for the future EDMS. The EDM team regarded the genre-based approach to determining the functionality needed in the future EDMS as a good foundation for doing this. The questionnaire vis a vis the desired functionality for the future EDMS was considered useful for shedding light on the capabilities of contemporary EDMS software packages. By viewing this questionnaire even employees with no previous expertise in EDMS packages could now imagine at least some of the technological capabilities available for the genres they were familiar with. The project was considered successful in clarifying and prioritizing the functional capabilities for the future system by balancing the identified needs and the genres. Hence the EDM team will be able to use the requirements definition in selecting alternative software packages for more detailed technical testing.

## 5.3. Evaluating the use of the method

Despite the fact that the representatives of the CSC were satisfied with the practical results obtained, both the EDM team and the researchers observed several potential areas for improvement in the method. We consider the method in relation to the practical results and the experiences of the researchers. Tolvanen [24 p. 212] has suggested four method engineering mechanisms consisting of the evaluation of : 1) the conformity of the steps taken during the action taking stage with regard to the method as planned beforehand, 2) the problem solving capabilities of the method with regard to the problem setting, 3) the modeling capabilities of the method with regard to the problem setting, and 4) the rationale for choosing the method in the initiative.

The method was used according to the steps planned in advance. The participants considered the diagonal matrix technique an effective and rigorous way to identify and name genres. The greatest difficulties were faced in gathering the metadata values of the genres, in encoding the metadata into digital form, and in analysing the metadata. Filling in the questionnaire, including putting down numerous document genres with several possible metadata values was considered too laborious by the participants. This task should somehow be made easier. As the great number of genres identified in the CSC was a surprise to the researchers, the encoding of the gathered metadata into a digital form was seen as laborious. Moreover, the considerable "softness" of many genres surprised the researchers and thus complicated and slowed down the analysis of the metadata, thus deviating from what had been planned.

Broad participation in the activities of the project was achieved. This was regarded as an essential factor in successfully obtaining comprehensive results. The results provided by the method met the practical goals of the target organization.

The diagonal matrices served as an illustrative tool for collaborative discussions to identify and name the genres. The diagonal matrix technique was also comprehended quickly by all participants, even those having no earlier experience in modeling information systems. However, ability to describe interrelationships between genres, i.e. genre systems [2], is lacking in this technique. The analysis tool implemented in an ordinary spreadsheet application offered a versatile way of organizing the genres according to any metadata value. The spreadsheet now offers a foundation for further development initiatives in defining the technological capabilities needed for particular genres, e.g. for identifying workflow management and versioning needs, and proper scopes for access rights. Thus the modeling capabilities of the method satisfied the needs of the initiative, except for modeling genre systems.

The rationale for constructing the steps in this method was to grasp the information resources of the CSC in a holistic and in-depth manner, but with reasonable effort. Despite the surprisingly great number of genres found and a greater than estimated amount of work required, the researchers considered the rationale of the method to be adequate in this case. However, in a larger organization, the method should be applied in each unit of analysis independently, each defining their internal genres and the genres exchanged with the others – otherwise it may not be possible for two or three assistants to coordinate this process in its current elaborate form.

## 6. Specifying learning

This section discusses the observations made of the target organization and aims at suggesting some of the ways in which the genre-based approach and method could be further developed to have greater applicability to other organizations.

### 6.1. Implications of the method tested in the study

The first trial of the method encourages us to develop it further. However, a couple of modifications are recommended. As mentioned, the ability to identify and model genre systems needs to be developed to illustrate interrelationships between genres.

The paper-based diagonal matrices could be replaced with an electronic tool to reduce the amount of manual

work. Such a tool could be supported by certain method engineering capabilities [e.g. 5, 24] to tailor the technique to the organization as necessary. The paper-based forms could also be replaced with an electronic tool to capture genre based metadata values in digital format for the analysis.

Means for discussing and enacting the identified "soft" information genres are needed before gathering metadata values. At the CSC, the researchers and the EDM team found about 150 important document genres, either already "hard" document genres or "soft" genres to be standardized towards "hard" genres. Thus this project would have saved considerable effort by scrutinizing and unifying the names and connotations of the most important genres before gathering the metadata about all 850 genres.

### 6.2. Implications for a genre-based approach to EDM in general

This study confirms that many organizational genres of documents and other information are implicit and soft, often implemented by various technological means. EDM should not be seen as a single technological application to be adopted in an organization but rather as a comprehensive framework for managing the information resources, including both human and technological systems, in a holistic way.

The study also confirms that an explicit genre repertoire with metadata values describing the genres provides a terminology capable of bridging discussion about the social and organizational aspects of EDM with its technological implementations. The genre-based approach and the method described in this paper also succeeded in bringing "soft" genres into the discussion about the future EDMS.

On the basis of our experience in this case, it seems likely that many genres first emerge as "soft" genres, later attracting greater awareness and becoming part of the common consciousness and formal procedures of the organization; i.e. become institutionalized. The new genres may be enabled by the adoption of new advanced technology to support existing communication needs, such as the use of an Intranet in the internal communication of the organization. In most cases, new forms of communication emerge with changes in the communication practices of the organization that can be fostered by development or elaboration of processes of the organization or by changes in the organizational culture or mission. In any case, the genres to be managed by an EDMS are likely to evolve continuously with new soft genres emerging, elaborating and institutionalizing into hard genres.

On this basis we might speculate that if EDMSs were developed only from the technological viewpoint they would easily include only the hard document genres recognized at the moment and mostly ignore the soft genres. Thus new forms of communication would easily become blocked by the existing technical implementation preventing the development of such new communication procedures as "non-standard" or "non-supported" forms of communication. In the short-term, this would probably decrease the use and effectiveness of the EDMS due to emergence of unofficial information management systems alongside it and, in long-term, could be fatal for information management in the organization in general.

On the basis on this study we can state that as a means to comprehend an information resource the genre-based approach seems to provide at least a complementary, or even an essential, counterpoint to high-level approaches that incorporate abstract concepts such as: 'technological infrastructures', 'strategic alignment', 'business processes' or 'maxims of information technology management'.

## 7. Directions for further research

This study reports the first trial of a method constructed for the needs of a target organization. Additional effort is now needed to generalize the method further. Identification of universal, contingent and situation-dependent metadata classes regarding genres is an interesting line of research especially in promoting methodical analysis of organizational genres of documents and other information. More illustrative methods for modeling organizational genres and genre systems, e.g. using an object-oriented approach to model genres with similar properties as object classes are also possible avenues of future research efforts.

## Acknowledgements

We wish to thank Esko Hakulinen, Tuomo Peltola and Airi Salminen for their efforts related to this research. We also owe our thanks to the personnel of the CSC, especially the Managing Director Matti Ihamuotila, for their inspiring support, participation and ideas. The Finnish Technology Development Centre has funded this research in METODI and inSGML projects [30].

## References

[1] R. Baskerville, A.T. Wood-Harper, "A Critical Perspective on Action Research as a Method for Information Systems Research", *Journal of Information Technology* 11, 3, 1996, pp. 235 – 246

[2] C. Bazerman, "Systems of Genres and the Enactment of Social Intentions", in A. Freedman and P. Medway (Eds.), *Genre and the New Rhetoric*, Taylor & Francis, London, 1994, pp. 79-101

[3] M. Bergquist, J. Ljungberg, "Genres in Action: Negotiating Genres in Practice", in *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences; Genre in Digital Documents*. IEEE Computer Society Press, Los Alamitos, CA, 1999. CD-ROM.

[4] J.C. Brancheau, J.C. Wetherbe, "Information Architectures: Methods and Practice", *Information Processing and Management* 22, 6, 1986, pp.453 – 463

[5] S. Brinkkemper, "Method engineering: engineering of information systems development methods and tools", *Information and Software Technology* 38, 1996, pp. 275 – 280

[6] W. Cheung, C. Hsu, "The Model-Assisted Global Query System for Multiple Databases in Distributed Enterprises". *ACM Transactions on Information Systems*, 14, 4, 1996, pp. 421 – 470.

[7] C. Hsu, M. Bouziane, L. Rattner, L. Yee, "Information Resources Management in Heterogeneous, Distributed Environments: A Metadatabase Approach", *IEEE Transactions on Software Engineering*, 17, 6, 1991, pp. 604-625.

[8] IBM, *Business Systems Planning: Information Systems Planning Guide*, International Business Machines Corp., Report GE20-0527-4, 4. edn., 1984.

[9] M. Ihamuotila, interview. E-mail communication Karjalainen-Ihamuotila, 25.05.1999.

[10] L. Kerschberg, D. Marchand, A. Sen, "Information System Integration: A Metadata Management Approach", in (eds.) C.A. Ross, E.B. Swanson, *Proceedings of the Fourth International Conference on Information Systems (ICIS)*, Dec 15 – 17, Houston TX, 1983, pp. 223 – 240.

[11] R. C. King, M. J. Shaw, M.J., "Information Infrastructure for Enterprise Coordination and Integration: Introduction to the Special Issue", *Journal of Organizational Computing and Electronic Commerce*, 9, 1., 1999, pp. 1 – 5

[12] J. M. Martin, *Information Engineering*. Volume 2. Prentice-Hall, Englewood Cliffs NJ, 1990.



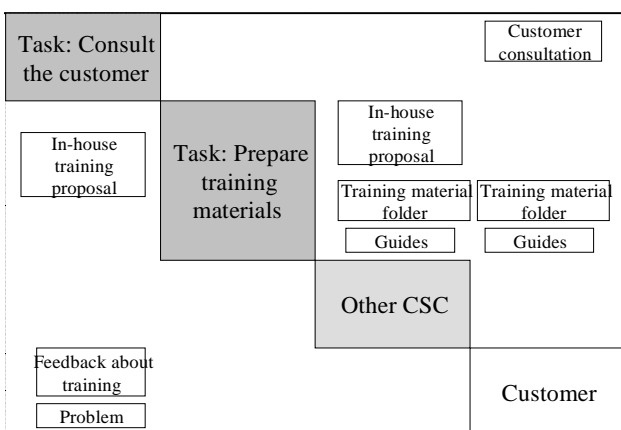
- [13] C. R. Miller, "Genre as Social Action". *Quarterly Journal of Speech* 70, pp. 151-167, reprinted in A. Freedman & P. Medway (Eds.) 1994. *Genre and the New Rhetoric*, London: Taylor & Francis, 1984, pp. 23-42.
- [14] L. D. Murphy, "Digital Document Metadata in Organizations: Roles, Analytical Approaches, and Future Research Directions", in *Proceedings of the 31st Hawaii International Conference on System Sciences: Digital Documents*, IEEE Computer Society Press, Los Alamitos CA, 1998. CD-ROM.
- [15] G. Nunberg, "Genres in Digital Documents: Introduction", in *Proc. of the 30th Annual Hawaii International Conference on System Sciences: Digital Documents*, Vol VI, Los Alamitos CA : IEEE Computer Society Press, 1997, pp. 2
- [16] W.J. Orlikowski, J. Yates, "Genre repertoire: The Structuring of Communicative Practices in Organizations", *Administrative Science Quarterly* 39, 4, 1994, pp. 541 - 574
- [17] T. Päivärinta, "A Genre Approach to Applying Critical Social Theory to Information Systems Development", in (ed.) C. Gilson, *Proceedings of the 1st Critical Management Studies Conference; Information Technology and Critical Theory –stream*, Manchester, England, July 14-16, 1999.  
URL: <http://www.mngt.waikato.ac.nz/ejrot/cmsconference/cmsdefault.htm>.
- [18] T. Päivärinta, P. Tyrväinen, "Documents in Information Management: Diverging Connotations of "a Document" in Digital Era", in *Proceedings of the 9th Information Resource Management Association International Conference*, Boston MA, May 17-20, IdeaGroup, Hershey PA., 1998, pp. 163 – 173.
- [19] K. Saaren-Seppälä, *Seinätekniikka prosessien kehittämisessä*. (Using the wall-chart technique in process development, in Finnish), Kari Saaren-Seppälä Ltd., Finland, 1997.
- [20] U. Schultze, R., Jr. Boland, "Hard and Soft Information Genres: An Analysis of two Notes Databases", in: *Proceedings of the Thirtieth Hawaii International Conference on System Sciences. Volume VI. Digital Documents Track*. IEEE Computer Society Press, Los Alamitos, California, 1997.
- [21] M. Shepherd, C. Watters, "The Functionality Attribute of Cybergenres", in *Proceedings of the 32nd Annual Hawaii International Conference of System Sciences: Digital Documents, Minitrack: Genre in Digital Documents*. Los Alamitos CA: IEEE Computer Society Press, 1999. CD-ROM.
- [22] K. Siau, M. Rossi, "Evaluation of Information Modeling Methods – A Review". In *Proceedings of the 31st Annual Hawaii International Conference of System Sciences: Modeling Methodologies and Intelligent Systems*. Los Alamitos CA: IEEE Computer Society Press, 1998. CD-ROM.
- [23] G.I. Susman, R.D. Evered, "An Assessment of the Scientific Merits of Action Research", *Administrative Science Quarterly*. Vol. 23, N:o 4, 1978.
- [24] J.-P. Tolvanen, *Incremental Method Engineering with Modeling Tools*. Doctoral dissertation, Jyväskylä studies in computer science, economics and statistics 47, University of Jyväskylä, Finland, 1998.
- [25] P. Tyrväinen, T. Päivärinta, "On Rethinking Organizational Document Genres for Electronic Document Management", in *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Los Alamitos CA, 1999. CD-ROM.
- [26] J. Yates, *Control through Communication: The Rise of System in American Management*, Johns Hopkins University Press, Baltimore, 1989.
- [27] J. Yates, W.J. Orlikowski, "Genres of Organizational Communication: A Structural Approach to Studying Communication and Media", *Academy of Management Review* 17, 2., 1992, pp. 299 - 326.
- [28] J. Yates, W.J. Orlikowski and K. Okamura, "Explicit and Implicit Structuring of Genres in Electronic Communication: Reinforcement and Change of Social Interaction", *Organization Science* Vol. 10,1, 1999, pp. 83-103.
- [29] E. Yourdon, *Modern Structured Analysis*, Prentice-Hall, Englewood Cliffs NJ, 1989.
- [30] Participatory organizations and projects:
- CSC. <http://www.csc.fi/>
  - TEKES (funding). <http://www.tekes.fi/eng/default.asp>
  - METODI project. <http://www.titu.jyu.fi/Methodi/>
  - inSGML project. <http://www.cs.jyu.fi/~airi/docman.html#inSGML>

**Appendix 1. The diagonal matrix technique [19]**



**Figure 2. An ongoing group session**

A group session is illustrated on Figure 2. First, white A1-sized sheets of paper are hung on a wall to form a big paperboard. The participants identify the subunits or the main tasks of the unit in question and other units with whom information is exchanged. The tasks and named units are written on colored cards, one task or unit per card. The cards are then placed on the diagonal of the board starting from the upper left corner. The participants then identify and name genres shared between the units. The names of genres are written on rectangular cardboard scraps, which are positioned with respect to the units that produce and use the genre. The resulting model (Figure 3) is read clockwise: an instance of a genre is produced by a unit which is horizontally at the same level as the genre itself and it is utilized by a unit that is vertically in the same column as the genre. Cards and scraps are attached to the board with pins or blue-tack mass. This way the model can be altered when needed.



**Figure 3. An imaginary (and simple) diagonal matrix**

**Appendix 2. List of genres including metadata values**

<b>Storage format</b>	.html, .doc			
<b>Media to share</b>	paper, internet		e-mail	paper, internet, phone
<b>Produced by (software?)</b>	MS Word		MS Word	Latex, Emacs, Netscape, E-mail, orally
<b>Operating system</b>	Win NT		Win NT	Unix, Win NT
<b>Approved by whom?</b>	Train. (PTM)		Train	Author
<b>Versioning? (y/n)</b>	n		n	y
<b>Changes during life-cycle? (y/n)</b>	y		y	y
<b>Retrieved by</b>	topic		genre	topic, content, customer
<b>Use freq. per week</b>	?		1	40
<b>User</b>	CSC, Customer		CSC, Train (CTC)	Customer, CSC, Customer
<b>Prod. freq. per week</b>	4		0.17	9
<b>Producer</b>	Train. (PTM)		Train. (PTM)	Train. (PTM), Train. (CTC)
<b>Genre</b>	Training material folder		In-house training proposal	Guides, Customer consultation

**Figure 4. Part of a list of possible genres and their metadata values**

Each row in Figure 4 contains the metadata class name on the left (for example "Producer"), and the metadata values on the following columns. The genre names are on the bottom of the columns. Each column responds to a genre identified. The list presented here is simplified to better demonstrate the idea. The actual genre list of the CSC included 850 genres and 19 metadata values for each.

The list can easily be (re)organized according to the desired metadata value of a genre in a spreadsheet application. For instance, if one wants to see which genres are intended for the whole organization's use she or he can filter and organize the 'User' column by the acronym used of the whole organization (f.ex. by the acronym "CSC"). On the other hand, the 'versioning' column (with values "yes" or "no") is useful for identifying which kind of documents should be put under automatic version control in a future EDMS.