The Dutch National Repository of Cultural Heritage

Contribution to the ELAG conference, Bucharest 26-28 April 2006

Caroline van Wijk, project staff member Digital Preservation, National Library of the Netherlands

Introduction

My name is Caroline van Wijk and I will be discussing the pilot project that explored the feasibility of a national digital repository for digitised cultural heritage.
From January 2004 until December 2005 a pilot project called TIFF archive was carried out by the KB, the National Library of the Netherlands. This project explored the feasibility of a national digital repository for long-term preservation of digitised cultural heritage at the National Library.
I will discuss the project's previous history, objectives, results and conclusions in this presentation. The emphasis will be on the technical objective of the project: the pilot system. I will also talk about digital preservation and the connection between the research area of digital preservation and a national digital repository for digitised cultural heritage.

Previous history

Digitisation
During the last ten years many cultural heritage institutions in the Netherlands have digitised their material on a large scale. The main objective of most of the digitisation projects was to give access to a large amount of cultural heritage collections via the internet.
An enquiry held in 2005 among Dutch cultural heritage institutions shows that over 5 million TIFF files have been created in the last ten years. In this enquiry the cultural heritage institutions expressed more digitisation plans, so this amount of TIFF files will be increasing to ten million files in the coming four years.

Archiving the digitised cultural heritage?
Best practices and guidelines for digitisation are in abundance after so many digitisation projects. However, well-formulated policy on archiving the created digital files lacks at most institutions. Many institutions store the master files on CD, DVD or hard disk without a maintained administration. Digital preservation of the created master files is not yet looked upon as a special point of interest by most cultural heritage institutions. As digital preservation is an underlying concept of the pilot project, I would first like to discuss what digital preservation is and why we need it.

Digital preservation
What is digital preservation?
Digital preservation can be defined as: "The act of maintaining information in a correct and independently understandable form, over the long term (OAIS)". Safe storage and permanent access are both part of digital preservation.

Why do we need it?
Fast technology developments, lack of standards and the fragility of a digital object causes digital information to become inaccessible. An example of data loss due to outdated software and hardware is the Domesday project by the BBC.

In the case of the Domesday project, digital archaeology has saved the detailed snapshot of British life in the mid-1980s. The project used the potential of multimedia current in the 1980s. However, the special computers developed to play the 12in video discs of text, photographs, maps and archive footage have become obsolete.
The obsolete hardware has been emulated by software so the original Domesday software can be executed on a modern computer platform. The reconstruction of the Domesday Project was such a huge operation that it can't be a blueprint for all inaccessible digital data.

So, if the problems involved in digital preservation are not recognised on time, digitised material will no longer be accessible in the near future. Of course it is possible to digitise the originals again in the future if the originals still exist. However, the high costs of scanning, organising digitisation projects and the growing need to reuse TIFF files for different purposes makes this option undesirable. Also, some institutions are inclined to digitise material not only for accessibility on the internet but for preservation of the original.
The Dutch government wants to be ensured that after digitisation of a collection, the digital objects will be reusable in the future. It fully supports the development of standards for long-term preservation of digital objects.

Pilot project
In 2004 the National Library of the Netherlands initiated a pilot project that explored the practicability of a national digital repository at the KB itself. Why organise a repository for digitised cultural heritage at the KB? The KB already makes use of a digital archive for long-term preservation, the e-Depot. It is used for the safe storage of electronic publications. To support this effort a digital preservation department at the KB is dedicated to defining policy on safe storage and permanent access.

When a national repository will be organised, the Dutch cultural heritage institutions do not have to reinvent the wheel themselves and have each institution separately bear the costs of a digital archive. I will tell you more about the digital archive and the digital preservation research at the National Library later on in my presentation.

We asked five cultural heritage institutions to cooperate with the project team on defining functional requirements and help with testing the pilot system during the project. The institutions that participated were quite divers: a museum, an archive, a university library, a documentation centre and the collection department of the KB.

Project objectives

Business plan
The project's two main objectives were the development of a pilot system and the writing of a business plan to asses the feasibility of a national service for safe storage.
The business model used in the plan is based on a "not for profit" model. The revenues, however, do cover the costs of the service. Among other things, the plan contains a cost and revenue calculation of a national service for long-term preservation, a definition of the target consumer groups, market research, user needs and risks.

Pilot system
The other objective of the project was the development of a pilot system. This pilot system explored the feasibility of the technical implementation of a national digital repository based on the e-Depot.

In cooperation with the five institutions participating in the project functional requirements for a TIFF archive were formulated. The TIFF archive should enable safe storage of and permanent access to deposited digital collections for cultural institutions. It is not designed to be an image database that can be used by anyone looking for a high quality digital image.

One part of the development of a system was to establish which metadata are required for long-term preservation of TIFF files. During the pilot project the focus has been on TIFF to keep the scope of the project workable. However, the system and the data model can easily be enhanced to accept other file formats such as JPEG2000.

A few institutions mentioned that they would not maintain digital copies of their collections themselves if they had deposited their collections to the TIFF archive. However, the institutions would

have to be able to access their deposited collections on a daily basis, because the master files are used in their services to customers.

In summary the requirements of a national digital repository amount to:

- A function for depositing collections and corresponding metadata
- A way of searching in deposited collections
- A request function for specific TIFF files or a complete collection of TIFF files
- a delivery function for the requested files
- A security system that should prevent the unwarranted use of collections by other institutions or third parties.
- The service should be available on a daily basis and independent of the location of the institutions.
- Development of a data model that facilitates long-term preservation and permanent access to digital images.

As I have mentioned before, the National Library makes use of an operational long-term digital archive called the e-Depot. As the e-Depot was the starting point for the TIFF archive pilot system, I will discuss the e-Depot and digital preservation at the KB first before I proceed with the implementation of the pilot system.

Digital preservation at the KB

e-Depot

In its capacity as the Dutch deposit library the KB is responsible for all publications appearing in the Netherlands. As the number of electronic publications is rising, the KB has embraced the important task of preserving and providing long-term access to Dutch electronic publications. It has developed a digital information archiving system, called the e-Depot. The e-Depot was originally intended for archiving publications by Dutch publishers, but the scope has been broadened to include international publications as well in 2002. The deposited publications consist mainly of electronic articles in the Portable Document Format. The e-Depot articles can be accessed by downloading one article per request in the reading room of the KB.

At the heart of the e-Depot system is a technical component called DIAS (Digital Information and Archiving System), developed by IBM and based on the OAIS (Open Archival Information System) Reference Model. This Reference Model establishes a common framework of terms and concepts which comprise an Open Archival Information System (OAIS). It is also a technical recommendation for use in developing a broader consensus on what is required for an archive to provide permanent, or indefinite long-term, preservation of digital information. The e-Depot system at the KB is one of the first operational digital archives based on the OAIS, worldwide.

Digital preservation research

In addition, the KB's research initiatives are geared towards ensuring long-term access. Research is being done on the digital preservation strategies emulation and migration. Emulation aims to render the digital objects in their original form and to preserve all functionality of the objects. The migration strategy is focussed on converting the digital objects to provide access to them in a future computer environment.

The digital preservation department is looking into properties of file formats, preservation metadata and the development of a module called the Preservation Manager. The Preservation Manager contains information about how to render digital objects on the file format level. The Preservation Manager will provide the information necessary to be able to plan actions before a specific file type becomes inaccessible.

Pilot system revisited

Pilot system
Now I would like to turn back to the pilot system for the TIFF archive project. The technical challenge we faced was the extension of the existing e-Depot system and infrastructure.
The functional requirements have resulted in the following developments:

- Depositing digital collections by institutions is handled similarly to the way publishers deposit their journals: the collections and the corresponding metadata can be transferred by an institution to a secure FTP server at the KB. The processing that is needed before the collections can be ingested into the e-Depot is also similar to the workflow of publications.

- A website has been developed that offers the institutions search and request facilities. The website also offers functionality for selection of specific files or complete collections. The website has been developed based on standard KB technology and fits in with the existing infrastructure at the KB.

- For delivery of deposited collections to the institutions a batch delivery component has been added to the e-Depot system. Until the development of the batch delivery component only single articles could be requested. After requesting the files via the website, the requested files are looked up in the e-Depot system and a copy of the stored files is then placed on a secure FTP server. An email will notify the institution when all requested files have been placed on this server. The institutions have access to the FTP server to transfer their requested files back to their organisation.

- The authentication and authorisation component already used by the e-Depot has been extended to enable controlled access to files from outside the KB. The website for searching and requesting digital collections can be accessed after login. The authorization level of a collection can be assigned flexibly according to an institution's needs.

At the end of the project the functionality of the pilot system was tested by the participating institutions.

Data model
Another part of the project was to explore which metadata are required for long-term preservation of TIFF files. Long-term preservation of digitised images not only requires careful registration of what is called “preservation metadata” but also additional information on the digitisation process.
We took the NISO Z39.87 standard Data Dictionary for Digital Still Images as a starting point for examination of the required technical metadata. The first PREMIS draft has also been used to compose the data model. PREMIS is a data dictionary of core elements to support digital preservation.
Our aim was to compose a practical and extendible model, based on the standards I’ve mentioned before. The KB will develop the data model further when the TIFF archive will be taken into production.

Conclusions

We can conclude from the enquiry held among cultural heritage institutions that there is a market for a national digital repository. Another point of interest emerged from the enquiry: digital preservation is fairly unknown to most cultural heritage institutions.

The feasibility of a national digital repository is an important result of the project. Another insight that can be inferred from the project is the urgency for raising awareness of digital preservation and the pitfalls of a digital age among cultural heritage institutions.

The test results from the project show that the pilot system can meet all requirements for a central repository for digitised cultural heritage. Further action on the TIFF archive will be a project to develop the pilot into an operational service. This implementation project will start at the end of 2006.

I hope you have found the discussion of the pilot project for a TIFF archive at the KB a clarifying and perhaps even an inspiring experience.