

Rethinking approaches to recordkeeping metadata

Barbara Reed*

Barbara Reed of Recordkeeping Innovation is a consultant in the field of records, archives and information management with more than 25 years of industry experience in Australia and the Asia Pacific region. Previously she was an academic in recordkeeping at Monash University. Areas of special interest include digital recordkeeping strategies, recordkeeping metadata and standards development for whole-of-government initiatives.

As one who has been involved in the development of standards, and particularly metadata standards for records and recordkeeping, since the mid-1990s, and as yet another metadata element set for recordkeeping metadata is about to be published as a standard,¹ I find myself questioning the approach employed to date, why it has not worked and thinking a bit differently about how to approach recordkeeping and metadata. These reflections should be seen as personal opinion and as a work in progress.

The word metadata is now emerging from the lips of our politicians in quite astounding ways – ‘it’s just metadata’,² the politicians say. But any coherent response to the issues that are raising the social ire in relation to the wholesale collection – but more scarily, the ‘repurposing’ – of this metadata must start to address recordkeeping issues. And the January response of President Obama to the unmonitored and unsanctioned overreach of the National Security Agency’s metadata collection³ clearly places the argument about recordkeeping in the highest of public arenas for scrutiny. Metadata, and particularly metadata about records, which detail transactions and are routinely collected for business purposes, is the stuff of big data, data mining, semantic web relationship linking and other emerging analytic technologies coming down the line. Metadata is also at the core of the social concerns over privacy, the capacity of individuals to control their personal information and the acceptable limits of data linkage.

And yet, in this environment our recordkeeping metadata standards have not gained traction. Our key message is not being heard.

Every community is busy developing metadata standards for its own purpose – as is quite valid. But recordkeeping metadata – the stuff of the transaction – is not well articulated or understood beyond (or within?) our own community. Metadata is everywhere, the world is awash with metadata, the Internet works on metadata, our Google searches depend upon and create metadata, a 140-character tweet is dwarfed by the metadata it creates.

How well do we as recordkeepers really understand metadata as it is created, generated and exploited in the world of vastly distributed networks? Not well enough, is my contention, and nor do we understand that our approaches of imposing rigid and fixed control are just not going to work in this complex, chaotic, digital world. And this level of misunderstanding or inadequate understanding of the role of metadata in the

*Email: B.Reed@records.com.au

digital world comes from both sides of the lifecycle divide – both from the archives world and from the records management world.

Archival understandings of metadata – a characterisation

From the archives, we have a profession who is still mainly dealing with paper, as the paradigms have not really shifted from the paper realities in which organisations were creating records. Archival institutions, by and large, when they operate with a custodial threshold mindset, are operating about 20 years behind the reality of the creating organisations. So the paper mindset is still very strongly held in most institutions. If Bearman⁴ argued, provocatively but convincingly, that our archival methods were failing before the digital tsunami struck, imagine how impossible our unchanged archival aspirations are in the digital world. Metadata for the archival community is still largely being seen as handcrafted, post-hoc, descriptive metadata. Standardisation is a goal, but with a view to unifying practice, or to supporting enhanced information exchange between archival repositories. This model is, at its base, built on library practices. Impressive individual implementations can be cited using these protocols,⁵ but they are still largely reliant on handcrafted, ‘cataloguing’-based, post hoc approaches.

The work of the recently established International Council on Archives Experts Group on Archival Description⁶ is ‘charged with developing a formal conceptual model for archival description that identifies and defines the essential components of archival description and their interrelations to promote a shared understanding of archival description, to facilitate the development and use of archival descriptive systems, to enable national, regional, and international collaboration, in the archival community as well as with allied cultural heritage communities’. Why? When such a model already exists in the ISO world? But of course, that ISO model is called ‘metadata for records’ and never should the worlds of archives and records coincide, it seems!⁷

And yet, from the archival world, even with its post hoc descriptive representation of metadata, we have some great exemplars from which to build. We have the example of digital historians willing to exploit what metadata archival description can provide in intriguing new ways.⁸ We have one or two exemplars of how archival descriptive metadata can be linked to open data initiatives.⁹ But none of these sets of exemplars is wholly dependent on sets of perfect ‘nice’ ISAD G metadata. Rather they work with what they can get.

Records management

And from records managers, we have international standards, the ISO 23081 series. I suggest that these standards were established well before the records community as a whole had reached any maturity or broader understanding of how to deal with metadata. Individual jurisdictions, particularly in Australasia, have established metadata element sets to support the broader conceptual models of the ISO 23081 standard. But in almost all cases, the metadata element sets are disrespected by the product vendors, who profess ignorance. Recently the head of information management of one of the top five international records management software vendor companies professed complete ignorance of the existence of these standards, and then commented on their pithy and memorable name – perhaps a lesson here that we need a catchy title!

Professionally, as is the nature of the personality trait of our profession,¹⁰ these element set standards reflect an obsession with detail, and frankly fail to convey the

main point of recordkeeping in ways that can be communicated to other disciplines and professions. Perhaps we have been too prescriptive about ‘describing’ the content, underselling the pivotal role of relationships, and not effective enough in emphasising the process. What happens to the information as record are the core things that mean we can rely on and assert authenticity of transactions. We can manage content metadata via other means – mapping, bulk association of records with functions or activities, automatic analysis and extraction at point of transfer or receipt (ingest if you must) from systems. Is this likely to reinforce the divide between the archival and the records management world because archivists in custodial mode do not really care about this process metadata? Perhaps we have to wait longer until the custodial mode is not the only relevant model to demonstrate the core requirement for this ‘provable’ metadata. Our exemplars here should be from the data migration community using metadata brokers in innovative ways the few (but increasing) imaginative uses of metadata mapping techniques to overcome disparate metadata structures.

So, what is to be done?

As always, some of the issue comes down to expressing the core essence of record-keeping – why it is essential to deal with the who, when, what and where of transactions so they carry the authenticity and reliability of the actions they reflect or embody.

We need the archives world to stop thinking post hoc description and start thinking inheriting metadata. We need the records management world to stop insisting on the narrow prescriptive view of records metadata that many have, and embrace diversity, difference and multiple creating environments. We need the recordkeeping world to unite in seeing that their models for records need to be ‘harmonised’ or merged to create coherence so that records created in one environment will last in multiple environments (including archives) over time.

The ISO 23081 Part 2 conceptual models are the most powerful conceptual models that we have professionally. They can be communicated, although I would suggest that they, too, need to stress the process metadata a little more – it is that transactional process metadata (the stuff that the National Security Archive is using) that is the power of recordkeeping.

Do we need the prescriptive types of metadata element sets for records and archives? I am coming around to the opinion that we do not, and that adherence to those models is a hangover from the ‘command and control’ mentality of the paper past that is just unsustainable in the digital age. Rather, we need robust conceptual models that clearly communicate. This will perhaps stop the insistence that the recordkeeping components of some sets (where they have some notion of the recordkeeping needs) are simply ‘administrative’ metadata.¹¹ We need to communicate the ‘meta’ message about recordkeeping metadata.

Endnotes

1. Standards Australia, AS 5478:2014, ‘Recordkeeping Metadata Properties Reference Set’, publication pending.
2. An umbrella phrase for those used by a variety of politicians. See for example Nicola Roxon, Attorney General Commonwealth of Australia, quoted in Josh Taylor, ‘Roxon Calls for Cold Shower on Data-Retention “Hysteria”’, 5 September 2012, available at <http://www.zdnet.com/au/roxon-calls-for-cold-shower-on-data-retention-hysteria-7000003742/>, accessed 1 March 2014, or Senator George Brandis, Question Time, Australian Senate, 3 December

- 2013, available at <http://scott-ludlam.greensmps.org.au/content/questions-without-notice/questioning-attorney-general-about-surveillance-overreach>, accessed 1 March 2014.
3. United States President Barack Obama, 'Transcript of President Obama's Jan. 17 Speech on NSA Reforms', *The Washington Post*, 18 January 2014, available at http://www.washingtonpost.com/politics/full-text-of-president-obamas-jan-17-speech-on-nsa-reforms/2014/01/17/fa33590a-7f8c-11e3-9556-4a4bf7bcb84_story.html, accessed 1 March 2014.
 4. David Bearman, *Archival Methods*, Archives and Museum Informatics Technical Report No. 9, Archives and Museum Informatics, Pittsburgh, 1989, available at http://www.archimuse.com/publishing/archival_methods/, accessed 1 March 2014.
 5. For example, the work of the UK Archival Portal, Archives Hub, available at <http://archiveshub.ac.uk/isadg/>, last accessed 1 March 2014.
 6. EGAD – ICA, available at <http://www.ica.org/13799/the-experts-group-on-archival-description/about-the-egad.html>, last accessed 1 March 2014.
 7. ISO 23081-1, Information and Documentation – Metadata for Records – Part 1, Principles; ISO 23081-2, Information and Documentation – Metadata for Records – Part 2, Conceptual and Implementation Issues; ISO 23091-3, Information and Documentation – Metadata for Records – Part 3, Self Assessment Method.
 8. The works of digital humanitarians or digital historians are multiple, but I continue to cite my two favourites, the work of Dr Mitchell Whitelaw, see examples of his work on <http://visiblearchive.blogspot.com/>, accessed 1 March 2014 (particularly Archive Series Browser) and the work of Dr Tim Sherratt, see <http://discontents.com.au/>, last accessed 1 March 2014 or <http://invisibleaustralians.org/faces/>, last accessed 1 March 2014.
 9. See the work of the NSW Digital Archive Team in linking to government websites such as 'publications.nsw.gov.au'.
 10. Ann E Pederson, 'Understanding Ourselves and Others: Australian Archivists and Temperament', *Archival Science*, vol. 3, no. 3, 2005, pp. 223–274.
 11. Dublin Core and METS, for example, both have core recordkeeping metadata about the scheme subsumed under 'administrative' metadata.