Archives and Manuscripts, 2014
Vol. 42, No. 2, 187–189, http://dx.doi.org/10.1080/01576895.2014.911685

Routledge
Taylor & Francis Group

Building an integrated digital archives (Part II)

Richard Lehane*

Richard Lehane is an archivist at the State Records Authority of NSW. He is a member of the digital archives team, who are undertaking a three-year project to build a whole-of-government digital archive for New South Wales. Richard also works on State Records' Open Data project, <*http://data.records.nsw.gov.au*>, and new search engine 'Search', <*http://search.records.nsw.gov.au*>.

Integration is at the heart of the archival endeavour. You can read most archival methods (appraisal, arrangement and description, access) as being fundamentally about integration: the task of creating a coherent archives that incorporates disparate recordkeeping systems. Sue McKemmish contends that records are ever in a state of 'becoming'.[1] The continuum model suggests that the same is true for archives, that we are constantly re-creating archives as we integrate new records and recordkeeping systems with them over time. With paper records, this integration happened above the 'item' layer through the documentation of ambient and provenancial context (that is, descriptions of series, functions and so on). With digital records, we have an opportunity to support much deeper integration. David Bearman devotes a chapter of *Archival Methods* to this opportunity (and challenge):

> Over the past several years, the proliferation of on-line databases and machine readable information sources has made information scientists painfully aware that the problem of intellectual transportation across disciplinary perspectives is not resolved by making data available on-line or in full-text. Indeed it may be exacerbated, since in manual retrieval systems the human mind makes leaps across categories which are not supported by existing mechanisms in automated systems. As a practical matter, if we are to integrate a variety of externally developed databases into archival information systems in order to provide for retrieval without much in-house description of records, we need to determine how we can best make large machine readable data stores, consisting of a variety of sources, each collated for particular purposes and audiences, accessible through a single user interface.[2]

In my presentation to the Archival Methods workshop I described how we are approaching this task at State Records NSW. Our goal is to create an integrated digital archives that:

- has simple, intuitive interfaces that can be used by general users to explore the digital archives as a coherent whole;
- ensures that government agencies can continue to rely on their digital records post-transfer and continue to use them, ideally seamlessly (by connecting their current business systems to the digital archives-as-backend);

*Email: Richard.Lehane@records.nsw.gov.au

- integrates across, and not just above, agency recordkeeping systems with a capacity for structured querying of the contents of the whole digital archives.

What we want to avoid is a digital archives that is merely a container for siloed systems that must each be approached and interrogated individually.

State Records NSW is adopting a metadata mapping strategy as the centrepiece of our integration strategy for the digital archives. This means identifying the metadata and structured data in digital recordkeeping systems, mapping to a controlled schema and transforming it, to enable search and access against a common set of terms. Interestingly this is an approach that Bearman identifies and discards in *Archival Methods*. His two objections are that it is labour intensive and semantically messy (mappings can never be precise and you can easily end up distorting meaning).[3]

To Bearman's first charge we admit that yes, mapping and transforming metadata according to a common schema is labour intensive, but in fact our whole approach to digital preservation at State Records NSW is labour intensive (our approach is to develop customised migration plans for individual systems, rather than attempt to construct a single automated workflow for everything). Since we propose taking this case-by-case approach anyway, we can incorporate the mapping of metadata and structured data into the migration methodology without adding a great deal of additional labour to projects.

To Bearman's second charge we can offer only a partial defence. A mapping strategy that would indeed be semantically brutal is mapping to a fixed vocabulary (defining a preferred schema at the outset, and then forcing agency-created metadata and structured data to conform to that schema). We are not taking this approach at State Records NSW. What we propose instead is to create a metadata registry that is a growing vocabulary of preferred terms with defined constraints. We will record preferred terms in this metadata registry and, when we encounter metadata and structured data that match those preferred terms, we will map and transform it accordingly. But it is an open vocabulary so that when we discover metadata and structured data for which there are no appropriate mappings, we can register new terms (preferably based on existing vocabularies but, where necessary, coining new terms). The metadata registry will be an online resource that users can consult when constructing queries for the digital archives. It will also be available as a best-practice guide for New South Wales agencies to consult when they are considering what metadata terms to use in new recordkeeping systems. The metadata registry makes heavy use of linked data technologies; the terms themselves will comply with the RDF data model and we anticipate supporting SPARQL queries over the digital archives.

My reason for giving this paper's title a 'Part II' appendage is that this metadata mapping approach only addresses part of the challenge of an integrated digital archives. It takes us some of the way but leaves a critical piece out: we also need to place the systems within the digital archives in the wider context of the archives as a whole. At State Records NSW, this means integrating the digital archives with the rest of the state archives collection. But perhaps I should have used 'Part 1.01' instead as I have not left myself room for more than a sketch here ... which in fact is convenient because we have not settled on solutions for this problem yet, it is still very much a work in progress and what follows are mostly personal opinions.

In Australia we are fortunate that our relational model of description and the basic entities of the series system provide a solid foundation for describing digital archives. But we should not rest on our laurels. Archival institutions around the country are presently considering the deployment of a new generation of descriptive systems and it

would be foolish to just re-create the old systems with new technologies and not consider what changes can be usefully made to our current descriptive practices.

The key change that I think needs making is the removal of many of the unnecessary rules and constraints in our current systems. Rules such as that an agency must have a single name, or a series only one descriptive note. Rules that constrain the types of relations we can assert between entities and the recordkeeping structures that we can represent. Rules that prescribe and limit the attributes we can attach to those entities. I am not suggesting that we abandon all constraints: fixed points (Hurley's datums[4]) are necessary for uniquely identifying and relating entities. But our underlying models can be simplified (as Hurley shows with his deeds, doers and documents[5]), we can enable greater flexibility in relations and structures (to better fit recordkeeping systems) and, just as State Records NSW's metadata registry is an open schema that allows the addition of new terms over time, I think we can similarly open up the schemas of our descriptive systems to support change and adaptation over time.

A more flexible and open descriptive system would provide space to experiment with different ways of documenting digital recordkeeping systems. At State Records NSW, for example, we are researching the histories and use of digital recordkeeping systems during migration projects and writing screenshot-illustrated descriptions of those systems. These texts could be squeezed into existing fields in our database such as 'descriptive note', but why? In fact, why impose a limited set of categories ('administrative history', 'biographical note', 'archivist's note' and so on) at all on the stories that we can tell about archives? Why not allow archivists (and possibly users too) to attach arbitrary descriptive texts to any of the entities in our catalogue? This would approach Tom Nesmith's notion of a descriptive system that comprises descriptive data overlaid with 'essays' on diverse subjects and themes, and including essays contributed by users.[6] It would also provide us with a means to bring all of those additional finding aids that we create (indexes, subject guides and so on) into a single, inter-connected descriptive system.

Our descriptive systems should not be disciplinary straitjackets; they should be platforms that free us to write rich and nuanced documentation. State Records NSW's metadata registry shows that we can build systems that are open yet still controlled. We should do the same for this next generation of descriptive systems too.

Endnotes

1. Sue McKemmish, 'Traces: Document, Record, Archive, Archives', *Archives: Recordkeeping in Society*, Sue McKemmish, Michael Piggott, Barbara Reed and Frank Upward (eds), Centre for Information Studies, Charles Sturt University, Wagga Wagga, 2005, p. 20.
2. David Bearman, 'Chapter V: Intelligent Artifices: Structures for Intellectual Control', in *Archival Methods*, Archives and Museums Informatics, Pittsburgh, 1989, available at <*http://www.archimuse.com/publishing/archival_methods/*>, accessed 11 March 2014.
3. ibid.
4. Chris Hurley, 'The Hunting of the Snark: Searching for Digital Series', Sydney Recordkeeping Round Table, October 2011, p. 10, available at <*http://www.descriptionguy.com/images/WEBSITE/hunting-of-the-snark-search-for-digital-series.pdf*>, accessed 11 March 2014.
5. Chris Hurley, 'Documenting Archives and Other Records – A Guide for Dummies', August 2008, available at <*http://www.descriptionguy.com/images/WEBSITE/Documenting-archives-a-guide-for-dummies.pdf*>, accessed 11 March 2014.
6. Tom Nesmith, 'Re-opening Archives: Bringing New Contextualities Into Archival Theory and Practice', *Archivaria*, no. 60, Fall 2005, pp. 12–14.