

**ASSDA Melbourne Node  
LIEF 2009 Archival Workshop – 1**

**Day 1 – 11 March**

**MORNING SESSIONS**

**Introduction**

Workshop to take through the archival process – archival program or ‘intervention’.

Key issue for an intervention not around technology, but about what people do and how they think in terms of records management. A ‘whole business’ process.

Intervention relevant to the social sciences and the social science data archive – need to manage records in a broad sense, qualitative and quantitative. Also running interventions with the researchers themselves. The people running the node are not social science people. It is therefore relevant to consider it an ‘intervention’ into the social science space.

Interviews with Chief Investigators (CI’s) cover not just what we are trying to do, but also the limits.

Want to use the workshop to unpack the protocols and processes, and work on rebuilding the tool we use to create something better suited to the 21<sup>st</sup> century. How do we capture information? How do we store it? Use it?

Julie McLeod – had been thinking and writing a lot about qualitative data, and the issues which will come up in this workshop. Though not an archivist, she has been dealing with archival issues.

**Identified three major themes or challenges:**

- 1) **epistemological challenge** – about meaning and the problems of meaning; preservation of meaning over long (or short) periods of time. You have stuff – how do you make sure it is meaningful to people other than those who created it? How do you make sure there is enough information around the stuff to allow others to understand? Two aspects – diachronic: things that change through time, affecting our ability to understand something from another time; synchronic: at the same time – for example, stuff from one culture, with people outside that culture trying to understand it.
- 2) **Ethical issues** – for Julie, human subjects, children – privacy and other issues which need to be considered; frameworks of ethics agreements, etc. – set in specific contexts, but these contexts can change over time; epistemology = meaning, ethics = people; records are there because there are multiple stakeholders, and not every stakeholder will have a copy of the records in

which they are a part. Therefore, sensitivity to these stakeholders is an important part of the archival process and is, potentially, infinitely complex. As well as those in the records or creators of the records, the user is another stakeholder. In the past, the researcher had to be physically there with the records – this is changing.

- 3) **Practical issues** – whenever data or stuff is created, it is in a particular context, using the technology of the day. For analogue technology (eg. Paper based), the problems are less complex. For digital technology, or a combination of digital and analogue, records have become a lot more complex, formats and forms are unstable, and are evolving very quickly. A period of ‘hyper change’. If we want to preserve records over time, the content has to be independent of the technology in which it was originally packaged.

These three areas are highly interdependent on each other, and intersect/interact with each other.

## Accessioning

Systematic accessioning as a vital first step in telling the story of the records.

Ideally, go to where the records are actually being used, to where they are in their operational space. As soon as you start moving things in the operational space, or removing things from it, you are destroying context.

So accessioning is not just a mechanism for transferring things from one place to another, and it’s not a museum accession (where it is a register of the stuff), it is where the operational world and the archival world come together.

Idea is to go as close as you can to the operational zone (and accept what has happened up to that point) and be absolutely meticulous and painstaking from that point. Be as pedantic and disciplined as you can.

Starting point can be a general survey to get a broad idea of the pattern, structure and content of the records.

Keeping records in the order that they are found. This applies in the paper world – they are equally valid in the digital world, but how does this work?

The order in which things are found is not going to be (or rarely going to be) the order that things are found in the archive. Things have to go along a pathway from one to the other. But if you have people asking questions about an item or group of items at the collection end [the completed archive], you need to be able to trace back along the path to where that item was found [its original location, context, etc.].

**HDMS (Heritage Documentation Management System)** – a tool that was geared for paper based collections. We need to unpack it in detail, particularly from the quantitative data perspective.

One thing we don't have is a project event registration section – documenting the steps and events in the story of the records before the accession even happens. There is currently no place these can be documented *as they occur* – a systematic events table where these can be accumulated as each step happens (photos taken when you first see the collection, meeting with stakeholders, etc.); plus events between the accession and processing, etc.

Move to a more modular system in the next version of HDMS.

Could also do more with accession data, photographs, etc. in the output from HDMS – we are not doing much with the data at this stage. Accession data is to some extent one of the more 'private' parts of the archival process – need to be able to document whatever you see, in whatever language you need, without concern for who is going to see it or how it might be used. Given this, in what ways could this data be represented? What aspects of the accession notes or the accessioning process could be presented in the output; how and where?

**First principle** – unique identifying number for each unit within the accessioning (and archival) process. So Series, Inventory and Provenance all have their own unique identifiers, as well as each unit of the accession.

**Accession unit** – whatever the archivist decides is a sensible group of stuff to be together; some sort of logic, coherence, or similar which suggests the material should be grouped together to form a larger unit of records (eg. A filing cabinet of alphabetised folders; a group of audio-visual material; a collection of journals; etc.)

Sometimes this is relatively clear, other times it is very difficult. Sometimes it is necessary to select a logical point (eg. The top left 'corner' of a group of records) and start. Need to be absolutely meticulous and pedantic, with the understanding that, even if meaning is not immediately clear, meaning and structure will become apparent over time.

Accessioning with unique identifiers gives enough information to allow control over the records from the outset. If required, access can be run from this point due to this control.

**Issue of pre- or post-mortem.** The approach outlined here is primarily based on and has been developed based on end of career, end of life or end of era collections (for people or organisations).

In the case of ASSDA, we are going in mid-career (in many cases, at the height of their research peak) so we have to think differently about what and how we are going in to document.

So we need a project based form of accession – in HDMS we are not documenting all their records, we are documenting the records of a particular project (or particular projects).

The extent to which this is true can depend on how clearly the subject has demarcated their research or projects. Julie McLeod – less demarcation, more like her whole research life; John Wiseman – 46 registered projects they are running which are clearly delineated. And their projects are still running and ongoing, bringing issues of continuing extra data and addition of materials.

What happens to the accession when you are going into a live project? We can't afford to wait, and they don't want to wait – they want to know stuff is preserved now, and make it more widely accessible. The imperative is coming from the researchers.

What are the informatics we might need to add to what we are already using in the accession which we might need systematically when working in the live environment, with live projects, mid-project, even toward the start of a project?

Where required, we can create accessions after the event to try and go back and re-create the story of the records. Where did things come from? When was it donated? Who by? Where are there copies of it? Can keep adding to this information as you find more out about the records.

[Based on this, should there be some method of distinguishing between a 'pre-archival' accession and a 'post-archival' accession or story of the records? Information added before or after? Should it just be date based, or should there be a clear delineation within the database of 'pre-event' accessions and 'after the event' accession notes and details?]

Can back-fill and forward-fill to fill out the story of the records.

## **Provenance**

The provenance of the records as accessioned may not be the provenance of the records themselves.

[At present, HDMS only allows us to directly assign a single Provenance entity to an inventory item, series or accession.]

Provenance has been over-simplified in the archival world – records can go through a chain of custodians; and there are also hierarchical relationships between a network of entities (e.g. Records of the ESRC, within that records of projects, within that records of particular people, etc.).

Linking records to whose records they are is a very important part of record documentation.

In HDMS we need to have varied relationships between provenance entities and records (eg. An entity holding records, records about particular entities, creating entities, etc.).

**A nominative imperative** – everything must have a name or title as well as a unique identifier.

Name/address details etc. [in the Accession details] is a snapshot at the time of the accession – contact details etc. – not to be edited, changed or updated even if these details do change.

**HDMS** - should the list of format types (from Inventory) be included at the Accession stage as well? Would this be of particular use when accessioning digital information (e.g. Word docs, excel docs, etc.)?

Chances are the person taking the project forward will not be the same person as the one doing the accession. So you need enough information in the accession notes and details for the next person to be able to make sense of what has been done.

So, at the stage you are on (eg. Accessioning) do everything you possibly can, do things quickly, document everything, review documentation to ensure it makes sense, etc.

**Date Ranges** – the closer you are to the operational zone, the better the chances of getting good dates and date ranges.

**HDMS Flag** – would a field or flag be useful for items which should be digitised, or would be ‘first priority’ to digitise/image?

**Decisions for Access Dates, Archive check box, Retention, etc.** – should these have a provenance? Specific notes? Details of the ‘decision maker’? Or is it enough to just have these details included in the processing notes?

Processing notes not formalised at present. Is there a possibility to add a checklist? However, while considering this, we need to remember it is a ‘knowledge soak’ not a workflow management system. Need to ensure workflow management remains separate (particularly as HDMS is used broadly, for a variety of purposes, and by people and organisations with various – and in some cases quite strict – protocols).

How to manage workflow is a very interesting question, but the workflow manager needs to sit on top or around HDMS in a modular way, not as an integrated part of the system. Should a modular workflow management system capability be considered for ASSDA?

**HDMS – is what is there enough? What is missing?**

None of HDMS was created with the idea of accessioning or recording quantitative data sets. We want to identify quantitative data sets; and we want to identify qualitative data sets (what does this mean in the social science world?).

**THREE TYPES OF RECORDS: Quantitative; Structured Qualitative (focus group records, etc.); and Un-structured Qualitative records.**

Can we generalise on quantitative data sets (across social science, physics, other science based data sets)? Many have very basic meta-data sets at present. Are the DDI (Data Documentation Initiative) standards potentially useful for these data sets too?

**Complexity of quantitative data** – people do a lot of stuff to it, in order to be able to read it. How do we document the medium, the data protocols, and what people have done to data-sets?

What we are looking at will always be about contextual meta-data in sets – never a ‘mega-metadata’ space trying to hold every variable.

Stuff exists in it’s own epistemological, ethical and practical framework. If all this is documented, there are multiple paths to find stuff and researchers and users can make the decision to do what is needed in the future, if and when required.

## AFTERNOON SESSIONS

Need an effortless process for pushing a quantitative data set through (with all the metadata etc. already in place).

There is a public interest in the preservation of data – saying it is only up to the individual researcher is removing some of our responsibility. Eg. Destroying data is an expedient to get around dealing with the complex ethical issues involved.

Need to think in terms of a 100 year research program and how the records may fit into a program of this type – eg. Longitudinal data sets, etc. And these data sets are often about people and their condition (in collective sets).

### **Need to clearly distinguish between policies for retention and policies for access.**

Again, we need to be aware of the difference between dealing with end of life/end of career records for analogue collections, with a set of documents around managing and providing access to that collection for a repository vs. the process we are going through (mid-project, mid-research, etc.) – and the challenges raised as a result.

ARC agreement – that the CI's will be willing participants – for this project. But at some point in the process we reach a point where their stuff will be transferring out of their custody (or copies will be) so need agreements around what the conditions are and other issues.

We need to look at already-existing access policies, conditions of access, etc. from the relevant parties involved in this project – have a look at the strengths of these policies, and also at areas where they are possibly not quite there.

**OHRM (Online Heritage Resource Manager)** – used to map the bigger context in which stuff sits and set up a complex networks of inter-relationships between things, people, stuff, entities, etc.

Using the OHRM as we go to map and set up an enduring and scaleable register of the node as it develops, including people, entities, other nodes, etc. that we might want to know about or need to know about.

We will be aiming to capture information on eg. Our CI's as OHRM entries which can then be transferred to DDI. Not going to use DDI to store the data, just use it to transfer the data at a certain standard. This means the data will be independent from the standard.

So, within the ASSDA framework, if we have a quantitative data set with information about people, places etc. can you have links back to the authority source where that came from?

DDI will have a sub-set of all the data we have.

Ultimately can set up an 'encyclopaedia of entities' which we can all refer to.

At present we don't have data sets in the OHRM, or the CI's. But we can quickly add in all our CI's with a summary note (not a definitive biography, but details that are directly relevant to their data sets, why they are inter-connected and part of the project, etc.). We can also reference data sets, define accessions as collections and therefore as archival resources, etc.

**DDI 2** – what does it contain?

- document description (metadata about the metadata)
- study description (this comes up – doc description not displayed)

In both levels, have the document title, bibliographic citation, etc. version control  
Study description has abstract, dates, etc. access conditions, obligations of users and what is required of them once they obtain data, etc.

We need to incorporate the good elements of/functionality from **Nesstar** into the **OHRM**.

We also need to move from the hierarchical tree of the current ASSDA site (eg. Australian Studies>Health>Nutrition etc.) to a more networked relationship, with independent relationships, and where the source data set would exist as an object in its own right, linked to many other objects.

Authors as entities; depositors as entities; etc. etc. – to create access points to the data sets by various entities.

The metadata on the current ASSDA site is in an XML document – we need to be able to work with DDI 2; a lot of this data is essentially the same as in DDI 3 so we have to be able to move into this space relatively easily too.

So, what can we put into the OHRM which are good entities, and can be programmed to go into DDI 2, DDI 3, etc.?

Looking now at having a Fedora repository where the XML and DDI files are linked to each other. And we can have data going both ways – so the OHRM can potentially be a point of access. (Possibility at this stage.)

Need to get APS (Archival Program System?) involved to look at how the data is being mapped and controlled.

SPSS – that's where all the editing and cleaning up of data sets happens. Then it is transferred to APS, before being transferred to Nesstar.

**Need to try and get every data set into the public domain** – Google searchable, etc. so when people search for that data it comes up; need to be search-engine harvestable, linkable (ie. Specific url for each page) and citable.

### **Inventory Items**

With most of our projects (end of career, end of life, etc.) a lot of culling etc. has happened already, and/or a lot of the value proposition/appraisal has been done. What we are looking for is duplications, or the occasional thing we can't see any long term value in keeping them. Or things better off in the family archive or personal archive, as opposed to the 'professional' archive or archive of the researcher. These are registered in the accession, but can then be left with the family or similar without being taken through the rest of the process.

For ASSDA, we will be identifying those accessions we want to take through to inventory processing stage – high-priority items etc. The appraisal of value, culling etc. will not have occurred to the same level as an end of career or end of life archive and, therefore, not all items may need to be taken all the way through the process (at least straight away).

An inventory item should never be split up. However, there might be things put together in the physical world which are too inconvenient, or too large and difficult to image etc. These can be brought down to a level more convenient for the user, and easier to image.

Documenting an inventory item is like documenting a data set, but we don't have as many variables involved.

Linking back to the accession, the ideal situation is to document inventory items in the order as found in the accession – ie. The records as found. This is the first level of control – preserving the accession context in its entirety.

However, this is not likely to be a good structure to understand and use the records. This is where Series and Provenance become useful structures.

**HDMS** – looking at controlled language for Conservation field in the ASSDA project? Eg. Data-sets etc. – information which may be transferred across to DDI – to ensure re-use later.

**HDMS9** – re: contents/formats – want to rebuild to make it easier to use. Drop-down menus? Different selection process? What about digital data? Create specific formats for digital data types (need advice from quantitative people on how to describe or identify these digital data types/data sets etc.).



**Questions around archiving a hard-drive (or electronic group of files, eg. Data tree on a computer, storage, etc.).**

Go through the same process – the accession/snap-shot of the directory, as the data source, etc.

We are going to have to have the storage space to put the accessioned file systems from researchers, and ultimately somewhere to be kept permanently. Essentially, we need the temporary storage space (equivalent to holding boxes of analogue records) while the data is being worked on; and, beyond that, long-term/permanent storage space for the digital data taken through the whole process.

Need to convert eg. Word documents into something preservable (xml?). Do we inventory every item/document? Are there some we can group together? If we inventory each item will it take too much time and money? Or are there filing structures which will be the equivalent of inventory which we can bring together?

Or, ultimately, can we inherit better metadata from the users/researchers (harvesting into an inventory management system from the systems used by the researchers, as those systems collect more and more metadata)?

**ASSDA Melbourne Node  
LIEF 2009 Archival Workshop – 1**

**Day 2 – 12 March**

**MORNING SESSIONS**

**If a researcher offers their data set or research, what is the next step?**

We would take this through an accession process; and try to identify all the documentation and other records associated with the research; however, accessioning and bringing an archival program in for data around a project brings changes to the process.

One aspect we need to look at is what is digital, what is analogue, and what is accessible. For example, the email associated with a project may not be accessible for archival processes (it may be held in a central repository or similar). But this can vary from individual to individual.

We are interested in mapping the contextual space around the data – we are going through this with our CIs. They are working in a complex space, and we need to look at what we can and can't include, what we can get copies of, information we know is there but can't get access to, etc. For example, we may have to record that information exists in a certain form/location, but the actual information can't be included for various reasons.

From an epistemological point of view, we need to ensure we collect what we need to make sense of the data.

Say there is a quantitative data set – as part of the accession, we would do this as an accession unit in its own right, collect some information about it (how it is found in relation to other stuff, etc.), collect information about its nature.

Part of our work would be to then create an inventory of items/records we can inventory, and put these in some kind of structure. Make some sort of appraisal or judgement from what we accession to what we inventory – the inventory would not have everything from the accession. Establishing value and meaning would be done in consultation with those involved.

At this point, do we start to collect DDI type data in a fielded way? There is an opportunity to do this – we need to test it out.

The two other concepts to look at, moving from yesterday's discussion, are series and provenance.

### **Series**

Series is a construct the archivists bring to stuff to make it useable and understandable.

In the quantitative sense, the notion of Series doesn't make a lot of sense – data sets are brought together using some sort of classification system. They are more like attributes of a data set, rather than a series, and are a hierarchical way of bringing things together which may or may not be limiting.

'Loose' inventory items with attributes are (or can be) chaotic and difficult to understand. Archivists see patterns of behaviour, filing, record keeping etc. in heterogeneous collections where people have been keeping projects together, records together, etc. as people manage their career and life and work.

These are individualised – from an archival perspective we look to the records and their internal structures etc. themselves. We do not impose an external filing system on the records (unlike library cataloguing, for example). There is a lot of extra knowledge that can be brought to something because of where it is physically (this includes the filing location of digital records).

Doing an accession, you start to see what the Series are. And when using Series, the intent is to try and maintain the contextual integrity of the records. To some extent, if they are chaotic a sense of that chaos can be maintained; but, at the same time, the intent is to make the records usable/accessible.

**HDMS** – we are looking to enhance the Indexing function in HDMS to include 'types' for index terms (Person, Place, Organisation, etc.), and index terms having a unique identifier; also looking at setting up indexes across multiple collections.

**HDMS** - Access conditions – current default is to show no access conditions when a series or item is open. Is this something which should not be assumed – if eg. A Series is open, should it state "Access: Open" or similar?

There is also the practical side – guides and indexes are, ultimately, to paper. What about the purely practical side of storing these papers, acid free folders, shelf space, etc. The material has been judged valuable – we need to address the repository/storage issues now that ASSDA has expanded beyond a primarily digital focus.

There are a limited number of repositories available and many issues resulting from this. So, for Social Science data, there will be physical collections that have to be managed. What are we going to do?

### **Digitisation**

One thing that has changed is around digitisation. For some small collections we have digitised the whole collection, and left the records 'in the wild'.

There are also the standard dilemmas around paper based records – that you have to go to the physical records to view them; guides have to be good, etc.

Digitisation changes what is possible (and comes with its own set of issues). The imperative should be digitisation. The drive is that all stuff will be imaged.

There are data sets in previous collections (eg. Uren) which can be utilised – and this collection has been digitised/imaged.

Though there is a drive for imaging, funding is the issue. Also, before imaging can take place, you need some understanding of the collection as a whole. And the items need unique identifiers (inventory IDs). This does not mean the whole process needs to be completed before imaging can take place – but the ID is required.

### **Do images need to be searchable? In PDF? JPEGs?**

Current imaging practice – none of it is searchable by its content. It may be possible to employ high-end pattern matching search functions or similar, but this has never been investigated.

An imaged collection is such a huge, complex, textural, heterogeneous thing that if it could be searched it could create chaos. Therefore, the reliance is still on really good metadata.

The advantage of images is that users of the guide/researchers etc. could bring a form of annotation to the inventory descriptions, knowledge, etc. Information beyond (and distinct from) the archivist's descriptions. And these would be searchable.

An image not being searchable is also a useful barrier which assists with privacy issues – there is more work required to find it, there are less issues with stuff in the public domain, etc.

**Access conditions, and conditions of use:** these should be light obligations if an imaged collection is open; more detailed obligations (including a Logon environment or similar) for more sensitive information. Some items may be imaged but never be available online.

Fragile items can also be imaged for preservation purposes (so users do not need to handle the actual documents).

Stuart has a different view - believes scanners should be used, so registers can be made, etc.

It is a matter of picking the tool – depends on variations in size, format, bound or unbound, etc. And whether OCR (Optical Character Recognition) is required. But scanning is more expensive (and takes more time). It can also be problematic for fragile items, bound volumes, etc.

**Does each node need this equipment?** The camera is an essential tool in the documentation process, and the fact it can also be used as part of an imaging rig increases its versatility.

**Outsourcing?** Most of the time, we have to take the imaging rig to the records – which reduces the possibility of outsourcing or similar in the majority of cases.

Now, a lot of the drive is toward records being ‘born digital’.

Julie McLeod’s records – there is a lot of analogue stuff. Some she would like to be digitised. Field notebooks, admin files around the project, etc.

The advantage of accessioning everything we find is we don’t miss anything, we get the full picture. The advantage for ASSDA is that we then become quite selective as to what we take through the full process (inventory, into Series, etc.).

The additional material can also be revealing.

This can come out of our archival practice, working on full careers, life archives, etc. There is a lot researchers will do working on data sets that they won’t let us have; and a lot we don’t want.

For qualitative stuff, what constitutes data? Secondary sources? Newspaper clippings? The entire data set for the project is potentially huge – and the researchers may say ‘here’s my data set – people might find it useful’.

Also run into problems copying other people’s materials etc. It is the primary data we are working with.

We are endeavouring to find those things we can put in the public research space, in some form/system that is accessible. What goes into those end points is copies of stuff that is kept locally.

What we want to test out is how we make sure the metadata we need to go with those data sets we collect in a way that can serve a multitude of purpose. And so, if those metadata standards change (eg. DDI) we can draw on that information.

**Eg. CIV website/project (Community Indicators Victoria)** – it brings together a variety of sources and information. It is currently an active project; but it may not always be. Is it ‘archivable’? How do we do this? Look at finding a way of bringing the whole site and the data underneath it and keeping it ‘alive’ (even if off-line). Significant challenges here. There may also be underlying data which is really useful in its own right.

The CIV project is also very rich in terms of entities involved, etc.

At present, there are no specific ways of preserving a site and its underlying data sets. The first step may be to take the whole site and associated files from a server and store it – duplicate the machine, essentially. We also don’t know at this stage whether this site (and sites like it) harvest data; and this will affect the issues involved.

Need to contact the National Library etc. around how they do this, what stage they are up to, etc.

Need to be clear, from ASSDA point of view – what are we actually getting? For example, data sets with some contextual information.

So what are the expectations from ASSDA? Are we trying to get as much data as we can to preserve? Or are we doing the ‘add on’?

If we are going to build an enduring service, we need to start tackling some of these issues. Some are quite straight forward, others are problematic. We shouldn’t be shying away from starting to tackle some of these issues (even if we don’t have the time or resources to make this happen).

Preserving the CIV project/site is preserving a body of work, not the data sets. These are distinct things.

ASSDA is a social science data archive – preserving data from social science.

However, not all the data sets in ASSDA are social science – there are humanities, arts, social science, health, etc. Social Science data archive is a misnomer.

#### **ATSIDA – Aboriginal and Torres Straight Islander Data Archive**

How do you deal with an unstructured set of data (eg. Historical material, manuscript data, etc.) all relating to the quantitative history of the indigenous population? Statistical quantitative, tabulated material; and material on individuals.

A way to organise the material is to start with a publication, and use that as an entry into the huge amount of background material.

Material disrupted – start with the accession mapping, and use that information to reconstruct the order.

Sometimes, we need to take off our ‘archiving hat’ and put on our ‘records management hat’ – if there is a clear or known structure (particularly from the creator or custodian of the records) which has been disrupted for some reason, it can be restored.

Do the easy stuff first – if there are known series, groups of records, material from a particular archive, etc. these can be identified in advance of the inventory work and other more detailed work. These series can be registered, and put in the series description. So a lot of the mapping can happen, which can lead to the physical resorting and shaping.

But a systematic rebuild should happen after the accession mapping of the collection.

Can have an archival guide without inventory – no attached images, etc., but can have series descriptions and other information.

### **Citation**

Part of the challenge working in the digital environment is being able to map collected sets back to their source. It depends on whether material is citable, etc. Ultimately, the idea is all stuff in digital archives will be citable but at the moment this is not the case.

Accessioning is the least standardised part of the archival process. The level of detail can change significantly.

ASSDA – to pull the expertise from whatever node it is in and bring it together, share expertise and knowledge. Intensive workshops (eg. One week) can be effective.

A name or details of an individual – by linking in to large databases (birth records, health records, etc. etc.) bringing together information on that person’s life. Challenges involved in building these databases, the scale involved – but also that the data must be kept permanently.

Issue of finding the proper xml bucket to put information into so it can be shared and accumulated, but that is also system independent. The content has to exist outside the tool. Encoded Archival Context.

There is a need for a standardised place where information can be captured about people and their relationships to other people.

## **Archival Standards**

**International Council on Archives** – govern the archival world  
Organisations and individuals can be members of ICA.

One of its roles is to promulgate high level descriptive standards – International Standards on Archival Description, which defines the key elements of description.

Description of provenance entities needs to be managed separately from descriptions of the material itself.

**ISAD(G)** – International Standards on Archival Description (General)

**ISAAR(CPF)** – International Standard Archival Authority Record for Corporate Bodies, Persons and Families

No link between these and DDI – archiving and data archiving have lived in completely different worlds.

- Accession, Series and Inventory comes from ISAD(G) standards.
- ISAAR(CPF) is focused on record creating entities (people, organisations, places, etc.)

## **Ethics/access**

In ICA archives, there are too many stakeholders involved to get consent from all those engaged in the records.

Keeping name databases raises very different issues from the other databases and records we keep. How do we keep data on individuals which can be used by two or more very different systems?

Raises issues of keeping two files – one containing statistical data; the other containing name data.

Issue with names – should they appear in the inventory descriptions? They are in the accession data at the moment (for Julie's records) which remains private.

For end of career/end of life archives we collect these names and include them in descriptions etc. For project research, mid-project research etc. it is likely this will need to be changed.

What about the assurances provided as researchers? Vs. historians etc. Conditions under which these assurances are provided, etc.

Another option – providing information to researchers, but then also providing a set of pseudonyms or similar. Data also telling a story about the methodology – identifying by name, etc. suggests a different level or type of relationship with the research subject.

## AFTERNOON SESSIONS

Digital imaging procedures and process – a copy of the digital imaging manual (for the Canon PowerShot G9) can be sent out if people would like to see the process outlined in detail.

### **Accession – ANU (ATSIDA records)**

**Tentative dates: 9 – 11 June 2009**

3 teams, 3 cameras (imaging rigs?)

Push some things through to inventory

One central laptop

Cartons etc. for re-housing

Part of the accession process is improving the storage of the stuff.

Issues around storage, holding hard copy records, digitising and giving back, etc. Indications are ASSDA does not want to become a repository for collections of paper records.

Worst position we could get into is becoming an ‘archive of last resort’ for physical stuff. As a broker of high value collections which we provide to the archives, we will be welcomed with open arms (providing fully boxed, quality records, with a guide etc.) Offer is a complete, processed collection for them to hold and manage access.

Attitudes from repositories can change significantly depending on how the collection is offered, and the amount of work required by the repository (or vice versa – the amount of work done before the offer is made).

We need to work on finding repositories who are sympathetic to the sort of collections we are offering as a ‘collections broker’.

It is also possible to image whole collections of records.

As part of ASSDA, we will be building a network of repositories for stuff – each local node will need to have a repository infrastructure to manage all the other records as part of what they are doing. We know all universities have to deal with that.

We (Melbourne node) are going to use **Fedora** (Flexible Extensible Digital Object Repository Architecture [Fedora] Open Repository Software).

A lot of data is already digital, so the accession process is not actually going to physical records etc. Bunch of issues which need to be resolved around tools people use (Envivo?), whether data underneath it is exportable, etc. As we work through with CIs these issues will come up, and we will be articulating these issues.

**Re: funding, etc.** – the proposal can't be made properly until the accession has been done, because a mapping of the collection is required to know what is there and what needs to be done.

If you go through the accession process etc. and think it is not particularly worthwhile, you can put the ball back in their court re: funding etc. Whereas, if the accession process reveals really valuable material it becomes more collaborative, working with the person to find the resources and capability to work on the collection.

All the CIs say they are the tip of the iceberg. But we are limiting ourselves to one or two projects at this stage. It is also exploratory – in that we are looking for issues or challenges we haven't seen before.

### **'Early Intervention'**

How can we work with researchers from the start of a project, or with new projects, to manage projects and remove (or limit) some of the issues which may emerge further down the track? The longer we leave this, the more of these issues we will find further down the line.

We should start this now, dependent on resources, current knowledge, experience, etc.

Issues around informed consent – information given with the understanding it will be held for five years then destroyed. We can't follow up on this informed consent – hold for many years, then release? Never release? Is there an ethical responsibility to destroy if this was in the original agreement?

Need something around the public interest over-riding the original agreement? The problem is not that we haven't destroyed it, it's that we said we would then didn't. What are the ramifications of this?

We need to get into University ethics offices now and make sure these sorts of agreements are included from the start. Working with the NHMRC on guidelines for storing data.

“All possible efforts taken to ensure that those sensitivities and issues will not be compromised”, etc. – but this does not mean data and records will not be stored, with restricted access (where required) for research purposes, etc.

On a case by case basis there may be instances where we do have to destroy the data.

Those frontline issues around ethics etc. are really important.

The problem is not the ethics committees, the problem is the research community. Retention for five years does not mean they have to be destroyed after five years – it just means they have to be kept for at least five years.

Part of the problem is researchers haven't had somewhere to store these. There has been no repository.  
Question around working on a collections policy for ASSDA? Is it specific to nodes? Or is there an overall? Flagged as a good question for the executives.

Can we get summer scholars or similar to work in this area? Honours statistics students, or social science students? Can we tap into these sorts of people? One of the things stopping us from doing this is funding uncertainty – if there is a really solid program these sorts of things can be run, and this sort of work is ideally suited to it. Accession work can be done through the year, with set projects for students to work on inventory and imaging, or quantitative metadata projects.

The idea of internships etc. has been raised – but until our position in the University has been stabilised a little more it is not really possible. Also, if they are not paid then realistically it often doesn't happen. And here (Melbourne University) the wealthy faculties get the students. Money and infrastructure are required – and this could lead to some really good students and some really productive results. Particularly if they get their name on the data sets or publications related to the projects.

#### **Where to from here?**

- push the accessioning that we are doing
- see what issues are coming up
- get some reports out
- get the notes from this meeting on the website
- push the use of the guide to a collection, to try and focus on the 12-18 records of Julie, get series and provenance in place, and get a guide together and out into the public domain
- keep pushing the issue of 'what can we say about these records that can go into the public domain?'
- use this to work on how to transfer or code private information from the accession into publicly accessible inventory and series descriptions
- if we can get this guide in place, it will become part of the matrix of information about the data sets that will end up in the quantitative repository or the qualitative repository – here is the data set, and here is the guide on information surrounding that data set
- to do this we don't need to make any changes to HDMS as it currently stands
- we should be looking at using HDMS in accessioning these projects to investigate all the other issues around digital stuff etc., other fields which might need to be added, other functionality, fields in series, inventory, etc. to assist in dealing with these digital objects
- and running the next workshop around the ATSIDA data
- getting the DDI 2 xml from some of the existing collections, decode them, do an entity analysis of them, work out how to get that data into our OHRM, and work out how realistic it is to use HDMS as a data space