

## **THE PROCESS OF TRANSFERRING ELECTRONIC ARCHIVES TO THE NATIONAL ARCHIVES OF NORWAY**

**Terje Pettersen-Dahl** \*

**UDC: 930.253:004.3**

*Terje Pettersen-Dahl: The process of transferring electronic archives to the National Archives of Norway. Technical and Field Related Problems of Traditional and Electronic Archiving. Conference Proceedings, Maribor 4/2005, No. 1, pp. 121-127.*

*Original in English, abstract in English, summary in German.*

This paper deals with the process of transferring electronic archives. It covers the phases from the surveying of electronic government records to the storage of records for long-term preservation as electronic archives. The paper also includes a short list of standards and tools used in the (testing) process.

### **APPRAISAL AND PRESERVATION**

The first phase in the appraisal process will be the survey. Government institutions in Norway are required to inform The National Archives about their systems, but it is necessary even to visit and inspect institutions to get the full view.

Normally this means that we have to do appraisal on existing systems, which in many cases have no functions related to preservation. Valuable information can be lost due to lack of adequate capture functions in systems. This makes it impossible to get the history of the records. When records capture functions are absent in systems, preservation must be limited to snapshots of the records.

It may also be difficult to identify the logical records in the technical structure of the system, yet again due to lack of capture functions. And in some cases even system documentation is missing.

To enable appraisal The National Archives have started projects surveying complete sectors of government (police, education, tax, health, agriculture, etc.). One experience is that electronic systems are to a large extent shared by the institutions within each sector. The projects can either be confined to electronic archives, or comprise both paper and electronic archives. The projects work closely with the government institutions and their records management staff. The result of the projects will be either a plan defining permanent records and their transfer to The National Archives, or a plan of disposal. The project aims are to identify shared systems (with a central database), shared systems combined with local paper archives, identical systems used locally with local databases and unique systems (local or central).

We even aim to do appraisal before the systems come into use. Then we need to define the information (the logical records), which is to be preserved and transferred to The National Archives later on. In this way we can ensure that

---

\* *Dr. Terje Pettersen-Dahl, Adviser at The National Archives of Norway, Department of electronic archives, Postbox 4013 Ullevål Stadion, N-0806 Oslo, Norway.*

functions for capture and transfer (export) of records of archival value are implemented.

## LAWS AND REGULATIONS

Norway has a national archival law, which requires government institutions to transfer an archival copy of every system, which The National Archives find worthy of preservation.

The archival law is supported by general regulations on record keeping, appraisal and transfer to The National Archives. These regulations also make the Noark standard compulsory for electronic registries and office systems with registry functions.

Additionally the law is supported by exclusive regulations on transfer of electronic archives. These regulations support The National Archives metadata- and format standards. These regulations can be found in English at [www.arkivverket.no/noark-4/Regulation-ch8.pdf](http://www.arkivverket.no/noark-4/Regulation-ch8.pdf), and the regulations on documents can be found at [www.arkivverket.no/noark-4/Regulation-ch9.pdf](http://www.arkivverket.no/noark-4/Regulation-ch9.pdf).

The archival law defines the concept of archival record as documents being created as part of a professional activity ("business"), with the general exception for documents, which have no business functions or value as documentation. Office records can be electronic if the systems, procedures, media and formats for storage conform to the requirements in the regulations. A copy of electronic records should be transferred to The National Archive as a deposit when the system is out of business. If the system is in business for several years, it is customary to transfer deposits with 5 years intervals. In registry systems, which conform to the Norwegian Noark standard this is done by organizing the records in periods of 5 years. Complete records periods are then deposited successively as they are terminated.

According to the Archival law, the National Archivist gives specific rulings on preservation or disposal. A ruling on disposal implies that all records, which is not mentioned, are to be preserved. This works for paper archives, but rulings solely on disposal is not sufficient for electronic archives. Electronic systems require rulings on preservation (and more than just the mere name of the system is then needed). Rulings must identify the logical records, which are to be preserved if the system is a records management system. Otherwise rulings at least must identify the exact types of information or functions to be preserved.

It is not required to specify the extract methods for transfer in the rulings, but it must contain the necessary directives to define the extract method. It should be possible to confine transfers to logical records (or other defined elements), and to organize the extraction accordingly. It is preferable that defined records of archival value, is put aside successively in the system during production.

## CREATION OF EXTRACTS FOR TRANSFER

Plans resulting from The National Archives' appraisal projects define permanent records from the various branches of government. The plans make it known to the government institutions from which systems they have to produce extracts for transfer to The National Archives. (It is impossible to receive the total system, so we only want extracts of defined information from the systems.)

The institution itself will have to produce the extracts through their own staff or rely on hired consultants to do the job. The production process may require some adjustments in the individual permanent retention provisions. The extract will then be transferred to The National Archives as an archival copy.

Transfer extracts contain information from database tables only. The forms in which the information is presented in the system and the logic in these forms will be disposed. The information must therefore be separated from the system by which it is managed. This requires a process of reorganizing and reformatting before transfer to The National Archives. The information must be extracted into text-files (flat files). Additionally a description of structure and content (technical metadata) must accompany the files as documentation as well as other documentation - system documentation, user manuals, etc. With the metadata - and other system documentation - at hand we can later regenerate the original logical structure and import the files. It is important to emphasize that the extraction of transfer files must be done while the production system is still operational and manageable.

The National Archives is using XML syntax to describe structure and other technical elements in transfers. These technical metadata in electronic form make preserved electronic archives self-documenting. This makes it easier to regenerate the original (logical) database structure, and paves the way for user-services on preserved electronic records.

To support the process of documenting electronic transfers in government institutions, The National Archives distribute an electronic form (Arkadukt) to specify data structure in plain text. The form outputs an XML-description in the proper format (ADDMMML).

The most common database-systems (Oracle, MS SQL Server, etc.) of today are about to offer tools to automate documentation of database outputs in an XML-format. It will then be possible to convert description from another XML-DTD to ADDMMML, or it may be possible to document an output from the database-systems directly in ADDMMML.

Be aware that electronic documents in the transfers, has to be related to the new structure.

## **ARCHIVAL COPY**

When the transfer of an archival copy is received at The National Archives, the copy will be tested. The first test is to ensure that the archival copy is complete. In addition to the testing of the data-files it is tested that the transfer contains the necessary documentation.

In many cases technical metadata are received on paper - and have to be converted to an electronic format. Conversion from paper format to electronic format is then done by our own staff at The National Archives by using Arkadukt. This conversion has to take place before we can do a full test of the archival copy. When both the metadata and the data-files are electronic, the actual testing can be performed.

With the tool Arkade we are testing the consistency and correctness of the archival copy, unless the original system is a Noark-based registry. In Arkade we can also convert the data-files into the standard format, which we use for long-time

preservation (if needed). If we later should decide to change the format for long-term preservation, we can use Arkade again to do the conversion.

If the original system is a Noark-based registry, the tool ArkN3 (or ArkN4) is used for testing instead. In this case metadata describing technical structure are standardized and published, and not required as part of the transfer.

After the test, the archival copy is stored for long-term preservation, i.e. if accepted, otherwise we need to get a new and corrected archival copy, and then do the testing process once more.

## **STANDARDS AND TOOLS**

The National Archives have defined two standards to support the management of electronic archives:

- ADDMML to describe technical structure in transfers from database-systems.
- Noark, which defines functional requirements and transfer formats in registry-based records management systems.

In addition to these standards, The National Archives have developed several electronic tools for testing of archival copies and for description at various stages in the process.

### **ADDMML (ARCHIVES DATA DESCRIPTION MARK-UP AND MANIPULATION LANGUAGE)**

The National Archives' own standard for describing the technical metadata, which are to accompany transferred extracts from databases.

The first version of ADDMML was developed in 1998. The last upgrade took place in winter / spring 2005. Current valid versions of ADDMML are 7.3 (2002) and 8.0 (2005).

ADDMML is a markup language based on XML, and it is defined by a DTD.

ADDMML has three major parts - references, structure and processes. References is describing the context of the original system (contextual metadata), all though very briefly. Structure is the main part, and contains the description of the archival copy (technical metadata). Processes is not a part of the description, but gives us the possibility to specify certain instructions (processes) to be performed on the archival copy for testing (in Arkade). The structure in the Structure part, consist of 5 levels: dataset, file, record-type, field-type and codes. The dataset is the entire archival copy, which can consist of several files, each with several record-types, etc. A table in the original database is normally corresponding to file (relational databases). In some older types of databases (hierarchical, etc.) both the file and record-type level will be corresponding to the original table.

### **NOARK (THE NORWEGIAN RECORD MANAGEMENT STANDARD FOR SYSTEMS WITH REGISTRY FUNCTIONS)**

The first version of Noark was developed in 1984. The current valid version is numbered 4.1 (1999/2002). Version 5.0 is scheduled to be defined this year.

The scope of the Noark standard is office records, i.e. office documents embedded in record-keeping metadata. Office documents may be defined as letters, memos, internal notes, e-mail messages, etc., while record-keeping metadata means identification of records, time of records capture, content-, context- and structure information, etc.

Noark's functional requirements may be implemented in stand-alone registry systems or in general or specialized office systems, which include registry functions.

The basic idea of Noark is - from an archival perspective - to ensure that record-keeping systems conform to certain standards for record-keeping functionality and metadata, and to meet the requirements for keeping records and metadata over the short and long term. The benefits from a records management perspective are systems based on acknowledged principles for record-keeping available for use, as well as standardization benefits in connection with changes in organization, education of staff, system maintenance and transfer to a repository.

The Noark standard specifies the information elements (what metadata a Noark system must and can include), basic data structure (the structure of these information elements and the relationship between them) and functionality (what functions a Noark system must support - at least as a minimum). The standard does not specify the user interface (how the system communicate with the users), implementation of the data structure (technical design), functionality (beyond the basic requirements specified) or technical implementation in general.

All Noark systems must have export functions for record-keeping metadata and electronic documents. There are two different export formats, one for transferring the data from one Noark system to another, and one for transfer to the repository. For record-keeping metadata the export format is fully specified in XML. For electronic documents the export format must be one of the document formats, which are approved by The National Archivist for archival purposes. At present the approved archival formats are ISO 8859-1 and (plain text), TIFF, XML and PDF. All Noark systems must also have functions for importing data from other Noark-compliant systems. Imported metadata and documents will then be accessible as in the original system.

The next generation (5.0) of Noark should not be confined to registry-based records management only, but serve a plurality of records management functions. Noark should use a basic module, which is general and flexible enough for this purpose. The basic module should support a variety of business processes, and registry management should only be one of these. The basic module should also support multiple processes in parallel. The basic module will use ideas from ISO 15489, MoReq and the Norwegian health service registry standard (EPJ).

The next generation of Noark systems should also be able to interact better with specialized office systems. The potential of electronic document archiving should be exploited more extensively and different types of communication should be supported (the current generation only accepts e-mail).

Part I of the Noark standard is available in English on this link: [www.arkivverket.no/english/electronic.html](http://www.arkivverket.no/english/electronic.html).

## **ASTA AND ARKIMEDES**

Asta is The National Archives' system for describing and retrieving transfers of all kind. Asta (in today's version) do not handle electronic archives, therefore do we have an additional system for electronic archives called Arkimedes. A new version of Asta will be ready in 2005 to replace Asta (old version) and Arkimedes.

The purpose of the new Asta (as well as Arkimedes) is to have a logical description of electronic archives and to have repository management. The logical description conforms to the international standard for archival description - ISAD(G) of ICA. The repository management system must take into account that electronic archives have other characteristics than archives created and held on traditional media (paper and parchment).

## **ARKADUKT**

The National Archives' own developed system for registration of technical metadata. Arkadukt creates an ADDMML-file.

When the technical metadata related to the archival transfer is not electronic, it has to be converted from paper. This is done by registration of the metadata into Arkadukt to create an electronic metadata-file, in accordance with the ADDMML standard. The National Archives distributes this program free of charge to enable government and municipal institutions to create the electronic metadata-file themselves. (Otherwise this has to be done at The National Archives.)

Arkadukt is able to read old ADDMML-files and convert them to the current version of ADDMML.

Arkadukt is also able to perform logical testing of the metadata and their structure.

## **ARKADE**

Arkade is a system for testing transfers to The National Archives based on ADDMML.

Arkade is developed in SAS, and can perform various tests and analysis on the archival copy. Arkade controls the consistency between the actual data-files and the metadata, both that the metadata reflect and document the correct structure of the data-files, and that the metadata information itself is correct (i.e. number of records, etc.). Arkade can perform analysis on different levels, as well as checks and controls. Additionally a few special functions can be performed. Arkade can do conversion on the archival copy from one set to another (i.e. convert character-set, etc.), to create conformity on versions for long-term preservation. Arkade can also create a basis (SAS-dataset) for public versions of the archival copy.

## **ARKN3 AND ARKN4**

Systems for testing and presenting transfers based on Noark (version 3.0 and 4.0).

ArkN3 and ArkN4 import archival copies from Noark systems, and perform a consistency check on the data. Even these programs can perform a simple analysis.

ArkN3 and ArkN4 can also present the information for public use.

## ZUSAMMENFASSUNG

### ÜBERNAHME ELEKTRONISCHER AKTEN DURCH DAS NORWEGISCHE NATIONALARCHIV

Dieses Referat beschäftigt sich mit dem Vorgang der Übernahme elektronischer Bestände durch das Norwegische Nationalarchiv - von der Begutachtung bis zur Langzeiterhaltung. Die wichtigsten Punkte werden in einem Überblick erfasst, welcher entsprechende Maßnahmen für die Aufbewahrung (oder Skartierung), Übernahme, Reproduktion und Langzeiterhaltung der Dokumente beinhaltet. Das Referat befasst sich auch mit dem Einfluss von Gesetzen und Verordnungen auf diese Tätigkeit und bietet am Ende eine Beschreibung Norwegischer Standards und Erfahrungswerte.

*Terje Pettersen-Dahl graduated from the University of Oslo, Institute of Informatics in 1980. He started his career working for the Central Bureau of Statistics in Norway. From 1985 to 2000 he worked as a consultant for different private companies in Norway. In 2000 he joined The National Archives, where he is working with electronic archives.*