Building Digital Government by XML

Airi Salminen
Department of Computer Science and Information Systems
P.O. Box 35 (Agora), FIN-40014 University of Jyväskylä, Finland
airi.salminen@jyu.fi

Abstract

Continuing innovations in information and communication technologies offer powerful tools for building digital government but, at the same time, in many environments they have lead into a number of heterogeneous, expensive, and inconsistent solutions. XML offers a common metalanguage and terminology to develop means for system and data integration, and for gradual transfer to more consistent formats in information assets. The paper describes ways for the use of XML in public administration and gives examples of the use, particularly, in Finland. The paper introduces XML standardization levels and types in public administration. Experiences of the long-term standardization of the Finnish parliamentary documents will be described. The work for building Finnish digital government continues in finding ways for the data and system integration by means of metadata standardization.

1. Introduction

The globalization of communication networks and capabilities offered by web technologies are significantly changing the work and services in public sectors. The term *digital government* refers to the possibilities to utilize current and future information and communication technologies (ICT) effectively to build new kinds of services both for people working in public sectors and for people needing their services. The possibilities also create hard challenges. In multicultural environments like in the European Union public administration is a complex network of organizations, people, languages, information systems, information structures, rules, processes, and practices. Effective utilization of ICT requires explicit rules for communication and means for the integration of heterogeneous systems and information resources. XML is a tool for the purpose.

2. The use of XML in public administration

Extensible Markup Language, shortened XML [2], consists of a set of rules for defining and representing information as *XML documents* where information structures are indicated by explicit markup. The markup vocabulary and the structures specified for a particular domain create an *XML application*, a formal language for representing information of the domain.

2.1. Origins of XML

XML was developed from the Standard Generalized Markup Language (SGML, [10]) for supporting the management of heterogeneous information resources of the Internet and to facilitate communication between various software applications. Compared to SGML, the number of rules in XML is clearly smaller and its character set and resource identification mechanisms are planned for use on the Internet, for represententing information content written in all natural languages of the world. The simplicity of XML has encouraged active development work around XML, including both software development and development of XML applications and related languages. Where SGML has been primarily in use as a format for documents intended for human readers, for example, in the form of HTML documents, the use of XML has extended towards data interchange between software applications. In public administration, likewise in other domains, the use of XML can be divided into two major categories: the format for data interchange and the format for information assets. The information assets can be further divided into documents and metadata.

2.2. XML as data interchange format

Flexible data interchange is important in contemporary global and local networks connecting heterogeneous systems, data resources, and organizations. Following are typical cases for data interchange:

- integration of systems by a common user interface,
- integration of services by a portal,
- data exchange between software systems within an organization, or
- data exchange between software systems in different organizations.

In all these cases XML offers a syntax and metalanguage for developing a common format for data exchange. Data exchange between public and private sector organizations has earlier been implemented by the use of electronic data interchange (EDI) standards. For example, an EDI connection between the Finnish Parliament and a publishing house was created in the begin of 1990's. Traditional EDI standards are gradually replaced by XML-based formats both in commerce and in public administration.

An example of a public domain service utilizing XML for data exchange is Suomi.fi, a portal providing access to e-government services in Finland. In the portal the various public services are classified and described. Information related to counties is updated daily automatically by the interconnection of the portal and a system of the county union. In this interconnection data is exchanged in XML form [15].

XML as the data exchange format is used for improving e-government services. Together with the expansion of the European Union, integration of information repositories written in various European languages and public sector services built in different European countries is needed more than ever. There is also urgent need to improve data exchange between European organizations, for example, between customs offices and police forces in different countries, and legislative authorities at different levels of the European Union. Having a tool to interconnect existing, heterogeneous systems is important also for extending the life time of legacy systems and legacy information repositories. This is especially important within the European Union when new countries with their own legacy systems have joined the union and new countries will be joining also in the future.

2.3. XML as document format

Considering XML as the format for information assets in public administration, the assets fall in most cases into one of the two categories: the documents created in public administration processes and the metadata related to documents. In using XML (or SGML) for the documents, the approach of *structured documents* is adopted in document management of the organization. In the approach, a set of documents of a specific type defined by a Document Type Definition (DTD), or by some other schema definition language, can be regarded as a database for the schema. The constraints defined in the schema can be examined either at the time of document creation, or at the time of document storage. Queries utilizing the information structure may be used for information retrieval. External representation for the documents or their parts can be specified by separate stylesheets. There are various kinds of implementations for managing a collection of structured documents. Some implementations are based on file storage, others on XML extensions of traditional database systems or on native XML databases. The approach of structured documents has well-known benefits. The approach supports

- consistency and correctness,
- rich information retrieval capabilities,
- information reuse and multichannel publishing,
- independency of particular software providers, and
- long-term accessibility of information stored in documents.

These well-known benefits have motivated, already before XML existed, the initialization of a number of projects in public administration considering the transfer from old document formats to structured documents by means of SGML. Successful transfer to systematic structured document management, however, still seems to be rather rare. SGML is in use e.g., in the Norwegian Parliament and ministries [13, 26], the Supreme Court of Canada [20], the Tasmanian government [1], and in the Finnish Parliament and ministries [22, 24]. SGML has also been utilized in the preparation of the Budget of the European Union [5]. XML has been considered as a basis for document production, for example, in Estonia [11] and United States Congress [4].

2.4. XML as metadata format

As an asset in an organization, metadata about documents is as important as the documents themselves. Metadata is needed, for example, to find documents and information in them, to record information about the context where documents have been created, to maintain the access rights, and to maintain the accessibility of information recorded in digital form over time. Metadata related to other resources than documents is also important. For example, for being able to use XML as data interchange format between a set of software systems, data about the systems is needed. The World Wide Web Consortium

(W3C), coordinating the development of web and XML technologies, has published a model for representing metadata associated to any information resources, and to express the data in XML format. The model is called Resource Description Framework (RDF) [14]. Building metadata repositories to support the design of new, consistent e-government solutions has started in some countries. The Danish Infostructurebase (http://isb.oio.dk/info/) and the e-GIF service (http://www.govtalk.gov.uk/) in Great Britain are examples of such repositories.

Important metadata categories are contextual metadata, structural metadata, and semantic metadata. *Contextual metadata* is used to provide information about the context where resources are created and used. *Structural metadata* and *semantic metadata* again provide information about the structure and meaning of resources, respectively. XML-based presentations have been developed for all these types of metadata to support digital government solusions.

EULEGIS prototype system [18] provides an example of the contextual metadata. The system demonstrates how contextual metadata can be used to offer new kinds of information retrieval capabilities to distributed document repositories. The system is intended to retrieve information from the European legal databases, containing documents written in different European languages and produced in various European legal systems. Metadata describing European legal processes, organizations involved, and document types created in the processes has been described in XML format. The metadata is visualized to the user by graphical models. The models provide information about the European legal systems and they also serve as the user interface for retrieving documents created in the processes, stored in various European legal databases.

From the structural metadata examples are the XML schemas defined for standard document structures to be used for documents in public administration. Examples of such schemas are available, for example, in the Danish Infostructurebase (http://isd.oio.dk/info/) and the British e-GIF service (http://www.govtalk.gov.uk/).

At W3C, semantic metadata based on the use of XML and RDF has been found essential for turning the web of heterogeneous resources offered by the current Internet technologies into *semantic webs* serving better particular communities. Compared to the current web offering huge content repositories but very limited capabilities to meaningful and safe automated processing of the content, semantic web has rich metadata resources with information about other resources facilitating meaningful communication between applications and meaningful and safe processing of the information stored in the resources. The metadata is expressed in standardized languages understandable both by people in a community and applications used in the community. The terms and concepts of the community are expressed in *ontologies*. There are various ways to express ontologies. The document structures defined by DTDs or other kinds of schemas are examples of introducing the terminology for documents of a type and relationships of the terms in the document structure. A document structure, however, does not necessarily tell much about the meaning of the terms in respect to the content of the documents. The Web Ontology Language (OWL) [19] and RDF Scmema [3] provide XML-based notation to describe concepts related to the content of resources and relationships between the concepts. The ontology describing cultural collections at the Finnish museums is an example case where ontology is used to facilitate semantic information retrieval and to integrate different databases [12]. The ontology-based retrieval is implemented in the pilot version of the MuseumFinland portal at http://museosuomi.cs.helsinki.fi.

3. XML standardization levels

XML standardization in a specific organizational environment is part of more or less interrelated standardization efforts going on around the world. One of the challenges in the design, implementation and maintenance of XML-based standards for an organization or organization network is the evolving, multilevel standardization environment. Three major inter-related levels of standardization can be identified:

Universal standardization. At this level general rules for wide use for different application areas are developed. Universal standardization activities are coordinated by the World Wide Web Consortium (W3C) and a number of XML-related specifications intended for wide use have achieved the status of W3C Recommendation. Examples of those specifications are the XSLT intended for describing transformations of XML documents [6] and XML Schema [9] intended for defining document types by constraining mechanisms not included in the DTD mechanism of XML. The universal standardization concerns XML as well. Version 1.1 of XML has been published in February 2004 [2]. The most important change in the version 1.1 compared to 1.0 concerns the allowed characters in the names of the markup vocabulary. The rigid definition for names in the version 1.0 has been replaced by a broader definition.

Sectoral standardization. At this level rules for the purposes of a specific sector or application domain are defined. The rules are built on top of the universal rules and they are not necessarily intended to be implemented as such but to be used to tailor further specifications at the local standardization level. The developer is often a consortium of organizations of a specific sector but it may also be a person or a single organization, for example, a software company.

In public administration, sectoral standardization is going on both on the international and national level. At international level, the European Parliament has established the EU Interoperability Forum for developing XML-based rules for the administration in the European Union. Public administration organizations also test specifications developed for electronic commerce, for example, within the ebXML activities. The ebXML specifications concern, for example, business processes and core business objects. Communities in public administration are looking for the capabilities to develop their own ebXML-compliant specifications and by them possibilities to utilize the technologies developed for ebXML-based environments and to build new kinds of communication connections to companies.

National XML standardization includes in many countries special recommendations and e-government services. In Finland, the Ministry of Finance has given its recommendations for the use of XML in public sectors [27]. In Denmark, an XML Committee has been established to support XML standardization. The committee has published a Handbook for Standardization [7]. In the United Kingdom, a number of recommendations and XML schemas have been published at the Govtalk web site http://www.govtalk.gov.uk/.

Local standardization. At this level rules for a specific environment and for specific processes are developed and implemented. The extent of the effects of standardization in the environment varies greatly. The standardization concerning data interchange does not necessarily have any radical effects to the work of people in public administration processes. The standardization may however be effective for improving web services, for example, in the form of a new flexible portal. Standardizing documents produced and used in a public administration process may cause major changes, not only to document formats, but also to the work of people in the process, technology to be used in the process, and the ways people of the environment collaborate with each other. In the implementation of document standards a major effort is in managing the changes in document production. Experiences of the local level standardization in the Finnish Parliament and ministries will be discussed in the following sections.

At all levels, standardization is a continuing process. The changes in information technology, in organizations, and in processes cause needs to update earlier specifications and to develop new ones. Furthermore, since the specifications are dependent on other specifications, changes in one specification cause needs to change other specifications. These are facts important to take into account in starting to design XML-based solutions.

4. The case of the Finnish parliamentary documents

Standardization of the Finnish parliamentary documents was activated in 1994, before the existence of XML. Constant changes in word processing applications and incompatibilities of document formats caused problems in information exchange and extra work. Printing costs were high and there was uncertainty about the long-term accessibility of information in the parliamentary documents in electronic form. A project called RASKE was commenced by the Finnish Parliament and a software company in cooperation with researchers at the University of Jyväskylä [22]. Later Ministry of Foreign Affairs, the Ministry of Finance, the Prime Minister's Office, and a publishing house also participated in the project. Early in the project the structured document approach based on SGML standardization was chosen as the objective for the development.

An important achievement in the project was to create a framework where document standardization was not considered as development of document formats only, but as holistic development of document management environments related to business or administrative processes. It was realized that effective implementation of the structured document approach will evidently change, not only documents, but also work with documents in the processes, roles of people in the processes, and tools for working with documents. Therefore methods for analyzing and describing document management on a domain were adapted from existing information systems development methods.

The RASKE methodology provides various methods and guidelines for a document standardization project: methods for data gathering and requirements analysis, modelling methods, evaluation methods, models for describing the results in project reports, and guidelines for proceeding in a standardization project. The RASKE methods and experiences are reported, for example, in [21, 23, 24]. Example reports and articles produced during the project as well as articles reporting experiences from the development activities initiated in the project are accessible from the web site http://www.it.jyu.fi/raske/.

Figure 1 shows the RASKE model for a document management environment [21]. The figure shows two types of entities in a document management environment: activities depicted by the oval and resources depicted by rectangles. An *activity* is a set of actions performed by one or more actors in a process. In public sector the process can be, for example, the legislative or budgetary process in a country. Corresponding to their different roles in the activities, *resources* are divided into three different types: actors, documents, and systems. *Actors* are the performers of activities. An actor is an organization or a person.

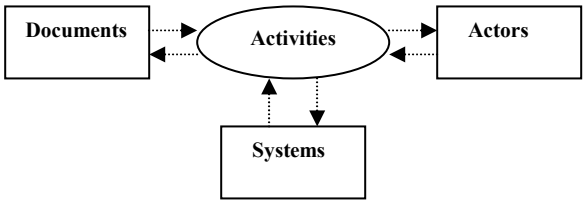

Figure 1. Components of a document management environment

Documents consist of stored data to be understood as information pertaining to the process. A document in its external representation is intended for human perception and it can be identified and handled as a unit in the process. *Systems* consist of the hardware and software applications used to support the performance of activities.

In the begin of a holistic development, one or more processes have to be chosen for the development domain. During 1995-1998, four domains were analyzed in the RASKE project: the enquiry process, national legislative process, Finnish participation in the EU legislative process, and national budgetary process. The analysis of each domain covered work processes, actors, documents, and systems in the processes, and user needs. After some preliminary analysis one or more central document types for a domain were chosen for standardization and preliminary DTDs were designed for them. For example, in the budgetary process the DTD development concerned the budget proposal. On the domain of national legislative process, a publishing house implemented a prototype document archive. The archive included three types of documents in SGML format: government bills, special committee reports, and Parliament replies. The prototype implementation was systematically evaluated by researchers. The results of the evaluation are described in [25].

The RASKE analysis cases were followed by projects where selected companies developed and implemented SGML solutions (DTDs and tools) for a specific subset of documents, and the Parliament and ministries redesigned their work processes. The first implemented document repository in SGML form was the archive of laws and statutes which was published by the Ministry of Justice in 1997. The archive was built by the publishing house printing the statute book. SGML-based budgeting system was implemented by the Ministry of Finance in 1998. In the Parliament, the adoption of SGML caused the most radical changes. Document production in the Parliament was gradually changed to structured document authoring. Changing the document production processes and authoring tools was a major reengineering effort. By the end of 2000, the SGML implementation covered all parliamentary legislative documents. The transfer from SGML to its subset XML is planned to take place in the near future.

The SGML implementation in the Finnish Parliament has had various effects. It has been characterized as the most important document management project in the Finnish public administration ever [16, 17]. The effects can be recognized both in the legislative organizations and outside them. In the legislative organizations the major changes have taken place in the document management of the Parliament. Outside them the major effects have been in the improved services on the Internet.

5. The case continues: From document management to content management

During the process of SGML standardization in the Finnish Parliament and ministries, XML evolved from SGML and the use of the Internet technologies extended from public services available on the Internet to intranet and extranet solutions all over the world. The network connectivity caused new demands for interconnectivity of the systems and data resources within networks, and XML was suggested as a tool for improving interconnectivity and data integration.

The positive results from the collaboration of researchers, the Finnish Parliament, and some ministries in the RASKE project encouraged establishing a new project called RASKE2 for the years 2003-2006 to support data integration and system connectivity in the Parliament and ministries. The goal in the new project is to identify and standardize the metadata most essential for improving legislative content management and related services.

The new project continues the holistic approach adopted in the RASKE project. The analysis domain is the national legislative process. The document management of the domain was analyzed at the time of RASKE and DTDs were developed for a set of documents of the domain. The RASKE framework (shown in Figure 1) has now been extended from the document management development framework to a more general content management development framework shown in Figure 2.

In Figure 2, *content items* represent, not only documents, but also other addressable units of stored data intended as information pertaining to the activities of the domain. A central tool for managing the content items is metadata, and systematic rules and procedures are needed to handle metadata resources as content items themselves. Therefore the content items in the model are divided into two categories: primary content items and metadata content items.

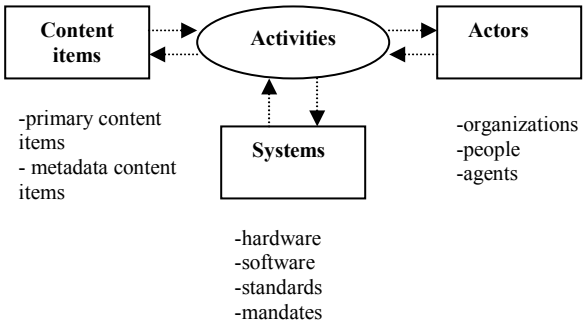

Figure 2. Components of a content management environment

Since a goal in the project is to support development of software-independent metadata standards, usable by various software systems, XML and RDF are the primary candidates for defining the metadata content items. The metadata content items may include, for example, Dublin Core [8] descriptions of documents and other resources, ontologies describing concepts used in documents, DTDs describing the structure of documents, database schemas, and style sheets.

Both in Figure 1 and in Figure 2, the broken lines show the information flow between activities and resources. Resources are regarded as information repositories in which the information produced in an activity can be stored or from which information can be taken and used in an activity. Information needed and produced during activities is stored in documents and other content items, in the heads and experience of people, in the organizational culture, and in systems. In the framework of Figure 1, XML is considered primarily as a format for documents. In the content management framework of Figure 2, XML may be used as a format for primary content items or for metadata content items. In addition to them, XML can be used as an information exchange format. In developing a content management environment towards flexible, standardized content management, the redesign of systems should lead to systems facilitating the import and export of data in XML format.

The RASKE methods were developed for analyzing and describing creation and use of documents in inter-organizational processes, their structures, the user needs related to the document management, and for evaluating new solutions based on the structured document approach. In RASKE2, the work has started for analyzing the legislative process as a content management environment. The RASKE modelling methods have been used to descibe the organizational framework of the legislative process and the document production in the process. New methods will be tailored for identifying and describing the most essential metadata in the process. User interviews have started to explore the major problems and needs related to the content management.

The preliminary results show that currently the metadata solutions in the legislative process are heterogeneous and strongly software-dependent. The length of the process varies from a few days to years. In complex cases the development of a statute has connections to several earlier national statutes and EU directives, collaboration of a number of experts is needed from various organizations, and the management of the collaboration and versioning of documents is a most challenging task without effective tool support. Some users have developed their private innovative solutions to support their work in the process. Analysis of these solutions is expected to help in the identification of the metadata most useful for people working in the process.

6. Challenges

XML offers a rich variety of possibilities but its adoption in public sectors requires extensive collaboration in standardization. Work is needed at all levels: international, national, and local. International agreements are needed to avoid extra work and to facilitate international communication. At national level, work on international level should be considered as a potential basis for the national recommendations or standards. The amount of efforts needed for standardization in a particular organizational environment is highly dependent on the availability of well-documented base standards. On the other hand, the base standards describing the real world never should be considered totally stable. For example, the set of country codes standardized by the ISO (International Organization for Standardization) and also used in XML to refer to countries needs continuous maintenance caused by the geopolitical changes.

The work needed for local level standardization in a particular organizational environment depends, not only on the availability of base standards and recommendations, but also on the type of XML usage. In using XML, for example, for data exchange between software systems, the solution may be a technical implementation. Even though the human users of the systems may participate in the design as user need experts, they do not need knowledge about the implementation details. On the other hand, adopting XML as a format for documents, understanding the changes needed in production processes and capabilities offered by the new solutions usually requires at least some knowledge about XML and about the approach of structured documents.

Experiences show that SGML/XML-based document standardization in inter-organizational processes is a tedious task. In the case of the Finnish parliamentary documents, the work in the first implementation projects proceeded very slowly; there was lack of SGML knowledge in the legislative organizations and lack of knowledge about legislative work among SGML consultants. The situation changed, however, during the work: in the later projects the resources needed were considerably minor than in the beginning. Changes in document production cause changes in the work of several groups of people in the organization. In the beginning, a group of people may seem to get benefits of the new practices while people in other groups may find their workload increase. Motivating the needs for changes and demonstrating future benefits is extremely important.

When the standardization concerns metadata, a problem evolved in the RASKE2 project is the lack of understanding of the concept of metadata. In system dependent solutions the terminology used for metadata varies, therefore to some people the term "metadata" may be totally unknown, for others its meaning varies. One of the goals of the RASKE2 project is to improve understanding of the metadata concept among people working in the activities of the legislative process and understanding its importance for flexible ITC support in the legislative process.

Experiences in the Finnish public administration emphasize the need for inter-organizational co-operation from the early phases of standardization. The involvement of many organizations in the process increases the complexity of the planning and implementation of new solutions. The need for profound analysis of work processes and resources in them in the beginning of the standardization has become clear. It is important to elicite information about the requirements and needs from all parties involved.

Continuous changes in specifications and software cause problems to all kinds of XML standardization. One of the hardest challenges is the vulnerability of the Internet. The lack of trust on the technology and people involved is causing disappearance of people from the Internet community. In the insecure Internet environment well planned services may remain without users. Therefore alternative and more trustworth network solutions have to be considered for digital government.

7. Conclusion

XML technology offers interesting and multiple opportunities for public administration to build digital government. The paper described various ways for utilizing XML and gave some examples from Finland. A challenge for getting XML into use is in the XML standardization. The local standardization in a particular environment always depends on other standardization activities. Syncronization of the standardization activities at different levels and by different groups may hinder reaching solutions stable enough to be effectively implemented.

The RASKE methodology discussed in the paper has been developed in long-term collaboration with the Finnish Parliament and ministries, and in the context of actual standardization work. As the result of the first phase, SGML was taken widely into use for parliamentary documents. The method development continues in the RASKE2 project to find means for metadata standardization and integration. The methods are tested on the domain of legislative process. In the case environment the metadata standardization concerns complex inter-organizational processes, people working in them, people needing information created in the processes, and people participating the processes by some other ways, for example, by commenting legislative proposals. A challenge to the method development is to find means to support the building of really useful services for people participating in the public sector processes and for people needing the services of public sectors.

References

[1] Arnold-Moore, T., Clemes, J., and Tadd, M. (2000). "Connected to the law: Tasmanian legislation using EnAct", *Journal of Information, Law and Technology*(1).

[2] Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., Yergeau, F., and Cowan, J. (2004). Extensible Markup Language (XML) 1.1. W3C Recommendation 04 February 2004. http://www.w3.org/TR/2004/REC-xml11-20040204/.

[3] Brickley, D. and Guha, R.V. (Eds.), RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation 10 February 2004.
http://www.w3.org/TR/rdf-schema/.

[4] Carmel, J. (2002). Drafting Legislation Using XML at the U.S. House of Representatives.
http://xml.house.gov/drafting.htm.

[5] Catteau, T. (1997). The European Union's budget: SGML used to its full potential. In *Conference Proceedings of SGML'97 US* (pp. 645-653).

[6] Clark, J. (Ed.) (1999). XSL Transformations (XSLT) Version 1.0, W3C Recommendation, 16 November 1999.
http://www.w3.org/TR/1999/REC-xslt-19991116.

[7] The Danish XML Committee, Handbook for Standardization, Version 1.0, 2003.
http://www.oio.dk/files/Standardization_Handbook_v1_20030203.pdf.

[8] Dublin Core Metadata Element Set, Version 1.1, Reference Description. http://dublincore.org/documents/1999/07/02/dces/.

[9] Fallside, D.C. (Ed.). XML Schema Part 0: Primer, W3C Recommendation 2 May 2001.
http://www.w3.org/TR/2001/REC-xmlschema-0-20010502/.

[10] Goldfarb, C. F. (1990). *The SGML Handbook*. Oxford, UK: Oxford University Press.

[11] Heero, K., Puus, U., and Willemson, J. (2002). XML based document management in Estonian legislative system. In H.-M.

Haav & A. Kalja (Eds.), *Proceedings of the Baltic Conference, BalticDB&IS 2002* (Vol. 1, pp. 321-330). Tallin: Institute of Cybernetics at Tallin Technical University.
[12] Hyvönen, E., Saarela, S., Viljanen, K., Mäkelä, E., Valo, A., Salminen, M., Kettula, S., and Junnila, M., A semantic portal for publishing museum collections on the web. *Proceedings of ECAI/PAIS 2004*, Valencia, Spain, 2004.
[13] Johanson, B. (1997). Medieuavhengig publisering i staten (MUP*)*. http://www.sgml.no/sgmlinfo/info3_97/mupref.html
[14] Klyne, G. and Carroll, J.J. (Eds.), Resource Description Framework (RDF): Concepts and abstract syntax, W3C Recommendation 10 February 2004.
http://www.w3.orf/TR/rdf-concepts/.
[15] Koli, A. (2003). XML julkisen hallinnon sähköisten palveluiden kehittämisessä. Insinöörityö, Espoon-Vantaan Teknillinen ammattikorkeakoulu.
[16] Kuronen, T. (1998a). Hajautettu dokumenttien hallinta (Vol. 41). Oulu: Oulun yliopiston kirjaston julkaisu.
[17] Kuronen, T. (1998b). Tietovarantojen hyödyntäminen ja demokratia (Vol. 174). Helsinki: Sitran julkaisu.
[18] Lyytikäinen, V., Tiitinen, P., and Salminen, A. (2003). Unifying access to heterogeneous document databases through contextual metadata. In S.A. Becker (Ed.), *Effective Databases for Text and Document Management*(pp. 93-107). Hersley, PA: Idea Group Publishing.
[19] McGuinness, D.L. and van Harmelen, F. (Eds.), OWL Web Ontology Language Overview, W3C Recommendation 10 February 2004.
http://www.w3.org/TR/owl-features/.
[20] Poulin, D., Lavoie, A., and Huard, G. (1997). "Supreme Court of Canada's cases on the Internet via SGML", *E Law - Murdoch University Electronic Journal of Law, 4*, 3 (September 1997).
[21] Salminen, A. (2000). Methodology for document analysis. In A. Kent (Ed.), *Encyclopedia of Library and Information Science* (Vol. 67 (Supplement 30), pp. 299-320). New York: Marcel Dekker, Inc.
[22] Salminen, A., Lehtovaara, M., and Kauppinen, K. (1996). Standardization of digital legislative documents - a case study. In M.S. Lynn (Ed.), *Proceedings of the Twenty-Ninth Hawaii International Conference on System Sciences*, Vol. 5 (pp. 72-81). Los Alamitos, CA: IEEE Computer Society Press.
[23] Salminen, A., Lyytikäinen, V., and Tiitinen, P. (2000). "Putting documents into their work context in document analysis", *Information Processing & Management, 36*(4), pp. 623-641.
[24] Salminen, A., Lyytikäinen, V., Tiitinen, P., and Mustajärvi, O. (2001). Experiences of SGML standardization: The case of the Finnish legislative documents. In J. R. H. Sprague (Ed.), *Proceedings of the Thirty-Fourth Hawaii International Conference on System Sciences* (pp. file etegv01.pdf at CD-ROM). Los Alamitos , CA: IEEE Computer Society.
[25] Salminen, A., Tiitinen, P., and Lyytikäinen, V. (1999). Usability evaluation of a structured document archive. In R.H. Sprague, Jr. (Ed.), *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society Press.
[26] Sundholm, E. (1997). The Odin: The central web server for official documentation and information from Norway.
http://www.ifla.org/IV/ifla63/63hole.htm
[27] Valtiovarainministeriö, XML-strategiatyöryhmä (2003). Julkishallinnon XML-strategia.