Invited Paper: MoReq2: a European Contribution to the Preservation of Electronic Records

Marc Fresko
Inforesight Limited
Kenley, CR 5HD UK

marc.fresko@inforesight.co.uk

ABSTRACT

MoReq and its successor MoReq2 are European model specifications of requirements for Electronic Records Management systems, often referred to as ERM or EDRM systems. There are several other specifications that set out to define model requirements for EDRM systems – notably the US DoD 5015.2 standard. However, the other specifications almost all are designed expressly for government bodies in one country[1]; MoReq and MoReq2 are differentiated in three ways:

1. they are designed to apply to all sectors (public, private and not-for-profit alike);

2. they apply to all members states of the European Union;

3. they include features that have been found to be valuable in practice, even though they are not strictly required for the theoretical management of records.

This paper concentrates on the third differentiator, and specifically on features that address digital preservation, notably:

- migration;
- "components";
- automated rendition;
- import and export;
- preservation metadata;
- XML schema.

General Terms

Management, documentation, design, reliability, standardization, languages.

Keywords

MoReq2, Model Requirements for the Management of Electronic Records, preservation, digital preservation, sustainability, records, electronic records, European Commission, XML.

---

[1] The singular exception is the recently-published specification from the International Council on Archives [6]. It is too soon to judge acceptance and impact of this publication.

1. BACKGROUND: MoReq AND MoReq2

1.1 Records

This paper makes frequent reference to "records". The formal definition of "record" is given in the international standard that defines Records Management [7] as:

"*Information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business*".

In less formal terms, "records" are documents that tell us what an organization (any organization, in government or in commerce) has done, how decisions were reached, what transactions have been executed, and so on. Importantly, records are *not* limited to those documents that are bundled up and sent to an archive; they include also those documents that are used every day to manage the organization.

1.2 Electronic Records Management

Records Management is a distinct discipline with a long lineage. Its history in Europe can be traced back to the Medieval era. And we have lived with the need to keep electronic records[2] ever since computers were first used in commercial applications[3], though in practice many organizations chose to keep paper or microform copy records instead.

The early history of keeping electronic records, and indeed of making them accessible over a long period, is trivial, as it applies to highly structured, or transactional data. Banks, insurers, manufacturers, airlines and others have routinely created, managed and (wherever necessary) preserved such structured electronic records without undue difficulty, without complaint, without massive loss (usually!), and certainly without multi-million dollar academic research programs. This is because these

---

[2] For reasons lost in history, we refer to "electronic records" instead of "digital records", even though we refer to "digital preservation" and not "electronic preservation". In principle we should prefer the term "digital records" to distinguish computerized records from "analogue electronic records" such as old audio recordings; in practice, we don't.

[3] The first computer known to be used in a commercial application was "LEO", Lyons Electronic Office I, in the United Kingdom in 1951. This 500kHz machine, armed with 2048 words of ultrasonic mercury delay line memory, was used for several applications including valuations, payroll and inventory management.

early electronic records were in formats that were straightforward and understood by (often designed by) the institutions in question.

However, electronic records rapidly became at once more complex and more widespread. The complexity arose because of the evolution of ever-more capable software for word processing, spreadsheets, project management, graphics, presentations and so on; and to compound matters, the formats used mostly were not only proprietary but secret. The widespread usage resulted from a series of technological waves – first office computing in general and PCs in particular, then e-mail, and then the web.

We now all know that virtually all records of enduring value are created, edited, used, modified, and stored electronically. Some records are still created and stored in paper form – correspondence from citizens and retail customers being the prime example – but they now constitute only a tiny fraction of the records created today. Combine with this the very real practical difficulties of storing paper copies of electronic records, the need to manage electronic records is immediately apparent – and thus the need to preserve them also.

Electronic records do not manage themselves, and certainly do not preserve themselves. Without active management, all manner of problems arise, from the change and deletion of records that ought to be fixed, through the inability to find records, to loss of access because of technological evolution.

All this gives rise to a need for systems that can manage electronic records so that they can be used on a daily basis as the information lifeblood of the organization – Electronic Records Management (ERM) systems. Because it turns out in practice to be convenient for the same systems to manage documents that are not formal records (mainly because many of them are destined to become formal records) these systems are also referred to as Electronic Document and Records Management (EDRM) systems.

The need for a formal standard expression of the capabilities of ERM dates back[4] to the "R/DIM" initiative[5] in Canada in 1996, a work which itself grew from a Canadian government initiative of the early 1990s. The resultant RDIM specification was widely read and must have been influential, but is rarely cited and difficult to find today.

More significant by far is the US Department of Defense's standard specification, usually referred to by its number as "5015.2" [11]. Catalyzed by problems observed with the management of electronic records of the first gulf war, it was first published in 1997, and updated several times since then (most recently in 2007). It has been extremely influential on both the capabilities of the software industry and in fostering the expectation of the user base. The former has engineered literally dozens of software application to comply with the various

versions of the standard, and the latter – at least in North America – demand compliance as a demonstration of the ability to manage electronic records.

Similar pressures in other countries gave rise to comparable specifications in other countries, including the "PRO" specification [10] in the UK (by far the most influential of this set), DOMEA in Germany, and many others in countries as far apart as New Zealand and South Africa. All of these specifications, without exception, were developed by organizations of national government to serve as guidance for other government bodies.

1.3 The MoReq Specification

The MoReq specification [2] is similar in concept to the other specifications. However, its genesis was different. It was conceived at an international level, to be applicable across national boundaries throughout the European Union; and it was from the outset intended to apply to all sectors, government and otherwise.

MoReq was conceived by the DLM Forum[6], an international organization concerned with archives and records management. Its initials "DLM" originally stood for the French words "Données Lisibles par Machine" – "Machine Readable Data" in English. However, in a bizarre reversal of acronym causality, the Forum voted in 2003 to change the meaning of DLM to the more anglophile "Document Lifecycle Management".

The Forum convinced the European Commission to fund MoReq's development, and as a result MoReq was produced by a team of consultants led by the author of this paper. It was published in 2001, and immediately became a success across Europe. Its success can be measured by the number of translations produced around the world: we have identified 8 full translations[7] and a handful of adaptations[8], some from countries well outside of Europe (South America, Asia). It was also used in the USA, for example at Indiana University.

MoReq differed from the other specifications. Because it was written by consultants, it took into account a wide range of experiences with both EDRM systems and with procurement specifications. Accordingly, observers viewed it as easier to read and understand than the others. More importantly, it also includes many practical requirements – features that are not strictly essential for a records management in theory, but which were deemed by the authors to be important in practical office settings. One such set of features was an early set of requirements for digital preservation.

1.4 The MoReq2 Specification

For all its rapid international success, MoReq was far from perfect. There were weaknesses in its metadata model; it was not maintained and hence lost its currency; and it lacked any external governance or compliance testing structures. At the same time, and driven partly by the enlargement of Europe, there was growing demand for a "better and bigger" version of MoReq, one

---

[4] The brief summary that follows is restricted to materials published in English. The author is not aware of earlier works in any other language.

[5] The R/DIM specification appears not to have survived on the internet except in the Internet Archive [1]. Interestingly, despite its age, it includes (in its requirements 3.3.2 to 3.3.8) features that relate specifically to digital preservation that are absent from most of the more recent specifications.

[6] See http://dlmforum.eu.

[7] Czech, French, Hungarian, Italian, Portuguese, Russian, Slovenianta, Spanish (two versions).

[8] Brazilian Portuguese (two versions), Dutch, Polish.

with an ongoing management regime. Accordingly, the DLM Forum turned anew to the European Commission, which again agreed to fund development.

So it was that MoReq2 was developed, during 2007, again by a team of consultants. The author of this paper again was in the lead, this time opting for a highly consultative process involving over 200 experts, interested parties and organizations from around the world. The finished MoReq2 was published in early 2008. Like its predecessor, it is intended to cater for all kinds of organization, in any European country. Longer and more detailed than any other specification, it also contains many practically-important requirements that are absent from other specifications, such as automatic numbering, exception processing and the like.

Importantly, MoReq2 was prepared with the support of most of the ERM/EDRM industry. Just some of the 44 companies who explicitly expressed their support are listed in Table 1.

Table 1: Selected vendors supporting the MoReq2 project

• Adobe	• Meridio (now part of Autonomy)
• CA	
• Capgemini	• Microsoft
• EDRM Solutions	• Objective Corporation
• EMC	• Open Text Corporation
• FileNet (now part of IBM)	• Oracle
• Fujitsu	• SAP
• Getronics	• Tower Software UK (now part of HP)
• IBM	
• Lockheed Martin	• Xerox

Like its predecessor, MoReq2 has prompted many translation projects: the French and Russian translations have already been published, with translations into Catalan, Korean, Romanian, Slovenian, Spanish and other languages well advanced.

It consists of the following publications:

- the specification itself [3];

- a comprehensive metadata model [3];

- a testing framework (test conditions, expected results) [4];

- an XML schema [5].

2. DIGITAL PRESERVATION FEATURES IN MoReq2

The general philosophy of MoReq2 is to include features that are necessary in a practical EDRM system, even if they are not strictly necessary for the management of electronic records. One set of such features addresses digital preservation. The features are described briefly here.

2.1 Migration

First, MoReq2 includes requirements for the migration of specified records to new formats, to support active management efforts for digital preservation. These requirements are mandatory.

2.2 Recognition of "Components"

Possibly the most significant novel feature of MoReq2 is its recognition of electronic objects that are smaller than electronic records. For example, web pages tend to be built up of text files, image files, style sheets etc[9]., each of which is treated as a separate entity by any operating system, but all of which have to remain tightly linked together in an EDRM system if the integrity and usability of the records are to be maintained. MoReq2 coins the term "component" to describe these objects, in the absence of any agreed term.

MoReq2 contains a sophisticated conceptual model to relate the various electronic and physical entities that make up a set of records stored in an ERM system. The model, expressed as an entity-relationship diagram, is a development of a similar model in MoReq, but is much more rigorous and powerful. A small extract from this model, in Figure 1, shows how components are related to records.
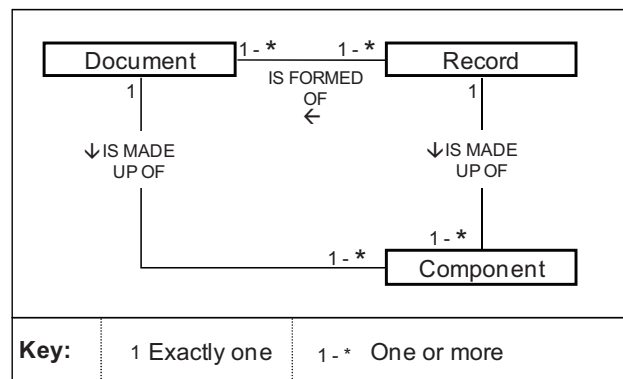

Figure 1: Relationship of components and records

This shows that all documents and records are made up of at least one component; and that some are made up of several. So, for example, a simple spreadsheet is made up of only one component; but a record consisting of linked spreadsheets is made up of several components. This paper could consist of a single object (if the graphics are contained entirely inline) or could be several components if the graphics are implemented by means of embedded links to external images.

Components play a critical role in digital preservation. To continue the example of websites, if it is necessary to preserve web pages made of many components (as almost all pages are) then it is the components themselves, and their links, that will need to be migrated – not the page "object" itself.

MoReq2 recognizes the need to manage components in a way that preserves integrity and usability. Uniquely, it also specifies requirements for migration. For example, if a website is made up of an HTML component plus GIF and PNG components, it might become necessary to migrate the GIF components (and nothing else), say from GIF format to PNG format. However, doing this would require the HTML tags that reference the GIF files to be changed to reference the new PNG files. This is unavoidable, but very counter-intuitive to Records Managers who (rightly) believe that the content of records must not be changed. The scenario of component migration is a rare situation in which a change to content (changing filenames from GIF to PNG in this example) is

---

[9] For example, the home page of the NARA website in early March 2009 is made up of one HTML file, 3 GIF files, 2 JPEG files, 7 Javascript files and 12 cascading style sheets.

a prerequisite for the maintenance of usability. MoReq2 contains several requirements that explain how migration processes need to take this into account.

2.3 Automated Rendition

MoReq2 requires that systems have the ability to render all electronic records to a preservation format of choice at the time of capture. So, for example, if PDF/A [8] is selected as the preferred preservation format for "four-cornered" records, then a compliant system would have to be able to render, automatically, all appropriate documents (so not audio, video etc.) into PDF/A format as soon as they are captured, subsequently managing both the original and PDF records in parallel. This relatively simple feature, which we believe to be totally novel to standard specifications, will provide a major boost to efforts to build repositories of preservable and accessible electronic records.

2.4 Import and Export Features

MoReq2 contains detailed requirements supporting the import and export of electronic records. Unlike most other specifications, the requirements explain in detail how not just the records themselves are involved, but also their metadata and audit trails. Details covered include error handling (in the event of duplicates or incomplete data structures – both major concerns given the complex hierarchical structures associated with records management) and protocols to confirm successful exchange. These features can be useful to export records from an ERM or EDRM system to either a preservation system or a newer system.

2.5 Preservation Metadata

MoReq2 contains a uniquely detailed metadata model [3]. It consists of 158 different metadata elements, each cross-referenced with the appropriate entities to which it can apply (component, record, file, class etc). In effect this defines 345 distinct metadata elements – more than any other electronic records management specification. The elements are chosen to provide all the metadata needed for full Moreq2 compliance.

Not only is the metadata model extensive, it is also detailed. Each element is described in some depth, in the form of a table. An example is shown in Figure 2 (note that the name of the element is structured to comply with ISO 23081 [9]).

M142: Use.technical_environment.file_format_version_original

	Obligation:	Mandatory	Occurs:	Once
Definition:	The version of the file format in which the component was encoded at time of capture.			
Populated:	Populated automatically by ERMS at time of capture.			
Source	See comment.			
Default	None.			
Use conditions:	Cannot be changed.			
Comment:	For this to be useful, it needs to be populated from a controlled vocabulary that is continually maintained as file formats evolve. Ideally, the vocabulary is taken from an established file format registry. At the time of writing, the leading file format registry in Europe is PRONOM, see http://www.nationalarchives.gov.uk/pronom. It is acceptable for this element to be combined with the element M133.			
Requirements	11.7.13			

Figure 2: Example of a MoReq2 metadata element

The example in Figure 2 is one of the simpler elements. Other metadata element descriptions specify not just the definition of the element, but how it is used, how it relates to other elements, and in many cases from where its values may come, how they are derived, and what values are valid. Contrast this with a comparable definition from DOD 5015.2, shown in Figure 3:

	Record Descriptors	
C2.T3.9.	Format	RMTF (Reference (y))

Above: Definition in DoD 5015.2

RECORD_Format_Code (New)
Codes indicating logical structure of the record.

Above: Definition in the reference "y" cited by DoD 5015.2

Figure 3: Example of a metadata definition from DoD 5015.2

2.6 XML Schema

The final piece of the jigsaw is the MoReq2 XML schema [5]. This expresses the metadata model in XML, and provides – hopefully – an unambiguous mechanism to allow the exchange of electronic records, including their metadata and audit trails, between systems. At this stage, it is just "hopefully" as there is no credible mechanism to test the correctness and completeness of the schema itself. However, we are hopeful that, as soon as software claiming to comply with MoReq2 becomes available, it will be possible to test the schema at least partially.

3. THE FUTURE OF MOREQ2

We already know that all of the major ERM/EDRM system vendors have MoReq2 compliance projects under way. Already, buyers in Europe are specifying MoReq2 compliance as (in some cases) mandatory. The near future will most likely see MoReq2-compliant products reach the market, first in Europe then beyond.

The original MoReq specification was very popular, but unfortunately it suffered from a lack of governance. Once published, MoReq had no ongoing supervision or management at all – it effectively was orphaned.

This time around, the DLM Forum has set up a solid governance structure to manage MoReq2 on an ongoing basis. It proved challenging to get this started, due in part to the fact that the governance is to be provided on a voluntary basis; but early indications are hopeful. The new MoReq2 Governance Board (MGB) will:

- oversee the software compliance testing regime;

- ensure the "brand" names MoReq and MoReq2 are not misused, and in particular ensure no false claims of compliance are published;

- monitor the accuracy and acceptability of translations and of the localization chapters;

- look after the maintenance and ongoing development of MoReq.

The MGB is made up of senior representatives of institutions from several countries with a strong interest in MoReq2. They are listed in Table 2.

Table 2: Composition of the MoReq2 Governance Board

- Hans Fredrik Berg, Senior Advisor, National Archives of Norway
- Marie-Ann Chabin, Proprietor, Archive 17 Consultancy (France)
- Marc Fresko, Director, (Inforesight (UK)
- Jon Garde, Principal Imagineer, Europe, Objective Corporation (UK)
- Ulrich Kampffmeyer , Geschaftsführer, ProjectConsult (Germany)
- Karl Mayrhofer, Product Owner, Fabasoft Records, Fabasoft (Austria)
- Thomas Rumi, MoReq2 Testing Team Manager, imbus AG (Germany)
- Jef Schram, MoReq2 Project Officer, European Commission
- Jože Škofljanec, Senior Advisor, National Archives of Slovenia
- Rory Staunton, Managing Director, Strategy Partners (UK)
- Martin Waldron, Chair, Inform Consult (UK)

4. LIMITATIONS?

Following the MoReq2 specification will not produce an archival management system. Nor will it produce a digital preservation system. What it will produce is a system that is useful to manage electronic records on an everyday basis – and one which also provides basic facilities to support active digital preservation management.

5. WHAT NEXT?

MoReq2 is not an end point in a trajectory, and is definitely not an end in itself. Rather, it is a point in a journey. As we all learn more, so it improvements will be made to MoReq. Probably the most important improvement we can think of is the integration of MoReq2 with other initiatives – DoD 5015.2 in the USA, perhaps VERS in Australia, certainly ISO 15489. There really is no plausible reason for this to continue as a solely European venture – and there is even less justification for the existence of different electronic records management standards on different continents. To be sure, uniting these different ideas would be challenging – but a challenge well worth winning.

6. REFERENCES

[1] Archives Canada, Records/Document/Information Management (RDIM): Integrated Document Management System for the Government of Canada, 1996. http://web.archive.org/web/19971021032801/www.archives.ca/www/english/mgr/order.html

[2] European Commission, MoReq: Model Requirements for the management of electronic records. INSAR (Information Summary on Archives) supplement VI, 2002, ISBN 92-894-1290-9. Also available at http://ec.europa.eu/transparency/archival_policy/moreq/index_en.htm

[3] European Commission, MoReq2 Specification: Model Requirements for the management of electronic records, update and extension, 2008. ISBN 978-92-79-09772-0. Also available at http://ec.europa.eu/transparency/archival_policy/moreq/index_en.htm

[4] European Commission, MoReq2 Test Framework, 2008. http://www.inforesight.co.uk/downloads

[5] European Commission, MoReq2 XML schema, 2008, http://www.inforesight.co.uk/downloads

[6] International Council on Archives, Principles and Functional Requirements for Records in Electronic Office Environments (in three volumes), 2008. http://www.ica.org/en/node/38972

[7] International Organization for Standardization, ISO 15489-1:2001, Information and documentation -- Records management -- Part 1: General

[8]   International Organization for Standardization, ISO 19005-1:2005, Document management -- Electronic document file format for long-term preservation -- Part 1: Use of PDF 1.4 (PDF/A-1).

[9] International Organization for Standardization, ISO 23081-1:2007 Information and documentation -- Records management processes -- Metadata for records

[10] The National Archives (of the United Kingdom): Requirements for electronic records management systems, 2002 revision. http://collections.europarchive.org/tna/20080108102455/http://www.nationalarchives.gov.uk/electronicrecords/evaluation.htm

[11] US Department of Defense, Electronic records management software applications design criteria standard, 2007, DoD 5015.02-STD. Also available at http://www.dtic.mil/whs/directives/corres/pdf/501502std.pdf