Economic and Design Issues of Large-Scale Multimedia Databases

Peter Olaf Looms
DRIVE
The Danish Broadcasting System
Islands Brygge 81
DK-2300 Copenhagen S, Denmark
telephone: +45 31 67 12 33 extension 6884/6885
telefax: +45 31 95 31 10

Abstract

The aim of this paper is to examine key economic and design issues facing three public institutions now working with multimedia - in particular image databases - and to highlight critical features which need attention while evaluating the feasibility of such a system.

1. Introduction

Museums have a two-fold purpose: to preserve aspects of our cultural heritage for posterity and to make this heritage available to society as a whole.

The transition from a nomadic way of life, or one based on agriculture and trade, to a society based on industry and services makes new demands on those responsible for providing this cultural memory. Our cultural heritage now encompasses not just artifacts such as archaeological remains, tools, works of art and books, but also sounds and images. The twentieth century has brought with it electronic means of recording, distributing and displaying the whole gamut of cultural expression. Electronic media also have a dual role as examples of cultural expression and as the means of disseminating culture in its widest sense. In this paper I should like to focus on the use of multimedia for the preservation of cultural heritage.

Large-scale multimedia databases in museums have been in regular use for more than ten years. Although the number of museums using them is in the hundreds, they are by no means a majority, as can be seen from surveys from North America and Europe.

The term " multimedia" is used here to describe any system where information in the form of text, figures, images, sounds or moving images is displayed on a screen. In some cases this will be a computer screen, in others a television monitor, where the use of computer technology is hidden from the user.

Unlike films or videotapes, which are linear in their structure, multimedia are interactive in the sense that the user can actively make his own way through the material. Each path is unique.

Multimedia, using combinations of text, sound and graphics, are relatively undemanding from a technical point of view, unlike systems containing large numbers of images. For this reason I would like to use three case studies involving images to examine some of the underlying economic and design issues for large-scale multimedia databases.

2. Multimedia Databases - the National Museum of Denmark

The National Museum of Denmark has been active in the use of computer-based image and text databases for the last five years. The Museum itself is organised into eight departments and the collections comprise approximately 1.5 million artefacts.

The basis for this work dates back to 1987 when a private organisation, the Egmont H. Petersen Foundation, provided a grant of some 30 million dollars to the National Museum, on the understanding that this sum would be matched by a grant from the Ministry of Culture, giving a total extraordinary budget of 60 million dollars. The aim was to transform the museum into a modern centre for the preservation and dissemination of knowledge about our cultural heritage. Nearly a fifth of the budget until 1994 was earmarked for a documentation project to automate collection management and to introduce the use of new information technology in exhibitions and services to the public. From 1994 on, the documentation section will be paid for out of existing museum funds.

2.1 Text

Prior to the beginning of the project, collection records were based on hand-written ledgers or file cards. The documentation department had to set up a database system allowing for the decentralised and retrospective registration of objects from all eight departments. The major problems to be overcome were:

- the lack of homogeneity of the paper-based records produced over a two hundred-year period
- whether to continue with the original classification schemes and transfer the content of paper-based records or to limit records initially to a basic register derived from a knowledge of the objects themselves
- the extent to which non-specialist manpower could be used during data input to keep down costs
- the number of data fields to be used for a given object

- ensuring consistent data inputs from different curators.

It is interesting to note that these problems are conceptual and intellectual rather than technical. Carsten U. Larsen, the head of the Documentation Department, reports that design issues, the need for in-service training, and manpower costs for retrospective text entry are by far the most significant challenges to be faced. The other key point is that of costs: given that there are 1.5 million objects at the Museum, the time required for text input is considerable. To date, 385,000 objects have been processed. Figures for the Ethnographical Department indicate that more than 4 minutes were required just to transfer information in its original form to the database. In some cases, several hours' work were needed to correct obvious errors or fill gaps in the records. Using the figure of 4 minutes, and with just over 110,000 artifacts to cover, a minimum of 5 man-years would be required. Text input for the whole collection would require somewhere between 75 and 220 man-years.

2.2 Images

To complement the automation of records, The National Museum also works with images. As the entire west wing of the museum had to be vacated in order to modernize the building, the collections of the Ethnographical Department were put into storage until this work was completed. Before this took place, a local area network was installed, and the department was equipped with computers. Each object was examined to assess its state of repair. Additional information about the object was recorded, after which it was photographed, identified with a label and bar code, packed and put into storage.

Eight teams produced a total of 110,000 35 mm colour slides during this period. Each slide contained a machine-readable identification code paired with the database collection number. The unmounted slides were spliced together in rolls of approximately 50, and transferred to 1"C videotape by DRIVE using a specially developed optical block. While the image (measuring 24 x 32 mm) was transferred, the numerical code in the remaining 4 mm of the frame was fed via a video camera to an optical character recognition unit. The code was paired with the time code of the tape and the information recorded on a computer. LaserDisc videodiscs were produced from the 1"C master tape, and the museum was provided with the address of each object (both the time code on the tape and the frame number on the video disc). This enabled the museum to keep track of each of the objects and to arrange for links between the database and a videodisc player connected to a given terminal. This process has been described in detail elsewhere. Film was chosen as the production medium rather than going directly to videotape or a recordable analogue or digital disc for several reasons:

- film offers extremely high resolution, nine times that of a television image
- the equipment required is easy and cheap to use
- film is a well-known and standardized storage medium allowing for automated transfer to some new medium within the next ten to twenty years as and when required.

As the museum works in the long term, this indirect route via film offers a high degree of security and flexibility for a slightly higher initial cost.

The major problems experienced in the production of the image database concerned quality control and optical character recognition. It is critical that the developed film be checked for adequate exposure and blemishes on a daily or weekly basis. With time, photo diodes can vary in output, or back plates containing this component can be dislodged, slowing the rate at which image transfer can take place from film to master medium.

To reduce such sources of error, DRIVE is producing a similar back plate to produce miniature bar codes on each frame instead of machine readable numbers. The procedure follows European Broadcasting Union guidelines for furnishing film material with time codes. To increase productivity and therefore reduce costs, the 1"C tape system has been replaced by a high quality component recordable video disc system from Sony. Videodisc masters are produced by editing onto Betacam SP (or a digital component format in the near future).

2.3 Multimedia and the public

To date, the records of more than 300,000 artefacts have been computerised and images from two departments - Ethnography and the Resistance Museum - transferred to videodisc. Work is in hand to make the image databases available not only to curators, but also to subject specialists and the general public. Subsets of the material are being used in interactive exhibition systems, the first of which was "Danish Resistance and the German Occupation of Denmark 1940-45." This contains 11,000 photos, posters, and objects and nine video sequences in Danish and English. Although originally conceived for use at the museum itself, the videodisc is now available commercially for use by schools and research workers in Denmark and abroad.

Following the Danish Resistance disc, further systems are being developed: "The Viking Age" (a collaborative effort with the York Archaeological Trust in England to be completed in 1992), an adaptation of "Siulleq", a multimedia database about the culture and people of Greenland to which the National Museums of Denmark and Greenland were significant contributors. A disc covering Mediaeval Church Murals, a prototype of which has been produced by DRIVE and the University of Copenhagen, may also be used when finished.

3. The Proposed National Picture Database at the Royal Library

A second case is that of the Royal Library in Copenhagen, which is the Danish national library. In addition to its collection of books, the Royal Library is responsible for managing some ten million pictures. The lion's share consists of photographs, but there are also significant collections of engravings and water colours.

Preserving such a collection for posterity is no mean feat. The photographic collections contain material from the dawn of photography to the present day, and the conservation of

plates and emulsions from the 19th century is a daunting task. Significant parts of the collections are not readily accessible as they were received as donations, often unsorted and without adequate indices.

The picture collections are widely used by historians, local historians and genealogists in addition to publishers, museums, and teachers in Denmark and abroad. Indicative of its use is the fact that, with few exceptions, pictures from the Royal Library are to be found in all illustrated historical publications about Denmark.

The collections also contain sizable quantities of photographs from other parts of the world, including, for example, Denmark's former colonies in the Danish West Indies, and reflect the pattern of contacts between Denmark and the outside world over the last century. The current situation is far from satisfactory and according to the Library management there are three main barriers to be overcome: geographical distribution, procedural matters, and the pictures themselves.

3.1 Geographical distribution

Most of the pictures are unique and can only be seen at the main building near the Danish Parliament. Due to lack of space - common to many other museums and archives - most of the pictures are in storage elsewhere.

3.2 Procedural matters

To find materials in the collection, users have to come in person to the Library during the opening hours of the picture collections, and a picture specialist from the Library is required to conduct a search. The user is present while this takes place, and this makes additional demands of security systems. Approximately 60,000 enquiries regarding photographs, pictures and maps are dealt with each year.

3.3 The pictures themselves

The pictures themselves constitute one of the problems: a search is conducted collection by collection, by and large by thumbing through originals themselves, each of which is kept in a brown envelope with handwritten details on the outside. For many of the collections there is no catalogue or index. A significant proportion of the images are not readily retrievable as they are in storage or are inadequately indexed.

3.4 A National Picture Database - a preliminary study

Following discussions between the Royal Library and DRIVE in connection with the SIULLEQ project about Greenland to which the library also contributed materials, a working party was set up in August 1989 to look into the feasibility of setting up a National Picture Database basedon the collections at the Library. The working party consisted of specialists from the Royal Library DRIVE and other departments of the Danish Broadcasting Corporation, and the head of the Documentation Department of the National Museum

as an associate member. The final report of the party dated June, 1990 reviewed the conservation and dissemination issues, provided a brief survey of comparable large-scale image databases and a proposal for a three-year preliminary phase to evaluate and decide on long-term conservation and dissemination strategies.

The report indicated the need to evaluate procedures for the organisation - recording, storage, registration and use of images - on the basis of which the specification for the final database could be based, by completing a cycle from design to production to multimedia products for use in the library and elsewhere, using a thematic selection of the existing collections. The working party indicated that setting up a National Picture Database would be extremely manpower-intensive, and would require funding in addition to that currently provided. On the other hand, the very existence of an electronic catalogue in place of searches using original material would have radical organizational implications. Manpower tied to monitoring visitors and conducting routine searches could be released to increase capacity. Wear and tear of unique originals would be reduced to a minimum, and searches would no longer be limited to the library itself. In line with experience at the National Museum in Denmark, the conservation laboratory of the Norwegian National Library in Rana and the Malm Museums' visual index project, the party recommended evaluating 35 mm still film, 35 mm and 70 mm motion picture film for working masters for images without negatives, supplemented by a high resolution digital format for aging glass plate negatives. The catalogue medium could be either an analogue or digital optical disc, and the choice was to be based on what was most widely accepted in the market place.

The total budget for this phase (work with 115,000 images) was approximately 18 million kroner (nearly 3 million dollars). 81% of the budget covers man-power, and 69% is manpower for database design, text entry and the handling and copying and storage of original materials. Less than 19% of the total figure is devoted to production facilities, and this proportion decreases as a function of the number of images. The preliminary phase of the project is expected to start within the next six months when the necessary funding as been found.

4. The Picture Archives at TELECOM Denmark

TELECOM A/S is part of the Danish Posts and Telecommunications that was privatised at the beginning of 1991. The telecommunications industry has a lengthy history in Denmark, and until last year the only systematic picture library was run by the information department. Photographs and films produced elsewhere in the organisation were produced and used on an ad hoc basis, as is the case in most organisations. TELECOM also has a museums group concerned with telecommunications past and present.

To come to terms with the dual tasks of preservation and dissemination, TELECOM set up a picture archive with the aid of an outside specialist. To date work the picture archive has:

- carried out a preliminary survey of the number and range of pictures, films, and video recordings within the organisation
- drawn up practical guidelines for the selection, use, and storage of visual materials by TELECOM staff
- raised awareness about the need to preserve material of historical interest about Danish telecommunications, and
- commissioned the pilot version of an electronic catalogue based on approximately 5% of the pictures known to exist and which meet minimum preservation requirements.

The purpose of the pilot is to evaluate and cost the procedures to be used for the full-scale project. The biggest uncertainty at present is the cost of establishing the provenance of much of the material and finding people with the necessary background knowledge to be able to describe the content of historical photographs. One idea being discussed is to invite retired employees to local meetings around the country to assist the archive in its detective work. The same idea has been used by our own organisation which occasionally prints historical photographs in the staff magazine which is also circulated to former employees.

The text database for the prototype is currently being implemented and the transfer of the first photographs will have taken place by the time this paper is published. For the prototype, the sample will be sorted by subject, size and format and recorded directly onto a recordable video disc by DRIVE. This allows material to be transferred more rapidly than would otherwise be the case and facilitates the addition of electronically-generated masks for pictures which do not have the 4:3 format of a television screen.

5. Critical Features of Large-Scale Multimedia Databases Needs

The introduction of large-scale multimedia databases often hinges on comprehensive modernization plans of which the computerisation of indices and registers forms a part. In other cases the focus is on conservation issues, and specifically for photographs, the need to distinguish between finding a picture and using it. Improved decentralised access, accompanied by organisational changes at the museum are also commonly-quoted arguments.

Whilst it is true that images on a screen can be no substitute for the real thing, this is not point. At any given moment in time, a small fraction of a museum's artifacts will be accessible: the remainder is in storage.

Image databases come into there own prior to seeing the original. A scholar, curator or a member of the general public has the opportunity to examine and compare in some detail individual pieces not only in the same room but in different collections. This prepares the way for a more efficient use of the originals themselves.

5.1 Finance

Instances of large scale databases being financed out of existing budgets are rare. The National Museum modernization program and the Royal Library project both require funding from extraordinary sources, and the 9 million dollar documentation project which started at the University of Oslo, Norway in February 1991 can only be carried out because subsidised manpower - unemployed staff hired to input text - is available.

5.2 The nature and scale of the task

At a recent conference, a spokesman from Kodak mentioned in passing that approximately 60 billion photographs are taken each year. As is the case with other media, it is clearly impossible to preserve everything for posterity: selection criteria are imperative.

Given the size of existing picture collections such as the Royal Library, hundreds of man-years required to establish an image database. Labour costs as a proportion of total costs increase as a function of the number of items. Decision-makers will therefore have to make trade-offs, being the number of images to be included in the system, the nature and amount of object registration. The critical aspect of object registration is to establish the minimum level of computerisation needed to replace manual records.

Experience from the National Museum shows that it is perfectly feasible to introduce object registration in stages, not only in the museum as a whole but also within the individual collections.

5.3 Conservation issues

Museums are entrusted with the task of preserving aspects of our cultural heritage for posterity - a collective memory. Yet all cultural artifacts deteriorate with time, some faster than others. There is therefore a premium on storage media which will still be intact and accessible well into the next century. As regards images, we are at a critical point in the transition from analogue to digital storage. As far as alphanumerical data is concerned, the transition is complete, whereas this is not yet so for sound and images. Images represent the biggest difficulty, as there are no universally accepted digital formats which offer the same combination of resolution, contrast, cheapness and longevity as film. It is likely therefore that film will continue to be used as an intermediate high-resolution format for the rest of this decade while the battle between analogue and digital storage in the outside world has been settled. If the multimedia database is required for an electronic catalogue or exhibition purposes only, the material can be transferred in automated systems to the delivery system currently in use.

5.4 Computerisation of text inputs associated with images

There is a wide range of taxonomies and classification schemata in use at museums around the world. While data covering the provenance and the physical aspects of an image

do not represent insurmountable difficulties, the content is often context-specific, and the context evolves in line with our conceptions of the past. There is therefore a premium attached to minimum levels of computerisation to facilitate exchanges between museums. If this information is not already in machine-readable form, it will consitute one of the major expenses in any database project.

