

## Digitisation Thread from [archives-and-records-australia](#) List (April – June, 2016)

From: Chris Hurley

Sent: Thursday, 14 April 2016 11:00 AM

Subject: FW: Check out "NSW state library to turn \$3bn collection over to private sector" on Archives Live

Extract from the story in [itnews](#) :

The State Library of NSW has offered the private sector access ... in the hope that digital operators might find a way to turn a buck from the resources online ... the State Library has still only managed to electronically preserve less than 1 percent of its total collection ... It is in a hurry to get the most fragile and vulnerable pieces captured electronically to preserve them ... and hopes there is a business model for making these resources searchable online ... Digital partners will be expected to provide all the staff, equipment and other resources required ... and to prove to the library that they can be trusted to care for highly valuable historical artefacts ... will not be able to claim any new copyright over the electronic versions of the resources, and ... any records created must remain in the public domain in accordance with public library policy.

1. This does not in essence appear to be different from PP arrangements already in place with Ancestry.com, etc. Each contract may vary in detail, however, re copyright, other rights (including access rights), royalties, redaction of access restricted material (where appropriate), duration of the contract and of elements within it, and for succession rights when the private partner fails, liquidates, or is taken over. In the light of recent history, it would be as well to make provision also in case the custodial partner fails, is liquidated, or is merged with someone else. Query: when is the “release” of dusty old resources from custodial dungeons onto the Internet going to start raising questions of copyright, privacy, and cultural sensitivity in records that we would prefer not be asked?
2. Such arrangements appear to be made one-government-institution at a time, with different partners (of each individual institution’s choice), and in accordance with contractual arrangements that are not standardised within a single jurisdiction, let alone across the whole Federal/State/Territory cultural sector.
  - a) Does this not mean an accumulation of arrangements over the documentary cultural heritage with a range of suppliers that are diverse, non-standardised, and potentially erratic?
  - b) It may be commercially advantageous but is this wise policy from a national ( as distinct from a jurisdictional or institutional) point of view?
  - c) Are there even standardised contracts (or statements of minimum requirements) when these deals are entered into, or is each institution sovereign and free to make its own arrangements?
  - d) Who is over-seeing them? I’m not talking about Treasury oversight of the contracts but oversight of the heritage policy and public interest aspects.
  - e) What limits (if any) should be placed on sovereign cultural institutions in regard to the digital exploitation of the resources they are entrusted with?
  - f) Most of them were set up in the pre-digital era? Does their mandate cover this or is it now time to re-examine their mandates? Is this a case of function-shift?
  - g) Should the assumptions behind their establishment as custodial bodies to remain unscrutinised as they move towards digital processes designed to “turn a buck”?
3. The existing PP arrangements are for low-hanging-fruit : viz. low-volume + high-use materials. How likely is it that the entirety (or even very much more than 1%) of our existing paper-based national cultural resources will ever be digitised (especially those at the low-use + high-volume end)? How is the “hope” that fragile valuable artefacts will be saved to be reconciled with “turning a buck”. How to resolve the conflict that culturally valuable artefacts may not be the ones that are most attractive commercially? Should the commercial exploitation of the resources be taken out of the hands of cultural institutions and given to another body within government (unencumbered by custodial responsibilities) to co-ordinate these activities across the sector?
4. What descriptive standards and requirements are being laid down for “making these resources searchable online”? Is a false assumption being made that institutions are responsible only for preservation and not for searchability? What integration will there be with the institutions’ own online portals? What is being done to use the potential of the technology to provide integrated searching beyond the custodial/contractual boundary established by each separate custodial institution and digital supplier? What thought has been

given to integrating access to born-digital resources with “making [digitised paper-based] resources searchable online”?

There are lots of questions around this process at SLNSW and elsewhere. These are just some of them (it seems to me).

All the best

**From: Chris Hurley**

**Sent: Thursday, 14 April 2016 3:57 PM**

**Subject:** digitisation vs description

Further to my post this morning re the [itnews](#) item on SLNSW proposing to contract out its digitisation programme, I have two further questions on a somewhat different tack (viz. the benefits of digitisation programmes that are not contracted out but continue to be paid for by the taxpayer). The story includes the following paragraph :

Despite nearing the end of a [10-year, \\$72 million digitisation journey](#), the State Library has still only managed to electronically preserve less than 1 percent of its total collection, which is made up of 6.3 million items like monographs, sheet music, newspaper collections, microfilm, videos, stamps, photographs and architectural blueprints.

1. Does SLNSW regard digitisation as an access or a preservation measure? Is preservation being mixed into the message to defend the indefensible (viz. a massive taxpayer spend on a boutique resource for the benefit of a favoured few)? The story speaks of “electronically preserv[ing] ... its total collection.” Is electronic preservation of resources that are not in any danger of loss and could be preserved physically for centuries cost-effective and/or justified? How much would it have costed to preserve physically the less than 1% that we have spent \$72m digitising? In any case is digital preservation a saving? It won't be unless SLNSW now discards the resources it has already digitised so the taxpayer does not pay twice – for digital and physical preservation of the same resource. Should government-funded heritage institutions abandon expenditure on digitisation for access, concentrate on digitisation for preservation only, and leave it to society to find a way of funding digital access without it being a burden on the taxpayer? Does spending \$72m to make less than 1% of SLNSW resources available to (probably) less than 1% of the population pass the “pub test”. What would the Sydney Institute say? Should we care what they say? What would be the total taxpayer spend if the approach taken by SLNSW over the last 10 years had been adopted for all our heritage institutions? Would the battlers on struggle-street (as they say) think it a justifiable expense when the money could go to pay for education, health, disability, etc. etc.?
2. After the low-hanging fruit and resources in immediate danger of physical decay, what priority should be given to digitisation over other possible avenues of expenditure of taxpayer money on cultural assets? On the basis of the figures quoted in the story, are we entitled to ask whether the balance between digitisation and description is right? If \$72m buys you digital access to less than 1% of your holdings, would it be better to spend a fraction of that amount to improve online access to better descriptions of un-digitised resources? I would just love it if this question initiated a debate over how a fraction of that amount could be used to improve online access to better descriptions of un-digitised resources.

All the best

PS I trust it is not necessary for me to say that I take the role of Devil's Advocate in some of my postings.

**From: Mark Brogan**

**Sent: Thursday, 14 April 2016 4:19 PM**

**Subject:** [archives-and-records-australia:1755] Re: FW: Check out "NSW state library to turn \$3bn collection over to private sector" on Archives Live

There are two parameters of SLNSW's proposal to provide private access that are likely to be a fly in the ointment as far as private sector collaboration is concerned. Firstly, if a profit making firm were to embark on digitisation as part of a partnership, apparently it will have no copyright over the digital source created and, secondly, it will not have the right to exclude others from the resource it creates (if this is the meaning of "any records created must remain in the public domain in accordance with public library policy"). The positions taken are familiar from like agreements elsewhere and basically protect custodial institutions from back door privatisation of holdings.

Perhaps as with Ancestry, the secret to making this fly from a commercial standpoint is aggregation and value added services behind a portal gateway. Ancestry has been very clever with this and quite entrepreneurial. But it will be a hard sell for SLNSW and as you have suggested, the value proposition of some digitisation targets will be greater than others, leading to the best assets being prioritised and most ignored.

**From: Andrew Waugh**

**Sent: Thursday, 14 April 2016 11:14 PM**

**Subject:** Re: [archives-and-records-australia:1754] digitisation vs description

Hi Chris,

Without going into whether \$72 million is good value for digitising less than 630,000 items, I'd make a couple of points.

The primary purpose of most digitisation is for access. But most GLAM organisations would digitise at 'preservation' quality. This is because most of the cost of digitisation is in the handling of the material (getting it out of storage, preparing it, putting it away), generation of metadata, and quality assurance. The marginal additional cost of digitising at 'preservation' quality over 'access' quality is relatively small, and has a number of advantages. These include additional uses of the digitised material (e.g. publication), future proofing access (i.e. higher bandwidth driving demand for higher resolution in access copies), and insurance against loss of the physical original.

Don't underestimate the value of the increased access provided by digitisation. I don't believe that it is a 'boutique resource for the benefit of a favoured few.' Digitisation means that people all around Australia can access material that previously could only be accessed by the favoured few who were relatively close and had the time to physically visit the GLAM. And Trove shows that they, do, indeed, access the material. The use of OCR of printed material has changed the game in historical research - One of the side effects of Tim Sherratt's recent '#fundTrove' twitter campaign was a view on the amazing ways Trove was being mined by people. It really was an eye-opener.

And Trove is the best answer to your question as to the relative value of digitisation/OCR/indexing versus description. The Trove approach is a far more cost effective and powerful method of providing access to the information in newspapers than the 100 years of desultory manual indexing that preceded it. For an archive, the question is more balanced. So many of the records are handwritten, which can be digitised but not OCRed or indexed. These digitised records need to be described to be accessible - expensive and not very effective unless the description matches the researcher's questions. But perhaps we need to spend more time (and money) digitising 20th century records (more likely to be typewritten) than 19th century records.

Finally, digitisation is a very effective preservation mechanism for the physical originals. To reverse Rothenberg: "Paper lasts for 400 years, or until someone unfolds it to read". The great contradiction of physical collections is that they survive best if no one, or at least very very few, actually use them. As soon as they become popular they start to fall to pieces. I can remember the condition of the books in the State Library, and I frequently see the circle of loose paper around researchers reading fragile records.

Yours, also, playing the devil's advocate :-)

Andrew

**From: Chris Hurley**

**Sent: Saturday, 16 April 2016 10:48 AM**

**Subject:** Re: [archives-and-records-australia:1756] digitisation vs description

**Does digitisation for access serve a marginal "boutique" market?**

Andrew says -

<<...I don't believe that it is a 'boutique resource for the benefit of a favoured few.'... Trove is the best answer to your question as to the relative value of digitisation/OCR/indexing versus description. The Trove approach is a far more cost effective and powerful method of providing access to the information in newspapers than the 100 years of desultory manual indexing that preceded it.>>

No one can doubt the popularity of digitising heritage resources. Football is also popular and, although we have yet to see the rise of heritage hooligans, it is legitimate to ask whether taxpayer money should be

applied to supporting either \$multi-million mass entertainment or \$multi-million mass heritage. But popularity and mass usage are not really the point. What challenges your value proposition and makes it boutique is the lack of penetration. If all our cultural resources were being digitised, your value proposition might hold but on the figures provided in this story it has cost \$72million to digitise “less than” 1% of SLNSW’s holdings. I have no way of knowing if these figures are correct and whether the holdings in question are the entirety of what is held at SLNSW or only a heritage portion of it, but either way we are dealing with big numbers. Assuming it is less than 1% of what (in someone’s estimation) should be digitised, it would therefore cost \$7.2billion to do the lot. Assuming SLNSW itself holds (let’s be generous) 2.5% of the nation’s documentary heritage that should also be digitised, then the bill for digitising everything would be \$288billion. I could build a Very Fast Train system for less.

It is a boutique resource not because few people use it but because, at this rate and at this cost, so little of it will ever be available digitally. Preferring digital access to less than 1% over access to descriptions of more than 99% is, I say, a boutique solution for a favoured few. Hence my question: could we get bigger bang for our buck with better descriptions? If you concede that question, you can’t avoid the follow-on: could we do online description better with the application of better brains and more resources? No prizes for guessing my answer.

Compared to traditional search room statistics, the numbers using digital resources are mind-blowing. They are a favoured few in relative terms, however, because (regardless of their absolute numbers) they are a minority group of hobbyists and researchers within the community whose claims on taxpayer funding have to be measured against competing needs and priorities in the total population. Some competing expenditure – on disability services, for example – also involves a minority of beneficiaries but the social imperative there is high. Education and health services affect almost all of us. What makes this comparison a fair one is twofold :

1. Heritage spending is usually so relatively insignificant that there is no point of comparison with big-ticket items but that is not so when \$288 billion is at stake.
2. Digital access to heritage resources is lower on my significance index (social significance, that is, not “cultural significance”) because access by alternative means, though clunky and less convenient, is still possible.

### **Is it possible to have two motives for digitising resources?**

Andrew says

<<...The primary purpose of most digitisation is for access. But most GLAM organisations would digitise at 'preservation' quality...digitisation is a very effective preservation mechanism for the physical originals...physical collections...survive best if no one, or at least very very few, actually use them...>> >>.

Yes, librarians and archivists can whittle and chew gum at the same time but they can’t have their cake and eat it too.

If \$288billion is too much to pay for digitising the nation’s heritage, choices will have to be made. The smaller the pot, the more likely it is that access and preservation will be in conflict: the low-hanging fruit that is digitised first won’t necessarily be the stuff most urgently in need of preservation. Since this is also the stuff least likely to be intensively used, and therefore less likely to fall to pieces in the researchers’ hands, the conflict of purpose may go some way to resolving itself. Some preservation advantage will undoubtedly come from access-based digitisation and this will relieve pressure on conservation programmes, but there remain two issues you will need to deal with if your argument holds :

1. Do we stop preserving resources that have been digitised and move conservation spending onto those which haven’t?
2. Do we stop spending money on physical preservation altogether, close down the conservation facilities, sack the conservators, and move all our spending into digital preservation?

All the best

**Subject:** [archives-and-records-australia:1758] Re: digitisation vs description

Chris asks

1. Do we stop preserving resources that have been digitised and move conservation spending onto those which haven't?
2. Do we stop spending money on physical preservation altogether, close down the conservation facilities, sack the conservators, and move all our spending into digital preservation?

These questions are phrased in a way that suggests that good digitisation does not require conservation of the physical resources involved, when the opposite is true. For good quality images to be made, the originals need to be in good condition. This was true for best practice microfilming and photography and remains true with digitisation. The cameras have changed, but not the process. Materials need to be cleaned, flattened and repaired to ensure the images can stand in for the original for most purposes. As part of an imaging or reformatting process, the originals can be rehoused effectively and efficiently, and stored away in case they are required or a new imaging project is required.

According to the SLNSW annual report 2014 -2015, they "digitised 53,000 pages of First World War soldiers' diaries, 5464 hours of oral history and 1.5 million pages of NSW newspapers, bringing the total number of 'turned digital objects' to 6.7 million." It's a little trickier working out what that actually means, because the same report says that they only created 160,000 digital images inhouse. Finding \$72million is even harder.

The same report also says

The Electronic Records Project (creation of an electronic collection catalogue) was first capitalised in 2013/2014 as an intangible asset.

The E-records project was completed in April 2014, of which \$21.5 million were capitalised as intangible assets. The intangible asset is the outcome of a project to create an online collection catalogue available to Library clients and improvement to collection retrieval and management processes.

The catalogue, built on knowledge and skills, is regularly updated for acquisitions and disposals. It therefore has an ongoing useful life. Hardware and software platforms may be upgraded due to technological obsolescence; however, the information created by the project has an indefinite useful life and is therefore not amortised.

The Digitization of Collection Project commenced in 2012/13. It will result in digital images of all collection items being created. The availability of these images will allow research to be carried out externally and increase access to information. As they become available to the public these digital images are capitalised. As at 30th June 2015 \$14.2 million has been capitalised as intangible assets. These digital assets have an existence and utility separate from the actual physical collection assets.

([http://www.sl.nsw.gov.au/sites/default/files/slsw\\_annual\\_report\\_2014-15.pdf](http://www.sl.nsw.gov.au/sites/default/files/slsw_annual_report_2014-15.pdf))

I'm guessing 'all collection items' is a bit of hyperbole, and actually means those items in the State Reference Library, Mitchell and Dixson collections for which copyright and other restrictions do not apply, and which are in a good condition or suitable format for digitisation.

At the moment, archives and libraries are digitising microfilm because the film is easy to copy and we have users who will accept black and white copies. This is changing rapidly, and I am aware of a project where original materials, that have been microfilmed and that microfilm digitised, are now being digitised in colour to meet new expectations. I suspect that, as the tools and techniques for viewing and working with images evolves, we will be seeing another push for digitisation at higher resolutions than we currently accept, and these materials will be dragged out again 5, 10, 20 years from now.

I am concerned that digitisation projects are being driven by 'niche' interests, and that the pool of potential digital content is shrinking rather than expanding. I agree that the way to make these resources more generally available is to provide good quality descriptions in the first instance, and then work towards making digital copies of those resources available. Given the costs of describing, conserving, copying, saving, migrating, redacting, republishing and preserving the digital copies, we know that full and complete digitisation is not possible. We need to think strategically about what is digitised and about how we get the

message out that if it's not on the internet, it's part of the iceberg of archives and other resources available physically.

Lise

From: **Chris Hurley**

Sent: **Sunday, 17 April 2016 11:42 AM**

**Subject:** Re: [archives-and-records-australia:1758] Re: digitisation vs description

Lise says

<These questions are phrased in a way that suggests that good digitisation does not require conservation of the physical resources involved, when the opposite is true. For good quality images to be made, the originals need to be in good condition.>>

The phrasing may bear that interpretation but that was not the intention. I was trying to probe the strategic thinking behind mass-digitisation for preservation by taking up Andrew's argument that digitisation may be a preferable way (economically and/or technically) to achieve preservation outcomes as distinct from access outcomes. As we know, cost is a significant component in digitisation for either purpose. Accepting that conservation is needed before digitisation occurs, the question remains: why should the taxpayer be asked to pay twice for the same result after it has occurred? Are they going to be persuaded to pay to preserve the originals in case they are needed for a new imaging project?

1. Pre-digitisation conservation is one thing; on-going preservation of the digitised "originals" is another. If digitisation is the preferred preservation strategy, should conservation work be limited to preparing material most desperately in need of digitisation? Should other preservation work be abandoned in favour of facilitating the preferred strategy of first digitising materials most at risk?
2. After digitisation (once pre-digitisation conservation has occurred) what do we do with the digitised "originals"? Why spend any more on their preservation? Why not discard them? A bonfire in Macquarie St, perhaps? Or, would that simply compound the problem of global warming?
3. Is a digital rendition a version or a transformation? Does the "original" have intrinsic value that must be preserved? What is it and why? The Mona Lisa, maybe. A record that may be needed for forensic examination, OK. But run-of-the-mill heritage materials? Doesn't preserving an "original" for its intrinsic value and digitising to preserve the content simply double the preservation costs? How do we sell that to the taxpayer?

At the micro-level, these questions are not new. As Lise suggests, they arose already with microfilming. Digitisation requires a re-think because the order of cost if anything other than a niche result is proposed means looking at offsets. \$72million is enough to raise these questions; more so when it gets into the \$billions. When did we last have a \$72million microfilm project on our hands? This thread is about the preservation justification for digitisation (as distinct from the other one I tried to start about digitisation for access, which hasn't taken off apart from Mark Brogan's sage remarks).

Nothing so far has persuaded me that preservation can be argued as a justification for mass-digitisation until these logistical questions are answered. You couldn't sign off on a \$multi-million (?billion) business case until such obvious cost offsets had been dealt with. Preservation as an incidental by-product of a digitisation programme undertaken for access purposes, sure. But you don't sell a \$multi-million (multi-billion?) strategy on incidental by-products. I accept that, politically, destruction of digitised originals would raise the fury of the heritage-mafia and may drown out logical analysis; but that is a different matter.

All the best

From: **Andrew Waugh**

Sent: **Thursday, 21 April 2016 12:02 PM**

**Subject:** Re: [archives-and-records-australia:1761] digitisation vs description

I started writing an answer, but it got too long and never got anywhere, so I'm trying again.

Chris asked... "Hence my question: could we get bigger bang for our buck with better descriptions [than digitisation]?"

No.

What Trove has shown us is that digitisation of a large corpus (even at low quality) coupled with OCR/searching completely change the scope of what is possible. It dramatically improves the ability to find resources. It dramatically reduces access costs for researchers. And researchers respond to this.

The question is more tricky for archives. So much of our collections are handwritten. This cannot (yet) be OCR'd, and so some description is necessary for to allow researchers to find things. So digitisation by itself not useful for an archive (but see below). This is why archival digitisation projects always include work to improve the description of the objects. This, of course, increases the cost of digitisation projects - what is portrayed as a 'digitisation' project is usually as 'digitisation and description' project. (And is often a 'conservation, digitisation, description, and rehousing' project at even greater cost. But I digress.)

But description by itself is almost as useless. Yes, it improves the ability to identify resources. But it does nothing to improve access. That's really a boutique resource - the benefit of the expenditure can only be realised by the very small number of researchers that can actually attend the archive concerned. This essentially means (some) locals, very well heeled researchers, and academics with a travel budget.

Actually, I think it's even worse than this as I don't think archives have fully processed the lessons of Trove.

I've said that so much of our collections are handwritten. So they are, but the proportion of handwritten vs typed records changes dramatically over time. In my experience (with a very limited range of materials), nineteenth century records are handwritten. By WWI, however, a substantial percentage of records are typed. Some series are almost completely typed, while others are a mixture of handwritten and typewritten material. Yet others remain completely handwritten.

I think we are missing a bet in our focus on digitising 19th century material. It would be interesting to experiment with digitising early 20th century material with a fair proportion of typewritten material. The typewritten material can then be OCR'd and used as a search access point into the files. You don't need to find every document in a search, what you need to find is the relevant files. And for that, you may only need some of the documents in the file to be OCRable.

andrew

From: **Chris Hurley**

Sent: **Friday, 22 April 2016 12:59 PM**

**Subject:** RE: [archives-and-records-australia:1763] digitisation vs description

Andrew's reply rather misses my point, demonstrated by his use of square brackets. My question was not whether description delivers a better access result than digitisation. If my question is going to be re-phrased, let me be the one to do it : could we get better bang for our buck with better descriptions [of the vast quantity of records that will never be digitised]?

I can argue that the comparison between digitisation and description is not as one-sided as Andrew suggests but that is nothing to the point and can be reserved for another day. If you read my posts you will find nothing that suggests that description delivers a better access result than OCR/search. My question was not about the technological advantages of digitisation but its economic feasibility. Assuming no one is ever going to hand over or find a commercially viable way of supplying the money needed to digitise the lot, much of the nation's non-digital documentary heritage is NEVER going to be digitised – regardless of whatever benefits it would confer if it could be paid for. It's not enough for Andrew to argue that OCR/searching produces a better access result than description; he has to explain how to pay for it.

I repeat, therefore, do we get better bang-for-buck by diverting some of the resources used to digitise a tiny fraction of the nation's documentary heritage into better descriptions of the larger proportion that never will be. You have to have descriptions in order to access the un-digitised proportion because that's all you have with which to find them. The relative merits of OCR and description are simply irrelevant to that question. The question(s) I want answered, therefore, are these. In light of the numbers cited in the article that kicked all this off -

1. How realistic is it to suppose that anything more than a tiny fraction of the nation's documentary heritage is EVER going to be digitised?
2. Should all our resources be directed towards mass digitisation projects that will never capture more than a tiny fraction?

- Should some of that money be spent instead on better descriptions of the much larger volume that will never be accessible any other way?

There may be answers. Maybe mass digitisation is a realistic possibility and I'm incapable of understanding how. If so, educate me.

PS. I'm waiting for someone (more daring than I) to suggest that what isn't worth digitising isn't worth keeping or a line of argument that leads to that conclusion.

All the best

From: **Chris Hurley**

Sent: **Friday, 6 May 2016 2:20 PM**

Subject: [archives-and-records-australia:1769] Data on digitisation

A few weeks ago I initiated a discussion on digitisation. The postings ended up pursuing a number of threads including -

- The rationale (access; preservation; both; other?)
- The pros/cons (better access; OCR vs description; toxic effect on the latter; de-contextualisation?)
- The strategic implications (proportions digitised; feasibility; boutique audience; priorities; resource allocation?)

I was especially interested in (3). In that respect, I used reported figures about the state of digitisation at SLNSW and extrapolated from there. Does better information exist about the kind of sector-wide quantities involved? Most online material is about pros & cons at the micro level. But question (3) needs to be explored (if at all) by reference to data relating to the sector overall about what proportion of analogue material has been digitised and what proportion still needs doing. I can't believe this data hasn't been collected and analysed by someone.

The article on which I based my initial posting spoke of <1% of SLNSW holdings having already been digitised. I have since come across a [European report](#) that gives more optimistic figures (see table 4.17 below), suggesting that European archives & libraries have already digitised 11-12% as against a perceived requirement of 47-48% yet to be done. This table is based on a 2014 survey. I wouldn't take these results at face value without looking into them more closely than I have been able to so far. I would want to know, for example, how comprehensive the coverage of the survey is and whether truly large holdings were surveyed at all. I am suspicious of the similarity between the figures for libraries and archives which is unexpected if truly large records repositories were surveyed. Is the quantum of "analogue...collections" finite or is it still rising? Adding together percentages from such diverse areas of activity to produce a "Total" without reference to the absolute quantities involved looks a bit dodgy. Size is measured by budget and staff numbers but a question on actual volume is "optional"! Another table from the Survey not cited in the Report (Figure 4.20) gives less optimistic figures of 6-8% for digitisation progress in archives. And I haven't figured out the implications of the remaining 40-42% that they say don't need digitising, but I see that figure for libraries & archives is more than double that for museums and others (which is curious). All this may be dealt with in the [Survey](#) when I get around to reading it closely.

In the meantime, it's really good to have any figures at all to work on. Is anyone aware of other sources for such data (especially for Australia and NZ)?

All the best

Figure 4.17: Estimated percentage of your analogue heritage collections that has already been digitally reproduced and still needs to be reproduced (n=1.179)

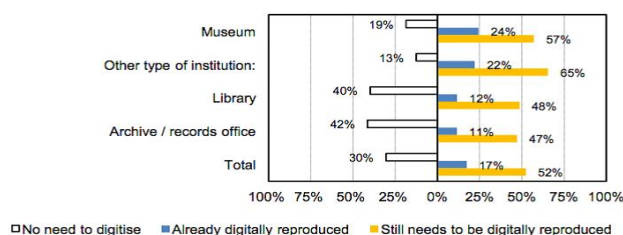
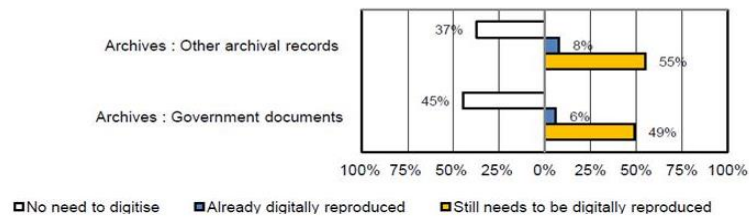




Figure 4.20: Archival resources – digitisation progress (n=64)



From: Chris Hurley

Sent: Tuesday, 17 May 2016 6:02 PM

Subject: Re: [archives-and-records-australia:1777] Digitisation and researchers

So far from being tangential, I think the issues raised in the article that Andrew cites are very relevant to a parallel set of issues for archivists, as well as historians. If anything, my original questions were tangential to them.

The author says, inter alia <<...we can now find information without knowing where to look...Web-based full-text search decouples data from place. In doing so, it dissolves the structural constraints that kept history bound to political-territorial units long after the intellectual liabilities of that bond were well known... Digital search offers disintermediated discovery...>>

My original point was that, so long as digitised resources amount to no more than a fraction of what is available online, we must look to improve the way undigitised resources are described online because descriptions are the only way they can be accessed. I am opposed to the idea that text searching replaces descriptive effort for this and for other reasons. Content searching alone misses the contextual knowledge unless that knowledge is captured in the metadata and lazy digitisation won't do that. In the case of undigitised records, searching can only be based on description, not content. It follows (in my view) for that reason alone that we must pay more attention to better description so that online searching for undigitised materials (for now, the majority of the resources we manage) is as good as it can be. Tim Sherratt repeatedly shows us how to mine and access descriptions of undigitised materials in ways not thought of by those who prepared the descriptions but those descriptions are only to be found on the web sites of custodians who have done little more (in my estimation) than place descriptions online based on practices meant for display in an undigitised world. I say we need to do more and reimagine how we describe undigitised resources for online access.

Pre-digital descriptions follow what the ICA standards call the "multi-level rule" which requires that descriptive data is pushed as far up as possible in an (assumed) hierarchy of descriptive levels. In practical terms, descriptive data that can be recorded at series level (for example) should not be repeated at item level. When such item descriptions are "decoupled" (as explained by the author of the article) the assumptions that underlay the description are fundamentally eroded. We have to deal with that and it is entertaining to watch how different custodians handle it on their web sites. But the consequences are more profound still. I have written elsewhere that our real treasure is not the materials we hold but the knowledge we possess about structure and context - knowledge about the resource, in other words, rather than knowledge contained in the resource itself. This is descriptive data that is most remote from item level descriptions and the most vulnerable to decoupling. The challenge is to improve descriptive practice so that our real treasure is displayed and available in online searching for all our materials (digital, digitised, and undigitised) alongside the tools that facilitate content searching.

The ultimate challenge is to ensure that online access is not only "across borders" but also across delivery channels, so that the same search hits all three. This relates to my original concern : that relying solely on the benefits of content searching of digitised materials is as remiss as simply uploading descriptions into an online home for which they were not designed.

All the best

On Tue, May 17, 2016 at 8:25 AM, Andrew Waugh <[andrew.waugh@gmail.com](mailto:andrew.waugh@gmail.com)> wrote:

Slightly tangential to the questions that Chris has raised, this is an interesting take on the impact of digitisation on professional historians, one of our user groups.

<https://ahr.oxfordjournals.org/content/121/2/377.full>

The author's argument is that digitisation and OCR are a sea change in historical researcher (for both well and woe), the more remarkable precisely because the historians have not marked it. The focus has been on the impact of 'big data' on history.

Andrew

From: **Chris Hurley**

Sent: **Tuesday, 24 May 2016 9:34 AM**

**Subject:** [archives-and-records-australia:1780] Digitisation: more on rationale and metrics

The article (below) offers a further insight into the rationale and the metrics of digitisation. It contains some very naïve misconceptions (confusing format and medium, for example, and supposing that the sum of a record is merely its information content) and it seems unlikely that Finland, Hungary, and Sweden are really digitising their entire non-digital archives and throwing away the hard copy, but, even if that were true, the numbers are startling. According to the article, NAFin has over 100Km of shelves now and expects another 130Km of hard copy records over the next 35 years. According to the article, the “prime motivation” for digitisation is cost-efficiency with improved access as a beneficial side effect. The policy questions I still don't seem able to find answers for include these :

- What is the total quantum of national heritage material that could or should be digitised (here or anywhere else)? For NAFin the answer seems to be 230Km by 2050.
- What proportion of that has so far been digitised, over what time, at what rate, and at what cost?
- At what rate is it currently being digitised? At what cost?
- What is the cost/benefit calculation for digitisation & migration vs hard copy storage & preservation over time?
- Does digitisation really eliminate the cost of hard copy records or do custodians wimp out and go on keeping hard copy after digitisation?
- What is the rate of un-digitised intake and does that growth rate fall behind, match or exceed the digitisation rate?
- What is the projected rate of un-digitised intake and for how much longer will it go on?
- When is it expected that the intake of born-digital material will bring the non-digitised growth rate into balance (at least) with the digitisation rate?
- What kind of integration is there between metadata schemes for digital/digitised resources and online descriptions of non-digitised materials?

[Why archives want to destroy their reams of papers and risk a digital dark age](#)

**The costs of storing tons of documents on miles of shelves means they're being banished from state archives. But is it too risky Tom Jeffreys investigates**

... “The National Archives are not about paper, but the data that the papers are covered with,” says Deputy Director General Markku Nenonen. While the National Archives do contain a number of rare and beautiful objects, its bread and butter is information ... The National Archives of Finland currently has documents covering over 100km of shelf space. Documents requiring a further 130km are expected to come in over the next 35 years ... The prime motivation behind digitisation is therefore cost-efficiency, especially with Finland is in its fifth year of recession ... However, digitisation is not just about efficiency. “Ultimately, this will be beneficial for citizens,” says Juha Haataja, Counsellor of Education at the Ministry of Education and Culture. “They will be able to access information much more easily.” For researchers, digitisation opens up new methods of searching, collating, and exploring information ...

*This article originally appeared in [The Long + Short](#), the free online magazine of ideas published by Nesta, the UK's innovation foundation*

From: **Chris Hurley**

Sent: **Tuesday, 24 May 2016 12:43 PM**

**Subject:** [archives-and-records-australia:1781] RE: Digitisation: more on rationale and metrics

This just in. There is an excellent article in the latest [Archivaria](#) (Spring 2016, No.81) entitled “Digitizing Archival Records: Benefits and Challenges for a Large Professional Accounting Association”. It doesn't actually bear on the metrics issue I have been raising but it does have some sensible things to say about the rationale and other follow-on matters, pointing out, inter alia, that the necessarily selective nature of

digitisation means that it complicates (enhances if you like) but does not replace existing archival management regimes. This is my point exactly – viz. that whatever we do with digitisation we'll still have a lot of stuff that we have to manage and make accessible and, further, that there can't be two access regimes. The authors recognise that digitised resources must be integrated with the non-digital to enhance accessibility overall and be linked with other resources across custodial boundaries. That last statement must also be true, surely, for online descriptions of non-digital resources. While acknowledging the benefits of text searching as an enhancement to accessibility, the authors do not fall for the hype that it replaces structure and context.

... Digitization adds another layer to the already inherent complexity associated with managing a physical archive ... Just as it is rarely feasible to retain every physical record generated by an organization ... it will be equally rare to digitize every record in the collection ... and the selection of items for digitization becomes a key variable in shaping the accessibility of the existing archival collection ... Enhanced access to archival records via digitization changes both the temporal and spatial relationships between researchers and records ... A reliance on a traditional records management (custodial) approach would have missed the opportunity to identify contextual and functional aspects of the resource and its place within the broader network ... Digitization improves access, but that access is only as good as the finding aids developed to support the process ... A further implication of the spread of digital archives is the scope for cross-repository linkages ...

From: **Michael Piggott**

Sent: **Monday, 23 May 2016 10:19 PM**

**Subject:** [archives-and-records-australia:1782] "The National Archives are not about paper, but the data that the papers are covered with"

Colleagues

Further to Chris' and others' recent contributions on digitization, see below. I seem to recall a version of this ruthless approach was practised in Singapore when copy+shred was via microfilm. As for the quote above, whether it is debind+digitize+repackage/store flat, or debind+digitize+discard, am I the only one who believes there is also "data" embodied in the material culture of the physical record worth preserving if the physical record itself is judged worth preserving?

Michael Piggott

Canberra

From: **Chris Hurley**

Sent: **Wednesday, 25 May 2016 3:35 PM**

**Subject:** Re: [archives-and-records-australia:1784] "The National Archives are not about paper, but the data that the papers are covered with"

One must assume they mean what they say. The Finns are quoted as arguing that the chief reason for digitising 230Km of records is to save costs on storage, with access and preservation as beneficial side effects. They may, of course, be misquoted. If saving on storage costs is the rationale it follows that they must discard the hard copies, otherwise it's crazy. You would go on paying to store and preserve the hard copy (the avoidance of which costs is your chief reason for doing it) while incurring new costs to digitise and migrate the copies. Respect the Finns for saying that (however much you may disagree with them, Michael) instead of blurring their rationale by giving all three reasons equal weight - as one sometimes sees in statements emanating from the woolly headed who don't confront the issue.

If you don't adopt the Finnish rationale, i.e. if you keep the hard copy, a three-pronged argument doesn't work anyway. If you keep the hard copy, there is no saving on storage so that leg of the rationale is voided. The second leg (saving preservation costs on the hard copy) is similarly voided since it makes no sense to keep the hard copy if you don't then preserve it. The reductio, therefore, is that (absent destruction of the hard copy) the only possible argument in favour of digitisation is better access (plus, in some cases at least, better preservation on top of what you have go on paying for preserving the duplicated hard copy). I'm not for a moment saying that better access through digitisation isn't a perfectly legitimate argument to make. But those promoting digitisation are seldom (so far as I can see) as clear-headed about their rationale as the Finns.

What I'm trying to get at is a coherent rationale for mass-digitisation. That must be the starting point for any sensible discussion of the implications. If you're talking about digitisation as a boutique added extra, costing relatively little, a coherent rationale may not matter. But when you're talking about 230K or the astronomical costs for SLNSW that started this thread off, I don't see how it can be avoided. And, the more you argue the benefits of the digitisation component of a heritage programme (however small it is, if not in cost, then as a proportion of your stuff) the more urgent it is to say how your digitised/digital content is being integrated with online descriptions of your non-digitised stuff.

All the best

From: **Tim Robinson**

Sent: **Wednesday, 25 May 2016 4:03 PM**

**Subject:** RE: [archives-and-records-australia:1785] "The National Archives are not about paper, but the data that the papers are covered with"

“If saving on storage costs is the rationale it follows that they must discard the hard copies, otherwise it's crazy.”

True. I have a memory of David Bearman saying (and my apologies if I am misrepresenting David) that if that if one goal is to preserve the information in records then scanning them and selling the originals is an option. The objects have a value, some quite high, that could be realised to pay for the preservation of the information.

Regards,

From: **Chris Hurley**

Sent: **Thursday, 26 May 2016 10:35 AM**

**Subject:** [archives-and-records-australia:1786] Targeted Digitisation

I've been going on a bit lately about mass digitisation. Just to remind ourselves of the benefits of targeted digitisation is this (very uplifting) article that's come up on Archives Professionals :

[From invisible to digital: digitising endangered historical documents in Brazil](#) by Courtenay J Campbell

All the best

From: **Andrew Waugh**

Sent: **Thursday, 26 May 2016 10:41 AM**

**Subject:** Re: [archives-and-records-australia:1787] "The National Archives are not about paper, but the data that the papers are covered with"

On 25 May 2016 at 15:35, Chris Hurley <[descriptionguy@gmail.com](mailto:descriptionguy@gmail.com)> wrote:

If you don't adopt the Finnish rationale, i.e. if you keep the hard copy, a three-pronged argument doesn't work anyway. If you keep the hard copy, there is no saving on storage so that leg of the rationale is voided. The second leg (saving preservation costs on the hard copy) is similarly voided since it makes no sense to keep the hard copy if you don't then preserve it. The reductio, therefore, is that (absent destruction of the hard copy) the only possible argument in favour of digitisation is better access (plus, in some cases at least, better preservation on top of what you have go on paying for preserving the duplicated hard copy).

This is not completely true, although digitisation/destruction would yield the greatest benefits of digitisation.

It's not true because physical storage does not have a uniform cost. Physical storage costs depend on the capital and running costs of the storage building. These, in turn, depend on where the building is located and how densely you can pack the stored objects. Both of these depend heavily on how often you expect to access the stored objects. Libraries have been thinking about this for about 15 or 20 years, and it's common to have off-site, high density, storage for items with little access requirements.

At one extreme is the typical archival building. It's located in a major city on land that is worth quite a bit. The storage density is reasonably low because staff have to be able to easily access any record. It's necessary to provide a reading room - with the consequent need for service facilities (toilets, entry areas, staff areas for the reading room staff...) Often the building itself is of quite high quality.

At the other extreme is a storage archive where access is expected to be rare. It's located in regional areas where land is cheap because the records are rarely accessed. Density is high because access is very rare. No

service facilities are provided beyond staff facilities - and these need only be those required to secure and manage the facility. A small digitisation facility would be useful (as re-digitisation would be the main access requirement), and somewhere where access to the Internet at a reasonable bandwidth was available. While building quality should be high (no leaking containers), it doesn't need the architectural quality commonly found in main archival access points

Incidentally, I think you are over estimating the 'preservation' costs. One of the (correctly) lauded positive characteristics of paper records is that paper \*generally\* deteriorates slowly. The major preservation challenge for most (but not all) paper records is the damage caused by handling during access. (I've said before, to deliberately misquote Rothenberg, "Paper lasts for 400 years, or until it's handled.") So the only preservation action often required for paper is simply to stop handling it.

I've occasionally thought about where I'd locate such a facility in Victoria. I worked out my main criteria would be risk of natural disasters. This would rule out anywhere with forest (even thin, poor forest) - too high a risk of bushfire. I'd avoid anywhere near a river or the sea (storms and flooding). The grain belts north of the Great Divide would be ruled out due to the risk of mouse plague. I reckon an area on the plains south/south west of Ballarat might be a goer, or in the rain shadow plains between Melbourne and Geelong west of the Princes Hwy.

andrew

From: **Chris Hurley**

Sent: **Thursday, 26 May 2016 12:13 PM**

**Subject:** RE: [archives-and-records-australia:1790] "The National Archives are not about paper, but the data that the papers are covered with"

Archives NZ (Wellington) stores records below sea level, on land resumed from the harbour, in a multi-level building that is on the fault line running down the country from north to south (along the Ring of Fire). The building violates about 90% of Ted Ling's criteria. When I was there I used to say to them that the only safe place for NZ's archival heritage was in Australia. That didn't go down at all well. But I suppose the Christchurch 'quake demonstrated that their records can survive even in the most hostile environment.

Some form of dormant storage could (as you say) reduce the cost of keeping discarded hard copy. You could, as Michael Piggott says, flat-pack the residue of the digitisation process and stack in it a virtually inaccessible place (the Finns are looking for storage 150Km or something like that from Helsinki). But why would you do that? Except for selected items with iconic, totemic, or intrinsic value or merit, why keep them at all? However remote and dormant, the site costs are only part of the story and here, I think, you are underestimating the cost of keeping physical records alive (if only on life support). If the records are to be kept at all usable, you have to pay for retrieval and delivery systems as well as site costs, wherever you put them, to say nothing of the cost of duplicating reference and access systems for both your digital and physical holdings. I repeat, why would you do that?

If, however, you strip the digitised physical records of access and retrieval, description, and availability you might as well bury them and be done with it. You may disarm criticisms such as that sounded by Michael when trying to justify mass digitisation but by only reducing (rather than eliminating) costs of physical storage then for every \$dollar spent on deep storage you reduce the cost-offsets of mass digitisation. You would have to be able to explain to the funding authorities why (on earth) you want to keep the physical stuff after it has been digitised. I am accustomed to likening archival repositories to morgues (places where you send dead records). Deep storage, where you house the unused and unusable detritus of a digitisation process in the hope that one day improved technology or funding will enable you to bring them to life again, would give me a new metaphor – repositories as cryogenic chambers.

I love the idea that archives could sell off the physical detritus of a digitisation process to help pay for the cost. Thank you Tim Robinson and David Bearman. After the sell-off, the dismembered "collections" would then reform elsewhere as re-incarnated Frankenstein monsters composed of fragmented bits from many repositories. In due course these new "collections" might themselves be digitised on the same principle and the whole process begin again. Delicious.

But seriously, folks, what I am obviously trying to do in this thread is frame the question. I think, if I keep probing, it will get to the point that we can agree (am I dreaming?) that digitisation is going to have a big impact on what I have called the low-hanging fruit (viz. at the low volume & high usage end of the

spectrum) and that vast quantities of the national heritage (towards the high volume and low use end) will remain undigitised – at least for a very long time. Thus framed, I would like the discussion to address what I regard as a central issue (an issue I tried to begin grappling with in my *Modest Proposal* by starting to set out the functional requirements for online global access). This involves stepping back (intellectually) from the technological wizardry available when dealing with accessing digitised/digital resources and integrating (in our thinking) access to both digital and non-digital resources and across custodial boundaries. I don't regard it as a satisfactory response to say that we -

- will go on providing online access to the non-digital as we have always done using descriptive practices developed for physical records and simply layer access to digital materials on top of that;
- will rely on content-searching at the expense of context and structure;
- have successfully integrated privacy, access, and redaction management regimes with global online searching.

While we're at it, we can also look for sensible ways of reducing the costs of physical storage (as we have always done). When we were bedazzled by electronic recordkeeping we thought of it as a special thing (we even had a SIG devoted to it). We have now sensibly refocussed and we regard electronic records as simply one more part of the suite of issues we have to deal with. That is how we should now be regarding digitisation.

All the best

**From: Cassie Findlay**

**Sent: Friday, 27 May 2016 11:58 AM**

**Subject:** Re: [archives-and-records-australia:1792] "The National Archives are not about paper, but the data that the papers are covered with"

Really interesting thread, thank you.

Of course there is a great deal of digitisation going on in govt agencies, both with destruction of originals and without, including for records identified as archives. There are (admittedly glacial) moves to do more proactive records release as well as FOI stuff being published online (and not just by the responsible agency, also by 3rd parties) and open data policies, along with (tentative) moves to harmonise archives/records laws & practices with these frameworks. So I'm definitely keen to make sure our discussion on maximising the opportunities for access & rich context for digitised and born digital records online is inclusive of these needs - across custody and control boundaries - as Chris suggests, if I understand correctly.

Yay for an extension to Modest Proposal!

Best

**From: Andrew Waugh**

**Sent: Thursday, 2 June 2016 11:31 PM**

**Subject:** [archives-and-records-australia:1801] The value of digitisation

An interesting view of the value of digitisation.

<http://historyonics.blogspot.com.au/2016/06/privatising-digital-past.html>

Essentially the author's argument is that valuable as digitisation is to research, it comes at a cost. This cost is the promotion of those stories that are made available by digitisation and the effective silencing of those stories that aren't. He goes further and states that the promoted stories are primarily about the west, white, and/or males, and that this normalises those specific histories.

I'd certainly agree that this is a major problem, I'm not necessarily convinced that this is a universal truth. Bulk digitisation has the ability to make available the hidden (previously unavailable) stories. Tim Sherratt's and Kate Bagnall's work springs to mind. As does the work on digitising gay and lesbian archives. But I'm not sure if these examples are notable precisely because they are (rare) counter examples to the dominant themes.

But the argument does highlight the logical outcome of Chris' question about the cost of digitisation and the consequent selection of what to digitise.

andrew

**From: Chris Hurley**

**Sent: Friday, 3 June 2016 4:54 PM**

**Subject:** RE: [archives-and-records-australia:1804] The value of digitisation

Strong stuff this :

... This is a remarkable thing – and as a historian it changes what I do in remarkable ways; but if you were to stand back and create a national, or even a global policy that selected what should be digitised and what should be available, it would look very different. In digitising this particular history we have inadvertently made it more exclusive and more conservative – ever more dead white western men – to the exclusion of the rest of the world ... in the last twenty years as part of ‘digitising’ the past we have both privatised our inherited culture and given a new hyper-availability to a subset of that culture – to the leavings of the same old rich dead white, men ... In the analogue world, archives and libraries were run ... as components of a national system of knowledge and memory. But in the rush to create a digital version of this we have simply handed the stuff ... to major corporations ... we have ensured that the objects read and desired by a western, educated elite – with money to spend in response to all those adverts – will be the kinds of objects that will be most easily available ... etc. etc.

I’m not sure I would say that libraries and archives were run in the analogue world as a “system” or that in pre-digitisation times resources were made available impartially and without privilege. But I would agree that some sort of underlying assumptions once existed about what libraries and archives were for, how access should be provided, to whom, and on what basis. These assumptions were, I believe, largely shared by those running them and working in them. There were always gaps between aspiration, rhetoric, and execution and the professions were sometimes even aware of those gaps and tried to close them. They were talked about in the schools and discussed at conferences (not enough, perhaps, but there was a self-conscious awareness that these were relevant matters for consideration, action, and improvement). It’s depressing how much our self-awareness has shrunk to a narrower focus on how rather than on why.

Mass digitisation could be said to have changed nothing – just a better how to satisfy the enduring why. But, in the rush to digitise, those same underlying assumptions - concerning what we are about – are being unselfconsciously eroded without sufficient consideration. We were once guardians of the documentary heritage and promoters of shared values as to its management and use. When digital assets are available for commercial exploitation or harvesting by third parties, not part of that fellowship, it is unrealistic for the guardians (even assuming they had the nous to understand the issue, the will to reach a shared conclusion, and the guts to carry it through) to expect to control consequences which are being driven by forces largely outside their power. What is emerging is an online gallimaufry of selected digital assets uncoupled from their source and available globally, utilising generic internet search capabilities. Those assets are largely beyond the reach of the source providers as to how they are managed, portrayed, and accessed. Some would argue that this decentralisation is a virtue (cf. [Peter van Garderen’s post](#) cited by Lise) because it liberates the assets from the dead hand of the source providers, implicitly denying the need for anything to replace a “national system of knowledge and memory”. Alongside all that, custodians with adequate technical and resource capability are deploying the digital assets they hold (some of which aren’t in sufficient demand to be commercially exploited or opportunistically decentralised) together with online descriptions of their own un-digitised assets but without comparable online search and discovery tools for their own descriptions, let alone in support of global discovery. Beyond that again, some of the third party providers with aspirations to being a one-stop-shop are harvesting online descriptions of un-digitised materials as well as high-value digital assets. That’s not to say that all this undirected digital activity is a bad thing (and the fact that it is undirected might, as Andrew suggests, mitigate some of the “exclusivity”).

What is a bad thing is making “digitisation” the answer to every question including the question “what are we doing and why are we doing it?” I’m sorry to keep harping on this but it really does come back to asking some fundamental questions, such as those set out in the [Modest Proposal](#). What are our functional requirements for global access? What do we want to achieve? What are we doing it for? And, once we’ve agreed what we should be doing, how does what we are in fact doing measure up? This is not about hoarding our assets and withholding them from commercial and third party exploitation, or about disallowing decentralisation initiatives, it is about asking whether those efforts (however meritorious) meet our requirements and, if not, what to do about it. Add to this an aspiration to do more than provide access to the gathered heritage resources we hold and the need for clarity of purpose becomes pressing because there’s lots for us to disagree about in all this.

As to the “selection of what to digitise”, that is a question that only occurs when digitisation is always the answer. I would like that question reframed as “what do we want to make available digitally?” Some suggestions:

- There must be seamless global access to all resources (both digital and non-digital).
- Inclusivity: Access must be available to material in the hands of barefoot archivists as well as rich, technologically endowed sources of content.
- Access must be provided for ungathered records as well as resources that have been gathered into an archival programme (the wholistic requirement).

As to how it is done, I would like us to reflect on content vs context search and discovery issues. Some suggestions:

- Search and display must provide for depth of description and display.
- Allowance must be made for the protean character of records (their changeability as to content and context).
- Provision must be made for authenticity (e.g. source vs rendition and transformation) and for differentiation (parallel provenance).
- Provision must be made for differential access and redaction (not just crude closure/release).

What I fear is digitisation coming to be seen as a fulfilment of our shared assumptions about access instead of a challenge to them. In the world of yesteryear, it was sometimes said that the finding aids shouldn't “push” researchers towards a conclusion (sometimes called spoon-feeding). They should have to struggle with the finding aids and find the stuff for themselves. Reference guides must not provide them with “easy” answers that might tempt them away from the struggle to find unfrequented resources unaided by us. That was allied to the view that finding aids should be objective; one of my major disagreements with Sue McKemmish and Marg Burns was over whether or not to introduce the term “patriarchal” into our descriptions of the Victorian Government in the C19th - an argument that could have been resolved with a dash of parallel provenance. Well, those views belong to another time and place, but I think they resonate with the issue raised by Tim Hitchcock. As to the even cruder matter of access to the un-digitised, there could hardly be a greater distortion of the research process than setting up systems that effectively lay down pathways frequented by most of our users by-passing the bulk of our heritage assets.

All the best

**From: Andrew Waugh**

**Sent: Saturday, 4 June 2016 9:45 AM**

(I just realised I wrote, but never sent this...)

Chris,

Ahh, my mistake. When you asked about the better 'bang for buck' of description, I interpreted that as asking about a comparison.

In thinking about description, I eventually realised that 'bang for buck' (or return-on-investment) is exactly the issue. ROI is the ratio of the benefits to the costs. From the point of view of ROI, it matters not if a course of action is cheaper if the benefits are reduced in proportion. It is easy to focus on the headline cost, and much more tricky to calculate the value of the benefits.

As I argued in my previous email, I believe the benefits of digitisation are significant. This is due to 1) massively improved access, and 2) reduced cost in preserving, providing access, and (potentially) storage of the physical original. It's very hard to quantify some of these benefits, of course.

On the other side, description is a many and varied thing, and the costs and benefits of description must vary enormously. The cost and benefit of understanding and producing a description of a complex set of record systems would be different to producing an individual Series text which would be different again to item level listings.

It is clear to me that the case for description, like digitisation, would be on a series by series basis. I certainly do not see that description would automatically be a useful technique for those series that aren't valuable enough to digitise. From an ROI point of view, it's quite possible that the records would not be worth describing either.



andrew

(As an aside, my view of the value of descriptions is coloured by the fact that the quality of the descriptions varies enormously. On the day I started writing this response, the NAA tweeted:

#FlashBackFriday Digital#infomanagement circa 1945 – the humble computer control room NAA B4498 128D6 [pic.twitter.com/voPjyN5EUO](http://pic.twitter.com/voPjyN5EUO)

Now this is complete nonsense. It's not the control panel of a computer - in fact I would suggest it controls the flows of material in an industrial process. But the date is nonsense as well. A colour photo in 1945 - possible but unlikely. The style screams the '60s or '70s - florescent light, wood veneer cabinets, the style of instrumentation, even the telephone handsets.

A quick look at the referenced description in the NAA

(<http://photos.naa.gov.au/photo/Default.aspx?id=9716244>) indicates that I shouldn't blame NAA public relations too much. They were just regurgitating what the item description said it was. Who knows where that came from; probably the agency.

The series description

([http://recordsearch.naa.gov.au/SearchNRRetrieve/Interface/DetailsReports/SeriesDetail.aspx?series\\_no=B4498](http://recordsearch.naa.gov.au/SearchNRRetrieve/Interface/DetailsReports/SeriesDetail.aspx?series_no=B4498)) says that this came from the Trade Publicity Branch - Film and Photographic Library which existed under various names from 1951 to 1986. It further claims that the series contents date from 1945 - which suggests where the purported photo date came from.

All this is a roundabout way of saying that descriptions can be nonsense. And, worse, that people uncritically believe the descriptions.)

From: **Chris Hurley**

Sent: **Sunday, 4 June 2016 4:01 PM**

**Subject:** Re: [archives-and-records-australia:1806] digitisation vs description

Andrew –

I think we are closer now to a meeting of minds but still at cross purposes.

First, as a critic of archival descriptive practices for more than 30 years, I yield to no one on the proposition that “descriptions can be nonsense”. But even digitised assets bear descriptions and where they are inadequate it is a problem for the digitisers also - ameliorated to be sure when coupled with text searching - but would that apply to a photograph as in your example? That leaves us agreeing that item and series level descriptions could be a lot better. In fact, if I were arguing that resources should be diverted from digitization and put into better item/series descriptions (which I am not – see below) your example rather tells against your own argument since the image in question obviously needs to be better described for its full value as a digitised asset to be realised.

Second, I am not in fact making that argument. I don't say that it is better to spend money on describing resources discoverable online instead of spending that money on digitising them. At any rate, not on describing them at the item/series level. Description, as you say, "is a many and a varied thing". I am perhaps at fault for not making myself clear. The descriptive endeavour into which I would argue resources should be diverted (the proposed calculation is purely hypothetical since the funding just isn't set up that way, this is about prioritisation) is of two kinds:

- at the contextual level, where I believe the kind of framework needed to support global access arrangements of the kind I have alluded to in the intervening postings in this thread can be developed; and
- at the structural level (relationships essentially), where I believe the implications of online discovery, both within and across descriptive programs, have not been properly understood and dealt with.

Exactly how is something I have some ideas on (indeed I have written on little else for many years) but I don't want to talk about that just yet in this context because that is implementation and my theme for now is that we must first settle on some functional requirements before moving on to the implementation. But one thing you can be sure of : simply presenting boring old series descriptions and interminable item inventories is not what I had in mind. My case for description is emphatically not "on a series by series basis".

So, we are still at cross purposes if you think I want to argue the merits of item or series level description as against digitisation. The idea that heritage assets not worth digitising aren't worth describing either is a profoundly interesting one. I think I alluded to it (or something very much like it) in an earlier post as a courageous proposition. If the stuff isn't worth digitizing or describing, is it even worth keeping? My assumption throughout has been that the assets held by archives and libraries need to be made accessible and are worth the effort to do so and that the debate is about how best to do that. It follows that some kind of descriptive solution is necessary for the vast bulk of undigitised material alongside the digitisation effort directed at the smaller component of the national asset. Not describing (or not even keeping) what isn't worth digitising is a completely different strategic proposition and that would involve a completely different set of calculations to the ones we have been discussing.

All the best

**From: Barbara Reed**

**Sent: Monday, 6 June 2016 9:36 AM**

**Subject: :** [archives-and-records-australia:1808] Digitising and historians (and archives/technology)

Hi all

In the wake of Chris and Andrew's really interesting exchange, Katherine Jarvie (Kathy Gallen) posted a link to this interesting provocation on Twitter this morning.

<http://historyonics.blogspot.com.au/2016/06/privatising-digital-past.html?sref=tw>

And a reminder that there's great some very interesting stuff on archival systems of tomorrow from Dan Gillean at ACA <http://www.slideshare.net/accesstomemory/technologie-proche-imaginng-the-archival-systems-of-tomorrow-with-the-tools-of-today> and Peter van Garderen's keynote on Archives, Technology and Innovation <https://www.youtube.com/watch?v=FL3UL90cG-k&feature=youtu.be>

Gratuitously

**From: Gene Melzack**

**Sent: Monday, 6 June 2016 10:13 AM**

**Subject: RE:** [archives-and-records-australia:1809] The value of digitisation

Hi Andrew,

A similar topic was addressed by Lisa Nakamura in her keynote address to the iPres conference last November, which can be viewed online: <https://vimeo.com/145196536>

She uses as a case study the 'guerilla' digitisation of a feminist text with access provided through the tumblr social media platform. The talk discusses the legal and social barriers that prevented the formal republication and digitisation of this marginalised text through official channels and the unofficial methods used by the similarly marginalised modern audience of the text to digitise and share the work with one another.

It's interesting both for shedding light on some of the factors that affect selection for digitisation within archives and libraries and for the access model promoted by 'guerilla' community digitisation and archiving, which speaks to Chris' question about the purpose of digitisation. In this case study, digitisation is not simply for the purpose of providing access to an archival object of historical interest, but also of providing new digital life to an object that is of ongoing interest as a living document. In general, physical objects must remain with an archive if they're to continue to exist as an object of study, but digital objects can be reused, repurposed, remixed, and recreated and can thus more easily become an integral part of contemporary culture in their own right. Not to say that there haven't always been artistic ways of reimagining and reinterpreting archival material, but that further enabling this kind of digital remix culture may well be a factor we want to consider when considering the purpose and value of digitisation.

Regards,

**From: Andrew Waugh**

**Sent: Wednesday, 8 June 2016 6:15 PM**

**Subject:** [archives-and-records-australia:1813] More on digitisation

This is a nice popular summary of the issues:

[http://www.nj.com/hudson/index.ssf/2016/06/digitizing\\_genealogical\\_records\\_not\\_as\\_easy\\_at\\_it.html](http://www.nj.com/hudson/index.ssf/2016/06/digitizing_genealogical_records_not_as_easy_at_it.html)

(originally tweeted by inside history).

The benefits, the costs (although not explicitly labelled as costs), and the need for description.

From: **Chris Hurley**

Sent: **Wednesday, 15 June 2016 8:57 AM**

**Subject:** [archives-and-records-australia:1816] More on mass digitisation

It is [reported from NZ](#) that objections are being raised to destruction of service records after digitisation – on operational rather than heritage grounds:

Submissions closed yesterday on the proposal to move to a fully digitalised record after just one month. Some Vietnam veterans ... say the government has a track-record of getting rid of evidence ... Many veterans already struggled to get their existing medical conditions recognised because of the mysterious holes in their records ... Auckland barrister Charl Hirschfeld ... said they were right to be worried. Some records are always lost in transferring them from one format to another, either due to oversight or policy, he said. Annotations and notes on the reverse of documents are not scanned, odd inserts are discarded. "Past experience shows when files are dealt with to rationalise them in an archive sense, parts of the file never make it onto the new format version ... However, Returned and Services Association chief executive David Moger, says said there were safeguards in place and the way in which they were being digitised meant they could be more "accessible".

This debate underscores some of the points canvassed recently on this list:

- if you don't destroy the hard copy, some of the cost advantages are lost;
- but some, for whatever reason, will want that anyway;
- digitisation makes the data more accessible;
- unless it is done well, evil may befall.

It is unclear from the report:

- whether the digitisation would go ahead in any case if the decision is to keep the hard copy;
- what search and discovery tools would be in place if both hard copy and digitised copies were kept available for use.

How interesting that the phrase "in an archive sense" is being used to connote rationalised management instead of heritage preservation; and depressing that the archivists are seen as possible co-conspirators in getting rid of the evidence.