



IBM and the Digital Library

IBM has a long-standing record for facilitating the storage and retrieval of high volumes of data. With respect to the growing interest for and use of digitised documents and pictures in the last decade, IBM introduced the Digital Library technology. Among the first implementations in the cultural heritage environment, joint projects have been realised with the Vatican and the State Hermitage Museum in St. Petersburg. As a result, a substantial part of their collections of paper-based documents, pictures and objects can now be accessed electronically.

As the National Library of the Netherlands (KB, Koninklijke Bibliotheek) was faced with the problem of preserving large amounts of digital documents for the long term, a collaboration between KB and IBM developed and in September 2000 the KB and IBM Netherlands agreed on a project to design and implement a "Depot for Netherlands Electronic Publications" (DNEP). A Study to the requirements for Long Term Preservation was an integral part of the complete effort.

Both the KB and IBM were committed to develop a general solution for the storage and long term preservation of high volume electronic information. The documents originate from two sources: from digitisation of paper documents and from new media types published directly in digital form.

In essence an electronic publication consists of three components: the bit stream, the logical format in the bit stream, and the functionality needed to decode this logical format. Even complicated publications break down into these three components. In order to achieve preservation of and access to electronic publications for the long term, three steps are necessary:

1) Archiving the Electronic Publications

Archiving means to assign an identifier, to file the electronic publication in a controlled environment, and to produce descriptive and technical metadata.

2) Maintain the bit stream

The second step is preserving the digital object by maintaining the bit stream complete and in its original structure. As deposit and national libraries act as a last resort, this should be done in a pro-active way.

3) Guarantee long term access

With a printed publication this is not an issue, because the information can be readily read without any specific tool or functionality. This step is unique for digital preservation. Without taking this into account we may archive and maintain an electronic publication, but we

cannot guarantee access to and correct interpretation of the stored information for the long-term.

Usage of standard components (hardware and software) and compliance with the Open Archival Information System (OAIS), the OSI standard for digital archives, should achieve general applicability of the developed DNEP solution.

IBM provided and installed a complete solution including hardware, software and services to establish a comprehensive digital content solution for storing and protecting millions of published documents for the future. The business solution is based on IBM Digital Library, now called Content Manager. IBM Content Manager stores the documents in Object Servers, and, for retrieving purposes, the related metadata in a Library Server. The system has been designed to handle over 500 TB of data. Electronic documents are put into the system through a custom-made input/indexing module. DNEP is completely web enabled which means all user interaction is taking place through a web browser.

This part of the project, developed in close collaboration with the IBM Content Manager Development Laboratory in Boeblingen, Germany, and Santa Teresa, California, has just been completed. It is the first realisation of such an actively maintained amount of electronic multimedia data in the world. The solution will be enriched by IBM with new functionalities, resulting in a new version called DIAS (Digital Information Archiving System). It will commercially be made available for other institutions as well.

Concurrent with the implementation of DNEP, KB and IBM started a Long Term Preservation study to extend the knowledge on long term preservation techniques. The study's main objective is to investigate the functionality needed to preserve the digital information in DIAS for the long-term e.g. for hundreds of years. This Long Term Preservation topic is a line of research within the IBM Research Division and is the main challenge for this KB/IBM project.

The first part of the study has been finished and has resulted in the specification of the Preservation subsystem functionality which is scheduled to be included in the next DIAS release and of which a proof of concept successfully has been tested at the KB premises. KB and IBM Research will jointly continue the research and development activities in collaboration with other interested parties.

For further information please contact:

IBM The Netherlands
Etta Pouw
Communications
Tel: +31 20 513 4228
E-mail: etta_pouw@nl.ibm.com