



Strategy for archiving digital records at the Danish National Archives

January 2013

| | | | | | | | | | |
|--|---|--|---|---|---|--|---|---|--|
| | | | | | 0 | | | | |
| | @ | | | | | | | 1 | |
| | | | | § | | | | | |
| | | | | | | | 0 | | |
| | | | 1 | | | | | | |

Strategy for archiving digital records

This document describes the strategy of the Danish National Archives for the receipt and preservation of digital records¹. Digital records are the non-published information that is created or provided in digital form in conjunction with the activities of an authority, organisation or private individual.

The aim of the strategy is to support the overall objective of the Danish National Archives, as laid down in the Danish Archives Act, which is to ensure the preservation of records that are of historical value, or that serve as documentation of matters that are of significant administrative or legal importance for citizens and authorities. The vision is to ensure:

that digital records are preserved so as to maintain their authenticity, and so that they can be found and reused.

Challenges

Digital records have different vulnerabilities, and present different challenges, to physical records. The media on which the data is stored can be degraded, and the technology that is required in order to be able to read the media may become obsolete. In the same way, the formats in which the digital records are held may become obsolete, and knowledge of how to interpret the data, and its context, may be lost. Preserving digital information for the long term, in a form that makes it reusable, requires some deliberate choices to be made in terms of methods, technologies and documentation.

Digital preservation must also take economic considerations into account, such as the authorities' costs of the production and submission of information packages, and the archive's costs of preservation and dissemination.

Basic choice of strategy

Archives that receive digital records may choose to either receive and preserve the digital material in the original formats (native formats) and possibly migrate (normalise) the material as required in connection with dissemination; or to migrate the material to a preservation format either before or in conjunction with receipt. On migration at the time of receipt, the intention is to ensure preservation in a format that is more suitable for long-term preservation than the production format.

The basic strategy choice faced by preservation institutions is whether to pursue an *emulation strategy* or a *migration strategy*. This choice is vital to determining how the work of digital preservation is otherwise organised.

The emulation strategy is based on being able to continuously develop programs and operative systems, for example, that emulate and function in the same way as the programs that were used to create the digital records. This makes it possible to preserve the original bit flow intact, in the expectation that it can be continuously preserved and made available on all future platforms, since old programs can be run on modern machines. The emulation strategy is therefore attractive in order to ensure authenticity, as no changes are made to the original digital material. It does entail a risk, however, because there is no documentation that it will be able to function on a broad scale, i.e. as a preservation method for all types of digital documents and databases that are used in public administration. An emulation strategy will also entail considerable costs, and the method is more suitable for the preservation of other types of digital material than administrative documentation, such as computer games.

¹ The strategy does not include the research data that is received, archived and disseminated by Dansk Data Arkiv (the Danish Data Archive)

The Danish National Archives have therefore chosen to base the digital preservation on a migration strategy. The strategy entails that the creators of the archives must migrate digital records to a few, well-defined standard formats identified by the Danish National Archives for the purpose of long-term preservation. This migration may be very invasive for the material, and it is therefore an advantage, in preservation terms, to ensure that migration takes place as soon as possible after the material is created. This increases the opportunity to find appropriate tools for export and conversion, and it will also be possible to check the conversion output against data in the system in which the records were created.

The migration strategy furthermore entails that data, documents and the structures used in the Danish National Archives' preservation work must, from time to time, be migrated to new formats and structures. Any migration entails a risk of data corruption, and should therefore only take place when necessary, based on the weighing of the costs and risks of migration against the costs and risks of retaining the data in the existing format.

Non-system dependent preservation

Data preservation according to the Danish National Archives' strategy must be non-system dependent. This means that the reuse and interpretation of data must not be dependent on continuous access to the system in which the data was originally created. It must be possible to interpret and re-use data in other systems.

Non-system dependent naturally does not mean that data in the formats and structures that we today describe as non-system dependent can be retained in this particular form forever afterwards. Nor does this mean that records can be deemed to be independent of specific technologies that are widely used in a number of systems. Information technology is developing continuously and, sooner or later, the digital records collected during a particular extended period will be considered to be so dependent on a particular technological environment that the data has to be migrated to a new preservation format.

The object of digital preservation

In a traditional archive, the records are preserved in the form in which they were created. This means that the original piece of paper, and so on, that was actually created as part of the administrative process, is physically preserved. The term "original" cannot be applied in the same way to digital records. Whether data is extracted from tables in a database or digital documents, a *representation* of the content is preserved in the preservation format. It is also necessary to ensure that it is assessed which data is to be preserved; that it is decided which formats the data is to be stored in; and which documentation is required for the data to be reused.

The object that is preserved in a digital archive is thus primarily *data* or *information*. When records are submitted to the archives, it is the task of the Danish National Archives to ensure that the relevant information created as part of a particular administrative process is preserved in a way that accurately reproduces the information that the authority created and had at its disposal. The key aspect is the preservation of authentic information, even if a different technical representation has to be found for archive purposes (i.e. migrated to archive formats). In this respect it is important that the information is sufficiently documented for it to be interpreted and reused independently of the system in which the data was originally created.

Implementation

The concrete implementation of the strategy in the Danish National Archives rests on three pillars²:

- Early identification and approval of systems for submission purposes

² This implementation is focused on national authorities. For municipal and regional authorities, as well as private creators of archives, the same preservation principles apply, but a number of variations apply to the concrete implementation of, among other things, the identification and submission cycle.

- Frequent submission in non-system dependent format
- Ongoing planning of preservation and periodical migration to a new preservation format

Identification

National authorities must notify new IT systems to the Danish National Archives. In conjunction with this notification, the Danish National Archives performs a preservation and disposal assessment, so that when the new IT systems are implemented it has been clarified whether data is to be submitted or not. Systems with data that is worthy of preservation must be approved, so as to ensure that, at an early stage of the systems' lifetime, the authorities have made sure that data can be exported to the Danish National Archives' submission format.

Assessment of whether data is worthy of preservation is undertaken by the Danish National Archives' archivists. There are no statutory requirements that specific types of data must be submitted to public archives, and the archivists do not take their decisions based on fixed rules, but rather an objective assessment of the value of the data.

The assessment is based on the data's

- Reusability, i.e. whether the data can be used to illustrate concrete historical and societal issues. This is not necessarily the same as whether the data is credible or legally valid.
- Representativeness, i.e. whether the data covers a sufficiently large population, in terms of either period of time or number of individuals.
- Singularity, i.e. whether data is unique or redundant in relation to other data.

At intervals, the Danish National Archives inspect the authorities' use of IT systems, so as to ensure a good overview of which systems are used in public administration, and which systems contain data that is worthy of preservation. This overview is a necessary precondition for the planning of the regular submission of data.

Frequent submission

Data is submitted at frequent intervals, and normally every five years. In contrast to paper-based records, which are not submitted until after they are no longer in administrative use, data is submitted frequently, and without consideration of administrative use and status. The most important consideration in terms of digital preservation is to ensure that data and documents are submitted to the archives before they become technologically obsolete, and before certain data is erased in accordance with other legislation. It is also necessary to secure the required documentation while the knowledge is still held by the person or organisation that created the data. It is therefore not possible to wait to submit data until the administrative process has been completed, and the authority no longer requires the data.

This approach means that data can be held, at the same time, in the authority's administrative systems, and by the Danish National Archives. It also means the several copies of the data may be held by the Danish National Archives, because submissions from a system that has existed for more than five years will often contain data that was also included in earlier submissions. These issues can be handled pragmatically, but serve to emphasise that the traditional archivistics' concepts and models cannot always be applied to a digital world without making any changes.

Data is always preserved in a relational database structure, while any documents are stored in a file structure. A database information package may consist of:

- An extract of all data from all tables in the database
- An extract that omits data from purely system-technical tables or other tables that have no value in a non-system dependent context
- A specific extract of core data from a database or a loosely connected system

If the information worthy of preservation is not held by the creator of the archive in a relational database, but in a hierarchical folder structure, for example, the information must be converted to a relational database structure in conjunction with submission.

Data is submitted in non-system dependent information packages in which data, documents and documentation are combined, to ensure self-documentation to the greatest possible extent.

National authorities are subject to an obligation to submit records that are worthy of preservation to the Danish National Archives. As an alternative, municipal and regional authorities may choose to create their own public archive. The standard laid down by the Danish National Archives for the submission of digital records is a national standard that is used on the submission of all public digital records, irrespective of whether they are submitted to the Danish National Archives or to another public archive.

Physical preservation

The digital records' bits and bytes must be preserved intact. To protect the records as well as possible, the Danish National Archives uses distributed digital preservation, whereby data is found in several identical copies on several different types of media, both optical and magnetic, at several different geographical locations.

The OAIS standard

In preservation institutions all over the world, the OAIS standard (*Standard for an Open Archival Information System*) is becoming increasingly important as the common frame of reference for the preservation of digital records. The Danish National Archives' strategy for long-term digital preservation was developed before the OAIS standard and is thus, in principle, independent of the OAIS standard. However, the standard's functional entities are well-suited to describing the Danish National Archives' implementation of digital preservation.

The OAIS standard's functional model provides an overall view of a data flow from the records creator to the archive, and from the archive to the user.

In the OAIS standard, an archival version of data to be submitted to an archive is called a SIP (*Submission Information Package*). In preservation terms, the archival version is called the *preservation version* or AIP (*Archival Information Package*) and for issue to the archive's users a *dissemination version* or DIP (*Dissemination Information Package*) is used.

The information packages must contain the information required to be able to interpret and disseminate the content. To ensure that the data can always be read, a digital archive must conduct ongoing planning of preservation, including the required migrations at the right time.

Planning of preservation

In accordance with the OAIS standard, the Danish National Archives conducts ongoing planning of preservation, which among other things defines when the preservation standard is to be changed, and when migration to new formats is to take place.

In principle, a change in preservation standard must take place as infrequently as possible. It is also important, however, that the standard is adapted to the current requirements of both the authorities and the archives. A need to revise the preservation standard may arise, for example, when a public administration unit begins to create and store digital information that is worthy of preservation in formats that cannot be satisfactorily migrated to the preservation formats used so far. A need can also arise on a shift in technologies for data processing, so that submission can always take place as effectively as possible, based on the current technological level.

Migration of existing collections to a new preservation format should take place within a few years after the change in preservation format, so that data is migrated while the organisation's knowledge of and practical experience with formats and related programs is still intact. In the experience of the Danish National Archives, it is cost-intensive to maintain knowledge and systems for the handling of preservation formats that are no longer used on a daily basis for receipt and testing. There will therefore be an increasing risk of jeopardising the data's integrity if it is stored in formats that are no longer used actively by the organisation. In order to ensure preservation of all public records, on a change in preservation format the Danish National Archives will actively promote the migration of the information packages submitted to other public archives.

A vital condition for preservation planning is that the Danish National Archives collects knowledge on an ongoing basis of how public authorities and other record creators utilise IT technology to maintain the documentation of their activities, and which preservation solutions are developed by other preservation institutions and other parties. On this basis, the Danish National Archives continuously adjusts the implementation of its strategy, so that the vision remains attainable, and within its reach.