doi: http://dx.doi.org/10.23925/1980-7651.2018v21;p13-20

The other side of the story: challenges and opportunities in building sustainable online knowledge resources

Ailie Smith[•]

Abstract

With more resources becoming freely available online through digitisation projects and publication of online public knowledge resources, the possibilities for connecting previously separate pieces of information have increased and some of the untold stories in the history of science can emerge. However there can be problems in relying on external knowledge services to connect pieces of the story. There are also gaps in the scientific biographies of many notable people, which may be filled by resources from other. This paper explores some of the possibilities and challenges associated with producing and sustaining online knowledge resources.

Keywords

History of science; History of Australian science; Digitization; Biography

[•] eScholarship Research Centre, The University of Melbourne, Australia. ✉ ailie.smith@unimelb.edu.au. The present paper derives from a presentation at symposium Doing History of Science in a Digital, Global, Networked Community: Tools and Services Linking Scholars and Scholarship, 25th International Congress of History of Science and Technology, Rio de Janeiro 22-29 July 2017.

Introduction

In 2010, the *Encyclopedia of Australian Science*[1] was one of the first datasets harvested by the National Library of Australia's *Trove*[2] service. This continued the activities of linking the *Encyclopedia* into a broader web of public knowledge resources and allowed the biographical information in the *Encyclopedia*, focussing on the scientific careers of its entries, to be linked to resources focussing on other aspects of their lives. With more and more resources becoming freely available online, through digitisation of newspapers and primary source material such as archival collections, and initiatives such at the *Internet Archive*[3] and the *Hathi Trust Digital Library,*[4] the possibilities for connecting previously separate pieces of information have increased and some of the untold stories in the history of science can emerge.

However there can be problems in relying on an external knowledge aggregation service to connect pieces of the story. In 2016 *Trove* faced funding cuts,[5] and its capacity to continue to play this important role into the future was put in doubt. In addition to this there are still holes in the scientific biographies of many notable people. There are stories that may never have been recorded in the biographical details of scientists before that are more challenging to uncover. We may need to look to other disciplines to see where the lives of prominent scientists have intersected with disciplines to fill in gaps in their biographies. This paper explores some of the possibilities and challenges associated with producing and sustaining online knowledge resources.

Background

The *Encyclopedia of Australian Science* (EOAS) is an output of the University of Melbourne's eScholarship Research Centre (ESRC). ESRC was originally established in 1985 as the Australian Science Archives Project (ASAP) and was located within the Department of History and Philosophy of Science at the University of Melbourne. ASAP was established to address a problem that was identified with the records documenting the history of science in Australia not being kept. ASAP set about locating, documenting and finding homes for archival collections relating to Australian science.[6] Alongside this, a project was run to locate archives of individuals relating to the history of science in Australian institutions and to publish an index to these records. First published as a print volume - *Guide to the Archives of*

---

[1] eScholarship Research Centre, "Encyclopedia of Australian Science," accessed 26 March 2018, http://www.eoas.info/.
[2] National Library of Australia, "Trove," accessed 26 March 2018, https://trove.nla.gov.au/.
[3] "Internet Archive: Digital Library of Free Books, Movies, Music & Wayback Machine," accessed 26 March 2018, https://archive.org/.
[4] "HathiTrust Digital Library | Millions of Books Online," accessed 26 March 2018, https://www.hathitrust.org/.
[5] Paul Daley, "National Library's Trove: A Great Digital Democracy under Threat," *The Guardian*, 14 March 2016, http://www.theguardian.com/books/2016/mar/14/national-librarys-trove-a-great-digital-democracy-under-threat.
[6] Ailie Smith, & Gavan McCarthy, "The Encyclopedia of Australian Science: A Virtual Meeting of Archives and Libraries," *The Australian Library Journal* 65, no. 3 (July 2, 2016): 191–202, https://doi.org/10.1080/00049670.2016.1212318.

Science in Australia: Records of Individuals - in 1991.[7] The information gathered for the publications was soon developed into an online resource, known as *Bright Sparcs*, which was published online in 1994.[8]

In 1999, ASAP became the Australian Science and Technology Heritage Centre, or Austehc. The Centre continued the work of ASAP, producing a companion resource to *Bright Sparcs* in 2000, known as *Australian Science at Work.* While *Bright Sparcs* documented people, *Australian Science at Work* documented organisations and corporate bodies involved with science, technology, engineering and medicine in Australia.[9] The two resources developed and grew in parallel, as separate resources, connected only by a shared set of bibliographic and primary source references.

Increasingly, Austehc was working with disciplines outside history and philosophy of science. As part of some organisational restructuring, Austehc moved out of the Department of History and Philosophy of Science and into the University Library in 2007, becoming the eScholarship Research Centre. Using informatics structures established for projects such as *Bright Sparcs*, ESRC continued to work with a wide variety of collaborators to create web-based public knowledge resources.

In 2010, *Bright Sparcs* was amalgamated with *Australian Science at Work* to become the *Encyclopedia of Australian Science*, a single register of the people, organisations and industries involved in developing Australia's scientific, technological, engineering and medical heritage. Also in 2010, the *Encyclopedia of Australian Science* was one of the first resources to be harvested by the National Library of Australia's *Trove* service.[10] *Trove* is an information aggregation service that brings together data from a number of resources. Alongside public knowledge resources such as the *Encyclopedia of Australian Science* sits an extensive bibliography, picture collections and a growing collection of digitised and transcribed Australian newspapers.

The other side of the story

It is notable that a number of the people documented in *Encyclopedia of Australian Science* for their contribution to science have also appeared in a variety of other contexts within resources for other disciplines. For example geologist and palaeontologist Dorothy Hill (1907-1997) appears in both the *Encyclopedia of Australian Science*[11] and *The Australian*

[7] Gavan McCarthy, *Guide to the Archives of Science in Australia: Records of Individuals* (Port Melbourne: D.W. Thorpe in association with Australian Science Archives Project & National Centre for Australian Studies, 1991).
[8] Smith & McCarthy, "Encyclopedia of Australian Science."
[9] Ibid.
[10] Gavan McCarthy, Ailie Smith, & Michael Jones, "Looking Beyond the Archive: Utilizing Encoded Archival Context in a Broader Societal Context," *Journal of Archival Organization* 12, no. 1–2 (April 3, 2015): 143–64, https://doi.org/10.1080/15332748.2015.1001206.
[11] G. J. McCarthy, "Hill, Dorothy (1907 - 1997)," *Encyclopedia of Australian Science*, accessed 28 March 2018, http://www.eoas.info/biogs/P000494b.htm.

Women's Register.[12] Surveyor Edward Charles Frome (1802-1890) appears in the *Encyclopedia of Australian Science*,[13] *Arrowsmith's Australian Maps*,[14] *Australian Dictionary of Biography*,[15] and *Design and Art Australia Online*.[16] Each of these resources focuses on a different aspect of the life of Frome, from his career as Surveyor General in South Australia, to his contribution to maps produced by London-based Cartographer John Arrowsmith, and his work as an amateur artist. Links between these resources are minimal, and not all are featured in *Trove*.

Historians of science now have access to a much wider range of resources and there is potential to bring together previously disconnected pieces of information, enhancing – or perhaps challenging – the existing biographical narratives. However, the lack of connection between resources means that these links may still be missed, and resources will continue to develop as disconnected siloes of information, rather than a networked knowledge landscape. *Trove* has provided a valuable piece of knowledge infrastructure in aggregating several key resources in the Australian context, however they can only harvest from a finite number of resources so there will be others that are not included.

Missing pieces and making connections

Increasingly full text versions of historic publications are being made available through resources, such as the *Internet Archive* and the *Hathi Trust*. This is giving us greater access to resources and information that was previously locked in libraries and archives. Newspaper digitisation projects are also introducing the possibility of discovering previously unknown information about people that has until now only been only accessible through time-consuming manual searches. Resources such as the National Library of Australia's *Trove* and *OCLC WorldCat*[17] help to aggregate this information.

Looking for information about botanist George Bentham (1800-1884) uncovered an example of how increased digital access to information can lead to potential new discoveries. Bentham was based at the Herbarium in Kew, England, and produced *Flora Austaliensis*, a seven volume catalogue describing Australian flora. This was something that Ferdinand von Mueller, the government botanist in Victoria, Australia, had wanted to produce himself, however rather than go into competition with Bentham, he was forced to

[12] Elle Morrell, "Hill, Dorothy (1907 - 1997)," *The Australian Women's Register*, accessed 31 March 2018, http://www.womenaustralia.info/biogs/AWE0064b.htm.
[13] Helen Cohn, "Frome, Edward Charles (1802 - 1890)," *Encyclopedia of Australian Science*, accessed 29 March 2018, http://www.eoas.info/biogs/P005549b.htm.
[14] Dorothy Prescott, "Frome, Edward Charles (1802 - 1890)," *Arrowsmith's Australian Maps*, accessed 29 March 2018, http://www.asmp.esrc.unimelb.edu.au/biogs/E000140b.htm.
[15] B. C. Newland, "Frome, Edward Charles (1802–1890)," in *Australian Dictionary of Biography* (Canberra: National Centre of Biography, Australian National University), accessed 29 March 2018, http://adb.anu.edu.au/biography/frome-edward-charles-2070.
[16] "Edward Charles Frome b. 7 January 1802," *Design and Art Australia Online*, accessed 29 March 2018, https://www.daao.org.au/bio/edward-charles-frome/.
[17] OCLC, "WorldCat.Org: The World's Largest Library Catalog," accessed 28 March 2018, https://www.worldcat.org/.

cooperate on Bentham's publication.[18] Bentham has an entry in the *Encyclopedia of Australian Science*[19] as well as an entry in the *Australian Dictionary of Biography* (ADB).[20] He is recognised as an important figure in Australian history because of his work on documenting Australian flora, however both the ADB and EOAS entries give no indication that he ever visited Australia – his work was done from Kew with samples collected by others. However a newspaper death notice states that he was "formerly of Melbourne"[21], suggesting that he had in fact spent time in that Australian city. Without searchable digitised newspapers, a brief note like this is easily missed. This could prove an interesting addition to Bentham's story, but is it true? Or did the newspaper misinterpret his close collaboration and correspondence with Melbourne-based botanists to mean that he had spent time there? Further investigation is required to uncover the truth.

While broader access to digital information can make researching the history of science and accessing historic resources easier and lead to new discoveries, it may also be proliferating or resurfacing inaccurate or incorrect information.

A further challenge that comes with this increased access to information and resources is how these resources work together. Are we being provided with better access to publications that may be difficult to find because they are out of print, or is the proliferation of digitised books just producing more disconnected siloes of information?

Using George Bentham's major work covering Australian botany, *Flora Australiensis*, as an example, the quantity of data available online becomes apparent. There is a digitised copy in the *Internet Archive*.[22] The metadata for this resource links to a copy of the book in the *Biodiversity Heritage Library*,[23] which is the same copy publication, as demonstrated by the matching annotations in the digitised versions on each website. The *Hathi Trust* also has a copy of the publication, which is not available as full text due to copyright restrictions,[24] this links to a catalogue entry in *WorldCat*.[25] *WorldCat* aggregates information from a wide range of sources and lists a number of editions of the publication. The record for the eBook edition links to versions in the *Internet Archive*, *Hathi Trust*, and *Google Books*,[26] which also has a full digitised copy of the work. There are some connections between the various

[18] N. T. Burbidge, "Bentham, George (1800–1884)," in *Australian Dictionary of Biography* (Canberra: National Centre of Biography, Australian National University), accessed 26 March 2018, http://adb.anu.edu.au/biography/bentham-george-2979.

[19] G. J. McCarthy, C. R. Moje, & N. Walsh, "Bentham, George (1800 - 1884)," *Encyclopedia of Australian Science*, accessed 26 March 2018, http://www.eoas.info/biogs/P000226b.htm.

[20] Burbidge, "Bentham, George (1800–1884)."

[21] "Death of Mr. George Bentham," *Evening Journal* (Adelaide, SA : 1869 - 1912), 20 September 1884.

[22] George Bentham, & Ferdinand Jacob Heinrich von Mu?ller [sic], *Flora australiensis?a description of the plants of the Australian territory [sic]* (London : L. Reeve and co., 1863), http://archive.org/details/mobot31753000235504.

[23] George Bentham, & Ferdinand von Mueller, *Flora Australiensis:A Description of the Plants of the Australian Territory*, v. 1 (London : L. Reeve and co., 1863), https://www.biodiversitylibrary.org/item/3669.

[24] George Bentham, & Ferdinand von Mueller, *Flora Australiensis: A Description of the Plants of the Australian Territory.* (London, repr. Amsterdam: Reeve and Co., Asher, 1863), https://catalog.hathitrust.org/Record/100143032.

[25] "Flora Australiensis a Description of the Plants of the Australian Territory.," OCLC WorldCat, accessed 27 March 2018, http://www.worldcat.org/oclc/621172286.

[26] George Bentham, & Ferdinand von Mueller, *Flora Australiensis: Ranunculacea to Anacardiacea* (L. Reeve and Company, 1863), https://books.google.com.au/books?id=vEIPAQAAIAAJ&source=gbs_navlinks_s.

resources, but there also appears to be a lot of duplication of effort and a disjointed network of one way links and isolated siloes. In some cases you need to know that the resources exist in order to discover content in them, because they are either not being indexed by search engines such as Google, or are buried after pages of other results.

An added complication is that *Flora Australiensis* was not a one off publication, but a serial that ran to seven volumes. Some of the resources listed in this paper neatly group the volumes together in their metadata interface, while others treat them as unique publications, requiring a combination of luck and perseverance to locate the other volumes of the publication.

All of this work can also be continuing to create disconnected siloes of information. This will often result in the duplication of work, and also the potential to overlook the opportunity of discovering new connections and stories that enhance what is already known.

Another challenge that can be created in the connected world is that of having too much information, or information overload. Using the example of Dorothy Hill - an Australian geologist and palaeontologist who was the first female to be elected as a Fellow of Australian Academy of Science, as well as the Academy's first female president -[27] it is possible to demonstrate that an increase in available information has made it more difficult to find relevant information. Figure 1 is a screen shot captured in 2012 of part of Dorothy Hill's entry in *Trove*.[28] It isuggests digitised newspaper and journal articles that are possibly related, and the majority of the suggestions are clearly about the right person. Figure 2 shows the same web page in 2018. This time, while the number of search results in the digitised newspapers has increased from 147,445 to 339,849, none of the suggested newspaper articles are relevant. This could be because the search algorithm changed in the intervening years and does not give as precise results now. But it is also very likely that this is because in the last six years the volume of digitised newspaper content has increased exponentially and there is now a lot more noise to get in the way. The more information there is to search, the harder it can be to find relevant information.

[27] McCarthy, "Hill, Dorothy (1907 - 1997)."
[28] "Hill, Dorothy (1907-1997) - People and Organisations," Trove, accessed March 26, 2018, https://nla.gov.au/nla.party-754082.

Figure 1. Screenshot of Dorothy Hill's entry in *Trove*, 2012

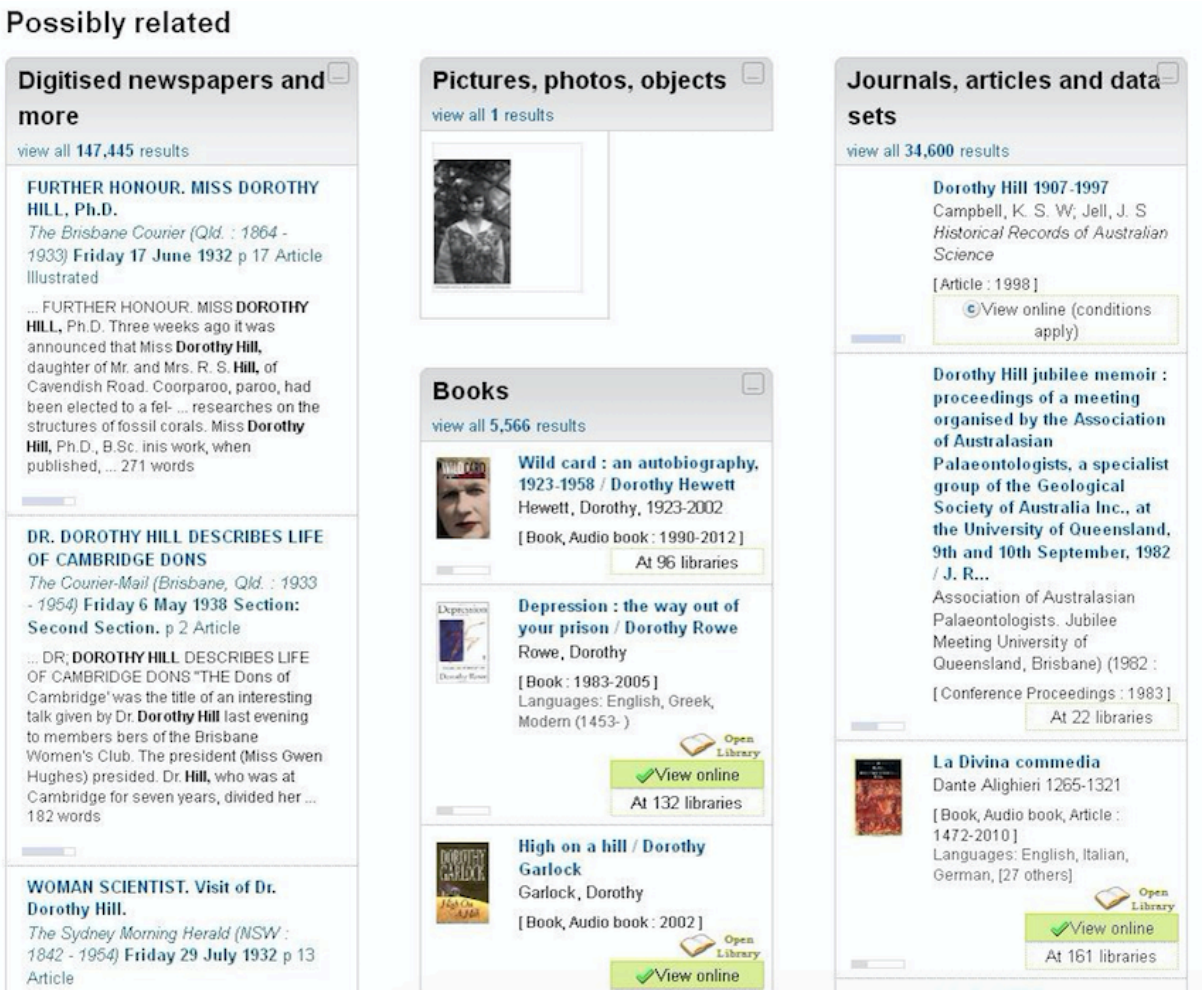

Figure 2. Screenshot of Dorothy Hill's entry in *Trove*, 2018

Digital fragility

The reliance on third parties to maintain and provide access to resources, while convenient, opens up the risk that resources that researchers have come to rely on may change over time or disappear. Maintaining such resources requires ongoing effort and funding. While they may provide end users with a simple and free service to access and interrogate a wide range of sources, behind the scenes extensive resources in terms of time, personnel, infrastructure and funding are being expended to provide the services.

In 2016 it was announced that due to funding cuts to the National Library of Australia, the ability of the Library's *Trove* service to continue to be maintained, developed and grow was under threat.[29] The many users of *Trove* came out in force and a social media campaign ensued, under the hash tag #fundTrove. While some additional funding was provided later in the year,[30] this highlighted the risk that the digital resources we have come to rely on may have finite lifetimes. The campaign to fund *Trove* brought out many and varied stories about how people were using the service and what they had been able to uncover through it. Perhaps we need to be more proactive about telling the stories of how we use these resources to keep reminding ourselves, and others, of their value.

Conclusion

Since it first went on the web in 1994, *Bright Sparcs* – and later the *Encyclopedia of Australian Science* – has played a role in the development of an increasingly connected digital public knowledge landscape. Knowledge and resources have increasingly moved out of libraries and archives and into people's computers. Research has been greatly aided by digitised resources. Where previously searching through newspapers and publications would have taken a great deal of manual effort in the print-based world, searches in the digital world can be conducted in seconds.

But with these developments and improvements has come a reliance on resources that may not stand the test of time and be available, or available in the same form, in the future. There is a need to support and promote the value of resources such as *Trove*, to reduce the risk of them idling or disappearing in the future. There is also a need to ensure that there are citations between, and aggregation of, digital public knowledge resources in order to fully embrace the capacity of the digital world to enhance and complete the disconnected biographies and biographic details of prominent scientists.

---

[29] Daley, "National Library's Trove."

[30] "MYEFO Sees Funding Channelled to National Library," *ABC News*, 20 December 2016, http://www.abc.net.au/news/2016-12-20/national-library-of-australia-gets-funding-for-trove-in-myefo/8136738.