

Long-term preservation of electronic records
from official agencies in Iceland

Eiríkur G. Guðmundsson
National Archives of Iceland

The ICA - SUV seminar in Reykjavík
13.-20. September 2006

14th September 2006

Ladies and gentlemen!

It is a pleasure to be able to talk to you today about an issue that is very important for archives today.

1.

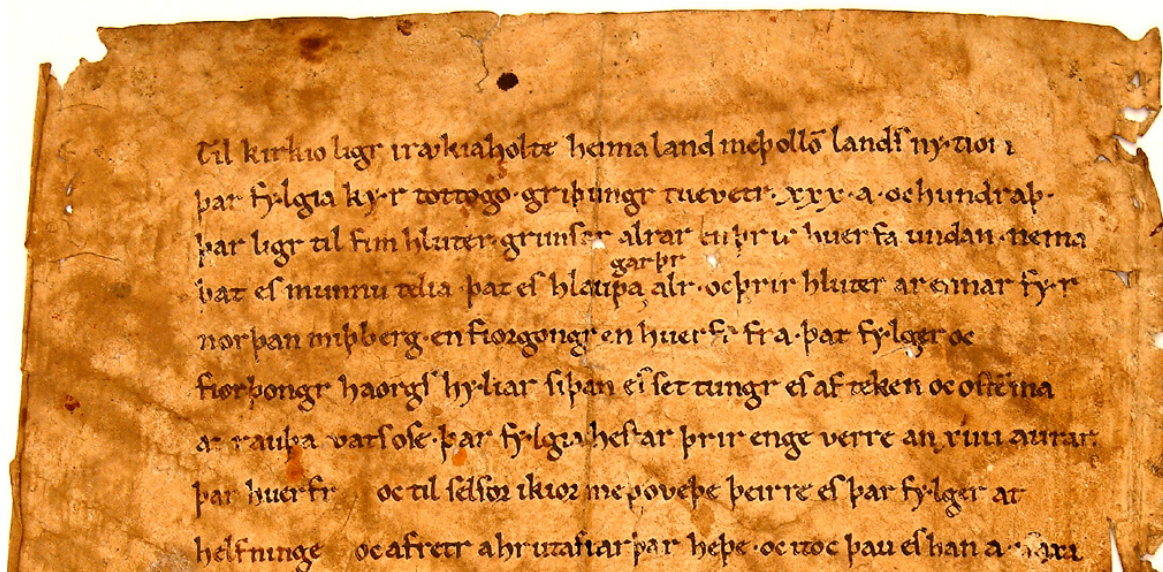
I will speak from the point of view of an employee of the National Archives of Iceland (NAI). The NAI preserves records from official agencies. Simply put, those who receive more than 50% of their operating income from the state treasury are obliged to transfer their records to NAI in due course. The same applies for the municipal-sector, if they do not have a regional archive to deliver to.

We all participate in archival work of some sort and we all know that official archives exist in order to preserve different information for various reasons. Information such as government resolutions, interactions between subjects and official authorities, communications between government and municipalities, companies, foreign states, to name a few examples. Official authorities have for centuries recorded information on society in different form and for various reasons. We have inventories from the middle ages, gift certificates, territory border declarations, tax reports, censuses, parish records, catechetical list, estate evaluations, sentences and court documents, various reports and letters, etc.

Records in archives reveal the coherence of past resolutions and form a basis for logical decisions based on resolutions of the past, which can be found in older documents. The information found in archives are relevant to citizens and authorities alike, right now as well as in the future, and ensure the rights of both and at the same time serve as a material for the history of citizens and society.

2.

On the screen you can see a part of the oldest document preserved in the NAI. It dates back 800 years. We believe that the eldest part of this document was written around 1185. This document (Reykholt Church Inventory) is in good condition and easily readable. It was written on a rather reliable media, i.e. parchment.

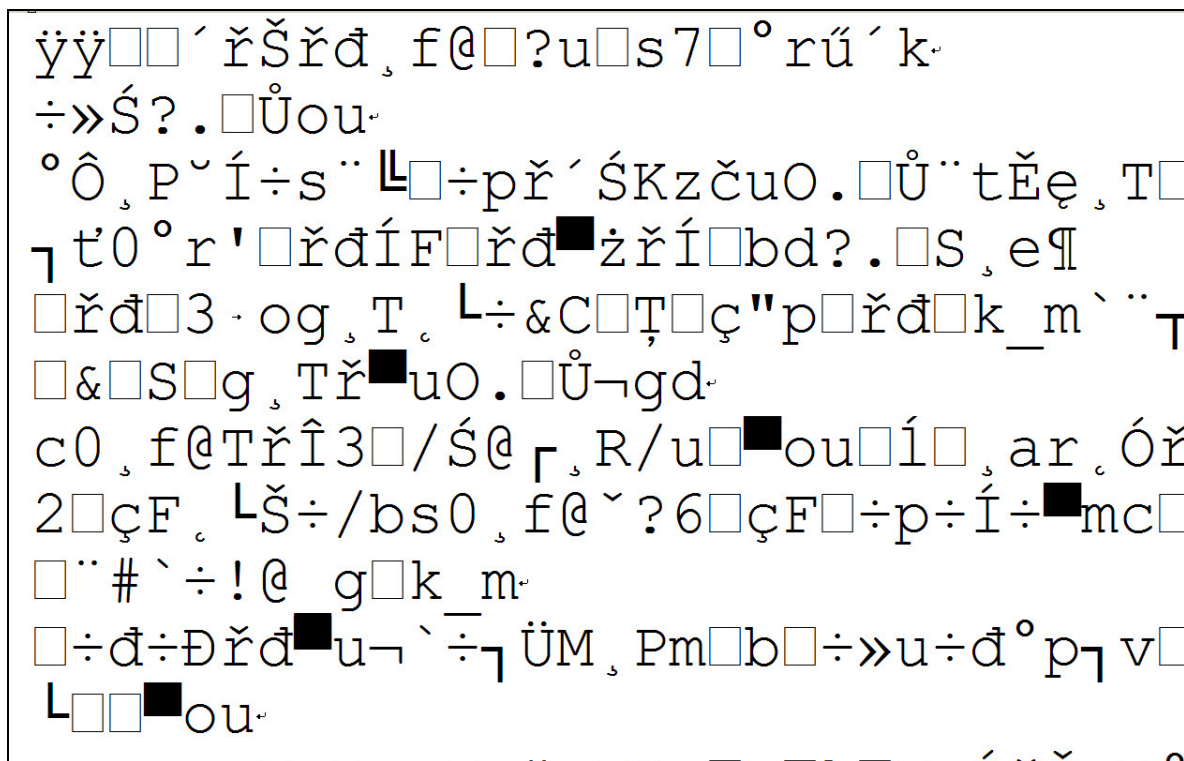


Church Inventory (Icelandic: máldagi) is a document that lists the property of a church, both land and other property. In our context, it is important to realize that this document is still relevant today. It has recently been used in an extensive case that has been going on for almost a decade and will most likely go on for another decade at least. The objective of this case is to determine who owns the uninhabited highlands of Iceland. The Icelandic government claims that a big portion of that land should be declared national territory. From the settlement period around 870 it has been the practice in Iceland that land has been a private property of those who utilized it, be they farmers, churches, or communities. In this situation, present landowners look to the archives for safe sources to support their claims to ownership of their land.

I mention this in order to illustrate that old records can supply information of important rights of citizens and society for a very long time. When we speak of long-term preservation, we are not referring to a few years or decades, but centuries. This holds true for all records, paper documents, and electronic files alike. We all understand that the problem facing modern archives is the long-term preservation of electronic records. How can we ensure that electronic records can still be readable after 800 years, or even only 100 years?

3.

The question is based on the fact that computer files are sensitive in many respects. Computer files are made with technology that needs electricity both when the files are made and used. The hardware and software used are subjected to rapid changes and new discoveries, mostly governed by the market, and both hardware and software are often inadequately made.



I suspect that we have all, at one time or another, been horrified when seeing a picture like the one we see on the screen. We intended to read an old file that we made and saved on our old computer or on a diskette a few years ago. In the meantime, we upgraded our

equipment and then we were unable to read our old file. In fact, it takes some effort to upgrade computer files so they are readable at any time. Moreover, if the individual has some trouble preserving computer files so they can be read few years later, then it is clear that the problem facing an entire nation is enormous.

It is obvious to everybody that this problem must be solved and computer files must be preserved in a reliable manner so that they can be read in the future. The fact is that most nations are still in the early stages of tackling this problem and comparatively few archives have actually accessioned electronic records in an organized manner so that it can be ascertained that they can be read at a later date. I am not talking about tape archives in this context, as it is very likely that it would be very expensive to make sure that they can be used in the future.

4.

In recent years specialists have talked about three possible ways of tackling the problem of preserving electronic records.¹

The museum-method means that an archive collects computers, operating systems, and programs of all times, so that computer files can be read using the equipment they were made with. Is this a good method? No. It is obvious that this method is impossible in the long run. It is very expensive, if possible at all, to collect all this equipment and make necessary upgrades. We would end up with a vast unusable computer collection and an archive of unreadable files.

According to another method the archives acquire or make programs that are able to emulate the operating systems and programs that originally created the computer files in question. This method of emulation does not require that computers and programs are collected, but makes use of emulation instead. This approach is not good enough either as a general solution, if only because one cannot ensure that all programs can be emulated. We can not rely on the willingness of software producers to allow access to the code of their software and sometimes computer programs are not adequately documented.

Migration is by many considered the only feasible method today. Migration means that files are exported from computer systems and preserved in an system independent manner in accordance with international open standards. If the format of files is known, then we can be sure that those files can be read later and can then be adapted to whatever computer system that may be relevant at the time.

This method means that archives must make well-defined demands to the structure of the data-systems and specific demands to accessions from such systems. Each provenance should make its own archival version of its files according to a precise technical description, and then transfer it to the archive according to a defined process at a defined interval. Then the archive will be able to make those files accessible, convert them to new formats, and transfer them to new media according to best practices at any given time.²

¹ See Danielsen, Jan and Mortensen, Ulla: *Strategier for elektroniske arkivaliers tekniske overlevelse. Tænkkelige strategier for langtidsopbevaring af elektroniske arkivalier. Statens Arkivers migreringsstrategi.* Arkiv 5, 2005, p. 78-84. The authors claim that the terminology the Danish State Archives use are very close to the ideas of Jeff Rothenburg (and refer to: *Avoiding Technological Quicksand. Finding a Viable Technical Foundation for Digital Preservation*, ECPA 1999).

² Here one could also mention the xml approach. It is for example researched as a preservation approach, along with migration and emulation, in the Netherlands. See Potter, Maureen: *Researching Long Term Digital*

5.

There is a long-standing tradition for the collaboration of Nordic archives. Collaboration on electronic records began in the mid 1980s and has tributed to the fact that the state archives of Sweden, Denmark, and Norway are foremost among the nations of the world in this field.³

In 1998, it was decided that NAI would seek collaboration with the Danish State Archives and obtain permission to use their methods. The main reason for this is that the Danish State Archives was the first national archive in the world to start accessioning electronic records in 1996, according to specific regulation and technical standards, and of course historical, cultural and archival reasons pointed that way.

In short, collaboration between NAI and the Danish State Archives began shortly afterwards. In 2005, a written agreement was signed between NAI and the Danish State Archives. This agreement guarantees the use of the Danish method for NAI as well as access to computer programs to test accessions of electronic records, consultations, etc. This agreement ensures access to knowledge and tools and it saves the NAI a lot of work and money.

Just to mention, internationally the employees of NAI participated in the making of policies and standards of the ICA Technical Committee 2001-2005. Open standards and system independent methods such as migration are favoured there.

6.

The regulations on electronic data systems and accessions proposed by NAI are in three parts.⁴

First, there are regulations on electronic databases that are meant to ensure that NAI is notified before databases are put into use so that NAI can determine whether it should accession the database in question, how often and when. This applies to systems that record and collect information in an organized manner. Examples of this are The National Register, The Register of Cars, and Land Registry Database .

Secondly, there are regulations on ERMS for official use where cases and documents are recorded based on an specific filing system (electronic journals). These regulations are more comprehensive and concern the structure and role of the ERMS. NAI should be notified of the use of such systems and they must be approved by NAI.

Preservation Approaches in the Dutch Digital Preservation Testbed (Testbed Digitale Bewaring)

<http://www.rlg.org/legacy/preserv/diginews/diginews6-3.html#feature2> (Valid on 24. oct. 2006).

³ See To Preserve and Provide Access to Electronic Records. TemaNord 1996:549. Nordic Council of Ministers.

⁴ As mentioned they are a translation of the danish set of rules. See the danish version

http://www.statensarkiver.dk/sa/stat/lov/20020308ci_anmejo24.pdf (Valid on 24. oct. 2006).

http://www.statensarkiver.dk/sa/stat/lov/20020308ci_anmere23.pdf (Valid on 24. oct. 2006).

http://www.statensarkiver.dk/sa/stat/lov/20040311bk_aflelek.pdf. (Valid on 24. oct. 2006).

See also Nielsen, Anders Bo, Preservation of Electronic Records: Experiences from Denmark. A lecture at „The International Congress on Archives“ Vienna,, August 2004.

http://www.wien2004.ica.org/imagesUpload/pres_190_NIELSEN_A_DEN01E.pdfhttp://www.wien2004.ica.org/imagesUpload/pres_190_NIELSEN_A_DEN01E.pdf. (Valid on 24. oct. 2006).

Thirdly, there are regulations on the accessions of electronic records from databases and ERMS. These regulations are the most detailed and include a technical description for the making of an archival version of records from the data-systems described above. A considerable part of the rules on the making of archival versions deals with formats, processes, and technical matters.

As already explained the main principle is to export records from data systems in a system independent manner. Open standards should be used as well as widely used formats and media. Detailed descriptions of the creation of records, their use, and registration should accompany the records. The archives will handle upgrading and conversions to future formats.

The only problem with this method is that the records cannot be examined or used in the future as the creator used them on a daily basis, but that holds true to some extent for paper documents as well.

7.

Lets look at a simple overview of this preservation model:

The model

<i>Data creator</i>	<i>NAI</i>
<i>-reports a ERMS or database to NAI.</i>	<i>-agrees the use of a ERMS.</i>
<i>-creates an archival version every 1-5 y.</i>	<i>-appraises the ERMS/database.</i>
<i>-maintains own archive for at least 30 years</i>	<i>-decides the pace of deliverance.</i>
<i>-provides access for 30 years.</i>	<i>-tests the archival version.</i>
	<i>-stores at least 2 copies in different locations.</i>
	<i>-maintains and migrates.</i>
	<i>-provides access after 30 y. or sooner</i>

- Data (records) creator reports a ERMS or database to NAI.
- The NAI evaluates the structure and the filing plan of The ERMS, comments on it and agrees on the use of it. Appraises the database, that is decides if, and which part af the data in a particular system is to be archived at the NAI and how often archival copies shall be delivered.
- The agency creates an archival version of the data (which is to be archived) according to NAI regulation. This will happen about every 5 years for the ERMS and every 1-4/5 years for a database, depending on if and when data is renewed and erased in the database.

- The NAI tests the archival version of the data for compliance with the regulation or standards if you wish. And when the archival version is flawless, The NAI stores at least 2 copies of it on different types of media, stored at different locations.
- Then ofcourse The NAI maintains the accessioned electronic archival material and migrates them due to technical changes, new formats etc.
- The data creator maintains its own data in an electronic archive for at least 30 years (probably this will be lowered) and provides public access to the data according to regulations laid down by law for at least 30 years.
- NAI provides access to the data when it is no longer accessible at the organization that created it. The version of the elcetric data in the archives is probably going to be looked upon as the original, because of the high security.

8.

If I just browse through the formats for archival versions as they are in the regulations at hand.

Text files should be saved using the ISO 8859-1:1987 character set and should also be converted to TIFF-images. Images and scanned documents should be in the TIFF-format. Sound should be saved according the MP3 standard (ISO 11172-3). Video clips should be saved according to the MPEG2 standard (ISO 13218-2).

Using the TIFF-format for the archival version does not mean that the creator of records must use TIFF on a daily basis. It is possible to convert for example PDF documents to TIFF format when the archival version is made. Thus, the size of documents should not be a problem in the daily routine.

These specifications will of course change in the course of events and then those records will be converted to other and, presumably, better, formats.

Information in databases is exported to relational tables, which are stored as text files.

Metadata conforms to EBNF and W3C XML Schema 1.0, which will make the use and presentation of data easier. Descriptions of databases should be accompanied with common SQL queries used for retrieving data from the databases.

9.

All this creates a different setup for official agencies and archives with new tasks and costs.

The creators of records must deliver electronic records at a more frequent interval than paper documents, which must be accessioned when they are 30 years old. And also keep their data for 30 year in a good electronic archive.

There will be two copies of each set of records one in the agency one in the archive. The one in the archive is not as accessible as the one in the agency since this version is system independent and needs some extra effort to be used. The problem of using the accessioned records remains unsolved at the moment, but the development of appropriate software is underway in the Danish State Archives as well as in the state archives of Norway and Sweden.

10.

The present situation is as follows:

In 2004 the NAI conducted a survey on the records creators of the state.⁵ 354 of the biggest agencies were sent questions. 71% answered and amongst other things we found out that about 50% of them use ERMS. These agencies have almost 600 databases. It should be noted that this is not the total number of agencies. So there is a big task ahead of us.

The regulations are ready and are in for final evaluation of the government. Hopefully they will be approved and then we can start accessioning electronic archives next year.

Presently the NAI operates two test projects testing accession of electronic records for the period of 2005-2006. One of them involves the Ministry of Education and the making of an archival version of its subject catalogue. It has only just begun. The other test project is almost completed and it involves the Internal Revenue Directorate and the making of an archival version of a database.

These test projects are operated in order to test the methods prepared by NAI based on the Danish regulations I have loosely described. By operating these projects, we intend to gain knowledge, tools, and programs as well as gaining experience in accessioning and acquiring archival versions of electronic records.

So far our experience, is the same as our Nordic colleagues, and reveals that the creators of records, and system designers, must take great care in designing and development of ERMS and databases and take into account the perspective of long term preservation already at the design level. In other words to get good result in preserving electronic data you need cooperation between records creators, system producers and archivists.

It is, therefore, very likely that NAI's regulations for accessioning of electronic records will considerably improve the handling of records in the official sector for the benefit of everyone concerned. This will surely also be the experience elsewhere.

It should be noted that what has been described here marks just the beginning of long-term preservation of electronic records. Methods, standards, formats, and media will change like everything else in the course of events. But the archives can not wait for the final solution for solving the long-term preservation of electronic records, that moment will never come. They must acquire knowledge now and a valid method and start accessioning this fragile modern data before it gets obsolete.

Thank you.

Eiríkur G. Guðmundsson
Director of Information Technology
National Archives of Iceland

⁵ A report on the survey was published (in Icelandic): *Rafræn skjala- og gagnavarsla ríkisstofnana. Könnun Þjóðskjalasafns á skjalavörslu ríkisstofnana árið 2004*. Maí 2005. It is accessible on the web; http://www.skjalasafn.is/Syningar/img/Rafræn_skjalavarsla_rikkisstofnana_2004_lokagerd.pdf. (Valid on 24. oct. 2006).