Strength below and grace above: the structuration of records

Chris Hurley

> Thus all below is strength, and all above is grace.
> John Dryden (1631-1700) *Epistle to William Congreve*

> The voyage of discovery is not in seeking new landscapes
> but in having new eyes.
> Marcel Proust (1871-1922)

I am greatly honoured to be your keynote speaker this afternoon. "Keynote" is a musical term defined as "the first and harmonically fundamental tone of a scale", hence it has come to mean "the fundamental or central fact, idea, or mood". A keynote address is "designed to present the issues of primary interest to an assembly ... and often to arouse unity and enthusiasm". Whether the note I shall sound will be key to your deliberations I cannot say, but I can say this : it is not my role to give you answers to the issues that are posed in your challenging programme, even if that were within my capabilities. In the next two days you will explore questions that our profession has scarcely begun to answer effectively. There is no consensus. There is no beaten path. Those issues have been troubling us since computers and the Internet began to change everything. Because I think our profession has not made the progress it should have, you are here to find answers, if you can, not to hear them from me or anyone else.

Before I leave this terminological introduction, some of you may be wondering about the word "structuration" in my title. It is an unusual word even in English and I'm not sure how it will translate. I can assure you that it is listed in the on-line Webster with a pedigree going back at least as far as 1925. It means "the interrelation of parts in an organised whole". Let me say then that the central idea or mood of this key note address is that structuration lies at the heart of what we do and that it must somehow remain part of how we respond to new challenges. A quotation from John Dryden has given me my title. He was admiring the graceful literary accomplishments of younger writers like Congreve which he believed were built upon the strong bedrock inheritance they received from an earlier generation of artists which they had embellished with their own graceful adornments.

The strength of the archival narrative is the coherence we maintain in related events and circumstances as told in documents. Here is an example, an old one, that some of you may know. It concerns the contents of an email – an actual email taken from a real email system[1]. It is a message comprising two words only : "Well done!" Standing alone like that they mean very little. When you put them in context, however, when you can see them as part of a transactional chain of events which is being documented, they become much more meaningful (**Figure One**). The message comes from the White House email system. They'd tried unsuccessfully to delete it to escape detection. The email dates from the Congressional Investigation into the Iran-Contra Affair during the Reagan Administration. Poindexter was the National Security Adviser. Pierson was a subordinate telling him how Oliver North had successfully misled the Investigation by lying. "Well done!" was Poindexter's response (in an email to North). In that context, the meaning is much clearer[2].

Now let me introduce you to the little truck (**Figure Two**). It represents a traditional view of archival work. Records are generated in many places and, when no longer required by their creators, they are transported on the back of the truck to an archives where they are preserved for use by someone else. The little truck might be a metaphor for de-

contextualisation. When records are still in the environment of their creation, there is much contextual knowledge, about the circumstances of their creation and use, that is known but not written down. It exists only in the minds of the creators of the records - the living finding aids. When the records leave their native environment that understanding is lost. The records become, like the email message, content without meaning.
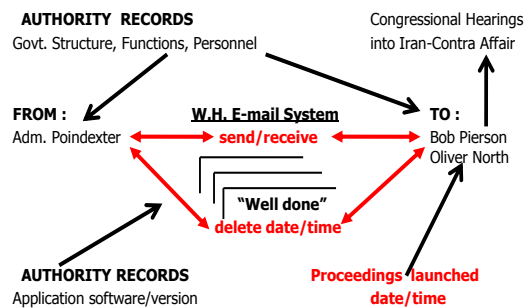
Information in context...

AUTHORITY RECORDS
Govt. Structure, Functions, Personnel

Congressional Hearings
into Iran-Contra Affair

FROM :
Adm. Poindexter

W.H. E-mail System
send/receive

TO :
Bob Pierson
Oliver North

"Well done"
delete date/time

AUTHORITY RECORDS
Application software/version

Proceedings launched
date/time

4

Figure One

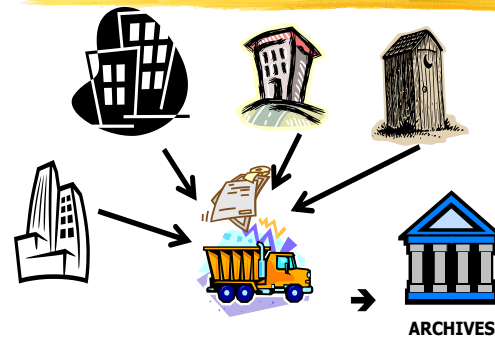## The little truck

ARCHIVES

5

Figure Two

When archivists intervene to re-contextualise records passing across the archival boundary[3] this does not happen. Records from different sources are not mixed up and we write down knowledge about their origin, their purpose, who was involved in their creation, and what they were used for[4]. With an archivist at the wheel, the little truck does not then rip records out of context. It transports them into to a larger world – a virtual world and not (as we once believed) a physical place. There, although they are no longer in a place of their own but in space shared with records from other places, archival methods keep them separate, organised, and meaningful. The challenge we face is to free them from the prison of physical place, the boundary of a single repository, and inter-weave them, descriptively, within the narrative to which they belong.
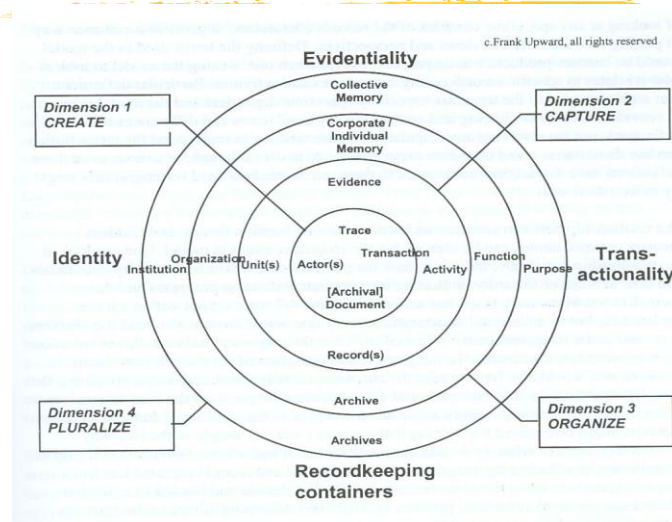
In 1981, when I joined the National Archives of Australia, then called the Commonwealth Archives Office, my mentor was Peter Scott. The Archives was still imbued with the spirit of Ian Maclean. Under those influences, the archival enterprise was a part of recordkeeping, connected with records management in what subsequently came to be termed the continuum of recordkeeping[5] (**Figure Three**). Archives and records management were like the string and wind sections of an orchestra. They could not function independently of each other without impairing the whole performance.

This was the view expressed at the 2009 Conference of the Australian Society of Archivists in Brisbane, by the ICA's Secretary General, Mr David Leitch. He said, inter alia, that the merger of archives into conglomerate heritage collections along with libraries and museums "must" be resisted because archives had a unique requirement, through records management, to keep a connection with creators[6]. The history of Australasia's government archives programmes can be written as a saga of separation from the public libraries. This was done partly on Mr Leitch's argument that a connection has to be made with creators and partly that the management of hard-copy archival materials was quite different to the management of hard-copy library materials (except for manuscripts)[7]. Governments are interested in records management today not because it sustains a link with archives but

because it supports information resources which are being re-purposed for public use on-line regardless of inherent archival value in the traditional sense. Information accessed on official websites is contextualised in a way, but it is not being done well and it is not being done by us.

At the same 2009 Brisbane Conference, there was a session on FOI developments in Queensland pointing to new directions in the management of on-line resources. Queensland, it was claimed, is moving from a "pull" to a "push" approach in FOI. They mean that the Queensland Archives is becoming more involved in what are essentially records management tasks in order to improve public sector recordkeeping so that resources held in ministries and agencies may be better available on the Internet[8]. Archival access has always been push, not pull. Records are released in advance of request and you don't have to identify who to ask because archival description gives a single access point (the finding aids) for the whole of government. One of the reasons FOI access was differentiated from archival access when FOI was first introduced was the perceived inability of the former to provide a whole of government gateway or search engine that did not first require you to identify who to ask[9].

Recordkeeping continuum

Figure Three

Earlier this year, Australia's Information Commissioner released an Issues Paper, laying down a blueprint for developing the Government's on-line resources. Like other papers before it, it urges seamless access to on-line services in place of fragmentation caused by maintaining separate web-sites for each ministry and agency. He wants to see on-line "re-use" of data held by government for different purposes than those for which it was collected or generated. He lists eleven projects already in place where such re-use is occurring. One of them is the National Archives' on-line database for records of military personnel[10]. This is a huge digitisation project undertaken by National Archives and heavily used by genealogists and others. The other projects are run directly from the ministries and agencies themselves. They include -

- an on-line interactive data base maintained by the Bureau of Statistics,
- spatial and environmental data from several agencies (including a co-operative venture with the New Zealand government),
- geo-scientific data gathered to support land use and management, and
- an on-line resource showing the location and details of more than 14,000 public toilets around Australia gathered to support health and ageing professionals and now made available for use by everyone.

I am a senior citizen now and that last site doesn't seem so funny as it might have once.

Because government services are still packaged up into parcels available on web sites for every ministry and agency, just like on-line access to archival resources are packaged on archives websites, you cannot easily get through a single gateway to search the content of many web sites at the same time or to obtain a government service[11]. Archives possess publicly available information which is now being progressively made available on the web through on-line finding aids and digitisation, but like government services they are not generally available via a dedicated portal or gateway. Instead of becoming models of data re-use archivists are being swamped by sloppy, low-grade access tools that do not meet what are (or ought to be) archival requirements for structure and context[12].

The public is getting on-line access to government-held data gathered for one purpose and used for another purpose by people other than its creators. This is a workable definition of archival access. But, in ten out of eleven listed examples, the Archives is not involved. Can anyone doubt that this kind of non-archival access is going to grow? Under the old paradigm, the data would first be transferred by the little truck into the Archives, processed, and then made available by the Archives, using Archives' systems. In the UK, a division of the National Archives is now dedicated to the distributed access of public sector information[13]. The UK website, data.gov.uk, now offers a single gateway for accessing re-usable public sector data – data that is being accessed long before any question of removing it to archival control even arises. Can anyone doubt that the old paradigm, letting the dust settle for 20 or 30 years until archivists have worked on it and then release the stuff to the public, is finished? In these examples, the Archives is out of the loop on 90% of the projects cited. But it is worse than that because archives are doing a poor job on the sites they do manage in the work of standards, metadata, and contextualisation, our key structuration tools, putting out flat content – just like everyone else - that users must largely contextualise for themselves[14].

The idea that we are involved in both archives and records management has led to some confusion of mind. In the world of Information & Communications Technology (ICT)[15], recordkeeping plays a fairly minor role. But at the 2008 ICA Congress in Kuala Lumpur, when progress was reported on ICA's Recordkeeping Requirements Project, the standard being drafted listed 15 or so principles, most of them data management and information management principles that we simply don't control professionally[16]. We could say a lot more about how to manage evidence as our contribution to these principles : articulating in detailed ways what is meant in two of those principles dealing with "evidence of business activity" and "good recordkeeping". But we cannot teach those involved in data and information management about most of the ICA's other 13 or so Guiding Principles[17], even if (as in the UK) the National Archives has a primary co-ordination role in data re-use.

At the 2009 Brisbane Conference to which I have already referred, another paper dealt with the Australian Recordkeeping Functional Requirements Project (the one subsequently taken up by ICA). It was stated (almost as an aside) that access had been left out of scope because it was regarded as "generic"[18]. They were saying that access doesn't belong to "us" (the world of recordkeeping) but rather to "them" (the world of generic ICT). We just use the web-based access tools that technology provides us and that is all there is to it. This is wrong. Access to archives is not a generic process belonging to ICT. It is part of the archival enterprise.[19]

We overstate our role in Information & Communications Technology when we equate good recordkeeping with the entire body of information and data management principles. Our role is narrower than that. But we under-estimate our role when we exclude access from it. There are, to be sure, "generic" access tools of which we are the users, not the administrators. But to simply cast something as crucial as access aside to be handled for us by these "generic" tools is going too far. The technology is not ours, we are users of that. But how we use the technology and how we structure the data so that the access tools are used to greatest advantage is very much our business.

In Canada, the once separate National Archives is now subsumed into the Library & Archives of Canada (LAC) – in part to more closely align Canada's heritage resources and make them more accessible. About 10 years ago, Ian Wilson, then Canadian Archivist and subsequently first head of LAC, gave an address in Wellington, New Zealand, to a Conference of the Records Management Association. He said, if I recall correctly, that archives programmes must move into digitisation or perish. I have no doubt he was correct and that, to remain relevant, archives must be able to enter cyberspace and convince funding authorities of their relevance and value in that arena, like any other heritage resource. Similar amalgamation has recently occurred in the Australian State of Tasmania, in the Northern Territory also, and between the Archives and National Library in New Zealand[20]. The records management role remains intact in Tasmania and New Zealand - within the merged entity. The Northern Territory, on the other hand, has divided the two roles into completely different and organisationally unrelated areas of responsibility on the argument that it was necessary to align the archives "with other collection and heritage focused agencies"[21].

Last year, the National Archives of Australia proposed shutting down three of its regional offices – in Darwin, Adelaide and Hobart – as a cost-cutting measure. The closures were modified after an uproar ensued. Arguing in favour of the closures it was asserted that on-line access and digitisation mean that on-site access was no longer so important[22]. In Session One of your Conference the question will be raised whether digitisation of collections "requires that the institutions and services which have custody of the collections connect in temporary or permanent networks seeking for their sustainability"[23]. If the tools and methods used to manage and deploy digitised archival content are the same as those used to digitise and display other heritage resources, our archives repositories risk becoming just another quarry for digitised content, often indistinguishable, depending on how it has been googled, from other information resources available on the net.

Earlier this year an article was published in the Literary Review of Canada entitled "National Archives Blues : Is a Precious Canadian Asset being digitised to Death?". It lamented the

decline of on-site services and yearned for the good old days of "nondescript rooms with big tables, straight-backed chairs and large windows, and very little going on other than an attendant or two dispensing information at the reference desk and a few people sifting through papers". This could be viewed as a luddite protest against change, but I think it highlighted the contrast between resource discovery and what I call the "fossicking" or "rummaging" component in archives research. Rummaging is the final stage of archival discovery and finding aids were built accordingly – getting the searcher to the place where fruitful rummaging can occur[24]. Rummaging is about finding stuff you don't know you're looking for until you find it. Archival description gets you to the best place for rummaging, makes connections between the stuff you find there and other stuff (stepping stones from where you've found gold to other places worth looking in), and it provides contextual explanations so you understand what it is when you find it. Providing access to people who can't easily visit and fossick has always depended on the quality of archival description. Family historians have always been amongst the most sophisticated users of original documents and the best rummagers. One of their associations advertises on Australian TV with the slogan : "You don't have to know what you're looking for." I just love that as the epitome of archival discovery. Our on-line search tools should be designed to support research into the unknown and provide contextual understanding, not just assist in the identification of particular instance records amongst commonly sought document types[25].

Digitisation makes archives available to great crowds of people on the Internet who want content, who are gratified by the common discovery tools available there for all kinds of content, not just archives, and who wouldn't dream of using a recordkeeping system to find it. For them, archives are just another resource to be ransacked for information they find interesting or useful. Many of them do not have the same search needs and expectations as our traditional users but they will increasingly make up the great majority of users of digitised content, and their requirements must be met. They expect discovery tools of the kind they are familiar with when using other kinds of stuff and they expect discovery of archives to be seamlessly integrated with that other stuff. They are not prepared to search and fossick, they expect an instant response to their search query displayed for them on screen in a format they are familiar with, they will not tolerate a multi-step process, they don't want to see anything that is more than 10 minutes out-of-date, and they probably won't go below page 3 of any results report. We have to jostle with other information providers for their attention. Otherwise we are out of the game. How do we do that and also go on meeting the legitimate needs of research?[26]

In the early 1990s, I joined the ICA's Ad Hoc Commission for developing descriptive standards. While I was there, we completed the first edition of *ISAD(G)* and a first draft of *ISAAR(CPF)*. It was not an altogether happy experience for me (and not much fun for them either, I suppose, having me there) – but that is another story. The rationale behind ISAD included developing a rule-base to guide the development of archival software (now culminating in the AtoM software being sponsored by ICA) and providing the basis for something very like the archival equivalent of what librarians used to call a union catalogue. The standard would enable resource sharing by facilitating on-line searching to bring together hits from multiple archival repositories.

In preparation for this paper, I went to Google and carried out three searches by personal name. The names I looked for were -

- Louis Riel (a leader of the Metis revolt in Canada)
- Simon Bolivar (a notable figure in South American history)
- Lachlan Macquarie (a Governor of NSW, often called the 'Father of Australia')

At the top of page 1 in each case (or very close to it) was a result for Wikipedia and a result for images of the person sought. In the first 3 pages in each case, there were numerous results for articles, encyclopaedia entries, curricula, teaching materials, and fact sheets dealing with the person. In some cases these were clearly based on archival sources but there were few links and those links that were given were often broken. Some of the fact sheets were issued by archival repositories and these often but (astonishingly) not always had links into the on-line search engine for that archives. There were no links on pages one to three to actual electronic or digitised records. An information sheet about Riel appeared from *The Northwest Resistance : A database of materials held by the University of Saskatchewan Libraries and the University Archives.* From the Home Page it is then possible to explore other resources available on that site including an on-line guide to the Special Collections and to the University's Archives and to a searchable database of bibliographic records about the Rebellion. Most of the material is part of collections of books, pamphlets, and manuscripts. My guess is that this is an important resource on Riel but represents only a fraction of what is available in Canada.

I found links to no archival repository materials dealing with Bolivar. In Australia, the National Dictionary of Biography is on-line and usually appears close to the top in any search there by personal name,. These entries are based on archival research and list published sources but have no links to archival resources. On the first three pages of my search results there were fact sheets from a local historical society (with links to online resources, including a cemetery register) and from the NSW State Records Office with a link to its own on-line search engine. Macquarie University (named after Lachlan) has a page linking to the University's own collection of Macquarie Papers and there are history pages put up by the NSW Parliament and the Australian Antarctic Division.

On-line access to archival resources is not coming from the first 3 pages of a Google search or from similar search engines but from websites dedicated to searching archival holdings[27]. These are usually sites put up by a single archives repository. Why not have sites that enable or facilitate on-line access to more than one repository – the original aim of ISAD? There are such sites. In Australia, the hard-copy, loose-leaf Guide to Australian Manuscripts (a union list of manuscripts held in non-government repositories) morphed into RAAM (Register of Archives and Manuscripts) which was the on-line equivalent. This was not about searching the on-line resources of the institutions but about finding special descriptions of holdings submitted for entry into RAAM. It is now superseded by TROVE[28], an on-line resource maintained by the National Library and incorporating entries from several repositories as well as guidance to non-archival sources. But it excludes a lot, and provides no more than an on on-line link to government archives sites, major non-government archives sites, and the sites of other major libraries - to say nothing of registries of births, deaths, and marriages, land registries, and the Patents Office.

The nearest equivalent to a library union catalogue is a register of archives. The most venerable that I know is the British Register[29]. Austria has one also[30] and there may be others. Perhaps the closest thing to what the developers of ISAD had in mind is the Archives Canada site[31]. This is appropriate because the dream of networked access to

archival resources expressed in ISAD owes much to Canadian influence[32]. Within such collective spaces, based on alliances between archival repositories, a kind of archival social network, we should be able to develop search engines that focus on the discovery of archival resources. An alternative model would see such initiatives based on alliances with other cultural heritage programmes (such as libraries and museums). The two models are not mutually exclusive but the question is whether the development of web-based discovery paths focussed on the needs of archival access is compatible with alliances with non-archival programmes.

What is clear is that twenty years after ISAD was conceived, you still don't get good direct access to archival resources using Internet-wide search tools (and I doubt you ever will). They will always be crowded out by other stuff. Most access comes by finding your way to a site maintained by an archives repository and using their search engine to look inside the holdings of that repository, not across the holdings of several repositories.

I seemed to get more relevant Google hits on the Australian topic than for Canada or South America[33]. There is a technical reason for this. It is because I was googling from Australia. If you google the same query from two different locations, you get different results. I am told[34] that the algorithms that Internet search engines use to locate and organise results are highly classified commercial secrets. Up to 200 "signals" I understand can be used to produce results, including user location and web history. These signals can come from both the user profile and from the target resource. Unscrupulous operators spam listservs, apparently, to lift the profile of their web pages and fool the search engines into giving them a higher place in the order of results. I am not suggesting that we resort to such methods to get a higher profile for archives web pages, but such understanding helps us to appreciate the functionality of on-line searching so we can use it better and it may be worth looking at it in Session Four.

Search results do not depend only on the search terms chosen and the quality of the description of the target resource. They vary also with the profile of the user doing the research. This is a fundamental shift in the balance of power between the user and the provider of information. The Internet moves the power to shape information away from the provider and into the hands of the user. As the Egyptian President found this year, the Internet weakens the power of those who control information to shape the narrative. We can no longer construct pathways along which users will approach archival resources we describe or control the way they will be used. In Egypt, twitterers and bloggers were able to by-pass an authoritarian regime and construct a new reality for themselves despite anything that the government could do. Archival resources, once they are released in cyberspace, will be used in ways that we cannot anticipate and cannot determine. Our materials can be combined with other resources to produce quite unforeseen results[35]. With linked data, users will be prompted towards targets found by other users with the same search profile and this can be used by us to point them towards related entities.

There are lessons for us in the WikiLeaks saga. The original model, based on the Wikipedia approach, was for leaked documents to be referenced and authenticated by the "general public" who, it was hoped, would then be contributors to public issues debate. It turns out that the public have not got the time, resources, or interest to do it. So they had to turn to a relatively small team of specialists for authentication and rely on established media sources

to channel the information[36]. The difference between –pedia and –Leaks is made clear by Julian Assange when he says : "If you censor important material of this type, we're not just gonna criticise you, we're going to take the material that you tried to censor and we're going to spread it all over the world and we're gonna stick it in our archives in a way that it's never going to disappear".

To do that, they've had to learn archival techniques, to control how documents are released, and to manage access. At the beginning, Assange said they had to release everything or nothing, otherwise the authenticity of the release would be compromised. We could have told him different. Following criticism about the possible dangers of releasing raw classified material, they began redacting it prior to release – just like archives programmes do with declassification.

It used to be said that the camera does not lie but, with Photo Shop, every image is suspect. We try to provide authentic information, in part by preserving context and structure. The Internet is lousy with bad, unauthenticated information, urban myths, and downright falsehoods. Just ask scientists involved in climate change or medical professionals dealing with parents who refuse to have their children vaccinated. They track it down, disprove it, and expose it but they never eradicate it or eliminate the harm. Of course, there was bad information before the Internet - Area 51 and the alien invaders, for example - but the difference between authenticated data and nonsense is much harder to pick in cyberspace.

We intend for structured information to be approached the way that we, the providers, determine. When releasing US reports and cables earlier this year, Assange argued that it was necessary to orchestrate the way media outlets used the material : "What is new is us enforcing co-operation between competitive organisations that would otherwise be rivals to do the best by the story instead of just doing the best by their own organisation". But the media resisted this and conducted their own sorting, fact-checking, and analysis. The breakaway group, OpenLeaks, has now rejected this approach and is taking a more passive role purely as an on-line conduit to deliver material anonymously to the media. Third parties are constructing tools to assist in searching and accessing the leaked cables – cf. Aftenposten cool cable tool[37].

Some archives are now experimenting with Wiki as a tool in archives search and use[38]. Have a look at Bright Sparcs[39], a sort of register of archival resources in the field of science and technology, and at other applications on the Online Heritage Resource Manager (OHRM) page at the University of Melbourne[40]. Have a look at what family historians are doing on their web sites[41].

Here is a painting entitled *Family of Henry VIII* (**Figure Four**). What is wrong with this picture? It shows Henry with his son, later Edward VI, and Edward's mother, Jane Seymour. But Jane Seymour died giving birth to Edward. They could never have stood side-by-side like this in real life. To the side are Henry's daughters, Mary and Elizabeth, each of whom ruled after Edward. But there are five wives missing, along with Henry's parents, his two sisters, from one of whom later Kings of England descended, and one of Henry's illegitimate children who lived into adulthood. Both Elizabeth and Mary were declared illegitimate at one time, and their mothers are no shown. Why are they there and the illegitimate son not? It is an allegory representing the English succession. Edward is the only unblemished heir so only his mother is shown. Mary and Elizabeth are shown to the side because they were only

to succeed if Edward had no children but another illegitimate child, ennobled as Duke of Richmond, was never included in Henry's succession plan and is not shown. This portrayal is like the description of a Fonds because it shows the component parts of the whole standing in relationships to each other that they never had in real life and because the relationships shown represent only one possible view of the subject on display.

Family of Henry VIII
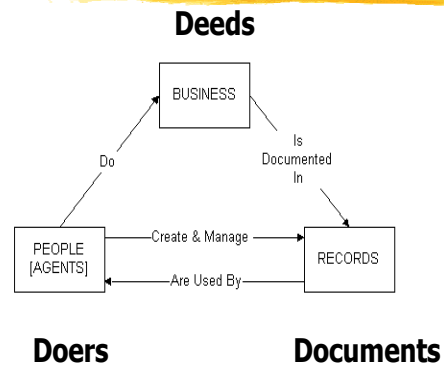

17

Figure Four

Three Entity Model

19

Figure Five

This is not how to deploy archival resources in cyberspace. Our descriptions need to be dynamic displays of descriptive data that enable multiple approaches to the same resource[42]. The ICA descriptive standards will now support this more dynamic approach – inadequately, to be sure, but with enough flexibility to enable it to be done. We need to focus less on the descriptive rules and more on the techniques that are used to manage descriptions and provide access to them. Our rules tell us how to fill in the values (the content) of the properties assigned to entities that we describe. But they are woeful in defining the kinds of entities we describe and the construction of an organised view of them in relation to each other. We need to develop much more sophisticated typologies by way of entity definition and relationship modelling - ways that will support our need to display archival resources by reference to the coherent whole of which they are part.

If we are going to launch our resources (or descriptions of our resources) for searching within the larger boundaries formed by web-based alliances with other repositories or with other heritage programmes, we must pay more attention to the architecture of the descriptive systems that hold and display our data, not simply to the way we capture data content to populate the properties we assign to them. I don't just mean the IT systems we use to hold our data, I mean the descriptive systems we use to capture and manage our stuff. Archival resources have to be displayed dynamically, not merely as artefacts frozen by the accidents that led to them being deposited into our custody and preserved there by us in the archival equivalent of a mausoleum. We must prepare them descriptively for synthesis with related resources consigned into the same descriptive space by other repositories and, if the user wills it, with unrelated resources from anywhere in cyberspace[43].
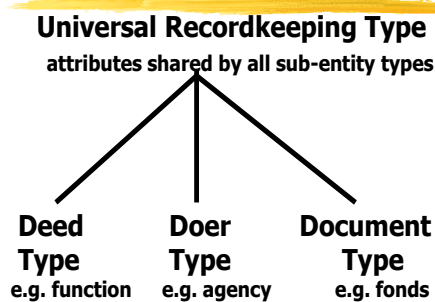
This three-entity model (**Figure Five**) corresponds, more or less, to what can be found in the ICA descriptive standards (or could be found, at any rate, before they made them password

protected) and also in some of the metadata standards applying to recordkeeping. Using this model, archival descriptions can be packaged into units corresponding to

- the documents themselves (anything from a single documentary object to an entire fonds),
- the agents of action or "doers" who undertake the actions the documents record, including the recordkeeping actions taken to make and keep evidence, and
- their activities or "deeds".

We can describe not just a heap of documents on the floor, like a family portrait of Henry VIII frozen in time, but an entire recordkeeping process or a complete family history.
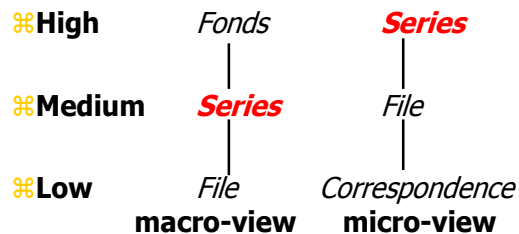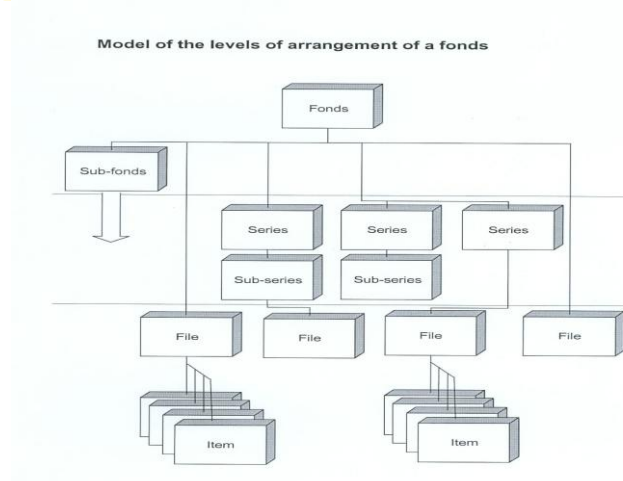

Figure Six

Figure Seven

Although this is called an entity model, it in fact shows entity types (**Figure Six**). That is why it was wrong for ICA to issue a fourth standard for describing archival repositories. Archival repositories are just another instance of the Doer type along with persons, corporations, and families. We are inside the recordkeeping process, not outside of it. All entities of the same type will have a common set of attributes but some will have additional attributes that only they have. The common attributes belong to a super-type which I call the Universal Recordkeeping Object[44].

Although entity types share a common set of attributes, therefore, they are extensible. This means that, although attributes belonging to different examples are common, in some cases inherited from super-types, they can each be extended to include attributes that belong to only one sub-type alone. Thus people, corporations, families, and archival repositories, though they are all Doers, may each have extensible attributes particular to each sub-type. In my view, recordkeeping entities have only three mandatory attributes in common : an identification code (to ensure that they are unique), dates (because all recordkeeping is time-bound), and relationships (because no record stands alone)[45].

Our entity types are scaleable[46] (**Figure Seven**). Their meaning derives from placement within context and not from any *a priori* definition. The placement of an entity into a coherent context, into the network of relationships it has with other entities within the same context, is what tells us what it is. In this example, the series is at different levels in two different aspects of a description. This is not a case of two different series in different circumstances. This is the same series, in the same set of circumstances, viewed and understood from two different user perspectives[47]. What changes are the relationships and the resulting

perspective within which the description is understood, The thing itself remains fixed descriptively, but our understanding of it changes with the perspective taken of its context. This means that the rules upon which such views of the data are taken are recursive, greatly simplifying system design.

The multi-level rule

Model of the levels of arrangement of a fonds

23

Figure Eight

For this to occur, the Multi-Level Rule (**Figure Eight**) has to be discarded[48]. The Multi-Level Rule supports a representation of archives that is like the Family of Henry VIII. Everything is frozen in fixed relationships between entities which they probably never had in real life. The Rule and the view of archival description which it embodies is based on the idea of logical taxonomies, of the kind that are useful in managing other kinds of information resource – books, for example – and useful, to be sure, in all information discovery. Archivists need contingent taxonomies, however, not logical ones to describe things as they actually are, not as we want them to be[49] (**Figure Nine**). We use logical taxonomies when we develop controlled language to populate access points and we also use them to develop our entity typologies and relationship models but when describing what people actually do our descriptions cannot follow logical typologies because that is not how people behave.

Our world, the world we must describe to others, is full of contradiction and confusion. Here is the Duke of Wellington (**Figure Ten**) encrusted within only a fraction of the multiple layers of activity and purpose that surrounded him. There is no pre-ordained pattern or authority file we can consult to determine how all this information should be packaged and presented. We have to figure out what should be depicted as entities and what are attributes and how to accurately and truthfully depict the relationships that will make them meaningful[50]. Military historians speak of Waterloo as a battle between Wellington and Napoleon. But it was a battle between armies not individuals. The British public may have thought that a British Army under Wellington's command fought that day alongside European Allies whereas the Europeans may have viewed it as a British contingent within an Allied Army of which Wellington was Commander-in-Chief. Do we put Arthur and Kitty Packenham together as a

family unit on their own or as part of the Wellesley or Packenham families, or do we portray an extended family comprising both? It hardly matters, so long as we use the same entity-type definitions and the same relationship models.
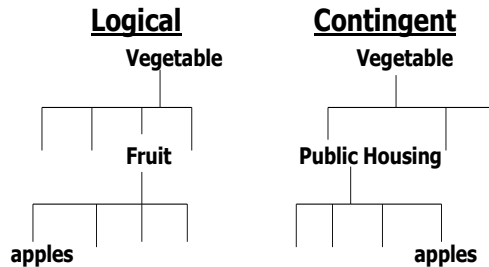
Taxonomies

Logical

Vegetable

Fruit

apples

Contingent

Vegetable

Public Housing

apples

Entity-type definition

Wellesleys   Kitty   Packenhams

Br.Army   violinists

Allies   Pr. Minister

Ambassadors   etc. etc.   Indian Army

24

25

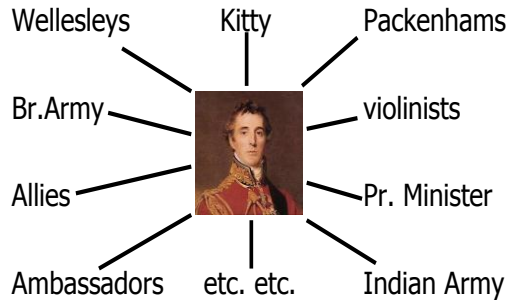**Figure Nine**   **Figure Ten**

How the entities then come together to tell a coherent tale is determined by consistent rules applying to their formation and management. In this example (**Figure Eleven**), "ambience" refers to the context of provenance[51] - the contextual framework within which the actual generators of records are to be found. It means the direction or purpose of a Deed as distinct from what actually occurs (the action). If you are the object of the action, you can't tell what is going to happen to you in a particular instance at the level of ambience. A public housing programme, for example, may be carried out in various ways. Only at the activity level will you know if you are going to receive a rent subsidy or be accommodated in a publicly owned dwelling.

The glue holding entities together in a coherent whole is relationships. Some people I know want to handle relationships as entities, mainly as a device for attaching metadata to them. That's OK but the work of entities and of links between entities is different. I am happy to surrender "relationship" and use the term "link" rather than have a fight over the word. Let's call them the Links-Formerly-Known-As-Relationships (LFKR). Because they are so important, we should spend time over creating a typology of LFKRs rather than arguing over terminology. Previous/subsequent relationships (inherited by and inherited from) and superior/subordinate (owns and owned) are commonly used and understood. For want of something better, we are compelled to call all other relationships "otherwise related". If we want effective systems for entity management, we will have to do a lot better[52].

Recordkeeping relationships are reciprocal. It is necessary to be on guard against attributes that look like relationships but aren't. John and Laura have a relationship (**Figure Twelve**). John has attributes (blond hair, blue eyes, and a firm jaw). Laura is equally attractive. They fall in love, live together and get married. Then, sadly, John falls out of love. This highlights the difference between relationship types. Being married and living together are reciprocal. It is not possible to say John is married to Laura without it being necessarily true that Laura is also married to John. But it is possible that Laura's love for John is not reciprocated. Being in love, it turns out, is contingent. It may be an attribute held in common but it is correctly typed as an attribute rather than a relationship in this diagram. Looked at with an

13

archivist's eye, however, being in love can be reciprocal if it has one of the magic ingredients that turns mere data into records – viz. dates. When John returns Laura's love, the link is reciprocal and, therefore, a relationship, but when John no longer loves her it becomes one of Laura's attributes and not a relationship. Contingency and dates are what we use to avoid the descriptive equivalent of the Fundamental Attribution Error[53] (FAE) – ascribing traits to an entity that are, in fact, determined by context. Data management rules can be used in this way for recordkeeping but they seldom are. We need to take those rules and work out very carefully how we are going to use them.
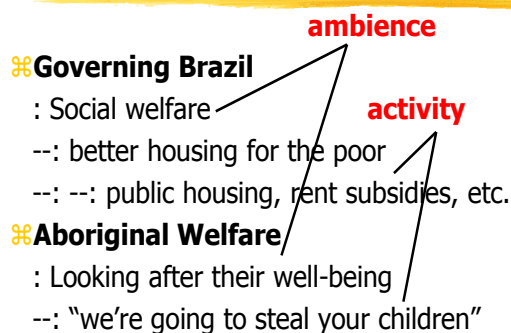

Defined by functionality

ambience

⌘**Governing Brazil**
  : Social welfare
  --: better housing for the poor
  --: --: public housing, rent subsidies, etc.
⌘**Aboriginal Welfare**
  : Looking after their well-being
  --: "we're going to steal your children"

activity

26

Figure Eleven

Reciprocity (John & Laura)

blond hair   brown hair
      in love   married
blue eyes   green eyes

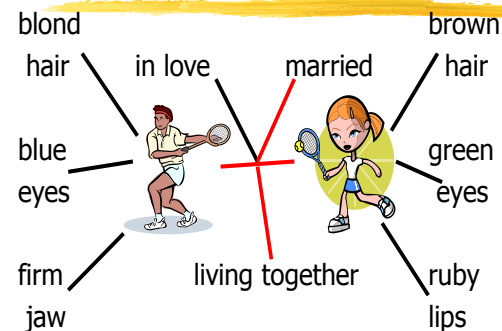firm jaw   living together   ruby lips

28

Figure Twelve

Our data can be viewed flexibly and approached by myriad pathways without losing the richness of its contextual meaning (**Figure Thirteen**). Here we see, at the centre of picture, a Doer carrying out a Deed using Documents. This might be any action linked with the documentation and with knowledge of who is undertaking it. Surrounding it we have links in every direction with entities both like and unlike to establish the meaning of what we have found. Whichever entity has been the target of our search, we are able to follow backwards or forwards in a chain of previous and subsequent entities of like type. We can also go up to see collectivities of like entities to which the target entity belongs, or down to see sub-entities of which it is composed. Superior and subordinate entities may also be of unlike type to show provenance of one kind or another – as when we show the Doer who is the creator of the Document. We should think of this as an interactive map of relationships. You can move the mouse over it and place any of the related entities into the central frame. The network of relationships will then re-form themselves into a new pattern correctly positioning the new target entity inside a different ganglion of links with other entities.

Some of you may be thinking that all this involves more effort and resources than can reasonably be justified. Here are some comments on that :

1. <u>An archivist's gotta do what an archivist's gotta do (get over it)</u>
    The only people who raise this objection are those who don't believe in their hearts that it has to be done. It's like saying : "You need a triple by-pass but that's a lot of trouble and expense so we're giving you an appendectomy instead." If you believe that it has to be done, you find a way to do it.

2. <u>You can only do your best</u>
    This is a methodology, not a work programme. The method does not fail if you don't make all the possible links and separate out all the possible entities. All description

is selective. Things are left out. We make trade-offs between what is useful and what we are capable of delivering. You do as much as you can. You make choices about what is most helpful and what can be done with the resources available.

3. Share the load by working collaboratively

   We can help each other out. In 1986 (almost exactly 25 years ago) I proposed that the Australian archival community combine to produce a "register" or handbook of contextual data at the higher strata of description[54], pre-packaging descriptions for higher level entities of all kinds – Documents, Deeds, and Doers – for use by us all. I estimated that a team of experienced descriptive archivists could prepare such a high-level descriptive superstructure for Australia in about a month. Instead of doing this work for ourselves singly, each of us would just plug in. Workloads would be reduced, high level contextual description would be standardised, and access portals to all repositories facilitated by synthesising the pathways that would thereafter be available in common when approaching the actual records[55]. I further estimated that an archives programme with upwards of 200 fonds should have no difficulty making the necessary connections in less than half a day. I do not doubt that if this had been implemented, Australia's archival portal would today be pre-eminent.

4. Let someone else do the work

   We can use other peoples' frameworks. In the example I gave of the Duke of Wellington, I noted that he played the violin. The question posed was whether "violinist" should be used as an attribute or an entity. In either case, we can look to the work of others to help us. A definition of violinist should be available in several taxonomies or thesauri developed by other information professionals that we can simply link to – we don't have to construct such tools for ourselves within the world of recordkeeping. And if we make violinists an entity, there are likely entity management systems for professional groups or employment categories (statistical bureaus, for example) whose work we can use.
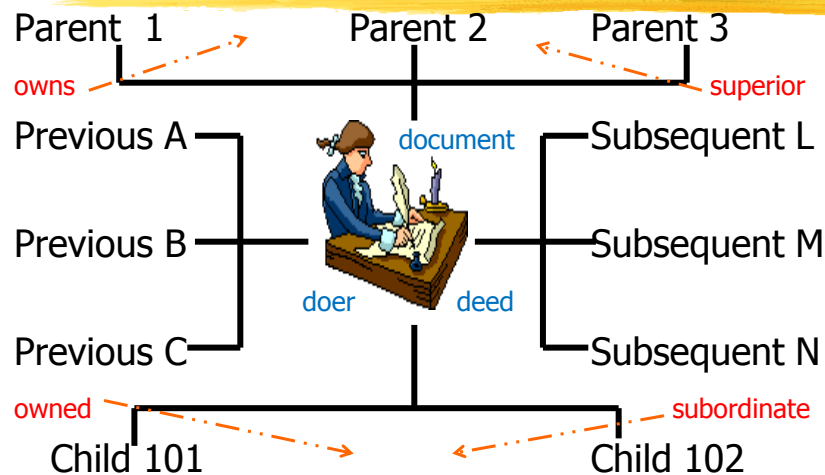
5. Be part of the recordkeeping continuum

   War, they say, is too important to be left to generals. Well, description is too important to be left to archivists. It makes records useable – all records, not just historical ones. I do not only mean better descriptions of records that already exist; I mean also making better records in the first place. Refocus on formation of the record, not just management of the artefact. Become champions of recordkeeping metadata and of managing records within a single process that includes creation, use, and archiving. Expect records to leave their native environment with much of the descriptive work already done at the micro- or sub-structural level regardless of whether they are being migrated to a new application as legacy data or transferred to an archives repository. Understand that re-use of records at the point of creation is now part of the archival enterprise. Become advocates for better recordkeeping so we're not dealing with transfers of inadequately documented electronic records. We can't change the world, but humans are recordkeeping mammals and we know how to do it better. Sooner or later, good recordkeeping will assert itself in cyberspace because it is the right thing to do. Maybe, like cement[56], the secret will be lost for a while and re-discovered later.

Digital resources will be de-contextualised from the moment they are born unless they are immediately taken across an archival boundary where the relationships necessary for their survival as records are preserved and kept. This is what we do. That is our strength. We

know (or once knew) how to do that. Records are transported into an archival realm to preserve meaning but also to ensure that meaning as well as content lies at the end of the road to discovery. For us, then, grace lies in finding better ways to do that imaginatively and effectively. Over the next two days, that is the road to discovery along which you will be travelling.

Structuration (how/when)

Figure Thirteen

Ladies and Gentlemen, I invite you to climb aboard the little truck, seat yourself behind the driving wheel, and start your engines.

END-NOTES

[1] http://www.nytimes.com/1987/07/17/world/iran-contra-hearings-north-s-lies-well-done.html?pagewanted=1

[2] This meaning is not undisputed. Poindexter was convicted but the conviction was later over-turned. The point is that the content of a record is only part of its meaning. In this case, that meaning can only be derived by relating the e-mail to other e-mails and to knowledge of the circumstances in which they were generated.

[3] The archival boundary need not be the thresh-hold of a building labelled "Archives", nor a single step in a process. It has been apparent for decades that the intervention of an archival repository between the generation of electronic records and the systems for preserving them and making them accessible is unnecessary. In the world today, there are many archival boundaries – beginning with the transportation of the record from its point of creation or modification into storage provided by the application in which it was generated. New boundaries are crossed each time it is copied, updated, or communicated and with each upgrade or migration from one application to another. The record crosses these boundaries even if it never reaches a digital archive or an archives repository or, rather, it crosses into an archives with each new state of being. The archives is no longer a mausoleum for dead records, just the latest in a series of transmutations that ends only when the record is finally destroyed and disappears altogether.

---

[4] This is archival description, the moral defence of archives, and the little truck can be used as a metaphor also for this archival enterprise.

[5] The continuum is not another name for a life cycle. It is an intellectual framework for understanding recordkeeping activities and re-aligning them to adjust to changing circumstances.

[6] Why this connection could not be maintained within a conglomerate was not made clear. When asked whether this meant he disapproved of the LAC merger in Canada, he said one had to allow for local conditions. When asked if this meant he would not disapprove of such mergers elsewhere provided the connection with creators was maintained, he said he did not think general principles were appropriate (presumably because they don't accommodate local conditions). His rationale was based on what he described as the continuum of creation, management, appraisal, and preservation. Now, a life cycle by any other name is a life cycle still, even if it is called "continuum". The continuum model does not imply a linear or operational connection between the component elements of the recordkeeping enterprise. It is an intellectual connectivity. But is it truly the case that archives programmes should :
   1. engage in both records management and archival curation, and
   2. not be merged into larger heritage programmes?
So long as someone is looking after other parts of the recordkeeping enterprise and ensuring the survival of an adequate "archival" remnant why can't the archives be the passive recipient? And if the archives can be a passive recipient and preserver, why can't someone different be charged with facilitating public access, at least to its digital resources?

[7] The role of the archives in records management was sometimes justified solely on the basis that it was needed to make sure the archival remnant survived for archival preservation, ignoring the fact that this is an almost trivial part of the recordkeeping enterprise. Some modern standard-setting has the same feel about it : getting the stuff in shape for archival transfer. But in a world of data re-use, "archival" now means much more than that.

[8] Push access is not intended to be information made available in response to a request (pulled out of government against its will), but information made available on the web for re-use by the public (pushed out even if no one asks for it).

[9] Subsequent developments in descriptive standards and metadata should be leading to consideration of supplying such a mechanism for FOI under the push model, rolling up archival finding aids and enabling installation of a comprehensive, up-to-date, and future-proof portal for whole of government access to official information regardless of its archival status. But this is not happening. Instead, records management is being combined with web tools and metadata models to achieve the same result.

[10] As I understand it, the rationale for custodialism is that archival custody is required to protect records that no one else cares about. But digitisation and data re-use changes that. My friend, Frank Upward, was personally responsible for the transfer of these military records when he worked at National Archives. With the benefit of hindsight, he asks whether moving them from Defence into Archives was such a good idea. These records are low-hanging-fruit – with enormous potential for re-use. Had they remained with Defence, it is certain that they would have been digitised anyway by that Department and Archives saved years of effort, money, and diversion of staff resources. Archives could have worked on less sexy records that no one is interested in caring for – the true purpose (if custodialists are to be taken at face value) of archival effort – and they would have been spared the notoriety that comes to them as owners of such a high-profile digitisation project. On this argument, we should not be "competing" with the data sources for digitisation kudos but directing our efforts at preserving the un-loved records that no one else cares about. It will be argued that, if

archives don't compete, then their funding will decline and they won't be able to carry out their high-value work but this is a devil's bargain. At any rate, it is a debate we should be having – but we aren't.

[11] When I received the invitation to come here, I found my passport was about to expire. To renew it I had to deal with three government agencies. This meant undertaking three separate processes. But from my point of view, obtaining a passport is one process, not three. Of all frequently used services, this must be one of the simplest. Yet it has still not been streamlined.

[12] For now, there is nothing to choose between the fragmented state of on-line government services and the fragmented state of on-line retrieval for archival resources. In our world, the parallel to replacing this with service-oriented access will be global discovery of archival resources regardless of custodial arrangements. The Issues Paper talks about "re-use" of government data. This means that the storehouse of information gathered by ministries and agencies to carry out their functions should be repurposed and made available on-line to be made use of by citizens, students, educators, research bodies, and the private sector undertaking research or other work.

[13] See 2010 report on re-use of public sector information at http://www.nationalarchives.gov.uk/information-management/legislation/directive-and-regulations.htm

[14] I do not condemn anyone for the work currently being done. We must crawl before we can walk. People are doing the best they can with the tools they have. What must be condemned is smugness and self-satisfaction, being content to make archives available within the limitations of the tools available, not aspiring to live up to the functional requirements for recordkeeping. Doing better requires us to be profoundly dissatisfied with the limitations within which we must now work, even while we are doing the best we can within those limitations.

[15] Encompassing Information Technology, Information Management, Document Management, and Data Management.

[16] It is someone else's turf and we should keep out. In order to be "relevant" we are running the risk of embracing aspects of data management that go beyond simple recordkeeping requirements. If the word "data" or "information" (or "data management" or "information management") is substituted for "records" and "recordkeeping" in most of the principles set out in the ICA Standard, or the Australian version from which it is derived, what you have is the basis for a perfectly good data or information management standard. You could eliminate all of them except two and replace the remainder with a single statement : that electronic records have to be managed in accordance with appropriate data and information management standards. Efforts have been, and are being, made I know by many archives programmes to upgrade records management and metadata standards but progress is glacial and, I fear, problematic.

[17] But those people need to hear a lot more from us about the two Guiding Principles that do concern us and we are the ones to tell them. We need to learn what we are good at and play to that strength, avoiding the risk of drawing too wide a net and getting into areas where we can't compete.

[18] This indicated that someone had thought about it and reached a conclusion about what was generic and what was not. No criteria were given so it remained unclear how the judgement was reached and whether it was repeatable.

[19] Do we maintain records, digitise them, release or redact them, and then let someone else provide the means by which people use them? Or is there something special about archival resources and the way they are accessed and used? Is it possible that these tools may be used in different, more imaginative, more

graceful ways by those who have the wit to adapt them to hitherto unforeseen uses? There is confusion, in other words, about where we fit in to all of this.

[20] Defining archives services and separating them from libraries became the prime directive for two generations of archivists in Australia and New Zealand. At almost the very moment when the last of them (Queensland) "escaped", the tide turned and now the trajectory is back towards re-amalgamation with other cultural services.

[21] In all three cases, the alignment and streamlining of heritage services and the rationalisation of funding was used as justification for these mergers.

[22] It was not entirely clear, but it seemed to be argued that remote archives could be accessed because resources were increasingly being digitised and, added to a digitisation on demand service, this would preclude the need for network of repositories.

[23] Digitisation has several different meaning. It means developing our finding aids into data bases accessible on-line to assist people in finding archival resources. But it has generally come to mean something else - scanning paper records into digital format so that people can access them remotely. This is a form of what was once called records publication. Reproductions of documents, calendars, and microfilming all involved duplication and distribution which facilitated preservation but was also used to distribute copies of records to other locations and allowing multiple simultaneous use. DIgitisation involves a quantum leap because digitised resources can be displayed on the world wide web, magnifying their exposure, and access to them can be obtained using web-based tools. Archival resources employ the same delivery and use mechanisms as resources from other kinds of heritage programmes such as libraries, museums, and galleries. Very often on-line access is to documentary objects, not to a database. The number of hits is therefore directly comparable to the number of times a paper-bound object is accessed in a search room. Burgeoning on-line access to the digitised portion of an archives' holdings will transform its unit costs. This makes it imperative for archives to divert resources into digitisation and consequently aligns them much more closely with other purveyors of digitised cultural heritage. Most of our archives programmes seem all too anxious to do this anyway. The result is that digital access expands enormously by comparison to non-digital access. In a debate a few years ago over the diversion of resources away from on site reading rooms, LAC came perilously close to saying that it could no longer justify giving "disproportionate" resources to on-site services when on-line access now far outweighed it quantitatively.

[24] This is not serendipity, which is the unexpected find of something where it shouldn't be – the Lincoln letter telling Meade how to follow-up after Gettysburg misfiled into a box of stores requisitions and lost for the next 100 years, for example.

[25] Digitisation now provides an alternative to personal inspection. The actual and potential volume of digital access (which is qualitatively different to traditional research as well as quantitatively greater) establishes a rationalisation for withdrawing or down-grading access services based on personal inspection and traditional research.

[26] We might try to influence the development of generic discovery tools with some of the unique insights of archival description but I doubt that is very likely to happen. Distributed digital access undoubtedly opens up archival resources to a vastly greater number of people and alternative uses that were inhibited if not precluded by the need to consult original documents in difficult-to-get-to search rooms. Let us call these the "new users". While archives programmes MUST expand into this new market to survive and justify the expenditure (mostly public) that supports them, the swamping of old archives users by new archives users does not justify devaluing the "traditional" tools and processes that (until they are replaced by better ones that support access to both digital and non-digital and access to both digitally) increasingly serve the needs of

only a small fragment of our users/uses. I say this, in part, because I firmly believe that it is in re-invented tools and processes to serve the needs of old archives users that we will also find the means to satisfy the needs of pre-archival users of migrated content. The digital programmes, in other words, should not (not yet at any rate) simply be lumped in with programmes for old users/uses and made the basis for decisions about resource allocation on the basis of numbers. Making digital and digitised resources available (to say nothing of making them available digitally) involves much more difficult questions that simply digitising the stuff and applying a search-engine unless the search engine is designed to archival standards. This is not about foregoing any of the advantages of digitisation. Recognising that we have yet to develop the tools and processes needed to make access to digitised records a satisfactory alternative to access to the originals, it is about attacking smugness and figuring out how to accomplish the archival mission in digital space. It is not a luddite argument and it is not about denying the advantages of digital access.

[27] Australia has a national government and eight State and Territory Governments. All of these (or very nearly all) have their own websites, each with its own search engine. If you can get there you can search in nine different places for what you want. Add to these a corresponding number of state libraries, non-governmental archives, special collections and manuscripts repositories and you are getting towards 50 to a100 places to look on-line. And we are small country.

[28] http://trove.nla.gov.au/

[29] http://www.nationalarchives.gov.uk/nra/default.asp

[30] http://www.oesta.gv.at/site/6415/default.aspx

[31] http://www.archivescanada.ca/english/index.html

[32] These are effectively what used to be called an extra-net – i.e. an enclosed web-based space outside of each participating institution but separated from the entire Internet.

[33] Perhaps in Australia most Google searches are about sport, beer, and gaming so the occasional search on a scholarly topic produces better results. Or, it may just be that Australia doesn't have that much history so more of what little there is finds its way to the top.

[34] My thanks go to Barbara Reed, Joanne Evans, Richard Lehane, Mark Stevens, and Andrew Waugh for the information in this section.

[35] The traditional approach to archival materials through finding aids and then into a search room where documents are presented sequentially within the network of associations that tell the story of their provenance and links with other documents cannot be assumed when we present archival materials on-line. Like President Mubarak of Egypt, we will find that we no longer hold the levers used to control how people approach and interpret the information : the finding aids that tell them what is available, the search rooms that constrain how it can be used, the arrangement that directs the sequence in which it is viewed.

[36] I am told that Wikipedia has no higher an incidence of error than Encyclopaedia Britannica, but it is the verifiable techniques of reference checking that makes Britannica a reliable source (despite the errors). Although Wikipedia is now partly referenced, it is still made up of offerings from the public – in stark contrast to Wikileaks which trades in original documents authenticated by using, amongst other things, provenance, diplomatics, and the balance of probabilities.

[37] http://www.aftenposten.no/spesiasl/cablegate/?lang+en

[38] We cannot yet know how on-line access is going to transform the way documentary resources are discovered and used. Archivists have been slow to think through the implications and to adapt to the challenges and opportunities involved in this new way of doing business.

[39] http://www.eoas.info/

[40] http://www.austehc.unimelb.edu.au/ohrm/

[41] e.g. http://shops.ecorner.com/epages/sag.sf/en_AU/?ObjectPath=/Shops/sag/Categories/Databases and http://www.ancestry.com.au/

[42] Australia is the world's largest island. It is also the world's smallest continent. Perhaps that is why those of us who are born there are used to looking at things from different angles.

[43] Some archivists think the last word has been spoken on this subject. They are wrong. We have scarcely begun. For more about this, go to the Records Continuum Research Group web site at Monash University. You might begin with Documenting for Dummies.

[44] This used to be the HERO (Hurley's Enduring Recordkeeping Object) but I was persuaded to change it in the service of modesty. Now my friend and colleague, Frank Upward, is trying to persuade me that it is not "enduring" at all but rather "perduring". An enduring object in philosophy, he argues, "is one that is the same now as it will be in another time or another space [whereas] a perduring object stretches out differently over space time". That would make them HARPO's (which Frank thinks is a lot more fun) – Hurley's Archival Records as Perduring Objects. If for no other reason than I want at this point to run hastily out of the room, I am happy to stay with URO for the moment. Frank suggests I would enjoy an entity/type treatment of perduring objects to be found in Michael Loux's 1999 book, *An Introduction to Contemporary Metaphysics.*

[45] Name is often nominated as a mandatory attribute for all information resources and it is likely that our resources will usually have a name but I doubt that it is necessary for a recordkeeping purpose. Nevertheless, if such optional attributes are universally found in all entities, there is no reason not to bring them up as attributes of the universal type – especially if that makes for greater harmony with other non-archival information resources.

[46] My use of the term scaleability has been objected to – both for the way I spell it and the way I use it. But I say this : If we regard different entities of the same entity-type as being differently "sized" (like Russian dolls, each larger one encompassing all of the smaller ones in graded dimensions), then the entity-type (the Russian doll template rather than a specific doll) is scaleable because it can, without degradation of functionality, accommodate them all (be adjusted to produce a doll of any size). The most common use, I grant, is about adjustment for increases/decreases in volume rather than size. But it can be used more generally for "anything whose size can be changed" :

> **(1)** A popular buzzword that refers to how well a hardware or software system can adapt to increased demands. For example, a scalable network system would be one that can start with just a few nodes but can easily expand to thousands of nodes. Scalability can be a very important feature because it means that you can invest in a system with confidence you won't outgrow it.   **(2)** Refers to anything whose size can be changed. For example, a font is said to be scalable if it can be represented in different sizes.   **(3)** When used to describe a computer system, the ability to run more than one processor. Webopedia

Our entity-types are scaleable because they are adjustable for size and so the systems used to manage them can be scaled in the technical sense to "expand" with the number of Russian dolls we have to deal with without materially adjusting functionality. Here are some other definitions :

> 1. Facility, plant, or unit whose size, performance, or number of users can be increased on demand without a penalty in cost or functionality.   2. System designed to handle proportionally very small to very large usage and service levels almost instantly, and with no significant drop in cost effectiveness,

functionality, performance, or reliability. Scalable systems employ technologies such as automatic load balancing, clustering, and parallel processing. www.businessdictionary.com

Let us call an entity a "unit" for the purposes of this definition. In my scheme, a distinction is made between an Entity Type (e.g. Document) and specific instances of that Type (e.g. fonds, series, item, documentary object). The Document (or "unit") can be hard-wired in to our application for writing, managing, and delivering recordkeeping descriptions. If I have to describe a fonds I can use the Document Type in my system to do so. But because the Document Type is scaleable (i.e. its use can be adjusted "on demand without penalty in ... functionality") the same Document type can be used (with minor variation under the extensibility rule) to describe a series, a file or a documentary object. That is to say, we can design systems whose functionality can be used to handle a variety of specific instances of that Type without changing the functionality within the system design. Let us call my scheme a "system". As described above, I think it may fairly be said to be able to "handle" entities of different dimensions "with no significant drop in ... functionality, performance, or reliability." It might even be said to employ parallel processing (thereby encompassing parallel provenance, perhaps?).

[47] On the macro- view, the series is a middle level entity belonging to a Fonds and containing files. On the micro- view the series is a collectivisation of files, each of which is made up of correspondence. On the second view, the series is doing the work of a Fonds. Using extensibility, the description of the series can be employed either as a middle level entity related to the Fonds or as a high level entity related to a File. The same entity-type and the same descriptive terms are used.

[48] I really can't understand why it is still there in the ICA standard. Revisions to the first edition have made it an anachronism even in ISAD, let alone in useful practices based in ISAD.

[49] We do not construct logical classifications into which information is organised to facilitate retrieval and understanding. We are documenting how people actually behave, rather than how we think they ought to behave. Human beings are illogical, inconsistent, perverse, and often deceitful. If they subsume public housing under vegetable matter then that is how our systems must represent it, however illogical that may be,

[50] . Your description of it will not be the same as mine. If you doubt this, find different cases where parts of the same Fonds have ended up in different repositories and see how successful (or otherwise) archivists have been in applying their descriptive rules to the same material to produce identical results. But this is not important. What is important is that the entity-types we use to contain the information, however you decide to organise it, should be the same entity-types as I use – even if I package the same data differently within them.

[51] Because so little work has been done on all of this, it is necessary sometimes to make up one's own terminology for want of something better.

[52] The December 2010 issue of ICA's Bulletin (*Flash*) reports that the responsible Committee (CBPS) is working on "the description of the relationships between the different types of entities" (p. 22). One's hopes that this will lead to a break-through are not raised, however, by the information that work on a conceptual model is being left until 2012-2016.

[53] It could, of course, be argued, that Fundamental Attribution Error (FAE) affects most attributes – since description is subjective. But we should only care about it when it matters. We do not have to eliminate all FAEs, just those that are relevant to recordkeeping. Our skill lies in identifying which those are, not in mind-numbing analysis of the inconsequential.

[54] The kernel of this idea goes back to one of Peter Scott's cherished but unrealised ambitions.

[55] A kind of super-fonds would be established for, let us say, the Archives of the Government of Victoria, with sub-fonds for the main administrative and juridical divisions. The same would be done for other jurisdictions. Any archives would be able to plug into the top level of their own descriptive work from "underneath" into the contextual superstructure thus provided. I envisaged the same thing being possible for the private sector with divisions for Health, Mining, Pastoral Industry, etc. Such classifications, in fact, already exist in areas such as statistics. It would be easy to mirror them and augment them for descriptive purposes.

[56] The story that the Middle Ages mislaid the formula for cement is something of an urban legend, but it makes a useful illustration.