

Unpublished paper prepared by Ross Spencer, prior to the Born Digital and Cultural Heritage Conference in Melbourne. Presented, Friday 20 June 2014

Audio: <http://playitagainproject.org/wp-content/uploads/Panel-F-Ross-Spencer.mp3>

Fly Away Home: Pilot Transfer of Born-digital Records at Archives New Zealand

*Ross Spencer, Aaron Braden, Jan Hutař, Mick Crouch,
Sean Mosely; and Talei Masters*

The end-to-end transfer of born digital records is a challenge many national archives are facing; from appraisal through to the digital preservation and access of records. Archives New Zealand established a pilot transfer work package in 2011 to understand the processes and structures of digital transfers as well as the resourcing implications for agencies and the organisation itself. The package hoped to concentrate on a small number of transfer sets from a handful of pilot agencies. This would result in a profile of skills, tools, file formats and technologies required for different types of accession as well as an understanding of time frames and risks involved with agencies and documentation for the entire breadth of a digital transfer. While the original vision has remained consistent, the practicalities of the organisation's statutory mandate meant that controlled transfers gave way, in-part, to substantive transfers with greater priority and importance, such as those of the recent Royal Commissions conducted by the New Zealand Government. This paper provides a case study exploring the ambitions of an ideal transfer process; the stark reality of transfer outside of a controlled work package; the relationship between Archives New Zealand and the various agencies involved in the transfers, as well as reflecting on its own infrastructural positioning. The paper will also look at lower level technical challenges of dealing with real data, real volumes, and the overarching issue of the digital preservation of that material. The paper will summarise Archives New Zealand's progress in this area and its future targets.

Keywords: Archives, case study, transfer, appraisal, assessment, analysis, digital preservation, government, inter-agency cooperation

In 2011, Archives New Zealand established a package of work that would take transfer sets of records from selected government agencies in order to better understand the processes and structures of digital transfers as well as the resourcing implications for agencies, and Archives New Zealand. *The deliverables*: Pilot Transfer Project Plan; transfers from 3-4 government agencies; test sets of records from government agencies; and a final report to feed into requirements document for a larger programme of work, the Government Digital Archive Project (GDAP).

It is now 2014 and GDAP succeeded in establishing the Government Digital Archive. Providing access to digitized material it contains 73 thousand 'intellectual entities' - digitized records, consisting of one million 850 thousand individual files. The programme closed before the archive was extended to support the archiving of born digital records. At the time of writing, Archives New Zealand is yet to ingest a born digital transfer into its long-term preservation system.

Within the paper we will look at how the project evolved within a short period of time, discussing various challenges and conflicts of interest which saw expectations and requirements change. Resisting the urge to be entirely introspective, the paper will place a greater importance on the discussion of how, out of a closed programme of work, the digital continuity team at Archives New Zealand have come full circle to attempt, once again, a small number of born-digital transfers. Each successful ingest will provide lessons to be fed into the improvement of the next; further still, it will help us to develop a longer-term understanding of the future, and how we want to handle born-digital records.

From Pilots to Reality

A digital archive stores digital records and provides mechanisms to make them available for consumption by the public. In the context of Archives New Zealand and the programme of work that was the Government Digital Archive Programme, these are born-digital records from government. Capability to manage born-digital didn't exist at Archives New Zealand in 2011 and so the pilot sought out sets of records with archival value that were small, manageable, and importantly, unrestricted. Authentic transfers that would provide the organisation with the background needed to understand the descriptive process behind born-digital transfer e.g. testing an item based model, developed to partially automate, and improve the quality of description for born-digital records; or selection of archival metadata; this would help to reveal gaps before tackling larger, more complex collections.

Other outputs would be an export metadata schema; transfer 'skill profiles'; understanding of tools required by agencies; understanding of formats used by a set of government agencies; understanding of the timeframes involved in digital transfers; documentation of technical specifications and requirements; processes and procedural requirements, and documentation of the costs of transfers for Archives New Zealand.

Agencies identified in the work package were:

- Archives New Zealand
- Private Office of the Prime Minister
- Treasury
- Commerce Commission
- A yet to be decided site that had a 'SharePoint' installation¹

From each we could collect digital transfers with a diverse range of characteristics to test any assumptions we had to guide the development of capability moving forward.

The depth and number of outputs expected from them is reflected in the number of transfers Archives New Zealand attempted to take on.

It is with hindsight that we can look at the expectations of this list and understand that the number of risks and issues that could arise with even a well-chosen list of controlled transfers is numerous. Control was the operative word – together it was thought that these collections represent small, manageable, *unrestricted* transfers that we felt we had the greatest chance of helping to keep the project low-risk.

Also in hindsight, we can look at the goals of the pilot transfers, specifically 'Test sets of records from government agencies', and consider how sustainable the original position was likely to be in the public sector where digital was already on its way.

A risk identified at the time was illuminating, regardless of insight: 'Scope of this project could exceed the current resources in the programme team'. It can be understood now that the amount of effort

¹ SharePoint is a web application platform commonly leveraged as an Electronic Document and Records Management System (EDRMS)

for a single transfer, let alone five separate transfers would indeed have pushed the resources available to Archives New Zealand.

Initial Transfers Project

Within a year the transfers proposed to be taken by Archives New Zealand had become:

- The Pike River Coal Mine Tragedy Royal Commission²
- Canterbury Earthquakes Royal Commission³
- Office of the Prime Minister
- Human Rights Commission
- Commerce Commission
- Waitemata District Health Board

The profile of these transfers was radically different from being small, manageable, and unrestricted - a single extract from the Pike River Royal Commission representing, *372,228 files* and *66,398 folders* at 209 gigabytes – the majority of it, restricted material⁴.

Working on such challenging accessions, even with some sitting in the custody of the same parent department provided Archives New Zealand with powerful lessons that we would take beyond the life of the Government Digital Archive Program.

Lessons Learned

A summary, of the lessons that we are most equipped to speak on at the closure of the GDAP project, and as new work begins is as follows:

Technical Implication – The work package was focused on a descriptive transfer process and creating the documentation surrounding that. There was no elaboration on the technical side of a digital

² Pike River Royal Commission website: <http://pikeriver.royalcommission.govt.nz/>

³ Canterbury Royal Commission website: <http://canterbury.royalcommission.govt.nz/>

⁴ Definition of Restricted Material in the Archives New Zealand, Advisory Notice A6; 'Making Access Decisions Under the Public Records Act': http://archives.govt.nz/sites/default/files/a6_2.pdf

transfer. As GDAP ended this had little impact, but it is clear a technical thread must sit throughout a digital transfer from assessment through to ingest.

Numbers and Complexity – The smallest collection we were dealing with most actively was 11,504 objects sitting at approximately 5GB in size. The largest was over 350,000 objects, approximately 200GB in size. Managing such an amount of data for ingest into a digital repository brings with it many different issues.

Working with Agencies – In some regards we were fortunate to be working on two sets of records held within the same parent department of Archives New Zealand, The Department of Internal Affairs (DIA). Specifically, we would be dealing with the DIA Records team and the Technology Services and Solutions team. The relationship brought with it many lessons that require serious consideration about how we might work with external agencies moving forward.

Digitisation a Distraction – The Digital Continuity department works across the organisation and one of its responsibilities at present is providing advice about digitisation and handling the ingest of digitised material. While we approach the digitisation of records with a strong digital preservation angle, the work takes two full time staff away from the focus of the ingest of born-digital material and therein lies the trade-off.

Engaging with Business – Transfer of knowledge about born-digital transfer across the organisation is paramount, but engaging with the business from the beginning of an information gathering process might not be the correct approach as it means that learning, and mistakes made are felt organisation-wide. A smaller specialist team, capable of ironing out issues and *then* sharing the knowledge throughout the business can potentially have a far greater impact on the support, and the desire of an organisation for developing new capability.

We will continue in the remainder of this paper to provide a more detailed description of what we found.

Technical Implications

The original work package sought different technical challenges as evidenced by goals such as the desire to work with a 'SharePoint' installation. And indeed, initial phases of the project achieved a lot

technically, including the delivery of a 'transfer metadata schema'; validation mechanisms for the schema; and a mechanism for ingest of digitised material of limited file formats.

It is clear though that the package was focused on surveying the landscape - creating business processes and documentation - but without a complete view of the technical implications of a digital transfer.

But understanding this before the fact would also have taken significant analysis in itself and would have required a thorough understanding of digital archiving and digital preservation issues – products of experience. This is a theme that this paper will seek to revisit. Can we find a way to deliver evidence to inform future decision making that does not require every permutation of a transfer to have been managed up front?

One must consider that in many scenarios, any technical issues that might arise have the potential to grow explosively. For example, the research required identifying an unknown file in any one collection, from sourcing a specification to developing an identification mechanism and registration in a community database is an unknown quantity, and remains different for each distinct format one might find. If we decide against taking the troublesome files; what does that process look like to a government archive? And how long does it take to follow that process compared to the former scenario?

As we gain experience post-GDAP, we know that we simply couldn't have guessed some of the potential technical issues; the identification of zero-byte files in an accession, for example; and what action we need to take with them. Such files should probably not have been maintained in the original record keeping system, and point to the need for a further technical assessment, pre- or post- traditional archival appraisal. The temporal placement of such analysis in the process is brought into question due to the apparent complexity of assessing digital records. Further discussion of this matter is far beyond the scope of this paper.

Applying a paper-paradigm to digital transfer would be the ideal world scenario - all objects would fundamentally look the same and behave the same way, but it isn't realistic. Digital comes in many shapes and sizes, each requiring subtly different treatments. Because of this we learned that it is preferable to see technical analysis happen throughout each stage of the digital transfer process. Technical analysis when making disposal decisions through to technical analysis on ingest.

Numbers and Complexity

The majority of our experience, understanding the *potential* complexity of digital collections, comes from two Royal Commissions transfers – The Pike River Coal Mine Tragedy and The Canterbury Earthquakes Royal Commission. It is here that we have gained most insight into the potential volumes and counts of material we might eventually have to transfer.

The two collections were *11,500+* objects and *350,000+* objects sitting at over *5GB* and *200GB* in size, respectively.

In more detail, the Canterbury Earthquakes Royal Commission files were exported from a Lotus Notes Document Management System⁵. The export at the time was incomplete, but the following represents the statistics of the objects we were working with:

Total files: 11805

Size: 5414 MiB/MB (Megabytes)

Total container objects: 13

Total files in containers: 301

Total unidentified files: 123

Percentage of collection identified: 99.0

Percentage of collection unidentified: 1.0

Total extension ID only count: 24

Zero byte objects in collection: 2

The total unique file formats found in the Canterbury Earthquake Royal Commission content management system was **55**.

The Royal Commission into The Pike River Coal Mine Tragedy was organised into *two* separate document management systems which in itself would create its own challenges. The smaller document management system was another instance of the Lotus Notes Document Management System. The statistics for the file export from that system are as follows:

⁵ IBM website for Lotus Notes DMS: <http://www-01.ibm.com/software/lotus/products/dominodocumentmgr/>

Total files: 24190
Size: 5139 MiB/MB (Megabytes)
Total container objects: 121
Total files in containers: 2390
Total unidentified files: 1254
Percentage of collection identified: 94.8
Percentage of collection unidentified: 5.2
Total extension ID only count: 900
Zero byte objects in collection: 0

And the second document management system an instance of AccessData's Summation⁶:

Total files: 374264
Size: 209337 MiB/MB (Megabytes)
Total container objects: 68
Total files in containers: 2036
Total unidentified files (extension and blank): 3892
Percentage of collection identified: 99.0
Percentage of collection unidentified: 1.0
Total extension ID only count: 312
Zero byte objects in collection: 19

The total unique file formats discovered across the two Pike River content management systems is **105**.

We can see that even in the smallest collection, with 99% of the collection identified. We still have 123 unidentified files. From additional identification statistics we know that 23 of these unidentified files are likely to be plain text files based on extension only identification. One hundred other objects will require a lot more work with the tool used to identify file format giving very few additional clues as to what they are likely to be. In our largest collection we are looking at 3892 files with an entirely different file format profile again.

⁶ AccessData's Summation Home Page: <http://www.accessdata.com/products/ediscovery-litigation-support/summation>

Identification and understanding of what exists in a digital collection is the single most important part of a digital transfer. It tells us how to manage the objects, it tells us whether we have records or non-records, and informs us whether we need to treat these objects up-front, or whether we are clear to put them into a digital archive as-is. As mentioned, the work required to do this for unidentified objects is an unknown quantity. The implication for a small team of identifying 3892 unknown objects is quite incredible, although the work does become more manageable as we break the work into pockets of similar objects.

The trade-offs encountered with this amount of work is the most unfortunate consequence. As a team works tirelessly to push these objects through to a digital repository it might result in being unable to record information in order to feed knowledge back into the digital preservation community whose tools we are often relying on in the first place - for example, by contributing to file format definitions and signatures. Expediting ingest can be done by taking advantage of speedier approaches to format identification such as file extension matching. Such a technique can provide a rough identification that we can record in the repository but this will come at the cost of truly understanding the digital object which is at odds with the most basic digital preservation principles.

With a copy of the Royal Commissions digital material temporarily in Archive New Zealand's possession we were able to gather intelligence that might prove useful in future transfers work. Study of the objects in the commissions was done on a completely scientific basis for the purposes of transfer, with absolute respect to the private and restricted nature of them both.

The volume of objects we had in the collection lent itself to understanding how we might approach the analysis of digital objects in large collections when we do eventually accession them. Our primary goal became to fill the gaps left by, DROID⁷ which provided the details on format identification discussed previously. Using tools we could easily combine using shell scripting techniques in Linux such as Tika⁸, Jhove⁹, and File¹⁰. We tried to collect as much additional data about the objects as possible. The dataset would become a triage dataset that would help us to then conduct further

⁷ The National Archives home page on DROID: <http://www.nationalarchives.gov.uk/information-management/our-services/dc-file-profiling-tool.htm>

⁸ Apache Tika home page: <http://tika.apache.org/>

⁹ Jhove on SourceForge.net: <http://sourceforge.net/projects/jhove/>

¹⁰ Documentation on Linux File command: <http://linux.die.net/man/1/file>

analyses on individual objects as required. A preliminary investigation of the potential approach we would take is outlined by Ross Spencer's blog on the Open Planets Foundation website¹¹.

With initial attempts to run the analysis over the network to our storage hosting failing inconsistently, and without pattern we placed the objects temporarily and securely on local networked storage. We saw immediate improvements in the time taken to run processes, and on completion we had statistics on how long the entire analysis took to complete, generated using 'time'¹²:

```
real: 2949m21.680s
```

```
user: 22m29.436s
```

```
sys: 39m35.964s
```

This equates to approximately **49 hours** to complete the analysis across the entire Pike River Collection.

A further analysis running just the Tika metadata extraction analysis engine on its own generated the following statistics:

```
real: 6071m11.579s
```

```
user: 11644m41.133s
```

```
sys: 324m28.397s
```

Approximately **101 hours**.

The timings can be improved on by using more programmatic techniques, but our approach was developed in short order and represented a low barrier of entry for us and potentially for other memory institutions wrestling with the analysis of similar collections on using commodity hardware and software.

If it works, and within reason, the work we do analysing digital objects is not time-dependent. In this instance we are creating a triage dataset over 49 hours, and that's before we begin to analyse it

¹¹ A Nailgun for the Digital Preservation Toolkit: <http://www.openplanetsfoundation.org/blogs/2014-02-24-nailgun-digital-preservation-toolkit>

¹² Documentation on Linux Time command: <http://linux.die.net/man/1/time>

further, and before we even start work understanding individual objects based on that. For example, should we need to run the analysis again on a collection this size, to take advantage of any performance improvements in the software stack used, it would take another two days. At that point, we are beginning to push the limits of what can be considered a practical exercise or not.

Once the triage work on the Royal Commissions had been put aside to focus on post-GDAP efforts we found an even greater need to report on what we had previously found. In this new period for Archives New Zealand, months after the original analyses were documented we needed once again to report on the collections. Reports written by different individuals, with different writing styles, using different language, and potentially asking different questions of each collection do not provide a suitable basis for comparison. Further, reports derived by hand from computer generated listings are prone to error.

Understanding the difficulties in communicating not just the findings of different collections, but how one might report on a single collection; consistently, and accurately, led to the creation of a tool that can query any export from the DROID file format identification tool and output the results in a clear and repeatable manner in plain text. The tool is extensible, and open-source, and is discussed in more detail on the Open Planets Website¹³. One of the biggest hopes for the tool is to help standardise the language used in the analysis of digital collections as its use becomes more widespread. The summary results referenced in this paper have been generated by this tool.

The lessons learned during this experience are:

- We now know that we are likely to receive accessions *at least* as large as this and we need to be able to deal with that.
- Tools are needed to support the entire process, and those tools need optimizing.
- There are many more potential analyses of digital objects not discussed here for brevity's sake.
- We need to make the most of what we have while we have the ability to conduct research.

¹³ An Analysis Engine for the DROID CSV Export: <http://www.openplanetsfoundation.org/blogs/2014-06-03-analysis-engine-droid-csv-export>

- At this point we are still only talking about further technical assessment and analysis to aid ingest; we still haven't completely understood the needs to 'digitally preserve' these objects, further, how we generate and make available access copies of the material.
- This work is being done outside of the digital preservation system. We do not yet believe we have a system suitably equipped to address the number of issues we are facing. It is believed, however, that this work will inform the improvement of those services over time.
- It is not always possible to throw computing resources at a problem - our instinct is to see a greater emphasis placed on the optimization of individual tools, perhaps even re-writing them to be tools native to particular environments to remove overheads using code compiled down to platform-specific machine code at run-time.

Working with Agencies

Progression on GDAP saw documentary requirements with their own technical implications put in place for the accessions we were working on. An example requirement might be the population of an XML listing conformant with an export metadata schema.

Before an agency is able to do this, it first has to be able to collate the records, ensuring their integrity in the process. Much of the technical work on the accessions up to the point of closure was finding understanding how to complete an export from a particular system and accompanying metadata.

As technical challenges arose, compliance with our own technical requirements became more difficult. Testing had to be done iteratively but with large exports occurring overnight, and testing re-completed the next day. The Royal Commission accessions are in the custody of the Department of Internal Affairs (DIA) Records Team. This means that Archives New Zealand and DIA share the same IT vendor. The relationship for the purpose of the transfers alone was not as synergistic as one might hope.

With diverse projects and responsibilities across a large government department it simply wasn't feasible for our shared vendor to dedicate appropriate full time equivalents. Dedicated focus and communication could only happen when other responsibilities freed up to allow us that time.

Further, developers provided by shared IT vendors are excellent generalists, but not digital archiving specialists. This concoction led to challenges seeing the simplest requirements pushed over the line, before we even tackle further technical challenges discussed before. We saw signs that there are issues we will still have to tackle when the work of digital transfer moves forward with agencies that do not sit in such proximity to home, or have even lower resourced vendors.

But overall it was *our* expectations that were challenged most working toward the transfer of the Royal Commissions. Even giving an agency the tools to do things like schema validation we realised that as a standalone instrument without expert knowledge it would not be a suitable tool to use. A tool can help narrow down what to fix if it finds something wrong but it can't always say how to fix them. And a tool can't always pinpoint issues; bad checksums, inappropriate values in fields, or values in the wrong fields, for example.

Without expert knowledge, anyone would struggle to understand what to do to fix this type of issue. An exception of this form needs human input to resolve and in the majority of cases this will likely be both a digital archivist from the team working on the digital transfer at Archives New Zealand and the client's records manager. Requirements then need to be communicated to the vendor.

We were fortunate enough to have development capability in our IT teams, even if it wasn't dedicated to the transfers alone. With agencies sitting outside our department and without such resource we will face a different set of challenges again,

In future transfers we will see multiple different systems with different export capabilities, customisations of those systems lead once again to an even more varied landscape.

How we will interact with external agencies is still to be resolved. Better documentation of requirements and more robust tools will help. One approach might be to create a lower technological barrier with our requirements, as done by The National Archives, UK, when accepting CSV based metadata¹⁴. One approach might be to sit alongside records managers and IT departments for the duration of a transfer process to always be on hand. There are many issues to be resolved, but hopefully many different ways to approaching their solution.

¹⁴ CSV validator – a new digital preservation tool: <http://blog.nationalarchives.gov.uk/blog/csv-validator-new-digital-preservation-tool/>

Digitisation a Distraction

Through the course of GDAP, the Digital Continuity team would provide advice and guidance elsewhere. One of the biggest pieces of work they engage with is the different digitisation projects across the organisation.

The first phase of the GDAP project saw the development of ingest capability for digitised material. This supported the migration of already digitised material to the newly implemented Government Digital Archive from a temporary Fedora archive. It also supported ingest of newly digitised material with a limited range of file formats.

There were a number of tensions in this effort including the use of the digital preservation platform, Rosetta, as a file store. While digitisation is approached, more recently, with digital preservation best-practice in mind, the overhead of a digital preservation platform is additional validation and checking of objects that might not otherwise be necessary.

Other issues which arose, causing distraction included:

- A digital preservation system requires specialist support from team members equipped to support it.
- Poor quality assurance of early digitised material resulted in skills being pulled from the digital continuity team to focus on correcting issues of validity and inconsistency prior to ingest.
- A switch in digitisation image format part-way resulted in re-engineering of workflows, including deposit mechanism; file enrichment routines, and mode of delivery.
- These skills do not sit anywhere else in the organisation.
- Deadline based digitisation initiatives put department wide pressure on digital continuity staff to meet those deadlines disturbing work on born-digital transfer.
- Digitisation cannot be done independently of digital preservation principles, and so the creation and testing of digitisation profiles takes time and effort.
- Focus is pulled away from core digital preservation issues.

- As digitisation standards are both narrow and remain fairly static over time, we do not see the range of files we'd see by conducting research on the digital preservation of larger collections of heterogeneous material. As such, digital preservation capability evolves at a slower pace.

The summary therefore is that digitisation efforts will pull people away from other projects and that will likely be reflected in delivery times. On the flip-side, the work on digitisation informs us about the use of the digital repository and helps inform future digital preservation policy and decisions which will ultimately be reflected in the fully operational Government Digital Archive.

Digitisation can help improve tools and specifications, and see them implemented in workflows - the key, however, seems to be in finding the balance in maintaining digitisation efforts while promoting the continual development of born-digital capability.

Engaging with Business

Digitisation is just one example of where Digital Continuity would deal with the wider business, and further, introduce to it the capabilities of the Government Digital Archive. Working with the organisation on the initial transfers project was a different experience again.

The transition from more controlled digital transfers to less controlled, more complex transfers gave the entire business a lot to think about. There would already have been a lot for the organisation to learn regardless of this change. From new models of describing digital objects, to the process of extracting material, and understanding the content of collection - the business was building born-digital transfer capability from the ground up.

On one of the non-Royal Commission transfers, a gap existed in the understanding of how files would be taken from a shared drive and delivered to Archives New Zealand. While New Zealand government has guidance which can help with a very small part of that, the fact is that the experience didn't exist within Archives New Zealand and so had to be developed from within the Digital Continuity team by the Digital Preservation Analysts. It was a small piece of research, but research nonetheless, involving understanding, in *hindsight*, of the following:

- File transfer
- File system metadata
- Preservation of file system metadata
- Cross file-system transfer
- Write-blocking equipment
- Encryption standards
- Government approved encryption standards
- Encryption and storage techniques
- Media sanitization

The research process was non-linear, that-is, we can begin by saying that ‘we need a portable hard drive’ but then understanding of other requirements such as media sanitization and encryption generates more research, and importantly, following research, guidance that can be used on future transfers.

Demonstrating complexity to the business and managing expectations is a difficult task. The reasons for this might be:

- Potential lack of technical understanding
- Different teams with different milestones and different targets
- Other teams involved in the transfers may be working more closely with external clients

Working more closely with external clients is an interesting example where we see that lessons learned, and more importantly, mistakes made, aren’t just transferred from one team to another, but ripple through to external parties too, where the effects of that will be felt, good or bad.

The entire business needs to develop a methodology for approaching born digital-transfer but to take it on a roller coaster ride while that methodology is developed, where many different approaches are taken, and many mistakes are made along the way, is not a desirable approach.

The main point we took away, as a team, is that rougher lessons should be ironed out first. We need to find *our* understanding and spread that understanding organisation-wide. When we look at the work that we are doing now we will demonstrate that we are taking a different approach to engaging

with the business - learning and development now happens inside the team and is then communicated outwards. The hope is that it creates a more appropriate learning environment communicating concrete lessons and displaying fewer mistakes. The mistakes a recipe for potential confusion down the line. The approach is not to be at the cost of losing evaluation by colleagues outside of the team, as It is recognised that different perspectives and the testing of one's assumptions by others is key to the success of any initiative. We cannot simply work inside a bubble, or give the business any unexpected surprises. Our principle is to simply share, but not to *suffocate*.

E-Accessions

A few years prior to the Initial Transfers Project, Archives New Zealand became custodians of a number of small accessions, primarily sets of ministers' papers. The accessions have been maintained on a network storage device up until now. The ultimate aim was for the E-Accessions to be transferred into the Government Digital Archive on successful completion of the project.

On closure of the project it is now the goal of the Digital Continuity department to complete the transfer, ingest, and description of these files using the infrastructure that exists at present. This will ensure the objects are better managed, and better monitored for digital preservation risks. It will also be possible to provide access to the records where previously they haven't been accessible to the public.

There are five 'E-Accessions' of which the department is currently focusing on two. The first two have been selected for the perceived simplicity, the greater potential to see them ingested into the digital repository without issue, and for the modest number of lessons that can help move our understanding of born-digital transfer one step further.

The accessions are 175 files, 63.1mb and 1295 files and 317mb respectively. Issues presented by the collections include duplicate objects, a handful of unidentified files, and small pockets of system files such as 'Thumbs.db' objects for previewing objects in Windows Explorer folders - the latter an interesting example of files that might not have been accessioned following further technical assessment at the beginning of the process.

Given the point at which the Government Digital Archive Programme closed, we were left with only a partial implementation of a proposed Item Model. Challenges with both accessions surround

describing the collections appropriately in the Archway Catalogue. A number of proposals were developed by the senior archivists within the Digital Continuity team and organisation wide discussions helped to select the appropriate compromise for these accessions. While we are making a descriptive compromise with these two accessions, the learning developed by ingesting them will be used as evidence to inform decisions further down the line, for example, if the catalogue cannot adequately display these collections, demonstration of this will help back up any case put forward to management and our information technology vendor for further catalogue development work.

All of the work we do emphasises the use of evidence to develop future capability. Work on two E-Accessions will help to create new processes and reveal gaps where we need to develop even further for the next eAccession, the one after that, and so forth. Through an evidence based approach we can put ourselves into a position to understand how a project needs to evolve without trying to take too much on at any one time.

Expected lessons of the first two E-Accessions will be around how we deal with duplicate content, or dispose of system files using the tools provided to us under the Public Records Act 2005. We also hope to have developed new format identification signatures to feed back into the digital preservation community. We will have pushed our systems, and processes forward for the next eAccession. We will use the outcomes of the next eAccession to move things further once again.

The eAccession work sees the Digital Continuity team work much more closely with one another with weekly meetings between all team members guiding the progress of the work. This is in stark contrast to the original transfer work where the entire team was only involved intermittently; coordination of the other teams in the organisation guided by the two senior archivists. Previously the team couldn't have a complete overview of technical challenges of any one transfer, however now, and promoting the idea of a technical thread running throughout an accession, the entire team has an overview and input at all times. The team is comprised differently than when the original Pilot Transfer Sets work package was formed. Now, the four, more technical members of the team can guide the senior archivists on issues that might arise, and likewise, the senior archivists can balance technical input with knowledge of archival theory and practice.

Learning is completed inside the team, and advice sought from the organisation when it is considered diligent to do so. Progress is demonstrated when there is something to demonstrate with

every attempt to have ironed out any potential issues. Ultimately the team is learning to walk before they can run and the process will be repeated with the organisation learning to walk so it can run.

The work on born-digital transfers has come full circle in some respects and now imitates the pilot transfers work package by using small manageable accessions to develop knowledge. The key differences are a technical thread throughout the process and a smaller team concentrating on a smaller number of accessions. Together, the team are taking a pragmatic approach to develop a pragmatic set of processes and a *capability* that will see the E-Accessions transferred into the Government Digital Archive – Archive New Zealand’s first born digital transfers.

Conclusion

Through this work we have learned:

- We can’t tackle every problem at the same time.
- The scope of the original work package was far too broad.
- We must focus on one problem at a time.
- We must focus on fewer accessions at a time.
- Iterative learning is paramount.
- An evidence based approach is required to make well-founded decisions.
- We must do the simple things well.

And out of it we also have a better idea about gaps and what we need:

- A plan for future royal commissions to accession to remove non-records and duplication.
- Tools to support transfer of born-digital material from government agencies.
- Optimized tools in the digital preservation toolkit.
- Better support for the technical assessment of records alongside archival appraisal.

Asked recently, “what about the influx of digital formats that we are expecting” – it is clear that we have an instinct about what we are likely to face in the future but the temptation previously has been to wait for a system to be developed to handle the ‘influx’, all the time putting the past on hold a little longer. The results could potentially be dangerous. We may be left with unmonitored,

unmanaged files sitting outside of a digital preservation system; files which cannot be delivered to the public easily.

Instead of focusing on the future we can build for the future by doing our housekeeping, looking at these older accessions and using them to build our understanding of what we might do with the what we *might* receive. We have material in our hands now; we can't predict the future, not just yet.

References

Greenan, M (2011). Government Digital Archive Programme Pilot Transfer Project: Work Package, version 2. *[Internal Document]*

Department of Internal Affairs. (2014) *Briefing to the Incoming Minister of Internal Affairs*. Retrieved from [http://www.identity-services.dia.govt.nz/pubforms.nsf/URL/BriefingInternalAffairs2014.pdf/\\$file/BriefingInternalAffairs2014.pdf](http://www.identity-services.dia.govt.nz/pubforms.nsf/URL/BriefingInternalAffairs2014.pdf/$file/BriefingInternalAffairs2014.pdf)

Department of Internal Affairs. (2014). Review of Active Archives and Government Digital Archive Programme. Retrieved from <https://web.archive.org/web/20140716061126/http://www.dia.govt.nz/Active-Archives-and-Government-Digital-Archive-Programme-Review>

Department of Internal Affairs. (2014). Review of Active Archives and Government Digital Archive Programme. Retrieved from [https://web.archive.org/web/20140807153005/http://www.dia.govt.nz/vwluResources/Active-Archives-and-GDAP-Review/\\$file/Active-Archives-and-GDAP-Review-April-2014.pdf](https://web.archive.org/web/20140807153005/http://www.dia.govt.nz/vwluResources/Active-Archives-and-GDAP-Review/$file/Active-Archives-and-GDAP-Review-April-2014.pdf)