

I've recently been reading Richard Ovenden's *Burning the Books: A History of Knowledge Under attack* where he talks about the many threats facing the collections of archives and libraries. The book is largely about how, over human history, individuals, groups and states have either tried to control or to destroy information. It also talks about threats to information in the digital age, for example how governments are reducing funding to knowledge institutions partly in the misguided belief that mega-corporations like Google can do the work of making knowledge accessible more cheaply and efficiently than a government institution ever could. But there are other potential threats, or risks, that defy even our best intentions.

For government archives like ours, collections are built on transfers of records from the agencies that created and managed them. Of course, archives don't actually "collect" at all – government agencies accumulate records through the natural process of their business activity and those records come to us, preferably but not always, as planned transfers of records that have been appraised as having ongoing archival value. And how those agencies manage their records has profound downstream effects on archives.

**[CHANGE SLIDE]** It's the relationship between the National Archives and records' creators that I want to talk about today, as I think it relates directly to the issue of EDRMS preservation and its importance, and the threats and risks inherent in government recordkeeping and the approaches we've adopted over the years to mitigate these risks. I'll take you back in time a bit because the story of the National Archives' developing approach to government recordkeeping has quite a few lessons.

The key period when the National Archives began to develop a new approach to government recordkeeping and issue a slew of new standards and guidelines was the latter half of the 1990s and early 2000s.

The drivers for this work was twofold – one was the poor state of government recordkeeping at the time, and the other was a classic digital preservation issue. It might worth talking about the latter for a moment. **[CHANGE SLIDE]**

The digital Series that was in high demand from the 1970s through to the 1990s was oil exploration data – these were recordings of undersea explosions that were kept on magnetic tape in an offline storage environment. When someone wanted to access a particular recording we would retrieve the tape and connect it to computer infrastructure in our Sydney reading room. Of course, by the 1990s, the earlier tapes were inaccessible on current computers. **[CHANGE SLIDE]** The recognition that we didn't have the means to ensure ongoing access to digital information, plus the urgent need to improve government recordkeeping, prompted us to introduce a distributed custody model for agencies – so between 1996 and 2000 we told agencies we weren't accepting transfers of digital records – agencies would manage their information under a management regime developed by the National Archives. This distributed custody approach was a controversial decision and sparked several years of intense debate between 'post custodialists', led by the American David Bearman and the Australian academics Frank Upward and Sue McKemmish, and their opponents, whose main voices were the Canadians Terry Eastwood and Luciana Duranti,. By 2000 the Archives was accepting transfers of digital record again, and had commenced a digital preservation project that resulted in some influential work including the development of the Performance Model and an XML-based normalisation strategy.

**[CHANGE SLIDE]** This brings us back to government recordkeeping. By the mid-90s when we introduced distributed custody for digital records there was ample evidence that recordkeeping in

Australian government agencies was, as Adrian Cunningham expressed it in a 2004 paper, “going to hell in a handbasket” and that some radical intervention was necessary. The new approach to recordkeeping was reflected in the theoretical revolution of the records continuum in the 1990s, developed by Frank Upward and Sue McKemish at Monash University, and the introduction in 1996 of the Australian Records Management Standard (AS 4390), which in 2001 became the international standard 15489.

The standard introduced the so-called ‘functional approach to recordkeeping’. This means that records are linked to the organisation’s business functions, which opened doors to what was then the new worlds of functional analysis, functional classification and the identification of recordkeeping requirements and functional requirements. From that point on our concept of appraisal was to be very different to what it had been previously. The Archives became committed to appraising functions and activities, not records, in order to identify the need for records. Not only did the Archives’ appraisal attention shift from records to functions, but our view of what constituted ‘appraisal’ widened considerably. Appraisal was no longer simply a process of deciding how long to keep records – it had become the process of deciding what records need to be made and kept, and in what form and for what purpose. Why decide how long to keep records when those records have not even been created and you have no idea what records you should be creating?

In 2000 the Archives released a set of products, collectively called e-Permanence, to assist agencies in implementing the brave new world of functional appraisal. The key products were an exhaustive manual for Designing and Implementing Recordkeeping Systems, known as the DIRKS Manual, and the first version of the Commonwealth government recordkeeping metadata standard. A lot of other advisory products were released but these were the main ones influencing EDRMS development. And in fact while this work was going on, a Canberra-based recordkeeping software company, Tower, had become involved in the development of the records management standard, and it released the first version of TRIM, TRIM Captura, in 1998.

**[CHANGE SLIDE]** Our main aim in developing the DIRKS Manual was to exert a positive influence on the design and implementation of recordkeeping systems. The Archives believed (and still does) that if recordkeeping were to improve, recordkeeping systems would have to be improved from the ground up. If you have lousy systems, all your other recordkeeping efforts will probably fail. So the DIRKS Manual built on the records management standard to provide comprehensive practical guidance on developing and implementing a recordkeeping system via an eight step methodology. A critical success factor in all of this was the number of new functions based disposal authorities issued under the new methodology, and while a steady stream of them was being issued, an influential 2007 report into government recordkeeping found the DIRKS process complex and resource intensive for agencies, and in response the Archives introduced a streamlined 1 step process, designed to halve the time taken to develop a Record Disposal Authority and resulting in a shorter, streamlined RDA.

**[CHANGE SLIDE]** Since then 100s of RDAs have been issued to agencies and according to the last Annual Report about 80% of government entities have records authority coverage, so the strategy of adopting an easier, streamlined approach to developing Record Authorities seems to have worked. However, it is interesting to note, at this point in time, when we are looking to AI and machine learning technologies as the panacea for information management pain points, in particular to automate appraisal and disposal decisions – those technologies rely on computers being fed as much information as possible, while the trend has been to make records authorities shorter and less detailed.

**[CHANGE SLIDE]** The other critical standard was the Australian government recordkeeping metadata standard. Version one of the standard was released in 1999, and version 2 in 2008 as a completely revised multi-entity model. Version 2.2 was released in 2015.

The importance of metadata in the management of digital records is undeniable: a record cannot be considered a complete record unless it is permanently accompanied by recordkeeping metadata. Recordkeeping systems are metadata collecting systems.

While the pivotal role of recordkeeping metadata in managing digital records is recognised by the recordkeeping profession, its implementation in the real world is less clear. EDRMSs are sophisticated software systems, but implementations of metadata standards are quite patchy. They can be certainly be configured to ensure, at the very least, the creation and management of minimum recordkeeping metadata, such as that outlined in the National Archives' Minimum Metadata Set, described on our website as a 'practical application' of version 2.2. of the recordkeeping metadata standard. And it certainly looks like this development of minimum metadata standards has become a norm internationally. And yet, to fully harness the benefits of recordkeeping metadata, it is simply not enough to meet those minimums without which the basic reconstruction of a record is impossible.

And of course I'm talking here about EDRMS implementations - software as a service systems like M365 and Sharepoint are not really metadata collecting systems at all, although there are efforts happening in different regions to talk Microsoft into improving the metadata functionality of these products. It will be interesting to see how these evergreen products develop.

**[CHANGE SLIDE]** It's worth mentioning in the context of EDRMS standards another project that came out of the Archives' response to government recordkeeping. The generic specifications for electronic records management systems and for business information systems, and supporting guidelines, were published in 2006, and fed into a joint project by the International Council of Archives and the Australasian Digital Recordkeeping Initiative, which published a much simplified *Principles and Functional Requirements for Records in Electronic Office Environments*, published in 2008, and later endorsed as ISO 16175.

In trawling our EDRMS, Content Manager, to develop the content for this paper (and lamenting the state of record titling), I found one or two briefing documents or reports about these generic specifications that under the heading Exclusions listed "Preservation". And certainly while digital preservation gets a guernsey in the ISO standard it's really just to say that digital preservation should be addressed "with a dedicated framework at the strategic level." So there is a gap there, and to some degree this is what the DPC's EDRMS Preservation Taskforce is trying to fill.

One other point about the standards and guidelines produced by organisations like ours: when they are developed and published, they are necessarily aspirational in the sense that they reflect what we believe is best practice – in other words where we think government recordkeeping needs to go - rather than reflecting actual practice at the time.

If you like, these products informed a view of recordkeeping that reflected the world of academia: of textbooks and research and development laboratories. No one had ever done recordkeeping quite in that way before, so the suite of standards, tools and guidelines reflected how the Archives thought the new approach to recordkeeping could be implemented. I guess the jury is still out on whether that thinking has proved effective, certainly it was developed when similar ideas were emerging internationally and it has shaped a lot of thinking around the world.

**[CHANGE SLIDE]** Standards and guidelines are one thing, but getting agencies to adopt them is another. Since 2011 the Archives has issued a series of rolling government wide policies, 5-year-plans if you like, to push agencies along in digital transition and which include actions, targets and pathways, online self-assessment kits, annual surveys and so on. The first of these, the Digital Transition Policy was developed by the Department of Prime Minister and Cabinet with the National Archives as the lead agency, and released in 2011. The Digital Continuity 2020 Policy was released in 2015, and the latest policy *Building Trust in the Public Record* came into effect this year. The current policy has something of an emphasis on digital preservation, and in fact, to take the opportunity to spruik the Digital Preservation Coalition, the DPC did provide some excellent comments on a draft of the policy. There is a release schedule of products created by the Archives for government agencies, which include things like eLearning modules on topics like digital preservation, digital archiving, data formats and metadata.

So where are we now? I suspect it really depends who you ask. Certainly the results of our agency surveys show the whole of government policies are working. But as someone with a background in digital preservation and digital archiving what I consider a success factor are the number of Series in the digital archive from ERDMSs. And we don't have many – not many have crossed the 'archival threshold' as Luciana Duranti would say. A few from closed agencies, and those not particularly exciting ones like the Australian Wool Research and Promotion Organisation, a couple exports of metadata from EDRMSs, in effect Master control records for paper Series, and one or two other examples, either from TRIM or from Objective, which is the other widely used EDRMS in Australian government.

**[CHANGE SLIDE]** By far the largest EDRMS transfer was our own transfer from our corporate EDRMS, TRIM Context, in 2012, so quite some years ago. The statistics from that project are interesting. The project team comprised 4 people including three sentencing officers, and indeed the focus was on sentencing – preservation wasn't really a consideration because the records would end up in the digital archive, so why do anything up front? The project took about 8 months to complete, and sentencing, which was effectively a manual process took about 5 months. About 34,000 TRIM files or containers were sentenced comprising about a million records, though records were sentenced at TRIM file level, unless there was a good risk-based reason to go into the file and look at actual records. The proportion of records to retain as national archives was about 10%, quite a high proportion, and 100,000 records were ingested into our digital archive.

Now, should the Archives be worried we haven't got more transfers from EDRMSs in custody? I mean, in effect what we have now is distributed custody, though not distributed custody by design, but by default. Is that a problem? Well, I think we should be concerned. After all, an EDRMS is defined by its compliance with recordkeeping standards, so the functionality to appraise, select and dispose of records should be intrinsic to an EDRMS – transfer to an archive shouldn't be an obstacle, and yet it clearly is. Certainly our experience 8 or 9 years ago was that sentencing was largely a manual process, although some features of TRIM can automate some parts of the process. There are other factors that are a barrier to sentencing and transfer. We know that information management sections still have resource issues. In the Australian government there are 14 Departments and State and hundreds of agencies. There is a huge diversity of practice and recordkeeping maturity in agencies. Churn is high and staffing levels do not necessarily reflect the complexity of work. On top of this are some of the issues I've mentioned today – minimal EDRMS implementations, sentencing and disposal is still not a trivial process and so on. These are all potential threats to the record.

On the other hand, maybe the situation isn't all that bad. EDRMSs haven't been around for that long. As I mentioned before, Tower Software released TRIM Captura in 1998. The National Archives was an early starter and introduced Captura in 2000. Broader agency uptake was generally much later than that. Under the Archives Act agencies must transfer their records 15 years after creation or when their business life has finished, so we would expect agencies to have transferred who had implemented an EDRMS up to 2006, so there is some wriggle room. Certainly what we need to do is get more proactive out there in Agency Land, and this is our intention in the coming few years.

**[CHANGE SLIDE]** Finally, I want to get amongst the weeds of EDRMSs preservation and transfer. Standards and generic requirements are one thing, but they are principles based, much in the same way that the law has moved away from black letter law to principles based legislation. It reduces complexity and the need for constant revision, but it leaves a lot to interpretation. It's really only when you get amongst the weeds that you see practical preservation problems emerging that you didn't think of at the time. A couple of examples from our transfer of corporate records from TRIM into the digital archive:

- We didn't do anything about emails with attachments linking to a record in TRIM – these links are lost when records are exported from TRIM, and automatically generated metadata about the linked record – the record number and title for example - might not have been retained in the body of the email – therefore the record is incomplete. Even if the record number was retained, this doesn't mean that was the version of the record actually emailed – it could have been edited in the meantime - so context is lost.
- That raises another point. We transferred finalised records, not TRIM versions. Unless versions were captured as separate records, versions wouldn't be captured.
- A third issue - when we exported records for transfer, we kept emails in TRIM Outlook format, that is emails with extensions .vmbx or .mbx. These formats are basically plain text files, so not a format preservation risk, but attachments to these formats are base64 encoded in the body of the email, so this will become an access issue down the track. Later versions of TRIM, or what is now called Micro Focus Content Manager, has software that can migrate these formats to EML and MSG – something to keep in mind for future transfers
- We also didn't do a detailed analysis of formats in TRIM. An EDRMS, as well as document management systems like Sharepoint, can load pretty much any format, and as we've found that are lots of interesting formats in TRIM that we could have identified up front as needing, for example, better documentation, such as dozens of legacy Access databases, early AutoCAD files and so on.
- Finally, archival control of records in EDRMSs isn't trivial – good archival control depends on a number of factors. We reuse recordkeeping metadata for archival control, so the level of compliance with the recordkeeping metadata standard is critical. A related issue is the sophistication of your archival control system to properly manage record aggregations and representations – in other words, what do you do if the data model of your archival control system can't deal with a complex web of relationships? The solutions – replacing the archival control system, or introducing a new data model - are big, long-term projects.

Anyway, that's all I've got to say today. The broad point I've tried to make is that digital preservation runs across the record continuum – it's not just about keeping an object accessible; it reaches into questions about appraisal, of selecting the best record, the complete record, of ensuring metadata capture is adequate, and it reaches into the configuration and functionality of recordkeeping systems, and what is exported for transfer to an archive. The work of the DPC's

EDRMS Taskforce should be welcomed – anything that raises awareness of these issues is a good thing.