KA MUA, KA MURI[1]—LOOKING BACK TO LOOK FORWARD: DIGITAL PRESERVATION AND ORAL HISTORY WORKFLOWS AT THE ALEXANDER TURNBULL LIBRARY, NATIONAL LIBRARY OF NEW ZEALAND

Valerie Love, Alexander Turnbull Library, National Library of New Zealand

1. Introduction

The Alexander Turnbull Library was founded in 1918 by collector and bibliophile, Alexander Horsburgh Turnbull (1868–1918). Alexander Turnbull had amassed a collection of over 55,000 manuscripts, books, photographs, drawings, paintings, and maps, which he maintained in his purpose-built private residence. At the time, it was the largest private collection in New Zealand. Turnbull bequeathed his collections to the government to create a reference library in the city of Wellington. His vision was that the contents should not be lent out, but kept together 'as the nucleus of a New Zealand National Collection.' Turnbull's house was purchased by the government, and in 1920, the Alexander Turnbull Library was opened to the public as a national reference and research library.[2] In 1965, the Alexander Turnbull Library officially became part of the National Library of New Zealand, though it continued to be located at Turnbull House (and various other sites around Wellington) before moving to a new National Library building in 1987.

In addition to the original collections owned by Alexander Turnbull himself, the Turnbull Library now holds the archives and special collections for the National Library of New Zealand, and actively collects materials representing New Zealand's unpublished documentary heritage and Māori worldview, as well as materials relating to the South Pacific Islands and Antarctica. Currently, the Turnbull Library holds millions of items, both analogue and born-digital. The Turnbull Library also collects websites and conducts harvests to capture all web content in the .nz domain.

2. Origins of a Digital Preservation Programme at The National Library of New Zealand

While digital materials have existed in the Turnbull Library's collections since the 1980s, it was in 2003, in conjunction with the passage of the National Library Act of New Zealand in Parliament, that the National Library began to formalise digital collecting and digital preservation policies. The National Library Act outlined the mandate of the National Library, and of the Alexander Turnbull Library within it, stating that the Library is responsible for ensuring that the documents and collections in its care are managed, preserved in perpetuity, and made accessible in a way that reflects their status as documentary heritage and taonga[3] of Aotearoa New Zealand. The Act also specified that a document can be in any form, including: "… information recorded or stored by means of any recording device, computer, or other electronic device … and material subsequently derived from information so recorded or stored."[4] This new legislation required the Library to develop a robust born-digital collecting and preservation programme to ensure that New Zealand's contemporary cultural herit-

1 'Ka mua, ka muri' is a Māori proverb which represents the idea of walking into the future (which we cannot yet see), while looking back at what has come before.
2 J. E. Traue. Dictionary of New Zealand Biography, vol 2, 1993. Accessible at: https://teara.govt.nz/en/biographies/2t53/turnbull-alexander-horsburgh [Accessed 8 January 2018].
3 The word "Taonga" is the Māori concept of something to be treasured.
4 National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act 2003. Accessible at: http://www.legislation.govt.nz/act/public/2003/0019/latest/whole.html#DLM191962 [Accessed 8 January 2018].

age—including Māori taonga and indigenous cultural knowledge—would be cared for and preserved for the future.

In 2004, the National Library began a project to develop a digital preservation and object management system. In addition to providing long term storage and access to born-digital collections, it also had to support all of the digitisation work streams in Turnbull Library, including sound conservation and imaging services. It also needed to support the digital holdings of the National Library's published collections, including ebooks, journals, and music received via legal deposit.

The Library successfully applied for additional government funding for a project to establish a National Digital Heritage Archive (NDHA) programme. As part of the project, the Library formed a development partnership with the software company Ex Libris to build a full digital archive and preservation management system. As Jay Gattuso writes:

> During the transition from a fledgling project, to a fully featured digital preservation programme, the NDHA team needed to explore what it meant to "do" digital preservation as a day-to-day activity inside a national library. The products of this exploration can be found in much of the core functionality found inside Rosetta. Much of this exploratory thinking was deeply speculative while also leveraging heavily on prior work, with direction, suggestions, and hints coming from works that preceded the project. As the project team started to focus on the requirements of a preservation "system", there was very little established knowledge to draw from. This led the project team to speculate on what the future would look like, addressing such thorny issues as "what will we know about formats, and their associated risks?", "how will we know that we have content with an associated technical risk?", "how will we undertake risk mitigation activities on file-like objects?", "how will we process items and workflow tasks", "who is the intended audience of this work?" and so on.[5]

The resulting Rosetta[6] digital asset management and preservation system was launched in October 2008, based on the Open Archival Information System (OAIS) reference model and Metadata Encoding and Transmission Standard (METS)[7]. Rosetta offered functionality for ingest, preservation action, recording descriptive and administrative metadata, arrangement and structure maps for digital files, and the creation of access copies/derivatives.

The actual digital objects being preserved, known as Intellectual Entities (IEs), might be a single item such as a digital photograph. Or, they could be complex items comprising multiple files, such as separate tracks that together form the sound recording for an oral history interview, or scans of individual pages from a soldier's diary written during World War I.

5 Gattuso, Jay. Preservation Actions: where we started and where to do we go from here. Alexandria: The Journal of National and International Library and Information Issues. SAGE Journals. 6 September 2017. Accessible at: http://journals.sagepub.com/doi/pdf/10.1177/0955749017725437 [Accessed 8 January 2018].

6 Currently the Rosetta application is used by libraries and archival institutions around the world. See http://www.exlibrisgroup.com/category/RosettaOverview.

7 Library of Congress. METS: An Overview and Tutorial. See https://www.loc.gov/standards/mets/METSOverview.v2.html.

3. The National Digital Heritage Archive (NDHA) today

In the nearly ten years since adopting Rosetta, the National Library has developed a robust set of processes, protocols, and documentation for working with born-digital collections. Many of these guides are available on the Library's Digital Preservation Programme website at https://digitalpreservation.natlib.govt.nz.

Since 2010, there has been a marked increase in the number of individual files within an IE, reflecting the growing complexity of these digital objects, both digitised and born-digital.[8] As of 31 July 2017, the National Library's digital collections comprise 1.6 million intellectual entities; 14.4 million master files; and 30,321 web harvest entities containing over 231 million individual files.

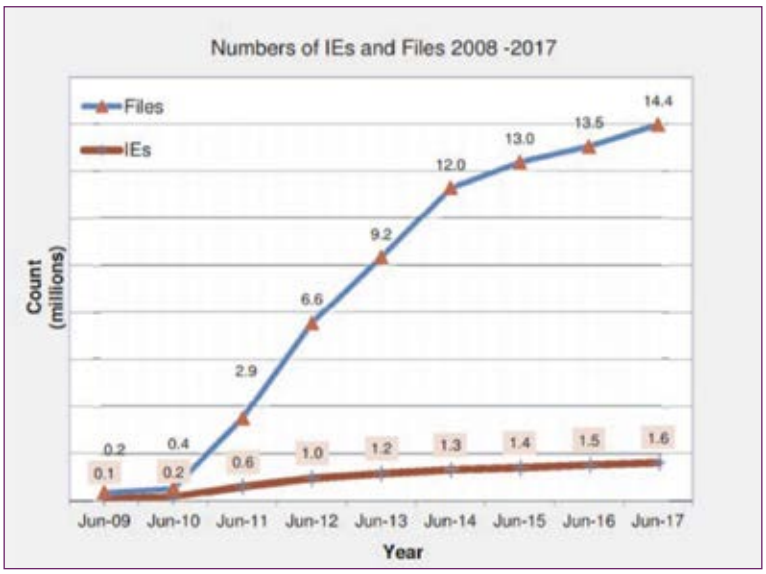

Figure 1. Number of IEs and Files, 2008–2017.

The digital collections in the NDHA constitute 204 Terabytes in total, a figure which has also been increasing at a growing rate over the years. 190 different file formats are represented in the collection, though the vast majority are standard formats such as tiff (7 million); xml (4 million); jp2 (2 million); .txt (319K); .pdf (238K); and .doc files (201K). Video and sound files are a much smaller subset of current holdings, with 29,000 .flac files; 26,000 .wav files; 17,000 .mp3 files; and video file formats comprising fewer than 1,000 intellectual entities.

8 Securing the future: Digital Preservation at the National Library of New Zealand, https://digitalpreservation.natlib.govt.nz/assets/NDHA/Home/NDHA-Booklet-end-Jul17.pdf [Accessed 8 January 2018].

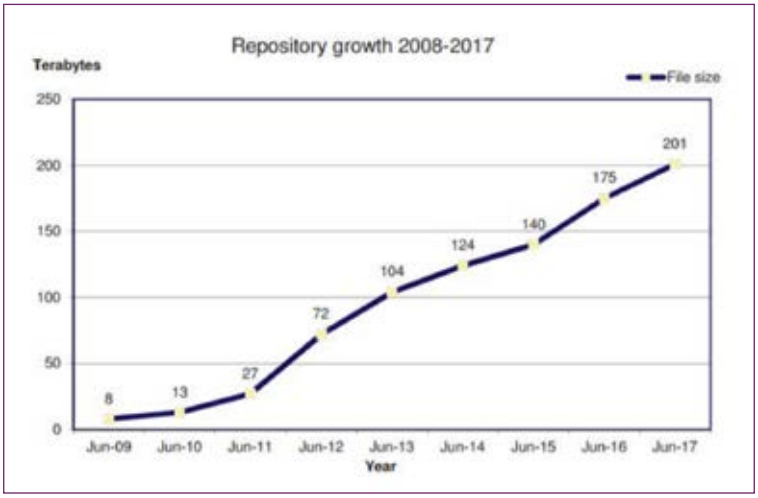

Figure 2. Repository growth 2008–2017.

In 2016, Turnbull Library launched a new collection management system for unpublished and archival materials, and the National Library developed a new set of strategic directions. The new strategic directions focused on three broad areas—reading, knowledge, and taonga. The taonga strategic direction included a vision that "Born-digital content reflecting contemporary New Zealand life and knowledge will be readily available for access and research." Like most collecting institutions, Turnbull Library has a backlog of unprocessed collections, including born-digital. As such, the Library began to re-evaluate its born-digital workflows in order to try to remove barriers to access in light of the new Library priorities.

4. Oral History Collections at the Turnbull Library

The Turnbull Library is the official repository for oral history projects that receive funding from Manatū Taonga, the Ministry for Culture and Heritage. Since 1990, Manatū Taonga has provided funding to nearly 400 community groups, individuals, and historians to gather and preserve life stories through oral history.[9] The Turnbull Library also commissions its own oral history projects to document contemporary events and politics. Over the past 10–15 years, oral history collections at Turnbull Library have transitioned from entirely analogue to almost entirely born-digital. However, many of the processing workflows have stayed relatively the same.

Turnbull Library currently holds close to 1,500 oral history collections on a range of topics—everything from interviews with Māori elders, to the history of ice cream making in New Zealand, to stories of sex workers, and more. The oral history collections comprise over 7,000 individual interviews, and approximately 16,000 audio recordings. Turnbull Library also generates oral history diary projects, which follow an interviewee over a set period of months or years. These collections hold the voices and experiences of New Zealand life that history might not otherwise remember.

9 See http://www.mch.govt.nz/funding-nz-culture/ministry-grants-awards/new-zealand-oral-history-awards [Accessed 8 January 2018].

Because the bulk of Turnbull Library's oral history collections are acquired from individuals and organisations, the more guidance and advice that the Library can provide to content creators before the files even come into the Library, the better digital content and metadata the Library receives. Our philosophy is that digital curation begins at the time of file creation. The Library actively engages with the National Oral History Association of New Zealand (NOHANZ), to run oral history training workshops on a regular basis, and to provide recording equipment for hire. The Library also provides oral history advice, including resources on file format recommendations and naming conventions, which are available on the Library's website.[10]

5. Born-Digital Workflows at the Alexander Turnbull Library

5.1 Acquisition

Digital files come to the Library in a myriad of ways—via physical media carrier, such as a portable hard drive, USB drive, optical disc, compact flash card, or via email or electronic transfer. A typical oral history collection may include a variety of files and formats, such as sound files in WAV and/or MP3 format; interview abstracts and a project report as Microsoft Word documents; PDF scans of agreement forms; and JPEG scans of photographs or other documents from the interviewee's early life. The Library asks donors to fill out a media deposit form at the time that the files are supplied to the Library, so staff can verify that all files they intended to provide have actually been received.

However, when the Library first receives physical media carriers, even with donor-supplied file lists to accompany them, it is impossible to know either the quality or the composition of the digital content. Receiving blank discs, or corrupted or poor quality files that are not suitable for long term preservation, as well as materials that are unrelated to a particular collection, is relatively common. So for these reasons, physical carriers of digital content are not accessioned immediately. Instead a proposed acquisition record is created for the materials, which records the type of physical carriers or other method of transfer to the Library, and documents the initial assessment of the collection.

Figure 3. Some of the materials received from the Sisters for Life oral history project, which interviewed women who graduated from the Christchurch School of Nursing in the early 1970s.

10 See https://natlib.govt.nz/researchers/oral-history-advice [Accessed 8 January 2018].

5.2 Technical appraisal

Next, the physical carriers and/or any files received electronically undergo technical appraisal. Technical appraisal allows the Library to assess the quantity and quality of digital files, and determine the processing and preservation needs of the material. This includes a virus check, the generation of a checksum for each file, and transfer of the files from the original media to the secure pre-deposit server for further analysis.

The Library then uses DROID[11], a file profiling tool developed by The National Archives of the United Kingdom, to identify file formats in a collection, and generates a file list to record the metadata for each file. For unusual or obsolete formats, staff determine if the files can be opened, or if they require proprietary software or the creation of an access copy in a standardised format. Even if the file format itself is relatively standard, there may be other technical issues such as illegal characters in the file names, which can require pre-conditioning before loading to the system.

Once the initial technical appraisal has been completed, collections that contain born-digital audiovisual content receive further analysis to assess the quality of the media files, noting any issues with levels, feedback, or background noise that could impact the ability for researchers to listen to the recordings. This information is documented in the ingest report, which is then supplied to the Curator and the arrangement and description staff who will be processing the collection. Files that do require pre-conditioning or preservation action prior to loading to Rosetta go into the work queue of the Digital Archivists or NDHA's preservation analyst respectively. For low quality, or non-standard files, the Library may request replacement files from the donor, where possible.

5.3 Content appraisal

Once the technical appraisal has been completed for both the audiovisual and non-audiovisual materials, the collection receives content appraisal, which determines exactly which materials are to be formally accessioned into the Library's collections. Content appraisal determines if there are any materials that should not be maintained as part of the collection, such as outtake or testing sound files; digital abstracts with the same intellectual content provided in different formats (e.g., MS Word and Adobe PDF), and any other supplied contextual information which may be used in creating the finding aid and descriptive records. Content appraisal is particularly important for maintaining the sustainability of digital archival collections over time. Recent additions to the oral history and sound collections at Turnbull Library comprise nearly four terabytes of data, and with ever-increasing amounts of content being created, it is necessary to make thoughtful and inclusive choices about what is collected and preserved.[12]

Beyond whether or not to actually retain materials for the collections, another key aspect of content appraisal is to determine the access rights for individual files that are accessioned into the collections. Currently three levels of access are available for unpublished digital

11 DROID stands for Digital Record Object Identification. Information on DROID is available at http://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/file-profiling-tool-droid/ [Accessed 8 January 2018].

12 There is extensive literature available on archival appraisal issues, and appraisal in a digital context. See *No Innocent Deposits: Forming Archives by Rethinking Appraisal* (2004) by Richard J Cox, and *'We Are What We Keep; We Keep What We Are': Archival Appraisal Past, Present and Future'*, by Terry Cook (2011), as examples.

objects—100 level open access (available online to anyone anywhere in the world); 300 level (only available onsite at the National Library within the Alexander Turnbull Library Reading Room), and 400 level (only available in the Alexander Turnbull Library Reading Room once permission for access has been granted by the donor or collection trustee).

The appraisal decisions are tracked using the file list, where any files not to be retained are noted. The file list also is used to designate the access rights for each file that is to be retained. Once the appraisal decisions have been finalised, the collection is formally accessioned, and the files are loaded from the pre-deposit server to the processing workbenches of the digital preservation system.

5.4 Loading of digital files

The National Library uses an in-house tool called INDIGO to load digital files to Rosetta. Interviews usually contain multiple sound recording files for single interview session, and these files are loaded together as a multi-file intellectual entity. This allows users to still see the original file names and listen to various parts of the interview, but maintains the context of the sound recordings individually as part of a greater unit. Where appropriate, the multi-file intellectual entities can be loaded with user-defined METS structure maps to define the hierarchy of the digital object, and create metadata labels for the sound recording file. Library staff can therefore apply headings to intellectually differentiate different files within a multi-file IE, such as technical checks or warmup questions from the main content of the interview. This allows greater transparency in content for researchers, who may or may not be interested in listening to sound check files.

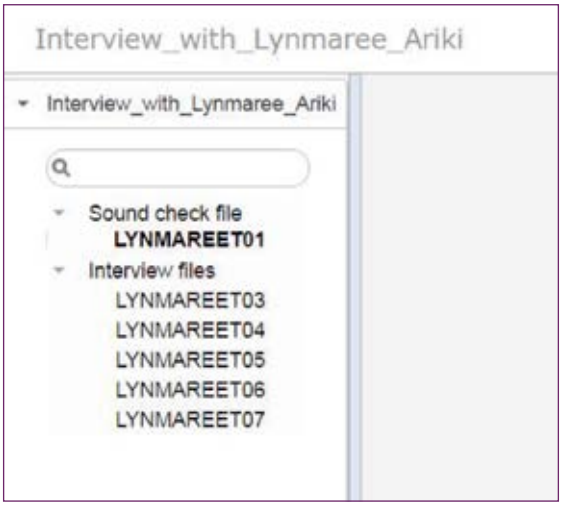

Figure 4. Example structure map for an oral history interview. The digital sound recording itself is a single intellectual entity formed by 6 WAV files.

5.5 Arrangement and description

All intellectual entities in the NDHA must be linked to a descriptive record, either from the National Library's main catalogue for published materials (books, serials, music, maps, etc), or from Tiaki, the Turnbull Library's catalogue for unpublished and archival collections.[13] The linking of born-digital objects to their corresponding descriptive record is carried out within the Rosetta processing workbenches. Within the Rosetta secure digital workbenches, the files can be safely accessed by staff without the risk of accidentally changing the file's content or metadata, and attached to the correct descriptive record. Maintaining the integrity of the digital content over time is core to the work of the NDHA.

The format of the descriptive records and finding aids for born-digital content depends on the collection itself being processed. For some digital objects, it makes sense to have a one-to-one object to record relationship, where each individual digital item has its own descriptive record. For other collections, multiple digital files may be linked to the same descriptive record, particularly if the context of the digital content is contingent on other materials, such as an oral history interview digital sound recording and corresponding abstract document. There is currently no limit to the number of digital objects that can be attached to an unpublished descriptive record, though until recently, the digital files attached to a single record had to all have the same access rights (e.g., all open access, or all requiring permission).

5.6 Discoverability and access by researchers

When the finding aid for the oral history collection and descriptive records for individual interviews are complete, the finding aid is published on Tiaki, and also discoverable via the National Library's federated search across published and unpublished collections.[14] Records with digital content display a link which opens the NDHA viewer, though for oral history collections, users must be onsite in the Turnbull Library's reading rooms to listen to the interviews. This policy regarding born-digital oral history materials is in place to protect the privacy of any individuals who may be mentioned in contemporary interviews. However, the Library does have open access audio collections that can be listened to online from anywhere in the world, such as the Te Upoko o Te Ika Māori radio collection.[15]

Ensuring the appropriate level of detail in finding aids for content can be challenging, whether born-digital or analogue. Oral histories frequently contain personal details about people in the interviewee's life, some of whom may have no idea that the interview materials even exist, or may not wish to have their experiences shared publicly. There can be unexpected connections in oral history collections, particularly in a relatively small country like New Zealand, which has a total population of less than five million. The Library ascribes to a policy for responsible discoverability in its finding aids and descriptive records—that is, providing enough information to allow a researcher to determine if an interview is relevant to their areas of interest, without compromising privacy of the interviewee and any other people mentioned in the interview.

13 See https://tiaki.natlib.govt.nz/ [Accessed 8 January 2018]. The Tiaki catalogue for unpublished collections uses EMu collection management software by Axiell.
14 See https://natlib.govt.nz/ [Accessed 8 January 2018].
15 See https://natlib.govt.nz/collections/a-z/te-upoko-o-te-ika-maori-radio-collection [Accessed 8 January 2018].

6. Workflow Management

The workflow for born-digital materials involves teams from across the Library, which can offer challenges in terms of oversight and workflow management. Staff are often balancing work on a variety of collections at the same time, and if a collection is not straight-forward, or contains files which require pre-conditioning or technical work prior for preservation and access, the collection may be delayed in getting to that end point of being available to researchers.

The Library aims to be transparent regarding any unprocessed arrears, and in these cases will publish a scant collection level record containing a note that the collection is currently being processed by Library staff, and to contact the library for more information. However, holding substantial numbers of unprocessed collections is not ideal for building good donor relations, or for serving the researchers or family members who are keen to access these materials. Unprocessed backlogs can also create added stress for staff who then have to explain to stakeholders why these collections are still in progress. There is always more work to do to reach the strategic goal of born-digital collections being "readily available for access and research."[16]

7. Conclusion and Parting Thoughts

Solid communication and documentation are key to managing born-digital audiovisual materials. Taking time to plan and evaluate is important, as it provides a baseline for decision-making, but it can be easy to fall into the trap of constantly assessing and planning, rather than actually doing the work. The longer collections await processing, the greater the risk of loss of content (or loss of context!). There may never actually be the "right time" or the ideal circumstances. So do what you can, when you can. It is essential to follow proper digital preservation protocols, but also important to identify the challenging aspects of the workflow and make space for reflection, and a bit of creativity.

If you do not know where to begin, look to the standards, and talk to colleagues at other institutions. Make the best decisions you can using the information and resources that you have available to you at the time. And be sure to document those decisions and why you made them. That makes it much easier to re-evaluate later if you need to.

With digital preservation, there is no finish line. The Turnbull Library is always working to improve processes and workflows in order to best care for the collections, and to provide quality service to researchers. It is always a work in progress. Having a strategy for managing born-digital collections, with robust workflows and the right tools in place can make that work a whole lot easier.

16 See https://natlib.govt.nz/about-us/strategy-and-policy/strategic-directions/taonga [Accessed 8 January 2018].

References

Consultative Committee for Space Data Systems (CCSDS) Secretariat (2012). *CCSDS Recommended Practice for an OAIS Reference Model*. Available at: https://public.ccsds.org/pubs/650x0m2.pdf [Accessed 6 November 2017].

Cook, Terry. (2011). "'We Are What We Keep; We Keep What We Are:" Archival Appraisal Past, Present and Future'. *Journal of the Society of Archivists* 32 no 2. Available at: http://www.tandfonline.com/doi/abs/10.1080/00379816.2011.619688 [Accessed 6 November 2017].

Cox, Richard J (2003). *No Innocent Deposits: Forming Archives by Rethinking Appraisal*. Scarecrow Press.

Gattuso, Jay. (2017). "Preservation Actions: where we started and where to do we go from here." *Alexandria: The Journal of National and International Library and Information Issues* 27, no 2. Available at: http://journals.sagepub.com/doi/pdf/10.1177/0955749017725437 [Accessed 6 November 2017].

Jacobs, Tracey and Sandra Falconer (2004). "Ka Mua, Ka Muri; Walking Backwards into the Future: Paths Towards Managing Māori Information in Archives". *Archifacts: Journal of the Archives and Records Association of New Zealand*, October 2004.

Library of Congress. Metadata Encoding and Transmission Standard (METS). Available at: https://www.loc.gov/standards/mets/METSOverview.v2.html [Accessed 6 November 2017].

National Archives, United Kingdom (2017). File profiling tool (DROID). Available at: http://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/file-profiling-tool-droid/ [Accessed 6 November 2017].

National Library of New Zealand (2017). Digital Preservation Programme. Available at: https://digitalpreservation.natlib.govt.nz/ [Accessed 6 November 2017].

New Zealand Parliament (2003). National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act 2003. Available at: http://www.legislation.govt.nz/act/public/2003/0019/latest/whole.html#DLM191962 [Accessed 6 November 2017].

Sajwan, Suman (2017). "Securing the future: Digital Preservation at the National Library of New Zealand." Available at: https://digitalpreservation.natlib.govt.nz/assets/NDHA/Home/NDHA-Booklet-end-Jul17.pdf [Accessed 6 November 2017].

Theimer, Kate (2015). *Appraisal and Acquisition: Innovative Practices for Archives and Special Collections*. Rowman & Littlefield Publishers.

Traue, J. E. (1993). *Dictionary of New Zealand Biography*, vol 2. Available at: https://teara.govt.nz/en/biographies/2t53/turnbull-alexander-horsburgh [Accessed 6 November 2017].