Routledge
Taylor & Francis Group

Digitisation of the WA Welcome Wall collection: a case study

Pauline Joseph*, Michael Gregg and Sally May

Pauline Joseph (PhD) is a Lecturer in Records and Archives Management at the Department of Information Studies at Curtin University. Pauline completed her PhD at the University of Western Australia in 2011. Her PhD research is titled *EDRMS Search Behaviour: Implications for Records Management Practices*. This study investigates the efficacy of electronic document record management systems (EDRMS) in enabling effective capture and dissemination of corporate information. The thesis examines the degree to which these systems are designed in accordance with the records management principles outlined in ISO 15489 to support the effective retrieval of records by knowledge workers.

Pauline's research interests are in the areas of design and implementation of EDRMS, information-seeking behaviour of knowledge workers, training and education of RIM services and programs for both knowledge workers and for the RIM profession.

Pauline's co-authored article entitled *Paradigm Shifts in Recordkeeping Responsibilities: Implications for ISO 15489's Implementation* published in *Records Management Journal* was selected as a Highly Commended Award Winner at the Literati Network Awards for Excellence 2013.

Michael Gregg was a journalist, writer, publisher, professional seaman, teacher and teacher–librarian, before returning to the field of his first degree, maritime history. He is currently the curator of the maritime history image collection and database designer and information manager for the Maritime History Department, Western Australian Museum, amongst other responsibilities.

Sally May is the Head of the Maritime History Department, Western Australian Museum. Sally graduated from Queensland University in 1983 and joined the Maritime Archaeology Section, Queensland Museum in that year. In 1985 she joined the WA Maritime Museum, Fremantle and was made Head of Department of Maritime History in 1995. Sally completed a postgraduate diploma in Heritage Studies, Curtin University, in 1998. In 1999 Sally was appointed the Exhibition Coordinator for the exhibition development of the new WA Maritime Museum, opened in December 2002.

This article is focused on the partnership between the WA Museum Maritime History Department and Curtin University's Information Studies Department on a retrospective digitisation project in 2012. The aim was to digitise the Welcome Wall project's paper records and link these records to the museum's online maritime history collection. The paper outlines and discusses the key business process questions that were considered prior and during the embarkation on the digitisation project. The conclusion shares lessons learnt from the project.

*Corresponding author: Email: p.joseph@curtin.edu.au

Introduction

Digitisation, also known as imaging or scanning, has been defined by *ISO 13028: Information and Documentation – Implementation Guidelines for Digitisation Rf records* as the 'means of converting hard-copy or non-digital records into digital format'.[1] Government and private organisations and memory institutions globally are embarking on digitisation projects for various business, library, cultural and archival service improvement reasons.[2] This is evidenced further in the statistics reported in *The Paper Free Office – Dream or Reality?* survey results published in 2012 by the Association for Information and Image Management (AIIM).

- 41% of organisations in the survey are using some form of digital-mailroom, either as a centralised operation or distributed at branch offices. 4% are outsourced.
- 20% of organisations scan half or more of their inbound mail at or before entry. A further 20% are more likely to scan at the point-of-process, and 29% scan-to-archive *after* the process.[3]

There are a number of articles published in *Image and Data Manager* on the implementation of digitisation projects in government and private organisations. These implementations are focused on digitising incoming paper records and using workflow modules in electronic document and records management systems to route these records to the action officers.[4] This form of digitisation is referred to as 'business process digitisation' and is different to 'project-based digitisation', which is best described as back-scanning in various guidelines and standards.[5] The difference between these terminologies is explained later in the article.

The literature cited generally pinpoint a number of business benefits for embarking on digitisation projects and the main reasons stated are to reduce or stop paper generation and move towards the digital office that offers greater accessibility, thereby increasing search and retrieval of corporate information. Digitisation, in turn, supports and improves the organisation's business processes to provide better accountability and governance of its information practices. Digitisation for preservation and risk management are also cited as significant reasons for digitisation projects in libraries and archives.[6]

What is missing in the literature is a detailed description of a case study in the Australian context for a project-based digitisation activity that adheres to best practice digitisation specifications outlined in its local jurisdiction. In the case at hand, the best practice specifications are those created by the State Records Office of Western Australia (SROWA). It is this gap in the literature our article addresses with this case study on the digitisation of the Welcome Walls collection.

Introduction to the Western Australian Museum Welcome Walls Project

Welcome Walls and similar monuments and edifices can be iconic features of museums concerned with the ethnic origins and identities of the communities they serve, especially in the New Worlds of the Americas and Australia. New York's Ellis Island Immigrant Walls of Honour and the Australian National Maritime Museum in Darling Harbour,[7] Sydney, are well-established examples that demonstrated the strengths and weaknesses of such projects.

In 2003 the Western Australian (WA) Government announced their intention to create Welcome Walls to honour migrants who came by ship through Fremantle; to have their names, the ship they arrived on and date of arrival, etched in panels erected

outside the WA Maritime Museum on Victoria Quay, Fremantle. Additionally, this information was to be accessible via the WA Museum's website, together with up to 50 words of text describing each person's migration experiences.

The Welcome Walls were to celebrate the 175th anniversary of the European founding of the Swan River Colony in 2004 and the diverse culture that resulted. Originally planned as a one-off project, stage 1, which ran from 2003–04, was so popular that a stage 2 was run in 2005–06 and stage 3 from 2008–10, which also included migrant arrivals at Albany. At the completion of stage 3, unveiled in December 2010, submitters had registered some 22,000 immigrant nominees on the Welcome Walls.

A considerable amount of records originated over the eight years of this project. An overview of these records and their past management is described in the next section to explore the records management issues leading to the digitisation project.

The Welcome Walls Project's records

The registration process involved the submitter completing a registration form and submitting it with a payment of AUD$66.00 to the WA Maritime Museum. Submitters were also able to purchase associated commercial products such as memento certificates, ship pictures and later a book *We Came by Sea: Celebrating Western Australia's Migrant Welcome Walls, Western Australian Museum*. The book includes the location of nominees' names on the panels.[8]

The definition of the terms 'submitters' and 'nominees' is critical to understand the key metadata entered in the registration forms that are the records. See Figure 1 for an example of the form and for an understanding of the required metadata.
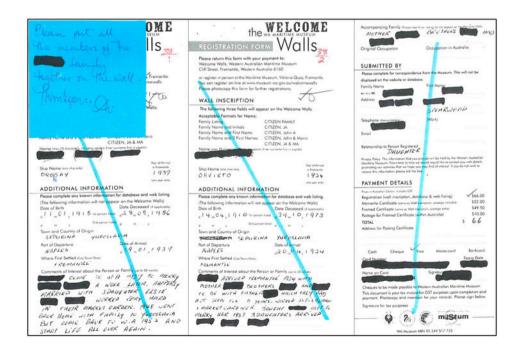

Figure 1.   Sample of a Welcome Wall registration form.

- Submitters refer to the person who submitted the form by completing details of their nominated family members and/or themselves on the form, and who paid the required AUD\$66.00 for the registration.
- Nominees are the people nominated on the form by their submitter, to be listed on the Welcome Wall panel and the WA Maritime Museum's Welcome Walls website.

A person could be listed as a nominee by different submitters. For example, different family members (son, daughter, niece and so on) of the same nominee could have submitted an independent form nominating them. This means the same nominee could be listed multiple times by different submitters. Where the submitters' details differed, such as arrival dates, names of nominees or spelling of the ship the nominees arrived on, it was not immediately apparent that there was duplication of the nominees and their name appeared more than once but with differing information.

Despite these issues, the records generated from the Welcome Walls project contain significant historical information about the attitudes, backgrounds and experiences of migrants arriving at the ports of Albany and Fremantle in Western Australia.

Background to the digitisation project

What follows is an overview of the information management issues leading to the digitisation of the Welcome Walls project.

User-contributed content

The challenges and lessons learned about user-contributed content in crowd-sourced settings in records, archives and libraries are shared and discussed by Steve Bailey and Kate Theimer.[9] Elizabeth Yakel points out that one of these challenges relates to the authenticity and credibility of user-generated content, which is applicable to the Welcome Walls records.[10] It could not be expected that the people (submitters) contributing and submitting data in the forms all had research experience and could or would use archival sources to verify the information they submitted to the Welcome Walls project, and indeed this has proved to be the case. Much of the information provided was anecdotal in nature, passed down over a generation or two and the dates of arrival, the name of the ship and so on, were frequently inaccurate and not verifiable by primary source records. These limitations of user-contributed content questioned the integrity of the data contributed by the Welcome Walls submitters on their registration forms. It highlighted the need to verify and quality-check this user-contributed content against refereed and published reference sources.

Outsourcing and data entry errors

The commercial department of the WA Museum undertook leadership of the Welcome Walls project and outsourced the information management activities that followed upon receipt of the registration forms. Given our late entry to this project it is unsure whether the expertise and involvement of the WA Museum's records and information management professionals was sought. However, comprehension of the data entry errors described next indicates no input from an information management professional was sought, not even to supervise and quality assure the data entry by the outsourced, non-records management savvy, data entry clerks.

Compounding the submitters' errors were the methodological and human errors created by the contract staff employed to manage the project. Firstly, they chose to manually enter all the data from the forms into a dynamic Microsoft Excel spreadsheet, rather than a database solution, with all the self-correcting, data-validating structure that a database would have employed. This meant that, for example, phone numbers were typed into an Excel 'number' cell, and were trimmed of leading zeroes and spaces, and in some cases, reformatted as consecutive numbers, making nonsense of the data. In other examples, by altering a date of arrival in one field to 1905 and then unconsciously scrolling the mouse down the column, some 20 or 30 dates were changed to 1905. Such errors are in addition to the inevitable typing errors and spelling mistakes in transcription. Furthermore, it was un-noticed that the 50-word histories were truncated by MS Excel to just 255 characters. In some cases only half the submitted text for each entry appears on the website. As the Welcome Walls, initially and at each subsequent stage, was conceived and managed as a discrete project, there was no continuity of staff or methodology. With each stage the contract staff changed and so too the data entry methods they used. Finding and correcting the resultant metadata errors became the focus of this Welcome Walls project.

When the third and final stage of the Welcome Walls project was wound up, the paper-based records generated by the project were placed in archival boxes and sent to the records management section of the WA Museum's head office. The records management staff registered the boxes with a basic synopsis of their contents into the HP TRIM system (TRIM) and the boxes were sent to off-site commercial storage. The Welcome Walls project's contract staff left and their computers were reallocated for operational work and only one electronic version of the final table of data was preserved in an MS Excel spreadsheet. This version was then uploaded to the Museum's website, including its errors.

Inheritance of the Welcome Walls project

Eventually, the Welcome Walls project was handed over by the commercial department of the WA Museum's head office in Welshpool to one of its branch offices, the maritime history department of the WA Maritime Museum in Fremantle. It is important to note that the latter is one of the many sub-branches of the WA Museum. The focus of the history department is the preservation and promotion of the maritime history of Western Australia within the context of the Indian Ocean.

The history department's involvement with the Welcome Walls project until its inheritance of the project was limited to providing the commercial department's contract staff with verified accurate information about the Fremantle arrival dates of ships, ship names (and their correct spelling), derived from a fully relational FileMaker Pro research database that integrates material from multiple sources, including researched and documented vessel histories ('vessels' table), shipping movements ('arrivals' table), images (vessels, people and places), related artefacts and memorabilia, and people (passenger lists and individual stories) and stories (anecdotal records and oral and press recordings of the experience).

Knowing the issues and experience of similar Welcome Walls projects, the history department was concerned about the quality of data being collected and displayed physically on the walls and virtually on the WA Maritime Museum's website. Moreover, there were, and are, ongoing requests arising from the information displayed,

ranging from requests for further information, through requests to facilitate family reunions, to demands that material submitted be removed or amended.

Early experience responding to these issues had demonstrated that it was near impossible to locate relevant paperwork when required, amongst the 44,000 records spread, somewhat randomly, across 18 archival boxes. A survey conducted by the AIIM in 2012 reports that respondents who have implemented digitisation state that customer or citizen response times are increased between two and three times faster and in some cases between five and ten times faster with access to digitised records.[11]

In the light of these issues, the head of the maritime history department contacted the department of information studies at Curtin University for advice and assistance on the management and digitisation of the Welcome Walls project's paper-based records.

Partnership with Curtin University

The lecturer for records and archives management at Curtin met with the senior curator and the curator at the maritime history department of the WA Maritime Museum to obtain an understanding of the issues and to scope the project requirements. It was decided to adopt a phased approach to this project and to involve Curtin's records and archives students where possible for learning with hands-on practical experience. Special practical and written assessments were tailored to credit students for their learning and participation.

An initial audit of the records and archives management practices at the WA Maritime Museum was conducted by two postgraduate students. These students also evaluated both the TRIM system used by the records management unit in the WA Museum's head office and the FileMaker Pro system used by the history department of the WA Maritime Museum to manage its maritime museum collection. This initial audit and review provided valuable insights into the different records management practices at the history department and WA Museum's operations and systems, which in turn assisted with planning the digitisation phase of the project.

In short, the WA Museum's head office used TRIM to manage the museum's paper records only, however at the history department (which is a branch office) TRIM was not implemented, hence not used. Given that the Welcome Walls project was managed by the commercial department located at the head office initially, these paper records were managed by the records unit at the head office. As such, the Welcome Walls project's paper records were already registered into the TRIM system at a very superficial level of registration: forms from letters A to C in one file, and then D to H in another file, and so on. However, the history department wanted to make the Welcome Walls contents currently hidden in the paper records accessible electronically on their FileMaker Pro system for two reasons: to correct and verify the metadata errors pertaining to these records already migrated from the MS Excel spreadsheets on the FileMaker Pro system; and, to link these digitised records to the existing online maritime museum collection on their FileMaker Pro system. Given these reasons, it was decided to manage the digitised Welcome Walls records for the project in the FileMaker Pro system instead of TRIM (the reasons for digitisation are discussed further below).

The focus of our paper is on the digitisation strategy used on the paper records and to capture them into the FileMaker Pro system so that they could be integrated as part of the electronic maritime museum collection. The paper records registered in TRIM remain the responsibility of the records management unit.

Strategies for the digitisation project of the Welcome Walls collection

To guide adoption of best practices for our digitisation project, *ISO 13028: Information and Documentation – Implementation Guidelines for Digitisation of Records*[12] was reviewed. The objective of the standard is to 'provide implementation guidelines for processes and policies for converting hard copy or non-digital records into digital format'[13] such as the: digitisation of paper records in a manner that enables the immediate and long-term accessibility and preservation of these records; maintenance of the authenticity of the digitised record to warrant legal admissibility; and the management of these records post-digitisation.[14] The digitisation guidelines are applicable for: business process digitisation for current and ongoing records; and/or, for digitisation projects of bulk legacy records. It is worth noting that the appended points that are relevant for our project are not in the scope of this standard:

- technical specifications for the digital capture of records,
- technical specification for the long-term preservation of digital records, and
- digitisation of existing archival holdings for preservation purposes.[15]

However, this is not a weakness of ISO 13028 as the various Australian state and federal archives have differing technical specifications and guidelines for digitised records and archives. The absence of coverage of the above bullet points is one reason that led us to adopt the archival digital preservation format specified in the *General Disposal Authority (GDA) for Source Records* for the long-term electronic preservation of these records.[16] The GDA approved in 2009 by the State Records Advisory Committee and published by the SROWA is among the many standards that enable WA state organisations to comply with the *State Records Act 2000* (WA).[17] The GDA mandates specifications for the digitisation of source records and the retention of the reproductions that meets the requirements of the State Records Act. Appendix 1 provides the minimum scan standard for digitised state records for colour and black and white: text, compound documents, drawing and photographs.[18] See Table 1 for an example of the different scan specifications provided in Appendix 1 of the GDA.

For the purpose of our digitisation project we referenced the GDA as a guide to confirm the master scan format required for long-term preservation of archival records as it is approved by the SROWA. Beyond these reasons, the GDA was not applicable to our project as it applies only to source records covered by the GDAs for administration, finance, human resource management and local government records.[19] Furthermore, the GDA applies to routine business scanning of source records while ours was a once-off retrospective scanning project. We also invited a senior representative from the SROWA to the museum, showed our records, discussed our digitisation strategy and consulted their advice for our project.

Table 1. Minimum scan standards for digitised state records.

Source format:	Scan resolution:	Scan type:	Scan format:	Master format:
Multi-page compound document (A2 or less) – text and graphics	300 dpi	Colour – 32 bit	TIFF v. 6 uncompressed	PDF/A
		RGB, 8 bits per channel	PDF	JPEG 2000

Having reviewed this documentation, it was obvious that our project was a digitisation project of the paper version of the Welcome Walls collection and not a business process digitisation project. The ISO 13028 standard defines a digitisation project as a 'retrospective, back capture of existing sets of non-digital records to enhance accessibility and maximise re-use'.[20] This is different from 'business process digitisation', that is defined as 'routine digitisation of records and incorporation into business information systems where future actions take place on the digitised record, rather than on the non-digital source record'.[21]

Having established the basics that we are working on the retrospective or back scanning of the Welcome Walls collection, we brainstormed and decided on the 10 key questions before embarking on the project. These are discussed below along with a commentary on what we did when we commenced the digitisation project. These key questions are suggested in the guidelines published by various Australian state and federal archives, such as the *Better Practice Checklist – 18: Digitisation of Records*[22] and *Just Digitise it: Information for Community Groups about How to Digitise Photographs and Paper Records*.[23] These guide questions are also covered in the digitisation workshops offered by the Australian Society of Archivists that were prepared by Recordkeeping Innovation Pty Ltd.

1.   Why digitise?

The identification of the business case is important before embarking on digitisation projects given the labour and cost resource implications of such projects. The ISO 13028 standard lists 11 benefits from digitised records in section 4.1.[24] Two of these key benefits applicable to our project are: greater and easier simultaneous electronic accessibility to the collection; and the increase in productivity when responding to citizens' information requests.[25]

The key benefit for the department from digitisation was to improve current data quality by verifying data submitted by the submitters. Improvements to the data quality would enable further research to be conducted once data was accurately registered and verified. Researchers want access to this data for reasons as diverse as conducting family and migrant history research, recording oral history and conducting training programs on cooking and languages.

The second key digitisation benefit is in increasing the online accessibility of the Welcome Walls project's information for both staff and, eventually, the public. There is ongoing correspondence with some of the submitters and nominees and the digitised version would enable staff quick and efficient access from their desktop to the collection. Online accessibility also enables more than one staff member access to the collection simultaneously.

Further digitisation benefits include the ability to develop the current knowledge and information about the Welcome Walls project. It would enable online links of this collection to existing maritime museum collections already registered on the department's FileMaker Pro system.

2.   What to digitise?

Once the justification for digitisation of the collection was established, it was next essential to scope what would be digitised. We agreed to digitise the following records that formed the Welcome Walls collection from the submitters:

(1) registration forms,
(2) correspondence with submitters and photos, and
(3) EFTPOS (Electronic Funds Transfer at Point of Sale) payment receipts.

We were contemplating whether to digitise the payment receipts as they were in an odd size (not A4). They were also in thermal paper and the contents had deteriorated to the extent that contents were frequently illegible, and the quality of the test scans were problematic due to the mottled background. Our investigations revealed that the museum's Centimen accounting system captures payment transactions so there is an official record of submitters' payments already. Hence, if there was a need for this record it could be retrieved from that system. The SROWA was appraised of this situation and we were advised that it was not necessary to scan the receipts given this information was captured in another business application.

Nonetheless, at the planning stage we decided to digitise the payment receipts in order to maintain completeness of the digital duplicate. However on first day of the digitisation process, we realised handling and collating the receipts was proving very awkward. It also slowed our digitisation process with paper jams on the scanner. We decided to omit digitisation of the receipts mid-way on the first day given this learning during the digitisation process about the difficulties with scanning the receipts, the knowledge that this information is captured in the museum's accounting system and SROWA's advice that scanning these receipts was not necessary for our project's purposes.

3.   What type of digitisation?

It was agreed that ours was a once-off retrospective digitisation project of the Welcome Walls collection and not an ongoing project. Also, that the output would be a 'static digital scan', that is, an image of the scanned item. Optical character recognition of the contents was not required or achievable given the forms were largely handwritten. Hence, we acknowledge that text within the scanned image cannot be searched. This is not an issue, as structured metadata fields are part of the FileMaker Pro system to aid searching.

4.   Digitise in-house or outsource?

We decided to digitise in-house for cost efficiencies, data sensitivity and information security reasons primarily.

Cost evaluations proved that it was cheaper to hire the scanners and digitise at the museum's premises with the help of Curtin's 18 archives students as part of their project placement assessment and learning. A period of six continuous working days was planned for the project. The 18 students were divided into two groups, with each group working for a period of three working days. Confidentiality agreements were signed by students and the lecturer prior to embarking on the project.

A number of different scanners and software packages were trialled intensively before the project, and it was found that none of the available packages met our entire requirements. In particular, most production scanners were optimised for bulk digitisa-tion of consistent format documents, and were unable to deal effectively with the mixed formats of our material. The convoluted feed paths of most of the scanners tested proved unable to either reliably feed mixed record stock (ranging from 300 gsm card

stock official documents and photographs, through 60–100 gsm forms and photocopies, to flimsy thermal paper faxes and receipts), or risked significant damage to the heavier paper records.

The Kodak i1220 vertical feed scanner proved to be by far the most reliable of the scanners we tested and was selected, combined with Kodak's Capture Pro software. These scanners offered a scan rate of up to 90 A4 pages per minute, effective detection of various sized documents, simultaneous duplex scanning, and page straightening. The straight, gravity-assisted feed path was undoubtedly the key to its more consistent document feed and detection performance.

However, considerable work had to be done to set up the software settings to provide the particular batch-processing settings we required. It was observed that very little of the available batch-scanning software was 'user-friendly' when asked to depart from the normal office scanning standards.

A contract was signed for the acquisition and services of the vendor. We negotiated hiring three scanners plus the scanning software installed in the three hired computers.

The technical expertise of the curator in our team influenced our decision for selecting the software that was to be provided with the hardware we had hired. This ensured the selected software integrated with the organisation's information technology (IT) infrastructure. We had intended that the hired computers and scanners would be directly connected to the museum network to allow simultaneous, real-time, multiuser access. However, government IT policy meant that instead the department's information systems consultants installed a one-way firewall to quarantine the transfer of the digitised contents from the networked computers to the department's information systems, meaning the hired machines could only be used for scanning, not processing and verification, which was carried out in the next room on the department's eight Apple Mac computers.

Having selected the equipment and resolving the labour matters, it was time to decide how the paper records were to be managed post-digitisation.

5.   *What happens to paper records after digitisation?*

When we scoped the digitisation project we planned for the long-term preservation of digitised versions. Hence, our decision to digitise to the PDF/A archive file format specified in the GDA.

The original paper records were fastened using archival paper clips and returned to original non-archival paper files and boxes in which they were previously stored. It was agreed to discard the payment receipts on thermal paper to prevent acid leaching into the other contents in the paper files and we arranged for confidential shredding of these receipts by the contracted off-site service provider of the museum.

The department will retain all other records on-site for reference if need be until all the digitised records are linked in the FileMaker Pro system. However, these files will remain as closed files.

Given that the archival status of the Welcome Walls record series was not determined at the time of the digitisation project, it was agreed that the records management unit would work with the SROWA to include the Welcome Walls collection of records in their functional retention schedule and then sentence these records accordingly. If this record series is deemed as archives, then an explanation needs to be made to the SROWA on the digitisation of this Welcome Walls collection. Given ours is a retrospective digitisation project, SROWA's permission needs to be sought to destroy paper originals as source records prior to destruction.

6.   What are the scanning rules and standards?

As stated earlier, the GDA[26] provided us with guidance when deciding the final master format of the records. As such, we adhered to the scanning standards outlined in Table 1 (adapted from the GDA for Source Records).[27]

As stated in column 1, Table 1, the registration forms are multi-page, double-sided compound documents, combining handwritten text and graphics, and are DL paper size (a third of A4 paper size). The front and back of the form were scanned and stored as separate images to allow flexible display later, that is, together (side by side) on the screen. This presentation format assisted later with metadata verification in the FileMaker Pro system.

The scan resolution was set to: 300 dots per inch; full colour; JPEG 2000 scan format; and PDF/A master format, as shown in Table 1, columns 2 to 5. Each scanner was calibrated every morning against a test card before production started.

It is worthwhile highlighting a couple of definitions and to discuss the implications on our digitisation project. Firstly, dots per inch refers to 'a measure of the resolution of a printer. It refers to the number of dots the printer is able to place in a linear one-inch space. The more dots per inch, the higher the resolution and the higher the printing quality.'[28] However, it results in larger file sizes therefore more storage space needs to be planned.

Secondly, the term 'scan format' in column 4, Table 1, lists examples of the common scan file output formats available in scanning software. That is, the file format PDF/A may not be a scan format available in a selected scanner's software but an alternative format like JPEG 2000 could be available. This was the case in our project.

In summary, adhering to the GDA, we digitised the access master copy to the PDF/A format. Initially, the records were scanned to the JPEG 2000 file format as the scanning software only permitted this format. An automatic batch process was developed to automatically convert the JPEGs to the PDF/A file format in the background. There was no loss of content or format during the conversion. This process saved us several steps in our digitisation work processes. We could have retained each individual JPEG 2000 image as it is also a specified file format in the GDA. However, PDF/A was selected to retain the multi-page context and content of the records. Another reason for converting to the PDF/A file format was to align with the international specifications for long-term preservation of electronic records in *ISO 19005-1: 2005 Document Management – Electronic Document File Format for Long-Term Preservation – Part 1: Use of PDF 1.4 (PDF/A-1)*[29] and *ISO 19005-2: 2011 Document Management – Electronic Document File Format for Long-Term Preservation – Part 2: Use of ISO 32000-1 (PDF/A-2)*.[30]

There was initially no requirement to create surrogate copies, as the access master copy would be backed up on the network and we also had a saved version of the JPEG 2000 as the preservation master format. However it was found that once the project was in full swing, the considerable network traffic generated by three computers saving high-resolution images to the server while eight other machines simultaneously opened them from the server caused unacceptably long server lag times. We therefore decided to create low-resolution thumbnails that would display in the first instance, while a click of a button would bring up the high-resolution image only when necessary. This resolved the problem immediately.

It was agreed that no image manipulation would be made, and as such no cropping of the digitised images was performed.

The GDA states that the contents need to be scanned to the original size.[31] However, the contents were not always scanned to actual size in our project. The reason for this is discussed in the conclusion under the sub-section *Scan to Original Size*.

7.   *How images will be registered in the scanning software?*

The scanning software was set to automatically number the scans using the box number allocated in TRIM, followed by an automatically generated sequence number separated by an underscore, for example, Box0030_00015.

This number is automatically imported into the FileMaker Pro system with the scan to create a 'document' which can then be viewed on screen and matched with the relevant existing data.

The department's FileMaker Pro database is a complex relational database, comprising separate tables for submitters, nominees, documents (the scanned records) and ship movements. There is a fifth table in the background that permanently retains the original MS Excel data as received by the department, so it can always be compared with the revised data. All five tables are automatically linked so, for example, when looking at a submitter the researcher can automatically see every nomination they submitted, and every piece of correspondence. Similarly, all documents submitted about a nominee is also viewable and when more than one person has submitted information about a nominee it is readily apparent that there is a duplication. Finally, it is possible to search a ship's arrival and be shown every nominee who arrived on that ship that day.

All fields are designed to automatically flag conflicts within the data, such as dates of arrival after the date of death, and so on.

8.   *Who owns copyright?*

The submitter's permission is not required to digitise the registration forms for the reasons outlined earlier under *Why Digitise?* as the form is owned by the museum.

By submitting the registration form and payment of AUD\$66.00 the submitter authorised the museum to publish specified details provided on the form on the Welcome Walls and on the museum's website. See 'payment details' in Figure 2. The submitter paid the museum to publish specific information: name of nominee, year of arrival and the name of the ship they arrived engraved on the Welcome Wall and to publish their 50-word story on the museum's Welcome Walls website.

Figure 2.   Payment details on the Wall registration form.

Submitters included photocopies of photographs, birth certificates, passenger lists and other key documentation pertaining to the nominee with some of the registration forms. All these original materials or documents were copied and the originals returned to the submitter. The copyright of all these contents is owned by the submitter. This means that the museum requires the submitter's permission prior to publishing any of these contents on the Web. Likewise, the submitter's permission is required prior to divulging any personal information to the public or on the Web or otherwise.

The registration form did not include a disclaimer about who owns copyright status of the contents provided and this is discussed later under the lessons learnt on copyright.

9. What are the digitisation processes and tasks?

The records spanning the Welcome Walls' three stages are stored in 18 A4-size archive boxes, containing approximately 2500 records for each stage and arranged in an inconsistent alphabetical order. It was decided to digitise following the physical order of the records, then to export the digitised records from the hired computers to the FileMaker Pro system.

Eighteen archives students were enlisted for the project. The students were split into two groups. The first group comprised eight students and the second group ten students. Each group worked for the duration of three days, split into three or four students per group. The lecturer, senior curator and curator assisted the groups when required, so they could keep up with their tasks. There were 18 boxes and 18 students, hence each student was assigned a box to complete or each group was assigned three boxes to work over the three days.

(1) To begin, all students commenced de-metalling the records in each paper file. These tasks were by far the most labour-intensive and time-consuming, but this is a precursor to start the digitisation process. Both the senior curator and curator started de-metalling two days prior to first day of the digitisation process before the students arrived. Each group had a box de-metalled already to enable some team members to start digitisation while another member de-metalled the next box of files.

(2) The records were then sorted in the following order: form, documentation (attachments to form such as birth certificates, photographs and so on), correspondence, receipts.

(3) Each record was numbered by the archive box number and assigned a sequential number, for example 370/1 meant box 370 and sequential number 1. This gave each record a physical location, enabling it to be easily relocated if required. During the digitisation process, student 1 continued with steps 1 to 3 while student 2 started step 4, and student 3 commenced step 5.

(4) The records were scanned in strict order and quality checks were done whilst scanning. Scanning was stopped if there were paper jams or when the notes attached to the registration form could not feed through the scanner.

(5) After the scanning, the records were reassembled into the order stated in step 2 and filed back to their original paper files and box. Archival paper clips were used to keep related contents together.

Steps 1 to 5 ended the tasks on preparation of the records for scanning and the actual scanning itself.

(6) The scanned records were then linked to the existing metadata on the FileMaker Pro system. Then verification was conducted to ensure the user-contributed metadata on the digitised records matched what was already captured in the system. Metadata errors were fixed and where interest summary information was incomplete it was typed in.

10.　*How will quality checks be done?*

We acknowledged when brainstorming the digitisation process that the Curtin students would have limited time for comprehensive quality checks. Hence, quality checks were confined to visual checks during the scanning process to verify that the image quality of the digital output matched the input and that the images were not skewed. Further image quality checks were conducted when linking scanned records to metadata in the FileMaker Pro system. It was agreed that more comprehensive quality checks needed to be done by the museum's staff or volunteers following completion of the digitisation project.

Conclusion: lessons learnt

For the WA Maritime Museum and Curtin University the digitisation project is a success, as it provided benefits for both organisations. To better understand and appreciate the intensity of the work undertaken, a total of 34,500 documents (contained in 18 archive boxes) were prepared (de-metalled and numbered) for scanning, 16,680 by the first group of eight students and 17,820 by the second group of ten students. Using three hired scanners, the total number of scans at the end of the project was 42,406. Of the 21,774 names contained in the FileMaker Pro system, representing nominations for display on the museum's Welcome Walls, some 6000 were linked to the scans.

By way of a small token of the museum's appreciation, the executive director of Fremantle Museums and Maritime Heritage personally awarded each of the students and the lecturer with a certificate of appreciation and a family boarding pass for 12 months' free entry to the WA Maritime Museum.

A post-implementation review of the project provided the appended lessons learnt.

Preparation of records for digitisation

The time and labour required to prepare the paper records for digitisation has to be planned when costing and planning for the project. Tasks like de-metalling, straightening paper folds, sorting the paperwork into desired order and aligning the records for feeding into the scanner are meticulous time- and labour-intensive work. However, if these repetitive tasks are not done correctly it will result in delays during the scanning activity. For example, staples and/or paper folds may cause paper jams or missed scans. The time taken to individually assign a consecutive number to each paper record was considerable, but proved to be an essential timesaver when a record had to be found and rescanned, or for any quality assurance requirement.

Master format

We could have left the scanned images in JPEG 2000 and not converted to PDF/A format, as JPEG 2000 is also an accepted archival format in the GDA.[32] A decision

was made to proceed with the PDF/A file format as it is an internationally recognised file format for the long-term preservation of electronic documents/records.[33] Also, it was easy to write a code to convert from JPEG 2000 to the PDF/A format.

Scan to original size

It would have been ideal if we could have scanned all the records to their original size, which is a requirement stated in the GDA.[34] Unfortunately, this became difficult when there were sticky notes and extra odd-sized items taped or stapled on to extend the A4 documents. The scanner selected would at times crop off the sticky notes attached at the tail end of the A4 document whilst scanning. Hence, we devised a work-around and where possible, either we moved the sticky note to a blank spot on the application form and scanned it in to keep it all together, or photocopied the sticky notes and other odd-sized attachments and pasted them onto a separate A4 document to be scanned. The GDA states that where the source records cannot be scanned to original size, then a resized scan can be made for accessibility only, and the original source records need to be retained.[35] Hence, in hindsight we should have taken a resized A4 photocopy of these pages for digitisation if the records management unit wanted to use these digitised records as source records in future.

Good form design for information capture

A number of the errors in the project could have been omitted if the initial form designed for capturing information from submitters was designed with more spacing and clearer instructions. The poor design of the form led to submitters adding post-it notes and other odd-sized paper attachments to complete the form. An option for more space to provide their information would have avoided these attachments and led to legible handwriting and eased the scanning process.

Copyright

In hindsight and as a risk management strategy, it would have been ideal if a disclaimer was included on the registration form stating that all copyright in the material submitted was to be assigned to the museum to streamline publication of all information submitted, the way it would be used and displayed and so on, into the future. Such a disclaimer would have provided the museum full copyright to publish the accompanying materials, such as photographs, on the Web without having to contact the submitters for their permission. This is another example demonstrating the need for organisations to consult and engage their records and archives professionals when embarking on new projects. It also indicates the need for our profession to be proactive and volunteer our expertise to the organisations' projects.

IT-savvy team member(s)

We emphasise the need to enlist IT personnel or have a team member who is IT savvy early in the project so that they are aware of the project requirements and can input their technical expertise to test-drive and select the scanning hardware and software for the project. We found the curator's input was valuable when evaluating the scanning

software and hardware to ensure the technical requirements for what we wanted to achieve with the digitisation project.

Limitations of off-the-shelf solutions

One of the unexpected problems was the difficulty of finding the ideal combination of scanning hardware and software at an affordable price. In our case there was only one clear choice of hardware that dealt effectively with our mixed media records, but that led to a compromise in the capability of the available software. It was also vital for the success of the project that we were using a user-configurable database solution (FileMaker Pro) so that we could work around unexpected problems immediately, rather than having to wait for a consultant to be called in.

Working with user-contributed content and content with data entry errors

In comparison, the prior process of preparing the content for digitisation by de-metalling and numbering the records was not as time-consuming and labour-intensive as the post-digitisation process. The latter involved physically linking each digitised record to the existing record on the FileMaker Pro system, then fixing the data entry errors made by the outsourced clerks followed by verifying the user-contributed data.

The data entry errors made by the clerks were easily fixed by visually cross-checking the metadata fields with the errors displayed on the left of the computer screen against the digitised image displayed on the right of the screen. However, more effort was required to address the user-contributed errors pertaining to the dates of arrival and the name of the ship but there was assistance from the FileMaker Pro system. The curator had programmed a computing code to automatically verify the user-contributed data regarding these metadata fields against authoritative data from other shipping tables on the FileMaker Pro system. Hence, a red flag was raised when there were inconsistencies in the user-contributed data with optional ship names and arrival dates in the system. This prompted the students to make a notation in the system that these user-contributed metadata were incorrect and needed to be checked with the submitters prior to correcting what was submitted. This exercise would be a separate follow-up action for the Welcome Walls project. This further highlights the issues involved in working with user-contributed data, and the time and financial resources it takes to get these rectified.

We are of the view that we have not altered the 'recordness' of the original metadata provided by the submitter, given we were only making a notation of the errors and did not amend it on the system. Likewise, when we amended the data entry errors by the outsourced clerks, we were not altering the 'recordness' as the initial data entry was performed incorrectly with little or no quality assurance performed. Additionally, the image of the registration form is linked to the amended metadata field and both these sets of information can be viewed side by side on the screen, further retaining the 'recordness' of the form.

The importance of records management

Stages 1 to 3 of the Welcome Walls project proved that, without good records and archives management programs and practices in place, information chaos will prevail. It further highlighted the need for records and archives professionals to be actively aware and informed about what is happening in their organisations so that

they can be proactive and market their expertise and services when new projects arise. This project highlighted the need for the profession to work closely with different parts of the organisation, in this case with museum colleagues in the commercial and marketing and finance departments, including professionals like curators and historians.

Another key lesson learnt in the absence of records and archives professional input was in regard to the risks associated with the outsourcing of key information tasks to staff who are not trained in information management. This is evidenced in the use of non-compliant or inefficient technologies like Microsoft Excel spreadsheets for data entry of user-contributed handwritten content. It also explains the inherited data entry issues stated earlier plus the poor quality assurance conducted prior to publishing vital information on the erected Welcome Walls. A further lesson learnt includes cost implications to rectify the poor information practices of untrained and outsourced staff.

Cost implications for digitisation projects

The project provided insights into cost estimates when embarking on digitisation projects. The quotes for the hardware and software ranged from AUD$3000 for one scanner for one week to the quote we accepted for AUD$1200.00 for hiring three scanners and three computers for 28 days. The latter quote included hiring the software for digitisation on all three computers. The actual operational digitisation tasks took place over six continuous days but the hardware and software were hired earlier to enable prior testing and preparation tasks.

The labour provided by Curtin University's lecturer and students for the project was pro bono and the labour costs presented here are for indicative purposes only. The hourly rate for contract staff shared by Information Enterprises of Australasia was used to work out the cost presented. The labour cost for preparing (de-metalled and numbered) for scanning of a total of 34,500 documents (contained in 18 archive boxes) by 18 students and a lecturer over six days is approximately AUD$20,000. This excludes the consultancy costs by the lecturer when planning and testing for the project prior to these six operational days. It also excludes the labour costs for the senior curator's or the curator's time spent on this project. Also omitted are costs of the museum's IT staff engaged to ensure firewalls and computer security were set up accordingly.

In total, this digitisation project's costs are approximately AUD$21,200 (hire cost of hardware and software AUD$1200 plus labour costs AUD$20,000).

It must be pointed out that the project team successfully scanned all the 34,500 documents but only had time to link, verify and amend data entries of 6000 of these scanned images to the 21,774 names contained in the FileMaker Pro system, representing nominations for display on the museum's Welcome Walls. To complete linking and verifying the remaining 15,774 scanned images to the names may cost another AUD$16,632 for these 18 students working eight hours over three days each for another week. This would make the grand total to complete this project approximately AUD$37,832 (that is, AUD$21,200 + AUD$16,632).

These approximate digitisation costs also highlight the cost implications of working on digitisation projects where user-contributed data needs to be verified for authenticity and credibility. It also draws attention to the financial cost for not managing records properly the first time.

Occupational health and safety

The location of the photocopier on a separate floor from where the scanning was conducted meant that people were constantly walking up and down the stairs. We had to run down the stairs to photocopy and run up to continue with the scanning. Some welcomed the opportunity to get up and stretch their legs but it did slow down the scanning process. Also, a little more working space would have made it more comfortable working in groups and with the records, boxes and scanners. The need for large working spaces with quick access to other office equipment when embarking on digitisation projects in-house also needs to be considered.

Future plans

Plans to further develop the Welcome Walls collection include measures to develop better access to the collection. Firstly, to work with the department's volunteers to link and verify the digitised records in the FileMaker Pro system. Secondly, to contact submitters and nominees and offer them the opportunity to have photographs and/or official documentation such as passports and passenger lists added to the virtual wall to enrich the experience. These materials will be returned once copied. Adding these materials to the collection would provide context to the migrants' experience. It will also enable integration of their voyages with the museum's artefacts and image collections. Thirdly, there are plans to enrich the collection by adding oral or video recordings with the nominees. Fourthly, to extend the museum's data visualisation technologies to publish the Welcome Walls contents on the public domain and use crowd-sourcing techniques to get the public to engage with the collection online.

Acknowledgements

Endnotes

1.   International Organisation for Standardisation, *ISO 13028: Information and Documentation – Implementation Guidelines for Digitisation of Records*, International Organisation for Standardisation, Geneva, 2012, p. 3.
2.   IG Anderson, 'Pure Dead Brilliant? Evaluating the Glasgow Story Digitisation Project', *Electronic Library and Information Systems*, vol. 41, no. 4, 2007, pp. 365–85, available at <*doi:http://dx.doi.org/10.1108/00330330710831585*>; D Birrell, M Dobreva, G Dunsire, JR Griffiths and RJ Hartley, 'The Discmap Project: Digitisation of Special Collections: Mapping, Assessment, Prioritisation', *New Library World*, vol. 112, no. 1/2, 2011, pp. 19–42, available at <*doi:http://search.proquest.com/docview/855901914?accountid=10382*>; 'The

Digital City Delivers in Paperless Push', *Image and Data Management Journal*, May/June 2011, available at <*http://www.idm.net.au/article/008571-digital-city-delivers-paperless-push*>; 'Digital Designs', *Image and Data Manager*, September/October, 2012, pp. 8–9; C Draycott, 'The Welcome Trust Medical Photographic Library Digitisation Project: A Case Study', *Journal of Audiovisual Media in Medicine*, vol. 23, 2000, pp. 165–70; A Hampson, 'Practical Experiences of Digitisation in the BUILDER Hybrid Library Project', vol. 35, no. 3, 2001, pp. 263–75, available at <*doi:http://search.proquest.com/docview/57450681?accountid=10382*>; 'Hyundai Motors Ahead with Process Automation', *Image and Data Manager*, September/October 2012, pp. 18–20; AM Zuraidah and A Ismail, 'Malaysian Cultural Heritage at Risk?', *Library Review*, vol. 59, no. 2, 2010, pp. 107–16, available at <*doi:http://dx.doi.org/10.1108/00242531011023862*>; 'Tasmania Hatches E-Govt Initiative', *Image and Data Management Journal*, January/February 2013, p. 33. All sites accessed 2 June 2013.

3. D Miles, 'AIIM Industry Watch: The Paper Free Office – Dream or Reality?', available at <*http://www.aiim.org/pdfdocuments/IW_Paper-free-Capture_2012.pdf*>, 2012, p. 4, accessed 10 January 2013.

4. 'Brickworks Builds a Digital Future with Efficiency Leaders', *Image and Data Manager*, May/June 2012; 'Capturing Census 2011', *Image and Data Manager*, 2011, pp. 16–28; 'Hyundai Motors Ahead with Process Automation'; 'The Digital City Delivers in Paperless Push'; 'Digital Designs'.

5. Department of Finance and Deregulation, A.G.I.M.O., 'Better Practice Checklist – 18. Digitisation of Records', 2004, available at *http://www.agimo.gov.au/archive/better-practice-checklists/digitisation.html*, accessed 28 November 2012; International Organisation for Standardisation, 2012; Public Record Office Victoria, 'Just Digitise it: Information for Community Groups about How to Digitise Photographs and Paper Records', 2011, available at <*http://prov.vic.gov.au/wp-content/uploads/2011/07/Just-Digitise-It.pdf*>, accessed 20 January 2012; State Records Commission of Western Australia, 'General Disposal Authority for Source Records', 2009, available at <*http://www.sro.wa.gov.au/sites/default/files/gda_sourcerecords.pdf*>, accessed 17 April 2012.

6. AE Bulow and J Ahmon, *Preparing Collections for Digitization*, Facet Publishing, London, 2011; D Roberts, 'Digitisation and Imaging' (Chapter 13), in J Bettington, K Elberhard, R Loo and C Smith (eds), *Keeping Archives*, 3rd ed., Australian Society of Archivists Inc., Australian Capital Territory, 2008, pp. 402–34.

7. Australian National Maritime Museum, home page, available at <*http://www.anmm.gov.au/ww*>, accessed 18 December 2012; The Statue of Liberty-Ellis Island Foundation, Inc., home page, available at <*http://www.wallofhonor.org/*>, accessed 18 December 2012.

8. N Peters, *We Came by Sea: Celebrating Western Australia's Migrant Welcome Walls*, Western Australian Museum, Perth, 2010.

9. S Bailey, *Managing the Crowd: Rethinking Records Management for the Web 2.0 World*. Facet Publishing, London, 2008; K Theimer (ed.), *A Different Kind of Web: New Connections Between Archives and our Users*, Society of American Archivists, Chicago, 2011.

10. E Yakel, 'Balancing Archival Authority with Encouraging Authentic Voices to Engage with Records', in K Theimer (ed.), *A Different Kind of Web: New Connections between Archives and our Users*, Society of American Archivists, Chicago, 2011, pp. 90–1.

11. Miles, p. 9.

12. International Organisation for Standardisation, 2012.

13. ibid., p. ii, paragraph 2.

14. ibid., p. 1, section 1.

15. ibid., p. 1, section 1.

16. State Records Commission of Western Australia, 2009.

17. Government of Western Australia, 'State Records Act 2000', available at <*http://www.slp.wa.gov.au/statutes/swans.nsf/be0189448e381736482567bd0008c67c/3988e10065ed24a948256a5d0004cf94?OpenDocument*>, accessed 12 July 2013.

18. State Records Commission of Western Australia, 2009, pp. 15 to 25.

19. ibid., pp. 5 and 6.

20. International Organisation for Standardisation, 2012, p. 3, section 3.6.

21. International Organisation for Standardisation, 2012, p. 2, section 3.3.

22. Department of Finance and Deregulation, A.G.I.M.O., 'Better Practice Checklist – 18. Digitisation of Records', 2004, available at <*http://www.agimo.gov.au/archive/better-practice-checklists/digitisation.html*>, accessed 28 November 2012.
23. Public Record Office Victoria, 'Just Digitise it: Information for Community Groups about How to Digitise Photographs and Paper Records', 2011, available at <*http://prov.vic.gov.au/wp-content/uploads/2011/07/Just-Digitise-It.pdf*>, accessed 20 January 2012.
24. International Organisation for Standardisation, 2012, p. 4, section 4.1.
25. ibid.
26. State Records Commission of Western Australia, 2009.
27. ibid., p. 22.
28. Department of Finance and Deregulation, A.G.I.M.O., 2004, p. 57.
29. International Organisation for Standardisation, *ISO 19005-1: 2005 Document Management – Electronic Document File Format for Long-Term Preservation – Part 1: Use of PDF 1.4 (PDF/A-1)*. International Organisation for Standardisation, Geneva, 2005.
30. International Organisation for Standardisation, *ISO 19005-2: 2011 Document Management – Electronic Document File Format for Long-Term Preservation – Part 2: Use of ISO 32000-1 (PDF/A-2)*, International Organisation for Standardisation, Geneva, 2005.
31. State Records Commission of Western Australia, 2009, p. 17.
32. ibid., p. 22.
33. International Organisation for Standardisation, 2005, pp. 33 and 34.
34. State Records Commission of Western Australia, 2009, p. 17.
35. ibid.