MONASH University
Information Technology

National Archives of Australia
Professional Development Forum
21 October 2008

Making metadata matter

Outcomes from the Clever Recordkeeping Metadata Project

Joanne Evans, Sue McKemmish and Barbara Reed on behalf of the CRKM Research team

www.monash.edu.au

Monash University has been privileged to be involved in collaborative research with the National Archives of Australia, State Archives institutions and the archival profession since the mid-1990s. Such collaborations are essential if research and development is to support innovations in recordkeeping in the digital age.

Outline

- **Metadata**
- **Recordkeeping metadata**
- **The metadata challenge**
- **Service Oriented Architectures (SOA) and recordkeeping**
- **Clever Recordkeeping Metadata Project**
- **Are we ready for SOA?**
- **Building recordkeeping infrastructure**

MONASH University
Information Technology

Maturing understanding of metadata in IT/IS

From metadata as 'structured or semi-structured data about data'

To metadata as data about:

- Data, data sources, data collections, data attributes, data models
- Systems, processes, components of processes
- System environments
- Software suites, programs, program fragments
- Specifications
- Events
- People and their roles in IT systems
- Organisations, departments and individuals

MONASH University
Information Technology

www.monash.edu.au

3

The IT and IS professions have progressed from simplistic definitions to a maturer understanding of metadata and its vital role in the digital age.

Recordkeeping and Library and Information Studies professionals bring a richer, pre-digital world understanding of metadata, but also mindsets and mental models which may no longer be relevant to metadata management in the 21st century.

Metadata is

- Structured data/information that describes an object in order to facilitate its understanding, management and use
- Itself data – it is in purpose/use that it becomes metadata
- Recursive – always metadata about the metadata

... → Metadata about the Resource Metadata → Resource Metadata → Resource

- Intrinsic
- Extrinsic
- Dynamic – accrues and changes as information objects move through space and time
- Complex

And has multiple purposes across different metadata communities

MONASH University
Information Technology

www.monash.edu.au
4

Key characteristics of metadata are listed on this slide. Metadata:

- Is itself data – so what we do to make data accessible, usable and re-usable, also needs to be applied to metadata

- Is recursive – as metadata is also data, there is always metadata about the metadata, about the metadata …

- May be intrinsic, part of an information object (e.g. email headers)

- Or extrinsic, external to it (metadata about provenance, related actions and people, e.g.)

- Is dynamic – accruing and changing as objects move through space and time

- Is complex – as it represents complex, multiple entity realities and intricate webs of relationships

- Has multiple purposes across different metadata communities (resource discovery, recordkeeping, geospatial, digital rights, preservation, document management, data management) who bring different perspectives on information objects to the table.

Metadata applies at various layers of granularity, aggregation and abstraction
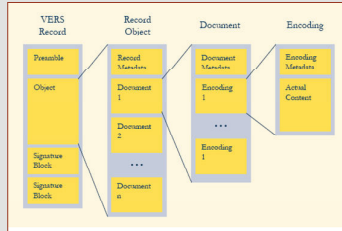
For example, VERS encapsulates metadata about the VEO, Record Object, Documents that make up the Record Object and their Encodings.

MONASH University
Information Technology

Metadata is defined in schemas and standards

- Metadata schema
  - The semantic and structural definition of a set of metadata, including the names of metadata elements, how they are structured, and their meaning
  - Also known as metadata set, specification or vocabulary
- Metadata standard
  - A metadata standard is a metadata schema that has been ratified by some authority for usage by a particular community

Recordkeeping metadata is

- Structured or semi-structured data about:
  - records at all levels of aggregation, their content, structure and context
  - related business and social functions, activities, processes, transactions, events
  - organisations, groups and individuals involved in records creation, management and use
  - recordkeeping functions, activities, processes transactions, events
  - mandates, including laws, standards, business rules
  - relationships
- Standardised information which identifies, authenticates, describes, manages and makes accessible, through time and space, records created in the context of social and business activity

MONASH University
Information Technology

www.monash.edu.au

7

Specifically in relation to recordkeeping, metadata can be defined as on this slide. In terms of digital recordkeeping, quality metadata plays an absolutely critical role in ensuring the creation, capture and ongoing management of the authenticity, integrity, reliability, accessibility and useability of records.

Cook's "Electronic records, paper minds …"

"We have 21st century digital technologies, but automated 19th and 20th century "industrial" systems"

"We need to start a revolution in IM, focussing on:
- Sharing and re-using information
- Evidence based policy and decision-making
- Building IM into our organisation's DNA – integrated into all applications."

National IM Skills Summit, 15 October 2008, Parliament House, Brisbane

MONASH University
Information Technology

www.monash.edu.au
8

The Government of Queensland sponsored IM Skills Summit in Brisbane on 15 October brought together CEOs, CIOs and leading professionals and educators from across all the Information professions (information managers, geospatial experts, statisticians, librarians, data managers, IT managers, records managers and archivists).

Two challenging statements made by keynote speakers (Laurence Millar, Deputy Commissioner and Government CIO, NZ State Services Commission, and Karen Pile, Director of Information Strategy and Services, UK Department for Business, Enterprise and Regulatory Reform, are quoted on this slide. They recall the title of Terry Cook's *Archives and Manuscripts* article of the 1990s …

The metadata challenge: moving from the Wright Brothers model of metadata management

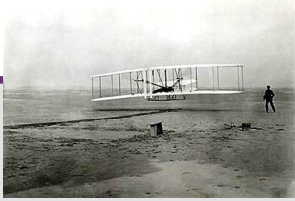MONASH University
Information Technology

www.monash.edu.au

9
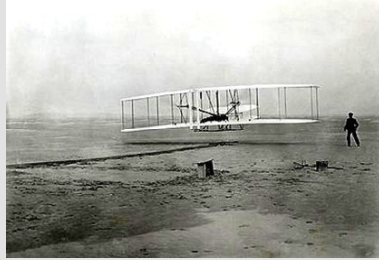
One way of thinking about the metadata challenge is to envisage us as on a journey from a Wright Brothers model of metadata management …

To a 21st century model

From:

- paper paradigms and mindsets

- "paper" standards – and many of our RM standards and instruments, e.g. disposal authorities still predominantly work within a paper paradigm

- automated paper systems – although EDRMS have evolved many add-on features and work through web interfaces, their functionality is still largely concerned with automating the registry and recordkeeping processes developed in paper registry systems of the 19th and early to mid-20th centuries.

- Unsustainable, unscalable, expensive and resource intensive manual metadata creation and use processes

- Stand alone systems and digital repositories, or hardwired applications that achieve a degree of interoperability in particular implementation environments only

- Metadata standards and schemas that do not support interoperability – arguable none of the standards and schemas developed so far do …

Towards

- **Digital paradigms**
- **Digital standards**
- **Digital recordkeeping processes and systems**
- **Sustainable, scalable, automated, metadata creation, gathering, sharing and re-use processes**
- **Integrated systems and federated digital repositories**
- **Metadata interoperability**

MONASH University
Information Technology

www.monash.edu.au

12

Clever Recordkeeping Metadata Project

ARC Linkage Project mid 2003-2006
- Chief Investigator Professor Sue McKemmish, Monash University
- Partner Investigators Professor Anne Gilliland-Swetland, UCLA, and Mr Adrian Cunningham, National Archives of Australia

Industry Partners and Collaborators

Project Links
- InterPARES - the MADRAS Metadata Registry
- ISO23081 and IT21/7 – Recordkeeping Metadata Standards
- NAA Redevelopment of Commonwealth Recordkeeping Metadata Standard

http://www.infotech.monash.edu.au/research/groups/rcrg/crkm

MONASH University Information Technology

www.monash.edu.au

13

To address some of these challenges, recordkeeping researchers at Monash came together with practitioners from National Archives of Australia, State Records NSW, and the Descriptive Standards Committee of the Australian Society of Archivists in an ARC Linkage Project. The project grew out of the earlier ARC SPIRT Recordkeeping Metadata Project, also in partnership with NAA and SRNSW, as well as the State Archives of Queensland and RMAA, and also engaged with a number of other initiatives.

Objectives of CRKM

- **To demonstrate how to move away from stand-alone applications and resource intensive, manual metadata creation and use processes**
- **Towards integrated, interoperable digital recordkeeping systems and processes**
- **And the clever use of metadata:**
  - created or gathered once
  - shared, used and re-used many times

MONASH University
Information Technology

www.monash.edu.au
14

The vision of recordkeeping metadata processes that create of gather metadata once, then share, use and re-use or re-purpose it many times can only be realised if we have:

- Reliable standards compliant metadata about the transactional, provenancial, jurisdictional, administrative and functional contexts in which records are created, captured and used, available to recordkeeping processes
- And intergrated systems environments which enable data or metadata to be shared between many different kinds of systems and re-used/re-purposed as recordkeeping metadata.

Neither pre-condition is yet in place, but …

Towards 21st century systems enabled by 21st century technologies
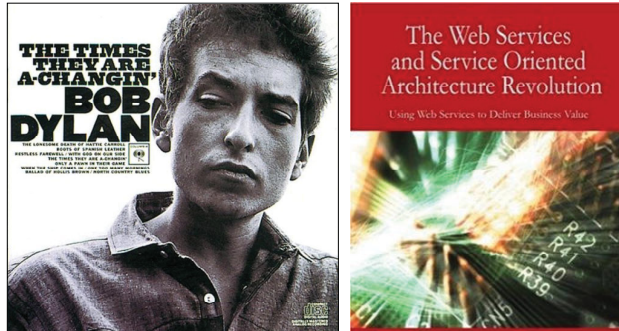
The times are changing and 21st century developments in the web services and service oriented architecture

Web services and service oriented architectures (SOA)

- **Fits emerging IT paradigms**
- **Enables integration into business systems**
- **Decomposes (but also potentially radically changes) means of delivering recordkeeping**
- **Suits the Web 2.0 world**

MONASH University
Information Technology

www.monash.edu.au
16

Given the focus of the research project on innovation and using the power of emerging technologies, we deliberately sought out the emerging architectural environments in IT in which to base our research prototyping.

This led us to the web services and service oriented architecture world. This is an emerging IT paradigm for building systems. It supports the type of work we were exploring in enabling the power of automated metadata, particularly through innovative ways of enabling access and reuse of functionality, and in enabling access to data residing in legacy systems not built from the ground up in the web enabled environment.

What we found in that world is very exiting indeed for recordkeeping. It is a text oriented world using multiple recorded instances of messages working in tandem at phenomenal speed. We began to think of different ways of delivering recordkeeping – as standalone services building together to form recordkeeping functionality. This way of thinking takes us away from the monolithic structures of standalone systems and introduces very powerful but quite different ways of approaching systems and functionality. We began to see how recordkeeping can take advantage of this environment and also how we need to redefine and repackage what we currently consider recordkeeping functionality.

This type of technology was an exiting place to play. It was new and different with very challenging requirements for our recordkeeping processes, but aligned to the technologies that have proven so popular on the web 2.0

platforms. At the moment there is considerable professional discussion about the impact of web 2.0 on the practice of records management. While the project did not explicitly explore that space, the techniques of web services that are the architectural basis of the working of the web 2.0 environment were definitely within our scope.

Web services

- **Not that new**
- **Moving to new uses**
- **Components of larger programmes**
- **Products in their own right**
- **Constructed to undertake repeatable and reusable bits of a process**
- **Reusable across many different applications**

- **'loose coupling' between application that uses service and service itself**
- **Can be public or private**
- **Bring issues:**
  – Granularity
  – Packaging
  – What will be useful?

MONASH University
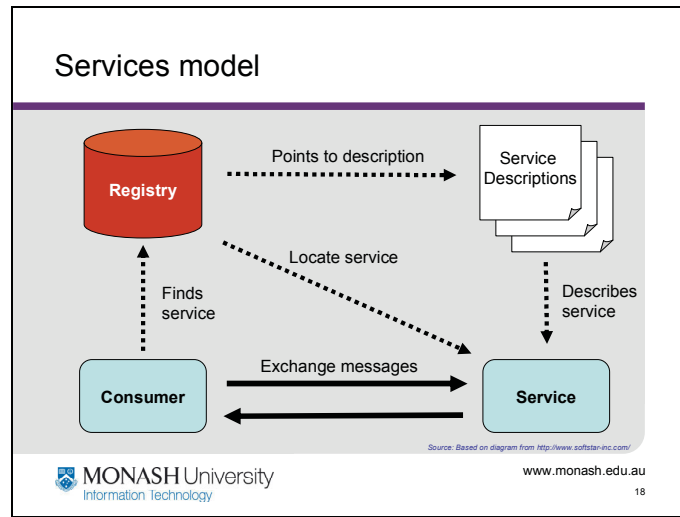Information Technology

www.monash.edu.au

17

Web services is one of those terms that we need to clearly define when we are talking about it, to ensure that we communicate like with like. Web services are not that new. They are the standards based way that distributed systems talk to each other and are one of the building blocks of the distributed systems technology that has been around since the late 1980s early 1990s.

But what we are talking about when we are discussing web services in the SOA world is a different version of web services, using the same sets of protocols. Here we are talking about services that are products in their own right. They are small chunks of programmes that have been packaged to deliver specific and replicable outcomes, and to work on all technology platforms. They are 'loosely coupled' which essentially means that they will work, if defined correctly, across many different types of applications. The services are not dependent on one specific application or mapping, but are able to perform their functions in all environments.

Web services need to be defined and designed to deliver small chunks of functionality. They need to be defined at an appropriate level of granularity – just what bit of action will they perform, and if they are to be really useful to multiple organisations, what layer should the functionality work at. How

should the functionality be packaged to deliver the required outcomes. What is replicable and reusable across organisations.

For recordkeeping this opens up really interesting territory. Just what would recordkeeping services be. How might we define recordkeeping as small, packaged components that deliver reliable and replicable outcomes but also enable us to build to more complex outcomes. The National Archives and Records Administration (NARA) has done some terrific pioneering work in this area, but we need to think the packaging and processes through from an Australasian perspective to meet our particular recordkeeping cultures and needs.

Services model

Registry — Points to description → Service Descriptions

Registry — Locate service → Service

Finds service

Consumer — Exchange messages → Service

Describes service

Source: Based on diagram from http://www.softstar-inc.com/

MONASH University
Information Technology

www.monash.edu.au
18

So how do they work? A consumer (in most cases, a programme) looks for a service that will deliver the required functionality. The consumer looks in a service registry. The service registry can exist either inside or outside the organisational firewall. Because of security concerns which have yet to be fully solved, often these registries are brought within the organisational boundaries, but we can see the beginnings of truly ubiquitous registries in the web, with the emergence of web services listings on sites such as Source Forge and others.

The registry has a listing of services it can deploy. These service descriptions are registered by service providers, along with issues involving rights to use, contractual obligations etc. The registry identifies the appropriate service description and then locates the service itself and passes information back to the consumer. The consumer negotiates with the service through the exchange of messages, and calls/invokes the service as is required, with the service performing the specific bit of functionality and passing the results back to the consumer, as required.

Describing these processes this way tends to minimise the fact that this is all happening at phenomenal, network transaction speed. The time lag between the stages of the process are minimal – a bit like sending an email message.

Service Oriented Architecture:

SOA is an approach to architecture whereby business services are the key organizing principles that drive the design of IT to be aligned with business needs

A paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains. It provides a uniform means to offer, discover, interact with and use capabilities to produce desired effects consistent with measurable preconditions and expectations

OASIS definition

MONASH University
Information Technology

www.monash.edu.au
19

So, once we have the notion of web services being able to define and deliver specific functionality, we need a way of knitting them together. And that is what the Service Oriented Architecture framework provides. This is a big picture way of redefining how an organisation's IT systems will be developed and work together. The definitions on the slide emphasise the business as the key driver, which represents a taking back of the technology to serve the business rather than the technology driving the business. This is a trend observable in all organisations as technology capacity matures. The OASIS definition also emphasises the distributed capabilities and multiple ownership domains – services may not be owned by an organisation, but invoked into their systems.

SOA provides a framework around which web services can be designed and delivered.

Reference model

- **AGIMO Architecture Reference Model, June 2007**
  - Lightly customised version of US OMB FEA
  - Provides common language for agencies involved in delivery of cross agency services
  - Supports identification of duplicate, re-usable and sharable services
  - Enables more cost effective and timely delivery of ICT services through repository of standards, principles and templates

MONASH University
Information Technology

www.monash.edu.au

20

So, where is SOA at present? Broadly speaking it has been adopted firmly in many government jurisdictions as the preferred framework of the future. In the US, the Federal Government established the Federated Enterprise Architecture Framework (FEAF) under the Office of Management and Budget 'to identify opportunities to simplify processes and unify work across the agencies and within the lines of business of the Federal government'. AGIMO established the Australian Reference Model in June 2007, which is described as a lightly customised version of the US model. There is value in consistency here as in other IT and business related worlds.

AGIMOs Reference model aims to:

- provides a common language for agencies involved in the delivery of cross-agency services
- supports the identification of duplicate, re-usable and sharable services
- provides a basis for the objective review of ICT investment by government
- enables more cost-effective and timely delivery of ICT services through a repository of standards, principles and templates that assist in the design and delivery of ICT capability and, in turn, business services to citizens[1].
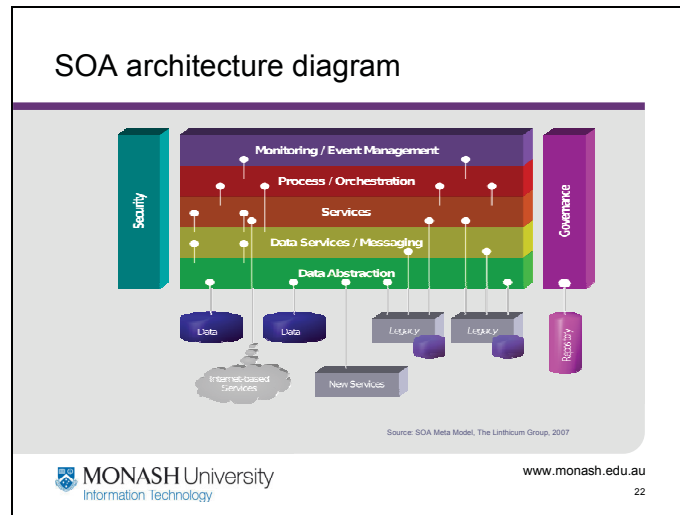
---

[1] Ibid

So, how successful is the uptake? Well it is an emerging framework. Many organisations say they endorse or are doing SOA. But in reality it is a significant shift in IT design and delivery. The rhetoric and the reality are often at odds. Organisations are all looking to this architecture but I would be wary of assuming that the rhetoric is necessarily backed by implementation reality

Gershon Review, October 2008

- Many private sector organisations have already embraced service oriented architecture (SOA), where ICT assets are aligned to business services in a standard, flexible and architected fashion, and also the benefits that come with it such as increased agility, re-use and reduced costs. Government organisations are slowly moving towards leveraging SOA in the development of their applications, but there is potential for the pace to be quickened and for further thinking to done at the whole-of-government level rather than agency level to maximise the potential benefits

MONASH University
Information Technology

www.monash.edu.au

21

The Gershon Review, the 'Review of the Australian government's use of information and communication technology' publicly released in October 2008 contains disappointingly little to support the SOA direction, but it does seem to endorse it as the way of the future. It also suggests that the uptake is more advanced in private enterprise, but that further development and thinking at the whole of government level will be required to maximise potential benefits. This was also the thrust of a number of submissions to the Review, supporting the continuation of the uptake of SOA in Commonwealth agencies.

SOA architecture diagram

Monitoring / Event Management
Process / Orchestration
Services
Data Services / Messaging
Data Abstraction

Security

Governance

Data Data Legacy Legacy

Internet-based Services

New Services

Repository

Source: SOA Meta Model, The Linthicum Group, 2007

MONASH University
Information Technology

www.monash.edu.au
22

So, SOA provides an architecture, a framework for the management of technologies which are constructed according to web service protocols. It needs a framework. There needs to be strong continuing attention to governance – who is doing this, are we all doing the same things, what is an acceptable contract for use etc, and of course for the security framework around these things.

This diagram is particularly interesting for two reasons relavent today. The first is the recognition that data stores and legacy systems are a vital part of the architectural landscape and this way of approaching IT enables the value of those assets to be continuously exploited and reused in new ways.

The second part of particular interest to us is the layer called 'Process/Orchestration'. This is the layer that defines what the business process is and governs the mechanisms and points at which web services are invoked. This is detailed business analysis and work process analysis. We can work with this using our existing skill base to determine the points of intervention for deploying records services.
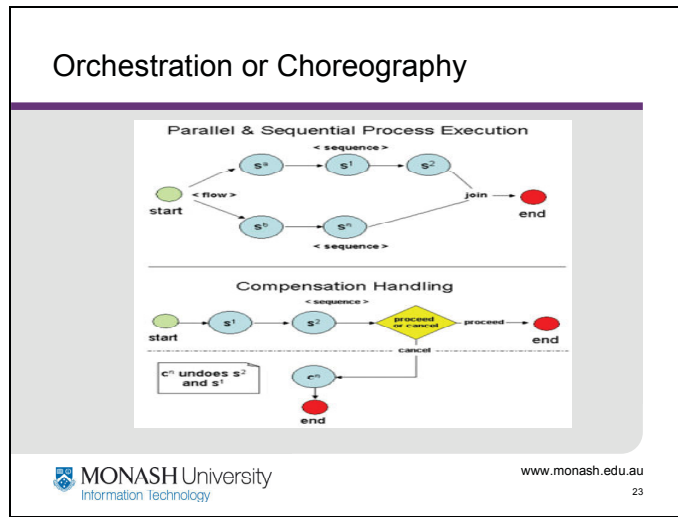
This slide represents an example of an orchestration (sometimes called choreography). This shows the type of process flows that are being designed and also provides us with some optimism for our ability to define specific records services into process design – to suit the business, not to suit the records technology!

$5 million SA tech project canned – Australian IT

Mahesh Sharma | *October 17, 2008*

THE South Australian government has pulled the plug on its $5 million records management system project, ending a five-year saga plagued by repeated cost blowouts, delays and confusion.

The project, commissioned by SA's Department of Families and Communities (DFC) in September 2003, was originally budgeted at $2 million but increased to $4 million after the South Australian Housing Trust (SAHT), which was funding the project, granted a request to double the spend.

However, SAHT froze the project in mid-2007 when the department requested an additional $1 million for the project.

Subsequently auditors KPMG was called in to evaluate whether any components of the project could be salvaged, but in May this year the firm recommended the records management system be scrapped altogether and that the department start from scratch.

Details of the project's demise emerged during a government budget and finance committee hearing earlier this week.

It remains unclear who is responsible for the failure of the project as there were two groups directly involved in the project - Housing South Australia and the department.

Housing SA runs services on behalf of the Housing Trust and had a service agreement with the department.

"There were a number of changes (to the project's management)," Housing South Australia corporate services director Dennis Huxley said at the hearing.

"At one stage it was in the Housing Trust, then it transferred to DFC, then we changed it back to Housing.

"A number of changes occurred over that time. The project has been floundering for some time," Mr Huxley said

MONASH University
Information Technology

www.monash.edu.au

24

We know that we must move beyond the electronic recordkeeping environment of today. EDRMS is not the end or the pinnacle of recordkeeping service delivery. Newspaper reports beating up failures, constant criticism of the clunkiness and non-user friendly nature of these products are continuous and undermining. The 2007 Ministerial Advisory Council Report 'Note for File' identifies the need to continue to examine and develop different ways of delivering recordkeeping. It identifies three future technology directions.

We need to rethink the ways of delivering recordkeeping, and this emerging environment provides a fertile new ground for exploration and development.

Opportunities

- **Away from expensive and fragile hardwiring of integration**
- **Services as a document-centric technology**
- **Orchestrations for business processes, including recordkeeping**
- **Possibility of transforming recordkeeping in the long term**
- **Interim strategies available for more immediate uptake**

MONASH University
Information Technology

www.monash.edu.au

26

So the opportunities and capacities of this emerging world seem very exiting and well worth further investigation.

Services are a document centric technology – they work on the exchange of messages – it is a selection (or appraisal) process to work out which messages to capture. Recordkeeping professionals are not yet participating in that conversation, but we could be and we should be.

The use of orchestrations and their similarity to the types of business analysis we need for recordkeeping offers us a great opportunity to deploy records services to do things like capture a record, move a record, perhaps lodge a record into a provenance string etc. Working out what the services are and what granularity of services to serve recordkeeping outcomes is the next urgent step

In the long term, we can begin to imagine very different possibilities for recordkeeping. Imagine recordkeeping functionality entirely delivered by components of programs invoked as and when required into our business processes. There may be no applications. There may be only an organisational repository or secure storage space, and even that might not exist within the organisational boundaries.

But in the short term, there are many opportunities to grasp in very practical ways.

It is a new way of approaching integration that moves beyond the current state of best practice which is to achieve integration into an EDRMS through handcrafted hardwired integrations. Anecdotally I was discussing this with a practicing records manager who was bemoaning the need to build these interfaces at a cost which was estimated at $10,000 per interface (which actually seemed pretty cheap). But, he explained he has 18 business systems that he needs to interface with. And then, each of the interfaces is fragile – movement of either the business system software product, design or build or in the EDRMS may well invalidate the individual interface requiring

The development and thinking behind the design of the CRKM metadata broker is one very practical, pragmatic and immediately useful tool to begin to build recordkeeping services. There are many more needed and the lessons we learned are relevant to future development in this space.
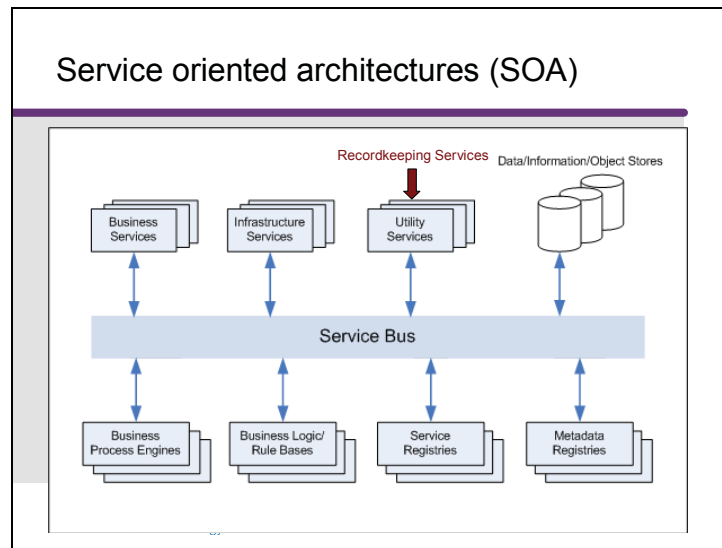
Service orientation results in the radical transformation of IT configurations, as applications are broken down into re-usable components, to be deployed on demand, when and where, they are required.

One way of illustrating SOA:-

- business processes, governed by business logic or rule bases, calling on business, infrastructure and utility services, with support from appropriate service and metadata registries, to drive their discovery and deployment at various levels of abstraction
- trend towards shared data/information/object repositories rather than being locked away in individual applications.

These kinds of architectures are becoming a reality through the maturing of web services technologies - lightweight communication and exchange protocols, which deliver the baseline interoperability, on which these frameworks can be developed.

Good news for recordkeeping:- [Continued]

- Not only is interoperability a given, but so is metadata.

- Particularly metadata about the business contexts in which recorded information is created and consumed.
- They have the potential to deliver integrated systems environments that allow for inheritance and re-use of metadata
- Recordkeeping can leverage off this framework rather than having to wear the cost of building it for itself.

To do this have to start thinking of recordkeeping functionality from a service oriented perspective:-

- [CLICK] Conceive as utility services which can be configured into business processes and …

Recordkeeping in SOA
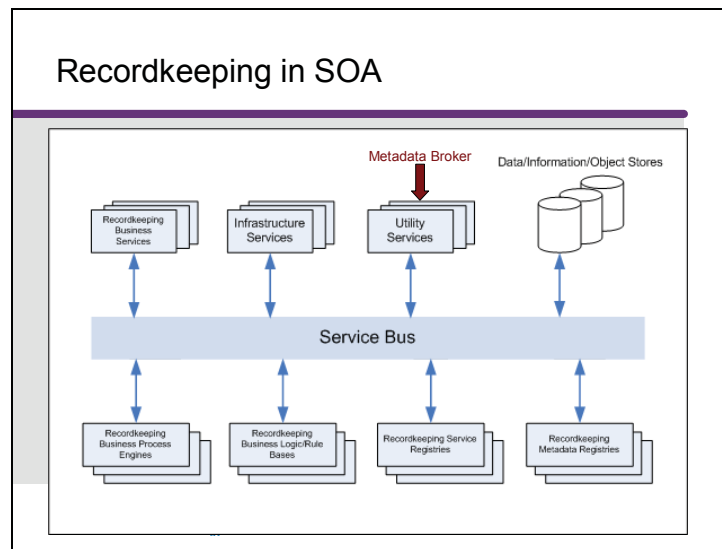
… conceptualise within a SOA paradigm as being made up of recordkeeping business process and rule bases, supported by registry and repository infrastructure.

So we need to think about how to go about designing and constructing appropriate recordkeeping services which can be configured into business processes.

- From low level - – for example to capture a record, ingest a record into a repository, retrieve a record, etc.
- To higher levels – appraisal driving records creation and recordkeeping metadata capture

Now what we did in the CRKM Project was to attempt to build a piece of recordkeeping services infrastructure …

- o [CLICK] a metadata broker as a utility service to translate metadata between standard schemas.
- o The idea was that building such a Broker, and applying it into a simulated scenario, would enable us to demonstrate description as automated capture and re-use of metadata, and in so doing address some of the sustainability and scalability challenges for electronic recordkeeping.

CRKM Metadata Broker
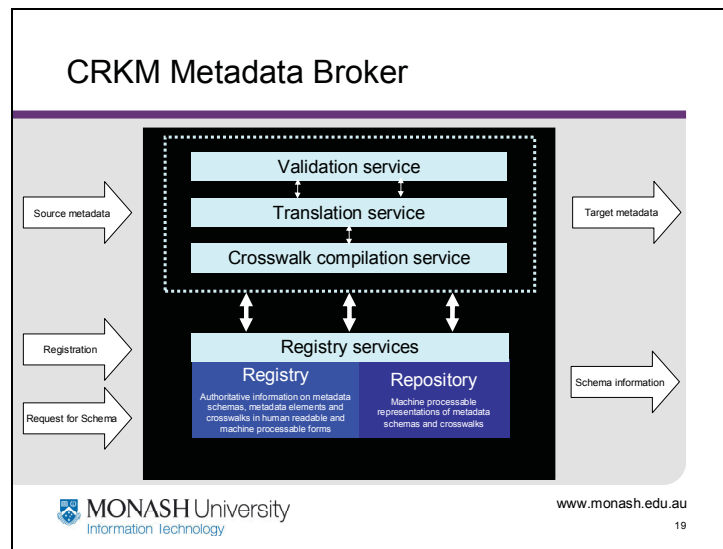
We conceived of the Broker as a cluster of services:-

- [CLICK] A Translation Service
- [CLICK] Which calls on a Repository of schemas and crosswalks
- [CLICK] Which are registered in a Registry …
- [CLICK] And accessible via Registry Services.
- [CLICK] We identified the need for a Validation Service to be associated with the inputs and outputs of the Translation Service
- [CLICK] And for a Crosswalk Compilation Service that would enable us to move beyond hand crafted crosswalks to the dynamic assembly of mappings between schemas.
- [CLICK] All needing to interact to provide their specific functionality.

Metadata Broker as Web Services Cluster
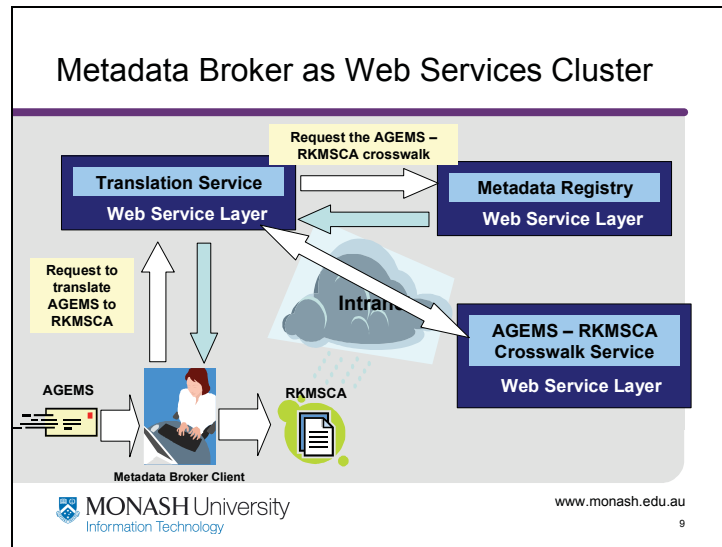
Walkthrough a proposed usage scenario for our Metadata Broker – moving an email into a records repository – where wish to translate metadata from the Australian Government Email Metadata Standard (AGEMS) to the Recordkeeping Metadata Standard for Commonwealth Agencies (RKMSCA).

In this case we want to re-use [2]:-

- o Sender, Receiver, Other Addresses as AGENT metadata
- o Security Classification as RIGHTS MANAGEMENT metadata
- o Message Id, Subject as IDENTIFICATION metadata
- o Keywords as SUBJECT metadata
- o Dates as HISTORY metadata
- o Reply to, references, attachments as RELATIONSHIP metadata

[CLICK] The process in which the need for such an action sits communicates with the Broker, invoking the TRANSLATION SERVICE component, with a request to translate metadata from the source to the target schema.

[CLICK] So the TRANSLATION SERVICE searches for a crosswalk to undertake this translation by contacting the REGISTRY SERVICE of the Broker. [continued]

---

[2] See Appendix 1 of http://www.naa.gov.au/Images/Email_Metadata_Standard_tcm2-911.pdf

The REGISTRY holds metadata about mappings between schemas (crosswalks), – e.g. their source and target schemas, how they are encoded, how they can be accessed, etc. – to enable their automated use.

[CLICK] The TRANSLATION SERVICE then receives and processes the metadata about the crosswalk, in particular establishing what it needs to do in order to invoke the crosswalk to carry out the desired translation.

[CLICK] As we were concerned with sustainability and scalability, the aim was for the Broker to be able to support a variety instantiations of crosswalks in machine processable forms. Crosswalks can themselves come wrapped up as a Web Service, with their complexity hidden from the client, and the ability to invoke them without having to install any particular applications.

[CLICK] Finally the TRANSLATION SERVICE sends the source metadata to the crosswalk for translation, with the target metadata instance sent back to the client, so that they can continue on with their work.

So what?
- Doesn't existing functionality allow for this?
    - Sure many EDRMSs are ODMA[3] compliant and allow for direct registration and re-use of data/metadata from the document in the EDRMS fields
    - but this only holds for those working in native Windows environments, and for those working with other ODMA compliant applications.
    - With the Broker we wanted to overcome this kind of technical constraint for a more universal solution.

- Are we expecting users to stop and do this every time they need to put an email into the EDRMS? [continued]

---

[3] ODMA - Open Document Management API

- No – [CLICK] the idea is that you would configure this into the business process so that it happens seamlessly from the (human) user perspective as they send off their email. [Continued]
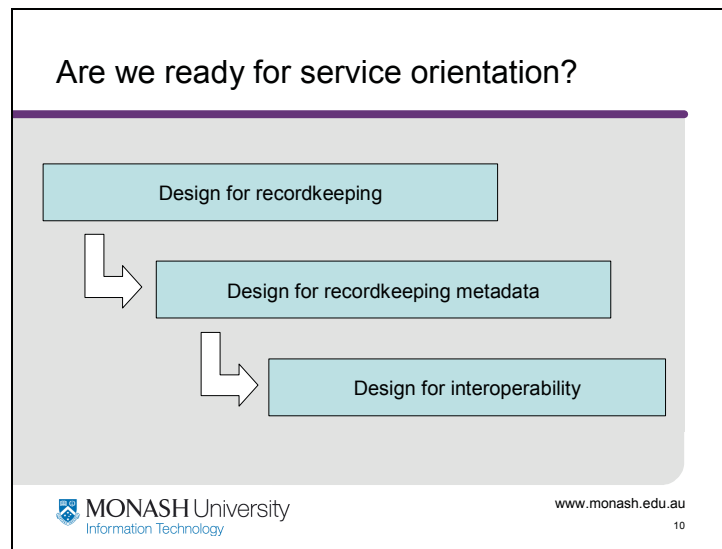
- So all emails go into the EDRMS?
  - No - only those where there is a recordkeeping requirement
  - that's where rule bases came into play to drive the deployment of metadata brokering in business processes
  - again the idea being as seamless as possible to the human user who is just getting on with their work.

Where we able to build this Broker and demonstrate its use in an appropriate scenario?
- No – for a number of reasons we couldn't, but reflecting on why this was the case was an important part of the action research process.

We ran into problems:-
- Firstly, with the robustness and rigour of the open source and web services technologies we were using, which curtailed what were able to achieve within the project timeframe and budget.
- Secondly, we uncovered some inherent interoperability constraints in records management and archival processes, technologies and tools developed for paper recordkeeping and for application-centric IT environments.

Are we ready for service orientation?

Design for recordkeeping

Design for recordkeeping metadata

Design for interoperability

MONASH University
Information Technology

www.monash.edu.au

10

Are we ready to realise the potential for automated recordkeeping metadata capture and re-use in service oriented environments?

The research undertaken within the Clever Recordkeeping Metadata Project suggests that the capacity of our existing recordkeeping metadata standards, processes and tools to capitalise on service orientation may be limited.

A key finding is the need to incorporate interoperability into the design of our standards, processes, tools and systems, not just expect to tack it on at the end.

This is in line with previous research and practical findings
- o Through the 80s-90s the need to design recordkeeping into digital processes and systems was identified,
- o which in turn led to understanding of the need to design in recordkeeping metadata
- o So the CRKM Project research is a further extension of this, finding that we need to consciously design for interoperability as well.

To participate in SOA …

- **Overcome paper thinking and dominance of paper paradigm**

Electronic Records, Paper Minds: The revolution in information management and archives in the post-custodial and post-modernist era

Terry Cook

MONASH University
Information Technology

www.monash.edu.au
11

What do we need to do to take advantage of the interoperability of service oriented environments?

- o confront some of the barriers to interoperability in our existing recordkeeping processes, standards and tools.

Firstly have to overcome dominance of paper paradigm:

- Vast majority of our recordkeeping tools and processes have been designed around handling and managing paper records.
- We've computerized them, and so transferred their inherent limited capacity for automated recordkeeping metadata capture and re-use into electronic environments.

In trying to develop scenarios in which to show the use of the metadata broker in records capture processes we were confronted by these limitations.

- We needed metadata about recordkeeping activities to be available in a form capable of driving recordkeeping processes and be inheritable by the records resulting from those processes.
- Raised the question of whether inputs to steps in the DIRKS process had more re-use potential than the outputs? e.g. recordkeeping requirements rather than disposal classes of records.

➢ We need to unleash the recordkeeping metadata in own processes and tools, and re-engineer them for interoperability. [continued]

- Not alone – with the emergence of SOA everyone is having to think about doing this.

- Lot to gain if we take on this challenge
  o Terry Cook – Electronic Records Paper Minds 1994 (still aspirational 14 years on)
  o 'reclaim our heritage (or birthright?) and become again central players in the world of both corporate memory and documentary heritage … [if] transform … from a physical and structure-centred mindset to one that is conceptual and process-centred.' (p. 305)

But it's also pretty tough as dealing with a hybrid world

- Evolving at different rates in different organisations, and within organisations some areas/activities becoming born digital faster than others

- Have to continue to support traditional (legacy) systems, at the same time as being able to support recordkeeping in the new ones.

- Tightrope – have to be on top of our own recordkeeping processes and their implications, and recognise when recordkeeping technologies become a hindrance rather than a help.

To participate in SOA …

- **Move beyond static resource discovery metadata models**

SPIRT Conceptual Models incorporated into ISO 23081:1

Mandates

Govern

Establish Competencies of

Business

Integrated in

Records Management business

Account for Execution of

Is documented in

Do

Do

People (Agents)

Create

Used by

Record, Manage, Enable use

Records

MONASH University
Information Technology

www.monash.edu.au
23

Secondly we need to move beyond static resource discovery models for recordkeeping metadata:

- Must continue to develop our understanding and refine our models so that they can adequately accommodate the complex, dynamic, recursive, multi-layered, multi-entity and relational nature of recordkeeping metadata.

Models in the ISO 23081 standard for recordkeeping metadata are examples of where have started to do this.

SPIRT Conceptual and Relationship Models of Records in Business and Socio-legal Contexts

- Outcome from the RKM Project – precursor to CRKM - Incorporated into part 1 of the ISO standard
- Codification of recordkeeping metadata in terms of the entities and relationships it applies to and encompasses.
- Tackles multiple entities involved in recordkeeping processes, multiple relationships and multiple aggregation layers
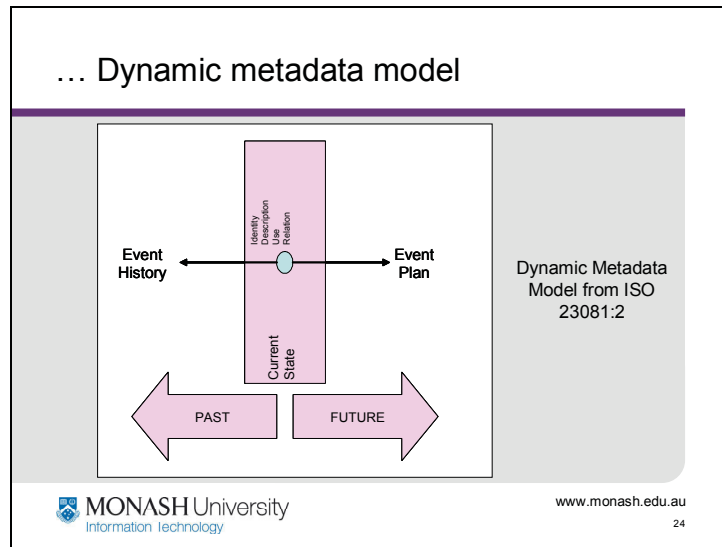
Dynamic metadata model from Part 2 of ISO 23081

- metadata about recordkeeping entities divided into IDENTITY, DESCRIPTION, USE and RELATION categories, reflecting current state of the entity.
- EVENT HISTORY capturing metadata about how got to current state
- EVENT PLAN capturing metadata about potential future actions, which when applied may change the current state and become part of EVENT HISTORY.

Useful start, but need to continue their evolution:-

- Encourage and learn from their implementation and feed that into the next cycle of their development
- Also use them to connect with other metadata communities – help them address recordkeeping requirements in their own standards.

To participate in SOA …

- **Recordkeeping metadata standards for machine rather than human processing in order to realise interoperability**
- **Standards development activities must:-**
  - be aware of standardisation activities in the IT sector in general, understand the forces driving their development, and be compatible with these activities,
  - address ontological and machine processability issues, and
  - seek an optimal balance between practical imperfections and theoretical models.

MONASH University
Information Technology

www.monash.edu.au
25

Thirdly, to participate in SOA, we need to extend these modelling activities to develop recordkeeping metadata standards more amenable to machine processing:

- Combination of standardising their machine encodings rather than just leaving it up to implementers,
- As well as tackling ambiguities and lack of precision in their semantics.

Findings from the project in relation to standards development activities in the recordkeeping sector:-

- Need to work in harmony with standardisation activities across the IT sector (e.g. can't ignore service orientation)
- Have to address their machine processability – need for formalized specification of their semantics, and through that process tackle the ambiguities
- But 'perfect is the enemy of the good' so have to strike the right balance between practical imperfections and theoretical models.
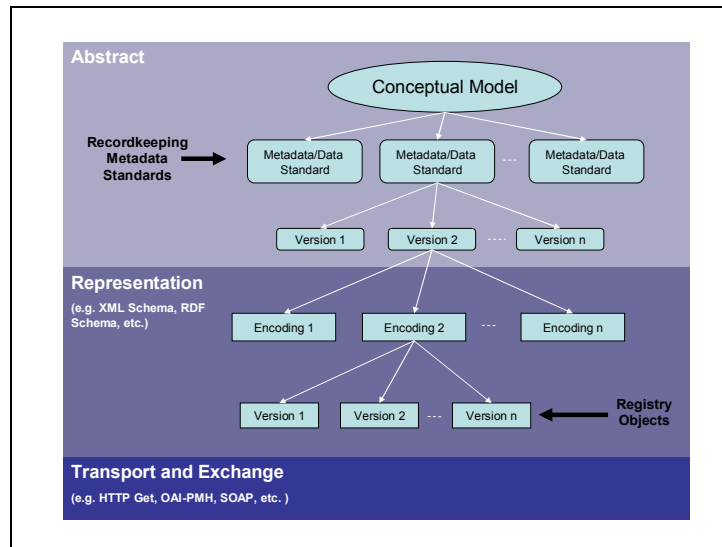
Illustrate this with the Layers of Interoperability Model used in the project:-

- Developed in 2002 by DELOS Working group on Metadata Registries to tease out metadata interoperability issues in the digital library community

- Simplified OSI model – (Open Systems Interconnection Basic Reference Model - abstract description for layered communications and computer network protocol design)

- At the bottom Transport and Exchange Layer
    - where the bytes get exchanged between machines
    - technical standards like http, OAI-PMH, SOAP etc. for machines to exchange messages and other stuff with one another

- At the top, Abstract Layer
    - Meaning of the stuff from human perspective

- In the middle, Representation Layer
    - Where represent the abstract in a machine processable form e.g. XML Schema, RDF Schema and other encoding languages

[CLICK] In the CRKM Project, developed this overlay:-

- Conceptual model underpinning the development of metadata/data standards, which are versioned through space and time [Continued]

- And can have multiple encodings, which in turn may also be versioned through space and time.

[CLICK] Our metadata broker requires representations of standards and the crosswalks between them in machine processable forms.

[CLICK] Problem is that majority of RKM standards pitched at the abstract level.
- Too much room for interpretation in their representation
- Allowable versioning in the abstract and representation layers means that have to develop lots and lots of encodings of schemas and crosswalks, which have limited application
- Flexibility comes at the expense of interoperability.

Some what perversely, getting better machine processable encodings means addressing ambiguities and lack of precision in semantics
- This requires more rigorous conceptual modelling along with better vertical integration through the layers.
- Model driven architectures –specify schemas and their mappings conceptually and automatically generate the layers – a long way off but have to start heading in this direction

In the meantime, have to have more rigorous tracking of relationships between standards, their versions and their encodings along with their relationships to other standards.
- Overlay provided the information model for the Registry component of the Broker
- The Registry metadata need to reflect all these layers in human and machine processable ways
- Conclusion that standardisation activities need to develop 'registry ready' products rather than expecting every instantiation to develop their own. Must be prepared to provide canonical encodings with appropriate descriptive metadata if want standards to deliver on interoperability

Standards and interoperability

- **Standards compliance does not guarantee interoperability**
  – Recordkeeping metadata standards are just a part of an infrastructure for interoperability

➢ **Balance between standardisation activities for best current practice versus standardisation activities to deliver better next generation practices**

MONASH University
Information Technology

www.monash.edu.au
11

- Through the CRKM Project came to understand that merely complying with a recordkeeping metadata standard is not going to realise interoperability.

- Service oriented architectures show that interoperability requires a number of interoperating standards, along with new kinds of infrastructure.

- These architectures don't spring up fully formed, they evolve through a mix of applying best practice standards and next generation models

  ⇨ Need to get the right balance between standardisation activities for best current practice,
  ⇨ Versus standardisation activities to deliver better next generation practices

- We have to learn how to operate in technological environments which are continually being defined, refined and redefined.

- We have to be prepared to lead and participate in recordkeeping innovation to meet digital and networking challenges.

Implementation Framework
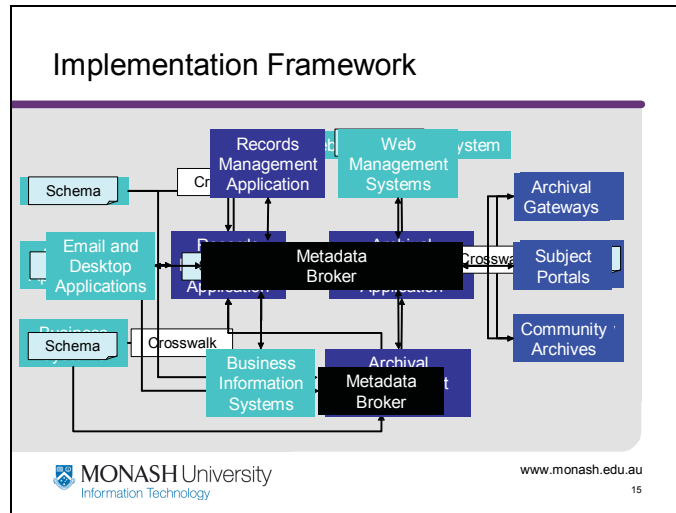
While this is all nice advice for future activities, how can findings from the Project aid in implementing recordkeeping metadata standards today?

- A tool like the Broker can be an agent of change.
- It could be used, in the first instance, to implement an existing recordkeeping metadata standard, from a compliance perspective, through managing the documentation of business and recordkeeping schemas that apply across an organisation.
- [CLICK] The Registry component of the Broker could document and store the versions of encodings of a recordkeeping metadata standard that apply in the various business, records management and archival control applications of an organisation, along with all the necessary crosswalks to and from these schemas.
- The explication of data and metadata schemas, and their mapping to the organisational recordkeeping metadata standard in the Broker, creates a framework in which compliance, and non-compliance, with the standard can be established and managed through time.
- [CLICK] A further benefit is that it could also be used to automate some movement of recordkeeping metadata. [continued]

- As business, records management and archival control applications develop Web services capabilities, then any previously hard-wired interoperability could be routed through the Broker instead.
- This has the potential to put control of relationships to the recordkeeping metadata schema into the hands of the recordkeeping function rather than the IT area.

Managing compliance with a recordkeeping metadata standard in such a way, may be a first step towards 'clever' recordkeeping metadata capture and re-use.

- Incorporating Web services into existing recordkeeping technologies is a useful start.
- It allows for a degree of practical interoperability in the short term, and acts as a spur to further system and process re-design.
- It also provides incentive for incremental development of the supporting infrastructure
- Realising new technical capabilities drives new conceptualisations, which in turn lead to the further development of technical capabilities, and so on.
- Through this ….

Innovate to service oriented future
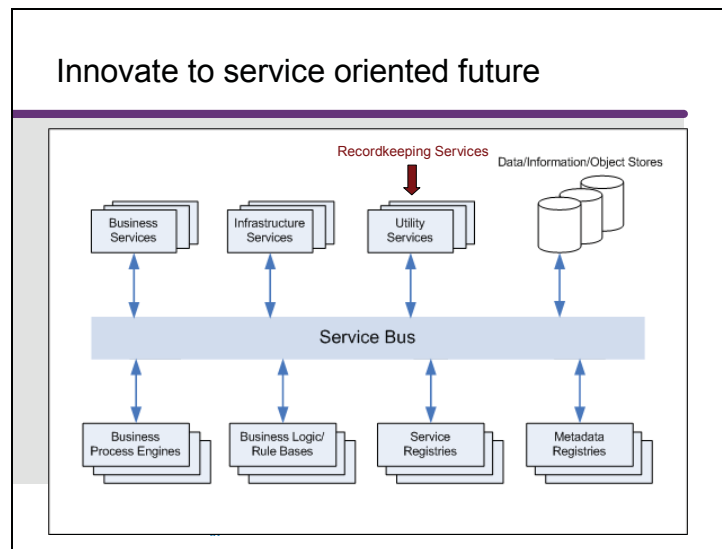
…. We can evolve through innovation to a service oriented future.

- Many challenges for research and practice in such a vision. It's no trivial exercise

- While digital and networking technologies threaten traditional practices, they also offer us the opportunity to undo some of the compromises forced upon us in the paper world. We may even by able to develop better recordkeeping and archiving systems, as we learn to take more and more advantage of digital capabilities for the use and re-use of information.

- **For more information see the Clever Recordkeeping Metadata Project Website**
- **http://www.infotech.monash.edu.au/research/groups/rcrg/crkm**

joanne.evans@infotech.monash.edu.au

Getting from the Wright Brothers in 1903 to the aviation industry in 2008 is only a straight line in hindsight. The reality is much messier - lots of research, discovery, dead-ends, innovation, successes, and failures

Think it is an exciting, but challenging, time to be a recordkeeping professional. Much to discover about what we need to know, and what we need to do, to meet the challenges of recordkeeping in the digital and networked era.

Personally hoping to be able to be involved in further research and development in this area – so that every digital and networking technology creates and keeps records as a matter of course.