Submitted on: 04.07.2016

Dr Perfectionist:
or How I Learned to Stop Worrying and Love the Quantity

Damien Cassidy
Digital Experience Division, State Library of New South Wales, Sydney, Australia.
E-mail address: damien.cassidy@sl.nsw.gov.au

Abstract:

With time running out to digitise large collections on magnetic media, how much can be done in that time that's left? The State Library of NSW is undertaking a project to digitise 5,000 hours of audio on Compact Cassette, Micro Cassette, Quarter Inch and DAT over a period of 6 months.
This paper will discuss the Quality Assurance strategy applied to address the questions, is that scale of digitisation sustainable, suitable and satisfactory?

Keywords: Quality Assurance, Digitisation, Audio, Magnetic Media, Oral History, New South Wales, Australia.

1 Oral History And Sound Recordings

Prior to c1991, the State Library of NSW had usually only been receiving oral history and sound recordings as part of larger paper based collections though some dedicated oral history collections were acquired. After 1991, the Library became more proactive in seeking material for its oral history collection.

Some highlights of the collections are:

- **The NSW bicentennial OH Project**
  - Interviews with two hundred men and women aged eighty years and over, who lived in New South Wales between 1900 and 1930.
- **The Ethnic Affairs Commission OH project**
  - Interviews conducted with Australians from non-English speaking backgrounds concerning the migrant experience of arriving in Australia in the early 1950s.
- **The Twelfth Hour Oral Testimonies Project**
  - 207 interviews with Holocaust survivors by the Australian Institute for Holocaust Studies.
  - These interviews document the memories of the participants' lives in Europe before the outbreak of war, during the Holocaust, including but not limited to the concentration camps, and their immigration to Australia before, during and immediately after WWII.

These recordings provide a different dimension to the Library's collection. Sound recordings contain rich storytelling and reveal nuances in mood, personality and language not always present in written works.

The Oral History and Sound Recoding collection at the Library are estimated at:

- ≈12,400 original items (and growing)
- ≈11,610 hours
    - 86% Cassette
    - 11% Reel to Reel
    - 3% DAT, CD, DVD, VHS…

2 A Brief History Of Oral History Digitisation At SLNSW

Prior to the 2014-15 fiscal year only a small amount of ad-hoc digitisation had been undertaken. This was often reactionary to meet a specific demand though some proactive digitisation was also done.

- Approximately 975 tapes
- 900 hours of digitised material
- 1,600 Archival Master files

Whilst these numbers appear small, they still accounted for almost 8% of the total collection being digitised, which is not a bad figure considering there was no targeted program or in-house capacity.

3 Digital Excellence Program

In July 2012, the NSW Government along with Library announced the Digital Excellence Program. This is a comprehensive 10year digitisation program aiming to transform the Library from one of Australia's most respected and valued Libraries to a library internationally renowned for digital excellence. $AUD22.4m was funded for digitisation in the first 5 years across a wide variety of the Library's collections. This would enhance the Library's preservation, access and online services, including an upgrade of the Library's catalogue & collection management systems and website. Another clear goal of this program was to develop industry capacity for digitisation.

The Library recognised that it didn't have the experience or knowledge in audio digitisation to go it alone. In September 2014, a consultancy report was commissioned using the services of the National Film and Sound Archive (NFSA). This document provided advice and guidance across a wide range of areas relating to audio digitisation including the development of a quality assurance framework, defining roles, responsibilities and technical standards.

To further enhance the Library's knowledge and experience and 'test the waters' in this area, a small pilot program was undertaken. It is a commonly used risk mitigation strategy to not 'put all your eggs in one basket'.

- 4 partners were selected
- 868 tapes were identified
- 560 hours
- 1,262 Archival Master files
- 7% of the total collection

At the conclusion of this pilot project, 14.9% of the total collection had now been digitised and many valuable lessons had been learned.

Following on from a recommendation of the NFSA report, the Library recruited a Digital Media Technical Analyst to increase in-house subject knowledge, which is me.

Now the Library had experience, knowledge and in-house expertise a rigorous selection process was undertaken to select our digitisation partners and we commenced the 2014-15 Mass Digitisation Project. These were ambitious projects seeking to digitise around 3,800 hours of Oral History recordings on Compact Cassette, ¼" Open Reel and DAT. It would represent about 40% of the total collection and the timing meant there was only a little over 2 months in which to do it.

Four partners were selected.

At the end of the project:

- 4, 904 Tapes
- 4, 968 Hours
- 8, 289 Archival Master Files
- +-40% of the total collection
- Average 90 hours of digitised output each working day.

At the conclusion of the 2014-2015 project 53.7% of the Library's total audio collection had now been digitised.

A similar project was undertaken in the 2015-16 financial year which has digitised a further 5130 hours. At the conclusion of this project, we have digitised 11,500 hours which is the around 90% of the total Oral History on tape collection. This is an achievement that we are very proud of but it was not all easy.

One of the many challenges facing the Library was how to ensure we were getting a quality product when there was so much material coming back, often at short notice. We had contractual obligations to our partners to accept the deliverables in a reasonable time frame to enable payment for their services.

I've often heard it said that, 'you spend as much time checking the work as you would just doing it in-house', but clearly in these projects, that was not an option. A clear, efficient and effective quality assurance framework would be required to ensure the deliverables of this project were fit for purpose.

4 Before We Begin – To Start Mass Digitisation

4.1 Quality Standards, Assurance and Control

To ensure a quality product, it is not just a matter of checking the quality of the outputs, but putting in place a framework that makes that quality achievable.

Quality Standards define the contractual obligations between the supplier (partner) and the customer (client).

Quality Assurance is the management of the quality aspects of the project. It involves setting expectations, assessing and mitigating risks and identifying and implementing the systems to allow those expectations to be met.

Quality Control checks to assess that the expectations are being met. This includes compliance and quality checks.

4.2 Quality Standards

A clear prerequisite for quality deliverables is to define what a quality deliverable is. This was documented in the contract as the Statement of Requirements. This is where we clearly set out the expectations of the vendors. The Statement of Requirements must;

- Support the objectives of the project
- Set the technical standards and specifications for the deliverables
- Set the technical standards and specifications for the equipment used
- Set the metadata standards and specifications
- Be clear and comprehensive
- Provide guidance where there is ambiguity
- Be suitable as an assessment benchmark

Our technical standards are based on international standards where possible, including *Guidelines on the Production and Preservation of Digital Audio Objects*, ed. by Kevin Bradley. Second edition 2009, IASA-TC 04.

4.3 Quality Assurance Strategy

Taking previous experiences into account, we set out to establish a documented QA Strategy. This strategy identified and focussed on areas of the highest risk of quality issues. The areas identified were:

- Communication
- Technical Infrastructure
- Operational Procedures and
- Quality Control Procedures both in-house and with the vendors

The QA Strategy highlights the need to establish clear expectations and to continually nurture a clear and open dialogue with our partners so we could mutually address any issues as they arose. I want to emphasise that it was key to the success of this program that we entered into more than a client, service provider relationship, but that we saw ourselves and the external digitisers as partners in the success of the program.

Another key principle applied in order to manage the scale of this project was to apply a 'trust level' to each of our digitisation partners. When dealing with the volume of digitisation that we were, we expected mistakes to be made but we also expected the majority of mistakes should be picked up before we saw them. Quality Assurance was a key requirement for our partners. Our in-house checks were to be a safety net.

The first step into establishing this trust level was an on-site inspection and assessment of the technical infrastructure and procedures.

We would start by checking 17-20% of the deliverables but as our confidence in their abilities grew, that rate could be lowered to 5-10%. This trust level is under constant review and is not assessed on a clear number of definable attributes, but drawn from my experience and knowledge to recognise consistently good work which demonstrates that our digitisation partners had adequate operational and quality assurance procedures in place.

The QA Strategy also identifies how we would select the items to be checked in the sample set. This would be based on an assessment of the digitiser's comments, issues that were identified in initial compliance checks, which may include specific non-compliance, or might just demonstrate a lack of attention to detail where the data supplied did not match the actual file, even if the file was still compliant. An example might be a spreadsheet entry of a sample rate that differed from the file. This would show a lack of attention to detail and we might look at that particular operators work more closely. We would check at least one file

from each collection of interviews and once we had selected files based on these criteria, and random selection would be chosen to fill any shortfall in the number identified as per the sampling rate.

The experiences learned in the Pilot Program informed significant changes in the Statement of Requirements and development of the QA Strategy as has the experience of concluding the first mass digitisation project informed changes to the current project. To maintain and improve quality, it is essential to be open-minded, to review, learn and constantly re-evaluate from each experience.

5 The Thick Of It – Checking The Deliverables

Once in full flight, the project was creating an average of 90hours of digitised material each working day.

We developed processes, tools and procedures to implement the SLNSW QA Strategy. In pushing through large volumes of material, at short notice, our obvious first question was:

'What can computers do for us?'

We needed to identify the systems and infrastructure required to effectively and efficiently assess the quality of the deliverables. There was no one-stop solution that would cover all aspects. A number of tools were sourced, or developed in-house.

- Semi-automated checks
    - In-house developed spreadsheet tools for checking and tracking compliance
- Integrated solutions
    - Dobbin for systematic audio quality analysis
- Single Function Applications
    - BWF MetaEdit to extract technical metadata

Then there was the next obvious question:

'What can't computers do for us?'

We still require people to make human choices, to look, listen and decide. The sample sets would dictate what was checked manually and compared to the original tape item. This selection is guided by automated checks but a human would assess an auto-detected problem as a true or false positive. For example, we received some files back which Dobbin had detected a hum on. The digitiser had also noted this hum as being present on the original tape. However, in a manual checks against the original tapes, there was no hum present. In discussion with the vendor, we identified that the issue was a technical issue in the cassette player used for digitisation.

For these projects, we engage an additional team member to assist in manual quality checks.

6 The Process

Each program will have its own objectives, resources and expectations so I don't intend to go into detail about the processes specific to our project but I will provide a brief overview.

- The completed wave files would be delivered to us on portable hard drives, along with spreadsheets, which we had originally supplied, completed by the vendor.
- Following a virus check, a content listing of the drive would be made
- The files are copied to a network location with checksum verification and to an internal drive on the Dobbin workstation.
- BWF MetaEdit is used to both check and extract core and technical metadata from the Wave files and to verify embedded MD5 checksum data. We require metadata about the digitisation

process to be embedded however we do not require external partners to embed any descriptive metadata.

- A in-house developed spreadsheet tool would import the spreadsheet data provided by the digitiser and compare it with the BWF MetaEdit data.
    - Check for matching filenames
    - Check Sample Rate, Word Length and other technical compliance and compare against vendor supplied data
    - Compare durations against vendor supplied data
    - Highlight discrepancies
- This process might highlight a missing file, but there may be an explanation provided, for example, a blank tape.
- At this stage, simple errors may be manually corrected, such as minor file-naming errors.
- All files are analysed through Dobbins Audio File Inspector and certain detections will sort files into pass and check folders.
- Initial results are checked using Dobbin Results Viewer interface to eliminate obvious false positives. For example, clicks that are clearly acoustic in nature.
- This information is used to mark specific files for checking against the original items and in some cases, specific disturbances are highlighted for checking. E.g. click at 0:30:22
- Where possible, originals are compared to digitised files to assess quality of reproduction and correct content, including missing content.
- These checks may highlight issues which require the scope of manual checks to increase
- The beginning and end of each side are checked to ensure the entirety of the recording has been digitised
- The overall quality is assessed
- Digitisers comments are verified
- Results are entered into a spreadsheet and identified issues are addressed with the vendor

7 Things can and do go wrong – providing an opportunity to learn and refine.

One of the key lessons we learnt early in the project was to plan for the scope of manual checks to increase. We highlighted a number of issues through random sampling. Once an issue had been identified, we had to assess the scope of the issue. In some cases, this meant sampling rates would go from 18% up to 40%.

We have to establish if the each issue is isolated or systematic. Is it a one-off poor transfer, or is there something procedurally or technically wrong. At some stage we have to decide if will we check every file that was done in a particular studio, or would it be more efficient just to re-digitise everything done in that studio or a third alternative, could we with live with the issue, especially taking any further risk to the original item into account. This is where an open and honest dialogue with the vendor really matters. To resolve issues requires concentrating on the resolution, not pointing fingers. We knew that this volume of digitisation was inviting mistakes. Our main aim is to have the same mistake made twice.

We also accept that in some cases our data was not always accurate or complete which lead to confusion and we found a few examples where files didn't meet our expectations, but those expectations had not been clearly defined or foreseen. Once again, we discussed these issues as they arose and have implemented these lessons in the review of requirements.

8 Conclusions:

We made a few mistakes along the way but nothing so bad as to derail the project. We didn't digitise everything we sought to but the issues were about the condition of the original items. The items had not had detailed condition assessments made and throughput of this project did not allow time for intensive conservation treatments.

Feedback both internally and from our external partners suggests that maybe we were pushing the envelope a bit too hard in earlier projects but what we have demonstrated to ourselves and others is that you can use out-sourcing model to achieve great volumes of digitised material if you have good teamwork in place. Things like a strong Quality Assurance framework, project management, external digitisers who partner and other logistical matters all played a key part in the finely tuned environment which supports mass digitisation within tight timeframes. I also feel that this project has not only delivered to us, but it has delivered a great deal of experience and development of archival audio skills in the industry. Through partnering with our digitisers, we have helped to develop industry capacity which I firmly believe will form an integral aspect of the sustainability of archival audio collections. At times we have had four companies running multiple studios in multiple shifts. This has helped the local industry develop in a time where archival audio skills are getting scarcer, we have helped deliver a significant number of newly skilled operators to the industry. It should also be noted that music studios are doing it tough in the day of bedroom producers. Institutions need digitisers and studios need work

To point out the elephant in the room, we at the State Library of NSW have been fortunate to receive the funding to allow us to do this project among many other digitisation projects. We are grateful to the NSW Government for their continued interest and support for the Digital Excellence Program.