Helen Bones

Linked digital archives and the historical publishing world: An Australasian perspective

1 | INTRODUCTION

Rather than being secondary to the creative role of the author, a writers' wider cultural and economic milieu is a vital component of the creation, dissemination, and influence of cultural artefacts such as books and printed texts. This presents a methodological problem, however, as the interlinked nature of the publishing world has often been overlooked in favour of stories of national development and heroic narratives of single publishers and authors. As a result, in the absence of a strong precedent in international publishing history, accessing information about the nature of these operations involves painstakingly piecing together fragments of information from nationally focused publishing histories and vast and scattered archive collections. How much more, though, could be learnt about society, culture, history, and memory if pertinent documents were presented as part of a constellation of documents or a network of texts where links between documents were used to recreate conversations and connections from 100 years ago? How might such results impact popular understandings of Australian culture and politics and history more broadly? The ARCHIVER project (Angus & Robertson Collection for Humanities and Education Research), based at Western Sydney University, is exploring these questions by developing a model for curating digitally accessible versions of print-based collections that has the potential to revolutionise humanities research. Using structured, linked metadata concepts, the resultant "Linked Archives" tool allows archival collections to be connected in new ways and facilitates complex meta-analyses of associated data between and across these collections to achieve results that would be impossible using traditional methods alone. At the same time, a key aspect of this approach is ensuring the technology builds on and complements more traditional humanistic practices and approaches to archive material and leads to real scholarly outputs and results, rather than merely claiming to supersede what has gone before. The "Linked Archives" technology has broad applications in terms of linking, discovering, and visualising content from historical manuscript collections.

The extensive publishing archives of Angus & Robertson housed in the State Library of New South Wales' Mitchell Library provide a fascinating window into the international literary world that they and other Australian publishers and writers inhabited and interacted with last century. Presided over by founder (with fellow Scotsman David Angus) George Robertson, and subsequently Walter Cousins and Robertson's grandson George Ferguson, the archive consists of meticulously preserved outgoing and incoming correspondence and thus comprehensively represents both sides of conversations relating to the publication of thousands of works. Angus & Robertson was formed in Sydney in 1886, and the company was an integral part of the publishing and bookselling scene in Australia throughout the 20th century. The

archive collection covers this whole time period from the beginnings of the company until a takeover in 1973 altered the outlook and many of the associated personnel. It has been estimated that the entire collection comprises more than 1,000,000 documents. As a result, the very richness and comprehensiveness of the archive presents a serious challenge due to its sheer scale. This paper will demonstrate the ARCHIVER project's unique approach to these challenges and its potential for the field of international book history and humanities research more broadly.

Large digitisation projects, while frequently recipients of significant injections of capital from funding bodies, can only go so far to address the problems inherent in rediscovering the underlying networks of book production. Staff at the State Library of New South Wales are currently involved with digitising the Angus & Robertson collection for accession in the library catalogue. Creating digital surrogates of fragile original material is valuable in terms of preservation as well as vastly improving accessibility if the content is freely available online. This is particularly important for collections related to the transnational book trade, as it reduces the number of trips a researcher needs to make to gather relevant material. However, even when the resources are easily available at our fingertips, researchers can still be left with the challenge of navigating huge corpora of material, and in some ways, this is little better than trawling through the physical archive. The added ease of access is tempered by the massive increase in available material. Without careful planning, digitisation can worsen problems that it is trying to solve.

"Linked Archives" is a specially designed repository and web interface that stores digital images of archival material and enriches them with structured metadata to allow ease of navigation and much more. Using a set of categories based on international library standards and ontologies, "Linked Archives" stores information about each individual document in a collection such as: who wrote it, its origin, when it was sent, and the identity and location of the recipient. Along with these details, important content is recorded using a controlled vocabulary of labels, or "tags" at the click of a button. Using these tools, a user can draw together all documents that mention a particular author or book or topic, or that contain tags related to a certain author, or place, and so on. In very large collections, like Angus & Robertson's, this can save a lot of time and yield results that would be unlikely otherwise. For example, a search for the author C. J. Dennis locates letters from the "Lothian" box of the collection (correspondence between Angus & Robertson and the Melbourne-based publisher Thomas C. Lothian). This includes a conversation about Dennis's iconic *Songs of a Sentimental Bloke*. According to the letters, a newspaper article had claimed that the "Bloke" was first rejected by Lothian before being taken up by Angus & Robertson, and George Robertson asks whether it could be possible that Lothian had refused it (Lothian and Co Ltd Melbourne, March 17, 1923). Because both inwards and outwards correspondence has been preserved, and items are presented in chronological order, the response appears immediately after. Lothian denies ever having seen the manuscript, saying he thought the article very "foolish and irreverent" (Lothian and Co Ltd Melbourne, March 21, 1923). Someone researching C. J. Dennis would be unlikely to look in a folder of Lothian correspondence, but using this tool quickly adds new intrigue to the story of the creation of the book. This is just within one collection, and we are in the process of adding more documents from connected collections such as the Auckland Museum's archive of material relating to New Zealand publisher Whitcombe & Tombs. This approach is considerably more powerful when applied across collections, institutions, or even

state or national borders. The software allows researchers to take advantage of digitised sources while also providing structure and tools to aid navigation.

By capturing whole collections, boxes or folders and associated information in their entirety, "Linked Archives" seeks to avoid another potential hazard for digitisation projects: the problem of selection. Designers of a digital archive need to consider whether the new arrangement will worsen issues of bias introduced by the manner of collection or the way the material is catalogued or arranged in storage, or create new ones. Collections of archives are already highly curated objects, and thus pose a danger to researchers who might take the inherent objectivity of primary sources for granted. A collection of papers represents what has survived numerous culling efforts—deliberate or accidental—as the creator or collector has chosen what to keep and what to discard or censor. In some cases, archives are just a random selection of fragments. The Angus & Robertson collection is vast but still only represents the material that the compilers wanted to be available to future generations. George Robertson culled and arranged his company's correspondence, saying in a 1929 letter: "I am weeding out worthless and uninteresting letters in our publishing department "archives" and by the end of the year they will be available" (Mitchell Library, 1990, ii). Some letters were also destroyed. What Robertson meant by "worthless and uninteresting" we may never know, but the gaps and visibly fragmentary nature of much physical archival evidence tend to be more difficult to ignore when viewing it in person.

Digitising archive collections adds many more opportunities to introduce new partiality to the way people view the archive and the information that is derived from it: something the "Linked Archives" system seeks to reduce. There are many reasons that digitisation projects are not exact replicas of the physical archive. Expensive digitisation procedures are usually reserved for key items that archivists think are of special interest, such as the letters of writer Henry Lawson from the Angus & Robertson collection. While it is a matter of practicality that digitisers must be choosy when dealing with huge masses of documents, something is always lost when items are removed from the physical context of the archive. "Linked Archives" includes containers within archives in their entirety (e.g., a whole box or folder), or ideally, where practicable, the entire collection. The process of capturing whole collection items and retaining their configuration allows "Linked Archives" to retain the act of browsing—as well as navigating straight to a document with a particular tag, it is also possible to browse backwards and forwards from that point in the collection as if you are viewing it in the physical archive. This is very important for context—the second page of a letter, for example, is sometimes incomprehensible without the information on the previous page (such as the identity of the recipient and the date).

The ARCHIVER project seeks to minimise additional problems of selection and bias by capturing whole collections or sections of collections and retaining the shape of the original archive. This avoids the distortions that result from taking things out of their archival context, but it is still important to be aware of the overrepresentation of certain types of sources and what is missing. The explosion in sources relating to World War I in recent years because of centenary commemorations, for example, must have led to heightened awareness of certain topics and issues relating to that period of time. And yet, as some scholars have warned, we are more likely to take for granted the objectivity of the results of digital searches because of the apparently disinterested nature of search algorithms and machine processing. The incomplete

nature of the material being searched is less obvious in digital form. Tim Sherratt advises that researchers need to "develop our critical capacities to be more aware of what we are not being shown and why" by search interfaces (Sherratt, 2016). The avalanche of newspaper content now available online means that many scholars rely on online repositories such as Trove (trove.nla.gov.au) rather than going through microfilms and paper copies in libraries, but what skewed data results from ignoring the newspapers that are not easy to access? The way that sets of archives or individual items are selected for digitisation is not democratic. Lara Putnam also warns of the misrepresentative picture gleaned from this proliferation of digital sources, claiming that it privileges ideas about the circulation of texts while obscuring those less likely to appear in printed sources (Putnam, 2016, 390). The kinds of sources that end up in libraries and archives are already heavily weighted towards the well-to-do and powerful, and it is important not to lose sight of the silences in between.

Retaining the shape of the collection also allows us to deal with another criticism levelled at large online collections of archives. Historians often talk of the valuable finds they make serendipitously when forced by circumstance to sift through physical documents in a library. For some, historical research is too easy when it is all just a search box away. There is a certain element of martyrdom and generational tension at play here. But it is also a fair point—where is the journey? As Graeme Davison has put it, digital research "gets us there more quickly but we skirt the towns and miss a lot of interesting scenery on the way" (Davison, 2009, 43.8). Here, the "Linked Archives" browse function is also useful. The ability to "browse" the archive page by page allows the user to come across things unexpectedly as if browsing in a library, or to encounter items that are peripheral to the main focus and could prove to be interesting or significant. The visual experience of browsing in the physical archive is preserved, if not the tangible aspect.

While something as unreliable as serendipity should not be enough to throw out the possibilities offered by digital explorations, traditional methods need not be entirely superseded by digital ones. It is not a zero-sum game, and the best approach retains the best of both worlds. There is a tendency towards fundamentalism on both sides of the argument—some of the more ambitious claims made by digital proponents of field-wide revolutions or rendering analogue approaches entirely redundant can inspire outright rejection by traditionalists. Although the digital era may be a time of "cultural–historical transformation" akin to the Industrial Revolution or the invention of moveable type (Burdick, Drucker, Lunenfeld, Presner, & Schnapp, 2012, vii), that does not mean that centuries of historical enquiry will be thrown out the window. "Linked Archives" is built on the principle that the best approach does not entirely replace traditional approaches with digital ones. Exclusively relying on search engines that provide instant information would cause a researcher to miss out on the centuries of historical scholarship knowledge needed to interpret the results. The context of information is important, and as Lara Putnam says, "The more far-flung the locales linked through our discoveries, the less consistent our contextual knowledge. The place-specific learning that historical research in a pre-digital world required is no longer baked into the process" (Putnam, 2016, 377). When viewing a digital surrogate through "Linked Archives," all of the information provided by the container and arrangement of the material is retained, plus additional content is available via linked tags and links to other repositories.

As Catherine Bishop argues, the availability of digital sources opens up the possibility of new serendipities. In response to Graeme Davison's freeway analogy, Bishop says that "using digital archives enables us to leave the freeway and visit all those little towns and that interesting scenery and still reach the destination on time" (Bishop, 2017, 778). The addition of full-text-searching capabilities across entire collections previously only available via individual physical parts in the library fundamentally changes the research possibilities. Optical Character Recognition technology produces computer-readable text from printed items (and in less developed form, hand-written material). This allows for both broad-spectrum meta-analyses of material relating to different groups or times, as well as incredible nuance and levels of granularity—a minor writer who barely features in secondary literature is suddenly findable, and unforeseen associations and connections may "serendipitously" emerge. The ability to analyse large volumes of text also allows researchers to trace vicissitudes across entire relationships or networks of relationships. Drawing on the Google Vision Machine Learning platform, "Linked Archives" has in-browser capabilities for OCR that generates a text version of each printed document (the vast majority of the Angus & Robertson materials are printed rather than handwritten).

Optical Character Recognition rarely produces perfect electronic renditions of texts, but it is more than adequate to be useful. The tool also allows users to correct the electronic text and save for future use. Even more excitingly, OCR allows us to semi-automate the process of adding tags to documents (by performing string matches with the list of keywords to generate suggestions). Merely having to oversee the tag additions rather than manually add every single one will make the process vastly more efficient. Ian Milligan has described the problematic, multiplying effects of even small inaccuracies in OCR—the combination of machine and human oversight used by "Linked Archives" avoids many of the pitfalls he describes (2013, 566).

Given the attractive availability of full-text searches via OCR, why then bother going to the trouble of adding the content tags at all? First, the OCR is not perfect, as mentioned. Where OCR is not available, such as at present with many handwritten documents, tags are entirely necessary. Second, tags enrich the data in multiple ways. The tags are a system of keywords based on the international standards laid out in the Resource Description Framework (or RDF), which was created to allow a standardised model for ontological representations of online resources using structured metadata (W3Schools, 2018). The tags refer to existing categorisations made to describe particular types of information, such as "foaf," or "friend of a friend," for describing people and their relationships (www.foaf-project.org/). Using this system allows for consistency and connectivity with other sources of information, but is also a specially designed ontology for the subject matter (developed in the absence of an existing vocabulary for classifying publishing correspondence, particularly in the Australasian context). Keyword tags relating to subjects, people, and publications allow all material relevant to a particular subject to be drawn together instantly across collections and media-types. For example, all documents containing reference to *The Songs of a Sentimental Bloke* will be tagged with a unique identifier under the classification "bibo:book,:" which refers to the Bibliographic Ontology Specification's method of categorising literary works (bibliontology.com).

Full-text searchable versions of digitised archives can only do so much. Adding metadata, in the form of standardised name and topic headings and controlled vocabularies for geographic locations and corporate entities, removes ambiguities that are inherent in primary material. We have used references to standard international vocabularies where possible, such as WorldCat (worldcat.org) for identifying books, which provides unique identifiers for different published works (to remedy problems caused by different books with similar names, misspellings and nonstandard ways of labelling). According to Bair and Carlson, "Studies have shown that over one third of keyword search hits would not be found if controlled subject vocabularies were not included in databases" (Bair & Carlson, 2008, 253). Using such vocabularies helps to provide context that may be lacking when faced with hundreds of search results and reams of information from all over the world. Tim Hitchcock warns of the dangers of full-text searching when results are "deracinated" from the careful information structure of libraries and archive collections, based on a "coherent collection of beliefs and systems for discovering and performing taxonomies on information" (Hitchcock, 2013, 14). Internet searchers risk losing further contextual information found in associated metadata.

One of the major innovations of the project is the development of a special ontology for the classifications of primary materials relating to publishing and the book trade, which only enhances this "framework of source criticism and classification" that Hitchcock is concerned about losing. While referring to international standards and retaining as much of the library's information as possible, it also allows for fine-grained classification of materials in a way that would not be possible using one of the more general library-focused schemes, such as the Library of Congress classifications (loc.gov/catdir/cpso/lcco). We have drawn on subject-specific glossaries and local resources such as the Australian literary database AustLit (austlit.edu.au), which also allows tags to be linked to sources of information (such as the context provided by an individual's entry in the AustLit database). For example, this system ensures that the George Robertson tagged in the system refers to the founder of Angus & Robertson, not the other George Robertson who was involved in the Australian publishing industry at the same time, and the source and unique identifier is embedded in the underlying code. The vocabulary specially caters for terminology specific to the writing and printing industries, but the list of keywords is dynamic and can be continually added to.

Within and across different collections, creating standardised tags for people, entities, and works of literature allows for written "conversations" to be recreated about particular publishing topics. It is not usually possible to "hear" both sides of the correspondence that happened around a particular literary activity or the creation of a work. But if multiple sets of publishing archives are connected and tagged for content, researchers can bring up, in chronological order, items relating to, say, *My Brilliant Career* by Miles Franklin, an Australian classic novel. This may have surprising results, and the potential to lead to what Bair and Carlson describe as "the discovery of new connections between the authors and recorded events—connections that may have taken years for researchers to discover, if ever" (Bair & Carlson, 2008, 253). The possibilities when artificially separated correspondence networks are reconnected for the purpose of study are vast and unprecedented.

The publishing world was (and is) an extensive, international network involving many people engaged in myriad activities, including publishers, writers, printers, agents, critics, readers, booksellers, and other practitioners. In order to best harness archival evidence of this network

and operations, we have added targeted metadata categories to aid with representing and analysing this information. In this way, the design allows for more potential outcomes than simply the opportunity to search across all the material. As Mark Byron has said, a well-designed digital project can go further than merely digitally mirroring the archive—it can "illuminate patterns and complexities in the material not otherwise readily evident in analogue form, presenting textual features at different orders of magnitude and in various modes (visual, statistical, diagrammatic, linguistic, and so on)" (Byron, 2014). Each letter or other item is tagged with special categories for "creator" and "recipient," and information about the organisational affiliation of the writer/receiver, and their geographical location is also collected. These data points can be treated as the "nodes" of a social network, with "edges," or ties or interactions, represented by the sending of letters to and fro. Alison Rukavina advocates for a social network model for the study of the international book trade (Rukavina, 2010)— "Linked Archives" enables this to be taken a step further by providing a mechanism to actually visualise and examine networks. Social network analysis conventions and software tools require the use of this terminology and for the data to be in a certain format, which "Linked Archives" can provide directly (Marin & Wellman, 2011, 11). The system produces data outputs in several formats that can be used to produce network graphs or other kinds of visualisations. These sets can encompass the whole collection or be targeted to certain areas to analyse the centrality of certain actors or the geographical spread of connections in particular decades. You can, for example, withdraw information about the correspondence network of George Ferguson and use it to produce a map or network graph of his letter-writing activities and examine its evolution over time (Figures 1 and 2).

"Linked Archives" has commonalities with other mapping projects such as Stanford's University's "Mapping the Republic of Letters" project, but extends their approach in unique ways. "Mapping the Republic of Letters" involves the capture of "spatial, temporal, and nominal attributes" of letters from the Electronic Enlightenment database with a view to mapping and analysing early modern intellectual networks (Chang et al., 2009; Electronic Enlightenment Project, 2008). Their Palladio tool is designed for this and available online (and was used to make the below visualisations). As their data do not include letter content and deals "primarily with the metadata about letters," it can tell us "only so much about the circulation of ideas," however (Edelstein, Findlen, Ceserani, Winterer, & Coleman, 2017, 407). "Linked Archives" is unique in applying a structured ontology to record topics and content, which allows targeted searches and resultant visualisations. For example, selecting only letters that contain content relevant to educational publishing using the content tag "educational/school books" means users can create a map or graph showing only this correspondence. In this way it is possible to map conversations about a particular book or author.

As well as connectivity, centrality, and other analytical measures that can be used to ascertain the ways in which these relationships evolved, an important innovation is the ability to ascertain tone from the correspondences. Can a breakdown in a relationship be traced to a particular rise in hostility or decline in friendliness in the correspondence, a trend that would not necessarily be visible from viewing individual cases? For example, the literary communities of Australia and New Zealand are much less interconnected in the present day than they were in the early 20th century. The general assumption is that this was inevitable, and the result of the gradual development of nationalisms along with the introduction of tariffs and protectionist measures for local industries (interfering with the international trade in books), but the story is likely to

be more complicated than that (Bones, 2013). An overall network view is great for getting a general overview of operations, but what if the interactions could be weighted in terms of the closeness or friendliness of the relationship? In social network analysis terms, this is the addition of "weight" as an extra variable in the network data. Being able to map the literary connections between the two countries and examine the tone of the relationship and change over time allows for added insight, perhaps confirming the impact of a particular event or suggesting new causes. Combined with historical information, these data can then be used to test long-held assumptions about literary development. Concrete data that represent trans-Tasman literary relationships, especially the nature and evolution of those relationships, will be hugely valuable for research in this area.
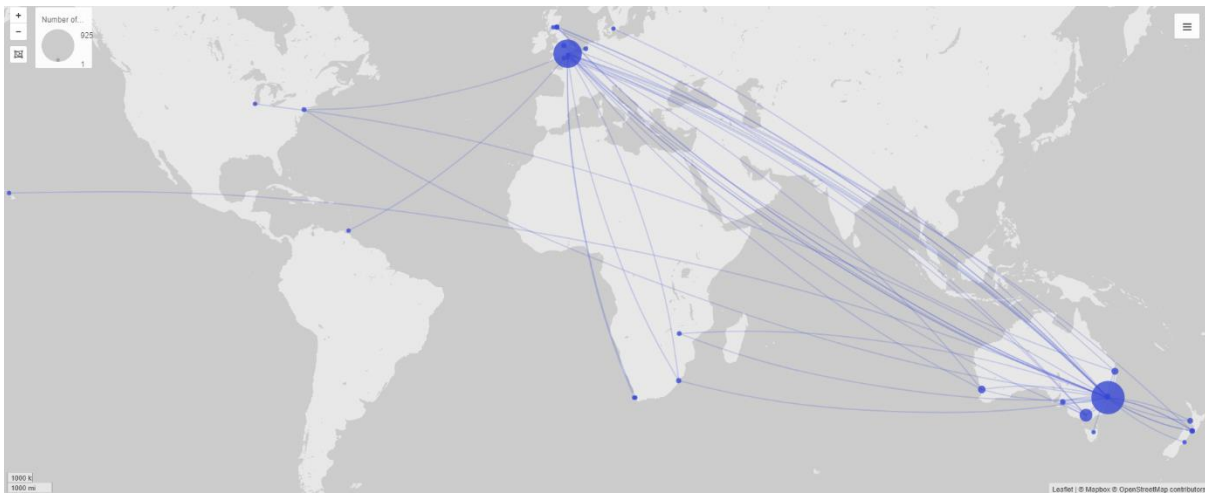

FIGURE 1 Example of a map showing George Ferguson's letter-based correspondence as represented by the current set of documents tagged in "Linked Archives." Constructed using Stanford University's Humanities + Design Lab's Palladio Tool: http://hdlab.stanford.edu/palladio/)

Determining the "tone" of an interaction is a subjective exercise, but we have developed a technique to keep this as objective as possible. Positive or negative signs can be used to describe the nature of relationships in terms of their relative strength or weakness (Easley & Kleinberg, 2010, 119) but a binary model is not adequate for complicated and wide-ranging business correspondence. We identified several categories to be included, from familiar (denoting a relationship beyond the work element), cordial (an on-going, genial relationship based around work), neutral (purely business correspondence), disgruntled, (a complaining tone) or hostile (being openly rude). Classifying the documents using only this taxonomy is highly subjective, however, so we have used the type of salutation as a more objective measure to determine the nature of the relationship (in combination with some other markers of familiarity or contempt). For example, "Dear Sirs" denotes a neutral, business-like tone, "Dear Mr Ferguson" a cordial business relationship and "Dearest George" a more intimate association. These classifications add a weighting to the data pertaining to each interaction, which can then be used to ask certain questions of it. For example, could proximity denote a certain kind of tie? Does the strength of relationship differ based on how close in geographical terms the nodes are? How does this change over time?

The most important thing about using a controlled vocabulary and structured metadata is that it allows the project to harness and further the potential of linked open data for connecting
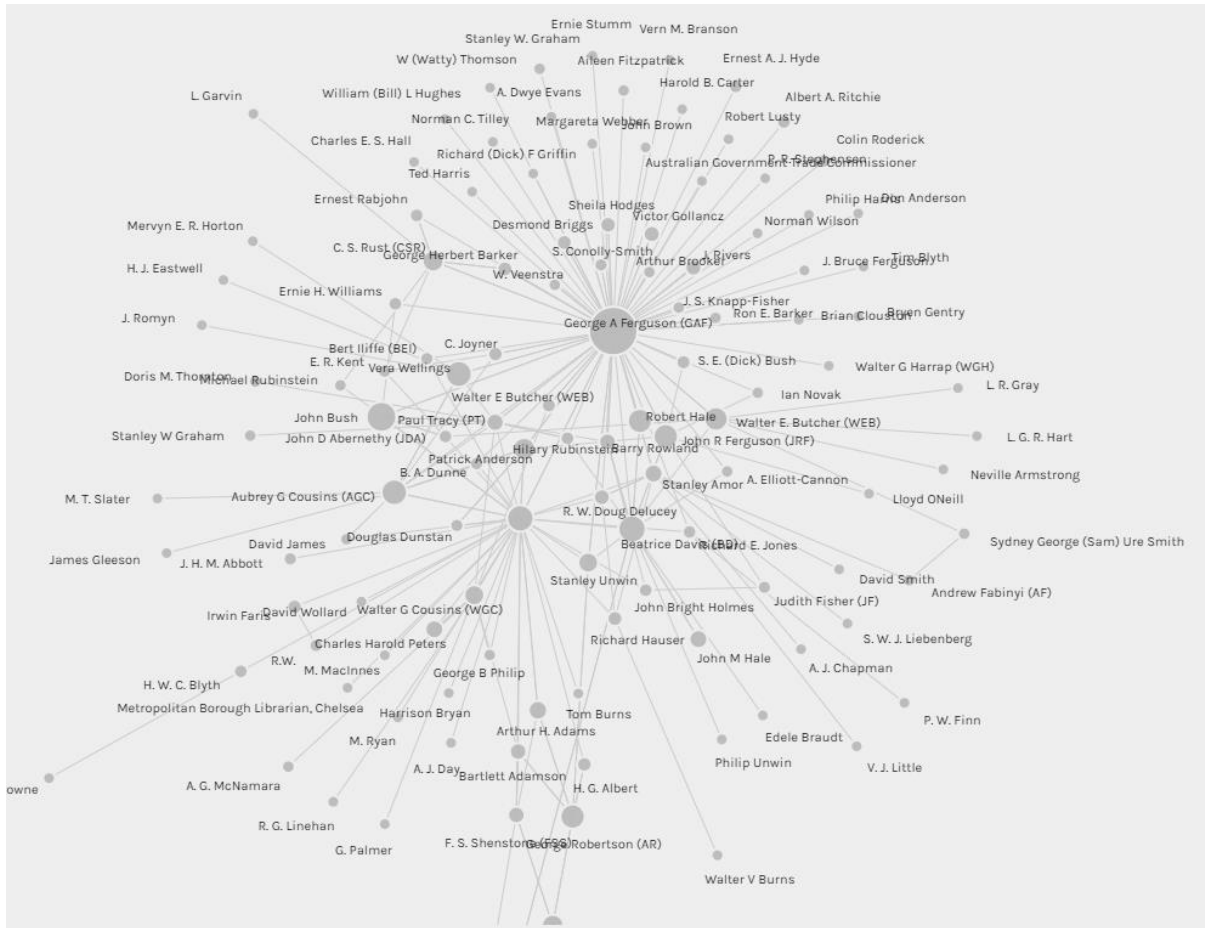
FIGURE 2 This is a network graph made using the same data and the same tool as in Figure 1

datasets online and making them easily accessible. Constructing a vocabulary and structure using the principles of linked open data as laid out in the Resource Description Framework allows our data to work with other datasets without having been specifically set up with this in mind. The RDF essentially sets out a way of describing metadata so that it can be understood and parsed by any other system built using linked data (Klyne & Carroll, 2004). Information is embedded in the underlying code that is readable by other systems that can decode linked open data formats. Each "tag" from our system captures information about the category and vocabulary the keyword is describing, stored in an XML-like expression called a "triple." The results of searches (metadata and tags associated with a given set of documents) can be downloaded in several formats compliant with these systems.

Linking together different projects and sets of data is crucial for scholarship in general, and in particular, a field like book history, which is interested in the global creation and flow of texts. It relies on big-picture analysis that involves large, expensive projects collecting big sets of data. Many of these projects are conducted to answer a single set of research questions and amass a considerable quantity of information that never leaves the cupboards or hard drive of the researcher. Linked data have the potential to make data useful beyond the scope of a single work, which is crucial for book history as, as Franco Moretti has said, "a field this large cannot be understood by stitching together separate bits of knowledge about individual cases, because it isn't a sum of individual cases: it's a collective system, that should be grasped as such, as a whole" (Moretti, 2005, 4). Even if the scope is limited to just the Australasian book trade, the field is too vast to be tackled by a single project, and results can be vastly intensified by

initiatives designed to link several together. "Linked Archives" suggests the possibility of breaking open archives that constrain past conversations about book history to what can be found within certain institutions. This is especially important in the context of transnational publishing operations where the activities of certain publishers have largely been framed within a national context (e.g., Alison (2009) on Angus & Robertson, Waite (2008) on Whitcombe & Tombs, with the notable exception of Ensor (2012)).

As well as the ability to link to other things, within its own dataset, "Linked Archives" is set up using carefully structured metadata to allow for the broadest range of research uses possible. The RDF defines data relationships as a "graph," meaning that no one resource or category has precedence over another. This avoids hierarchical subject categorisations, which require the use of proscriptive classifications for human data that seldom fit neatly into a single category. The background metadata are structured to allow for both broad and specific searches without presupposing search interests. For example, if you want to find out about conversations with and pertaining to a particular author, you can bring up the tag for that author, but also broaden the search to include related concepts. As opposed to confining concepts to certain categories of ever-increasing specificity, for example, Arthur H. Adams is an Australian novelist, or a New Zealand poet, the entity known as "Arthur H. Adams" will be associated with broader tags such as "poets," "New Zealand," "Australian," and "novelists," allowing the underlying ontology to better reflect the fluidity of human lives (via the Simple Knowledge Organization System (SKOS) "broader" class).

The sometimes ill-defined and highly self-reflective field of digital humanities suffers from a perceived or (at times) actual lack of engagement with "real life" humanities scholarship (Bode & Arthur, 2014, 2). This project represents a blending of traditional and digital methods to help bridge this divide. Digital humanities approaches do not have to be in an either/or battle with more traditional approaches. "Linked Archives" provides the potential for large-scale information processing of a "big data" project to reveal new historical contexts while not attempting to supersede the explanatory power of the archive. Thus, it assists with what Mark Byron describes as necessary: to "strengthen a middle ground in scholarly practice and digital design, where digital humanities and traditional scholarly methods might interact productively and thus extend their respective zones of capability" (Byron, 2014, 17). It brings with it the possibility of a large-scale textual and structural analysis of sets of manuscript materials that represents a similar revolution in archival studies as those afforded to literary studies by the rise of large-scale textual analysis (Jockers, 2013, 7).

REFERENCES

Alison, J. (2009). *Doing something for Australia: George Robertson and the early years of Angus and Robertson, publishers: 1888–1900*. Melbourne: Bibliographical Society of Australia and New Zealand.

Bair, S., & Carlson, S. (2008). Where keywords fail: Using metadata to facilitate digital humanities scholarship. *Journal of Library Metadata*, 8(3 (2008)), 249–262. https://doi.org/10.1080/19386380802398503

History Compass. 2019; e12522. https://doi.org/10.1111/hic3.12522

Bishop, C. (2017). The serendipity of connectivity: Piecing together women's lives in the digital archive. *Women's History Review*, 26(5), 766–780. https://doi.org/10.1080/09612025.2016.1166883

Bode, K., & Arthur, P. L. (2014). Collecting ourselves. In P. L. Arthur, & K. Bode (Eds.), *Advancing digital humanities: Research, methods, theories* (pp. 1–12). Basingstoke: Palgrave Macmillan.

Bones, H. (2013). New Zealand and the Tasman Writing World. *History Australia*, 10(3), 129–148. https://www.tandfonline.com/doi/abs/10.1080/14490854.2013.11668484

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T. S., & Schnapp, J. T. (2012). *Digital_Humanities*. Cambridge, MA: Massachusetts Institute of Technology.

Byron, M. (2014). Exercises in Battology. In P. L. Arthur, & K. Bode (Eds.), *Advancing digital humanities: Research, methods, theories* (pp. 15–27). Basingstoke: Palgrave Macmillan.

Chang, D., Ge, Y., Song, S., Coleman, N., Christensen, J., & Heer, J. (2009). *Visualizing the republic of letter*s. Stanford University. http://web.stanford.edu/group/toolingup/rplviz/papers/Vis_RofL_2009.pdf.

Davison, G. (2009). Speed-relating: Family history in a digital age. *History Australia*, 6(2), 43.1–43.11. https://doi.org/10.2104/ha090043

Easley, D., & Kleinberg, J. (2010). *Networks, crowds and marketing: Reasoning about a highly connected world. New York*, NY: Cambridge University Press.

Edelstein, D., Findlen, P., Ceserani, G., Winterer, C., & Coleman, N. (2017). AHR forum: Historical research in a digital age: Reflections from the Mapping the Republic of Letters Project. *American Historical Review*, 122(2), 400–424. https://doi.org/10.1093/ahr/122.2.400

Electronic Enlightenment Project (2008-2017). Bodleian Libraries, University of Oxford. http://www.e-enlightenment.com/.

Ensor, J. D. (2012). *Angus & Robertson and the British trade in Australian books, 1930-1970: The getting of bookselling wisdom*. London, New York: Anthem Press.

Hitchcock, T. (2013). Confronting the digital: Or how academic history writing lost the plot. *Cultural and Social History*, 10(1), 9–23. https://doi.org/10.2752/147800413X13515292098070. https://www.loc.gov/catdir/cpso/lcco/

Jockers, M. L. (2013). *Macroanalysis: Digital methods and literary history*. Urbana: University of Illinois Press.

Klyne, G., & Carroll, J. J. (2004). Resource description framework (RDF): Concepts and abstract syntax—W3C recommendation. http://www.w3.org/TR/rdf-concepts/.

Lothian and Co Ltd Melbourne (papers), ML MSS 314/53, Angus & Robertson Collection, Mitchell Library, Sydney.

Marin, A., & Wellman, B. (2011). Social network analysis: An introduction. In J. Scott, & P. J. Carrington (Eds.), *The sage handbook of social network analysis* (pp. 11–25). London: Sage. 2011

Milligan, I. (2013). Illusionary order: Online databases, optical character recognition, and Canadian history, 1997–2010. *The Canadian Historical Review*, 94(4), 540–569. https://doi.org/10.3138/chr.694

Mitchell Library (1990). *Guide to the Angus & Robertson archives in the Mitchell Library, State Library of new South Wales*. Sydney: Library Council of New South Wales.

Moretti, F. (2005). *Graphs, maps, trees: Abstract models for a literary history*. London; New York: Verso.

Putnam, L. (2016). The transnational and the text-searchable: Digitized sources and the shadows they cast. *American Historical Review*, 121(2), 377–402. https://doi.org/10.1093/ahr/121.2.377

Rukavina, A. (2010). *The development of the international book trade, 1870–1895: Tangled networks*. Basingstoke, England: Palgrave Macmillan.

Sherratt, T. (2016). Caring about access. *Paper presented at digital directions* 2016, National Film and Sound Archive, Canberra, Australia, https://dx.doi.org/10.6084/m9.figshare.4229402.v1.

W3Schools (2018). XML RDF. XML Tutorial. Accessed February 29, 2018. https://www.w3schools.com/xml/xml_rdf.asp.

Waite, N. D. (2008). *Books for a nation: The Whitcoulls story*. Auckland: Whitcoulls.