Proceedings of the NACCQ 2000
Wellington NZ www.naccq.ac.nz

Taonga On Line: Managing and Preserving Culture in a Digital Age

Trish Evans[1]
Kevin Wilkinson[2]

Museum of New Zealand Te Papa Tongarewa
Wellington, New Zealand
trishe@tepapa.govt.nz[1]

Massey University, Wellington, New Zealand[2]

ABSTRACT

The pace of change of information technology is not in dispute. We are all familiar with the way electronic storage media become obsolete. The hardware and software combinations that use these storage devices mutate even faster and, of course, one is of no use without the other.

While we may normally view this ferment of invention in a positive light and willingly accept an ongoing development burden for the sake of additional speed and processing features, the disparity between the rate at which information technology changes and the time scale of a permanent national collection gives cause for alarm.

This paper examines the features of collections management systems, as illustrated by Te Kahui, the Collections Management System of the Museum of New Zealand, Te Papa Tongarewa, and shows why migration of such a system to a new platform will be more difficult than for typical commercial systems. It also argues that, when the collection includes digital items, such systems may need to call on the services of a new discipline of historical computing.

Keywords
Information technology; archives; taonga.

1. INTRODUCTION

The pace at which information technology changes is a concern for any institution that uses computers to record and manage its critical information. An institution charged with maintaining a national archive feels the tidal pull of this change especially strongly. Unlike typical commercial organisations, the information it collects does not become less important with the passage of time. On the contrary, records gathered in the distant past may well be more important than the most recent ones. Decisions concerning the format and content of a database have lasting implications. One does not have the option of ignoring old and inconvenient records on the ground that they are of diminishing significance. Neither can one wait for anomalies to disappear without intervention as transactions are consolidated into quarterly and yearly summaries. In archival systems, history is valued for its own sake.

In addition, there are features of a collections database that make it especially difficult to migrate to a new hardware or software platform. This combination of extreme longevity of information and complexity of processing requirements is well illustrated by Te Kahui, the Collections Management System of the Museum of New Zealand, Te Papa Tongarewa. In the following section, we examine the features of Te Kahui that evidence the migration difficulties of such systems.

NACCQ
FOSTERING COMPUTING EDUCATION IN NEW ZEALAND

The effects of information technology on the preservation of cultural treasures is a broader problem than modernising the system that manages the archive, however. We close by examining the strategies that have been proposed for coping with change when a collection includes items that are themselves in digital form. We argue that digital items that define or constitute taonga will eventually require the emergence of a new discipline and a new industry: that of historical computing services.

2. THE MIGRATION DIFFICULTIES OF A COLLECTIONS MANAGEMENT SYSTEM

2.1 The Mission of Te Papa Tongarewa

The Museum of New Zealand Te Papa Tongarewa is a forum for the nation to present, explore, and preserve the heritage of its cultures and knowledge of the natural environment in order to better understand and treasure the past, enrich the present and meet the challenges of the future.

Te Papa is responsible for the provision of access to the national collections through the acquisition, management, conservation and study of items relevant to the Museum's purpose. The national collections are at the heart of the Museum. Display, interpretation and storage for some collections are now at a world-class level.

2.2 Collections Management at Te Papa

To assist with the care and storage of the nation's collection Te Papa uses an electronic collections management system. The Maori name of the database, Te Kahui, means an assemblage or a collection. The system is used to facilitate the management of museum collections of all types, from historical artifacts and art works to natural environment specimens.

The object-oriented database was custom-built for Te Papa six years ago and uses the text retrieval software BRS Search as the backend search engine with a Windows user interface written in Smalltalk. Images attached to records are stored in and drawn from Hyperwave software. Figure 1 shows why migration to a new system will prove challenging to say the least.

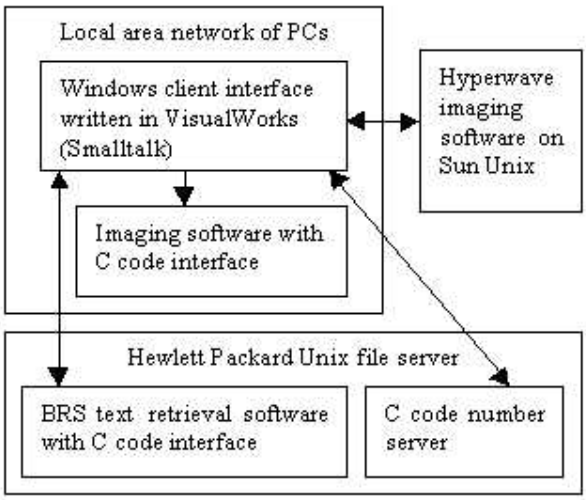

Figure 1. Te Kahui System Architecture.

2.3 System Modules

The database consists of various modules that handle aspects of the work of the Museum. These modules allow the storage and manipulation of information about items and the collection management procedures affecting them such as conservation, loans, movement, storage, insurance, and exhibitions.

The main module (called the Object File, menu item shown in Figure 2)

contains records about collection items and specimens and must contain descriptive data as a bare minimum but can also contain information about:

- acquisition of the item (when and where acquired, purchase price)
- whether an item has been deaccessioned (that is removed from a collection)
- restrictions on access to the item itself or its intellectual property
- insurance details - past and present
- where an item is stored permanently and temporarily, and what special storage requirements apply
- the movement of an item such as sending it out on loan or exhibiting it
- the condition of the item - past and present
- past and present treatment carried out on the item by a conservator
- whether an item has been or is on exhibition
- who catalogued and verified the data, along with who entered and updated the data on the database

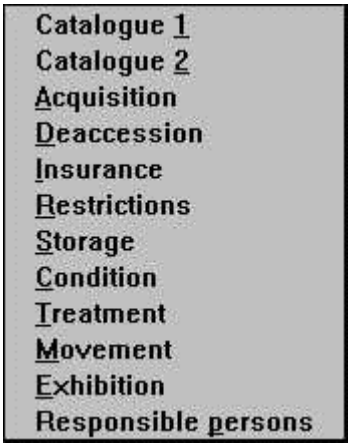

Figure 2. Te Kahui menu for Object File.

The modules are separate files but all link into each other (see figure 3.)

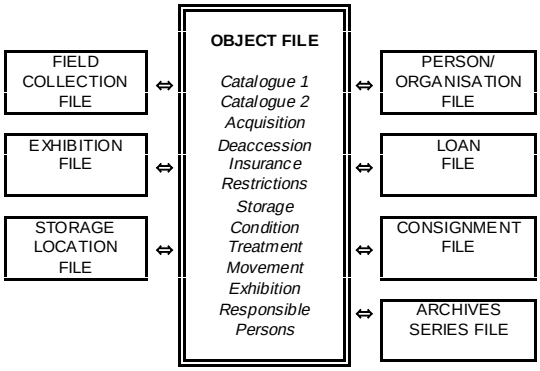

Figure 3. Te Kahui Files.

For example the module containing information relating to the collection of Natural Environment specimens in the field (called the Field Collection File) links into the Object File which provides information about the actual specimens collected. Similarly, the Person/Organisation File, which contains information on persons or organisations with whom the Museum has a relationship such as artists, donors and collectors, links into all files where there are name fields. These linked files reduce the number of fields users have to grapple with and mean certain data is entered once only.

2.4 System Volumes

Currently 460,000 items and specimens are entered onto the main module of the database with a further backlog of 2,000,000 items to be entered, largely from the Natural Environment collections. It will take up to twenty years to enter this backlog. At the same time new acquisitions or specimens collected on recent field trips are entered on a daily basis.

Other files contain the following number of records:

- Field Collection (stores information about where, how and when specimens are collected on field trips) = 89,000 records
- Storage location (stores the codes of areas within Te Papa down to draw and shelf level) = 11,300 records
- Person/Organisation (contains information about artists, donors, collectors, staff) = 5800 records
- Archives Series = (contains information about the Series of Archives held within Te Papa) = 1000 records
- Consignment (contains information about the movement of items within and in/out of Te Papa) = 670 records
- Loan (contains information about incoming and outgoing loans to/from Te Papa) = 460 records
- Exhibition (contains information about exhibitions within Te Papa) = 200 records

With this number of records one feature that has become indispensable is the global or multi-record update. This allows for large numbers of records to have data added, changed or replaced automatically rather than updating each individual record manually. This could be improved on immensely however by allowing more than one field at a time to be updated and to reduce the processing time for such updates. Presently such updates can never be completed within working hours due to the mammoth updating delays they cause! The database has been around for six years and standards have changed over time. Consequently, it is not unusual to have to change a piece of data for thousands of records.

As more and more items are entered processing power for searching and exporting data slows down too - the Museum now experiences limitations with report writing whereby only a % of very large reports will export to Excel.

The sheer volume of records and their associated histories can blow the limits of other applications too, for example the most commonly used way to supply data to customers is via Excel spreadsheets and on occasion Museum staff have wanted to export over 65,000 records to Excel, which is not possible with the current version.

The photography department currently has nearly 157,000 records entered, the Plants department over 71,000 records so the need for such large reports is not uncommon.

2.5 System Complexities

2.5.1 Disciplines

The database is a complex system that must cater for a number of unique disciplines at Te Papa. But while catering to the needs of individual disciplines it must also allow for the needs of more generic users from other departments. The use of the database by an Art Curator for example who uses it for research, cataloguing and verifying data is quite different to the use of it by someone working in the Enquiry Centre responding to public enquiries. It is complex also in that the two major areas the database supports - Humanities and Natural Environment - have quite different requirements for entry and retrieval of data. Think of the collection, storage and use of a fish caught at sea by local fishermen compared to the acquisition, storage and use of a Hotere painting. This centralised approach requires users to push different information into a standard list of attributes which can cause frustration, in fact some museums opt for separate databases to support these diverse disciplines.

The database therefore is designed to present tailored catalogue screens for different disciplines within the Museum, while maintaining a considerable degree of standardisation in terminology.

2.5.2 Security

Collections management databases for museums have unique security issues. Access rights to the database within Te Papa vary hugely from users on the exhibition floor with read only access to limited files and fields, through to users who have rights to restrict all other users including the system administrator from viewing records.

The design allows each of the fifteen disciplines to select from their own records but it is also possible to search the entire database using a separate login.

There are a number of other features of the database which assist with the management of collection items within Te Papa which show the unique requirements of such a database.

2.5.3 Hierarchies

Hierarchies are used for taxonomic names and to ensure accuracy and ease of data entry as individual fields of the hierarchy are validated and form a link, so when you enter a value at any level of the hierarchy the system automatically fills in all higher levels. Hierarchies may also be accessed independently of any collection item and can be printed out or exported electronically. Searching on hierarchies has caused problems in the past since it required the database to link lengthy strings of text over a number of separate fields.

2.5.4 Labels

Each discipline prints out labels of varying format and content and attaches them to their items. This ranges from a large label with limited data being tied to an artwork through to a tiny label with over 14 fields of data to be inserted into a small Molluscs specimen glass tube.

The problem for museum-specific databases is that there is a need and expectation that labels can be as easily created in the database as they can be in Word, but Te Kahui does not incorporate word processing software so all labels must be formatted by a programmer.

2.5.5 Images

As the number of records entered on the database grows, so too does the need for more images to be attached to records - in an ideal world all collection items would have an image attached. This in turn reduces the handling of fragile and rare items. The capacity of the database to display a set of images presently has its limits - as an example out of a result list of seventy two Colin McCahon works only fifty images may be viewed at one time. This feature will have to be greatly improved on in the future.

2.5.6 Histories

Histories provide a record of past information about an item and sit in the background of a record. They can then be accessed at a later date via a “Histories” menu. Histories on the database include past exhibitions an item has been in, condition and treatment reports carried out by a conservator, insurance details that change over the years, taxonomic names that change. Transaction history such as this goes on forever and is online unlike commercial databases where transaction history is taken offline. Basically as long as the data is of use to researchers it will remain online - which means indefinitely in museum terms.

Even items that have been deaccessioned (removed from the collection) remain on the database although are clearly marked as deaccessioned and history transaction information is included in all searches run on the database.

Within each discipline standards are implemented and there are a number of structured fields which have constraints placed upon them by the system. So if the data does not conform to those parameters, the database will not accept the information that is being entered. Two examples are date fields and fields containing the latitude and longitude of where a specimen has been collected.

2.5.7 Specialised Data Types

Unlike commercial systems, Te Kahui has no limit on dates. Although dates default to AD, artifacts date back to 8000 BC and 700 BP. Dates are strictly validated as they are a common field users search on. Often though the date of an artifact is unknown and so there must be room to give approximates, examples are "Circa" which means ten years before or after a date eg Circa 1900 stands for 1890 through to 1910 and "s" means the next decade or century eg 1960 s, 1900 s. Even a ? is acceptable in date fields eg 1987 ?

For specimens collected on field trips, a special format is used for latitudes and longitudes to show the degrees, minutes and decimal points of a minute. Thus to enter a latitude the format must be dd mm.ss N or S (N for north, S for South.)

2.6 Future Extensions

Long term, Te Kahui will need multimedia capabilities such as video clips and searchable attachments (eg Word documents), it may be bilingual, it will need a spell check, and it will need to incorporate bar-coding to track the movement of items and to carry out inventories - currently a mammoth job given the total number of items and specimens housed within Te Papa. Although a subset of fields and records is drawn from the database to "Te Papa Onscreen" (the multimedia database which is available to the public within Te Papa) it will also need to interface with the Museum's intranet and Web site.

It must also allow graphical images, numeric data and lengthy strings of free format text to be retrieved and displayed; include monitoring facilities and reports (for example, to monitor when items on loan are due back); have advanced searching capabilities and an online Thesaurus.

Te Kahui shows the intricacies of a museum collections management system and the unusual information requirements of staff within a museum - systems and requirements that commercial databases may not necessarily be able to accommodate.

Off the shelf systems specific to museums are available but support for these could be hard to find locally in New Zealand and there will need to be a high level of customisation.

A new system would have to comply with all requests in a more user friendly way with much more processing power and of course ensure smooth migration of the current datasets.

3. THE PROBLEM WIDENS

3.1 Preservation Issues

The propensity of information technology to transform itself every few years has wider implications for the preservation of cultural treasures than the difficulties it poses for keeping a collections management system up to date. When items in a collection are themselves digital in nature, their survival and integrity are threatened.

Firstly, electronic storage media may degenerate so quickly that information is lost. This can be prevented by the timely copying of an item to the same medium or to a new and, presumably, more stable and currently supported medium.

Secondly, even if the survival of an item is ensured, its integrity may be eroded as new technology replaces old. When items are digital, their "look and feel" is a significant part of their content, if not their identity. Copying them to a new format and accessing them by means of more modern software may well change their character and functionality.

It is useful to distinguish between the definition and constitution of a digital item. Some items can be defined by a digital record even though the record does not constitute it. A novel, for example, is defined by an ordered set of words and punctuation marks. While some copies may have special significance - copies of a first edition, for example - a novel is an abstract object that exists independently of its instances. A Tale of Two Cities, for example, is not constituted by any particular printed copy but by the ordered set of words that Charles Dickens published under that name. In Francois Truffaut's film Fahrenheit 451, futuristic victims of totalitarianism take it upon themselves to each memorise a famous novel to thwart the cultural genocide of the authorities, who are systematically burning all unindoctrinated works of art.

In such cases, a novel or poem might survive the destruction of all its copies as long as someone is able to reproduce the correct sequence of words and punctuation marks in the future.

Other items, by contrast, are constituted by their instantiation. They cannot exist unless a particular physical object, or set of objects, exits. This is most obvious in the case of computer hardware. One may decide that the original Apple Macintosh that appeared in 1984 with 128K of RAM is a design classic that deserves collection and preservation. If the last remaining such Macintosh were destroyed, the treasure would cease to exist. Historically faithful reproductions might be made, but the genuine article would have been lost forever. Software, too, may in future be considered as worthy of preservation for its own sake, in which its features, some of which may depend on the basic instruction set of a particular type of processor, must be preserved in working order.

3.2 Strategies for Longevity

The threats posed by rapid obsolescence of the storage media or software conventions of digital documents are starting to be recognised and strategies are being developed to give some protection from them. The main ones [following DRA (2000)] are:

- **Store.** Store the item in the hope that a future technology will be able to rescue it from oblivion.
- **Refresh.** Copy the item using the same kind of medium.
- **Transfer**. Copy the item to a more stable or more highly maintained medium, such as a CD or tape archive.
- **Freeze.** Copy the item to a very stable medium such as printed paper or microfilm that is probably readable by humans.
- **Migrate**. Copy the item to a currently supported format without trying to preserve the look and feel of the original.
- **Emulate**. Find software that will emulate the original in a currently supported hardware and operating system environment.
- **Preserve**. The same as migration except that a master copy is also stored in the original format which can be emulated once that becomes feasible.

1. **Store and Refres**h do not cater for technological changes, which are inevitable.
2. **Transfer** does, but because items retain their formatting, they require the original software to be retained or emulated and new software features are unavailable.
3. **Freeze** is only suitable for items that do not contain embedded functionality such as hypertext links, calculation logic, animation and the like.
4. **Migrate** caters for new hardware and software features but only at the cost of losing the look and feel of the original, which may be of scholastic importance.
5. **Emulate** is more promising, but it may not be available for the hardware and software mix required by a particular item. Alternatively, that which is available may be partial and compromised. For items that constitute a cultural treasure, emulation necessarily involves a loss of authenticity.
6. **Preserve** inherits the limitations of both Migrate and Emulate.

Even emulation, the most promising strategy, looks like failing in the long run. As time passes, there will be an ever increasing diversity in the hardware and software required to support the functionality of digital items. National archives depend upon the goodwill of the societies they serve for funding. The need for each institution to acquire and maintain a growing stock of commercially obsolete hardware and software to preserve its store of digital items will eventually prove prohibitively expensive.

3.3 Historical Computing Services

What is needed when authenticity is required and emulation fails, and what might emerge if the institutions charged with maintaining such collections are able to speak with a unified voice, is a new discipline of historical computing to support obsolete technology in much the same way as furniture restorers use historical materials and techniques to keep antique furniture in good repair.

Without a clear call from the Public Good sector, such skills are likely to be fostered by enthusiastic amateurs. The skill and energy levels of amateurs can be very high, of course, but amateurs follow their own interests and cannot be depended on to supply the right kinds of expertise for as long as they are required.

The commercial sector cannot be relied upon either. The preservation of cultural treasures is something that society demands whether or not doing so generates a monetary profit.

The lead, then, must come from the Public Good sector: libraries, museums and other institutions of collection and preservation that are funded either locally or nationally.

Motivation for cooperation should be high. After all, even if an institution does not currently hold a

document that requires an Acorn computer or a pre-1986 version of the Macintosh operating system, it may acquire one at any time in the future. Supporting a service industry that provides such services makes more sense than leaving every institution to maintain its own equipment and specialised knowledge. Perhaps the Internet can help by providing a cheap and convenient delivery mechanism of emulation software from a centrally maintained repository.

Specialist services to diagnose and repair obsolete hardware when this constitutes an item in a collection will be required whose practitioners are dedicated electrical engineers with the enthusiasm of an amateur and the dependability of a professional.

Such a call to preserve outmoded technology may suit the cultural climate. After all, it is possible to purchase ersatz copies of classic sports cars. Perhaps the same type of enthusiasm can be harnessed to provide historical computing services and all the more eagerly, one hopes, in view of the potential cultural benefit to be gained.

7. REFERENCES

DRA (2000). "A Draft Research Agenda for the Preservation of Physical Format Digital Publications". National Library of Australia. Accessed 26 January, 2000. http://www.nla.gov.au/policy/rsagenda.html