



Feedback on Digital Economy Future Directions Consultation Paper



Prepared by
Hitachi Data Systems
for the
Department of Broadband, Communications and the Digital
Economy

20 February 2009

COPYWRITE

All copyright in this paper is owned by Hitachi Data Systems and its licensors and no copying, adaptation or other use of any part of the paper is allowed without Hitachi Data Systems' prior written consent.

DISCLAIMER

Hitachi Data Systems has made every effort to ensure that the material contained in the paper is accurate and up-to-date as at the date of publication; however independent skill and judgement should be exercised when relying on the content of this paper. Further advice should be sought with respect to individual circumstances.

CONTACT DETAILS

For any queries relating to this document please contact:

Paul Bicknell
Principal Business Consultant

HITACHI DATA SYSTEMS

Level 12, St George Centre
60 Marcus Clarke Street
Canberra ACT 2601

Desk: + 61 2 6240 5502
Fax: + 61 2 6240 5599
Mobile: 0421141891
paul.bicknell@hds.com

CONTENTS

Introduction	4
C.1 Open Access to Public Sector Information	6
C.2 Digital Confidence	15
C.3 Developing Australia's Knowledge and Skills Base	17
C.4 Ensuring that Australia's Regulatory Framework Enables the Digital Economy	17
C.5 Digital Economy and the Environment	21
C.6 Measuring the Digital Economy and its Impacts	24
Appendix 1 - Legal and Regulatory Requirements: Data Storage	26

Introduction

Hitachi Data Systems (Hitachi) welcomes the opportunity to provide feedback to the Digital Economy Future Directions Consultation Paper, released by the Department of Broadband, Communications and the Digital Economy.

While Hitachi is known for its capabilities as an ICT vendor, Hitachi also maintains a formal business consulting arm with a cadre of senior, multi-disciplinary (CxO-level) consultants that have established a strong reputation for independent, 'above the line' (vendor-neutral) client engagements and trusted advice to the Australian Government.¹

It has been argued that the technologies that lie at the core of cyberinfrastructure can be separated into five general categories:²

- **High-performance Computing Resources.** Supercomputers and clusters of computers integrated to provide supercomputer-like performance to individual applications.
- **Cyberinfrastructure Applications and Tools.** General cyberinfrastructure applications and tools that support research but are not specific to a particular discipline (such as software for simulation, visualisation, job scheduling, data mining and statistical analysis).
- **Data Storage and Management Resources.** Large-scale data storage systems for real-time use and for archival purposes, as well as facilities, software, and procedures for periodic backup of data sets.
- **Advanced Network Infrastructure Resources.** High-performance on-site networks and their connections to off-site high-performance networks that support such capabilities as massive data transfers to and from clusters, real-time visualisation, and use of remote instrumentation.
- **Resources for Collaboration within Virtual Communities.** Facilities and support for teleconferencing, off-site collaboration and for operating remotely located instrumentation and related devices; support for identity management and associated middleware in collaborative research activities.

While Hitachi has expertise across a number of these cyberinfrastructure areas, we have chosen to concentrate our input where we can provide authoritative commentary - the data storage domain. This submission will therefore focus on Hitachi's area of subject matter expertise - the provision of data storage and management resources in support of the Digital Economy.³

The Digital Economy is founded on data, the use of data, and its carriage, protection, accessibility and efficient disposal. Whereas paper-based records are comparatively easy to control and manage through their lifecycle, digital data is far more transferable, prone to 'leakage' and harder to control.

¹ For example, Hitachi Data Systems was one of a select group of vendors invited to contribute to the Independent Review of the Australian Government's use of ICT, lead by Sir Peter Gershon. Our submission leant heavily on the Hitachi's data centre economics methodology, and on a range of government and industry benchmarking. Both our core assumptions and our methodology were independently validated. This submission earned Hitachi an invitation to bilateral discussions with Sir Peter Gershon; and a request for us to prepare supplementary modelling. The data centre component of the Gershon Review, as released, is almost entirely based upon the Hitachi contribution.

² "Higher Education IT and Cyberinfrastructure: Integrating Technologies for Scholarship," ECAR Research Study, vol. 3 (2008).

³ The discussion paper defines the 'digital economy' as the global network of economic and social activities enabled by ICT, particularly the internet. It includes commercial transactions, personal dialogue, information, entertainment and delivery of services. The term is often used interchangeably with 'internet economy' and 'information society'.

Exponential data growth is a function of doing business in the digital age. The quantity of public sector information (PSI) is increasing at an annual rate of 40-60 percent. While data is often perceived as 'intangible', this data must ultimately be stored on physical devices. The data explosion is changing the dynamic within the IT environment since the variety and complexity of data sources and the potential for using this data is increasing the need for high performance data storage.

Data that resides within the server for a split-second may need to remain accessible within the storage platform for decades because of its operational or business importance, or the need to comply with legal obligations. Poor storage performance can slow or deny access to data. These storage devices require management, maintenance, energy, space, security and sensitive disposal; often on a massive scale.⁴ Hitachi leads the industry in providing solutions to these issues.

Some organisations struggle with their custodial responsibilities for their data. There is a growing demand for services that ensure data is properly documented, placed into a well-known and well-defined format, and preserved over suitable periods of time. New technologies such as 'cloud computing' can provide innovative possibilities for data storage, but also create new financial, geographic and jurisdictional issues. Australia may need to establish policies around how PSI should be managed, located and protected in this virtual environment.

⁴ The strategic importance of data storage to the Digital Economy is underscored by Gartner research indicating that "on average, storage represents 26% of data centre Total Cost of Ownership (TCO) and 48% of a (client's) hardware budget".

C.1 Open Access to Public Sector Information

Unified communications is automating and combining many forms of human and device communications into a common user experience. Gains in efficiency can result through an optimisation of business processes, reducing latency, managing flows, and eliminating device and media dependencies. The simultaneous movement towards highly capable mobile devices, software encryption, remote access technologies, wireless networking and fixed-mobile convergence is also enhancing workforce mobility. These trends create innovative opportunities for effective use of networked ICT to gain access to computing capability, services and information.

Such connectivity also spawns new ICT implications and requirements such as the need for secure network gateways, filters and working storage. It also creates increased demand for access to information.

Critical to the success of any initiative to open up Public Sector Information (PSI)⁵ to use and reuse is that it be simple and cost effective to find, access and use. Creating the infrastructure, applications and processes to support PSI access has the potential to be expensive, time consuming and a distraction from core-services for government and public sector organisations. However, a policy that aims to make PSI available, without providing the means for organisations to make this data public, will likely lead to a confusing concoction of stand alone systems.

Many users want the ability to store data on network-accessible servers and to know that these bits are automatically and frequently backed up to one or more geographically disparate sites; they want confidence that their research will survive a disaster. They want storage as a highly reliable service, with a service operator checking the backups, arranging crash and disaster-recovery mechanisms, and handling the necessary periodic migrations from obsolescent hardware to newer technology.

Exponential Growth in Public Sector Information

The amount of data generated by public sector organisations continues to grow at an alarming rate. The amount of data created, distributed, and stored across the Australian Federal Government is growing at a compound annual rate of around 60 percent. If this compound annual growth rate is maintained, then the quantity of data to be stored by the Federal Government alone will increase over the next five years, from a very conservative estimate of about 25 Petabytes to around 250 Petabytes.

Generic drivers of data growth include such things as: DNA, Radio Frequency Identification (RFID) technologies, digital recording and shared entertainment content, BLOG sites and email. Not long ago, a 20 Megabyte mailbox was considered large. With Microsoft Exchange 2007, most organisations are now deploying environments where every user has a 2 Gigabyte mailbox.

⁵ The Consultation Paper describes PSI as including: Government produced data (such as Australian Bureau of Statistics (ABS) and geospatial data) and copyright protected materials (such as reports and other documentation). It can also include materials that result from publicly-funded cultural, educational and scientific activities.

Other major drivers for the growth in PSI include:

- CCTV.** Close Circuit Television (CCTV) has seen a rapid rise in digital content and could be a ‘sleeping giant’ of public sector data growth. CCTV camera solutions have traditionally involved analogue cameras stored on local tape. However migration to digital cameras and more sophisticated analytics combined with increased retention periods now mean that digital content management is now the biggest cost in CCTV surveillance systems. The table below shows typical CCTV storage averages across various industry sectors⁶.

Industry	Amount of DVS Storage Needed
City government	70TB
Big box retail (per store)	12TB
Sea port	70TB
Major train station	300TB
Large airport	280TB
Toll booth	60TB
Midrange casino	150TB
Prison detention center	50TB
K-12 school	20TB
Pharmaceutical company	45TB
Manufacturing company	15TB
Oil refinery	440TB

- Healthcare.** Australian healthcare organisations plan to leverage ICT to enable aggregation of existing stored information, and provide access to accurate and up-to-date patient data to health care professionals when they need it, and wherever they are located. However, these eHealth initiatives will also have a massive impact on the amount of data generated and maintained. Although the actual electronic patient medical record might be a relatively small document, the attached diagnostics and other information can greatly increase data size. File types and sizes in the Healthcare industry are on a par with the broadcast and media industry, since radiology, cardiology, mammography and pathology all generate large amounts of unstructured and fixed content data.
- Secondary School Education.** The Federal Government National Secondary Schools Computer Fund (NSSCF) will provide funds to deploy additional computers to schools.⁷ Similar ‘Smart Schools’ initiatives will provide students with digital classrooms equipped with large interactive computing displays and interactive whiteboards. Video conferencing and multimedia learning resources will enable distance learning for students and schools. Students will be able to access information and resources online, such as Video on Demand (VOD), podcast or streaming video. Students will also store information and their assignment

⁶ The general rule of thumb is about 10GB of digital content per day per CCTV camera. 100 cameras would therefore require ~1TB per day. The Industry trend is to retain video for at least 30 days. Hitachi calculations prepared recently for the Department of Parliamentary Services (DPS), agree with this order of magnitude.

⁷ Over 2500 additional PCs are expected to be rolled out in the ACT alone by June 2009.

work as a digital portfolio, enabling teachers to monitor students' work as they progress through their schooling.

- **Higher Education and eResearch.** The higher education sector and eResearch environment is undergoing transformational change. The Commonwealth Education Budget is currently \$11 billion - an annual increase of about \$5B. In turn, the Government expects efficiency dividends from the academic sector; including consolidation activities to improve economies of scale, and initiatives to make publicly-funded research data more accessible.
- The traditional divergence between researchers and central administrative /IT providers is being challenged:
 - Many research areas previously used grant funding to 'do their own thing' – snubbing centralised IT providers in the process. The central players, in turn, have had a poor track record for delivering on Service Level Agreements (SLA).
 - However, rampant data growth and increasing compliance and convergence complexity, such as the requirement to publish source data with research and to guarantee secure storage of unreproducible data, is now forcing individual researchers to hand over significant quantities of data to central IT. Initiatives are also in place to recover the wealth of electronic data previously stored on diskette, CD, DVD, and personal hard drives.
 - Central IT in turn needs to step up to the mark and become much more professional and capable, and to investment in more robust storage infrastructure.
- Collaborative and consolidation activities include major cross-campus and cross-university initiatives stemming from the national academic fibre network. There are also a number of significant international collaboration endeavours that will generate vast quantities of data. For example, Australia has been short listed in a major international collaborative bid to host one of the largest and most ambitious international science projects ever devised. The Square Kilometre Array (SKA)⁸ will be over fifty times more powerful than the world's largest existing telescope, and is expected to produce about an Exabyte⁹ of data per day.¹⁰

⁸ The principal collaborators in the Australian SKA project are the Australian Government, Western Australian Government and the CSIRO. The SKA bid builds upon the \$111 million Australian SKA Pathfinder project.

⁹ The prefix 'exa' means one billion billion. In decimal terms, an Exabyte is a billion gigabytes.

¹⁰ The digital data streams for each beam formed at an aperture array station in the SKA must be sent to the central processor for correlation. For the mid-frequency array, the total data rate per station is 16 Terabytes per second, and for the higher frequencies each dish generates 0.1 Terabytes per second. The total data rate for the SKA is approximately 4 Petabytes per second and involves distances of thousands of kilometres. For comparison, today's global internet traffic volume is estimated at 15 Terabytes per second, and is expected to reach about one Petabyte per second by 2015.

- A recent survey of over 250 academic and research institutions was asked to predict how the overall importance of each cyberinfrastructure technology would change over the next three years.¹¹ The greatest percentage of institutions identified data storage and management as most important. Significantly, more than three-quarters of respondents at research institutions said they anticipated a “great increase” in importance.

	RESEARCH INSTITUTIONS		TEACHING INSTITUTIONS	
		Percentage of Institutions		Percentage of Institutions
Importance to Research	Data Storage and Management	92.8%	Data Storage and Management	35.5%
	Advanced Network Infrastructure	88.4%	Collaboration within Virtual Communities	25.8%
	Cyber-infrastructure Applications and Tools	87.0%	Advanced Network Infrastructure	23.9%
	Collaboration within Virtual Communities	85.0%	Cyber-infrastructure Applications and Tools	18.4%
	High-Performance Computing	82.6%	High-Performance Computing	16.9%
Importance to Teaching		Percentage of Institutions		Percentage of Institutions
	Data Storage and Management	80.6%	Data Storage and Management	51.8%
	Collaboration within Virtual Communities	75.7%	Collaboration within Virtual Communities	45.5%
	Advanced Network Infrastructure	71.3%	High-Performance Computing	31.5%
	Cyber-infrastructure Applications and Tools	59.0%	Cyber-infrastructure Applications and Tools	31.5%
	High-Performance Computing	48.1%	Advanced Network Infrastructure	31.1%

¹¹ Source: “Higher Education IT and Cyber-infrastructure: Integrating Technologies for Scholarship,” ECAR Research Study, vol. 3 (2008).

Current Shared Data Storage Repositories

There are a number of robust mechanisms currently available for the shared storage and extraction of PSI.

- **Institutional Repositories.** Many institutional organisations maintain secure and accessible repositories and search engines for the long-term preservation of digital resources created by members of their institutional community. For example, the Australian National University (ANU) 'Demetrius' repository, powered by DSpace repository software,¹² stores two kinds of digital items (or objects):

Digitised items produced as surrogates for materials in analogue format (such as printed books, manuscripts, museum artefacts and video tapes).

'Born Digital' items originally produced in machine-readable form (such as e-books, scientific databases, sensor data, digital photographs and websites).

The digital items stored in Demetrius are organised into 'collections', which are in turn organised into 'communities' (or collection groupings).¹³ The ANU can thematically organise these digital collections and determine access policies and use.

- **Public Libraries.** Libraries have long had the mission of acquiring appropriate teaching and research materials, organising this material and making it accessible. These materials take a range of formats; and increasingly include digital collections. However, they have generally not included data collections other than those that have been organised and published as coherent collections (such as statistical data collections).

While many libraries are gradually reclaiming the collection management role for digital articles, they are only able to take on the small to medium size data collections. Large collections require more technical expertise and hardware capacity than is generally available within library environments. The scale and complexity of the infrastructure and operation necessary to preserve core electronic scholarly literature and other PSI exceeds that which can be supported by any individual university, library or institutional budget.

- **Other Repositories.** There are currently a number of other options currently for maintaining PSI, each with their own characteristics and issues:
 - **Publishers.** Some publishers are relied upon to maintain digital copies of published material. However, entities responsible for research data repositories need to be enduring, in the same manner as other collecting organisations such as art galleries, archives, libraries and museums. Hence, it is a somewhat risky strategy to rely on the longevity of a commercial entity such as publishers.

¹² The DSpace repository platform is a very popular and widely adopted repository within the higher education and research sectors. One of the reasons for the popularity of DSpace is that it is relatively easy to install and maintain, and has a ready made user-interface that integrates data management and access functions within the system's architecture. A strong international developer community has evolved to support DSpace and new features are being added constantly.

¹³ In DSpace terms, the fundamental unit of organisation is an Item; for example, a digital image, PDF file, data set, and so on. Items are grouped into Collections.

- **National Library.** A 'dark archive' of publishers' content could be held at the National Library of Australia (NLA).¹⁴ However, the concern associated with maintaining important content at National Libraries is the risk of actual or perceived interference in content management policies. Political separation is therefore important.
- **Outsourcing Collection Management to 3rd Party.** Another option is to outsource specific publication management to the publisher's own data bases, or other third party organisation (such as subscriptions to Athica Harbours or Portico) for an annual fee. However, while this approach does have some practical advantages, there are still a number of issues such as the need for local interfaces, the risks associated with proprietary format migration, and the risk to the collection were the provider to go out of business
- **LOCKSS.** LOCKSS (**L**ots **O**f **C**opies **K**ee**P** **S**tuff **S**afe™) is a cooperative international effort between publishers and libraries.¹⁵ The project was designed for scholarly journals by Stanford University. It supports an open source system allowing libraries to collect, preserve and provide access to material published on the Web. The system attempts to replicate the way libraries do this for material published on paper.

Cloud Computing

Many Australian businesses and individuals have realised the value proposition and embrace the benefits that the internet and other Digital Economy technologies enable. One of the most important recent developments in the Digital Economy is the emergence of 'cloud computing' as a successful business model.

Cloud computing suppliers are developing a number of models for delivery of infrastructure and application services in a transparently priced and highly agile manner. Cloud computing allows businesses to use application services without having to buy, maintain or support their own IT infrastructure. Cloud computing currently has several variations:

- **Software as a Service.** Software as a Service (SaaS) delivers a single application through the browser to thousands of customers using a multi-tenant architecture. On the customer side, it means no upfront investment in servers or software licensing; on the provider side, with just one application to maintain, costs are low compared to conventional hosting.
- **Utility Computing.** The idea of Utility Computing is not new, but is getting new life from Amazon.com, Sun, IBM, and others who now offer storage and virtual servers that can be accessed on demand. Early enterprise adopters mainly use utility computing for supplemental, non-mission-critical needs, but it has the potential to replace parts of the data centre, enabling users to stitch together memory, I/O, storage, and computational capacity as a virtualised resource pool available over the network.

¹⁴ The term 'dark archive' pertains to content that is only made available when there is a triggering event. Access to archived content is granted when specific conditions are met that cause titles to no longer be available from the publisher or any other source. Specific 'trigger events' may include:

- A publisher stops operations, or
- A publisher ceases to publish a title, or
- A publisher no longer offers back issues, or
- Upon catastrophic and sustained failure of a publisher's delivery platform.

¹⁵ ANU is the designated Australasian Host Server for LOCKSS.

- **Web Services in the Cloud.** Closely related to SaaS, Web Service providers offer application programming interfaces (API) that enable developers to exploit functionality over the Internet, rather than delivering full-blown applications. They range from providers offering discrete business services to the full range of API, such as Google Maps.
- **Platform as a Service.** Another SaaS variation, this form of cloud computing delivers development environments as a service. Users build their own applications that run on the provider's infrastructure and are delivered to other users via the Internet from the provider's servers.
- **Managed Service Providers (MSP).** One of the oldest forms of cloud computing, a managed service is essentially an application exposed to IT management rather than to end-users, such as a virus scanning service for e-mail or an application monitoring service. Managed security services fall into this category, as do such cloud-based anti-spam services and some desktop management services.
- **Service Commerce Platforms.** A hybrid of SaaS and MSP, this cloud computing service offers a service hub that users interact with. Most common in trading environments, such as expense management systems that allow users to order travel or secretarial services from a common platform that then coordinates the service delivery and pricing within the specifications set by the user.
- **Internet Integration.** Integration of cloud-based services is in its early days.

In-house infrastructure and applications will remain important, but the demand for cloud services is growing rapidly, particularly in a Digital Economy facing skills shortages, increasing power costs and ever more quickly changing markets.

While large enterprises will benefit from cloud computing capabilities, the relative benefits to Small to Medium Enterprises (SME) are prodigious. In contrast to large enterprises, SME rarely have sufficient capital, in-house skills or real-estate to establish and maintain substantial IT infrastructure. Cloud computing can obviate the need for large-scale capital investment and allow SME to gear more cheaply and effectively than if they have to establish their own infrastructure and systems from day one.

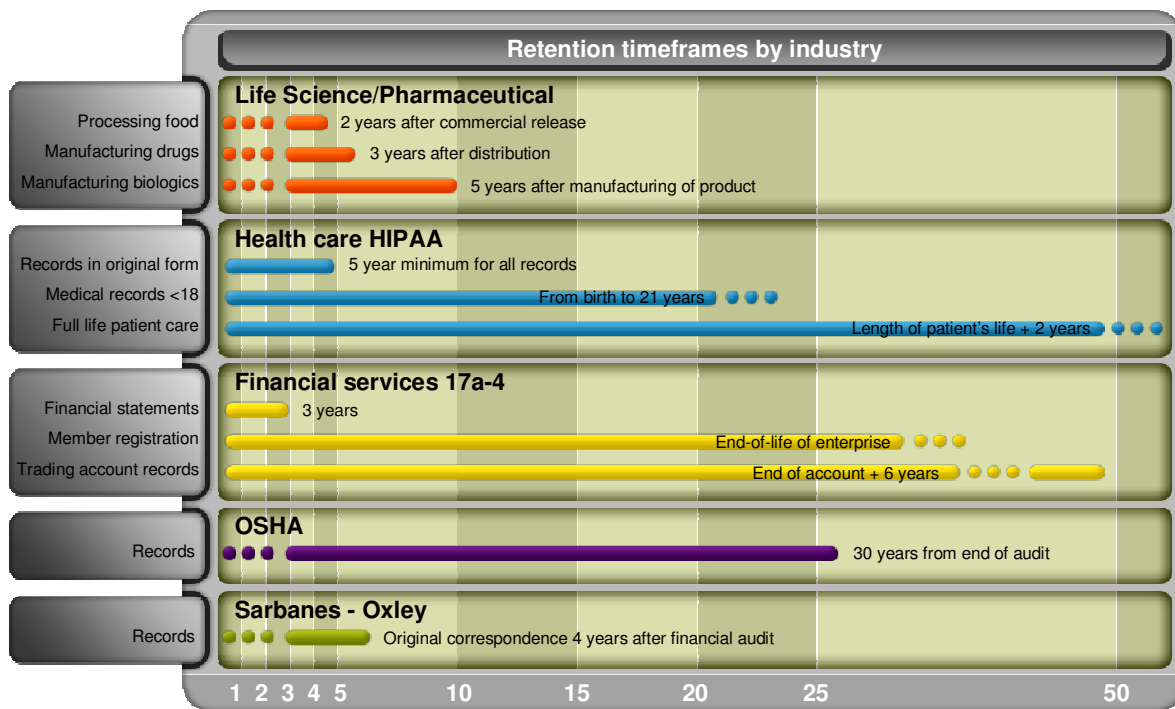
Cloud computing provides the ability to access data storage and other applications (such as email, office applications, CRM and even payroll and ERP) in the 'cloud', thus removing the need to acquire these applications or the infrastructure and skills to support them. It also provides organisations with the flexibility to deal with peaks in demand without the expense of under-utilised systems during off-peak periods.

Government should therefore encourage Australian SME to leverage cloud services for competitive benefit. It is equally important that the Government creates an environment where cloud services suppliers are encouraged to see Australia as an efficient and competitive base in which to establish cloud services infrastructure. In the absence of such an environment, Australian SME and the general public will still use cloud services but Australia will gain little of the infrastructure benefit.

Preservation and Curation of Public Sector Information

One of the enduring problems with storing information in digital format is ensuring continuing access to electronic resources. It is important that data is stored and made available in formats that are easily read (by human or machine as appropriate) and in a manner which facilitates re-use of the data in other applications and services (such as Web2.0 services). These formats need to be maintained in a readable format over time. As data-formats evolve and new formats emerge, it must be possible and cost-effective to transfer data into new formats securely and cost-effectively. Adherence to appropriate standards will help preserve access to digital information, however the technological standards themselves are also evolving rapidly.

Legislative and compliance obligations often mean that PSI must generally be maintained in excess of seven years. The forces of functional improvement, obsolescence, carrier failure and market imperatives mean that data storage systems rapidly become obsolete. The planning life of storage technology is generally about five years. Hence, most regimes will need the capacity to seamlessly refresh the data storage technology within the 'life' of the data. In essence, this means that most organisations require the capability to retain their data indefinitely.



The requirement to keep old software and hardware, operating systems and manuals, and to maintain relevant skills, generally makes preservation of obsolete technology an impractical strategy.¹⁶

¹⁶ Some organisations do maintain older equipment and software for the rare occasions when they need to access this data. The National Library of Australia maintains some retrospective capability with a wide collection of tape – tape players (including spare parts, gear ratios and cogs). The ANU is also considering building a functional 'museum' of devices (eg one of every known tape drive)

Typical preservation methodologies for addressing technological obsolescence include:

- **Migration.** Migration is a preservation methodology which involves transitioning content from one file format to another as technology evolves and file formats become obsolete. Digital information is migrated from one hardware/software configuration to another or from one generation of computer technology to a later one.
- **Emulation.** Emulation involves strategies to make future technology mimic technology of earlier generations.
- **Byte Preservation.** Byte Preservation involves simple storage of an unmodified stream of digital bytes without special provision for future display or functionality.

A central policy problem with data curation and preservation is that the cost of these activities often persists long after the project that generated the data (and any associated funding) has ended. Storage systems architecture, software and hardware must therefore be designed to allow for cost-effective system upgrade or replacement and easy data migration without risk to the content. For example, metadata is created and entered manually in the majority of current repository systems. This is unsustainable in the long term and leads to non-standard implementations. There is a need to integrate existing metadata extraction tools with repository software so that the systems ingest data and create metadata in a way that is largely transparent to the user.

Government should therefore establish, and enforce, an Australian metadata standard.

C.2 Digital Confidence

Many organisations have invested extensively in local infrastructure such as identity management and authentication/authorisation systems. These now often allow the very secure and flexible protection and sharing of local resources within defined communities. However, collaborative environments and virtual organisations cross organisational and national boundaries and often different sectors (industry, academia, government). Effective collaboration and resource sharing therefore requires inter-institutional trust (with associated business processes and regulatory compliance).

In parallel with these basic connectivity requirements comes increasing requests for access to richer forms of data content, including graphics-based imagery, video, presence-based applications and other collaboration resources. This all equates to large quantities of data that needs to be identified, located, authenticated, recovered and shared. Users of the Digital Economy need confidence in the veracity of this digital content.

Building Digital Confidence through Effective Content Archiving

Identity management and other authentication/authorisation systems are important to achieving digital confidence between organisations and recipients of public sector information. It is also critical that the digital content management regime is able to ensure confidence in the actual digital content.

It is ironic that, even when possible, migration of old data to newer formats and mediums is often not considered worth the effort involved. Many researchers are still more inclined to recreate data rather than re-use it, particularly if there is any uncertainty over the ownership or reliability of the data; or of the methodology behind the data collection. This uncertainty is also a function of the manner in which the data is collected and stored, and the veracity of the associated metadata.

If use and reuse of PSI is to be encouraged data owners need to:

- Ensure that the data has integrity.
- Ensure that the data is retained for the appropriate period.
- Make sure the data is not deleted, corrupted or lost.
- Optimise what is stored to avoid lots of duplicate copies of the same thing.
- Encrypt the data so that it can't be accessed illegally.
- Find the files that have been stored.
- Dispose of the data when it is no longer required.

Government should therefore work with Industry to establish common protocols and guidelines for content management and archive. These guidelines need to address:

- **Open Interoperability.** Many records management systems take a proprietary view of 'their' archive. However, open interoperability requires the ability to interface with all existing and future systems using standard file formats (such as XML and HTML) for 'future proofing'.
- **Single Archive Name Space.** All objects should be stored in a single, archive-wide global name space, using a well understood paradigm, and be open and easy to navigate with standard tools and applications
- **Search and Indexing.** It should be easy to find the files that are stored in a simple, sensible manner. This function needs to be independent of the creating application – as it may no longer exist. Similarly, a stand-alone storage system for each content management system prevents organisations from seamlessly searching and leveraging all content, no matter which application generates it, or to centrally manage this data.
- **WORM, Immutability and Retention.** The archive needs a 'Write Once, Read Many (WORM) file system where content is stored in an immutable format with the ability to set file-level retention. It also needs to ensure that other users, systems and applications cannot intentionally or inadvertently change data.
- **Scalability.** Need to be able to store billions of files and Petabytes¹⁷ of data; access this data quickly, and manage it effectively. Conventional data storage technology cannot meet the scalability requirements to manage storage growth with clusters of appliances sitting in front of static storage systems.
- **Protection.** Need a self-configuring and self-healing capability with automated policy enforcement, failover and ongoing integrity checks. Must be able to ensure that a specified number of replica copies are maintained to tolerate simultaneous points of failure.
- **Retention.** Must be able to prevent file deletion before the business/records retention policy retention period expires. These policies can be set explicitly or inherited
- **Authentication.** Policy-based object management must guarantee that archived data is authentic, available and secure. It must guard against corruption or tampering and ensure confidence in the integrity of the data.
- **Shredding.** When an item is deliberately destroyed, there must be no trace of the file recoverable from disk after deletion. Disks can be stolen, removed or equipment simply decommissioned – but the data needs to be comprehensively removed.



¹⁷ 1 Petabyte =1,000 Terabytes

The capability to address the above attributes exists now, but most organisations are deferring investment because data compliance is not being enforced. However, the cost of addressing these content management issues retrospectively will be much more difficult and considerably more expensive.

The Government should actively enforce legislative and regulatory compliance obligations for the content management and archiving of public sector information.

C.3 Developing Australia’s Knowledge and Skills Base

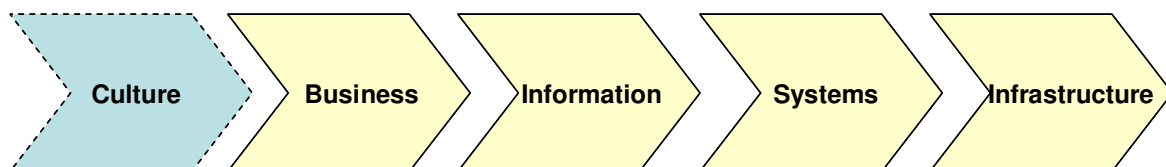
A key challenge of maintaining the cyber-infrastructure to support the Digital Economy will be the design and staffing of the organisations to:

- help the public access cyber-infrastructure services;
- assist organisations in managing their PSI —including observational data, the construction of research and reference collections, or data from analysis or simulation; and preparing this data for handoff to the appropriate data repositories and curators at the appropriate time;
- organise data for reuse, mining, and mashups.
- assist in the setup of virtual organisations and also to help with their breakdown.

Existing computing organisations and libraries can no doubt provide some expertise, but these resources need to be supplemented with expertise in disciplinary data, standards, and tools; and perhaps also with more capability for consulting on software, data, and information design.

C.4 Ensuring that Australia’s Regulatory Framework Enables the Digital Economy

Rather than focusing on cyber-infrastructure alone, it is essential to ensure that business strategy drives ICT infrastructure investment. It is necessary to understand (working backwards)the systems that this infrastructure supports, which in turn requires an understanding of the information needs of the organisation, and also a sound understanding of underlying business needs. In addition, it is often necessary to understand some tenets of the prevailing culture, since this culture informs core values, behaviour and governance structures and therefore the practicality of potential solutions for the Digital Economy.



Given that Business should drive IT (rather than the other way around) and the regulatory and legislative framework is a reflection of business needs, then the technical infrastructure should facilitate this legislative requirement. Therefore, how to we enable the regulatory framework so that it doesn’t impinge on the Digital Economy?

It is important that Government guidelines and policy for making PSI available are clear and simple to implement. Convolutioned policies will lead to long decision-making cycles and may prevent some

information being made available at all. Current legislative and compliance policy and guidelines are summarised at Appendix 1.

Government should also promote shared infrastructure and interfaces to provide storage for and access to PSI; and make it far simpler for organisations to make data available. By centralising the storage and distribution of PSI, it becomes easier to create and enforce standards for formats, lifecycles and information retention and protection. Centralised storage also allows the implementation of a single set of processes to protect PSI and ensure that it remains in readable formats in perpetuity.

Legal and Regulatory Compliance Culture

Data Ownership

Data and intellectual property ownership and security is an important consideration for the Digital Economy. However, Hitachi consultants have found that often the question “who owns the data” has not been seriously considered.

While administrative areas are relatively clear cut, in that Business Process Owners generally own their data (the HR Director owns HR data, and so on); in other areas there can be markedly differing views as to data ownership. There is also an implicit (but somewhat unstable) understanding that research data arising from a project belongs to those undertaking the research and that data is their responsibility.

The question is nevertheless important since the answer often determines the classification and value of the data to the organisation and the corresponding requirements for data storage. For example, within the same institution, Hitachi consultants found two organisations with completely different attitudes to data ownership:

- One organisation believed that it owned its data and that this data was very valuable and required to be accessible 24/7. The organisation therefore invested heavily in developing a robust and responsive data storage infrastructure, with reciprocal Disaster Recovery (DR) provisions via a sister institution.
- Conversely, another organisation considered its data as generally belonging to its members, but stored by the organisation on an ‘all care, no responsibility’ basis. While this organisation did backup its data, their business continuity plan focused more on the ability to restore the data storage service rather than the data itself. It instead adopted a pragmatic risk-management approach towards data storage, believing that the cost-benefit equation for more sophisticated solutions (such as automatic failover / UPS) did not warrant greater investment.

Both organisations chose a logical storage infrastructure **given their paradigm** as to ownership, value and custodial responsibility for their data.

Compliance Apathy

Many organisations are only marginally aware of the extent of legal and regulatory compliance obligations in relation to data management and retention. Some organisations adopt an 'ostrich mentality' claiming that their legal and regulatory compliance obligations are unclear and un-enforced so they will delay necessary investment.

Few organisations maintain formal policies pertaining to the storage, backup and security of data and how long their data must be kept. In general there is little consideration given to data format, storage, standards or metadata¹⁸ unless it is a critical constraint to a specific project.

Somewhat paradoxically, data management is often seen as a short term expedient rather than a long term obligation. Preserving data for the future is not part of the core business of researchers in the same way as intellectual property, copyright, publication, ethics, mentoring and training. Once a project is completed, data is rarely cleaned up, quality controlled or metadata added – there is no time or money to do so and the focus is on the next project.

Many data retention policies are considered un-enforced and unenforceable. There is no clear responsibility in many organisations for the enforcement and auditing of compliance. Centralised IT services typically do not extend to many of the activities associated with data management and curation. There is also a 'black hole' between physical and electronic records management policy and practice, which means that important data (and metadata) may be inadvertently lost or destroyed.

Regulatory Environment for Cloud Computing

Cloud Services

Cloud services can provide Australian SME with a great competitive advantage both locally and internationally. Key to encouraging SME to make use of cloud services is clear advice on what cloud services must provide in terms of protection of their and their customers' data to be compliant with Australian law.

One of the cloud's great advantages is 'location independence' but when data may be stored or processed in another territory, it is important that compliance of those services with Australian law can be confirmed.

Data protection and privacy laws can be complex and difficult to navigate. Simple advice to Australian SME on the protection of customer data in particular will go a long way to enabling them to make informed decisions as to the location of data and mitigate the risk of data-loss or exposure.

Australian privacy and data protection laws must ensure that data is only stored in locations, formats and systems which provide protection and redress for Australian citizens. Without assurance that Australian business is taking care to protect their data, citizens will be wary of making that data available or taking full advantage of the Digital Economy.

Cloud Infrastructure

In response to the various regulatory regimes existing in the world, but also to provide high-performance and resilient connectivity, cloud services suppliers (such as Amazon, Google, Microsoft and many others) are setting up infrastructure in multiple territories. There are significant

¹⁸ Metadata and standards are very important, as poor metadata reduces the chance of the data being discovered, and home grown standards minimise the chances of interoperability.

benefits to any nation to being the recipient of the investment involved in the build of cloud infrastructure or development of local cloud services suppliers.

Cloud services infrastructure can bring skilled jobs, infrastructure development and export earnings to Australia. Because of power requirements for data-centres it also brings the opportunity to invest in and develop alternative energy sources. The provision of local infrastructure, which is used by international customers, will drive utilisation of Australia's international network benefiting telecoms companies and encouraging further investment in local broadband connectivity.

Hypothetically, what if the Federal Government outsourced its data storage to Amazon?

Based on current estimates of Federal Government data holdings and the US \$0.15 cost per Gigabyte (each way), the government would expect to spend about US\$1.2 billion to outsource its data storage to an outsourced provider such as Amazon. Moreover, the data, and the money would go out of the country¹⁹.

Year	Data Holdings in GB	\$US GB /Month Data Holding	Holding Price per Year	\$US GB/ Month Upload	Upload Price per Year	\$US GB/ Month Download	Download Price per Year	Total per Year
2009	25000000	0.15	\$45,000,000	0.1	\$30,000,000	0.13	\$2,340,000	\$77,340,000
2010	40000000	0.15	\$72,000,000	0.1	\$48,000,000	0.13	\$3,744,000	\$123,744,000
2011	64000000	0.15	\$115,200,000	0.1	\$76,800,000	0.13	\$5,990,400	\$197,990,400
2012	102400000	0.15	\$184,320,000	0.1	\$122,880,000	0.13	\$9,584,640	\$316,784,640
2013	163840000	0.15	\$294,912,000	0.1	\$196,608,000	0.13	\$15,335,424	\$506,855,424
5 YEAR TOTAL			\$711,432,000		\$474,288,000		\$36,994,464	\$1,222,714,464

The Government should therefore ensure that Australia becomes an attractive site for cloud services infrastructure, it is critical that our regulatory regime is open, transparent and cognisant of the needs of international customers. Privacy and data-protection legislation should look to ensure the 'neutrality' of Australia toward foreign data stored in our territory. Wherever possible, Australian law should work to ensure that foreign companies and citizens can be sure that data stored or operated upon in Australia will not be accessible by Australian companies or government.

¹⁹ Upload % is 100% as all data must be initially uploaded. Download is an assumed 6% based on estimates that 3% of data changes daily and therefore needs a download and an upload. AMAZON S3 price in 2007 was USD0.15/GB/Mth

C.5 Digital Economy and the Environment

The Quality of Service Enigma

A key factor of the Digital Economy is the 'user Experience'. The Australian infrastructure must provide the anticipated Quality of Service if it wishes to be a serious player in the Digital Economy. Research has shown that a User will go elsewhere if they fail to receive responses within seconds (or even sub-seconds).

This Quality of Service 'overhead' directly impacts the environment. For example, one Google search may involve 2000 servers. Similarly, in the data storage environment, data may be 'striped' across multiple disk drives in a storage system and even across multiple systems in order to ensure the required responsiveness.

Multiple, small capacity disks may also be used in preference to larger capacity disks provide since smaller capacity disks provide faster response times. However, these smaller disks consume as much power and cooling as larger capacity disks, thus overriding potential environmental benefits.

Solid State Disks (SSD) have the potential to deliver massive gains in throughput, but expressed in terms of storage capacity and power consumption, they are also environmentally unfriendly.

Government should expect all service providers, whether Government or Industry, to employ 'smart technologies' in the data storage arena that automatically migrate data between faster and slower devices according to the Quality of Service required. These smart technologies should allow both manual and automated actions to maintain Quality of Service.

Government should establish an Industry Forum to identify the smart technologies currently available, and on the design board, to meet this challenge.

Impact on Data Centre Infrastructure

Growth in business systems and supporting data is challenging the physical capacity of many traditional data centres world-wide since it requires a corresponding growth in servers, data storage and data centre floor-space. The energy requirement to drive both this infrastructure and the increasingly powerful air-conditioning systems needed to keep the systems from overheating is massive - even with the most advanced systems in place. The weight and density of this equipment and cabling is also compromising the floor loading of many traditional buildings.

Gartner estimates that 70 percent of top Fortune 1000 organisations will require new data centres within the next 5 years. Hitachi believes that at least fourteen federal government data centres (approximately 10,000 square metres) will be due for upgrade, replacement or contract renewal within the next five years. Annual power costs alone for these data centres are conservatively estimated by Hitachi to grow from \$11.3 million in 2008 to about \$22.7 million in 2013 - a total cost of nearly \$100 million over the five year period.

Examples of the pressure on data centre infrastructure parameters include:

- **Storage.** Average data centre storage demand is increasing by about 30 percent per year. In many cases, data centres are doubling their storage needs every 18 to 24 months. The growth in storage demand is challenging the physical constraints on many data centres globally.

- **Servers.** Server blade technology is a major driver data centre infrastructure. Rack topology has changed significantly from an average of 4-6 servers in one rack to 48-96 servers per rack. Gartner estimates the compound annual growth rate (CAGR) in the average number of servers at 12 percent.
- **Heating Ventilation and Air Conditioning.** Data centres require stable thermal conditions with closely controlled temperature, humidity and air quality. Equipment heat output in modern high density data centres typically have very high heat loading and very uneven heat distribution.²⁰ Gartner anticipates that today's mix of older technology and blade technologies requiring a power density of approximately 1.5 - 2KW/m², will grow over the next five years to 3.5KW/m².
- **Data Centre Floor-space.** Most organisations are experiencing a significant increase in the deployment of new infrastructure equipment (servers, storage and networking equipment). Despite major advances in data storage consolidation and virtualisation, Gartner forecasts floor space growth for IT and plant to continue at around 5-10 percent per annum. The most telling measure is that the amount of floor space actively occupied by racks and equipment. For example, in a number of federal government departments, floor-space occupation doubled (15 percent CAGR) over the past five years, in part due to the implementation of duplicated service continuity and disaster recovery systems.
- **Weight and Floor Loading Support.** Data centre weight and floor loading constraints are being compromised by growth in the number and density of servers, storage and other communications equipment:
 - **Servers.** The typical weight increase from single server racks (at 4-6 servers per rack) to blade server racks ranges from around 380kg to 1200kg. Since the footprint of these racks has not changed significantly there is a corresponding increase in floor loading.
 - **Storage.** Weight increase for storage units is somewhat offset by increasing disk capacity. Nevertheless, enterprise-class storage units now weigh in excess of 4 tonnes compared to previous generations (~ five years ago), which weighed about 2 tonnes.
 - **Communications.** The density of communication switches is becoming more significant as an increasing number of Director-class units are installed; each unit has hundreds of ports, and weighs in at over 350kg.

Changes in the weight of component equipment can have a direct impact on data centre floor load requirements. For example, one of Hitachi's Federal Government clients experienced average floor load growth from 380 kg/rack to 1000 kg/rack over the past five years. A typical SAN storage unit in this department grew from 1980kg to 4300kg in the same period. The floor load capacity also impacts upon the ability to utilise finite data centre floor-space.
- **Power.** Industry analysts are in agreement that the new IT technologies in communications, servers and storage²¹ are rapidly increasing the ratio of power (and hence heat generation and cooling required) to the infrastructure footprint. Typical data centre electricity consumption (power and air-conditioning) increased 97 percent from 2000-2005 and now accounts for between 2-3 percent of total electricity consumption.

²⁰ For example, in 2000, the average computer rack produced about 1kW of heat whereas a current rack of about four fully loaded blade servers creates about 24kW of heat in an enclosure with just 0.7m² of floor area.

²¹ Storage generally accounts for 30 to 40 percent of the power used in datacentres. IDC estimates that storage accounts for 37 percent of overall datacentre power consumption

Server growth and data storage growth can generate corresponding increases in energy requirements.²²

- The United States EPA issued a report to Congress in 2007, stating that it anticipated a doubling of data centre power consumption over the next five years. This estimate in turn was based on evidence of a doubling of power consumption over the previous five years.
- The report identifies current consumption of data centres as accounting for 1.5 percent of total power consumption in the United States, and expects this figure to grow to 2.5 percent.
- The EPA also predicts that securing additional electricity supply from already stretched utilities will be a major threat to data centres in the next five years. EPA power consumption growth estimate corresponds to a CAGR of approximately 15 percent.²³
- Gartner anticipates that today's mix of older technology and blade technologies requiring a power density of approximately 1.5 -2KW/m², and will grow over the next five years to 3.5KW/m².²⁴

²² In 2006, there were 420 million hard disk drives in circulation worldwide, consuming the equivalent of 4330 megawatts of power. By 2015, it is expected that New York City will require an additional two power stations simply to accommodate the demand for data centre power.

²³ This estimate may in fact be overly conservative. Recent Uptime Institute research shows data centre energy consumption in the top quartile of sites researched grew 20-30 percent annually in 2006 and 2007. This rate far exceeds the EPA prediction. One Australian Federal Government department has experienced total power and cooling growth from 230 KVA to 850 KVA over the past five years; equating to 30 percent CAGR.

²⁴ As an example, over the past five years, one Federal Government department has seen ~1000 older technology servers requiring 230KVA on ~600m² grow to 2000+ mixed old technology and blade technology servers requiring 850KVA on 700m². This growth entailed an increase from 380W/m² in 2003 to 1.2KW/m² in 2008.

C.6 Measuring the Digital Economy and its Impacts

Whether they be maintained in-house, outsourced, or over the internet, ultimately, the data storage systems and servers which drive the digital economy are housed in datacentres.

The ICT sector of developed economies (including Australia) accounts for 2 percent of global greenhouse gas emissions. Similar reports place the power consumption of the world's largest datacentres at around 1.5 percent of global greenhouse gas emissions.

Australia's average power consumption rate in 2006-2007 was 23GW; peaking at 35GW.²⁵ This translates to Australian datacentres consuming some 350 MegaWatts of power (350MW) in 2006-2007. At AUD \$0.12 per kilowatt hour (kWh) this power consumption costs AUD \$370 million /year, and generates 3.25 Mega-tonnes of greenhouse gas emissions. This number is doubling every 5 years.

A dataset which should be considered as core information for the impact of the Digital Economy is annual datacentre power use and consequent greenhouse gas emissions. This dataset must also include a measurement of the efficiency of the respective datacentres.

Potential metrics include:

- **Power Usage Efficiency.** Power Usage Effectiveness (PUE) is a metric currently used to gauge data centre energy efficiency. PUE is determined by dividing the amount of power entering a data centre by the power used to run the computer infrastructure within it. PUE is therefore expressed as a ratio, with overall efficiency improving as the quotient decreases toward 1.

For example, the *Hitachi CoolCenter50* experimental data centre covers 10,000m² and is currently the world's most eco-friendly and power-efficient data centre. It has been designed to achieve a 50 percent reduction in power usage by 2013 based on an underlying PUE rating of 1.6 combined with extensive implementation of Virtualisation.²⁶
- Many datacentres in Australia are ageing, and have a relatively poor efficiency.²⁷ Hitachi estimates the average PUE ratio of these data centres at around 1.9 (which means that for every 1.0 kW of power used for computing, a further 0.9kW goes in cooling and general building supply. Building and migrating to newer, more efficient datacentres with an average PUE of 1.6 would save Australian datacentres 3.46 Megatonnes of greenhouse gas emissions, and AUD \$550 million over the next 5 years.
- **Corporate Average Data Centre Efficiency.** A more recent addition is Mckinsey's Corporate Average Data Centre Efficiency (CADE), which takes into account the energy efficiency of facilities, their utilisation rates and IT asset efficiency. This adds a useful measurement of IT utilisation to the picture.

²⁵ source AER Australian Energy Regulator Review

²⁶ Located in Yokohama, Japan, this state-of-the-art datacentre architecture has been designed to offer the highest levels of energy efficiency and reduce carbon emissions by 20 percent, while also lowering IT management costs. The facility also features cutting-edge products such as thermal hydraulic cooling devices, UPS systems and highly advanced power supply converters that work in concert to contribute significant power, cooling and space benefits. The datacentre also has a thermal storage unit on the roof of the facility that stores cool air at night and uses this 'free energy' to cool the systems during the day. <http://www.hds.com/assets/pdf/hitachi-green-technology-at-work.pdf>

²⁷ Some commercial data centres claim to have designed down to a PUE of 1.3 in their marketing materials, however these claims deserve close scrutiny.

- **Total Cost of Storage Ownership.** The Government needs to rationalise how best to reduce total cost of ownership (TCO) for storage infrastructure. Since the purchase price of any storage solution is generally 15 percent to 25 percent of the total cost of storage ownership, a wider view of TCO must apply the basic premise that price \neq cost.²⁸
 - This approach should also be cognisant that, regardless of the method or approach used for quantification, every organisation has a unique perception as to which savings are real (hard), and which are phantoms (soft).
 - Hard savings are typically those that can generate tangible financial variances, and can be counted or taken out of a budget. Soft savings are those that are recognised and appreciated, but may not be defensible in demonstrating that real monies are ever saved. However, hard savings to one organisation (such as floor-space savings from consolidation) may be a soft saving to another.

The Government should therefore look to Industry leaders such as Hitachi to develop guidelines and benchmarks for data delivery rates / power consumption and total cost of ownership for Tiered Levels of Service.

²⁸ Some IT planners assume that purchasing low-cost disk solutions will drive down the cost of storage ownership, however Hitachi contends that purchase price alone does not govern operating expenses or reducing the total cost of disk ownership. In fact a focus solely on cheaper disk solutions to has been observed to reduce short-term capital expense (CAPEX) but produce a negative long-term impact on OPEX.

Appendix 1 - Legal and Regulatory Requirements: Data Storage

Data storage practices must comply with a range of legal and regulatory requirements and records management standards.

Legislative Requirements

Commonwealth Government authorities are required to comply with a range of legal obligations in relation to their records and the information contained within them. There are three overarching and complementary Commonwealth laws that apply to all government records management: the Archives Act 1983, the Freedom of Information Act 1982 and the Privacy Act 1988.

- The [Archives Act 1983 \(Cth\)](#) prescribes access to, management, custody and disposal of Commonwealth records and establishes that the public generally has a right of access to Commonwealth records that are more than 30 years old. The Act prohibits the destruction of Commonwealth records without permission from the National Archives of Australia ('NAA'), unless otherwise specified under other Commonwealth legislation.
- The [Freedom of Information Act 1982 \(Cth\)](#) provides for public access to documents kept by Commonwealth Government departments. Importantly, this legislature enables the public to obtain access to electronic records. It also establishes a mechanism for the public to request and obtain amendments to records containing their personal information where they can prove their records are inaccurate or false.
- The [Privacy Act 1988 \(Cth\)](#) aims to prevent Commonwealth agencies from misusing information about members of the public. It specifies that agencies may only use information about individuals for lawful purposes and these purposes must relate directly to the functions or role of the agency. Further, the information kept by these agencies must be secure, accurate, relevant, complete, and not misleading. The Act gives people a right to see records kept by agencies about themselves. Notably, records over 30 years old are exempt from the Privacy Act (but covered under the Archives Act).

There are also a number of other legislative requirements that affect recordkeeping generally, such as the Evidence Act 1995 and the Electronic Transactions Act 1999:

- The [Evidence Act 1995 \(Cth\)](#)²⁹ addresses the use and admissibility of documents (including electronic documents) in court proceedings.³⁰ This legislature imposes a requirement on all Commonwealth agencies to ensure that their records management practices can withstand the scrutiny of the law.
- The [Electronic Transactions Act 1999 \(Cth\)](#) enables the use of electronic communications to satisfy obligations under Commonwealth law, however the obligations placed on agencies under the

²⁹ The Evidence Act applies as per following laws (in Australia)

- ISO 15489: the world standard for information and document management
- BSI DISC PD0008: British standard relating to 'Legal Admissibility and Evidential Weight of Information Stored Electronically'
- The Federal Rule of Civil Procedures: incl Rule 901 (a) covering the authenticity of information
- VERS: Victoria Electronic Records strategy
- XENA: The National Archives of Australia digital preservation approach for converting digital records into open preservation file formats based on XML

³⁰ Electronic evidence is anything that can be stored on computer (documents, emails, metadata, logs, audit, images, and any other information stored digitally). This evidence can be easier to tamper with or destroy. The number of cases where electronic evidence is used has increased in the last 5 years. Electronic discovery is a key to providing electronic evidence.

Archives Act 1983, specifically in relation to the preservation and disposal of Commonwealth records, continue to apply.

Most organisations need to create and keep not only information about transactions carried out via electronic means but also the evidence, in the form of records that captures the content and the context of these activities. This evidence therefore needs to document what transaction occurred, when it occurred, its location, the identity of the participants, its relationship to the business process for which it serves as evidence, and links to broader mandates governing the business of which it is part.

- **Sarbanes Oxley Act 2002 (US).** Despite being a ‘foreign’ law, the Sarbanes-Oxley Act may still have relevance for many Australian organisations, particularly where US research grants or commercial collaboration is involved. Complying with Sarbanes Oxley law requires that ‘organisations have the ability to produce, on request, authentic and reliable records³¹, as well as supporting documents. The primary goal of the SOX Act is to protect investors from corporate corruption and scandal. Its main feature is the stipulation of the accountability of management in order to ensure the reliability of corporate financial reporting.
- **Government Initiatives.** Government online access and service delivery initiatives, such as those presented in the *Australian Government Information Management Office (AGIMO) 2006 e-Government Strategy*, and the importance of electronic commerce provide added impetus for agencies to implement systems with reliable records management functionality. Such functionality ensures that Government records:
 - can be proven to be genuine;
 - are accurate and can be trusted;
 - are complete and unaltered;
 - are secure from unauthorised access, alteration and deletion;
 - are findable and readable; and
 - are related to other relevant records.

While in the traditional paper recordkeeping environment these legislative requirements are accepted and built into a records management system, the electronic environment makes it necessary to think anew about strategies to adopt that ensure records have the same degree of reliability, authenticity and useability as paper records previously.

In short, electronic recordkeeping systems are metadata systems, and metadata is vital to any good recordkeeping system.³²

Records Management and Metadata Standards

There is a range of relevant standards and guidelines that, from a general records management perspective, affect the way in which electronic records should be managed.

The standards and guidelines that are specifically endorsed and recommended by the National Archives of Australia (NAA) for use by government agencies include:

³¹ According to ISO 15489, a document is a recorded information or object which can be treated as unit. Records, in contrast, are defined as “Information created, received or maintained, as evidence and information by an organisation or person, in pursuance of legal obligations or transactions of business”. (Such as electronic and printed documents, images and emails).

³² The Australian Records Management Standard, AS ISO 15489, defines recordkeeping metadata as:

‘Data describing context, content and structure of records and their management through time...’

The National Archives of Australia further defines recordkeeping metadata as:

‘Structured or semi-structured information that enables the creation, management and use of records through time and across domains. Recordkeeping metadata can be used to identify, authenticate and contextualise records and the people, processes and systems that create, manage, maintain and use them...’

- **ISO/IEC 11179.** ISO/IEC 11179 is the international standard for representing metadata for an organization in a Metadata Registry. ISO/IEC 11179 is cited by the Australian Government Interoperability Technical Framework (AGITF) as the standard for definition of data elements.
- **Australian Standard for Records Management – AS ISO 15489.** The NAA has adopted AS ISO 15489 as the basis for its e-permanence suite of best-practice records management standards, policies and guidelines.
- **Recordkeeping Metadata Standard for Commonwealth Agencies (1999).** This standard details the metadata that the NAA recommends (as a minimum standard) should be captured in recordkeeping systems by Commonwealth government agencies. The Standard is designed as a reference tool for the design, selection and implementation of electronic records management and related information management systems.
- **NAA Digital Recordkeeping Guidelines.** These guidelines assist Government agencies on the creation, management and preservation of digital records. It also addresses a number of issues germane to digital records management including determining how long to keep digital records, storage of, securing, providing access to, and business continuity planning for digital records.
- **Australian Government Recordkeeping Metadata Standard (AGRkMS).** This standard describes information about records that the NAA recommends be captured in business information systems used by Australian Government agencies to create and capture records. It sets out the type of information that agencies should capture in a structured way to describe the identity, authenticity, content, structure, context and essential management requirements of records.

The NAA recordkeeping metadata standard and the digital recordkeeping guidelines are of particular relevance in the context of defining a transition from paper-based to fully electronic records.