# Making Metadata Matter: Outcomes from the Clever Recordkeeping Metadata Project

Joanne Evans, Sue McKemmish and Barbara Reed

**Dr Joanne Evans** was the PhD student attached to the Clever Recordkeeping Metadata Project. She graduated in October 2007 following the successful examination of her doctoral thesis, Building Capacities for Sustainable Recordkeeping Metadata Interoperability, and is currently undertaking further research into metadata and interoperability in eResearch and eHealth contexts as a Research Fellow at the eScholarship Research Centre at the University of Melbourne and at the Centre for Organisational and Social Informatics at Monash University.

**Professor Sue McKemmish,** PhD, is Chair of Archival Systems, Monash University, Director of the Monash University Centre for Organisational and Social Informatics, and Associate Dean Research of the Faculty of Information Technology. She is engaged in major research and standards initiatives relating to the use of metadata in records and archival systems, information resource discovery and smart information portals, Australian Indigenous archives, and the development of more inclusive archival educational programs to meet the needs of diverse communities. Sue McKemmish directs the postgraduate teaching programs in records and archives at Monash, has published extensively on recordkeeping in society, records continuum theory, recordkeeping metadata, and archival systems, and is a Laureate of the Australian Society of Archivists.

**Barbara Reed** is a practising consultant in the field of records, information and archives management as a director of Recordkeeping Innovation Pty Ltd, an Australian-based company delivering recordkeeping consulting and training services in Australia, and

increasingly in the Asian region. With a background in archives administration, long practice as a practitioner and consultant in the field, Barbara's interest is in transforming recordkeeping into the digital world and aligning recordkeeping with business. She has been a teacher and researcher at Monash University, written extensively on recordkeeping issues, and has been active in recordkeeping standards development work both locally and internationally.

*Across the information professions there is a recognition that meeting the challenges and exploiting the potentials of digital and networking technologies require a revolution in thinking and practice, in order to overcome barriers to interoperability in our existing systems. This is most apparent in attitudes to metadata, particularly recordkeeping metadata and its management. There is a need to shift from systems and tools dominated by the paper paradigm and mindset towards those that are 'born' digital. This article reports on the Clever Recordkeeping Metadata Project, which through prototyping of the 'create once – use many times' concept, tested the capabilities of existing processes, standards and tools to support metadata interoperability. It outlines key findings and outcomes from the project, particularly the need for recordkeeping to be reconceptualised in the light of the emerging service-oriented approach to the architecture of IT systems, and discusses the place of technological innovation in reforming our practices for a digital age.*

*This article is based on an Australian Society of Archivists 2008 conference paper, 'Discovery through Innovation: A report on outcomes from the Clever Recordkeeping Metadata Project'.*

## Introduction

> *We have 21st century digital technologies but automated 19th and 20th century industrial systems.*
>
> Laurance Millar, Deputy Commissioner and Government CIO, NZ State Services Commission[1]

Across the information professions – amongst information managers, geospatial experts, statisticians, librarians, data managers, IT managers, records managers, archivists, and others – there is a growing recognition that a revolution in thinking and practice is required to meet the challenges, and take full advantage of the opportunities, of digital and networking technologies. These technologies have the potential to enable us to share and re-use information in unprecedented ways, and in so doing, deliver massive productivity and capability gains. However this potential is proving slow to realise as the many barriers to information re-use designed into existing practices and systems become apparent. Recognising and overcoming these constraints, along with concurrently developing infrastructure that can support greater information portability and system interoperability are essential. Having reached the limits of the benefits from the automation of industrial systems, we now have to think digitally and networked from the start and foster system, tool and practice innovation.

Many communities of practice have come to realise the importance of metadata to achieving these ends. The information technology and information systems professions have progressed from simplistic definitions of metadata to a more mature understanding of its nature and vital role in the digital and networked age. Metadata is a key to enabling information objects to move through space and time in scalable and sustainable ways - but realising this requires cracking metadata's own scalability and sustainability issues. Hence there is a need to increase the capacity of our information systems to sustain quality metadata capture and enable its reliable re-use. Recordkeeping, library and information science professionals bring a richer, pre-digital world understanding of metadata which may be of benefit to other communities just beginning to identify and explore these issues. Yet at the same time our mindsets and mental models may no longer be relevant to metadata management in the twenty-first century, and so we have much to learn from others innovating in this area.

## Metadata – characteristics, divergences and convergences

> *The premise is that metadata is an essential, and the most important, component in advanced information systems engineering.*
>
> Keith Jeffreys[2]

The past decade has seen a maturing in the understanding of metadata and its importance to the retention, use and re-use of information in the digital and networked age amongst many communities of practice within the information professions. The simplistic 'data about data' definition has been replaced by more sophisticated ideas, in part due to the nature and role of data becoming more complex, as well as greater appreciation of the multifarious metadata needed to maintain data and enable its re-use across spatial and temporal boundaries.[3]

Metadata can be defined generally as structured data or information that describes an object in order to facilitate its understanding, management and use. The following statements outline some key characteristics of metadata that are increasingly applicable across metadata communities.

- Metadata is itself data. What we do to make data accessible, usable and re-usable, also needs to be applied to metadata.
- Metadata is recursive. As metadata is also data, there is always metadata about the metadata, about the metadata, and so.
- Metadata may be an intrinsic part of an information object (email headers are one such example), or extrinsic and external to it (metadata about provenance, related actions and people are examples).
- Metadata is dynamic. It accrues and changes as information objects move through space and time.
- Metadata is complex, as it represents complex, multiple entity realities and intricate webs of relationships. It also applies at various levels of granularity, aggregation and abstraction. This point can be illustrated by recent modelling work undertaken within the Dublin Core community which reveals these issues within a so-called 'simple' and 'flat' metadata standard.[4]
- Metadata has multiple purposes across different metadata communities (for example, resource discovery, recordkeeping, geospatial, digital rights, preservation, document management, data management) that bring different perspectives on information objects to the table.

## Recordkeeping metadata

Recordkeeping professionals are concerned with recordkeeping metadata – 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 and events;
- organisations, groups and individuals involved in records creation, management and use;
- recordkeeping functions, activities, transactions, processes and events;
- mandates, including laws, standards and business rules; and
- relationships amongst these recordkeeping entities.

Recordkeeping metadata identifies, authenticates, describes, manages and makes accessible, through time and space, records created in the context of social and business activity. 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.[5] The past decade has seen a number of research and practical initiatives undertaken to codify recordkeeping metadata requirements. These include a number of jurisdictional recordkeeping metadata standards, as well as the development of ISO 23081 as a metadata standard to support the ISO 15489 records management standard.[6]

Despite these activities, realising reliable, robust and comprehensive recordkeeping metadata management frameworks for electronic recordkeeping is proving elusive. Uptake of jurisdictional recordkeeping metadata standards has been limited, and when undertaken is often done minimally and with only partial coverage.[7] In addition there is still much to learn about the nature of recordkeeping metadata and how to effectively represent and manage its complexity. Working with other metadata communities may help us to tackle these issues. And in return we can bring our pre-digital understanding of managing the authenticity and integrity of information objects through time and space to bear on their metadata management requirements.

## Metadata challenges – from Wright brothers to aviation industry

One way of thinking about the metadata challenge is to envisage us as being on a journey from a Wright brothers model of metadata management towards fully articulated and matured frameworks. Table 1 summarises the move we need to make from systems and practices dominated by the paper paradigm and mindset towards those that are 'born' digital.

| From paper paradigms and mindsets ... | ... towards born digital systems, practices and recordkeeping tools |
|---|---|
| 'Paper' standards - many of our records management standards and instruments still predominantly work within a paper paradigm, for example, disposal authorities. | Digital standards - designed from a machine-processable rather than a human readable perspective. |
| Automated paper systems - although EDRMS have evolved that 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 nineteenth and early- to mid-twentieth centuries. | Digital recordkeeping processes and systems - recordkeeping processes that can be seamlessly integrated into business processes, incorporating automated recordkeeping metadata capture and re-use. |
| Unsustainable, unscaleable, expensive and resource intensive manual metadata creation and use processes. | Sustainable, scalable, automated, metadata creation, gathering, sharing and re-use processes. |
| Stand-alone systems and digital repositories, or hard-wired applications that achieve a degree of interoperability in particular implementation environments only. | Integrated systems and federated digital repositories. |
| Metadata standards and schemas that do not support interoperability - arguably none of the standards and schemas developed so far do. | Metadata interoperability - standards and schemas designed for interoperability. |

**Table 1**. From paper to digital paradigm

## Clever Recordkeeping Metadata (CRKM) Project

To address some of these challenges, recordkeeping researchers at Monash University together with practitioners from the National Archives of Australia (NAA), State Records Authority of New South Wales (SRNSW) and the Australian Society of Archivist's Committee on Descriptive Standards (ASA CDS) joined forces in an ARC Linkage research project from mid 2003 to 2006.[8] The Clever Recordkeeping Metadata (CRKM) Project aimed to investigate how standards-compliant metadata could be created once in particular application environments, then used many times to meet a range of business and recordkeeping purposes. The project grew out of the earlier ARC SPIRT Recordkeeping Metadata Project,[9] also in partnership with NAA and SRNSW, as well as the State Archives of Queensland and RMAA, and also sought to engage with a number of other initiatives.

In the CRKM Project, a systems development approach within an action research framework was adopted to address the research questions.[10] The idea was to use the building of a demonstrator of the 'create once, use many times' capability to develop understanding of the requirements which enable metadata to cross technical, spatial and temporal boundaries in automated ways. Building a demonstrator would test the proposition that metadata can indeed be created once and re-used many times, along with helping to identify what metadata is re-usable, as well as the tools, system configurations and architectures that facilitate re-use to enable us to move away from current resource intensive manual metadata attribution processes. In short, conceptualising and prototyping the environment and the tools to support automated metadata re-use would enable the 'create once, use many times' concept to be tested, limitations to be identified, and other insights to be gained.[11]

Through two iterations, the CRKM Project sought to build a prototype of a 'metadata broker', as a piece of infrastructure to enable metadata to be translated between schemas in an application independent fashion, and to demonstrate its use in a simulated real-world scenario. As befits the action research approach, each iteration involved systematic planning, action and observation, and reflection phases.[12] The first iteration was essentially a 'proof of concept', where a simple version of a metadata broker was conceptualised and built, in order to establish the feasibility of automated recordkeeping metadata capture and re-use, and identify sustainability and scalability challenges. This was followed by a second iteration where attempts were made to address issues raised in the first iteration in further prototyping.[13]

With digital and networking technologies challenging assumptions designed into traditional practices and tools, there is a clear need for recordkeeping research to foster innovation. The capabilities and limitations of paper technologies have shaped our current practices, so moving beyond their mere automation requires tools and infrastructure that exploit digital and networking capabilities. Digital recordkeeping and archiving practices can then be re-shaped around these new capacities. Pursuing the development of such tools in the context of a research project helps to mitigate risks as well as ensuring that the innovation can be well studied before moving on to widescale implementation. A feature of the CRKM Project's action research approach was reflective evaluation of the prototype from both researcher and practitioner perspectives.

## CRKM findings

A central finding of the CRKM Project is that the vision of recordkeeping metadata processes that create or gather metadata once, then share, use and re-use or re-purpose it 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
- integrated systems environments which enable data and metadata to be shared between many different kinds of systems and re-used or re-purposed as recordkeeping metadata.

The project also determined that neither precondition is yet in place – in either recordkeeping or broader systems environments. However the focus on innovation in the research design allowed the potential in emerging service-oriented approaches for the creation and configuration of information technology to be investigated. This led us towards web services and then onto the concept of service-oriented architectures (SOA).

## Service-oriented architectures

> *The initiative begins from the premise that what is wanted from IT systems is not 'monolithic applications' and 'information silos', but flexible 'services' ... that can be easily configured to provide improved capabilities and re-configured to meet changing needs.*
>
> JISC[14]

SOA is an emerging paradigm for constructing IT systems, in which applications and business processes are broken down into re-usable components, known as services, which can be brought together on demand to perform tasks. This delivers, not just interoperable data, but interoperable functionality. It may enable the realisation of the ability to dynamically configure technology into work and social processes - the holy grail of many an IT professional and business manager - rather than work and social processes being constrained by or built around the technology.

### *Web Services*

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. Web Services are a set of standards based around using XML as a neutral and standard way of representing structure in a machine-processable form, and Internet protocols for communication between services. Service interfaces can be defined using Web Services Definition Language (WSDL), messages can be exchanged between components using Simple Object Access Protocol (SOAP), and services can be dynamically discovered and located using the Universal Description, Discovery and Integration (UDDI) specification to provide databases or registries of service descriptions as illustrated in figure 1.[15] Web Services are the evolution of standards-based ways for distributed systems to interact with one another that have been around since the late-1980s. Delivery on a Web platform enables services to be implemented, and therefore be accessible, beyond traditional application and organisational boundaries. Ubiquity and pervasiveness are also potential flow-on effects.

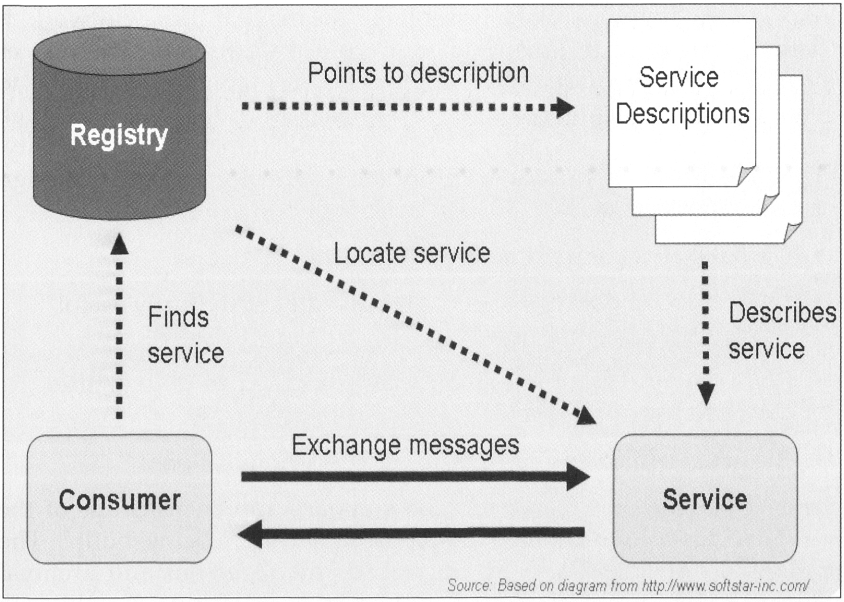


**Figure 1.** Web Services[16]

It is however important to note that Web Services and service-oriented architectures are not synonymous.[17] Web Services can be used to enable other types of architectures, and service-oriented architectures can be built using other technologies. Much of the current implementation interest in Web Services is based around their use in integration scenarios as a means to provide interfaces to legacy applications so that their data is more accessible. This allows for data in these applications to play a part in distributed computing environments. It delivers initial short-term interoperability, but will not of itself lead to the kind of adaptive and scalable systems that underpin service-oriented thinking. For the full benefits of the service-oriented approach to be realised, iterative deconstruction and componentisation of all aspects of the business and IT framework need to be progressively undertaken.[18]

Web Services in the SOA world are products in their own right. They are small chunks of programs that have been packaged to deliver specific and replicable outcomes, and 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. In order to do this, Web Services need to be designed to deliver small chunks of functionality. They need to be defined at appropriate levels of granularity. Design questions arise as to:

- What bit of action a service will perform?
- What layer should the functionality operate at to be really useful to multiple organisations?
- How should the functionality be packaged to deliver the required outcomes?
- What is replicable and re-usable across organisations?

The 'services rather than packaged software approach' is one of the foundations on which Web 2.0 environments are being built.[19] The potential impact of Web 2.0 on records management and archival practices includes the exploitation of social networking capabilities to organise and pluralise records. However the focus of the CRKM Project was on services that could support records creation and capture activities.

### *Services model*

How exactly does the services approach work? The classic services model, also known as the publish-find-bind triangle is illustrated in figure 1. In this diagram a consumer, usually a program, process, or service of some kind, has a need for a certain piece of functionality, and so looks for a service to perform the task in a service registry.[20] Service registries contain descriptions of services to enable their discovery, as well as providing details of how the consumer will need to interact with them. These service descriptions are registered by service providers and will also include information about rights to use, contractual obligations, and so on. The service registry can exist either inside or outside an organisational firewall. Because of security concerns, which have yet to be fully solved, often these registries are brought within the organisational boundaries, but with ever-growing uptake and maturity, truly ubiquitous registries on the Web are emerging.[21]

Once the consumer has located the desired service, and established the process of engagement, he or she can then automatically invoke and exchange messages with the service to accomplish the task. Describing these processes this way tends to downplay the fact that this is all happening at phenomenal, network transaction speed. The time lags between stages of the process are minimal - a bit like the instantaneous dispatch of an email message.

### *Service-oriented architectures*

With the notion of services being able to define and deliver specific functionality, service-oriented architectures provide a way of knitting them together. This is a big picture way of redefining how an organisation's IT systems will be developed and work together.

> *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.*
>
> Wikipedia[22]
>
> *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[23]

These definitions emphasise the business as the key driver, which represents information technologies serving the business rather than driving it. This is a trend observable in all organisations as technological capacity matures. The OASIS definition also emphasises the distributed capabilities and multiple ownership issues - services may not necessarily be owned by an organisation, to be invoked by their systems.

Figure 2 illustrates this new IT architecture. Business processes, governed by business logic or rule bases, call on business, infrastructure and utility services, with support from appropriate service and metadata registries, to drive their discovery and deployment. Utility services can be considered as general purpose functional units that, while not core business, contribute and are essential to getting the core business done. They therefore need to be capable of being deployed into any business process.

Infrastructure services are those services which are required to support the service-oriented architectural model and the IT infrastructure in general. A key example of infrastructure services is the functionality surrounding the registries in which service descriptions are published so that they can be discovered and utilised. Figure 2 also shows the trend in SOA towards shared data, information, or object repositories rather than each application maintaining their own. An extension of this is that SOA can accommodate legacy systems wrapped in Web Services as repositories, with their data available for re-use until resources are available for them to be replaced by a suite of services.

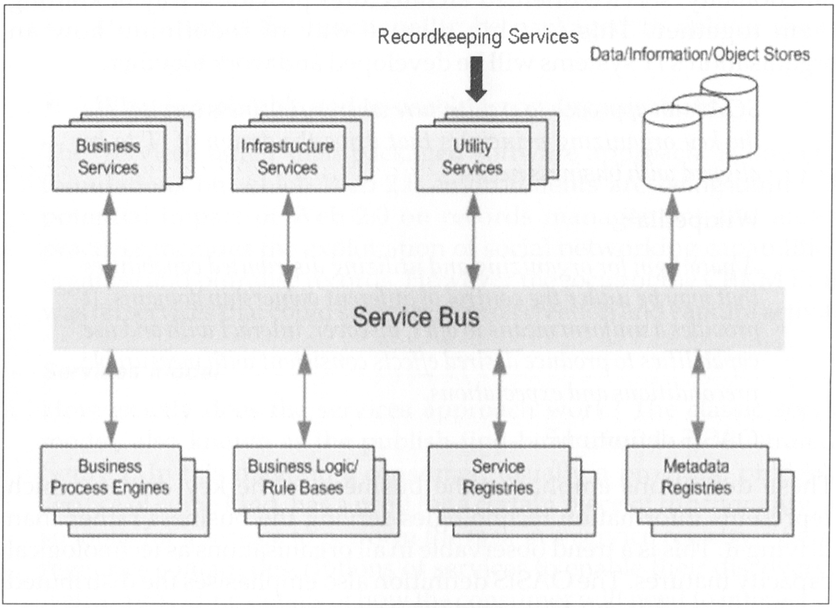


**Figure 2.** Service-Oriented Architectures[24]

### *Uptake of SOA*

Where is SOA at present? Broadly speaking it has been adopted 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 common processes that could be applied across agencies and lines of business. Australian Government Information Management Office (AGIMO) established the Australian Reference Model in June 2007, which is described as a lightly customised version of the US model.[25] There is value in consistency here as in other IT and business related worlds.

AGIMO's Reference Model aims to:

- provide a common language for agencies involved in the delivery of cross-agency services,
- support the identification of duplicate, re-usable and sharable services,
- provide a basis for the objective review of information and communication technologies (ICT) investment by government, and
- enable 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.[26]

How successful is the uptake? At this stage, it is an emerging framework. Many organisations say they endorse or are doing SOA. But in reality it involves 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 there is a need to be wary of assuming that the rhetoric is necessarily backed by implementation reality. As with any technological innovation it is the accompanying social and organisational changes which are the most challenging and have the most profound impact. Service orientation is an approach that applies to the design of technology, how work and business processes are configured, and the interaction between the two. Fuller articulation and understanding of the SOA model is therefore emerging as the architectures emerge.

The Gershon 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.[27] 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.

### *Implications for Recordkeeping*

In the CRKM Project, we found that Web Services and SOA have very exciting implications for recordkeeping.[28] Not only is interoperability a given, but so is metadata - particularly metadata about the business contexts in which recorded information is created and consumed. SOAs have the potential to deliver the integrated systems environments that allow for inheritance and re-use of recordkeeping metadata. Recordkeeping can leverage off this framework rather than having to carry the cost of building it for itself.

Web Services and SOA enable us to begin to think of delivering recordkeeping as service components which can be put together to form recordkeeping functionality. This takes us away from the monolithic structures of stand-alone EDRMS and introduces very powerful, but quite different ways, of approaching recordkeeping systems and functionality. In particular it opens up the real potential to design and drive recordkeeping from a business rather than a technology perspective.

This raises some interesting new issues. Just what would recordkeeping services look like in these environments? How might we define recordkeeping as small, packaged components that deliver reliable and replicable outcomes, and also enable us to build more complex outcomes? The National Archives and Records Administration (NARA) has done some pioneering work in this area,[29] but we need to think through the packaging and processes from an Australasian perspective to meet our particular recordkeeping cultures and needs.

In terms of our diagram of service-oriented architectures in figure 2, recordkeeping processes become key candidates for utility services, which could be orchestrated into business workflows, capturing evidence of business transactions as, and when, appropriate. Some examples of these recordkeeping services could include services to capture a record, ingest a record into a repository, retrieve a record, apply a disposition action, and so on. As illustrated in figure 3, recordkeeping itself also needs to be conceptualised from a service-oriented perspective. This involves defining recordkeeping services which are governed by policies captured

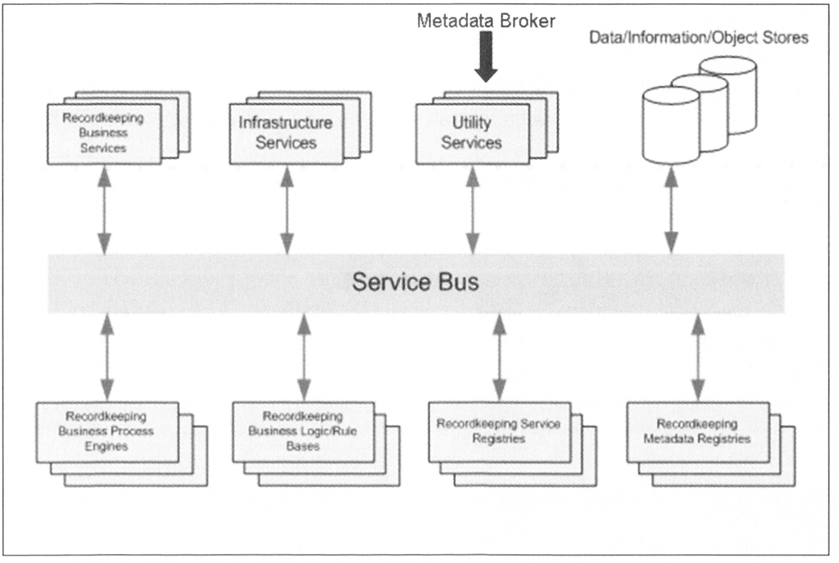


**Figure 3.** Recordkeeping in SOA

in recordkeeping business rule engines, and supported by recordkeeping service and metadata registries. Such services would be made up of various kinds of recordkeeping processes that serve both business and recordkeeping functions.

### *CRKM Metadata Broker*

In the CRKM Project we attempted to build a piece of recordkeeping services infrastructure – a metadata broker – as a utility service to translate metadata between standard schemas as indicated in figure 3.[30] The idea was that building such a broker, and applying it into a simulated scenario, would enable us to demonstrate how records description could be redefined as automated capture and re-use of metadata, and in so doing address some of the sustainability and scalability challenges for electronic recordkeeping. Through the process of innovation in conceptualising, constructing and attempting to demonstrate this new kind of recordkeeping tool, we discovered some interesting things about records, recordkeeping, recordkeeping metadata, recordkeeping practices and recordkeeping capabilities.[31]

### *Metadata broker conceptualisation*

The conceptualisation of the CRKM Metadata Broker, as illustrated in figure 4, was developed through two iterations of the prototyping process. The basis of the broker is a translation service which calls on a repository of schema and crosswalk objects. The repository is controlled by a registry in which metadata to facilitate the management and use of repository objects is held. Objects are deposited in the repository via registration processes which give them a unique identity and capture other descriptive metadata. Subsequent access to the objects and/or their metadata is via other kinds of registry services. These will include discovery services, which enable searching of the metadata to find a desired schema or crosswalk object, and retrieval services, which retrieve requested objects from the repository. A need for a validation service to allow for validation of the inputs and outputs of the translation service, along with a crosswalk compilation service to enable the dynamic creation of mappings between schemas was also identified.

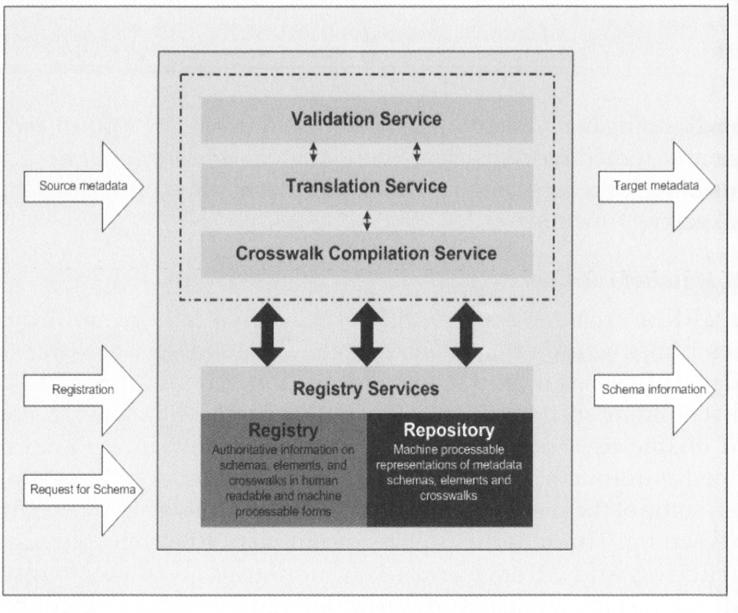


**Figure 4.** Conceptualisation of CRKM Metadata Broker as a cluster of services

### *Metadata broker usage scenario*

The use of the CRKM Metadata Broker can be illustrated with a simple scenario - moving an email into a records repository - as shown in figure 5. In this scenario metadata needs to be translated from the Australian Government Email Metadata Standard (AGEMS)[32] to the Recordkeeping Metadata Standard for Commonwealth Agencies (RKMSCA).[33] In this case, as detailed in the AGEMS standard,[34] we want to re-use:

- Sender, receiver, other addresses as **agent** metadata,
- Security classification as **rights management** metadata,
- Message ID, subject as **identification** metadata,
- Keywords as **subject** metadata,
- Dates as **history** metadata, and
- Reply to, references, attachments as **relationship** metadata.

The process with the need for the translation communicates with the broker, invoking the translation service component, with a request to translate metadata from the source to the target schema. The translation service then searches for a crosswalk to undertake this translation by contacting the registry service of the broker. The registry contains metadata about the mappings between schemas (that is, crosswalks) which includes details of their source and target schema, how they are encoded, where they are located, and so on. The translation service then receives and processes the metadata about the crosswalk, establishing the requirements to invoke the crosswalk and carry out the desired translation. As we were attempting to generate an open system, the aim was for the broker to support different instantiations of mappings between schemas in machine-processable forms. These crosswalks can themselves come wrapped up as a service so they can be invoked without having to install particular applications, as well as hiding their complexity from the translation service and the client. Such loose coupling would also enable their internal workings to be modified as required without undue impact. Finally the translation service sends the source metadata to the crosswalk for translation, with the target metadata instance sent back to the client.

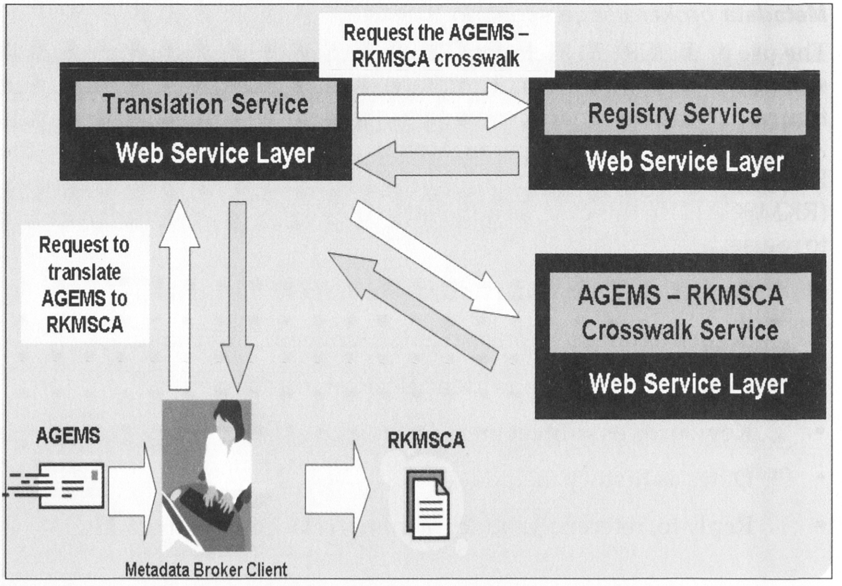


**Figure 5.** CRKM Metadata Broker usage scenario

So what? Doesn't existing functionality allow for this?

- Many EDRMSs are ODMA[35] compliant which allow for direct registration and re-use of data and 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. In the design of the broker we wanted to overcome this kind of technical constraint, due to its impact on sustainability and scalability. We need to look towards ubiquitous solutions, which have the potential to embed recordkeeping into all electronic environments.

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

- No – this action would be configured into the business process so that it would happen seamlessly from the (human) user perspective as emails are sent.

Would that then mean that all emails go into the EDRMS?

- No - only those where there is a recordkeeping requirement. That is where a recordkeeping business rules base and engine would come into play and drive the deployment of metadata brokering in business processes. Again the idea is for this to be as seamless as possible from the perspective of human users who just want to get on with their job.

## Readiness for SOA

Are we ready to realise the potential for automated recordkeeping metadata capture and re-use in service-oriented environments? The research undertaken within the CRKM Project suggests that the capacity of existing recordkeeping metadata standards, processes and tools to capitalise on service orientation is limited. A key finding from the project 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 aligns with previous research findings for recordkeeping in electronic environments, where the need to design recordkeeping into processes and systems was identified.[36] This has in turn led to better understandings of the need to include recordkeeping metadata requirements as part of the design process, and the need for recordkeeping metadata standards to specify these requirements. A further extension to this need, arising from the CRKM Project, is the imperative to design recordkeeping standards, processes and systems for interoperability, so that we can take better advantage of the capabilities of digital and networking technologies.

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

To accomplish these interoperability aims, existing barriers to interoperability in our existing recordkeeping processes, standards and tools will need to be overcome. The first major challenge is to move beyond the dominance of paper paradigm in our thoughts and actions.[37] Recordkeeping tools and processes have been designed around handling paper records, with the vast majority of digital systems initially developed to automate the handling and managing of paper records. Hence their inherent limited capacity for automated recordkeeping metadata capture and re-use has been transferred into electronic environments. Unleashing the recordkeeping metadata in recordkeeping processes, systems and tools will involve redesigning them around interoperability.

We are not alone in this endeavour - with the emergence of SOA, practitioners across all information professions are starting to think about how to reap the benefits from the flexibility and adaptability of the service-oriented approach. Iterative reconceptualising of processes is required in order to identify the most useful kinds of functional units that make for self-contained and coherent services capable of being assembled into complex and dynamic structures, 'with built-in potential for integration and evolution'.[38] Some of these components may already exist within applications, whereas others may be entirely new types of components.[39]

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

Secondly, moving beyond static resource discovery models for recordkeeping metadata is essential. This involves refinement of our models so that they can adequately accommodate the complex, dynamic, recursive, multi-layered, multi-entity and relational nature of recordkeeping metadata. The SPIRT Conceptual Models[40] and the Entity and Metadata Models of ISO 23081[41] are promising starts. However a more rigorous approach to their formalisation needs to be developed, tested and refined through instantiation and application in real-world environments. Only then can they play their intended roles as framework models from which implementation schemas are derived. There is a need for better understanding of the ways to represent dynamic, multi-entity, recordkeeping relationships in terms of schema structure and behaviour, and of what the flattening of models to create schemas should, and should not, entail. Standardisation activities also need to be informed by greater understandings of technologies for interoperability, of requirements for interoperability in electronic recordkeeping, and of the interplay between the two.

### *Develop standards for processing by machines rather than humans*

Thirdly, as part of these modelling activities, recordkeeping metadata standards more amenable to machine processing are needed. This requires standardisation of their machine encodings rather than just leaving it up to implementers, which in turn requires ambiguities and lack of precision in their semantics to be resolved. The importance of such approaches is illustrated in the Layers of Interoperability model

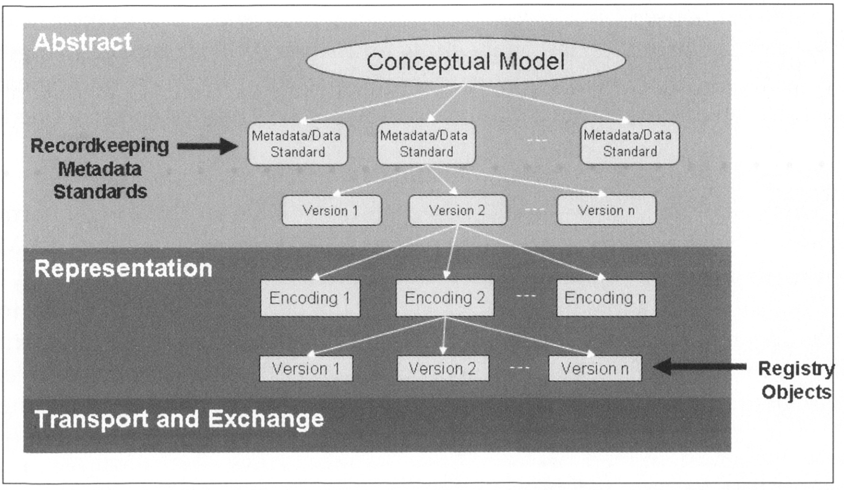


**Figure 6.** Extended Layers of Interoperability Model

developed in 2001–02 by the DELOS Network of Excellence on Digital Libraries, Working Group on Metadata Registries.[42] As shown in figure 5, most recordkeeping metadata standards exist in the abstract layer whereas the automated processing of metadata requires machine-processable representations.[43] Recordkeeping metadata standards developers need to provide canonical encodings of their schemas to reduce the potential for proliferation in the representation layer which is detrimental to interoperability. At the same time, more rigorous conceptual modelling is required to address inconsistencies, loose terminology and semantic imprecision in existing standards in order to develop schemas more amenable to machine processing.

### *Standards for compliance versus interoperability*

Merely requiring conceptual compliance 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 do not spring up fully formed. They evolve – through a mix of applying best practice standards and next generation models. It is critical to strike the right balance between standardisation activities for best current practice and standardisation activities to deliver better next generation practices. Recordkeeping innovation will be needed to meet the digital and networking challenges of operating in technological environments which are continually being defined, refined and redefined.

### *Implementing recordkeeping metadata standards*

Much of what has been discussed so far relates to the longer term transformation of recordkeeping practice. Can findings from the CRKM Project aid in implementing recordkeeping metadata standards today? We believe that a tool like the CRKM Metadata Broker has the potential to be an agent of change. In the first instance it could be used to implement a recordkeeping metadata standard, from a compliance perspective, through managing the documentation of business and recordkeeping schemas that apply across an organisation. 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.

A further benefit is that it could also be used to automate some movement of recordkeeping metadata. As business, records management and archival control applications develop Web Services capabilities, then any previously hardwired 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 redesign. It also provides incentive for incremental development of the supporting infrastructure. Realising new technical capabilities will drive new conceptualisations, which in turn will lead to the further development of technical capabilities, and so on. By taking this approach, we can evolve, through innovation, to a service-oriented future.

There are however many challenges for research and practice in such a vision. In realising recordkeeping as a suite of services we have much to learn about records and recordkeeping. In undoing some of the compromises forced upon us by the paper paradigm, we also may have the opportunity to develop better recordkeeping and archiving systems. Vital to realising this vision will not only be the development of the technical capabilities, but also better understanding of social and organisational dimensions and their interactions. Understanding where and how recordkeeping might utilise other IT and information management services and infrastructure, and where it might develop its own so that in turn it can be used by others, will also be vital.

## Conclusion

There is a tendency to take for granted the evolution of form, process, custom and law that has occurred over hundreds of years to give paper-based records systems their evidentiary capabilities and stable form. For recordkeeping in electronic environments the same kind of evolution has only just begun, and, as in the paper world, the technology itself will shape the possibilities. We hope that there is the opportunity for further collaborative projects, in the mould of the Clever Recordkeeping Metadata Project, to foster innovation that enables us to discover what we need to know, and what we need to do, to move us from Wright brothers to mature and robust digital recordkeeping systems and infrastructure.

### Endnotes

[1] Statement made as keynote speaker to the Queensland State Archives, National Information Management Skills Summit, available at <*http://www.archives.qld.gov.au/im_summit.asp*> accessed 7 February 2009.

[2] Keith G Jeffery, 'Metadata: The Future of Information Systems', in Sjaak Brinkkemper, Eva Lindencrona and Arne Sølvberg (eds), *Information Systems Engineering: State of the Art and Research Themes*, paper from Information Systems Engineering Symposium (ISES–2000), Stockholm, Sweden, 5–6 June 2000, on the occasion of the retirement of Professor Janis Bubenko jr, Springer-Verlag, London, 2000, also available at <*http://www.wmo.ch/pages/prog/www/WDM/ET-IDM/Doc-2-3.html*> accessed 8 September 2008.

[3] One such example comes from the e-Science community where a 2007 US National Science Foundation report on cyberinfrastructure has the following definitions of data and metadata:

> Data are any and all complex data entities from observations, experiments, simulations, models, and higher order assemblies, along with the associated documentation needed to describe and interpret the data.
>
> Metadata are a subset of data, and are data about data. Metadata summarize data content, context, structure, interrelationships, and provenance (information on history and origins). They add relevance and purpose to data, and enable the identification of similar data in different data collections.

National Science Foundation, *Cyberinfrastructure Vision for 21st Century Discovery*, March 2007, available at *<http://www.nsf.gov/od/oci/CI_Vision_March07.pdf>* accessed 12 August 2008.

[4] Andy Powell, Mikael Nilsson, Ambjörn Naeve, Pete Johnston and Thomas Baker, 'DCMI Abstract Model', Dublin Core Metadata Initiative, 4 June 2007, available at *<http://dublincore.org/documents/abstract-model/>* accessed 26 March 2008.

[5] David Wallace, 'Archiving Metadata Forum: Report from the Recordkeeping Metadata Working Meeting, June 2000', *Archival Science*, vol. 1, no. 3, 2001, pp. 253–69.

[6] A discussion of these initiatives can be found in Joanne Evans, 'Evaluating the recordkeeping capabilities of metadata schemas', *Archives and Manuscripts*, vol. 35, no. 2, November 2007, pp. 56–84.

[7] The Auditor-General, *Recordkeeping including the Management of Electronic Records*, Audit Report No. 6 2006–07, Australian National Audit Office, 12 October 2006, available at *<http://www.anao.gov.au/download.cfm?item_id=4FA72EF31560A6E8AAB9AFCAAB3EABB6&binary_id=37F0C8021560A6E8AA539CFB3D3BDC82>* accessed 19 February 2009.

[8] *Create Once, Use Many Times – The Clever Use of Metadata in eGovernment and eBusiness Recordkeeping Processes in Networked Environments,* Australian Resource Council (ARC) Linkage Grant in conjunction with Monash University, National Archives of Australia, State Records Authority of New South Wales and the Australian Society of Archivist's Committee on Descriptive Standards, Chief Investigator (CI) Professor Sue McKemmish, Partner Investigators (PI) Professor Anne Gilliland-Swetland and Adrian Cunningham.

[9] *Archival metadata standards for managing and accessing information resources in networked environments over time for government, social and cultural purposes,* Chief Investigators Sue McKemmish and Anne Pederson, Principal Investigator Steve Stuckey, funded by an Australian Research Council SPIRT Grant and industry partners State Records NSW, National Archives of Australia, Queensland State Archives, Australian Council of Archives and the Records Management Association of Australia (RMAA), 1998, available at <*http://www.sims.monash.edu.au/research/rcrg/research/spirt/about.html*> accessed 1 June 2003.

[10] Joanne Evans, Barbara Reed and Sue McKemmish, 'Interoperable Data: Sustainable Frameworks for Creating and Managing Recordkeeping Metadata', *Records Management Journal,* vol. 18, no. 2, 2008, pp. 115–29.

[11] Frada Burstein and Shirley Gregor, 'The Systems Development or Engineering Approach to Research in Information Systems: An Action Research Perspective', in Proceedings of the 10th Australasian Conference on Information Systems, Wellington New Zealand 1999, 1–3 December 1999, available at <*http://www.vuw.ac.nz/acis99/Papers/PaperBursteinGregor-173.pdf*> accessed 28 October 2003 (which can now be found at <*http://arrow.monash.edu.au/hdl/1959.1/41512*> accessed 19 May 2009). For the application of systems development as a research method in archival science see Joanne Evans and Nadav Rouche, 'Utilizing Systems Development Methods in Archival Systems Research: Building a Metadata Schema Registry', *Archival Science,* vol. 4, nos 3–4, December 2004, pp. 315–34.

[12] Valsa Koshy, *Action Research for Improving Practice,* London, Paul Chapman Publishing, 2005, and Jean McNiff and Jack Whitehead, *Action Research Principles and Practice,* RoutledgeFalmer, London, 2002, 163 pp.

[13] More information about the first and second iterations can be found in the research reports that have been made available on the CRKM Project website <*http://www.infotech.monash.edu.au/research/groups/rcrg/crkm/index.html*>.

[14] JISC, 'New animation makes sense of the e-Framework', 7 February 2007, available at <*http://www.jisc.ac.uk/news/stories/2007/02/news_animation.aspx*> accessed 13 February 2009.

[15] Venu Vasudevan, 'A Web Services Primer', O'Reilly Network, 4 April 2001, <*http://webservices.xml.com/pub/a/ws/2001/04/04/webservices/index.html*> accessed 23 March 2006. The design and development of Web services specifications is mainly being undertaken by the W3C. More information about the scope of the activity can be found on the W3C Web Services Activity website at <*http://www.w3.org/2002/ws/*> accessed 29 March 2006. The UDDI specification comes under the umbrella of OASIS, see <*http://www.uddi.org/*> accessed 20 October 2003 (which can now be found at <*http://uddi.xml.org/*> accessed 14 May 2009).

[16] David M Rubin, 'Introduction to Web Services', SoftStar Research, 6 August 2002, available at <*http://www.softstar-inc.com/Methodology/Softstar Web Services Presentation.ppt*> accessed October 2005.

[17] Kurt Viklund, *Service Oriented Architecture (SOA) Explained,* TietoEnator, 2004, <*http://www.tietoenator.com/binary.asp?GUID=C4AD1567-7905-48DA-958E-5C2CB092C507*> accessed 28 March 2006.

[18] L Cherbakov, G Galambos, R Harishankar, S Kalyana and G Rackham, 'Impact of Service Orientation at the Business Level', *IBM Systems Journal,* vol. 44, no. 4, 2005, pp. 653-90.

[19] O'Reilly, Tim (2005), 'What is Web 2.0 Design Patterns and Business Models for the Next Generation of Software', O'Reilly Media Inc. 30 September 2005, available at <*http://www.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html*> accessed 24 March 2008.

[20] In the services model, the term 'consumer' refers to a service requester. Depending on granularity, a service requester may be another service, business process or application that has a requirement to carry out particular functionality that is to be provided by a service.

[21] Web Services listings on sites such as SourceForge <*http://www.sourceforge.net*> are one example.

[22] Wikipedia, 'Service-oriented architectures', available at <*http://en.wikipedia.org/wiki/Service-oriented_architecture*> accessed 11 February 2009.

[23] OASIS, *SOA reference model, Reference Architecture for Service Oriented Architecture Version 1.0,* Public Review Draft 1, 23 April 2008, available at <*http://docs.oasis-open.org/soa-rm/soa-ra/v1.0/soa-ra-pr-01.pdf*> accessed October 2008.

[24] Adapted from 'Figure 5 Typical SOA Application Architecture' of Viklund, p. 14.

[25] Australian Government Information Management Office (AGIMO), *Australian Government Architecture Reference Models Version 1.0,* Commonwealth of Australia, June 2007, 196 pp., <*http://www.finance.gov.au/publications/australian-government-architecture-reference-models/docs/AGA_Reference_Models_Version_1.0.pdf*> accessed 8 February 2009.

[26] ibid.

[27] Sir Peter Gershon, *Review of the Australian Government's Use of Information and Communication Technology,* Department of Finance and Deregulation, Australian Government Information Management Office, Commonwealth of Australia, August 2008, available at <*http://www.finance.gov.au/publications/ICT-Review/docs/Review-of-the-Australian-Governments-Use-of-Information-and-Communication-Technology.pdf*> accessed 11 February 2009.

[28] Barbara Reed, 'Service Oriented Architectures and Recordkeeping', *Records Management Journal,* vol. 18, no. 1, 2008, pp. 7–20.

[29] NARA Records Management Service Components Program, *Functional Requirements and Attributes for Records Management Services,* National Archives and Records Administration, 7 December 2005, 115 pp., available at <*http://www.archives.gov/era/pdf/Functional-Requirements-and-Attributes-for-Dec07-2005.pdf*> accessed 27 March 2006.

[30] Joanne Evans, Barbara Reed and Sue McKemmish, 'Interoperable Data: Sustainable Frameworks for Creating and Managing Recordkeeping Metadata', *Records Management Bulletin,* vol. 18, no. 2, 2008, pp. 115–29.

[31] ibid.

[32] Chief Information Officer Committee, *Australian Government Email Metadata Standard AGEMS Version 1.0,* National Archives of Australia, December 2005, available at <*http://www.naa.gov.au/images/email_metadata_standard_tcm2-911.pdf*> accessed 14 August 2008.

[33] National Archives of Australia, *Recordkeeping Metadata Standard for Commonwealth Agencies Version 1.0,* Canberra, May 1999, 136 pp., <*http://www.naa.gov.au/Images/rkms_pt1_2_tcm2-1036.pdf*> accessed 14 August 2008. Note that this was the Commonwealth standard at the time of research. It was superseded in 2008 by the *Australian Government Recordkeeping Metadata Standard* (AGRkMS).

[34] See appendix 1 of *Australian Government Email Metadata Standard AGEMS Version 1.0.*

[35] For more information about ODMA (Open Document Management API) available at <*http://odma.info/*> accessed 14 August 2008.

[36] This idea was articulated in the early-1990s by David Bearman, amongst others, and subsequently developed in research projects such as the University of Pittsburgh's Functional Requirements for Evidence in Recordkeeping from 1993 to 1996 (see <*http://www.archimuse.com/papers/nhprc/*>) and the first InterPARES project from 1999 to 2001 (see *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project,* September 2002, available at <*http://www.interpares.org/book/index.htm*>, accessed 2 October 2003).

[37] Terry Cook, 'Electronic Records, Paper Minds: The Revolution in Information Management and Archives in the Post-Custodial and Post-Modern Era', *Archives and Manuscripts,* vol. 22, no. 2, November 1994, pp. 300–329.

[38] Viklund, p. 9.

[39] As an example, for a discussion of how business rules lend themselves to being useful services see James Taylor, 'Achieving Decision Consistency Across the SOA-based Enterprise', in *Web Information Systems Engineering – WISE*

*2005*, Proceedings of the 6th International Conference on Web Information Systems Engineering, New York, NY, USA, 20–22 November 2005, Berlin, Springer-Verlag, 2005, pp. 750–61, also available at <*http://dx.doi.org/10.1007/11581062_82*> accessed 13 December 2005.

[40] The *Conceptual and Relationship Models: Records in Business and Socio-Legal Contexts* have been published online as a deliverable of the SPIRT Recordkeeping Metadata Project at <*http://www.sims.monash.edu.au/research/rcrg/research/spirt/deliverables.html#conceptual*> accessed 13 July 2003. They are discussed in Sue McKemmish, Glenda Acland, Nigel Ward and Barbara Reed, 'Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema', *Archivaria*, vol. 48, Fall 1999, pp. 3-43.

[41] International Organization for Standardization, *ISO 23081-1:2006 Information and documentation – Records management processes – Metadata for records – Part 1: Principles*, 2006, 20 pp. and *ISO/TS 23081-2:2007 Information and documentation -Records management processes – Metadata for records – Part 2: Conceptual and implementation issues*, 2007, 34 pp.

[42] Thomas Baker, Christophe Blanchi, Dan Brickley, Erik Duval, Rachel Heery, Pete Johnston, Leonid Kalinichenko, Heike Neuroth and Shigeo Sugimoto, *Principles of Metadata Registries*, a white paper of the DELOS Working Group on Registries, DELOS Network of Excellence on Digital Libraries, 2002, available at <*http://delos-noe.iei.pi.cnr.it/activities/standardizationforum/Registries.pdf*> accessed 26 May 2004.

[43] Joanne Evans, Sue McKemmish and Karuna Bhoday, 'Create Once, Use Many Times: The Clever Use of Recordkeeping Metadata for Multiple Archival Purposes', *Archival Science*, vol. 5, no. 1, March 2005, pp. 17–42.