A COMPARISON OF MOREQ AND SÄHKE METADATA AND FUNCTIONAL REQUIREMENTS

PEKKA HENTTONEN, DEPARTMENT OF INFORMATION STUDIES, UNIVERSITY OF TAMPERE

Published in Records Management Journal, Vol. 19 No. 1, 2009, pp. 26-36, DOI 10.1108/09565690910937227

ABSTRACT

Purpose The article compares MoReq and SÄHKE specifications to see their differences and similarities. Both MoReq and Finnish SÄHKE set requirements to functionality and metadata of ERM systems and define a XML-scheme for exporting records from the systems. Thus, both specifications are quite similar in some sense. The article seeks to find out whether there is an easy way for harmonizing the both. This tells about obstacles and possibilities of European co-operation in the records management sphere.

Design/Methodology/Approach The study is based on textual sources.

Findings MoReq and SÄHKE are quite incompatible when one looks at data models, elements and functional requirements. There are also some similarities, but no clear path from one specification to another.

Practical implications There is no easy way for turning a SÄHKE-approved system into MoReq certified system (or vice versa). Harmonizing SÄHKE with MoReq would require major revisions in SÄHKE and also a policy change in Finland.

Originality/Value Because the SÄHKE specification is available only in Finnish, its content is largely unknown outside the country. There are no similar studies.

Keywords Records management, ERMS, MoReq, SÄHKE, metadata

Paper type Research paper

1 INTRODUCTION

The second version of the MoReq (Model Requirements for the Management of Electronic Records) was published in 2008. The first MoReq version was not formally adopted in European countries, but it had impact on national requirements as a best-practice standard (Waldron 2004). One of the national specifications is Finnish SÄHKE. SÄHKE has been in force from the beginning of the year 2006. Finnish public authorities are required to use SÄHKE-compatible electronic records management systems if they want to preserve records with permanent value only in electronic form and later transfer them to the custody of the National Archives Service of Finland (NAS).

The question is to what extent SÄHKE is compatible with MoReq. The first MoReq version was used as a source in SÄHKE development, but there was no intention to produce a MoReq compatible regulation. To see the possibilities and obstacles for the harmonization of European electronic records management, one must understand how SÄHKE and MoReq2 compare to each other. This is also important for vendors trying to produce software for the European records management market at large and for the future development of SÄHKE as a national specification.

2 COMPARISON

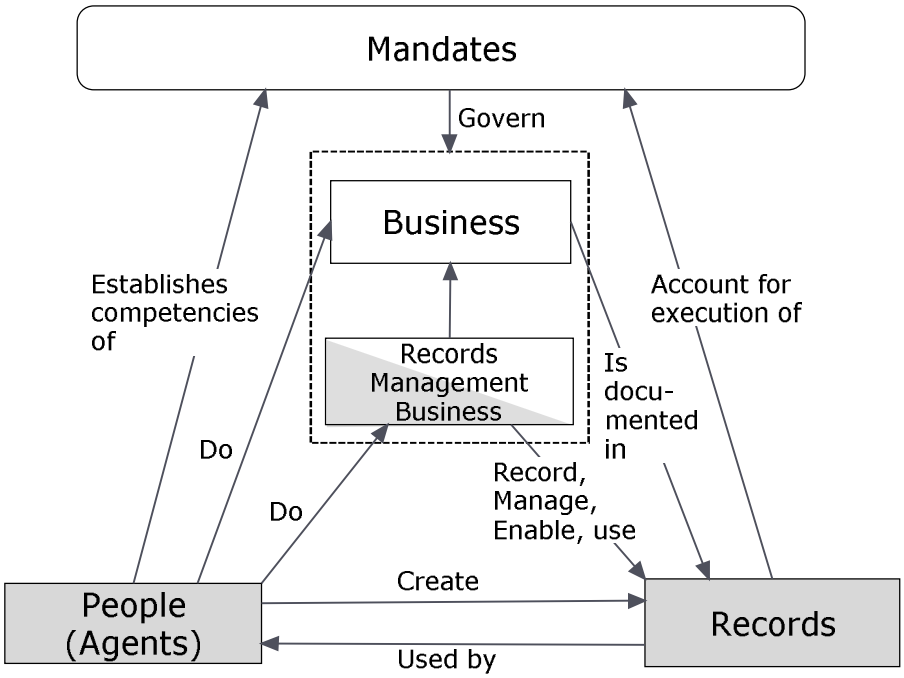
Comparison of SÄHKE and MoReq is based here on the text in the specifications and especially on the explicit similarities in functional requirements and metadata defined in SÄHKE. Because SÄHKE is older than MoReq2 it refers only to MoReq1. However, MoReq2 does not represent a fundamental change in the development of MoReq. The

sub-title describes it only as an “update and extension” and it is stated that MoReq2 “has been written to ensure, as far as possible, compatibility with the original MoReq” (*MoReq2 Specification* 2008). Hence, the results should apply also to MoReq2. If anything, MoReq2 has further set apart the two specifications, because explicit mappings between the two are broken. MoReq2 includes some new metadata elements that have equivalents in SÄHKE. Although this makes the specifications a little more compatible, it does not change the big picture of their relationship.

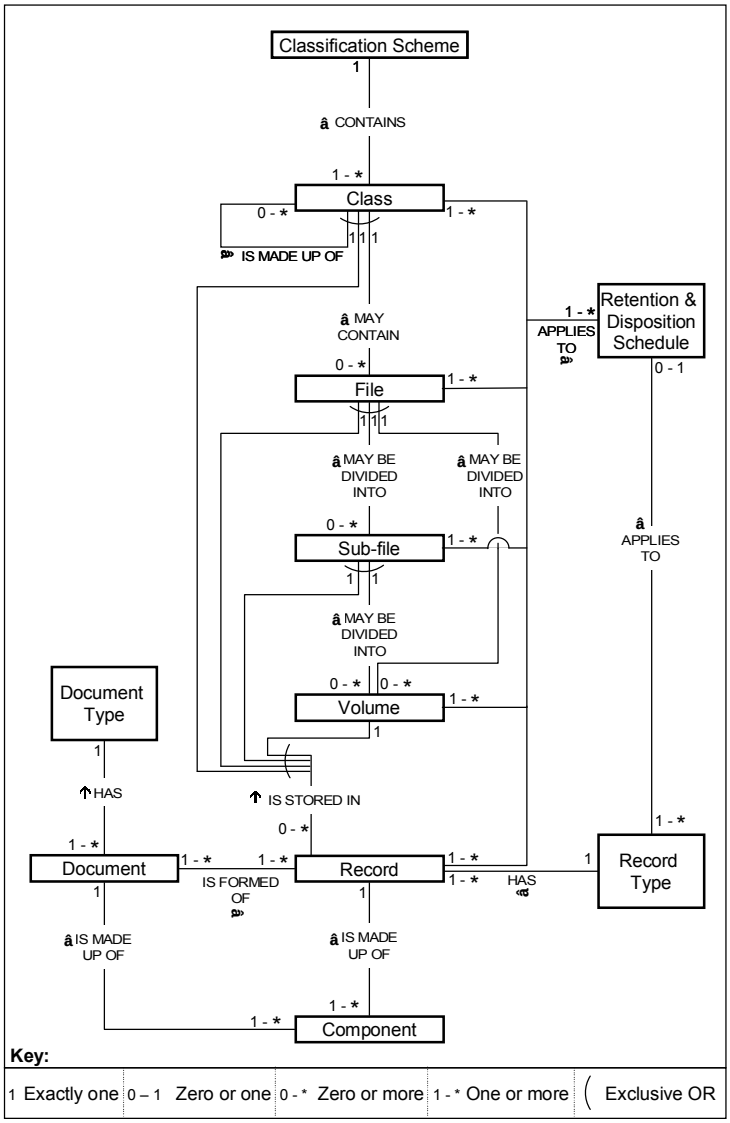
2.1 DATA MODELS

(Haynes 2004) discusses purposes of metadata and suggests a new five-point model to describe them (resource description, information retrieval, management of information resources, documenting ownership and authenticity of digital resources, interoperability). Despite this high level harmony, every metadata model represents a view of reality to suit a particular purpose. A purpose of a scheme determines its elements. The content, structure and level of a metadata scheme is dependent upon the context of the scheme’s creation (Duff 2001). Both MoReq2 and SÄHKE are similar in that they set requirements to electronic records management systems.

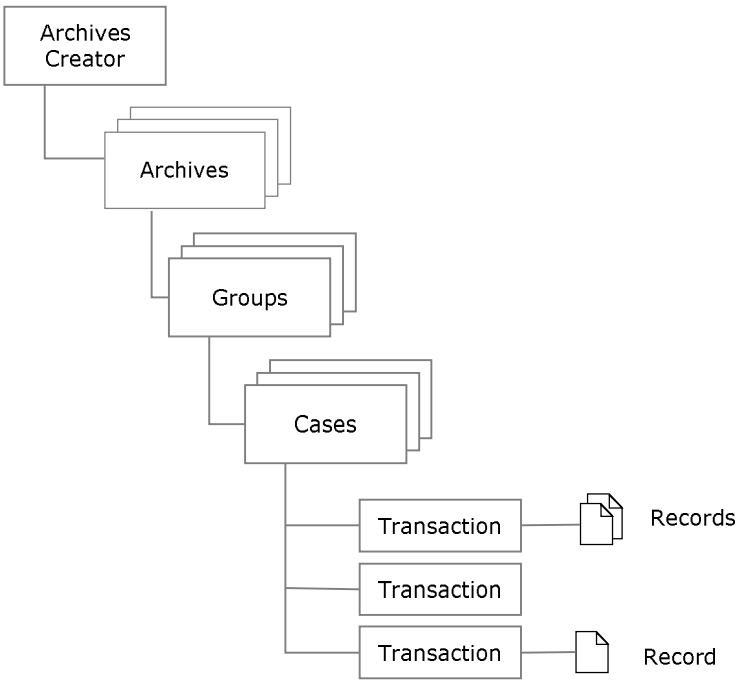
The metadata of electronic records management has been defined in the ISO 23081 standard (*ISO 23081-1. Information and documentation Records management processes Metadata for records Part 1. Principles* 2006). MoReq2 states that because it describes only computer system and not the entire record keeping environment, it covers only part of the metadata described in the ISO 23081: MoReq2 has metadata about Agents, Records, and Metadata Record, and partly about Records Management Processes (that is, entities in the lower half of the Fig. 1).



From bird’s eye view SÄHKE and Moreq2 are quite similar. Differences emerge when one looks at the data models. MoReq has been developed as an initiative of the DLM-Forum, which is a European wide organization of national archives, enterprises and research organizations having interest in electronic records management. Initially created and funded by the European Commission, DLM-Forum now operates as an independent body. The group behind the MoReq2 consisted of representatives and specialists from various countries. Below is the entity–relationship model of MoReq2 (*MoReq2 Specification* 2008, p. 22).



SÄHKE's background is purely national. It was developed as a project of the National Archives Service, which guides records management in the Finnish public administration. Therefore, SÄHKE closely reflects both local Finnish recordkeeping traditions and the needs of the National Archives Service. Because of its creator, SÄHKE is at the outset more inclined to long-term preservation than MoReq. "SÄHKE" is an abbreviation of the Finnish project name "Project for developing electronic information management and long–term preservation" (Moisio 2002). The hierarchy required by the SÄHKE data model is below (Arkistolaitos 2005a, p. 7, I).



The differences are obvious. The data structure in MoReq2 consists of entities of same kind: record aggregations joined together by a common classification scheme. Thus, MoReq2's model is quite consistent with the ISO 23081-2 standard. MoReq2 notes that good practice dictates that the structure should reflect business functions. However, the data model does not enforce this. SÄHKE hierarchy is a hybrid. It has entities not only from ISO's Records (*Records* and *Archives*) hierarchy, but also from Agents (*Archives Creator*) and Business hierarchies (*Transaction*). Some entities (*Groups* and *Cases*) are a cross, because they describe both aggregations of records and business of the organization at the same time.

MoReq2 introduced a new entity, *Component*, which is defined as "a distinct bit stream that alone or with other bit streams makes up a record or document". Already MoReq1 had the concept of *Volume* ("A subdivision of an electronic file or paper file"). Both are problematic when it comes to SÄHKE compatibility, because SÄHKE has no counter-part for these entities.

MoReq2 allows flexible combination of entities. For instance, it is possible to store records directly in a *Class*, without being in a *File*. SÄHKE model is very rigid. Although the upmost level (*Archives Creator*) is used only when records are transferred to NAS, all the levels in the hierarchy must exist and a *Record* can be linked to the hierarchy only via a *Transaction*.

SÄHKE model is understandable, if we look at its background. Finnish record keeping is traditionally based on registry systems. In a registry system, incoming and outgoing letters are registered when they enter/leave the organization. In registration a record is linked to a process. Management and citizens may use registries to follow what takes place in administration. A "case" is an administrative process with a definite beginning and an end. The registry tells what actions have been taken in any particular case and what records have been created in the actions. SÄHKE allows *Transactions* without a *Record*, because not always a record is created in the action registered.

SÄHKE is built to facilitate transfer of records to NAS. Hence, the entities in the data model can be directly mapped to the description levels (Archives Creator – Archives – Series – Archival Units) defined by the Finnish rules for archival description (Arkistolaitos 1997). This is visible also in some metadata elements which are structured to fit in the data model of the NAS archival database. The name of an *Archives Creator* is divided into two parts in metadata matching the two fields in the database. SÄHKE has also metadata elements for the previous/alternative names of the *Archives Creator*, in accordance to the archival description rules. This adds features which do not exist in MoReq.

2.2 FUNCTIONAL REQUIREMENTS

SÄHKE contains relatively few functional requirements. There are about 60 functional requirements in SÄHKE, whereas almost 800 in MoReq2. SÄHKE specification gives a source for most of the requirements. The sources include MoReq1, Functional Requirements for Evidence in Recordkeeping (the results of the so called "Pittsburgh Project"), the instructions of NAS concerning registration of records, instructions for data security in the Finnish state administration, and the metadata specification of the SÄHKE itself.

Comparing the requirements is mostly impossible or not meaningful, for three reasons. Firstly, although SÄHKE requirements are partly (in 25 cases) mapped to MoReq1, they are quite summary. In every case where MoReq1 is the indicated source of a SÄHKE requirement, the reference is to a MoReq1 chapter or sub-chapter instead of a single requirement (see table, below, *MoReq* 2002; *MoReq2 Specification* 2008; Arkistolaitos 2005b).

SÄHKE-requirement(s) (and their number)		Chapter in MoReq1 [modified name in MoReq2]	Number of requirements	
			MoReq1	MoReq2
2.1.1	1	4.1 Access	12	24
2.1.4	1	4.2 Audit Trails	12	16
2.1.5	1	4.4 Tracking Record Movements	3	deleted
2.1.9–10, 2.2.9, 2.5.4	4	10 Other functionalities	71	281
2.2.2	1	6.3 Types of Documents	6	deleted
2.2.7, 2.4.4	2	6.1 Capture	15	41
2.3.2, 2.3.4–8, 2.3.11	7	12 Metadata Requirements	24	24
2.4.1	1	5.1 Retention [and Disposition] Schedules	18	43
2.4.3, 2.4.5	2	5.2 Review [of Disposition Actions]	11	8
2.4.11, 2.4.14	2	5.3 Transfer, Export and Destruction	17	24
2.5.1	1	7 Referencing	7	16
2.5.2 – 2.5.3	2	8 Searching, Retrieval and Rendering	46	56

For instance, SÄHKE requirement 2.4.1 is mapped to chapter *5.1 Retention schedules*, which has 18 requirements in MoReq1 and 43 in MoReq2 (where its name has been changed to "Retention and Disposition Schedules"). Still, the purpose does not seem to be to sum up MoReq1 requirements into few sentences—most of the requirements in SÄHKE are quite specific—but to indicate similarities at a general level. Thus, the relationship between SÄHKE and MoReq requirements remains often quite indeterminate.

Like in MoReq, some of the requirements are not technical.[i] In addition, SÄHKE is sketchy at times. For instance, the requirement 2.4.1 could easily be split into four parts.[ii] This may be due to the fact that individual consideration is

always used when the fulfillment of requirements is evaluated. Finnish policy is not to certify ERM systems. NAS may give a government agency permission to keep permanent electronic records in a system that meets SÄHKE requirements, but the permission is given to the agency, not to the system vendor. Every agency has to apply for permission, even when the system has already been approved for use in another agency. Hence, permission is as well a statement about the general level of electronic record keeping in the agency as a seal of approval for its ERM system.

Because SÄHKE is a specification for case management systems used in the Finnish public administration (and not a general ERMS specification), it has requirements that are not meaningful outside this sphere. In addition to registry tradition, a distinguishing feature of Finnish records management is the use of AMS (*arkistonmuodostussuunnitelma*). AMS is usually translated either as a "records retention/management schedule" or an "archives creation/formation plan" (Kilkki 2004; Lybeck 1998). All short translations fail, because an AMS is actually a combination of classification scheme, file plan, records retention schedule and records management program. In electronic environment organization's AMS serves also as a repository for default metadata values assigned for records captured in its ERMS. In some respects the idea of AMS sets Finnish records management apart from that of other countries. Some SÄHKE functional requirements are about implementation of AMS.

As a small specification, SÄHKE requirements are mostly about the core of the electronic records management functionality: life cycle management, access rights, retention and disposal. SÄHKE does not at all address questions of collaborative working, digital rights management, workflow, offline and remote working, email or integrating ERMS with content or document management. The table above shows how MoReq2's chapter 10 (covering these areas) has been expanded from the first version: it alone contains now about five times as many requirements as SÄHKE.

2.3 METADATA ELEMENTS

SÄHKE has over 120 metadata elements,[iii] many of which can be used at several levels in archival hierarchy. In addition, there are metadata elements that are used only in a transfer of records to NAS. Altogether there are about 280 possible metadata element and entity combinations that may get a value. To express obligatoriness, SÄHKE uses a four part classification: an element value can be "optional", "obligatory", "obligatory if information exists", or "obligatory in eGovernment business processes". About 60 % of the values are more or less obligatory. SÄHKE tries to guide metadata production, not to just specify what element values are technically minimally required.

The metadata elements are compared in the specification to three other metadata standards: Finnish JHS143 (a recommendation of metadata elements to be used in public administration), Dublin Core and MoReq1. JHS143 and Dublin Core are about document level metadata. Information in the table below shows what metadata elements are linked to MoReq1 (Arkistolaitos 2005a). The table also shows what entities these elements are used to describe. Altogether only 21 SÄHKE-elements are mapped to MoReq1.

SÄHKE		MoReq1	
Metadata Element	Entities	Metadata Element	Entities
1.1 Title.mainTitle	Archives, Group, Case, Transaction, **Record**	*12.7.1 Identifier*	**Record**
2.1 Agent.role. mainCreator	Group, Case, Transaction, **Record**	*12.7.3 Author*	**Record**
4 Subject	Case, **Record**	*12.4.2 Name*	Class, File
		12.4.3 Desc. Keywords	Class, File

		12.4.22 Keyword-based name	Class, File
		12.7.2 Subject	**Record**
5.1 Description. mainDescription	Archives Creator, Archives, Group, Case, Transaction, Record	*12.4.4 Description*	Class, File
6 Type	Case, Transaction, **Record**	*12.7.7 Record Type*	**Record**
7 Language	Archives, Group, Case, **Record**	*12.7.27 Language*	**Record**
*10 Identifier **	Archives Creator, Archives, Group, Case, Transaction, **Record**	*12.7.1 Identifier*	**Record**
*11 Date **	Group, Case, Transaction, **Record**	*12.7.5 Date / Time*	**Record**
		12.7.8 Registration Date / Time	**Record**
11.5 Date. acquired	Case, Transaction, **Record**	*12.7.23 Date received*	**Record**
11.8 Date.sent	Case, Transaction, **Record**	*12.7.22 Date Sent*	**Record**
12 Version	**Record**	*12.7.26 Document Version*	**Record**
13 Source	Archives, **Record**	*12.8.2 Identifier of original record*	**Record** (extract)
*14 Relation **	Group, Case, **Record**	*12.7.24 Links to related records*	**Record**
15 Rights	**Record**	*12.7.25 Intellectual property rights*	**Record**
*16 Restriction **	Group, Case, **Record**	*12.7.11 Security Category*	**Record**
		12.9.4 User access rights	User
*17 Retention **	Group, Case, **Record**	*12.7.16 Retention schedule*	**Record**
*19 Preservation **	Archives, Group, Case, **Record**	*12.7.13 Preservation metadata*	**Record**
*20 Format **	**Record**	*12.7.13 Preservation metadata*	**Record**
20.4 Format. encryption	**Record**	*12.7.29 Electronic watermark information*	**Record**
*23 Storage Location **	Archives, Group, Case, Record	*12.5.7 Physical Location*	File, Volume

28 Signature	**Record**	*12.7.20 Electronic signature*	**Record**

SÄHKE specification does not state how SÄHKE and MoReq1 entities are mapped to each other: self–evident pairs are marked in the table 2 with bold letters. Examination of data models (see above chapter 2.1) already showed that only one entity, *Record*, is common to MoReq and SÄHKE. Consequently, it is hardly surprising that mapping usually takes place at Record level.

Two elements, *5.1 Description.mainDescription* and *23 Storage Location* are exceptions to this. In both cases, it may not be clear how mapping should take place. Element descriptions in SÄHKE suggest that SÄHKE *Groups* and MoReq *Classes* are roughly equivalent (Arkistolaitos 2005a, pp. 7–8). SÄHKE *Group* is equivalent to Series level in Finnish archival hierarchy, which is on its part equivalent with MoReq *Class*. However, these are open issues because mapping between SÄHKE data model and Finnish archival hierarchy on one hand, and record aggregations in MoReq and those in Finnish archival hierarchy on the other hand, is not entirely clear.

This is not the only problem of SÄHKE and MoReq metadata interoperability. Elements marked with an asterisk (*) in the table 2 cannot get values according the SÄHKE specification. Thus, in these cases a MoReq metadata element cannot be mapped to its SÄHKE equivalent shown in the table. Instead, it must be mapped to an unspecified sub-element of the indicated SÄHKE element. For instance, there are eight sub-elements of *20 Format* to map MoReq's *12.7.13 Preservation metadata* with: *extent, medium, fileFormat, encryption, compression, hardwareSupport, softwareSupport, and significantColor*. One has to solve these problems at implementation phase, if interoperability is needed.

It should be noted that the table above ignores repeatability and obligatoriness of metadata elements which may not be the same in both specifications. Also encoding schemes may be partly incompatible (this has not been studied here). Finally, how the elements are used and what is their semantic meaning may sometimes be different.

To sum up, a ready path for mapping SÄHKE and MoReq elements exists only for small minority of metadata elements. Even when it exists, there are likely to be problems and questions requiring implementation specific consideration.

3 DISCUSSION

The comparison shows that although SÄHKE and MoReq are both specifications for electronic records management, they have considerable differences. Entities assumed by the data model are not similar and also metadata elements differ. This is a result from their different background and purpose. Unlike MoReq, SÄHKE is not a general ERMS specification. MoReq is more technical than SÄHKE. MoReq2 is specifically written to support compliance testing (*MoReq2 Specification* 2008). SÄHKE, on the other hand, tries to guide electronic record management at a general level without restricting itself to technical matters.

Because of differences and the fact that the relationship between SÄHKE and MoReq is largely indeterminate or unclear, there is no simple path from one specification to another. From a vendor's point of view this means that turning a SÄHKE-approved system into MoReq certified system (or vice versa) is far from being a straight-forward process.

SÄHKE cannot be harmonized with MoReq with small modifications. Harmonization would require a major policy change and change in the mindset behind the SÄHKE. It would also mean a clear separation between purely local features and those prevalent at a European level. Thus, such an undertaking would probably require major revisions and partly rewriting the SÄHKE specification from a scratch.

4. REFERENCES

Arkistolaitos (1997), *Arkistojen kuvailu- ja luettelointisäännöt*, Helsinki.
Arkistolaitos (2005a), *SÄHKE-määritykset. Osa I. Abstrakti mallintaminen*, available at http://www.narc.fi/Arkistolaitos/sahkoinen/ (accessed 17.3.2008).
Arkistolaitos (2005b), *SÄHKE-määritykset. Osa II. Toiminnallinen mallintaminen*, http://www.narc.fi/Arkistolaitos/sahkoinen/ (accessed 17.3.2008).
Duff, W. M. (2001), "Evaluating Metadata on a Metalevel", *Archival Science*, vol. 1, no. 3, pp. 285-94.
Haynes, D. (2004), *Metadata for information management and retrieval*, Facet Publishing, London.
ISO 23081-1. Information and documentation Records management processes Metadata for records Part 1. Principles, 2006, ISO.
Kilkki, J. (2004), "Bearmania. Frosting Finnish archival practice with imported archival theory", *Comma*, no. 1.
Lybeck, J. (1998), "Electronic Records as Archival Material: an Overview", in *Musarum Minister. Studia in honorem Kari Tarkiainen. Kari Tarkiainen 60 vuotta*, Arkistolaitos ja Arkistoyhdistys ry, Tampere.
Moisio, R. (2002), "SÄHKEellä säilytystä kuntoon", in K. Kekki & O. Salminen (eds), *Digimaan kartta. Puheenvuoroja digitaaliseen tietohuoltoon*, Opetusministeriö.
MoReq, (2002), Model requirements for the management of electronic records, European Commission. Office for Official Publications of the European Communities, Luxembourg.
*MoReq2 Specification (*2008), CECA-CEE-CEEA, Bruxelles – Luxembourg, available at http://www.moreq2.eu/downloads.htm (accessed 16.2.2008).
Waldron, M. (2004), "Adopting electronic records management: European strategic initiatives", *Information Management Journal*, vol. 38, no. 4.

[i] For instance, see requirement 2.3.1: “There must be instructions for how metadata values should be used”.

[ii] 2.4.1 says ”The system must have functionality for determining retention periods and automatic destruction [of records]. By default the retention time must come from the classification scheme, but there must be a possibility to change the default value”.

[iii] The total number of elements depends on how you count them. Most elements are qualified and consist of main element and one or more qualified sub-elements (e.g. element *Coverage* may be either *Coverage.jurisdiction, Coverage.spatial* or *Coverage.spatial*). This number includes all the qualified and non-qualified elements that may have a value in an ERMS. In addition, there are elements used only to describe a transfer of records to NAS and the system from which the records are exported.