STANDING THE TEST OF TIME: BUILDING BETTER RESILIENCE INTO ONLINE ARCHIVAL DESCRIPTIVE NETWORKS

Helen Morgan, Ailie Smith and Joanne Evans

Abstract: The advent of the web in the nineties allowed archivists to make information about archival holdings available online. Rapid developments in web publishing software mean that many archival institutions have migrated their descriptive data to new systems, changing their online presence. This paper discusses the outcomes from a project to update the design and informatics of two online archival gateways, the *Australian Trade Union Archives*, and the *Encyclopedia of Australian Science*. The revitalisation of these archival reference tools enabled significant review and reflection on issues around their sustainability and the sustainability of online archival descriptions they linked to. This included investigating what impacts changes have on discoverability of material and how we can ensure online descriptions of archival resources stand the test of time.

Introduction

It is now just over two decades since the creation of the World Wide Web by Tim Berners-Lee and colleagues at CERN, the European Organisation for Nuclear Research, in 1989/90.[1] Berners-Lee proposed using 'a distributed hypertext system' to address CERN's (and a growing number of other organisations) problems with 'loss of information about complex evolving systems'.[2] The Web has rapidly evolved into a ubiquitous and pervasive information infrastructure which now underpins much of our social and professional lives. Now in its second generation, the Web has shaken archival, along with many other, endeavours to their core. It challenges the applicability of our appraisal, access and description models, and questions the relevance of our institutional frameworks and professional expertise. It has opened up a myriad of new possibilities, and raised new archival imperatives. Chief amongst these is how we develop our practices and structures for a complex evolving digital and networked information age. When it comes to the Web, how do we ensure that our usage of it is archival, rather than ephemeral?

Archivists began using the Web in the mid to late nineties to make information about their archival holdings widely available. Rapid developments in web publishing software mean that many archival institutions have already had to deal with the migration of their descriptive data from at least one system to another. Those going online in the late nineties might have used HTML editing

tools to manually craft their finding aids into a series of web pages, whereas now the process may involve web pages being dynamically generated from one of many content management systems available.

Not only did the Web allow for remote discovery of an institution's archival holdings, but it also opened up exciting new prospects for the creation of archival networks linking related records held across a number of repositories. *Bright Sparcs*, an online register of biographical, bibliographical and repository data about the people who have contributed to Australia's scientific, technological and medical heritage, was an early example (1994) of this kind of archival network.[3] Following the documentation and publishing model of *Bright Sparcs*, the *Australian Trade Union Archives* gateway (amongst others) was established in 2001. These registers were both produced and published using methodology and software – the Online Heritage Resource Manager (OHRM) – developed by a group of archivists formerly known as the Australian Science Archives Project, now the eScholarship Research Centre, based at the University of Melbourne.[4]

In 2010 *Bright Sparcs* merged with its companion register *Australian Science at Work* (about industries and organisations) to form the *Encyclopedia of Australian Science*.[5] At the same time the *Australian Trade Union Archives* gateway was reviewed, its data audited, its design freshened up and the underpinning informatics brought into line with our latest thinking on the presentation of archival descriptive data. The impetus for this work came from the recognition of both of these registers as important contributors to scholarly research infrastructure and, specifically, the desire to export data using the EAC XML-Schema and establish OAI-PMH repositories for them to enhance this contribution on a national level.[6]

The revitalisation of these archival reference tools enabled significant review and reflection on issues around their sustainability and the sustainability of the online archival descriptions that they linked to. This conference paper examines this experience to address the question of how we might enable online descriptions of archival resources to stand the test of time.

We do, however, have to be careful when examining our past practices to not be too quick to damn them with 20/20 hindsight. Through their honest, open and critical examination we seek an understanding of how our thinking, systems, and capabilities have evolved. And where they have not (and should have), why not? There is no sign that the pace of technological and social change in the online world is abating, so adaptive, dynamic and emergent need to be key characteristics of our techniques, tools, systems and selves. Thus this conference paper also aims to explore ways in which to dialogue, discuss and critique in search of better practice models.

Australian Trade Union Archives

The *Australian Trade Union Archives* gateway was established in response to concern about issues relating to archives in the national research infrastructure and the call 'to enable the development of cooperative documentation strategies by archival institutions'.[7] Approximately 80% of archival records relating to trade unions in Australia are held by the Noel Butlin Archives Centre at the Australian National University in Canberra and by the University of Melbourne Archives. The remainder is held by between twenty to thirty other university and state repositories throughout Australia. In 1999, at the time this project was proposed, the best way to get an overview of archival record holdings in Australia was to search the National Library of Australia's *Register of Australian Archives and Manuscripts* (RAAM) online database.[8] RAAM had been established in 1997 and included data originally published in the *Guide to Collections of Manuscripts Relating to Australia* over the period 1965 to 1995.

A central task of the original *Australian Trade Union Archives* Project was to build the historical contextual framework about creators, custodians and other stakeholders necessary to better understand the records of Australian trade unions. This was the detail lacking in RAAM, particularly the relationships between creators as they changed over time. Without this information, as the project proposal noted, 'the usefulness of minute books, membership lists, correspondence, newsletters… is severely limited'.[9] The data in RAAM, collected over a long period and supplied by the collecting institutions, was a key starting point for information about the record holdings.[10] This data was brought into the *Australian Trade Union Archives* Project's database and augmented through further research. In particular, citations to online finding aids and descriptions to record holdings in online catalogues were actively sought out and recorded in the database, while at the same time explicating the complex network of relationships between organisations, people, records and other resources gleaned from the records of holdings and where necessary the records themselves. Fifty-nine per cent of archival repositories and collecting institutions containing trade union material had descriptive data about their holdings online during the original phase of this work in 2001.[11]

The resulting online gateway, listing almost 2,500 biographical entries (trade unions, peak bodies, government agencies and people) with references to more than 1,400 archival resources was published in 2001. The perennial problem of funding meant that the data wasn't continuously curated – no new entries were added – but the gateway continued to provide a valuable central access point to information about labour history for researchers and scholars.

Providing national scholarly research infrastructure

When the National Library decided to combine its multiple discovery services, including RAAM, Picture Australia, Libraries Australia and the fledgling People Australia project into one single discovery service (Trove), the eScholarship Research Centre provided one of the first sets of external biographical data (about Australian women) to be harvested into it.[12] The involved the use of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to share EAC data.[13] *Bright Sparcs* was identified as another potential contributor of data. The Centre was successfully funded under the University of Melbourne's Scholarly Information Innovation grant scheme to further the goal of enabling 'the harvesting of contextual metadata into scholarly communication networks and management frameworks'.[14]

We focused on the three online registers *Bright Sparcs*, *Australian Science at Work* and *Australian Trade Union Archives*. We had long identified the need to bring data about the people involved in Australian science (*Bright Sparcs*) together with data about the scientific industries and organisations (*Australian Science at Work*) to better identify and expose their integral relationships. The two science-related registers were combined into one – the *Encyclopedia of Australian Science* – and successfully harvested into Trove.

These entity data types (people and organisations) had been together in *Australian Trade Union Archives* since its inception; however, the data model in use was still the original one we began publishing with back in 1994, which, in terms of discrete units of information, presented the archival data only from the biographical point of view. This needed rectifying in order to create data of a standard ready for OAI-PMH harvesting.

Working with minimal free form archival descriptive metadata

The *Australian Trade Union Archives* Project, like other projects undertaken at that time using the OHRM system, carried the compromises of a flat paper world (abbreviation, repetition of data and brevity all round) into the online environment. Harvesting had taken what was available from RAAM, which in turn had taken what was available from old paper guides. Commonly this was a repository's controlling number and description of the contents of record groups along with the name of the repository – archival data at its barest. A record was created in the OHRM database for each unique description of records relating to trade unions, only gathering the minimal data required to make sense in our system *at this time*, the major addition being links to online finding aids or links into repositories' own catalogue records about the holdings.

In the original *Australian Trade Union Archives* (2001-2010) web pages, each biographical entity had its own page with biographical data, linked to another single page listing all the data collected about all archival holdings, grouped by repository, relating to that entity. The relevance of the archival holdings, indeed often the content of the archival holdings (versus the formats), was derived from their listing on a page bearing the title of a biographical entity (Figure 1).
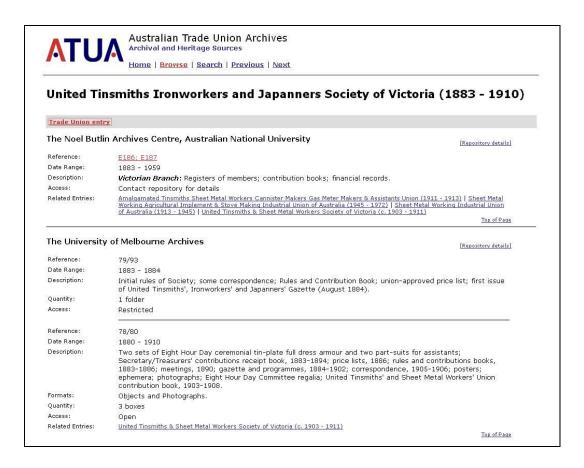

Figure 1: Original web page for archival holdings related to the United Tinsmiths Ironworkers and Japanners Society of Victoria trade union, including reference to NBAC records (E186; E187).

Minimal descriptions were preferred.[15] The entity's name (or sub-entity's name, such as a branch) was sometimes added in bold to descriptions if the original description was unclear about which records related to which entry in a record group where multiple entries (provenances) were documented. Many descriptions of records simply followed the format: 'Minutes; correspondence; circulars'.

In the model that we eventually moved to, each records description (captured uniquely in the OHRM database) was expressed in the web output on its own page, supporting a more open view of the data, unconstrained by but supporting multiple views and relationships (Figure 2).

ARCHIVAL RESOURCES DETAILS

Sheet Metal Working Agricultural Implement & Stove Making Industrial Union of Australia - Records

Title	Sheet Metal Working Agricultural Implement & Stove Making Industrial Union of Australia - Records
Repository	The Noel Butlin Archives Centre, ANU Archives Program
Reference	E186; E187
Date Range	1883 - 1959
Description	Victorian Branch: Registers of members; contribution books; financial records.
Access	Contact repository for details

Related entries

Trade Unions

- Amalgamated Tinsmiths Sheet Metal Workers Cannister Makers Gas Meter Makers & Assistants Union (1911 - 1913)
- Sheet Metal Working Agricultural Implement & Stove Making Industrial Union of Australia (1945 - 1972)
- Sheet Metal Working Industrial Union of Australia (1913 - 1945)
- United Tinsmiths & Sheet Metal Workers Society of Victoria (c. 1903 - 1911)
- United Tinsmiths Ironworkers and Japanners Society of Victoria (1883 - 1910)

Figure 2: New web page for NBAC records (E186; E187) related to the United Tinsmiths Ironworkers and Japanners Society of Victoria trade union, a predessessor union to the SMWAI&SMIU.[16]

A biographical entity page now only includes a summary description of related archives, and links to more detailed pages for each related archival holding (Figure 3):

Archival resources

The Noel Butlin Archives Centre, ANU Archives Program

- Sheet Metal Working Agricultural Implement & Stove Making Industrial Union of Australia - Records, 1883 - 1959, E186; E187; The Noel Butlin Archives Centre, ANU Archives Program. Details

Figure 3: Detail from new (biographical) web page for United Tinsmiths Ironworkers and Japanners Society of Victoria, providing summary description for NBAC records (E186; E187).[17]

The biggest issue in moving to this more open and flexible model was the 'simple' fact of only 21 per cent of archival descriptions having had title metadata captured during the original phase of the project.[18] Not only did our system at the time not require this data but the (very necessary) unique numeric identifier appears to have been deemed adequate for archival management and access purposes in the pre-Web world originally supplying the data. Under our earlier publishing model we addressed this by taking advantage of the ability online to

link to related biographical entity pages. Where titles had been ascribed to archival holdings this was usually the ubiquitous 'Papers' or 'Records' – meaningless on their own.[19] In Figures 2 and 3 replace all occurrences of the words 'Sheet Metal Working Agricultural Implement & Stove Making Industrial Union of Australia' with [Untitled] or leave it as 'Records' and the implications of the problem are clear. To address this, where no title existed, we established the protocol of title equals creator (last active provenance) followed by the word 'records'.[20]

Are our finding aids documents to be read from start to end, or are they databases to search and explore from any entry point?

The value and integrity of linked data

Recognising the role these online archival descriptive networks play in research infrastructure, we have an organisational commitment to maintain the information we publish. Persistent, citable URLs are critical to providing continuity and integrity of service for users. When, after 16 years, we merged *Bright Sparcs* with *Australian Science at Work* to form the *Encyclopedia of Australian Science*, we established permanent redirects from the original URLs, anticipating the likelihood that material had been cited.

At the same time we undertook a major audit of the *Australian Trade Union Archives* data, checking the reference and finding aid links collected back in 2001, and adding new links discovered during the title audit. The majority of the original links were broken and it took much frustrated slogging through online catalogues to both verify the location and existence of holdings and discover new citable links/URLs to them.[21]

Taking the broken link/URL as a starting point revealed that most institutions who had migrated to newer systems had failed to provide even the simplest of redirects to a useful page that could help users find the data they had been sent in search of. We encountered many 404 error pages which simply said 'page not found' or 'request cannot be completed'. Some occasionally redirected to a search page. In other cases even locating a search page proved difficult. We attempted to locate the archival holdings based on the descriptions and reference/manuscript numbers originally gathered, an exercise which revealed that sustainability of archival descriptive metadata is not only a recent web-related problem. Searching by provenance gave the best results.

In preparation for this paper, we revisited the URLs for archival resources to compare them with the results from two years ago.[22] Looking at the revised data in 2012 showed that not only were there more unique URLs assigned to archival records, but a much larger proportion were still working, or were

redirecting to a relevant page. Figure 4 shows a comparison of URLs between the two analysis periods.
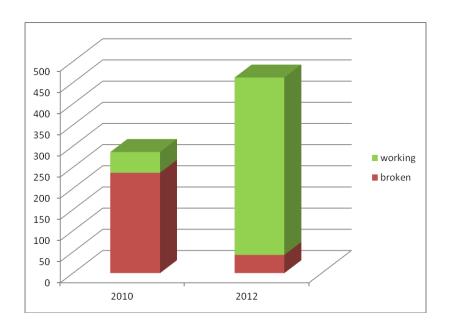

Figure 4: Overall comparison of archival resource URLs in the *Australian Trade Union Archives,* 2010 and 2012.
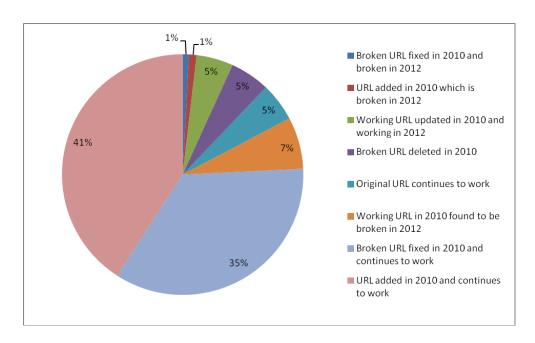

Figure 5: Comparison of URLs for individual archival resources between 2010 and 2012.

Analysing the fate of individual URLs across the analysis periods showed us that the most common outcome was that new URLs were added in 2010 to resources that were previously not linked to a website, and these URLs were found to be still working in 2012. A slightly smaller percentage of URLs were the group that

were found to be broken in 2010 and subsequently edited, and continue to be working URLs in 2012. These two groups make up over three quarters of the total URLs. There were also cases where URLs that had been working in 2010 were now broken, and where broken URLs had been edited or added, but the new URLs were now broken. In more positive news there were a small percentage of URLs that had been working when they were checked in 2010, and still continue to function in 2012. Figure 5 displays a breakdown of the comparison of individual URLs across the two periods of analysis.

Where to?

But in another ten years, will an audit of links to online catalogue records and finding aids reflect the only slightly more comforting situation of Australian repository data revealed by our two year audit (2010 to 2012) or the sorrier situation of the data noted over the period 2001 to 2010? Will more archival institutions and archivists have become the champions that the records they hold in trust deserve?

Endnotes

[1] http://info.cern.ch/

[2] Tim Berners-Lee, 'Information Management: A Proposal', March 1989, May 1990, http://www.w3.org/History/1989/proposal.html

[3] *Bright Sparcs* was first published online by the Australian Science Archives Project in 1994. It was subsequently published by the Australian Science and Technology Heritage Centre (1999-2006) and the eScholarship Research Centre (2006-2010). *Bright Sparcs* was amalgamated with *Australian Science at Work* to form the *Encyclopedia of Australian Science,* eScholarship Research Centre, http://www.eoas.info (2010-).

[4] The Online Heritage Resource Manager (OHRM) has been developed over a number of years by staff members at the University of Melbourne's eScholarship Research Centre and its predecessor organisations. It provides a framework for mapping contextual information, including published, archival, heritage and digital resources – http://www.esrc.unimelb.edu.au/ohrm/

[5] *Australian Science at Work*, published by the Australian Science and Techology Heritage Centre (1999-2006), and the eScholarship Research Centre (2006-2010). From 2010 it formed part of the *Encyclopedia of Australian Science*, eScholarship Research Centre, http://www.eoas.info/

[6] Encoded Archival Context for Corporate Bodies, Persons, and Families (EAC-CPF) – http://eac.staatsbibliothek-berlin.de/. Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) – http://www.openarchives.org/pmh/.

[7] Resolution 2, 10th National Scholarly Communications Forum, National Archives of Australia, Canberra, November 1999, which continued '[University of Melbourne to initiate an ARC funding application]' – http://www.asap.unimelb.edu.au/asa//aus-archivists/msg03483.html – accessed 26 July 2012.

8 RAAM was established online by the National Library of Australia in 1997. It was decommissioned in 2009-10 and redirected to the Library's Trove discovery service. For a copy of the original press announcement about RAAM see the H-Net Discussion Network H-ANZAU list, 27 June 1997 – http://h-net.msu.edu/cgi-bin/logbrowse.pl?trx=vx&list=h-anzau&month= 9706&week=d&msg=BHDhjW9eP/lRPAr2WNLGPA&user=&pw= – accessed 26 July 2012.

9 Research Infrastructure Equipment and Facilities Scheme Application for the Australian Trade Union Heritage Resource Gateway, 23 April 2001, held by the eScholarship Research Centre, The University of Melbourne.

10 Subsequent difficulties in locating records noted in RAAM indicate that sustainability of archival descriptive metadata is not only a recent web-related problem.

11 840 of a total 1420 archival resources (not unique repositories) had online citations, of which 6.6% (56) were redundant links back to RAAM or the Directory of Archives in Australia.

12 Data from the Australian Women's Register – http://www.womenaustralia.info/

13 For more on this process see Basil Dewhurst, 'People Australia overview', ARDC Party Infrastructure Project, National Library of Australia wiki, 1 June 2010 – https://wiki.nla.gov.au/ display/ARDCPIP/People+Australia+overview – accessed 26 July 2012. See also http://www.openarchives.org/pmh/

14 Scholarly Information Innovation Grants 2009 application, Dr Joanne Evans, eScholarship Research Centre, The University of Melbourne.

15 The original data entry protocols also reflected the difficulties encountered in working with archival descriptions from multiple repositories and divergent descriptive practices.

16 http://www.atua.org.au/archives/Larc01205.htm – accessed 27 July 2012.

17 http://www.atua.org.au/biogs/ALE1049b.htm – accessed 27 July 2012.

18 1122 out of a total of 1420 archival resources had no title metadata.

19 By title we mean words that describe the level of holding documented in the *Australian Trade Union Archives* database, from collection level down to item level.

20 In consultation with the two largest providers of archival data to the project, Noel Butlin Archives Centre and the University of Melbourne Archives. In the case of individuals, title equals the name of the individual followed by the word 'papers'.

21 We would like to thank eScholarship Research Centre staff member Melissa Downing for her work on the *Australian Trade Union Archives* title and link audits, along with Maggie Shapley, Noel Butlin Archives Centre, and Sue Fairbanks, University of Melbourne Archives.

22 There were several cases where a number of resources had been linked to a single higher level URL for the repository where the archives were held. We have decided to omit exact duplicates of URLs in cases such as this, and focus on unique URLs.