Recordkeeping and Archiving in the Cloud. Is There a Silver Lining?

Sue McKemmish
Centre for Organisational and Social Informatics
Monash University, Australia
sue.mckemmish@monash.edu

Summary

There is a rapid uptake of cloud computing in many places around the world. What are the implications for recordkeeping and archiving? Cloud computing offers attractive benefits including significant cost savings, efficiencies, flexibility and scalability, as well as opportunities for the innovative development and delivery of new records management and archival services. Recordkeeping and archiving in the cloud also carry significant risks associated with security, privacy, integrity, authenticity, accessibility and digital continuity, as well as issues relating to commercial continuity and the lack of transparency of cloud services. This paper provides an overview of the current cloud computing environment, different models and types of cloud services, and related recordkeeping and archiving benefits and risks. It uses a case study approach to explore the strategies that two archival authorities in Australia are pursuing in their role as records management standard setters. They include risk assessment approaches and the development of checklists and other tools to guide the evaluation and selection of cloud services, risk management, and contract negotiation. The paper also briefly references the European Union's Cloud for Europe initiative. Finally it calls on recordkeeping and archiving communities to take a pro-active approach, as both consumers and potential service providers, to influencing the future of recordkeeping and archiving in the cloud.

Key words: cloud computing, recordkeeping, archiving, risk management

Introduction

This paper explores the question: what are the implications for recordkeeping and archiving of the rapid uptake of cloud computing? It provides an overview of the current cloud computing environment, different models and types of cloud services, and related recordkeeping and archiving benefits and risks. It uses a case study approach to explore the cloud computing strategies that the National Archives of Australia and the Public Record Office of Victoria are pursuing in their role as records management standard setters. The case study is based on analysis of the policies, strategies and guidelines for government organizations developed by archival authorities to optimize the benefits and miti-

gate the risks associated with records in the cloud. National and state government archival authorities in Australia have a dual role in relation to the regulation of current recordkeeping for good governance and democratic accountability, as well as preservation of government archives as part of Australia’s collective memory and cultural heritage. The case study focuses on their role in relation to the regulation and promotion of current recordkeeping. However, the two roles are inter-related as best practice recordkeeping, particularly in today’s digital environments, is critical to the long-term preservation of archival records of continuing value. Over the past few years the National Archives of Australia and Public Record Office of Victoria have developed strategies relating to cloud computing which are based on risk assessment and management approaches. The paper goes on to present a categorization of risks, and a list of questions relating to cloud services for organizations planning to put their records in the cloud. It is based on an analysis of the policies, guidelines, checklists and other tools issued by these two archival authorities to guide the evaluation and selection of cloud services, risk assessment and management, contract negotiation and the monitoring of cloud services, Finally the paper references recent developments in the European Union, in particular the Cloud for Europe initiative. In conclusion, it calls on recordkeeping and archiving communities to take a proactive approach, as both consumers and potential service providers, to influence the future of recordkeeping and archiving in the cloud.

Records are defined in this paper in accordance with AS/NZ ISO 30300 as:

> Information created, received and maintained as evidence and as an asset by an organization or a person, in pursuit of legal obligations or in the transaction of business (ISO 2012, p. 9).

Archives are defined as records of continuing value. Digital records and archives take multiple forms, come in many formats and are captured in many places, including email systems, websites, blogs, wikis and other social media, business systems including databases, and multimedia systems, as well as Electronic Document and Records Management Systems. They are stored in personal and corporate devices and equipment (on desktops, laptops, tablets and mobile phones, organizational servers), in data stores and digital repositories, including archival digital repositories. In organizational settings, their storage and management might be in-house or outsourced to external service providers. Increasingly records may be managed and stored in the cloud.

Cloud Computing

The US National Institute of Standards and Technology (NIST) defines cloud computing as:

> a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provi-

> sioned and released with minimal management effort or service provider interaction (Mell & Grance 2012, p. 2).

Cloud computing offers attractive benefits including significant cost savings, efficiencies, flexibility and scalability, as well as opportunities for the innovative development and delivery of new services. It also carries significant risks associated with the security, privacy, integrity, authenticity, accessibility and digital continuity of data and records in the cloud. There are also issues relating to commercial continuity and the lack of transparency of cloud services that impact on recordkeeping and archiving.

Public cloud computing offers a pay-as-you-go business model as an attractive alternative to large-scale capital investment in software, platforms and infrastructure, and paying an in-house workforce to manage them. Software-as-a-service (SaaS) delivers business applications hosted by a provider over the web. Platform-as-a-Service (PaaS) provides custom application development or deployment environments in which applications can be built and run on service provider systems. Infrastructure-as-a-Service (IaaS) provides virtual infrastructure components such as servers, storage and network access. Computing resources are accessible everywhere, anytime through diverse media and devices and thus support a mobile and flexible workforce. A cloud services provider pools and dynamically configures its computing resources to meet the needs of multiple clients, enabling rapid scaling of service to meet demand and optimised use of resources.

The rapid uptake of cloud computing is evidenced by the widespread personal and business use of public cloud services such as Microsoft drop boxes, Google drive, Apple's iCloud, gmail, Facebook and other social media, as well as contracted public, private and community cloud services customized to individual business needs. Public cloud services operate on a "multiple tenant" model over the public internet with facilities shared by multiple users. In a private cloud an organization has exclusive use of cloud infrastructure, and in a community cloud, several organizations from a community with common concerns (e.g. related to security, privacy, accountability or jurisdiction) share cloud infrastructure (Mell & Grance 2012, p. 3). Private and community clouds may be managed internally or by a third-party and hosted internally or externally. If they are managed and hosted externally by a third party, private and community cloud services may share more of the benefits and be vulnerable to many of the risks associated with public cloud computing. A hybrid cloud is composed of two or more clouds (public, private or community) and may deliver the best of both worlds. Hybrid clouds and to a lesser extent community clouds potentially enable organisations to realise the benefits of a public cloud (multi-tenancy and pay-as-you-go), but with the added level of privacy, security, accountability and standards compliance usually associated with a private cloud.

Risks associated with the processing, storage and management of data, and more specifically records in the cloud, particularly the public cloud, include security, privacy, integrity, authenticity, accessibility and digital continuity, as well as issues relating to commercial continuity and the lack of transparency of cloud services. The degree of risk and possible consequences for data and records vary for different models and types of services, with the highest risks and most serious consequences generally associated with software and platform services in the public cloud, and relatively low risks associated with infrastructure services. However some of the risks associated with the location of data stores and servers, and trans border data flows are the same regardless of the model or type of service.

The market-speak and images associated with cloud computing provide us with a somewhat misleading mental model which belies the physical reality of the vast server farms owned by Google (US, Finland, Russia and Germany, expanding into Chile, Hong Kong, Taiwan and Singapore), Microsoft (US, Hong Kong, Singapore, Ireland, Brazil), Amazon (US, Dublin, Brazil, Japan), Facebook (US, Sweden), and Apple (USA). From another perspective "cloud" may be an apt term – as there are concerns about the extent to which the operations of service providers are "clouded" in secrecy. There has been a lack of transparency and accountability. For example, many large service providers have policies on non-disclosure of the location of data farms (apparently on security grounds, although server farms are hard to hide as apparently the largest ones are visible from space). More specifically exactly where data is stored, and whether it is being moved around and transferred across borders may not be disclosed to you (see for example Microsoft's explanation at http://www.microsoft.com/online/legal/v2/en-us/MOS_PTC_Geo_Boundaries.htm). The use of the terms public and private cloud computing, especially in marketing, can also be opaque as the model offered may not always be what it purports to be:

> Be sure that the deployment model offered is what it appears to be and not a marketing ploy whereby a vendor offers differently priced packages of the same services. (PROV 2012, p. 16)

Australian Responses

In 2013, a whole of government policy embracing cloud computing was issued by the Australian Government Information Management Office (AGIMO 2013). The policy was framed as part of the National Digital Economy Strategy and National Cloud Computing Strategy, and highlighted the potential contribution of cloud computing to the national economy, and the leadership role the Australian Government should play in the adoption of cloud computing in all sectors. The policy aims to "help agencies adopt cloud computing to boost innovation and productivity", and places on government agencies "an explicit obligation to consider cloud services when procuring their new ICT requirements"

(p. 3), taking into account best value for money and adequate management of associated risks. It also prescribes the migration of public web sites to public cloud services:

The Australian Government will be a leader in the use of cloud services to achieve greater efficiency, generate greater value from ICT investment, deliver better services and support a more flexible [mobile] workforce. (AGIMO 2013, p. 3)

In terms of future directions, in line with developments in other parts of the world, the policy points to the possible adoption of a more centralised approach in the form of an Australian Government community cloud, and the possible development of whole-of-government arrangements whereby cloud service providers would submit tenders to become preferred suppliers. In the tenders they would need to demonstrate that they could meet Australian Government requirements. In the meantime, individual government agencies are directed to a range of documents that provide advice and recommendations relating to assessing and managing the risks associated with cloud computing, particularly involving the security and privacy of data processed and stored offshore. These documents include the *Protective Security Policy Framework*, the Defence Signals Directorate's paper on *Cloud Security Considerations* (Department of Defence 2011), and AGIMO's guidelines relating to *Privacy and Cloud Computing for Australian Government Agencies*, and *Negotiating the Cloud: Legal Issues in Cloud Computing Agreements* (AGIMO 2013, pp. 11-12). Following these recommendations and guidelines, personal data and security classified data should only be processed, stored and managed by cloud services located in Australia *and* under Australian jurisdiction – otherwise compliance with Australian security requirements and privacy law cannot be guaranteed.

AGIMO's policy also points to the National Archives of Australia's policy on *Records Management in the Cloud* (NAA 2011a) which was developed in the context of an Australasian Digital Recordkeeping Initiative (ADRI) advisory paper on identifying risks associated with cloud computing, assessing the level of risk for different records (for example secret, confidential, commercially sensitive and personal records), how to perform due diligence when selecting a provider, establishing contractual arrangements, and monitoring their implementation (ADRI 2010). The National Archives' policy states that:

> Cloud computing poses both benefits and risks for Australian Government agencies. Gains in cost and efficiency need to be weighed up against the risks associated with privacy, security and records management.

For records management, it is essential for agencies to consider:

- where records will be stored – there may be risks to Australian Government records when they are stored outside Australia

- the value and nature of the records – the higher the value of the records the more control there needs to be over their management to ensure their integrity, authenticity and reliability
- the risks that may arise – different models of cloud service provision will present different risks to records
- whether risks can be satisfactorily mitigated – this may depend on the ability to negotiate contracts and agreements that address the risks and meet legislative obligations (NAA 2011a, np).

The National Archives policy specifies the need for informed risk assessments and NAA provides a *Check List* (NAA 2011b) for use in assessing risks, assessing and selecting service providers and negotiating contracts. It also specifies the obligations of government agencies to ensure that records in the cloud are:

- governed by the Archives Act 1983
- authentic, accurate and trusted
- complete and unaltered
- secure from unauthorised access, alteration and deletion
- findable, readable and returnable
- related to other relevant records (NAA 2011a, np).

State government archival authorities around Australia have also issued policies and/or guidelines relating to cloud computing. The Public Record Office of Victoria is leading the way. On the basis of extensive research on the implications of cloud computing for recordkeeping (PROV 2012), the Office has issued an excellent *Cloud Computing Guideline* (PROV2013a) and a set of very useful *Tools*, including a risk assessment template and matrix for cloud computing environments, a contract checklist, a requirements checklist and a mapping of intersecting requirements which affect decision making (PROV 2013b). The *Guideline* states that the appropriate treatment of records in the cloud should shape decisions relating to:

- The level of service required
- The contract conditions imposed
- The audit regime adopted
- Restrictions on the physical location of the cloud servers
- Restrictions on the selection of providers by country of registration
- Restrictions on the kind of cloud environment selected (public, private or community) (PROV 2013a, p. 5).

In developing strategies, policies and guidelines for the cloud, the archival authorities frequently reference the suite of international, national and archival authority standards, policies, strategies and guidelines that are already in place relating to digital records and recordkeeping. There is a hidden assumption that best practice digital recordkeeping is already implemented in government organizations, and that good quality records and metadata are already being cre-

ated and managed in digital systems – that the organization's recordkeeping and records are "cloud ready". However as evidenced in reports by Audit Offices, Ombudsman's Offices, Privacy Commissioners and other watchdogs in a number of Australian jurisdictions, there are many cases of poor quality recordkeeping in digital environments, in particular a lack of recordkeeping functionality in business systems and databases.

The Recordkeeping Implications

A close analysis of the issues papers, policies and guidelines published by the National Archives and Public Record Office of Victoria highlights a range of risks relating to processing, storing and managing records in the cloud (ADRI 2010, NAA 2011, 2013a, 2013b, PROV 2012, 2013a, 2013b). On the basis of the analysis, seven risk categories have been identified, and two checklists of questions have been developed to assist government organizations planning to put their records in the cloud, particularly the public cloud. They are presented in Table 1 below. One checklist of questions is designed to help an organisation to evaluate whether potential service providers can meet an organization's requirements relating to their records in the cloud. The other checklist of questions relate to the "cloud readiness" of an organization's recordkeeping management frameworks and digital recordkeeping.

Table 1. Recordkeeping Risks and Checklists

RISK CATEGORIES	CHECKLIST FOR SERVICE PROVIDER	CHECKLIST FOR ORGANIZATON
Location and Legal Jurisdiction Note: Risks in this category are particularly relevant to personal, security and confidential records	Can the service provider disclose precise information about location and movement of data/records? Can the service provider negotiate a contract that guarantees compliance with the laws in the client's jurisdiction so that the client can meet legal obligations re privacy, security, FOI and legal disclosure?	Does the organization know what laws apply to data, information and records in the jurisdiction(s) in which its records will stored (e.g. evidence laws, disclosure laws, privacy laws, FOI laws, national security laws)? Would those laws apply to the organization's records? Do the organization know whether those laws put it in breach of the laws in its own jurisdiction (e.g. privacy, disclosure, FOI, data security)? Does the organization know, for example, that the US Patriot Act applies to all data in storage facilities and server farms operated by US companies, regardless of where in the world they are located? Can the organization identify high risk records in terms of security,

		privacy and confidentiality? Does the organization have policies and risk management strategies relating to its use of social media in public clouds?
Transparency Accountability Governance Note: relevant to all records.	Are the operations of the service provider transparent? What governance, accountability, internal and external auditing and reporting arrangements are in place? Are third party sub-contractual agreements compliant with client requirements? Are there risk management strategies in place to mitigate unauthorised actions on records, security breaches, disasters resulting in loss or damage, the possibility of cyber criminals, terrorists, or spies hacking or scraping service provider systems?	Does the organization have in place a policy and processes for contract negotiation and monitoring compliance with contractual arrangements?
Protection of Rights in Records Note: This category is particularly relevant to personal and confidential records, and records that might be subject to copyright and intellectual property rights.	Who owns records in the cloud service? Is the privacy of data subjects protected as required by the law in the client’s country, including trans-border movement of data? How is third party access to client records managed, for example if required by a government watchdog organization in the jurisdiction in which the records are stored? Can the service provider support disclosure and access according to the rights regime in the client agency’s jurisdiction?	Can the organization specify its rights management requirements relating to records? Can the organization identify which records might include information that requires rights management?
Recordkeeping Functional Requirements Note: Typically, requirements specify that records are authentic, accurate, reliable, complete, discoverable, accessible, retrievable, readable, and persistently linked to metadata relating to their content, structure, context, management and use.	Do service provider systems have recordkeeping functionality to manage records and their metadata? Do systems create and manage metadata about service provider action on records to ensure their authenticity and integrity? Can systems ingest, migrate, convert, refresh, destroy and extract records according to client specified standards? Do service provider systems protect records from unauthorised	Can the organization specify recordkeeping functionality and metadata requirements for the different types of records to be moved to the cloud? Do in-house systems meet recordkeeping functional requirements? Are records in in-house systems and applications “cloud ready”? Are organizational systems capable of exporting and ingesting records with their metadata into/from storage and management systems in the cloud?

This category is relevant to all records and their metadata.	actions, e.g. illegal destruction, unauthorised access and use, hacking or scraping, and security breaches? Do they have best practice disaster recovery functionality? Can the service provider guarantee that no copies of client records are retained after termination of the contract? Can the service provider return all records and associated metadata in specified formats?	
Digital Continuity Note: particularly relevant to records of continuing value.	Can service provider preserve authentic records over long periods of time?	Can the organization specify its archival preservation requirements?
Vendor Lock-in Note: relevant to all records.	Does the service provider use proprietary systems and formats that potentially lock data into the service provider's servers?	Can the organization specify its requirements relating to proprietary/open source software and formats?
Commercial Continuity Note: relevant to all records.	What contingency plans does the service provider have in the event that it goes out of business or is taken over by another company?	Does the organization have adequate due diligence strategies and procedures to check the commercial viability of service providers and contingency planning?

Cloud for Europe

There has been growing concern in the European Union with the consequences of the fragmentation of the European public sector cloud computing market – namely that EU requirements relating to data ownership, integrity, preservation, data protection, privacy, transparency and accountability have little impact, services are not well integrated, and public sector clients do not get the best value for money.

> [A EU Justice] Opinion examines issues associated with the sharing of resources with other parties, the lack of transparency of an outsourcing chain consisting of multiple processors and subcontractors, the unavailability of a common global data portability framework and uncertainty with regard to the admissibility of the transfer of personal data to cloud providers established outside of the EEA. Similarly, a lack of transparency in terms of the information a controller is able to provide to a data subject on how their personal data is processed is highlighted in the opinion as matter of serious concern (EU Justice 2012, p. 2)

In response, as part of the Digital Agenda for Europe, the European Cloud Partnership (ECP) of public sector users and industry experts aims to shape the

market in Europe and stimulate a European cloud industry that can meet public sector requirements.

> The ECP aims at driving the first steps towards better public procurement of cloud services in Europe, based on common definitions of requirements and possibly eventually going as far as joint procurement across borders …
>
> Pooling public requirements could bring higher efficiency and common sectoral requirements (e.g. eHealth, social care, assisted living, eGovernment services) would reduce costs and enable interoperability. The private sector would also benefit from higher quality services, more competition, rapid standardisation and better interoperability and market opportunities for high-tech SMEs (European Commission 2013).

The Partnership has put in place a range of consultation strategies to engage stakeholders in determining the requirements of the Cloud for Europe, presenting an opportunity for the pro-active engagement of archival institutions, recordkeeping standards setters and the recordkeeping and archiving community in ensuring that requirements relating to records and archives in the cloud are fully addressed.

Towards Archives 3.0

Stančić, Rajh and Milošević (2012) introduce the concept of Archiving-as-a-Service and the need to transition beyond "custody" from a "postcustody" to a "postcustody 2.0" paradigm as illustrated in Figure 1 (which reproduces Figure 16 (p. 124) with the kind permission of the authors).

They put forward four scenarios for archiving in the cloud:

1. Service providers are responsible for control of archived content without much interference and additional control taken by creators and archival institutions;
2. Creators invest much effort in additional control of non-standardized services;
3. Services are standardized through best practices and creators recognize the importance of choosing providers consistent to these practices;
4. Archival community is actively involved in the new concept of archiving and influence providers' practices.

They conclude that the fourth "postcustody 2.0" scenario is "probably the best way to ensure long-term protection, preservation and usage of electronic content created and archived today" (p. 123), and urge the archival community to be proactive "in the formation of the new, preservation-aware cloud services" (p. 124). The Australian case study discussed earlier in this paper is an example of a "postcustody 2.0" approach. In the Australian context, government archival institutions are engaged in regulating and advising government agencies. Through the standards they set for records in the cloud, they also hope to influence the development of whole-of-government preferred supplier procurement

arrangements, and the cloud services offered by providers. At this stage in the evolution of cloud services, no Australian government archival institutions would sanction the placement of records identified as being of continuing value in the cloud. However, it is possible to envisage future scenarios where archival records might be managed and stored in the cloud, for example the establishment of an Australian government community cloud in line with developments in Europe, the UK and Canada, in this case serviced by providers compliant with the relevant laws in Australian jurisdictions.

Figure 1. Changes of Archival Practice

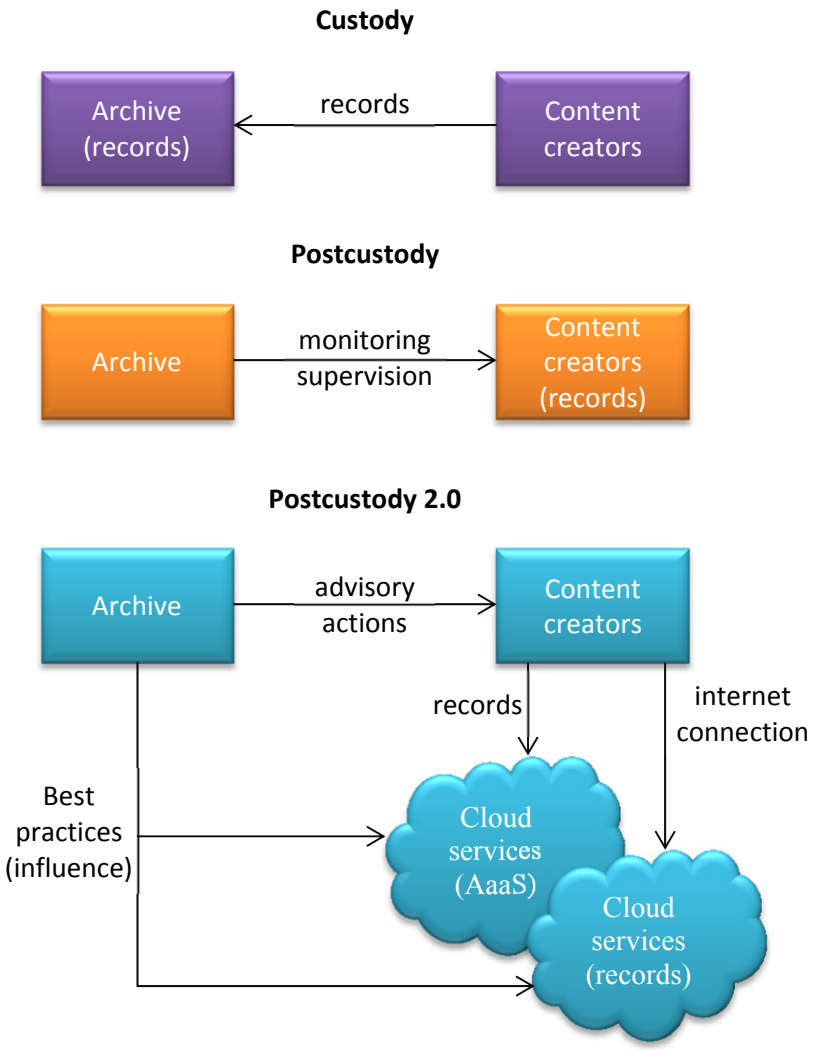

Figure 16. Changes of Archival Practice

Conclusion

Somewhere beyond custody, in the archival multiverse, lies the vision of Archives 3.0, taking advantage of the benefits and opportunities of cloud computing to build community clouds, e.g. of national, state and community archives in the EU or Australia. In this scenario the policies, standards, strategies, guidelines and tools being developed by archival authorities like the National Archives of Australia and Public Record Office of Victoria contribute to the development of broader requirements that embrace the recordkeeping and archiving needs of all those involved in the community cloud partnership. To realize this vision, recordkeeping and archiving communities need to take a pro-active approach, as standards setters, clients, and potential service providers, to shaping cloud services, and the future of recordkeeping and archiving in the cloud.

References

Australian Department of Defence. Cloud Computing Security Considerations. Canberra: Australian Government, 2011.

Australasian Digital Recordkeeping Initiative. Advice on Managing the Recordkeeping Risks Associated with Cloud Computing. Canberra: ADRI, 2010. http://www.adri.gov.au/products.aspx (12 October 2013).

Australian Government Information Management Office (AGIMO). Australian Government Cloud Computing Policy: Maximising the Value of the Cloud v. 2.0. 2013. http://agict.gov.au/sites/default/files/Australian%20Government%20Cloud%20Computing%20Policy%20Version%202.1.pdf (9 October 2013)

European Commission. Justice Data Protection WP Article 29. EU Justice Opinion 05/2012 on Cloud Computing. 2012. http://ec.europa.eu/justice/data-protection/article-29/documentation/index_en.htm (12 October 2013)

European Commission. Digital Agenda for Europe, European Cloud Computing Strategy – CloudforEurope. 2011. http://www.cloudforeurope.eu (12 October 2013)

Mell, P; Grance, T. The NIST Definition of Cloud Computing, National Institute of Standards and Technology. 2010. http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf (1 October 2013)

National Archives of Australia. Records Management and the Cloud. NAA, Canberra: NAA, 2011a. http://www.naa.gov.au/records-management/agency/secure-and-store/rm-and-the-cloud/ (12 October 2013).

National Archives of Australia. A Checklist for Records Management and the Cloud. Canberra: NAA, 2011b. http://www.naa.gov.au/records-management/publications/cloud-checklist.aspx (12 October 2013).

Public Record Office Victoria (PROV). PROV Cloud Computing Guideline 1: Cloud Computing Decision-Making. North Melbourne: PROV, 2013a. http://prov.vic.gov.au/wp-content/uploads/2013/06/Cloud_Computing_Guideline_1.pdf (12 October 2013).

Public Record Office Victoria (PROV) 2013, PROV Cloud Computing Guideline 2: Cloud Computing Tools. North Melbourne: PROV, 2013b. http://prov.vic.gov.au/wp-content/uploads/2013/06/Cloud_Computing_Guideline_2.pdf (12 October 2-13)

Public Record Office Victoria (PROV). 2013, PROV Cloud Computing Policy. North Melbourne: PROV, 2013c. http://prov.vic.gov.au/wp-content/uploads/2013/06/Cloud_Computing_Policy.pdf (12 October 2013)

Public Record Office of Victoria (PROV). Cloud Computing Implications for Records Management. North Melbourne: PROV, 2012. http://prov.vic.gov.au/wp-content/uploads/ 2012/04/Issues-Paper-Cloud-Computing.pdf (12 October 2013)

Stančić, Hrvoje; Arian Rajh, Arian; Milošević, Ivor. "Archiving-as-a-Service". Influence of Cloud Computing on the Archival Theory and Practice. // *The Memory of the World in the Digital Age: Digitization and Preservation* / Duranti, Luciana ; Shaffer, Elizabeth (ed). UNESCO, 2013, 108-125.

Standards Australia; Standards New Zealand. AS/NZ ISO 30300: 2013: Information and Documentation – Management Systems for Recordkeeping – Fundamentals and Vocabulary (ISO 30300: 2011 MOD). Sydney: Standard Australia, 2012; Wellington: Standards New Zealand, 2012.