Author Accepted Manuscript version of chapter 10 in Lewi, H. (Ed.), Smith, W. (Ed.), vom Lehn, D. (Ed.), Cooke, S. (Ed.). (2019). *The Routledge International Handbook of New Digital Practices in Galleries, Libraries, Archives, Museums and Heritage Sites*. London: Routledge, https://doi.org/10.4324/9780429506765

Hacking Heritage: Understanding the Limits of Online Access

Author: Tim Sherratt (University of Canberra)

In 1995, an Australian government plan for digital innovation highlighted some exciting possibilities that lay ahead for the cultural sector (Department of Industry, Science and Tourism, 1995). Access to collections would be 'simplified' through the creation of an 'Electronic Smithsonian'—a portal to bring together the holdings of national cultural institutions:

> For the user this home page access will be like walking electronically down an avenue of all our major museums or galleries. People will be able to find out about the collections, their significance and context, and use interactive links to other institutions, as well as to access digitalised images.

Two decades later, the United Kingdom (UK) government's Culture White Paper (Department for Culture, Media and Sport, 2016) envisaged a similar pathway for users, while seeking to make the UK 'one of the world's leading countries for digitised public collections content': 'We want users to enjoy a seamless experience online, and have the chance to access particular collections in depth as well as search across all collections.'

Digital technologies continue to offer a beguiling vision of universal access. Everyone, everywhere will be able to find and use our cultural collections. Hidden riches will be revealed. Obstacles to discovery and exploration will be removed. Technology, it is often assumed, can push collections across a threshold—it can make them *open*.

But access is never truly open. Voices are suppressed or lost. Information is withheld or restricted. Priorities are set. Technology fails. Yes, placing collections online does create exciting new opportunities for engagement and use, but if we focus on the threshold moments, on the expansion of scope and scale, we draw attention away from the forces that control and shape the cultural record. We make it harder to see what is missing.

As the digital GLAM sector expands worldwide, it is important to critically examine the *meaning* of access. Access to what and by whom? If cultural institutions hope to avoid replicating existing biases, we need to understand the limits and the possibilities of online systems. This chapter explores what we can discover about access by peeking behind existing interfaces—by hacking them in creative and constructive ways. Using examples from the Australian cultural sector, I suggest that we can turn digital technologies against themselves to question the way access itself is manufactured and controlled.

The Promise of Online Access

There is plenty of evidence that the web, coupled with digitisation, optical character recognition (OCR) and search, has transformed the way we find and use cultural heritage collections. The delivery of historical Australian newspapers through Trove is one of the most obvious and powerful examples. A 2013 user survey indicated that Trove use across Australia generally corresponded with the national population distribution (Ayres, 2013). People who had never visited the National Library of Australia in Canberra, or even one of the state libraries, are now able to dive deep into their newspaper holdings. Digital delivery has lessened physical isolation. Similarly, the National Archives of Australia (NAA) has pursued digitisation as a solution to the 'tyranny of distance' (Ling & McLean, 2004). In its 2015–16 Annual Report, the NAA (2016) noted that while 111,526 records had been viewed in its reading rooms, 10,579,254 records had been accessed online.

This is not simply a matter of convenience. People who might never identify as 'researchers', who might never have thought of visiting a major cultural institution, can explore their collections without having to brave the intimidations of architecture or the questioning of gatekeepers, however well-intentioned. Nor is the digital simply a replacement for books or microfilms. The application of OCR and full-text search to large text collections such as Trove's digitised newspapers may now seem commonplace, but that does not make it any less transformative (Hitchcock, 2008). The change is both quantitative and qualitative. More people are using more resources, but they are also using them differently—navigating patterns, traces and fragments in a way that would be impossible for even the most hardened microfilm operator (Putnam, 2016).

Using digital technologies, GLAM institutions can expose the vast number of collection items that will never make it into physical exhibitions. Online collection databases give people the freedom to ask their own questions and embark on their own adventures of discovery (Cameron & Robinson, 2007). Likewise, institutional authority can give way to new modes of collaboration, as demonstrated by the growing

proliferation of online crowdsourcing projects in the cultural heritage domain (Ridge, 2014). An increasing number of institutions are providing openly licensed, and easily downloadable, high-resolution collection images (Kapsalis, 2016). Opportunities are being opened, not simply to consume collections as audiences or visitors, but to create with them.

Collections are being encountered, shared and used beyond the frame of the institutional website. Instead of waiting for visitors to arrive, Twitter bots set objects loose in places where people already congregate. Services such as Trove, DigitalNZ, the Digital Public Library of America and Europeana bring together millions of items from GLAM organisations to open new avenues for discovery. However, the purpose of large-scale aggregation is not simply to build better search interfaces. These services provide new platforms for sharing, collaboration, enrichment and re-use of cultural heritage collections (Sherratt, 2013). Aggregation enlarges the scope and meaning of access.

There is no doubt that digital technologies have changed the way we find and use cultural heritage collections. As the previous examples demonstrate, online access has opened GLAM collections around the world to new audiences and new questions. However, as the number of online collection items continues to grow, as exciting new interfaces emerge, as more and more organisations share their collection data, it is all too easy to view access as a technology-fuelled march towards some ideal of openness—towards the fabled 'seamless online experience' in which the riches of our cultural institutions are arrayed for easy consumption.

But seams are not simply obstacles to a smooth user experience; they are reminders that our online services are themselves constructed (Sherratt, 2015b). How and why does information become 'open'? And when does it remain 'closed'? By focusing on the technological drivers, we obscure the resourcing decisions, ethical judgements, political controls and historical processes that define the boundary between open and closed and construct our experience of access.

The Limits of Online Access

As Tara Robertson (2018) reminds us, 'not all information wants to be free'. Ethical considerations around privacy and consent should inform decisions about what to digitise. Australian GLAM institutions, for example, generally recognise that access to Indigenous cultural collections should be subject to community consultation and control. The ATSILIRN Protocols, first published in 1995 and updated in 2012, state that access to secret, sacred or sensitive materials requires careful management in

the online environment. Digital knowledge management systems such as Ara Irititja, Keeping Culture and Mukurtu have been developed in consultation with Indigenous communities to provide culturally appropriate controls over access. Kimberley Christen (2012), one of the developers of Mukurtu, argues against 'false choices' between open and closed systems, and notes that 'general calls for "open access" undo the social bearings of information circulation and deny human agency'. Access can be withheld for good reasons.

Digitisation itself has a history, rooted not just in recent technological developments, but in much earlier efforts to expand the reach of access. As Tim Hitchcock (2016) points out, early targets for digitisation were canonical texts microfilmed by commercial firms in the predigital age:

> In other words, what happened in the twentieth century—the aspiration to create a particular kind of universal library, and to commercialise world culture (and to a 1930s mind, this meant male and European culture)—essentially shapes what is now available on line.

However, decisions about what to digitise are only the latest in a series of selections, omissions, erasures and accidents that have shaped the holdings of our cultural institutions. 'As spaces of power', Rodney Carter (2006) argues, 'the archive is riddled with silences'. Collections are formed by exclusion—by decisions about whose lives and voices matter. Online access is built atop generations of absence and loss. It comes with a responsibility to consider whose experiences are missing from our list of search results. As Lara Putnam (2016) suggests, we need to 'size up the absence'.

The ability to type a few words in a search box and find relevant resources still seems miraculous. But we do not know what we cannot find. The apparent omniscience of online discovery systems is maintained by their ability to hide their biases and failures. The reality is different. Safiya Noble (2018) has shown how search algorithms reinforce inequality. Ian Milligan (2013) has pointed at the limitations of OCR when applied to historical texts. Search interfaces lie, OCR is flawed and metadata is incomplete and inconsistent.

Discovery is just the starting point. Having found an item of interest, what can you do with it? How easy is it to share a persistent link, or download a high-resolution image? Can you obtain catalogue data in a machine-readable form for large-scale analysis? Online interfaces make assumptions about the needs and desires of users. They do not merely provide access; they construct it by defining the types of interactions we can have with collections.

It is sometimes claimed that the growth in online collections has resulted in a 'democratisation' of access. More resources are available to more people. But what we see, and how we see it, are the result of decisions made by someone else. What power do we really have?

Most of Trove's digitised newspapers were published before 1955. This is not because of a lack of content or resources, but because of restrictions imposed by copyright legislation (Sherratt, 2015a). Similar stories can be told around the world (Terras, 2015)—copyright skews online access away from contemporary history. The protection of property constrains our vision of the past.

Legislative barriers are frequently erected around archives. Australian government records are expected to be opened to the public after 20 years. But there are limits. Records created by courts and the parliament itself are treated differently, and the *Archives Act 1983* defines a series of exemptions that can be invoked to withhold records (National Archives of Australia, n.d.). Most countries impose limits on what we can know about the workings of government—the right of access is defined within systems of classification and control. Political power, bureaucratic processes and professional practice all play a role in determining what we can access.

Online collections have a history. Digital access is the product of analogue processes—of institutional policies and individual judgements. Our search results are not manufactured by algorithms alone. They are created by many small acts of human imagination, initiative, obstruction and neglect. But if the development of online access is not an onwards march towards some ideal of 'openness', what is it? How can we track, analyse, understand or change it?

Many coding languages provide mechanisms for 'introspection'. This allows us to discover something about the properties of a program as it is run. Interfaces to cultural heritage collections offer similar opportunities to peer inside and observe the processes that deliver resources to our browser. Perhaps the simplest example is the number of 'total results' that most search interfaces display. While we cannot take this at face value—among other things, it is dependent on the configuration of the search index—it does tell us something about the collection as a whole. And that is a convenient place to start.

Search Engines Lie

I created the QueryPic (2012) tool to visualise searches in Trove's digitised newspapers. You enter keywords as you would in the normal web interface. However, instead of viewing a list of search results,

you see a chart showing the number of matching articles, year by year. QueryPic lets you look for patterns across the complete newspaper corpus.

One QueryPic chart examines the question, when did the 'Great War' become the 'First World War'. It is a good example of how the tool can be used to track changes in language. However, if you look closely, you will notice a small bump in the usage of 'First World War' around 1916. How could this be? If you dig down through QueryPic to the relevant articles, you will find that 'First World War' appears not in the newspaper text, but in tags added by Trove users. By default, Trove searches user tags and comments as well as the articles themselves.

Digital access to cultural collections is not just delivered by websites. An increasing number of cultural heritage institutions provide direct access to collection data and images through application programming interfaces (APIs) or downloadable files. Such data are said to be 'machine-readable' or 'machine-actionable'—instead of being displayed as a product for human consumption, like a web page, they are represented in a form that computers can understand and manipulate. Made available in this form, collection data can support the development of new research, applications and analyses. The *Collections as Data* (2017) project seeks to develop guidelines, requirements and use cases for the sharing and re-use of these sorts of data.

QueryPic is built using the Trove API. It is a simple example of what becomes possible when cultural institutions make their data available for re-use. However, as the 'First World War' example reveals, machine-readable data inherit the limits and biases of the systems and processes that created it. Trove's search indexes are tuned for easy discovery, not large-scale data analysis. There is currently no way to exclude tags and comments from a search using the API. In some ways, this compromises the usefulness of QueryPic. It certainly highlights the need for visualisations to be carefully interrogated. You could also argue that this 'blip' or bug reveals features of the system that are otherwise difficult to see.

Safiya Noble's (2018) analysis of the ways in which search engines reinforce existing prejudices began with a simple Google search for 'black girls'. Matthew Reidsma's (2016) examination of bias in the related topic suggestions of the Summon discovery service began by simply logging user search terms against the topic suggestions. In both these cases, existing search interfaces were turned upon themselves. Through thoughtful observation and experiment, the interfaces offered information about their own biases.

The accuracy of OCR is a major issue for services like Trove that offer full-text search across large collections of historical documents. Trove, unlike some commercial services, does at least expose the results of its OCR processing so you can get a feeling for how messy it is. But how does this messiness affect your search results?

Here is a simple experiment to try right now. Go to Trove's digitised newspaper interface and search for 'tbe'. How many results are there? A little browsing will quickly reveal that 'tbe' is a failed OCR attempt at 'the'. When I last tried this search, 'tbe' was found in 14,996,286 articles (about 7% of the total). You could easily repeat this search across particular newspapers, locations or periods to see how error rates compare.

The 'truth' of search engines is formulated within the limitations of technology and content. Yet they express no doubt and offer no qualification. They encourage us to believe that they are comprehensive and accurate.

The proceedings of Australia's Commonwealth Parliament from 1901 onwards are all available online. The original volumes of Hansard have been scanned, and the text marked up in XML—one file for every sitting day. All this content is searchable through the ParlInfo database, but the interface is not easy to use. For example, it is difficult to simply browse a day's proceedings. In 2016, I wrote a computer script to search ParlInfo and download all of the underlying XML files from 1901 to 1980. The harvested files were shared through a GitHub repository (Sherratt, 2016a), enabling researchers to undertake large-scale analysis of political language and events.

However, there was a problem. The filenames of some XML files did not conform to the standard pattern, and when I looked inside, I found they were empty. This meant that some sitting days did not appear in search results—they were effectively invisible. After further investigation, I discovered that 94 days were missing (Sherratt, 2016b). The gaps were most pronounced in the Senate between 1910 and 1919. For example, 21 of 47 sitting days in 1917 did not appear in search results. Anyone relying on ParlInfo for research into political responses to the First World War would have missed significant slabs of content. The Parliamentary Library has since replaced the empty XML files and is undertaking further analysis.

On another occasion, while trying to update details of Australian Security Intelligence Organisation (ASIO) surveillance files that I had previously harvested from the NAA's online database, I discovered that about 400 files had disappeared from the public search interface. The NAA explained that this was an unintended consequence of a recent reorganisation of their holdings.

These things happen. Complex processes fail. The point is not to apportion blame, but to recognise failure as an inevitable part of the experience of online access. While it is unlikely that the missing files would have been noticed by someone using the standard search interfaces of either system, traces remained. These traces tripped up my harvesting programs. Just like the blip in QueryPic, they pointed to anomalies—they helped me 'size up the absence'.

Search interfaces to cultural heritage collections are not simply services to be consumed; they construct our experience of access. But they also offer glimpses of their inner workings that we can observe and change. What happens when we approach search engines as sites of experimentation and play? Or when we read them as historical documents awaiting close analysis?

Not Everything Is Digitised

What do 200 million newspaper articles look like? Or 300 km of government records? Part of our problem in understanding access is the difficulty of grappling with the scale of our cultural collections.

As Mitchell Whitelaw (2015) argues, the view offered by search boxes is a narrow, miserly slice of our rich collections. He and others have been exploring the possibilities of 'generous interfaces' that encourage exploration by putting collection items up front. Such experiences are no less constructed than a set of search results and can just as easily deceive their users. However, they offer a different approach to the challenges of scale—trusting users to interpret big, abstract pictures, instead of just consuming a stream of bite-sized chunks. Big pictures can prompt us to ask different types of questions, to address the meaning of a collection *as a collection*.

Here is one way of seeing 200 million newspaper articles.

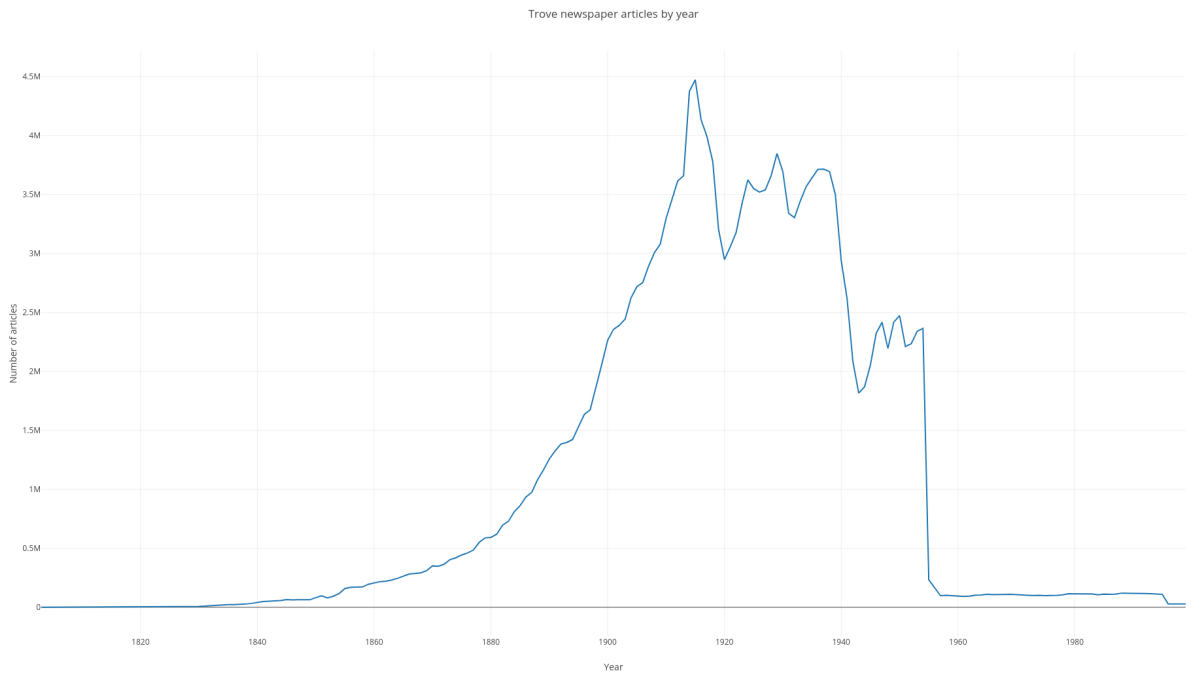

Figure 1: Number of digitised newspaper articles available in Trove by year.

This chart simply shows the number of digitised newspaper articles per year in Trove. You can make your own version of this chart using QueryPic. Simply paste 'http://trove.nla.gov.au/newspaper/result?q=+' into the 'Query url' box and select the 'number of articles' view. Of course, depending on when you undertake this experiment, your results might be quite different. New articles are added to Trove all the time. Has the picture changed?

In 2017, when I constructed this version, there was a significant spike in the number of articles around 1915. Why? Did something notable happen in 1915? No—it is not a result of the war … or at least not directly. It is a product of funding and priorities. In the lead-up to the centenary of the First World War, collaborating libraries decided to focus digitisation dollars on newspapers from the war period. Eventually, the spike should flatten out as other gaps are filled, but it is a useful reminder of how online collections are shaped by politics and practicalities.

Creating a big picture of the NAA's holdings is rather more challenging. Currently, the NAA only provides API access to some First World War service records—once again in support of the Anzac centenary. To obtain any other data from RecordSearch, the NAA's online database, it is necessary to reverse engineer the interface and extract structured information from web pages. This process is known as screen scraping. Over the years, I have scraped large amounts of data and digitised page images from

RecordSearch, sharing both the results and the code (Sherratt, 2012b, 2016d). Screen scraping is inefficient and prone to error, but it is also an example of how access can be negotiated and changed *from the outside*—web pages can be transformed into data; online collections can be opened to computational analysis.

In late 2016, I harvested data from 63,711 series in RecordSearch and aggregated information about the numbers of files described and digitised in each series. About a third of all series have at least some item descriptions and about a fifth have some items digitised. But how is this distributed across the whole collection? Item descriptions make records findable. Digitisation provides instant access. Together they determine the shape and texture of the collection as it is experienced online.

Total series	63,711
Series with item descriptions	20,735 (32.5% of series)
Series with digitised items	4,634 (22.3% of series with descriptions)
Total items described	10,727,214
Total items digitised	1,769,967 (16.5% of items described)

To explore the way access is distributed across the National Archives, I categorised each series using top-level government functions (Sherratt, 2017). These functions are defined in a thesaurus maintained by the NAA, and include things like transport, employment and immigration. In RecordSearch, functions are associated with government agencies rather than individual series, so there is some fuzziness in my groupings. Nonetheless, it is possible to create some collection-level snapshots.

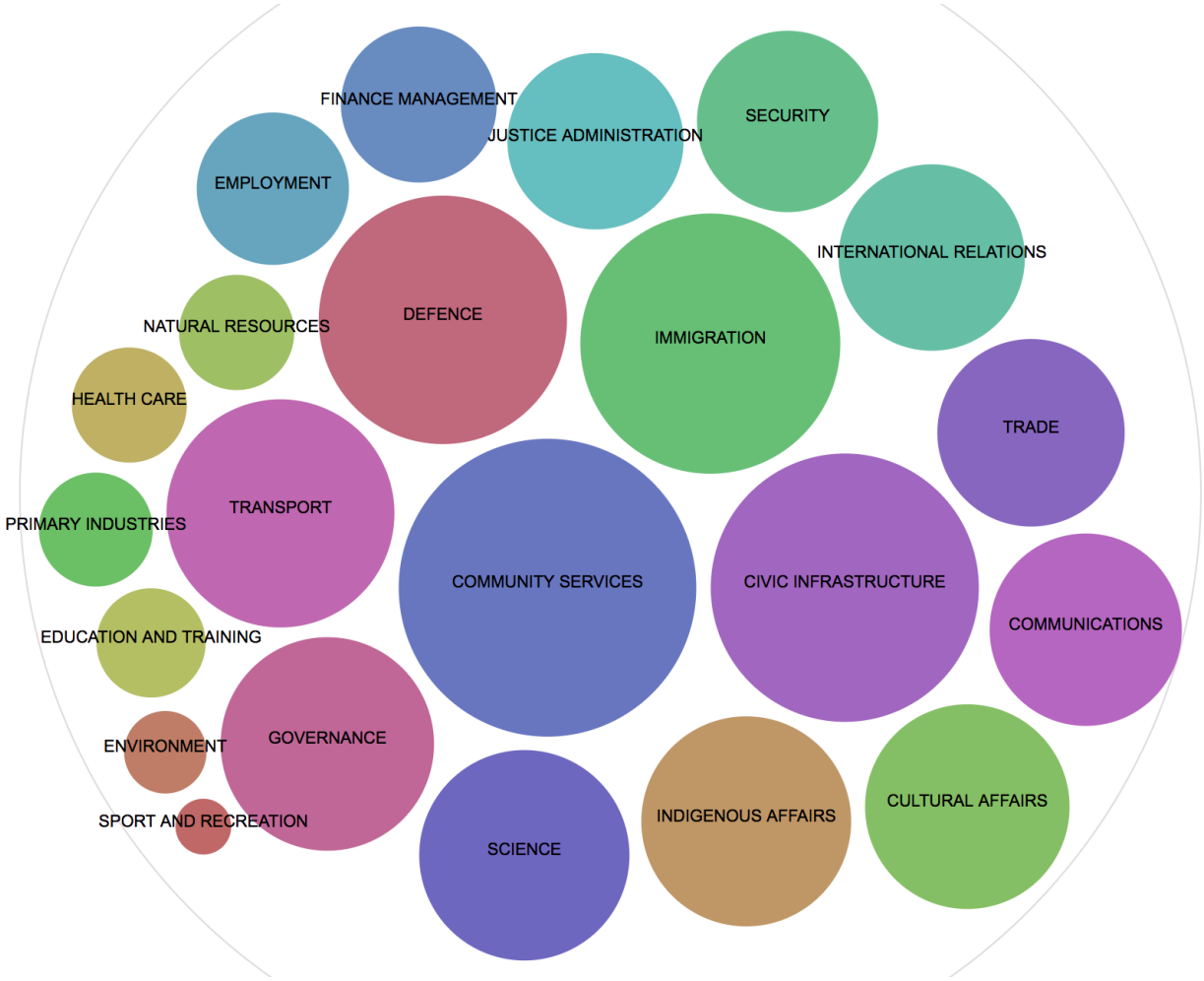

Figure 2a: Top-level functions in the National Archives of Australia by quantity (shelf metres).

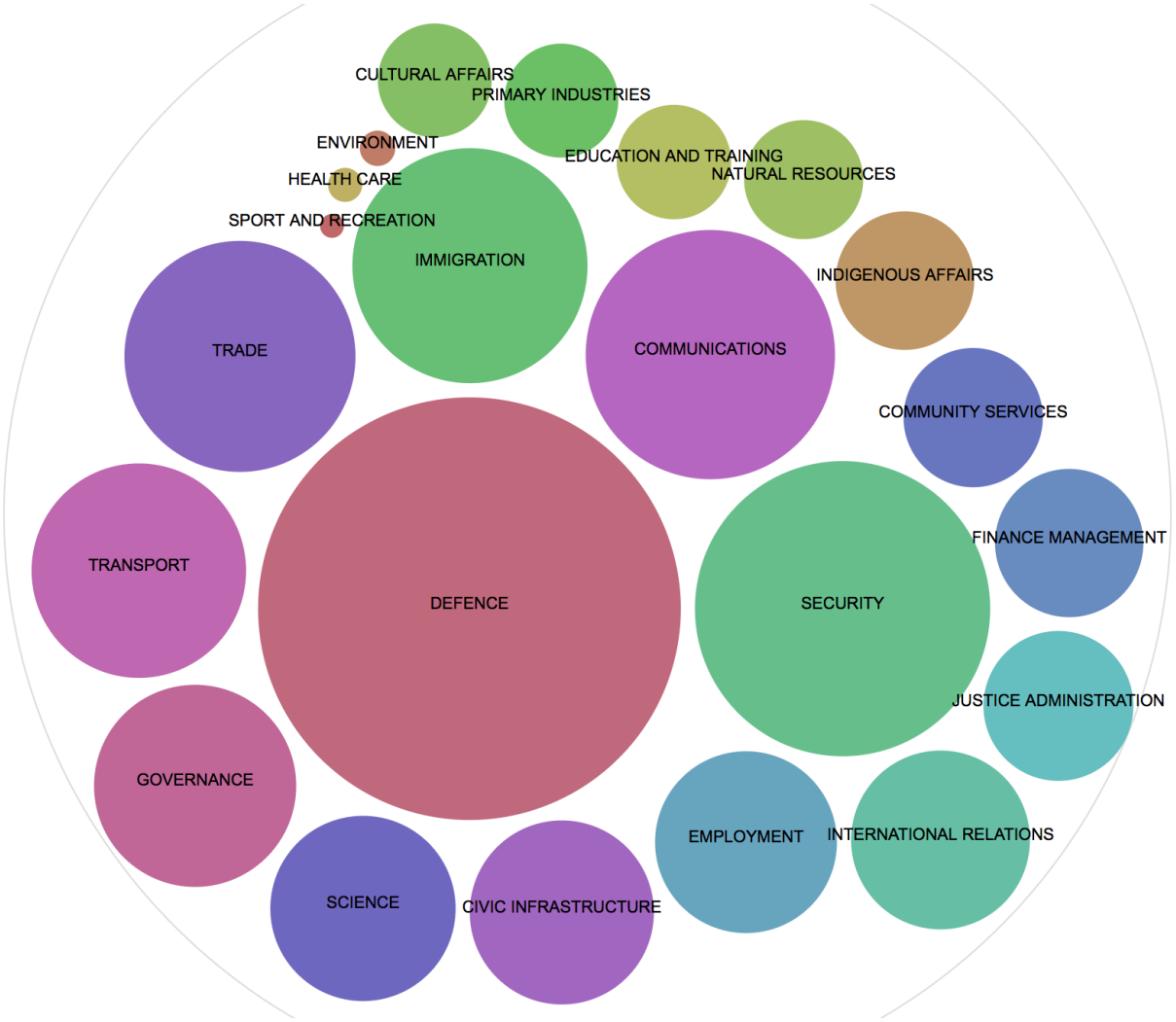

Figure 2b: Top-level functions in the National Archives of Australia by number of items digitised.

If we view the subject groupings by the amount of records in each (measured in shelf metres), the distribution seems fairly even. But the picture changes when we focus on the number of items described or digitised. ‘Defence’ becomes particularly prominent, while areas such as ‘community services’ and ‘cultural affairs’ seem to shrink. The prominence of defence is really no surprise. Service records are heavily used by family historians, and in 2007 the Australian government funded the digitisation of all 375,000 First World War service records in what was branded ‘A Gift to the Nation’. This one investment has had a significant and enduring impact on access to records held by the NAA.

There will always be priorities in digitisation programs. There will always be short-term funding opportunities related to specific initiatives or events—and there is nothing wrong with that. It is just that these biases and distortions are not obvious to someone typing queries into a search box.

Marilyn Lake (2010), Carolyn Holbrook (2017) and others have described how Australian government funding of educational resources related to the First World War has promoted a particular vision of Australian history. Digitisation of war-related collections needs to be examined in this context—what perspectives are privileged by easy online access?

In many countries, the boom in family history has had an impact on digitisation priorities, encouraging new collaborations with commercial providers like Ancestry (Kriesberg, 2017). Access to certain collections has a distinct dollar value, a capacity for on-sale to eager genealogists. As Barbara Reed (2014) asks, do organisations understand the compromises they are making by entering commercial arrangements? Do *we* understand how such priorities affect the range of stories we can tell about the past?

Digitisation shapes our perceptions of reality. The more we have in digital form, the easier cultural heritage collections are to find and use, the more likely we are to assume that everything (or at least everything important) is online. Researchers are used to asking questions about context and completeness when working with historical sources, but the power and scale of digital technologies draws attention away from their limits. To overcome this, we need to take scale seriously and analyse digitisation priorities in the context of a bigger, unknowable whole.

Access Is Not Always Open

The first of January in Australia each year has become an annual celebration of access. Both national and state archives release previously closed files for public scrutiny, and the media fills a slow news day with secrets from governments past. The opening of a new batch of cabinet records by the NAA attracts particular attention. However, the media coverage generally overlooks that this is a routine bureaucratic process governed by archives legislation.

With each new year, the 'open period' (now set at 20 years) creeps forward, and many more records are potentially available to the public. The cabinet documents are given a head start, pushed through the process of access examination in preparation for their big day. But most records have to wait for a researcher to request them before they are considered for release. Access examination checks the contents of a record against a list of exemptions defined by the *Archives Act*. These are reasons why files should not be made public (e.g. privacy or national security). Most records pass this test and are opened, or partially opened, for all to see. But a small percentage remain closed.

Here is another experiment to try. Go to RecordSearch on the NAA website and select the 'Advanced search' for 'Items'. Go to 'Access status' near the bottom of the search form and choose 'Closed'. RecordSearch will display details of the records you are not allowed to view.

Since 2016, I have been undertaking my own new year's ritual—harvesting details of all files in RecordSearch with the access status of 'closed'. In January 2018, there were 11,235 closed files (Sherratt, 2018). How many did your search return? I also compile and share lists of files that were newly closed in the past year. It seems only fair that the celebrations surrounding the cabinet release should spare some time to remember the files that did not make it through. Access is about what we cannot see as well as what we can.

You might have noticed that RecordSearch not only records the access status of a file, but the date of and reasons for the decision. Usually the reasons are specific exemptions defined by the *Archives Act*, but there are some additional categories. I used the 2016 harvest to create a new interface that lets you look for patterns and connections in a way that is impossible within RecordSearch itself (Sherratt, 2016a). Here, for example, is a chart that shows the number of closed files associated with each 'reason'.

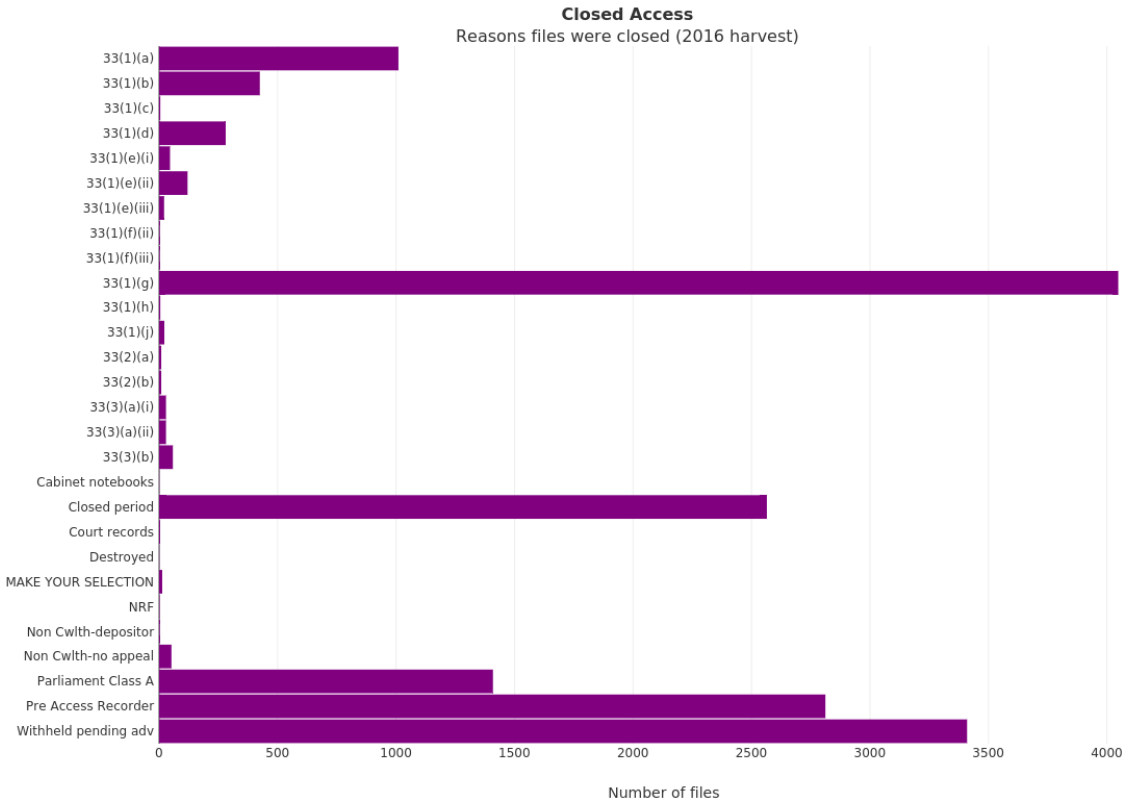

Figure 3: Reasons cited for closing files in the National Archives of Australia.

The most commonly cited reason is Section 33(1)(g) of the *Archives Act*, which relates to individual privacy. Two other heavily used categories, 'Pre access recorder' and 'Withheld pending adv', are not defined anywhere in the Act. 'Pre access recorder' was used on records that had been closed before the introduction of the *Archives Act* in 1983. It all but disappeared from the 2017 harvest, as the NAA changed the access status of these files to 'Not yet examined'.

'Withheld pending adv' tells a more complex story. It is used when records are referred back to the government agencies that created or controlled them for advice on whether they can be made public. This process can take months or even years, so 'Withheld pending adv' is used as a marker to indicate that a file is part way through the examination process. It is closed, but not finally closed. The use of this marker is evidence of a glitch in the system. The Act states that access decisions will be made within 90 days—there is no provision for extended consideration by agencies. This will be changed by proposed amendments to the archives legislation; thus, the category might disappear from future harvests.

How closed is 'closed'? Files marked as 'Withheld pending adv' sit in a sort of archival limbo—neither open nor closed. Using data from the 2018 harvest, I was able to identify which files were finally released to the public in 2017 and calculate how long they had been waiting. The average was three years and 77 days (Sherratt, 2018). They might not be finally closed, but they are effectively closed.

In the case of the NAA, the meaning of access is defined by legislation. It is assumed that records older than 20 years will be opened to public scrutiny—the justifications for withholding access are called 'exemptions' for a reason. But that is not the end of the story. Access examination is a complex process involving bureaucratic practice, individual interpretation and the public right to access. By making regular harvests, I am hoping to expose this as a historical process—to identify changes over time in the way access is constructed.

The data I am harvesting is all publicly available through RecordSearch. However, the existing interface does not allow you to aggregate information about sets of files. Like Trove, its purpose is discovery, not analysis. When we free data from existing collection interfaces, new possibilities emerge.

The NAA holds many thousands of records documenting the workings of the *White Australia Policy*—a system of immigration control designed to keep Australia 'white'. By extracting portrait photographs

from identification documents in the NAA, I was able to create *The Real Face of White Australia*, showing the faces of people who lived under this racist system of surveillance and control (Sherratt & Bagnall, 2018). It is both a visualisation and a discovery interface—a way of navigating the records though the people inside.

The History Lab at Columbia University has assembled a huge database of declassified government documents. Many of these include redactions—sections of text blacked out for security reasons. By identifying redactions and comparing redacted and non-redacted copies of the same documents, the History Lab team is revealing new patterns in state censorship (History Lab, n.d.). Similarly, I extracted many thousands of redactions from the surveillance files of ASIO and turned them into a discovery interface (Sherratt, 2016c). The redactions became a gateway to the files they were intended to obscure.

One way of exploring the meaning of access is to see what collections look like when we turn them inside out. What happens when we focus on what is closed or censored, when we look at what we are not allowed to see?

Hacking Access

It is easy to get excited about the possibilities of online access for cultural heritage collections—new audiences, new uses and new opportunities to demonstrate value and relevance. But there will always be limits. The idea that our efforts are aimed towards a 'seamless' online experience that brings everything together is a dangerous mirage.

In *The Theory and Craft of Digital Preservation*, Trevor Owens (2017) warns 'whatever discovery system or interface you use today is temporary'. Instead of relying on a single point of access, Owens argues for 'multimodal access and use', where collection data is shared in a variety of forms, and new interfaces are created both by GLAM institutions and their users. This approach engages with the complexities and contradictions of access. There is no solution, no off-the-shelf system—all we can do is create, play, build and critique to explore access in the making. As Bethany Nowviskie (2016) suggests in her talk on 'speculative collections', there is an opportunity to shift the temporal orientation of our libraries and archives away from a closed and linear past towards an exploration of what might be.

This is not a job for cultural institutions alone. Collection users need to view themselves as more than simply the beneficiaries of access. Researchers generally accept that their use of primary source material comes with an obligation to critically engage with its context and meaning. Why should such obligations

diminish online? Contexts are multiplied through digitisation, aggregation and indexing. We should treat interfaces as archaeological sites, digging down through layers of technology, descriptive practice and institutional history to understand what is delivered so conveniently through our browsers.

This chapter has provided some examples of how these sorts of excavations might start—from simple experiments using the search box, through to large-scale data harvests, and the creation of new interfaces. This is not intended as a structured research program, but as an invitation to start hacking. Mark Olson (2013) describes a 'hack' as something that 'transforms the effectivities of socio-technical systems, making them work, or *un-work*, often in new and unexpected ways'. Olson explores how adopting a hacker ethos can enlarge the field of humanities practice:

> A hack can be elegant or kludgy, authored from scratch or patched together and remixed—the important thing is getting things done, pushing the boundaries of what the humanities can do, what effects it can have in the world, and where.

Hacking the systems that construct and control access to our cultural collections is at the core of humanities practice in the early 21st century. As online collections continue to expand, we need to carve out spaces that resist the weight of scale and foster alternative perspectives. As interfaces grow in sophistication and complexity, we need to stage playful and pointed interventions that reveal their limits and empower critique. We do not all have to be coders, but we do have to take code seriously. We have to take what we are given by collection databases and change it.

Hackers might work inside or outside a cultural institution—this is not about us and them. This is about recognising that for all the resources, intelligence, skill and care that institutions invest in their online resources, these resources will never be perfect, they will never be finished, they will never be open. Owens, Nowviskie and others have situated work on digital collections within an ethics of maintenance and care. We are in this for the long haul. Once we give up the dream of universal access, we can admit the limitations of our systems and set about the never-ending work of repair.

References

ATSILIRN. (2012). *Aboriginal and Torres Strait Islander protocols for libraries, archives and information services*. Retrieved from http://atsilirn.aiatsis.gov.au/protocols.php

Ayres, M. (2013, July). *Singing for their supper: Trove, Australian newspapers, and the crowd*. Paper presented at the IFLA WLIC 2013, Singapore. Retrieved from http://library.ifla.org/245/1/153-ayres-en.pdf

Cameron, F. & Robinson, H. (2007). Digital knowledgescapes: Cultural, theoretical, practical, and usage issues facing museum collection databases in a digital epoch. In F. Cameron & S. Kenderdine (Eds.), *Theorizing digital cultural heritage: A critical discourse* (pp. 165–191). Cambridge, MA: MIT Press.

Carter, R. G. S. (2006). Of things said and unsaid: Power, archival silences, and power in silence. *Archivaria*, *61*(Spring), 215–233. Retrieved from https://archivaria.ca/index.php/archivaria/article/view/12541

Christen, K. (2012). Does information really want to be free? Indigenous knowledge systems and the question of openness. *International Journal of Communication*, *6*. Retrieved from http://ijoc.org/index.php/ijoc/article/view/1618

Collections as Data. (2017). Home. Retrieved from https://collectionsasdata.github.io/

Department of Industry, Science and Tourism. (1995). *Innovate Australia*. Retrieved from http://webarchive.nla.gov.au/gov/19961102131909/http://www.dist.gov.au/events/innovate/itt.html

Department for Culture, Media and Sport. (2016). *The culture white paper* [White paper]. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachmentdata/file/510798/DCMSTheCultureWhitePaper3.pdf

Hitchcock, T. (2008). Digital searching and the re-formulation of historical knowledge. In M. Greengrass & L. Hughes (Eds.), *The virtual representation of the past* (pp. 81–90). Farnham, UK: Ashgate.

Hitchcock, T. (2016, 6 July). The digital humanities in three dimensions [Blog post]. Retrieved from http://historyonics.blogspot.com.au/2016/07/the-digital-humanities-in-three.html

Holbrook, C. (2017). Adaptable Anzac: Past, present and future. In *The Honest History Book* (pp. 48–63). Sydney, NSW: New South.

Kapsalis, E. (2016). *The impact of Open Access on galleries, libraries, museums, & archives*. Smithsonian Emerging Leaders Development Program. Retrieved from http://siarchives.si.edu/sites/default/files/pdfs/20160310OpenCollectionsPublic.pdf

Kriesberg, A. (2017). The future of access to public records? Public–private partnerships in US state and territorial archives. *Archival Science*, *17*(1), 5–25. https://doi.org/10.1007/s10502-016-9268-6

Lake, M. (2010). How do schoolchildren learn about the spirit of Anzac? In M. Lake & H. Reynolds (Eds.), *What's wrong with Anzac: The militarisation of Australian history* (pp. 135–156). Sydney, NSW: New South.

Ling, T. & McLean, A. (2004). Taking it to the people: Why the National Archives of Australia embraced digitisation on demand. *Australian Academic & Research Libraries*, *35*(1), 2–15. https://doi.org/10.1080/00048623.2004.10755253

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

National Archives of Australia. (n.d.). *Access to records under the Archives Act–Fact sheet 10*. Retrieved from http://naa.gov.au/collection/fact-sheets/fs10.aspx

National Archives of Australia. (2016). *Annual Report 2015–16*. Retrieved from http://www.naa.gov.au/about-us/publications/annual-reports/2015-16/index.aspx

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York, NY: NYU Press.

Nowviskie, B. (2016, 27 October). *Speculative collections*. Retrieved from http://nowviskie.org/2016/speculative-collections/

Olson, M. J. (2013). Hacking the humanities: Twenty-first-century literacies and the 'becoming other' of the humanities. In E. Belfiore & A. Upchurch (Eds.), *Humanities in the twenty-first century: Beyond utility and markets* (pp. 237–250). Basingstoke, UK: Palgrave Macmillan.

Owens, T. (2017). *The theory and craft of digital preservation*. LIS Scholarship Archive. https://doi.org/10.17605/OSF.IO/5CPJT

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

Reed, B. (2014). Reinventing access. *Archives and Manuscripts*, *42*(2), 123–132. https://doi.org/10.1080/01576895.2014.926823

Reidsma, M. (2016, 11 March). *Algorithmic bias in library discovery systems*. Retrieved from https://matthew.reidsrow.com/articles/173

Ridge, M. (Ed.). (2014). *Crowdsourcing our cultural heritage*. Farnham, UK: Ashgate Publishing.

Robertson, T. (2018). Not all information wants to be free: The case study of *On Our Backs*. In P. D. Fernandez & K. Tilton (Eds.), *Applying library values to emerging technology: Decision-making in the age of open access, Maker spaces, and the ever-changing library* (Publications in Librarianship #72) (pp. 225–239). American Library Association. Retrieved from http://eprints.rclis.org/32463/

Sherratt, T. (2012a). *QueryPic*. Retrieved from http://dhistory.org/querypic/

Sherratt, T. (2012b). *recordsearch_tools*. Retrieved from https://github.com/wragge/recordsearch_tools

Sherratt, T. (2013, October). *From portals to platforms: building new frameworks for user engagement*. Paper presented at the LIANZA 2013 Conference, Hamilton, New Zealand. Retrieved from http://www.nla.gov.au/our-publications/staff-papers/from-portal-to-platform

Sherratt, T. (2015a). Asking better questions: History, Trove and the risks that count. In Phillipa McGuinness (Ed.), *Copyfight* (pp. 112–124). Sydney, NSW: New South Publishing. Retrieved from http://discontents.com.au/asking-better-questions-history-trove-and-the-risks-that-count/

Sherratt, T. (2015b). Seams and edges: dreams of aggregation, access, and discovery in a broken world. Presented at the ALIA Online 2015, Sydney. Retrieved from http://discontents.com.au/seams-and-edges-dreams-of-aggregation-access-discovery-in-a-broken-world/

Sherratt, T.2016a). *Closed access*. Paper presented at DH2015, Hobart. Retrieved from http://discontents.com.au/closed-access/

Sherratt, T. (2016b). *hansard-xml*. Retrieved from https://github.com/wragge/hansard-xml

Sherratt, T. (2016c). *Investigating the Hansard black hole*. Retrieved from http://timsherratt.org/research-notebook/historic-hansard/notes/investigating-the-hansard-black-hole/

Sherratt, T. (2016d). *recordsearch-series-harvests*. Retrieved from https://github.com/wragge/recordsearch-series-harvests

Sherratt, T. (2016e). *redacted*. Retrieved from http://owebrowse.herokuapp.com/redactions/Sherratt, T. (2017). *Viewing the NAA through functions*. Retrieved from http://timsherratt.org/research-notebook/aggregated-archives/notes/naa-functions-view/

Sherratt, T. & Bagnall, K. (2018, in press). The people inside. In K. Kee & T. Compeau (Eds.), *Seeing the past: Experiments with computer vision and augmented reality in history*. Ann Arbor, MI: University of Michigan Press.

Sherratt, T. (2018, 2 February). *Withheld, pending advice*. Inside Story. Retrieved from http://insidestory.org.au/withheld-pending-advice/

Terras, M. (2015). Opening access to collections: The making and using of open digitised cultural content. *Online Information Review*, *39*(5), 733–752. https://doi.org/10.1108/OIR-06-2015-0193

Whitelaw, M. (2015). Generous interfaces for digital cultural collections. *Digital Humanities Quarterly*, *9*(1). Retrieved from http://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html