National Digital Repository for Digital Still Images in the Netherlands

Astrid Verheusen, Caroline van Wijk; National library of the Netherlands; The Hague, The Netherlands

Abstract

This paper addresses the outcomes of a project executed by the National Library of the Netherlands to explore the possibilities of a National Digital Repository for long-term preservation and permanent access to digital still images of cultural heritage institutions. The project was carried out in 2004-2005.

Digitisation activities in the Netherlands

During the last ten years Dutch cultural heritage institutions have digitised large parts of their collections with the main aim to increase accessibility to cultural heritage through availability on the Internet. An enquiry held in 2005 shows that cultural heritage institutions in the Netherlands together have produced approximately 5 million TIFF files. Because the majority of these institutions have plans for digitisation in the near future, the amount of TIFF files is expected to increase to more than ten million over the next four years.

Due to this progress in the field of digitisation a profound knowledge about best practices has been developed. There are, however, growing concerns about the storage and permanent access of the products of digitisation. In most cases high quality TIFF files serve as master files. The majority of the cultural heritage institutions do not have well-formulated policy on archiving these master files and are not aware of the problems of digital preservation. Often the focus lies on producing a website to make the images available. If action is not taken, digitised material will no longer be accessible in the near future.

Although it is possible to digitise originals again in the future, the high costs of scanning and the growing need to reuse the TIFFs for different purposes make this option undesirable. Furthermore, institutions are starting to show a tendency towards digitising material for the purpose of conservation of the original. Therefore, the Dutch government is now stressing the need to ensure that the expensive digitised material can be used again when needed in the future and pushes towards establishing standards to safeguard long-term preservation of the digital objects.

Consequently, the National Library of the Netherlands initiated a project to investigate the possibilities of long-term storage of TIFF files in a so-called 'TIFF archive'. The project had two main objectives: the development of a pilot system for the storage of TIFF files and the formulation of a business plan to assess the feasibility of a national service for storage.

Functional requirements

Functional requirements for a pilot system were composed in cooperation with five cultural heritage institutions. Among the participating institutions were several libraries, a museum, an archive and a documentation centre. Each institution provided one or two digital collections for testing the functionality of the system. The diversity of the institutions has brought about a system that is widely usable for long-term preservation of and permanent access to digital still images.

Analyses reveals that the main functional requirements for a National Digital Repository for digital still images are exclusive and secure access for cultural heritage institutions to stored collections, on demand access to these collections and on demand delivery of collections to the institutions. The desired level of authorisation for access to the collections varies between the institutions. One institution is adamant about exclusive rights for searching and retrieving it's own material, while other institutions are more inclined to make their collections searchable by all TIFF archive members. Some institutions are considering exclusive storage of their digital collections in the National Digital Repository at the National Library without retaining private copies. Because their daily activities imply frequent use of the digital collections, on demand access to and subsequent delivery of these collections is of vital importance. As the original point of departure for the system was the long-term preservation of and future access to digital objects, the requirements based on the workflow of the institutions were challenges for the development of the TIFF archive.

e-Depot

Starting point for the technical implementation of the National Digital Repository is the digital archive called "e-Depot", which has been operational in the National Library of the Netherlands since 2002. The e-Depot system is composed of DIAS (Digital Information and Archiving System), developed by IBM and based on the OAIS (Open Archival Information System) Reference Model [1], with the addition of components specific to the National Library. The e-Depot has been developed to sustain the important task of long-term archiving and permanent access to Dutch electronic publications [2]. Archiving agreements have been signed with national and international publishers since 2002. Deposited publications consist mainly of electronic articles in the PDF format. The e-Depot articles can be accessed by downloading one article per request in the reading room of the National Library.

To secure the possibility of rendering the stored publications in future times the National Library and IBM jointly developed the Preservation Manager component. The Preservation Manager facilitates managing the information of the line-up of software and hardware environments which are needed for rendering the publications stored in DIAS [3]. Current research projects focus on digital preservation strategies such as emulation and migration. The DIAS system will be extended with new functionality to support preservation strategies.

Pilot system

The technical challenge of the pilot project was to extend the existing system and infrastructure. The aforementioned functional requirements of the institutions have resulted in the addition of an authentication and authorization component and a batch delivery component to the e-Depot system. Deposited digital collections can be searched for and requested by a web interface after login. The authorization level of a collection is assigned flexibily according to requirements. Complete collections can be retrieved by one request. The collections are delivered by use of secure file transfer protocol.

The pilot system has been developed in the test environment of the e-Depot system. The participating institutions have tested the pilot system functionally during the project. The test results show that the functionality of the e-Depot system can successfully be extended to meet the requirements for a National Digital Repository for digital still images.

Preservation metadata

Aside from the addition of extra functionalities to the e-Depot system, another main objective of the project was to establish which metadata are required for long-term preservation of TIFF files. The NISO Z39.87 standard Data Dictionary for Digital Still Images [4] was taken as a starting point for examination of the required technical metadata. Metadata extraction tests were performed on several TIFF files using the extraction tool Jhove [5]. Comparison of the Z39.87 elements with the Jhove output showed that a large amount of required information was not provided by metadata extraction. Specifically the elements describing the relation between the analogue original and the digitised version could not be extracted. This information could not be provided by the institutions either. The relation between the original and the digitised version is vital to the preservation of digital masters. Next to Z39.87, the first PREMIS draft was also used to compose a data model for the National Digital Repository. PREMIS is a data dictionary of core elements needed to support digital preservation [6]. When operationalisation of the pilot system commences the TIFF archive data model will be further developed.

Business plan

The TIFF archive is designed according to a “not for profit” business model. In this respect the project is innovative in the world of the cultural heritage institutions: it is the first time a business plan was written for a repository for the long-term storage of TIFF files. The business plan focuses on target customer groups, user needs, risks (technical and other), and costs and revenues of the service.

Market

In March 2005 an enquiry was held among 300 archives, libraries, museums and documentation centres to establish if there is a market for a national digital repository, as well as to ascertain the demands of potential customers. Approximately 100 institutions responded. The results did not show any significant differences in response between the various types of institutions.

From the 100 responding institutions 74 percent is in possession of TIFF files. 30 Percent holds more than 10.000 TIFF files. CD-Rom, DVD and hard disks are the most frequently used storage mediums. 57 Percent of the institutions do not have a policy towards save storage of the TIFF files, whereas 43 percent do. Almost all institutions have plans towards digitising more material in the near future.

The institutions were asked if they would make use of a central repository for digitised material. 42 Percent of the institutions gave an affirmative response; 38 percent wavered. When the institutions were asked why they would not make use of a central TIFF archive, they mentioned several reasons. Some institutions did not see the need for a central repository because they are of the opinion they can manage the storage of TIFF files themselves. Furthermore, they use the files often and doubt whether a repository will be able to deliver the files quickly. Another reason is that they do not outsource the storage of their paper files and do not see why they need to outsource the storage of their electronic files. Institutions that were not convinced of the need for a central repository had many questions concerning the service. They wanted to know how fast the performance would be, had questions about conditions and had doubts about copyright, authenticity, accessibility, the National Library as deliverer of the service and most of all about the costs of the service.

The response indicates that many of the institutions are not always aware of the issues concerning safe storage and permanent access. Sometimes institutions confused simple storage with long-term storage and digital preservation. The institutions that do see the use of a central repository are very demanding towards the performance of the service. Lastly, nearly all institutions have concerns about the costs. However, an astonishing 62 percent of the institutions that did see the need for a central service were aware of the high costs and were willing to pay.

Costs and revenues

It is difficult to find a standard model that covers all relevant expenses of digital storage and preservation, in order to calculate the costs. The model developed for the business plan is based on the costs that were made for the storage of two million articles in the e-Depot. Personnel, hardware and maintenance costs of storage in the e-Depot were taken into account. Costs for the development of strategies for preservation were not included. The outcome shows that safe storage is costly. The average storage costs of a single file amount to approximately € 0.20 per year. Costs decrease as the number of stored files increases. The business model drawn up for the national repository charges institutions with a one-off sum for the deposit of a collection and a yearly amount for maintenance. The height of the sum depends on the size of the collection. Although the model is not for profit, it does ensure the revenues cover the costs of the service.

Conclusion

The results of the pilot project confirm that there is a market for a national repository for digitised materials and - although storage is relatively expensive - costs and revenues of the service can be balanced. Furthermore, the project has demonstrated that occurring technological problems can be resolved. On the basis of these results, the National Library of the Netherlands has decided to further develop the service. In 2007 the national digital repository is expected to be operational.

References

[1] "Reference Model for an Open Archival Information System (OAIS)" Blue Book Issue 1 (January 2002).

[2] J.F. Steenbakkers, "Digital Archiving in the Twenty-First Century: Practice at the National Library of the Netherlands", Library Trends Vol. 54 (1) (Summer 2005).
E. Oltmans & H. van Wijngaarden., "Digital Preservation in Practice: The e-Depot at the Koninklijke Bibliotheek", VINE - The Journal of Information and Knowledge Management Systems Vol. 34 (1), 21-26 (2004).

[3] E. Oltmans & R.J. van Diessen & H. van Wijngaarden, "Preservation Functionality in a Digital Archive", Procedings of the Joint Conference on Digital Libraries, Tucson, Arizona, USA (June 7-11 2004).

[4] NISO, National Information Standards Organization, Z39.87 “Data Dictionary - Technical Metadata for Digital Still Images”, proposed draft standard (2005).

[5] JHOVE, JSTOR/Harvard Object Validation Environment. Information available at: http://hul.harvard.edu/jhove/index.html.

[6] PREMIS, PREservation Metadata: Implementation Strategies Working Group, Datadictionary for Preservation Metadatata (May 2005).

Authors Biography

Astrid Verheusen completed her MA in history at the University of Leiden in 1992. In 1993 she followed a postgraduate programme on history and computing at the University of Leiden. From 1994 tot 1999 she worked at the Institute of Netherlands History in The Hague as an information specialist. In 2000 she worked at Sdu Editors as a content-analyst. Currently she is employed at the National Library of the Netherlands as a projectmanager at the Research & Development and is involved in projects concerning digitisation and digital preservation.

Caroline van Wijk received her degree in Arts from the Gerrit Rietveld Academie in 1998 before turning to software developing. In 2000 she took a course in Java programming and consequently worked as a software engineer in internet applications development for several years. Since 2004 she is employed at the National Library of the Netherlands and assists the National Digital Repository project..