BORN AGAIN: DIGITISING THE ANTHROPOLOGY PHOTOGRAPHIC ARCHIVE

Rose Holley*
University of Auckland
Private Bag 92019
Auckland, New Zealand
Ph +44 9 373 7599 Ext 83984
Email: r.holley@auckland.ac.nz

INTRODUCTION

This paper describes how the University of Auckland's unique and valuable Anthropology Photographic Archive was digitally preserved, gaining a new lease of life after many years of neglect. The joint venture between the Anthropology Department and the University Library has resulted in the creation of a digital image database containing unique photographs of social anthropology and archaeology activities in New Zealand and the Pacific Islands since the 1950's. The database has great research potential for students and staff at the University and also people worldwide who are able to access and use the resource. I will describe how the project was undertaken, the issues involved and show the results.

BACKGROUND TO THE ARCHIVE AND PROJECT

In April 2002 the Anthropology Department contacted the library for assistance and advice on managing a 'small' archive within their department. They were concerned particularly about the deterioration of the archive. I made an appointment with the departmental photographer to assess the archive. It was rather a sorry sight that met my eyes. The archive was housed in a small storage room that had no environmental control. It doubled up as storage space for office equipment. About 80,000 negatives and some photographs were crammed into filing cabinets in non-archival enclosures in no apparent order and were visibly in a state of advanced deterioration (also confirmed by the strong smell of vinegar in the air). They dated from the 1950's. This is part of the report I filed:

"The filing system is not fully clear at this stage. In the archive itself items are in filing cabinets in 4 areas: negatives, prints, transparencies, and uncatalogued material since 1990. The negatives are arranged by photographers name and are separated by format 35mm and 4 x 5. The prints are arranged by location and/or subject. Not all negatives have a matching print and vice-versa. The correlation between the prints and negatives is unclear. There are approximately 50,000 negatives and approximately 60-80,000 items in total. There are some old card catalogues which refer to filing cabinet numbers 1-7 which seem to have been rearranged since then. In the 1970's a DOS Inmagic Database was created to catalogue and cross-reference the items. The system (version 7.2) was last used in 1992 or 1996. The hardware is on an old pc from the 1980's. On opening this up the battery has leaked and corroded the motherboard. The IT dept has been asked if they can save any information or the database from the hard-drive. 4 floppy disks have been found with data on – not clear what this data is, and it doesn't appear to have the database structure on it. We are trying to retrieve this data. One floppy was in the safe with the cameras – so obviously important at some stage. There should be a hard copy print out from the

database, but at this stage it hasn't yet been found. Without the database retrieval of photos and making sense of the system is very difficult. There may have been a thesaurus. There is a print out of subject headings. Whether or not any of this data can be restored or saved will radically affect the way the situation is handled."

Unfortunately the data could not be saved. In addition as environmental conditions in the archive were monitored, and taking into account the current state of deterioration we anticipated that we only had about 3 years left before all of the negatives would be beyond recovery and in effect lost. We immediately set about attempting to move the archive to cold storage or more suitable environmental conditions but for various reasons this did not eventuate. In the meantime the University Library proposed to apply for special funding to preserve and increase access to the archive by use of digital technologies. In October 2002 the Library and the Anthropology Department was jointly granted a small fund of money to achieve this.

PROJECT STEPS

Project Team
A team was established which comprised of library and anthropology staff members. I was the project manager, the departmental photographer was responsible for the archive, and additional staff were employed that had previously worked in the Anthropology Department and were familiar with the content of the archive. The first issue that arose was being able to effectively retrieve items of value that were known to be in the archive because the arrangement system was not understood, and being able to identify uncatalogued items that may be of value.

Inventory
The initial step was to create an inventory of the archive because we had no usable catalogue/archive system. The inventory was completed by an ex-member of staff and basically listed what was in each filing cabinet. Because speed was of the essence the inventory took the form of a notebook recording:
- Photographer, location, subject.
- Approximate number of rolls of film/negative numbers/format.
- Approximate date
- Condition or other notes of importance.

When it was finished the collections of most value, and those most at risk were highlighted.

Copyright and Cultural Sensitivity
Although the Anthropology Department owned the archive it was necessary for them to establish copyright ownerships and identify any potential cultural sensitivity issues. After seeking legal advice the department sent letters to known photographers or their surviving relatives asking for their permissions for digitisation. Since many of the photographers were well known in the anthropology field it was not too difficult to trace them. In addition a statement was drafted outlining acceptable use of the images which included a cultural sensitivity statement.

Image selection

Image selection was a time consuming stage of the project and took far longer than anticipated. We also had a tight deadline due to budget agreements. Subject experts (ex members of staff) used the inventory to identify suitable subject material that was at risk. They removed the items from the archive and indicated which negatives they felt should be digitised from negative strips. These were then passed to the photographer who checked that the ones with best image quality had been selected. In most cases only a few negatives were selected from whole sequences to give a representation of that subject or place. It was very difficult to do the selection from negatives even when using a lightbox because the images were so small, inverted, black and white, and often there would be many that appeared to be identical on a strip e.g. groups of villagers, archaeological digs showing excavations. It took 180 hours to select 5000 images.

The photographer created a sheet to be attached to each negative strip, which clearly indicated which images had been selected. This was invaluable for the scanning firm and also made it easier to return items to the archive afterwards. At the time of selection no description details were created which in hindsight would have made the description stage so much easier.

Image scanning
The scanning was outsourced. We obtained 3 quotes and selected 'Desktop Imaging' in Wellington for the contract. They visited the archive to get an idea of the situation and we discussed how we would submit the images to them and our scanning specifications. They were happy to work around our requirements. A summary of the project specification is below:

Transport and Handling of Original Material
- Negatives to be transported to Wellington in a secure pelican case to provide an airtight secure environment for the material during transit.
- Notification of dispatch and acknowledgement of safe receipt.
- Material not to be transported over weekends to minimise risk during transit.
- Negatives removed from the pelican case upon arrival in Wellington and stored in the vault.
- Negatives to be kept and returned with selection sheets in correct order.

Scanning
- Tiff Image (master). 3000 pixels across the long dimension of the image (outputting 300dpi over 10 inches) print quality. Not corrected for dust or scratches. [Black and white images 8-bit greyscale].
- Jpg images (for web). Generate image and thumb from the tiff - one at 800 pixels across longest dimension (outputting 200dpi) and thumb at 200 pixels (outputting 96 dpi).

File Naming and Structure
- Discs to be labelled in sequence e.g. UA_APA_01 (University of Auckland. Anthropology Photographic Archive, CD01)
- Structure to be 3 folders: master (contains tiffs), access (contains large jpg), thumbs (contains thumb jpg).
- File names to follow our instruction

Delivery Format
- On Mitsui CD ROM's following our labelling guidelines attached (no stickers on

---

> discs, use acid free card in the jewel case).
> - CD's returned separately to the original images.
>
> *Quality Control*
> - Any original items not scanned or named according to the above guidelines will be redone correctly at no cost to us within an agreed time frame.

We spent a considerable period of time discussing how the files would be named – meaningful or numeric? Eventually we agreed to make the file names the location of the negative in the archive. We therefore created a filing sequence like this:
Cabinet x, drawer x, sheet x, frame x   e.g. H_1_137_014.tif.
This served two main purposes. Initially it enabled easy re-filing of items back into the archive, secondly it would enable easy retrieval of similar items from the archive if a database user requested 'do you have any more like this?'

Quality Control and Storage of Images
On receipt of the CD's we immediately burnt an additional copy for archival purposes. The 5000 images generated 60 CD's. About this time the University Library was reviewing the current storage of digital resources and future requirements for digital storage. The result of this was the purchase of a Dell PowerVault 770N storage device with a 143GB hard drive. This was very timely for us and we were able therefore to load all of the archival images onto the storage device.

The photographer then quality controlled the images. We were very pleased with the work the scanning firm had done and there were minimal errors. However we were keen to check every image having had not so good experiences with different scanning firms on other projects. Quality assurance checked that:
- All negatives sent were returned.
- Images were the correct way round.
- File name, type and size/resolution correct.

Content Revealed – Born Again
It was only at this stage that the library staff working on the project began to really get a feel for the content of the archive. We were able to view some of the returned images and they were quite arresting. Some of the anthropologists had been very good photographers and images of local villagers who had never encountered white people before had a great deal of depth and held a lot of value, even if we were unsure what the significance or content of the photograph may be. It was at this stage that we began serious discussions about who should have access to the digital images. It had originally only been intended to provide access to selected university staff members and PHD students who currently had access to the physical archive. This was a relatively small group of people and the archive was certainly undervalued and under used due to difficulty of physical access, uncertainty about content and difficulty in retrieval and use of images in the negative format. However seeing the images in beautiful large clarity on a pc screen changed all that. For all those involved in the project it now took on a new lease of life and the archive was in effect 'born again'. The decision was made to allow the digital images to be public with the condition that they be used for research purposes only. In addition it was hoped that this might generate further interest for more funding.

The photographs broadly fell into two groups- archaeology in New Zealand and social anthropology in the Pacific. Noted anthropologists, archaeologists, and linguists were Ralph Bulmer, Andrew Pawley, Jack Golson, Bruce Biggs, Judith Huntsman and Anthony Hooper. In addition at a later stage of the project it was decided to include a large collection of photographs depicting Lapita pottery.

Database Design

Because we had no useful data or previous catalogue we decided to design a web accessible database in-house using Inmagic software. We wanted it to serve two functions:

- A delivery mechanism for the preserved digital images.
- A current archive system for new acquisitions.

At an early stage of our discussions we had noticed that the photographer was now working almost exclusively in digital format and the new images, which would previously have been stored in the archive, were now kept on department drives in a rather random fashion. In the past staff had returned rolls of films and cameras to the photographer who had added them to the archive, now with digital cameras the process had changed.

To this end a flat structure database was designed with 60 fields. A considerable amount of time was spent in consultation between the University Library and the Anthropology Department to discuss the structure, purpose and use of the database. The Library's considerable expertise in designing databases and interfaces was invaluable.

Description and Data Entry

Some anthropology staff members remembered the old Inmagic database from the 80's and were adamant that it didn't work properly with regard to being able to retrieve data effectively and this database should be better. Piecing the story together we thought this was because no authority files had been used and indexing was 'free text' i.e. over the years different people had catalogued things in different ways using different terms with no control. The Library was therefore very keen that authority lists should be used on several fields but particularly for place names, people's names, dates, and subject. Although this was agreed in principle, in practice this was one of the hardest issues to overcome in the whole project.

The data entry was done not by cataloguers or library staff (who were used to following guidelines and clearly understood the purpose of authority lists) but by academic subject experts. The subject experts needed to do data entry because there was very little written data in the archive and we were relying on the experts to remember or be able to interpret the content of the images. Their subject knowledge was invaluable. However they felt constrained by the use of drop down lists, and had differing opinions on even the simplest things like which countries should be in which world regions – Micronesia or Macronesia? We stressed the importance of adhering to an authoritative source and being consistent in data entry, so that data retrieval would be effective, but at times it was definitely a battle. Our other major hurdle was over the subject index. The options were:

- To use an existing subject index such as Anthropological Index Online (AIO)
- To use an existing subject index such as AIO and modify it.
- To modify the hard copy index we had found in the archive.

- To create our own subject index from scratch (the hardest option).

The department did not agree to any of the above but instead opted for free text keywords. This was partly due to the fact that we had limited time for data entry. The Library reluctantly agreed to using free text as an interim measure on the understanding that subject indexing would be a priority and done properly on the proposed next phase of the project. At the end of the data entry process all the project team members had learned a lot from the experience. In retrospect the University Library should have perhaps insisted on certain things, being the experts in the field of information retrieval, although in partnership agreements this can sometimes be difficult.

Apart from using authority lists the expert staff also struggled with data entry because when the images had been loaded into the database they became 'out of sequence' and 'out of context'. In the end they often went back into the archive to look at the original negative strips to understand the context and sequence. This was hard since they were working remotely from both the archive and the library most of the time. In the future we plan to enter the description information into the database at the time of selection and scanning, now that the system is fully implemented. This will considerably streamline and speed up the whole process. Some field notebooks were found in the archive that contained hand written notes by researchers detailing content of some films or trips. We had hoped that these would give us valuable information for data entry but they were often barely legible, abbreviated and patchy, only some being really clear and typed. It would have been necessary to spend much more time than we had available to fully utilise these resources. For these reasons it took much longer to complete a full description for an image than we had anticipated, 30 minutes rather than 5 minutes. We therefore had to compromise and only enter brief details in the key fields. On the whole the data entry for archaeological images of excavations was much more straightforward and quicker than that for anthropology because of the nature of the images. We were able to speed up data entry by using skeleton records as a batch process where information was the same for many images e.g. Lapita pottery.

Splash Page and Search and Results Interface Design
Once the data entry had begun and all the images were loaded we could think about the end design for the public interfaces and how we anticipated users might want to search. We had already gained a valuable insight by designing the data entry and search screens for staff. Because we still felt the archive was largely hidden and the content unknown we were keen to have a very simple interface but with strong visual impact giving the user an idea of the content of the database. We designed some screens using stunning photographs of people depicted in the archive, but were subsequently told these could not be used due to privacy issues. We then considerably toned down the visual impact of the interface, making it look more 'research' like and only using one photograph of a piece of well known Lapita pottery. The Anthropology Department were extremely pleased with this design and this was the chosen one.

Three search screens were created – a simple keyword search, a browse search, and an advanced search. These led to brief thumbnail results screens, and then the large image with full details.

THE MAIN ISSUES

All joint partnerships and major projects have issues along the way and on the whole this project went smoothly and according to plan. Our timeframe did slip slightly due to things outside of our control. Issues we encountered were sometimes expected such as differing opinions over splash page interfaces, and other times took us by surprise such as reluctance to use a subject index. Working with non-library people who were not schooled in basic information retrieval processes and requirements required a great deal of patience and explaining. Principles were often only understood when seen in practice or after errors had been made. Equally our lack of knowledge about the content and subject matter of the archive was frustrating at times. From the Library's point of view the main issues were:

- Authority Lists (Lack of understanding by academics/data entry staff of the importance of using them and their impact on end user searching. Lack of agreement on choosing/creating a thesaurus).
- Data Entry (Difficulty in describing sets of images that were out of sequence and out of context where many looked the same e.g. groups of villagers, excavations. The need for consistency in data entry not appreciated resulting in typing errors, use of singular and plural forms, use of synonyms, principles of specificity not followed etc).
- Staff working remotely from the archive and the library.

PROJECT RESULTS

Two years after the University Library had been called on for help we had successfully managed to:
- Obtain funding.
- Digitally preserve 5000 at risk images.
- Deliver 5000 images to the public in a web accessible database searchable by keyword, title, date, researcher, photographer, place name, activity type, category, collection name, site reference, peoples, file name, and record ID number.
- Provide advice and recommendations for ongoing management of the archive (including legal deposit document and environmental conditions).

FUTURE

The work done was only the tip of the iceberg. The life of the archive is limited and further recommendations have been put forward for a phase 2 of the project. The Library believes that expanding current descriptions and using a consistent thesaurus/subject index for existing images is a priority. It seems unlikely that enough funding will be obtained to preserve and increase access to the whole archive so the Anthropology Department is currently considering their priorities:
- Preserving more of the archive – those most at risk?
- Focusing on those of highest research value?
- Expanding the 'interest' value of the collection?

It will be interesting to monitor the usage of the database and see how this effects the use of the physical archive, and shapes the future plans.

CONCLUSION

The joint partnership and the project was a win, win situation for both parties. The anthropology department achieved their aim of getting some of their archive digitally preserved, and the Library achieved their aim of increasing access to a valuable collection while also gaining more expert knowledge in the digital field for future projects. I feel the project was a success due to the fact that both parties were quite clear about the joint objectives all the way through the project and also this was the third time the Library had undertaken a joint digital project with a faculty so had previous experience in potential partnership pitfalls and inter-departmental project management. We would sincerely like to continue our partnership with this project and build on what we have achieved so far.

USEFUL LINKS
Digital Project Information:
http://www.library.auckland.ac.nz/about/genlib/digital_projects.htm

The Database:
http://www.library.auckland.ac.nz/databases/alt/apa

Copyright and Cultural Sensitivity Statement:
http://www.library.auckland.ac.nz/databases/alt/apa/content/conditions