

Turning the inside out

Author : Tim Sherratt

Tagged as : [access](#), [archives](#), [ASIO](#), [redactions](#), [surveillanceWhite Australia](#)

Date : October 24, 2016

Keynote presented at [Forging Links](#), the Annual Conference of the Australian Society of Archivists, Parramatta, 19 October 2016.

Please cite as: Tim Sherratt, 'Turning the inside out', presented at the Australian Society of Archivists Annual Conference, Parramatta, 2016, .

This is RecordSearch, but not as you know it.

This is my hacked version of [RecordSearch](#), the online collection database of the National Archives of Australia. Unlike the regular version it displays the number of pages in each file. But more interestingly, if you search in Series ST84/1 you see more than metadata - you see the people inside.

discontents

As Barbara Reid noted in her article <http://discontents.com.au>, [Passport control people overseas](https://www.abc.net.au/news/2015-08-28/passport-control-people-overseas/6711144) are imbued with people'. [Series ST84/1](http://discontents.com.au) goes by the fairly benign title, 'Certificates of Domicile and Certificates of Exemption from Dictation Test, chronological series'. But of course the Dictation Test was the administrative backbone of a racist system designed to exclude people who did not fit the widely-accepted vision of 'White Australia'. ST84/1 is full of people just trying to live their lives under the weight of the White Australia Policy.

The [certificates in ST84/1](#) allowed people, born or resident in Australia, to return home after travelling overseas. If your 'whiteness' was suspect and you had no certificate, you would be subjected to the Dictation Test, and you would fail. The certificates usually include photographs and handprints – they are compelling and confronting documents. But you have to dig through layers of metadata in RecordSearch to see that. Or do you?

About five years ago, [Kate Bagnall](#), a historian of Chinese Australia, and I were thinking about ways of drawing attention to these records. In a little over a weekend, I harvested about 12,000 page images from ST84/1 and ran them through a facial detection script. The result was ['The Real Face of White Australia'](#).

You may have seen it before. It's had a remarkable life, travelling around the world as an example of how we can use digital tools to see records differently. But of course the power is in the faces themselves – in the connections we make through time. We cannot escape their discomfiting gaze.

You may think that the certificates in ST84/1 are merely a form of identity document. But remember that in the early years of the 20th century, passports were still evolving, and the use of photographs and fingerprints for identification were generally confined to prisoners and criminals.

The sociologist Richard Jenkins talks about identity not as an essence, a noun, but as 'something that we *do*, a process of identification'.¹ But self-identification is constrained by broader systems of categorisation, or 'social sorting', that decide who belongs, who is a threat, who needs to be watched.

The records in ST84/1 were embedded within a system of surveillance that extended outwards from Australia's ports to the offices of shipping companies around the world, and inwards to anyone who seemed out of place in White Australia. Technologies of identification and surveillance do not simply enforce boundaries, they create them. Their existence demonstrates why they are needed. These records did not document identity, they defined it according to a set of racial categories.

Modern parallels are not hard to find. Last year in a bungled operation that became known as '#borderfarce', immigration officials planned to prowl the streets of Melbourne on the hunt for illegal immigrants. The focus of border surveillance once again turned inwards, to those who seemed out of place. Watching in horror as events unfolded on social media, I helpfully pointed people in Melbourne to a convenient source of identity documents.

In Melbourne this weekend? Get your Border Force identity papers here:
[#borderfarce #wap pic.twitter.com/j2TrkKEPkR](http://t.co/H3dIXHZy6q)

— Tim Sherratt (@wragge) [August 28, 2015](#)

The link I tweeted was not to RecordSearch, but to an [experimental interface](#) where Kate and I are continuing to think about ways of exposing the bureaucratic remnants of White Australia.

So far I've harvested metadata from more than 20,000 files and downloaded around 150,000 page images. Amongst other things, I'm working on an updated wall of faces.

Most recently I created a way of sorting and [viewing pages by their orientation](#) – by the ratio of height to width. Why? Kate wanted an easy way of finding birth certificates which, in this period, tended to be short and wide. It was a [simple little hack](#), but it revealed the records in a very

different way.

There's a big bunch of birth certificates, but there's also envelopes, photographs, and an assortment of slips and notes. It makes you think about how we see the written world through the frame of 'portrait' orientation. It's perhaps also worth noting that RecordSearch displays thumbnails as cropped squares - shape is subsumed to the regularity of the grid.

The records in our landscape view are no more accessible than they were before, but they can be accessed *differently*. In his contribution to the ASA's 30th anniversary symposium, Eric Ketelaar described the relationship between archival access and human rights - highlighting the importance of access not only to democratic accountability, but also to our rights as 'victims' of official surveillance: 'As human beings, subjected to the panoptic sort of governments and private enterprise, we have the right to know'.² Ketelaar concludes by stating that 'access is not the actual use of archives, access enables use'. I'd like to extend this a bit. What my own work reveals is the complex relationship between access and use. Not only does access enable use, use changes what we mean by 'access'.

My bio nowadays describes me as a 'historian and hacker'. 'Historian' describes my orientation to the world - I see the past in the present. 'Hacker' refers to the tools I use to make connections through time. Hacking is creative and positive, despite what the mainstream media might say - it's about finding solutions, exploring alternatives, and pushing the limits of what's possible.

My RecordSearch hack is just a [little piece of Javascript code](#) that you can install in your browser - it changes the pages as they load.

Userscripts, as they're called, allow anyone to alter the way webpages look and behave within their own browser. They give users greater control over their online experience, but they also open opportunities for experimentation. A lot of my own research is guided by questions like: What would happen if I...? What would it look like? What would change? What would I feel? Userscripts are one way of playing with the complexities of access.

Surveillance too can be hacked. If you're concerned about technologies such as facial detection you'll be pleased to know that thanks to the [work of artist Adam Harvey](#), not only can you confuse detection algorithms, you can make a dramatic fashion statement.

Gary Marx, a major figure in the field of surveillance studies, has catalogued the ways in which individuals can resist the growing encroachments of surveillance. Amongst possible tactics he identifies 'discovery' - the attempt to undercover the scope of surveillance.³ Access to records can empower such acts of everyday resistance, but in other ways surveillance and access are more alike than opposed. Both start from a place of concealment. Access cannot be given unless it is first restricted. Both depend on asymmetries of power. Decisions about what we can know are ultimately made by others. Access is as much a process of control, as it is an act of release.

This is not necessarily a bad thing. I think we can all agree there should be limits to access, particularly relating to individual privacy and cultural sensitivity. But just as 'identity' is defined

through acts of 'identification', so access is elaborated through instances of deployment and use. Access is not a state of being, it's a process to be negotiated. And so the question is, what can we know about what we can know?

On 1 January this year I harvested the metadata of [all the files in RecordSearch](#) with the access status of 'closed'.

These are records that the access examination process has determined should be withheld from public scrutiny. While the files themselves can't be seen, RecordSearch does tell us a fair bit about them, including when the decision was made and the reasons behind it. Unfortunately you can't search or filter on this data, so it's difficult to look for patterns within RecordSearch itself.

I've taken the data and loaded into [a new site](#) where you can examine it from a [number of different angles](#). You can explore the reasons why files were closed, the series they came from, the age of their contents, and the dates when decisions were made. It's more of a workbench than a discovery interface, and it's likely to change as I ask different questions of the data.

Of course, the outlines of the examination process, including the grounds for exemption, are defined by the Archives Act. So what more is there to know?

[Section 33 of the Archives Act](#) does indeed spell out 17 reasons why records can be withheld from public access. But [the data from RecordSearch](#) includes an additional 11 categories. Some, like ['Parliament Class A'](#), relate to other definitions under the Act. Others, like ['MAKE YOUR SELECTION'](#), tell us something about the RecordSearch interface. But two of the most heavily cited reasons - 'Pre-access recorder' and 'Withheld pending advice' are not defined under the Act or anywhere that I could find on the Archives' website.

Being an archivally-educated audience you can probably guess what these labels refer to, but if you need a little help you can look at when the access decisions in these categories were made.

The majority of decisions on ['pre access recorder'](#) files were made *before* the introduction of the Archives Act in 1983. I checked this with the Archives and they confirmed that these records were examined before the existence of the Act. They explained that 'pre access recorder' was used when the original exemptions couldn't be mapped to those later defined under Section 33.

Conversely, most decisions on [‘Withheld pending advice’](#) files were recorded in the last five or six years. If you look at the series that contain the most files citing this reason you can see that almost half come from [A1838](#) – DFAT’s main correspondence series. As I’m sure you’ve realised, these are files that have been referred back to agencies for advice. And DFAT has been particularly slow in responding. They’re listed as ‘Closed’ on RecordSearch even though their access status has not been finalised. They’re not, however, [included in the count of ‘closed’ files](#) that the Archives reports in its annual summary of access outcomes.

This is probably fair enough, but if you search the closed files you can see that 1,467 of them files have been waiting for more than three years for a final decision. They might not be officially closed, but for a PhD student wanting to see them they are effectively closed.

My point here is not to be critical of the National Archives, or even of DFAT. What I’m interested in is the inevitable gap between legislation and practice. Access examination is subject to a range of influences and constraints, resourcing amongst them, and needs to be understood not as the application of a set of rules, but as a process that is historically contingent. A human process.

Earlier this year I harvested [several gigabytes of parliamentary proceedings](#) from the Australian Parliament’s ParlInfo database and created my own version of [Historic Hansard](#).

As you do.

In the process of harvesting the files [I discovered that data was missing](#) for about 100 sitting days – most of them from the Senate between 1912 and 1919.

There’s no conspiracy at work here, it’s just some sort of processing error. However, Parliament staff weren’t aware of the problem, and it’s unlikely that anyone would’ve noticed using the web interface. You can’t find what you can’t find. Fortunately, Parliament staff are now working on a fix, but if you’ve been relying on ParlInfo for access to debates relating to World War I, you might want to do some more checking.

These things happen. Systems go wrong. Mistakes are made. Again what interests me is not finding who’s to blame, but exploring the gap between design and outcome, between ideal and reality. This is the gap where access is made and experienced. A gap that can only be understood through the complexities and contradictions of use. Access does not exist until its limits are tested. It’s not a process of opening, it’s a constant ongoing struggle over the very meaning of ‘open’.

And that’s a good thing.

We’re here at this conference to explore the possibilities of ‘forging links’. But of course collaborations don’t have to be comfortable to be constructive. The struggle over access may sometimes be tense, frustrating, and annoying, but it is also productive. Users of archives do not just consume access, they create it.

About a year ago I fired up my RecordSearch harvester and downloaded the metadata and page images for [most of the ASIO \(Australian Security Intelligence Organisation\) files publicly available](#) through the National Archives. I ended up with about 70gb of images. These are mostly dossiers on individuals and organisations – odd collections of gossip, published articles, and records of surveillance.

I’ve made this data available for anyone who wants it. Some of the images were recently used in the GovHack open data competition to create the [‘Cute Commies’](#) site.

To be honest, I didn't have a clear purpose in mind when I harvested the data. It was another one of those 'What would happen if?' moments. I was, however, thinking generally about possible points of comparison between the ASIO files and the archival remnants of the White Australia Policy. Both built systems of identification, classification, and surveillance in which recordkeeping was crucial.

Kate and other historians of Chinese Australia have noted that the administration of the White Australia Policy was not uniform or consistent. Similar cases could result in quite different outcomes depending on the location and those involved. Understanding this is important, not only for documenting the workings of the system, but for recovering the agency of those subjected to it. Non-white residents were not mere victims, they found ways of negotiating, and even manipulating, the state's racist bureaucracy. In her work on colonial archives, Ann Laura Stoller identifies this 'disjuncture between prescription and practice, between state mandates and the manoeuvres people made in response to them' as part of the 'ethnographic space' of the archive.⁴

How do we explore this space? One of the things I've found interesting in working with the closed files is the way we can use available metadata to show us what we can't see. It's like creating a negative image of access. Kate and I have been thinking for a number of years now about how we might use digital tools to mine the White Australia records for traces, gaps, and shadows that together build a picture of the policy in action. Who knew who? Who was where and when? What records remain and why?

The workings of ASIO, on the other hand, are deliberately obscured. Many of the files in the Archives include a note explaining why details have been withheld. [Some warn](#) that the 'public disclosure of information concerning the procedures and techniques used by ASIO' would enable people of interest to formulate counter-measures 'based on an analysis of ASIO modus operandi'. David Horner's recent history of ASIO notes that he was required to remove ASIO file references from his footnotes 'because of the nature of ASIO's filing system, which itself is classified'.⁵ We don't even know how many files ASIO has on people and organisations, although David McKnight suggests that it's somewhere in the hundreds of thousands.⁶ My harvest includes about 12,000 files.

Just like systems of racial classification, intelligence services exist within a circle of self-justification. The fact they exist proves they need to exist. We are denied information that might enable us to imagine alternatives. And yet as limited as the provisions under the Archives Act are, we do have access.

How can we use this narrow, shuttered window to reverse the gaze of state surveillance and rebuild a context that has been deliberately erased. Just as with Closed Access and the White Australia records can we give meaning to the gaps and the absences? Can we see what's not there?

This is one of the questions being explored by [Columbia University's History Lab](#). They've created the Declassification Engine - a huge database of previously classified government documents that they're using to analyse the nature of official secrecy. By identifying non-redacted copies of previously redacted documents, they've also been able to track the words, concepts and events most likely to be censored.

The History Lab's collection of documents on foreign policy and world events is rather different to ASIO's archive of the lives, habits and beliefs of ordinary Australians. But I'm hoping that they too can tell us something about the culture that created them.

I'd intended to have a wonderfully compelling suite of examples and arguments to demonstrate today, but time has run short. Instead I have a set of half-baked experiments which sort of look a bit interesting. But perhaps that's better. It's important to me to try and be open about my own processes. I share my code and data, and I've started documenting most of what I'm up to in a [open research notebook](#). If access is a struggle, then we should be sharing our stories of loss and

frustration, and not merely celebrating our victories.

All of these experiments are online in some sort of form. So please explore.

Experiment A is nothing more than a [browse interface to all the digitised records](#) I've harvested. It's just a clone of my work with the White Australia records, but I think there's real conceptual power in the ability to browse.

Experiment B started with a problem. From RecordSearch I could harvest data on access status and find out how many ASIO files were in each of the three categories - Open, Open with Exception, and Closed. But how much of the 'Open with Exception' files are actually open?

Most of the files include a summary which tells you how many pages have been completely or partially exempted. That's great, but did I really want to open up 12,000 files and manually scan for summaries? By [playing around with the Tesseract OCR engine](#) I've created a simple filter that extracts text from the images and searches for words like 'exemption', 'archives', and 'folio'. I now have a good sized collection of summaries awaiting data entry...

Experiment C began as another attempt to quantify the scale of exemption. The summaries told me how many pages had redactions - bits of information like names and ids that are blacked out, or sometimes even cut out of the page. But if I could identify individual redactions I could both test the summaries and create a new measure of openness... or redactedness...

Looking for redactions in ASIO files — that looks hopeful... pic.twitter.com/heeWMAQiwe

— Tim Sherratt (@wragge) [September 11, 2016](#)

Through trial and error I developed a computer vision script that did a pretty good job of finding redactions - despite many variations in redaction style, paper colour, and print quality. It took a couple of days to work through the 300,000 page images, but in the end I had a collection of about 300,000 redactions. Unfortunately about 20 percent of these were false positives, so I spent a number of nights manually sorting the results.

But the false positives themselves are sort of wonderful... pic.twitter.com/Jn9l04bUaY

— Tim Sherratt (@wragge) [September 29, 2016](#)

My redaction finder still needs a lot of refinement, and plenty of errors have slipped through. But, within the files that are currently digitised, the scale of exemption seems about ten times greater than Margaret Kenna estimated when giving evidence to Parliamentary Joint Committee on ASIO in 2000. She thought every file contained about 10 exemptions 'be it a word or a folio or a paragraph'. I'm seeing an average of about 100 redactions per file.

I've started adding information about the size and position of the redactions to my database and aggregating this data by page. When I left Canberra, the script was still running, but you can [explore](#) the current standings in my top 50 lists of the [most redacted files](#) and [pages](#).

Once the data processing is completed you'll be able to filter files by the amount of area blacked out, or the total number of redactions. Many more opportunities to see what you can't see.

Experiment D was an attempt to build a composite image of all the redactions to visualise what

parts of a page were most likely to be removed – something like a heatmap. It [sort of worked](#), but by the time I'd added all the redactions I had nothing but a very large black blob.

My composite of 170,000 [#ASIO #redactions](#) has opened a portal to another universe...
pic.twitter.com/39iQewjst4

— Tim Sherratt (@wragge) [October 8, 2016](#)

A rethink is required...

Experiment E had two aims. First to highlight the visual character of the redactions themselves – there's a strange sort of beauty in a massed collection of blobs. Secondly, just as with the Real Face of White Australia, I wanted to turn the files inside out. Instead of being dead ends, I wanted the redactions to be discovery points, signposts, ways of exploring the files.

It's online now so play. You can view a [random sample](#) of redactions, or [browse page by page](#) through the entire collection.

Talking about her own ASIO file in the book *Dirty Secrets*, the politician and academic Meredith Burgmann noted that the 'blacking out process seems totally arbitrary and for the reader terribly frustrating, like reading a detective novel with the last page torn out'.⁷ But in hunting for redactions I found they could also bring moments of unexpected joy. It seems that someone got a bit bored and has left us with a [glorious collection](#) of [redaction art](#).

discontents

Working for the commons? Or just overworking as redaction finds its way to its accuracy.
<http://discontents.com.au>

It's interesting, and perhaps ironic, that the removal of information has given me an identifiable data point that I can potentially track against other characteristics of the files. Can I identify patterns by time or topic?

Apparently ASIO assessments have become less conservative over the years - I can test this by looking at changes in redaction rates over time.

I also want to explore the context of redactions. By expanding the window around redactions and OCRing the result, I hope to identify the words that occur most commonly appear near redactions.

Those of you coming to the workshop on Friday will hear more about some of the tools and technologies I've used in these experiments. But I wanted to give a brief overview today because **this is access.**

Digital tools and technologies give us the opportunity to use databases like RecordSearch as archaeological sites to sift through layers of metadata in search of new connections and meanings. **This is access.**

We can turn digitised collections inside out, revealing the people, the processes, the structures, the form. **This is access.**

We can reveal the processes through which records are controlled, concealed, and withheld. **This is access.**

Access is not a deliverable or a product. It's a struggle for understanding and power - not just to see, but to see differently.

This is RecordSearch but not as you know it.

Experiment F is a userscript that [puts the redactions back](#) into RecordSearch. Access is an honest acknowledgement of its own limits, and an invitation to push beyond.

1. Richard Jenkins, 'Identity, surveillance and modernity: Sorting out who's who', in Kristie Bell, Kevin D. Haggerty and David Lyon (eds), *Routledge Companion of Surveillance Studies*, Routledge, 2014, p. 159. [[↵](#)]
2. Eric Ketelaar, 'Access, the democratic imperative', *Archives and Manuscripts*, vol. 34, no. 2, November 2006, p. 73. [[↵](#)]
3. John Gilliom and Torin Monahan, 'Everyday resistance', in Kristie Bell, Kevin D. Haggerty and David Lyon (eds), *Routledge Companion of Surveillance Studies*, Routledge, 2014, p. 407. [[↵](#)]
4. Ann Laura Stoller, *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense*, Princeton University Press, Princeton, 2009, p. 32. [[↵](#)]
5. David Horner, *The Spy Catchers: The Official History of ASIO*, Allen & Unwin, Sydney, 2014, p. 581. [[↵](#)]
6. David McKnight, 'How to read your ASIO file', in Meredith Burgmann (ed.), *Dirty Secrets: Our ASIO Files*, NewSouth, Sydney, 2014, p. 38. [[↵](#)]
7. Meredith Burgmann, 'The secret life of B/77/26 (and friends)', in Meredith Burgmann (ed.), *Dirty Secrets: Our ASIO Files*, NewSouth, Sydney, 2014, p. 454. [[↵](#)]

Share this:

- [Click to email this to a friend \(Opens in new window\)](#)
- [Click to print \(Opens in new window\)](#)
- [Click to share on Twitter \(Opens in new window\)](#)
- [Click to share on Facebook \(Opens in new window\)](#)
- [Click to share on Google+ \(Opens in new window\)](#)
-

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).