Markus Merenmies 1.9.2005

Archival requirements in record keeping systems

1. Introduction

Preservation of electronic records and ensuring their future usability is a widely recognized major problem. The need for unification of the ways of preservation and presentation of electronic records and the corresponding need for technical requirements for transfers of such records to custody are important issues in the National Archives Service (NAS) of Finland. The NAS has through Sähke-project developed requirements for records management systems defining how records should be created. The project is now developing the reception process and the digital archives as well as services for the electronic records.

The Finnish Archives Act (831/1994) defines ‘record’ as a written or pictorial presentation or such an electronically or otherwise produced presentation, which can be read, heard or otherwise understood with the aid of technical equipment (831/1994, 6 §). Electronic records determined to be preserved permanently shall be drawn up and the information therein shall be recorded with materials and methods appropriate for long-term preservation as separately ruled by the National Archives Service (831/1994, 10 §).

2. Requirements for the records management

2.1. The Sähke-project

The National Archives of Finland launched in 2001 the so-called SÄHKE-project in order to find solutions to problems related to the preservation of electronic records. Attention was at that time focused on electronic records management systems (ERMS). A 'records management system' is defined by the project exclusively as a type of system, which includes a registry of the cases and their related documents within the system.

The National Archives of Finland published in 2003 requirements as a result of the SÄHKE-project in order to ensure the integrity and usability of records in electronic record management systems (ERMS) during every phase of the life cycle of records. In 2004 a pilot venture of the SÄHKE-project was launched, in order to test those requirements in practice by implementing a records management system consistent with them. A second version of requirements was published in the February 2005. ERMS must fulfill these requirements in order to ensure that the

records in the ERMS can be delivered to archival institutions in electronic format, and be accepted by them.

An other important task of the SÄHKE-project is to determine what characteristics should be assigned to electronic document preservation in order to ensure its legal value. Registration metadata and the collecting of them are important matters that largely affect the legal value of such documents. The SÄHKE-project is focusing on the archival aspects of the system.

As a result of updated requirements of the Sähke-project the National Archives is updating the norm on preservation of electronic records. This norm covers whole records management process from the creation of records to transferring them to the archives and how information systems, which include electronic material which should be preserved permanently, should be handled.

2.2. Planning the preservation

In Finland each authority must have a records management schedule for its records management. The schedule contains directives about the registration, filing, retention periods and publicity of records. Records are grouped according to the functions to which they relate. In a schedule which is drawn up to the accuracy of record title. The life-cycle of a document is considered in advance when a records management schedule is drawn up and authorities have to determine the retention periods of their records. The National Archive Service determines which records shall be preserved permanently.

Thus, the functional requirements will concentrate on the building and maintaining of records management schedules and associated metadata. The system shall assign the metadata as automated as possible and the assignment shall be guided by default values planned in the records management schedule. Consistent structure in ERMS ensures that default values, like preservation period and publicity, can be assigned automatically. Functional and metadata requirements include several requirements which presume that the management of records and the archival functions are planned in advance properly.

Our requirements and Moreq requirements for the management of electronic records have many similarities, but we believe that our requirements concerning guidance of preservation by using records management schedules would benefit the Moreq standard. According our requirements the records management schedule has very active guiding role in the process; it is not just a passive list of retention periods.

2.3. Sähke-requirements

The Sähke-requirements consist of three separate models: operational, technical and abstract, each focusing on ERMS from a different point of view. A functional model describes the requirements of managing the life-cycle of records, i.e. how records should be handled in ERMS and how metadata should be gathered. The second part is an abstract model describing the required metadata model. The technical requirements constitute a third part, which describes how records are transferred to the archives. Technically it describes standards to be used for files and how documents should be converted.

3. Transferring electronic records to archives

Permanent preservation of electronic records requires permission from the National Archives. When permission is applied for the authority must assure that ERMS fulfil SÄHKE and other requirements about registering records.

Records which should be preserved permanently in electronic format have to be delivered to archives. In order to make it possible we had to develop methods and interface for transferring electronic records to archives in a consistent way. As a starting point we decided to limit the transfers to records and related metadata, and to postpone the transfers of functionalities or ”look-and-feel” of the ERMS.

The functional requirements include the production of a XML structured delivery file. It is of vital importance for successful reception that the delivered material is well formed and in accordance with the regulations of the National Archives. The archivist's role in the system is to produce a standards compliant delivery file, check its compliance to the system’s internal data and seal the file for delivery. Before making the delivery file, the archivist has to perform the disposal of non-long-term material with the assistance of required retention metadata.

A key requirement which ensures that electronic records can be delivered to archives is the metadata model which is used for controlling processes and describing the records. Functional processes must support saving values for defined metadata fields of the cases and related records so that integrity and authenticity can be guaranteed permanently.

Electronic records with corresponding metadata and documents are transferred to archives using xml-structured delivery files. We have an defined xml-schema which defines required metadata fields. All the records are converted to tif-images during the preparation process of the delivery file. Metadata of the record has a link which connects actual records to it.

We suggest that Moreq also adds definitions of structured delivery files to its requirements.

4. Preservation system

The SÄHKE-project has defined a unified model for producing electronic records and metadata to be preserved permanently in electronic format. Permanent preservation of electronic records requires permission from the National Archives. When permission is applied for the authority must assure that ERMS fulfil SÄHKE and other requirements about registering records.

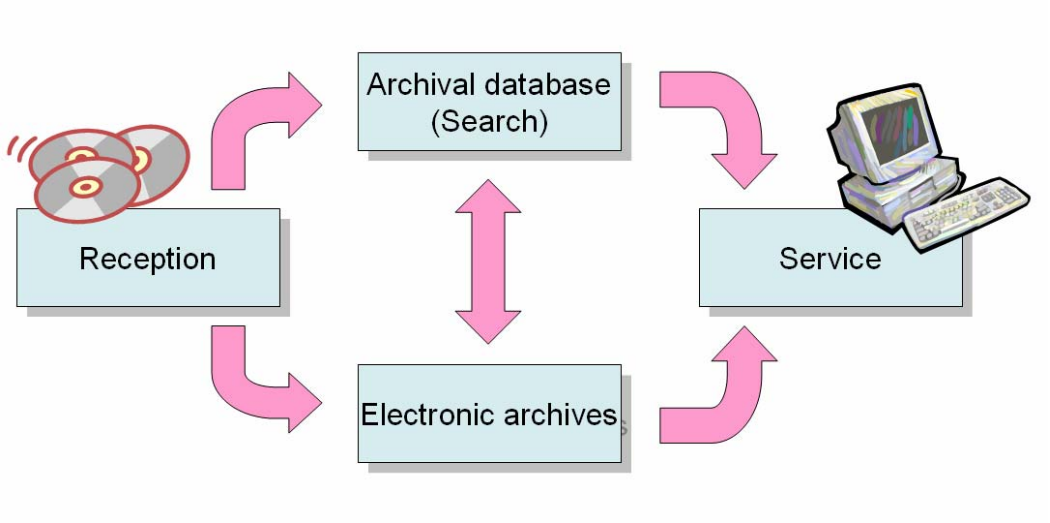
The next step in this process is to develop a reception process, electronic archives and services.

The perception process starts with the production of a delivery file from the records management system. This delivery file consists of both xml-files for metadata and tif-images for records. Due to the aging problems of the modern word-processing formats the National Archives of Finland has decided to accept records only if they are converted to tif-images (CCITT/ITU-T Group 4, tif v. 6.0). PDF format is not accepted as a delivery format.

Metadata values of the cases and records will be stored as xml-structured files, and connection between records and metadata of the records is made by using file-

references in the metadata. The SÄHKE-project has defined an abstract model of the information that should be stored into records management systems. This metadata model serves as a base for the structure of the delivery file.

An electronic archive and service system according to our plan consists of four logical parts: reception, storing, search and service.



Reception

When a delivery file is received in an archive, the first step is quality assurance. The descriptive metadata of delivered material must be checked in order to ensure that it is in accordance with the delivered material. Metadata values of the cases and records must also be checked, to ensure that all the required fields are properly filled. This can be done automatically, by using xml-testing tools against defined xml-schemes. The quality of the conversion of the records has to be controlled by random selection of a few records to ensure that its technical quality is acceptable.

The reception part of the system receives delivery files and performs quality assurance for the material. Descriptive metadata of the material will be stored in a search database (VAKKA) for retrieval purposes and finally both registration metadata and the records themselves will be stored in the electronic archives.

When the delivery file has passed quality check it will be stored in the electronic archives. The whole delivery file will be stored as a “media image” in the off-line electronic archives. Storing delivered material this way makes it possible to ensure that delivered material can be protected against any changes. In addition to this storing in an off-line “media image” store the delivered material will be stored also in the on-line archives database. Descriptive metadata will be stored in the archival search database and records will be stored in the file system so that they can be linked together.

Archival database

The search functionality in an archival database supports both traditional “paper based” archives and electronic archives equally. All the required metadata values for searching records are stored automatically from the delivery files.

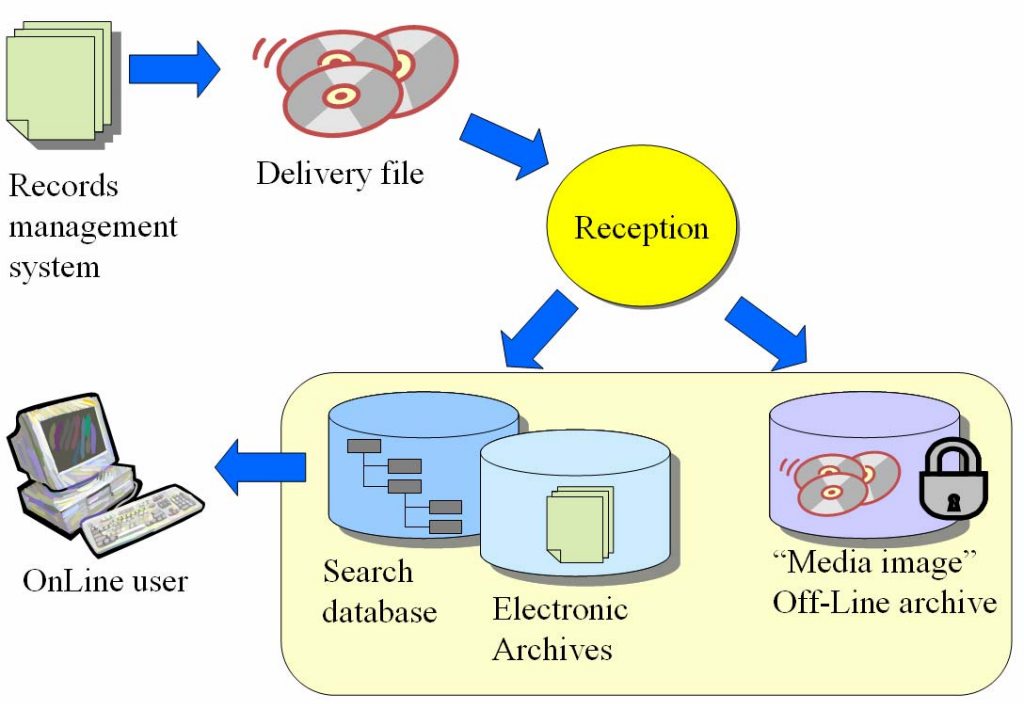
Electronic archives

The electronic archives stores both registration metadata and records. This system consists of computers, databases, storage systems, an information management system and a network. The National Archives of Finland has chosen migration to modern standards as a preservation strategy for electronic records. Continuous monitoring and transformation to modern standards is the basic principle of preservation to ensure usability and permanence of records in the future.

Service

The administration in Finland is based on the principle of publicity. Researchers have the right to use and to study all official records of the public administration or in the custody of the National Archives. The right of access to public records is affected by certain exceptions prescribed on the grounds of national interest and the protection by law of the individual or society. This principle of publicity imposes preconditions to preservation and presentation too. The principle of publicity requires also the electronic archives to provide proper information service. This on-line access to the electronic records can be made through a search database. Researchers can search for electronic records in the same way as they search for paper ones.

The service part of the system offers an interface to the electronic archives together with search functions.



5. Conclusion

We all agree that strong proactive efforts are required. Preservation cannot be seen as a passive "wait-and-see" style action. We strongly emphasize the importance of the creation part of the life cycle. The system shall assign the metadata as automated as possible and such assignment shall be guided by the default values planned in the records management schedule.

The present Sähke-requirements focus on functional, abstract and technical requirements for electronic records management systems in order to produce electronic records, especially to assure that integrity and authenticity of records can be archived. Different kinds of electronic data, which should be preserved permanently, are however produced in various kinds of formats, like databases and registers. There are also other different electronic formats for special purposes, like electronic drawings, 3D-models and others which can only be used in electronic format, but cannot be archived in a same way as common records. Further research should be made with the aim of developing archival methods for these problematic electronic files and creating similar processes for transfer of those to archives.

The National Archives of Finland aims to offer high quality and centralized preservation and service solutions for public administration. At present there are strong efforts if Finland to develop national IT-strategy and we are emphasizing the benefits and requirements of the archival of electronic records in the electronic format.

To define functionalities of the reception process and service requires strong efforts from NAS. International cooperation and development of Moreq standard is very valuable asset which guarantees cost effective and competent solution and also guarantees the development process also in the future.

Kansallisarkisto
Riksarkivet
Rauhankatu 17, PL 258, 00171 HELSINKI
Fredsgatan 17, PB 258, 00171 HELSINGFORS
Puh./Tel. (09) 228 521
Fax (09) 176 302
kansallisarkisto@narc.fi
www.narc.fi