

Metadata: a contestable concept?

Barbara Reed, Director Recordkeeping Innovation Pty Ltd

Firstly, thank you for the invitation to participate in this event honouring the work of our very esteemed colleague Hans Hofman. It is a great privilege to be asked to contribute to this event, and a pleasure to be here.

This also explains the topic which I have chosen to address. Hans has been a long time advocate of recordkeeping metadata and the transition of records/archives to the digital world, and it seems appropriate to address this topic which has been one of the foci of his national and international work over the last 15 years. As I too, have been involved in the recordkeeping metadata world for much the same time, and have been challenged recently by those who have said – so much promise, why hasn't the world imagined in our early work eventuated. So I've taken this opportunity, in the spirit of provocative thinking from the recordkeeping roundtable tradition, to revisit and perhaps look to reinventing, some of the thinking and implementation practice surrounding recordkeeping metadata.

Metadata: a contestable concept, or Recordkeeping metadata in the news.

So firstly, let me explain the context of the title of my talk. Like many others, I have been intrigued and provoked and challenged by the transformation of the conversation about records which has occurred in the last 3 years on the front pages of our newspapers. Since WikiLeaks released the 'collateral murder' video in April 2010, but more particularly since their release of the Diplomatic cables from November 2010, records have been front page news. This has continued through to the Snowden leaks in May 2013, and prompted a major ongoing social debate about records, metadata, the role of governments, transparency, accountability, privacy, the nature of networked communications and increasingly citizenry expressing distrust of their governments provoking engagement and activity from data activists.

Over the past 3 years in collaboration with a number of colleagues I have been trying to make sense of our professional role as recordkeepers when confronted by these events and activities which are absolutely and centrally located to our professional mandate – the business of recordkeeping. We have been presented with problems, issues, dilemmas but also interesting new takes on how to access and how to interpret digital records. I and a number of my colleagues have tried to use these examples to reinvigorate a debate about our professional role and its sustainability with our current professional methods. I accept, grudgingly perhaps, that our profession is never going to be an activist profession, however, the dangers in ceding ground completely, either by complete abrogation of responsibility, unwillingness to engage, or lack of imagination seems like an opportunity being missed as records are front page news. These events also illustrate and allow us to explore what it means to do recordkeeping in a networked, connected, truly digital world: one that is fundamentally being invented on the run, with technology capability driving actions which are isolated from social debate, and transforming the world well away from our paper based norms.

Cassie Findlay will talk about some of the work that we've tried to do under the banner of The Recordkeeping Roundtable. That professional forum, in which – with no mandate whatsoever – we try to prompt professional debate, has resulted in the issuing of two papers focussing on Access and on Appraisal, issues which arose from a one day workshop we convened. Today, Cassie will speak on those issues, but I thought it timely to turn attention to one of Hans' long term interests – recordkeeping metadata.

Not only are records front page news, but so too is metadata. Consider these fairly random examples, all coming from the context of collection of bulk metadata from telephone transactions:

'It's actually keeping what's called metadata; so, a record potentially of the time that a phone call is made, the time or fact of an email being sent.' *Nicola Roxon, 2012*¹

'...because the term 'metadata' is not a term of art; it is a term that means different things to different people. I can tell the honourable senator that, during the parliamentary inquiry to which I referred in my answer to his first supplementary question, a number of witnesses gave different definitions of what they understood to be metadata. I myself, on the basis of having been informed by the evidence of those several witnesses during the last parliament, thought that the Prime Minister's description of metadata as essentially billing details was a perfectly accurate, shorthand description of what is a contestable concept.' *George Brandis 2013*²

I start with these two quotes, coming as they do from the mouths of very own Australian Attorney General (both immediately past, and present). They are echoing views expressed world- wide from politicians. Don't worry – it's just metadata – it's not the content of the message, phone call or other communication transaction. But of course, we know that this is being hugely disingenuous – deliberately, or ignorantly? We know the richness of metadata, the stories that it can tell. Recordkeeping is about the metadata. If this metadata was so innocent, why is it being sought so vigilantly? An old, well known case is that of the visualisation of telephone call metadata relating to the German Green MP, Malte Spitz³. So bizarre is this claim of 'its only metadata' that Cory Doctorow, the internet activist has written a fictional mock drama involving Alan Turing, his bureaucratic superior and Winston Churchill metaphorically referencing 'just the metadata' in the context of WW2 espionage:

¹ Nicola Roxon, Attorney General Commonwealth of Australia, quoted in Josh Taylor 'Roxon calls for cold shower on data-retention 'hysteria'

<http://www.zdnet.com/au/roxon-calls-for-cold-shower-on-data-retention-hysteria-7000003742/>, 5 September 2012

² George Brandis: Senator George Brandis Question Time Australian Senate 3 December 2013 <http://scott-ludlam.greensmps.org.au/content/questions-without-notice/questioning-attorney-general-about-surveillance-overreach>

³ Der Zeit <http://www.zeit.de/digital/datenschutz/2011-03/data-protection-malte-spitz> 26 March 2011

‘All we can tell with this analysis is who is speaking, what equipment they use to speak, whom else they speak to, who the messages are addressed to, the subject of discussion, the places where all parties are when conducting discussion, which equipment they use.’⁴

But of course, metadata is a bit of a mysterious word. And using it, allows some degree of obfuscation. In particular it is the ‘contestable concept’ of George Brandis from which I have taken the title of this talk that I would draw attention to. And, of course he is not wrong. Metadata is contestable. It always needs another word to contextualise it. What type of metadata are we talking about? Is it the label on a soup can, or something else. For recordkeeping this is a particularly important thing to reflect upon. It is difficult to divorce discussions about metadata from discussions of recordkeeping more generally. The metadata world is awash with standards and specifications. As it should be, because metadata schema reflect the specific requirements of the community that develops them, and these should be discrete, because they address different requirements driving the need to develop the specific sets. So metadata schema are solidly imbued with the world view or conceptualisations that drive their development. How then, do we discuss recordkeeping metadata schema without discussing the underlying conceptualisations that drive our community. I will return to this point. But back to the ‘front page news’ :

In sum, the program does not involve the NSA examining the phone records of ordinary Americans. Rather, it consolidates these records into a database that the government can query if it has a specific lead, a consolidation of phone records that the companies already retain for business purposes. Obama 2014⁵

I am therefore ordering a transition that will end the Section 215 bulk metadata program as it currently exists and establish a mechanism that preserves the capabilities we need without the government holding this bulk metadata. This will not be simple. The review group recommended that our current approach be replaced by one in which the providers or a third party retain the bulk records, with government accessing information as needed. Obama 2014⁶

These two quotations are from Obama’s speech on 17 January announcing the results of his review of the NSA data surveillance program. There are a number of interesting things in these quotes. Firstly, it is unequivocal that we are talking about records. And metadata as records. Records that are already created for ‘business purposes’. While that claim is somewhat contentious to internet companies who are ceasing to need this granular individual data to support their changing billing models, none the less we are talking about transforming existing records into new records and

⁴ Cory Doctorow, ‘Metadata – a wartime drama’ The Guardian, 5 July 2013

<http://www.theguardian.com/technology/2013/jul/05/metadata-wartime-drama-privacy>

⁵ United States President Barack Obama, ‘Transcript of President Obama’s Jan. 17 speech on NSA reforms’ The Washington Post, 18 January 2014, http://www.washingtonpost.com/politics/full-text-of-president-obamas-jan-17-speech-on-nsa-reforms/2014/01/17/fa33590a-7f8c-11e3-9556-4a4bf7bcd84_story.html

⁶ United States President Barack Obama, ‘Transcript of President Obama’s Jan. 17 speech on NSA reforms’ The Washington Post, 18 January 2014, http://www.washingtonpost.com/politics/full-text-of-president-obamas-jan-17-speech-on-nsa-reforms/2014/01/17/fa33590a-7f8c-11e3-9556-4a4bf7bcd84_story.html

reusing this data. And it is also interesting that a serious social concern was the impetus to review the arrangements and improve oversight, as announced by Obama.

Metadata is of critical social importance to the conduct of business, created in the course of doing business. In its current state of evolution it is clearly being largely driven by technological capacity, and outstripping the social norms and acceptable practices. The technological imperative is driving the collection of data, and under the rubric of big data, also determining the retention of that data. At least one commentator has speculated on the rise of the 'beneficent archivist' :

'Enter the idea of the "Beneficent Archivist," some entity that's entrusted to copy every piece of digital data that we create as a society, public and private, every day and store it, safely, for future generations. At a later date, when society identifies individuals that we want to learn more about, we can go back to the archive and start digging.'⁷

So where does our professional recordkeeping metadata thinking fit?

We entered the discussion about metadata very early. Professionally we had a pretty good start. Archival descriptive standards were agreed as early as the 1990s with the ISAD(G) standard first published in 1993/4. When the Dublin Core was being discussed, the Australian SPIRT Recordkeeping Metadata Project⁸ posited that it was not sufficient for recordkeeping and developed a research set of recordkeeping metadata. We've had research projects in this field since 1994. We've had international guidance in the form of specifications and standards since 2001⁹.

But hiding within that statement is a serious clash of world views. The original Australian recordkeeping metadata work was conceived and created following the records continuum worldview. That is significantly different to the world view expressed in the ISAD(G) type standard. The emphasis of the first is to stress relationships, multiple points of implementation, and dynamic ongoing processes of creation. The ISAD (G) view is applied to 'dead' end products, echoing views from within the custodial archives and the (I would argue), pragmatic perhaps, but fictional world, of stable fonds. It applies when all the processes of records formation are finalised, and epitomises the static records end product approach.

The Australian work was incorporated into the ISO Records Management community resulting in the ISO 23081 Metadata Standard for Records. To date, we have been somewhat reticent about stating such claims loudly. But now, I think it is time to be explicit – there is a different world view being proposed. Let's have a proper professional debate! The recordkeeping metadata standard represents a dynamic and relational way of thinking about records and can be applied in many different layers and sites of implementation. Through circumstance, I suspect, when the ISO Records Management community were establishing recordkeeping metadata standards there were very

⁷ [Matt Baxter-Reynolds](http://www.zdnet.com/does-the-nsa-make-the-case-for-a-beneficent-archivist-7000016918) 'Does the NSA make the case for a "Beneficent Archivist"?' ZNet June 18, 2013 <http://www.zdnet.com/does-the-nsa-make-the-case-for-a-beneficent-archivist-7000016918>

⁸ See <http://www.infotech.monash.edu.au/research/groups/rcrg/projects/spirt/>

⁹ Note ISO 23081 was first published as a Technical specification in 2001.

ISO 23081, Part 1: Information and documentation – Records Management Processes – Metadata for Records Part 1: Principles, 2003

ISO 23081, Part 2 Information and documentation – Records Management Processes – Metadata for Records Part 2: Conceptual and Implementation issues, 2007

limited numbers of individuals and organisations engaged in the debate and thinking. This is a reflection both of the end product focus mentioned above, and the largely paper based thinking which is still driving the profession generally. The recordkeeping standards are based on a world view that is continuum based. Of course, being continuum based, we think that these can be adapted to the end product view if required, but that is a dumbing down of a far more powerful set of capabilities enabled by the metadata standard. So, we have, by circumstance, achieved adoption of a Standard which is a very powerful enabling tool, but it comes with a degree of complexity.

The set of Standards issued as the ISO Metadata for Records standards (ISO 23081) are far perfect, and it is now quite old. In Part 2, conceptual models were presented which I think continue to have validity, and it is that part of the standard that I think the power of the approach is best articulated and presented. The conceptual metadata model is dynamic, complex and relational. It may never be implemented by our own community. But we need something like it to drive the implementation of recordkeeping in any of the 'contested' environments, referred to by our Attorney General. It needs to be a statement of professional approach, relevant to all other communities who need to maintain records of activity. We need a means of communicating the 'meta' metadata approach to recordkeeping.

While I speak of Australian approaches as being the records continuum approach, it is worth noting that uniformity of views cannot be guaranteed in any community, the Australian recordkeeping community included. Within that community I was charged with the job of extending the work of the International Standards and defining a recordkeeping metadata element set. This was a frustrating process, perhaps better if it actually got finished quickly when it started back in 2005. But, in this area particularly, the perfectionism and the detail orientation of the recordkeeping community came to the fore. Various aborted attempts to move beyond the detail obsession left the work unfinished for far too long. And as the years dragged on, my own interest and thoughts on the usefulness or otherwise of this metadata set diminished significantly. But it was being pursued by the archival institutions. It was almost like a tick the box checklist item – got an agreed Australian metadata standard, tick. And now we've got one – AS5478. I dislike the outcome although I am complicit in it. I understand that we need a canonical set for implementation, for interoperability and mapping purposes, and I think this work will potentially have some uses in that context. But, I fear that it is thought of the answer for implementation, when I think its pragmatic uses will be elsewhere. I fear our community have not engaged sufficiently with recordkeeping metadata. Most of the conversations which I have relating to recordkeeping metadata are still looking at implementing in a single implementation environment, within the 'fast paper' world of the EDRMS or content management system, and not engaging with other 'meta' uses of a recordkeeping metadata standard.

As Hans will no doubt discuss, the ISO 23081 standard is up for revision in its routine 10 year review process. There are indications that the requirements of those focussed on end product management are now beginning to use metadata, so the concept is permeating more widely to the implementing audience, and the complexity of the continuum based standard is being seen as a hindrance. I do think it could be usefully reviewed, but I suspect my ideas of a communicating, 'meta' metadata standard might not be widely accepted. It's a different approach, and given that our existing approaches have had minimal impact, maybe it's worth thinking about.

There is no doubt that our recordkeeping metadata standards are not particularly well understood nor well implemented. Here are some suggestions as to why:

- We were 'permitted' to establish recordkeeping metadata standards through ISO because few were engaged with the organisational realities of implementing digital recordkeeping, and perhaps more see (and still see) the utility of metadata relating to end products only. Few were engaged and therefore the community could be lead.
- Because few were engaged, there was little conceptual and implementation follow through.
- The myopia of recordkeeping professionals kept them focussed on jurisdictional priorities rather than experimentation.
- The end product orientation on records objects conceived as something separate from business and what implementation uptake there was concentrated on a 'document' focus rather than transaction focus.
- No one institution really took implementation on. The guidelines were for someone else. In Australia, National Archives were the most logically positioned but were still largely enmeshed in the 'fast paper' environment of most organisations.
- We underplayed the strong conceptual focus underlying the metadata models. The records continuum focus is a different focus than that of the stable end product focus.

So now is the time to focus our thinking on what type of metadata standard do we need professionally, to support what functionality, and how do we raise the discussion in a collaborative and cooperative way?

So what do we want?

Here are a few of my suggestions:

1. Conceptual models

We need to engage with other communities with our requirements for metadata. And the communicating principle is not the detail oriented specification of technically correct metadata elements, although it is perhaps good to be able to produce that when required, but rather it is the conceptual models that we need to communicate. I am firmly of the belief that articulating the requirements of metadata as:

- Point of capture/creation metadata
- Recordkeeping process metadata (defining all the actions that take place on the record object over time).

And then further defining the type of metadata relevant to each by the nature of the entities needed

- Agents
- Business
- Records

We need to stress different layers of aggregation and different sites of implementation. And its not audit trail data, its not 'administrative' metadata. Both are common expressions used to try to define recordkeeping metadata.

2. Engage our community in the discussion

We need to engage the whole spectrum of the recordkeeping community in explicit and quite complex conversations about recordkeeping metadata. Digital systems are going well and strong – there are digital records supporting banking, tax, health and a plethora of other systems: that is to say that business is being done and transactions are being successfully undertaken. But there are bits missing, and it by concentrating on those gaps that we are likely to have most success. We have largely failed to articulate the power of our recordkeeping process metadata approach – but that is actually the metadata being used in the social discussion and exploitation of internet and telephone data. We need to make connections.

While there are terrific things being done in the archival description community, we need to find a way to broaden the discussion so we don't communicate at cross purposes. Digital recordkeeping is different to paper based recordkeeping. We can't apply paper based thinking to the digital world. We can create metadata about people, places and events from creation. We can create metadata records that are themselves as powerful as the content that we are still privileging as archives.

And it's not about digitisation and archival institutions are focused on digitisation as digital records to the extent that their understanding of the fluid technological environment in workplaces is increasingly divorced from reality. Very little usable advice is issued about ensuring the record is maintained in appropriate forms in workplaces to be available for the future. The obsession with end products – record objects – infects innovative approaches that will take us more strategically into the future. In the longer term that approach is doomed to ensure that archives, definitely - and possibly records managers - will remain passive recipients of stuff, chronically underfunded and unloved.

We need organisational champions. The CIDOC-CRM data model¹⁰ for museums is not a million miles away from the types of relational models we are seeking in recordkeeping, although applied in a specific (and different) community. That metadata model was codified in 2006, but didn't really gain traction until it was adopted and championed by the British Museum. It is now being referenced quite widely in European cultural heritage initiatives.

We need active research partnerships focussed on recordkeeping metadata and its implementation. It is not sufficient to have the ideas out there, we need active exploration of them – and that is not happening. There is a little bit of work being done through postgraduate research at Monash University, but it is too little. In the UK, the work of Jenny Bunn¹¹ is making active connections between archival description and recordkeeping metadata drawing on much of the formative early work of the late 1990s and the 2000 Archiving Metadata Forum, for which Hans was an organiser. But many of our best and brightest researchers are puddling away at bespoke digital preservation issues. I'm not trying to denigrate the fabulous efforts of individuals, but are these post-hoc projects really our sole focus of attention, particularly in the world where enthusiastic amateurs and those

¹⁰ The CIDOC Conceptual Reference Model <http://www.cidoc-crm.org/index.html>

¹¹ Jennifer Bunn: 'Developing descriptive standards: a renewed call to action' Archives and Records 2013, Vol 34, No 2

from other communities seem to be doing many things better. I am referring here to the work of the Internet Archive, and Jason Scott with his JSMESS emulation environment¹².

3. We need to understand and appreciate the work of other communities

There are many more, but here are three likely communities which share some of our interests.

Digital Humanities

There is great work being done in the field of digital humanities and with clever data implementers working with archives. This data is open and freely available. Because of this, there are many great projects we can learn from. Rather than proscriptive sets, perhaps we need to understand how the use of APIs, visualisation techniques and data quality and standards affects what can be constructed. Following on from the big data approaches which make and mine relationships, we should be studying and learning (where we can) from the data analytic approaches used in creating the relationships which are horizontal, not simply vertical according to predefined archival practice.

Data management

We share some concerns with the data management community. In particular their interest in 'data lineage' is about data's origins and where it moves over time. This term can also describe what happens to data as it goes through diverse processes. Data lineage can help with efforts to analyze how information is used and to track key bits of information that serve a particular purpose. It is used in business intelligence applications, and increasingly in data warehousing applications used for decision support. Even the big boys, such as IBM, get this requirement and have tools that support such requirements – but it is perhaps seen as a niche requirement.

We need to be working more broadly with the data management profession. They share many of our aspirations and our frustrations. Starting way back with the notion of data dictionaries in the 1980s and initial attempts to define metadata, which like many of our professional attempts ended up too detailed. For organisations focussed on 'getting things done' large and complex projects which involve projects where the payback is down the line, many of these initial data management initiatives also failed. Like our recordkeeping attempts, data management continually reinvents itself – now master data management, now data warehousing, ETL (extract, transform, load) – what we would call migration etc., but fundamentally always going back to their core concepts. Data managers are concerned with our stuff. Expressed in different language, there is real synergy to be had.

Research Data management

Similarly the data management community is interested in managing the metadata around data sets over time, including their source, explanations of their field values etc. We find the emergence of provenance management systems in research data organisations. And we can see some innovation

¹² blog.archive.org/2013/10/25/microcomputer-software-lives-again-this-time-in-your-browser/

happening in the management of datasets in research data management initiatives, now interestingly apparently only about university and library research sets. Within Australia there is a degree of optimism that I can summon as at its core, the ANDS (Australian National Data Set) model is based on the information continuum developed at Monash, attribution buried down in the early development of the models, but still, it is there¹³! Within research establishments the concern about provenance is becoming clearer, although perhaps not always articulated as provenance. These more extensive provenance systems seem to be largely internal systems where they exist, but there are indications that this is becoming a wider concern. The W3C provenance data model, largely defined during 2013, without any archival or recordkeeping input is a fabulous example of our lack of wider engagement¹⁴. We do actually know something about provenance!

What we have to offer, and perhaps to explore with other communities is not the data field, or the data set level of aggregation, but something in the middle. And that's the rub – its not a fixed level in the middle, it's the fields that represent the transaction taking place. But it's a great opportunity to create common ground.

4. Identify opportunities

Within organisations in Australia we are seeing more concerted attention to information management, now being developed less in confrontation to recordkeeping but more embrative of the fundamental requirements of recordkeeping, accountability, traceability, authority of information, reuse, sharing, retention. Our professional views are finding increasing resonance with this area, when creatively expressed. As IT systems are virtualised, and outsourced to the cloud, our IT colleagues are facing the reality that the equipment, and even the software that used to be their major focus is no longer the focus of organisations. The data in systems and the capacity to manage that data is far more important than the systems themselves which are transitory. In the words of Tim Burners Lee *'Data is a precious thing and will last longer than the systems themselves'*.¹⁵

We have data that is rich in relationships. Yet, we've done very little intellectually as a building block to identify and exploit our knowledge in this area. Within Australia we have acknowledged that more work is needed in the definition of relationships for many years. We know that this is incredibly powerful. Just think about the use of relationships on the web. It's the core of linked data, it's the core of semantic web. We need to be expressing recordkeeping metadata in RDF and linking it to semantic web implementations. Potentially we have professionally let another opportunity slip because of the lack of people to play seriously in the game. Philosophically we can connect with strong currents of thinking making powerful links to the work of Bruno Latour¹⁶ and Gabriel Tarde in their exploration of monads, a powerful concept for recordkeeping thinking.

Recently I have encountered some commercial vendors exploring the notion of stand alone metadata brokers to act as an independent services to move information (records) between

¹³ Treloar, A., Groenewegen, D. and Harboe-Ree, C. "The Data Curation Continuum: Managing Data Objects in Institutional Repositories", *D-Lib Magazine*, Vol. 13 No. 9/10, 2007, doi:10.1045/september2007-treloar.

¹⁴ W3C Provenance Data Model, April 2013, www.w3.org/TR/prov-dm/

¹⁵ Widely attributed to Sir Tim Berners Lee, there appears no authoritative statement to be cited.

¹⁶ Bruno Latour, "The Whole is Always Smaller than Its Parts': A Digital Test of Gabriel Tarde's Monads' *British Journal of Sociology*, Vol 63, No 4, pp591-615 <http://www.bruno-latour.fr/node/330>

systems. This technique is being used also in the research data and information sharing environments. This is good news.

Conclusion

Metadata is in the public mind, but not really a focus in the recordkeeping community discussion. We need to reinvent our approaches to digital recordkeeping in all environments, not simply being content or focussed on what happens inside the custodial walls of our archival institutions. Metadata may be 'a contestable concept', in the words of our Attorney General, but people have demonstrated that they care. The growth of data activism and the increasing attention to managing personal data on the web are all indicators that there are plenty of seeds from which to sell our recordkeeping message.

To do that effectively in the complex and ever changing technology environment we need to focus our message. Not on the old stuff, but in the creation and responsible management of authoritative data for individuals, organisations and societies. We need to revisit the professional thinking of Hans Hofman and others and reengage with the debate about recordkeeping as an approach. Should we enable a different way of thinking about recordkeeping, or are we professionally satisfied with the static end product focus? I would clearly have a view to a dynamic, flexible and forward looking view. If we need to recraft the message, then lets do so. Regardless of whether you call it recordkeeping informatics (which I do), information management or something else, professionally we have a great deal to contribute to the current issues of concern across multiple disciplines, and in the service of sustainable authoritative sources of information. Just lets get outside the custodial walls, away from our obsession with static end products, engage with others who are actually doing what we purport to do, and learn!