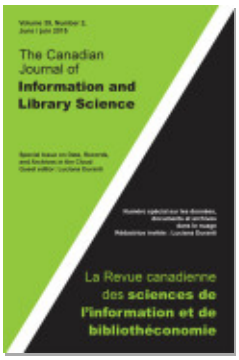
New Technologies, New Challenges: Records Retention and Disposition in a Cloud Environment / Nouvelles technologies, nouveaux défis: Conservation et déclassement des documents dans un environnement de nuage informatique



Patricia C. Franks

Canadian Journal of Information and Library Science, Volume 39, Number 2, June juin 2015, pp. 191-209 (Article)

Published by University of Toronto Press
DOI: https://doi.org/10.1353/ils.2015.0011

New Technologies, New Challenges: Records Retention and Disposition in a Cloud Environment

Nouvelles technologies, nouveaux défis : Conservation et déclassement des documents dans un environnement de nuage informatique

Patricia C. Franks
School of Information, San José State University
patricia.franks@sjsu.edu

Abstract: This article describes core records retention and disposition functional requirements extrapolated from relevant standards and guidelines and from responses to a questionnaire developed to gather information about retention and disposition functionality built into cloud services. Results of a survey completed by 168 records and information professionals are analysed to identify cloud records retention and disposition challenges. Gaps between the functionality required and that provided in selected cloud environments are identified, and recommendations to mitigate the challenges posed are offered. Future research plans are shared.

Keywords: records, retention, disposition, cloud services, functional requirements

Résumé : Cet article décrit les exigences fonctionnelles de base pour la conservation et le déclassement des documents, extrapolées à partir des normes et des lignes directrices pertinentes ainsi que des réponses à un questionnaire élaboré dans le but de recueillir des informations sur la conservation et le déclassement en tant que fonctionnalités intégrées dans les services informatiques en nuage. Nous avons analysé les résultats d'une enquête réalisée auprès de 168 professionnels de la gestion documentaire et de l'information afin d'identifier les défis de conservation et de déclassement des documents. Nous avons identifié des écarts entre les fonctionnalités requises et celles qui sont prévues dans les environnements informatiques en nuage sélectionnés. Nous formulons des recommandations visant à atténuer les défis posés et nous partageons nos projets de recherche à venir.

Mots-clés : documents, conservation, déclassement, services informatiques en nuage, exigences fonctionnelles

Introduction

Businesses and government agencies employ cloud services to take advantage of the benefits offered, such as increased operational efficiencies, accessibility, collaboration, security, reliability, and opportunities for innovation. Individuals within organizations charged with managing the records and information residing in the cloud understand, however, the myriad challenges presented when control

is relinquished to a third-party provider. One way to minimize risks associated with content stored in the cloud is to employ only an enterprise-hosted private cloud to communicate, collaborate, and conduct transactions. To do so, though, would sacrifice the opportunity to engage the public. Organizations, therefore, are increasingly turning to hybrid solutions.

By 2017, nearly one-half of all large enterprises are expected to be engaged in hybrid (that is, public/private) cloud computing (Babcock 2013). A hybrid cloud is an integrated cloud service using both private and public clouds to perform distinct functions within the same organization. A 2014 survey of 1,068 technical professionals revealed that "hybrid and multi-cloud implementations continue to be the end goal for the enterprise: 74 percent of enterprise respondents have a multi-cloud strategy, and 48 percent are planning for hybrid clouds" (Weins 2014).

Regardless of the implementation model, it is essential that organizations are able to "trust" that their records residing in the cloud can be retained and disposed of in accordance with the same requirements that govern the retention and disposition of records stored within the enterprise (Franks and Doyle 2014, 52).

Methodology

In June 2013, InterPARES Trust, a multi-national, interdisciplinary research project funded by a five-year partnership grant (2013–18) from the Social Sciences and Humanities Research Council of Canada was launched to explore issues concerning digital records and data entrusted to the Internet. Several studies were approved to investigate diverse facets of the larger research agenda, among them retention and disposition in a cloud environment research project. The project committee determined that a qualitative approach should be taken to answer two research questions:

- How does the use of cloud services affect our ability to retain and dispose of records in accordance with the law and other applicable guidelines?
- What can be done to mitigate any risks arising from the gaps between our ability to apply retention and disposition actions to manage records residing within the enterprise and those residing in the cloud?

Three data collection methods were employed to achieve these objectives: (1) content analysis of national and international standards and guidelines to identify functional requirements for retention and disposition of records stored in electronic systems; (2) analysis of data gathered through publicly available cloud service information and interviews with cloud service representatives to understand the retention and disposition functionality offered; and (3) an analysis of the responses to an online survey of records and information management professionals to understand the retention and disposition challenges faced when records and information reside in a cloud environment.

Literature review

Five themes that emerged from a review of the literature are risk analysis and risk management, legal issues, information governance, new approaches to manage retention and disposition in the cloud and records classification.

Risk analysis and risk management

Allison Grounds and Benjamin Cheesbro (2013) cite eDiscovery risks due to the mismanagement of retention policies and the inability to implement legal holds successfully in the cloud environment. Peter Géczy, Noriaki Izumi, and Köiti Hasida (2013) purport that two major risks posed by a hybrid cloud are inherited from their public cloud segment: data security and loss of control. Of the top ten cloud computing risks identified by Amab Dutta, Chao Alex Peng, and Alok Choudhary (2013, 44), two impact the organization's ability to adequately govern content residing in the cloud: difficulties in changing cloud vendors in the event of service dissatisfaction (vendor lock-in) and enterprise data re-migration difficulties at the end of the cloud contract.

Legal issues

When managing information in the cloud environment, retention and disposition no longer entail local storage but, rather, global and cross-border storage locations with multiple jurisdictional laws, especially related to data privacy. To respond to this dilemma, some cloud providers locate physical data centres in various geographic regions. According to Masooda Bashir and Jay Kesan (2011), contracts and terms of service agreements do not protect customer data from misuse of data or disclosure of data to third parties by cloud service providers. Iulia Ion and his colleagues (2011) observe that the expectation of privacy is not typically written into cloud provider service agreements. Cloud users potentially do not even know if and when their data are accessed by other users. For instance, on 23 July 2013, the Supreme Court of the State of New York (2013) ordered the execution of 381 search warrants directed at subscribers of Facebook, authorizing the district attorney and its investigators to search and seize digital information uploaded by hundreds of individual account users and stored within Facebook's servers. As a result of the fungible nature of digital information, the ability of a user to delete information instantly, and other possible consequences of disclosure, the court ordered the search warrants sealed and Facebook not to disclose the search and seizure to its users. This decision has implications for customers of hosted cloud services—such as Google Docs and Amazon's Cloud—in that the court found that, as a mere "landlord" or custodian of the customers' records, Facebook had no "legitimate expectation of privacy" in the customer's or client's records.

Information governance and records in the cloud

In 2009, ARMA International identified eight Generally Accepted Recordkeeping Principles® that could be applied to records residing in the cloud. To do so, organizations must address, among other issues, a persistent preservation strategy

and disposition practices that ensure removal of both data and metadata (Hoke 2011). The Information Governance Maturity Model, which was also developed by ARMA International (2010), describes transformational retention programs as those that, among meeting other criteria, apply retention to all information in an organization, not just official records, and a transformational disposition process as one that covers all records and information in all media. Disposition is assisted by technology and is integrated into all applications, data warehouses, and repositories.

To maintain effective information governance for records residing in a public cloud, "preservation of metadata" and "enforcement of retention periods" should be included as two key components of service agreements and contracts (Blair 2010). A private cloud can offer retention and disposition capabilities that public clouds do not. For example, Hewlett Packard (HP) Autonomy's private cloud utilizes a cloud-based suite of meaning-based governance solutions that enable the organization to enforce defensible governance in archiving, eDiscovery, compliance, data protection, and records management (HP Autonomy 2013).

Emergence of new approaches to handle research and development in the cloud

As a result of the variety of cloud models, products, services, and vendors, new approaches to retention and disposition challenges will take a variety of forms. Currently, product documentation reflects that the data centres of most cloud vendors are designed to be compliant with physical and network security, but very few of those investigated for this study offered more than limited retention and disposition functionality. Yang Tang and his colleagues (2010) propose file assured deletion (FADE) encryption technology to implement and execute retention and disposition policies. This technology will also facilitate complete data withdrawal when switching vendors. Hitachi Data Systems explains that Hitachi Content Platform (HCP) ensures retention and disposition in the cloud environment, enables litigation hold or release, and provides assurances for data segregation in a multi-tenancy environment (Ratner 2013).

Few of the cloud products or services reviewed are designed to provide long-term retention. Jan Askhoj, Shigeo Sugimoto, and Mitsuharu Nagamori (2011) suggest remodelling the Open Archival Information System (OAIS) with a platform-as-a-service (PaaS) layer, a software-as-a-service (SaaS) layer, a preservation layer, and an interaction layer to preserve records in the cloud (2011). One vendor reviewed, Preservica, offers active preservation solutions based on the OAIS model that are available in cloud-hosted and on-premise editions. Preservica supports workflows to automate bulk ingest of exported DSpace, CONTENTdm, SharePoint, and Outlook packages, advanced website harvesting, and the bulk ingest of digitized content (http://preservica.com/).

Records classification

Many electronic records systems identify the disposition status and retention period of the record at the point of capture and registration, a process that can

be linked to business activity-based classification. The classification terms are applied to the aggregation (that is, a file or container); individual records contained in the aggregation inherit the classification terms. When the classification scheme is mapped to retention requirements, inherited classification facilitates the retention and disposal of aggregations of records.

International Organization for Standardization (ISO) 16175, Module 2, 3.6.1 on Disposition Authorities specifies that an "electronic records management system must by default ensure that every record in an aggregation is governed by the disposal authority(s) associated with that aggregation." It further states that the electronic records management system must, for each aggregation, automatically track retention periods that have been allocated to the aggregation; and initiate the disposition process." According to ISO 16175, more than one disposal authority may be associated with an aggregation. If so, all retention periods specified in these disposal authorities must be automatically tracked and the disposal process initiated only after the last of all of the retention dates have been reached.

By contrast, a system that adheres to MoReq2010®'s (2010) principle that "classification determines destiny," closely associates classification with retention and disposal. Following this principle, each class has an associated disposal schedule and each record inherits its disposal schedule by default, from its class. A record is subject to no more than one disposal schedule at a time, but the default disposal schedule inherited from its class can be overridden. Each record within an aggregation may have a classification different from other records and, therefore, be due for disposal at different times. Following the principle of "bottom-up destruction," an aggregation can only be disposed of when all of its contents have been destroyed and the aggregation is closed. Aggregations need not have disposal schedules; only one disposal schedule is required—the one associated with the record (27).

In practice, probably no more than 5 percent of all digital records created or received by organizations ends up in classified aggregations in recordkeeping systems. The rest are stored, unclassified, on network drives, in email folders and, increasingly, in the cloud (Warland and Mokhtar 2012). Organizations seeking to extract knowledge from big data and legal firms seeking to locate relevant documents during a review process are investigating new technologies to make the process of sorting information less taxing; one such methodology is predictive coding. Predictive coding is

> the use of machine learning technologies to categorize an entire collection of documents as responsive or non-responsive, based on human review of only a subset of the document collection. These technologies typically rank the documents from most to least likely to be responsive to a specific information request. This ranking can then be used to "cut" or partition the documents into one or more categories, such as potentially responsive or not, in need of further review or not, etc. (Austin 2010)

The goal of predictive technology in eDiscovery remains the same as described in 2010, but technology and the view of the courts have evolved in recent years.

In a 2014 eDiscovery decision in the case of *In re Domestic Drywall Antitrust Litigation*, US District Judge Michael Baylson emphasized that counsel should use predictive coding and other "computer based programs" to help prepare their cases for trial (Favro 2014).[1] This same technology is increasingly used to automate electronic records management processes (Skamser 2013).

Records management standards and guidelines for electronic systems

The first step in understanding the challenges posed to retention and disposition in a cloud environment is to identify the functional requirements systems should possess to control retention and disposition of records hosted within the enterprise. Retention and disposition functional requirements for electronic records management were extrapolated from the following standards documents: ISO 15489 on Information and Documentation—Records Management (parts 1 and 2); ISO 23081 on Information and Documentation—Records Management Processes—Metadata for Records (parts 1, 2, and 3); ISO 16175 on the Principles and Functional Requirements for Records in Electronic Office Environments (parts 1, 2 and 3); Department of Defence (DoD) Electronic Records 5015.2 on Management Application Design Criteria Standard; and MoReq 2010®.

Records systems are designed specifically to manage records, either by hosting them in a dedicated repository or by controlling records residing in another repository. According to ISO 15489–1,

> record systems should be capable of facilitating and implementing decisions on the retention or disposition of records. It should be possible for these decisions to be made at any time in the existence of records, including during the design stage of records systems. It should also be possible, where appropriate, for disposition to be activated automatically. Systems should provide audit trails or other methods to track completed disposition actions.

The term retention in relation to electronically stored information (ESI) is the act of storing electronic information for a specified, predetermined period based on its value. The retention period is based on several factors, including the organization's operational needs; governing statues, laws, and regulations; legal issues such as the duty to preserve records for current or future audits; and historical or research needs. The organization's official policy for retention is expressed in the form of a records retention schedule and supporting procedures.

According to ISO 15489–2, any records created or captured need to have a retention period assigned so it is clear how long they should be maintained. All records within a records system should be covered by some form of disposition authority, from records of the smallest transactions to the documentation of the system's policies and procedures. Retention periods should be stated clearly and disposition triggers clearly identified. For example, "destroy x years after audit" or "transfer to the archives x years after last transaction completed." As ISO 15489–1 specifies,

> records systems should be designed so that records will remain authentic, reliable, and useable through any kind of system change, including format conversion, migration between hardware and operating systems or specific software applications, for the entire period of their retention . . .
>
> When a records system is discontinued or decommissioned, no further records may be added to the system, although they should continue to be accessible. Records may be removed from the system in accordance with retention and disposition guidelines in force, or with conversion and migration strategies. The process of discontinuing systems should be documented, as such detail will be required to maintain the authenticity, reliability, usability and integrity of records still held within that system, including conversion plans or data mapping.

Not all records reside in dedicated records systems. Some reside in electronic document and records management systems, enterprise content management systems, email systems, systems specific to the organization's business, and in a variety of cloud hosted services such as social media, cloud storage, and business applications. According to ISO 16175, Module 3, business systems must prevent the destruction or deletion of electronic records and associated metadata, alone or in conjunction with other systems, except when records are legally authorized for disposition. Business systems must also support the disposition of records in compliance with disposition authorization regimes, which includes the following:

- allowing the definition of disposition classes, which can be applied to electronic records, either through the internal functionality of the business system software or via an automatic or manual external mechanism;
- ensuring the definition of each class includes a disposition trigger, a retention period, and a disposition action;
- supporting the following disposition actions: review, export, transfer, and destruction; and
- allowing retention periods to be defined from one day to an indefinite length of time.

Additional business system functional requirements specified in ISO 16175 include allowing disposition classes to be applied to records and associated metadata and where applicable to aggregations of electronic records; recording all disposition actions in a metadata profile; allowing a disposition freeze to be placed on the electronic record, aggregations of records, and associated metadata; preventing the deletion or destruction of records subject to a disposition freeze; and providing the ability to remove a disposition freeze to a system administrator or other authorized user.

In addition, business systems must alone, or in conjunction with other systems, allow for a review of the records before the application of a disposition action. Disposition metadata can be used to trigger the automated processes and should be retained for electronic records that have been transferred or destroyed. Finally, according to ISO 16175, the system should be able to produce a report

detailing the disposition activity, identifying records that were disposed of and those that were not successfully destroyed.

Cloud service retention and disposition functional capabilities

The functional requirements described in the previous section were analysed and then categorized according to actions related to the disposition authorities. A questionnaire (shown in Table 1) was devised to evaluate cloud services. This questionnaire assumes that the system under review contains records and that both a classification scheme and disposition authority are in place.

The task of determining the existence of retention and disposition functional requirements in cloud systems is complicated by the variety of cloud service models (for example, infrastructure as a service (IaaS), PaaS, and SaaS), cloud deployment models (for example, private cloud, public cloud, and hybrid cloud), and cloud vendors (for example, IBM, AMAZON, and Rackspace). To complicate matters further, participation in public social media results in content stored in social networks in the cloud.

A 2013 Forrester consulting survey of 154 US IT decision makers at 500+ employee companies asked the question: "What best describes your organization's current use/implementation of cloud services?" (Forrester 2013). SaaS was selected by 78 percent of the respondents, storage/backup as a service by 75 percent, and disaster recovery as a service by 70 percent of the respondents. Intelligence/analytics as a service was selected by 67 percent of the respondents, and business process as a service by 62 percent. One of the services missing from the responses to the survey was records management in the cloud. When employed in a private cloud, even one provided by a third party, this is the solution that provides the greatest degree of control over an organization's records.

More than twenty cloud services, shown in Table 2, were investigated to determine the retention and disposition functionality of each.

One of the products, HP TRIM (now HP Records Manager) was deployed as a "solution-as-a-service" to make the management of government records by the Oregon secretary of state's office more transparent. The product is designed to the international records management standard, ISO 15489:2001, and to elements of ISO 16175: *Principles and Functional Requirements for Records in Electronic Office Environments* and is DoD 5015.2 certified. Housing this solution in a private government cloud hosted by Synergy Data Center and Services in Oregon required the services of a technology integrator, Arikkan Incorporated (http://www.autonomy.com/products/hp-records-manager).

Retention and disposition functionality integrated into the offerings of the remaining cloud providers is less robust, as would be expected. Extensive examination of publicly available information and personal contacts with several cloud vendors revealed answers to some of the questions on the research and development functional requirements survey. For example,

Table 1: Questionnaire of retention and disposition functional requirements (for use when evaluating specific cloud products/services)

No.	Questions	Yes	No	Do not know
	Privacy and Security Considerations			
1	Does the vendor allow independent audits of systems and processes?	☐	☐	☐
2	Is the content encrypted when in transit to the cloud?	☐	☐	☐
3	Is the content encrypted when at rest in the cloud?	☐	☐	☐
4	Are the physical servers located within a jurisdiction approved for your organization?	☐	☐	☐
5	Are the backup servers located within a jurisdiction approved for your organization?	☐	☐	☐
	Establishing Disposition Authorities			
6	What indexing capability is supported (can it accommodate customers' taxonomy for indexing)?	☐	☐	☐
7	Can retention periods be applied?	☐	☐	☐
8	Can destruction be automated?	☐	☐	☐
	Applying Disposition Authorities			
9	Can a disposition authority (retention and disposition specifications) be applied to aggregations of records?	☐	☐	☐
10	Can records be locked down for viewing only?	☐	☐	☐
11	Can records be retained indefinitely?	☐	☐	☐
12	Can records not in an aggregation be destroyed at a future date?	☐	☐	☐
13	Can records not in an aggregation be transferred at a future date?	☐	☐	☐
	Executing Disposition Authorities			
14	Can records be deleted according to the retention/disposition schedule?	☐	☐	☐
15	Can backups be deleted according to the retention/disposition schedule?	☐	☐	☐
16	Are users alerted to conflicts related to links from records to be deleted to other records aggregations that have different records disposition requirements?	☐	☐	☐
17	If more than one disposal authority is associated with an aggregation of records, can these multiple retention requirements be tracked to allow the manual or automatic lock or freeze on the process (for example, freeze for litigation or freedom of information request)?	☐	☐	☐
	Documenting Disposal Actions			
18	Are disposal actions documented in process metadata?	☐	☐	☐
19	Can all disposal actions be automatically recorded and reported to the administrator?	☐	☐	☐
	Reviewing Disposition			
20	Are electronic aggregations presented for review along with their records management metadata and disposal authority information so both content and records management metadata can be reviewed?	☐	☐	☐
21	Can records be marked for destruction, transfer, and further review?	☐	☐	☐
22	Are all decisions made during review stored in metadata?	☐	☐	☐
23	Can the system generate reports on the disposition process?	☐	☐	☐
24	Is the ability to interface with workflow facility to support scheduling, review, and export transfer processes provided or supported?	☐	☐	☐
	Integration			
25	Is the metadata schema compatible with other systems, such as enterprise content management or records management systems?	☐	☐	☐

Table 2: Cloud services investigated for this study

Cloud Storage	RM Software and add-ons	IaaS	Litigation Support and E-Discovery
Dropbox for Business	Collabware	Century Link (Tier3)	NextPoint
Egnyte	Gimmal	GoGrid	CloudNine
One Drive for Business	HP Trim	Rackspace	
Archiving Solutions	**Collaboration / Content Mgmt**	**Long-term Digital Preservation**	**Backup and Data Protection**
ArchiveSocial	SharePoint Online	Archivematica	CrashPlan
Google Vault (email and chats)	Office 365/ Exchange/Linc Online	Preservica	HP Autonomy Live Vault
Symantec Enterprise Vault			
Smarsh			

- Rackspace (http://www.rackspace.com/) provides cloud-enabled managed hosting of the public and private clouds it designs, builds, and runs for clients. Some of the information gathered that applies to retention and disposition relates to its email hosting solution. Nine copies of each message are held across multiple data centres. Users have access to archived email without having to ask the IT team. Users can locate and recover deleted emails.
- Smarsh (http://www.smarsh.com/) provides archiving and compliance solutions to archive email, social media content from public and private social platforms, internal communications, mobile communications, and web content. Smarsh supports e-discovery searches or more advanced supervision workflow, can automate and implement legal holds and retention policies, and enforces internal governance policies for recordkeeping, supervision, and data protection. Rules can be created and then configured to take automatic action (for example, flag, classify, delegate, and apply a legal hold or a retention policy) on messages that match the criteria.
- CrashPlanPro (http://www.code42.com/business/) provides backup services for business. CrashPlan works across platforms and operating systems to backup data automatically. Users can restore data to any device on their own. Data are secured from end to end, encrypted at the source, during transit and at rest (in storage). The data can be stored in a private cloud within the enterprise or hosted in a managed private cloud, public cloud, or hybrid cloud. Administrators can enforce data retention policies, implement legal holds, specify backup scheduling, and adjust security settings.

Summaries of all of the cloud services evaluated will be published in a separate report by the InterPARES Trust project team in late 2015.

Analysis of data gathered related to cloud services

Information gathered using the questionnaire designed to evaluate cloud services revealed the following information. Questions 1–5 relate to vendor services.

Approximately 57 percent of the cloud services encrypt content residing in the cloud, and 71 percent provide encryption for content while in transit. Approximately 50 percent allow independent audits of systems. Approximately 38 percent have physical servers located within a jurisdiction approved for the client, and approximately 33 percent have backup servers located within an approved jurisdiction.

Questions 6–8 relate to establishing disposition authorities. The cloud services explored did not refer to disposition authorities, but 71.4 percent allow retention periods to be applied. Destruction is automated in 47.6 percent of the services. Indexing capability is present in 61.9 percent of the cloud services studied.

Questions 9–13 relate to applying disposition authorities and locking down records for view only. Less than half (47.6 percent) of the services allow a disposition authority (retention and disposition specifications) to be applied to aggregations of records. Only 52.4 percent of cloud services allow records that are not in an aggregation (individual records) to be destroyed (42.8 percent) or transferred (42.8 percent) at a future date.

Questions 14–17 relate to executing disposition authorities; the dedicated records management solution as a service (HP Trim) and add-ons for SharePoint (Gimmal and Collabware) meet all of these requirements. Those that provide e-discovery or compliance services allow for the deletion of records and backups according to a retention/disposition schedule (disposition authority) and allow legal holds to be imposed. The responses in this section reveal that 76.1 percent allow records to be deleted according to a retention/disposition schedule, and 57.1 percent allow backups to be deleted according to the retention and disposition schedule. However, only 9.5 percent of the services alert users to conflicts related to links from records to be deleted to other records aggregations that have different retention requirements, and 33.3 percent allow multiple retention requirements to be tracked to allow the manual or automatic lock or freeze on the disposition process if more than one disposal authority is associated with an aggregation of records.

Questions 18–19 relate to documenting disposal actions. This functionality is rarely mentioned since the model of most cloud providers focuses on retention of content of their clients and not disposition. However, 57.1 percent of the services document disposal actions in process metadata, and 57.1 percent automatically record disposal actions and report them to the administrator. In some cases, the metadata exported is descriptive and does not include operational metadata added while in the custody of the cloud provider.

Questions 20–24 relate to reviewing disposition. Dedicated records management solutions will possess the functionality that allows electronic aggregations, their records management metadata, and disposal authority to be reviewed and records to be marked for destruction, transfer, or further review; they will also store decisions in metadata. Most other systems will generate reports, and a few can also interface with a workflow facility. Only 19 percent of the solutions reviewed preset electronic aggregations, their metadata, and disposal authority information to be reviewed; 28.6 percent allow records to be marked for

destruction, transfer, or further review; 23.8 percent store all decisions made during the review in metadata; 61.9 percent provide system-generated reports on the disposition process; and 38 percent provide the ability to interface with a workflow facility to support scheduling, review, and export transfer processes.

Question 25 is related to integration. At least one of the vendors expressed frustration when discussing the metadata schema used in their products since no one industry standard exists. Only 33.3 percent of the services reviewed indicated they use a metadata scheme compatible with other systems, such as enterprise content management systems or records management systems. In some instances, third-party providers develop connectors that allow integration of cloud services with other products. One example is a connector called Vega Unity available from Vega, a consulting firm, to merge Salesforce Cloudbase with ECM repositories, file systems, databases, and workflow systems. In another case, Preservica includes multiple connectors to allow content to be ingested from ContentDM, DSpace, Outlook, Lotus Notes, and SharePoint.

Records and information professional user survey

The third and final data gathering exercise involved creating a web-based survey and inviting records and information managers to participate. The survey was opened from 9 February to 29 March 2015, and 168 useable responses were collected.

General participant information

Records managers comprised 60.84 percent of the respondents, followed by information governance professionals at 10.24 percent. Business executives, archivists, and information technology specialists each made up 2.41 percent of the total respondents, followed by information officers at 1.81 percent and legal professionals at 0.60 percent.

The majority of respondents, 37.13 percent, work in the government sector followed by professional and technical services at 8.98 percent and finance and industry at 8.38 percent. Additional industries represented are education (5.39 percent); mining, quarrying, and oil and gas extraction (5.39 percent); construction and manufacturing (4.19 percent); healthcare (2.99 percent); wholesale trade/retail trade (1.80 percent); and media arts and entertainment (0.60 percent). Organizations with more than 5,000 employees made up 26.67 percent of the respondents followed closely by those with 1,000 to 5,000 at 24.24 percent. The remaining 49.09 percent of respondents were employed in organizations with less than 1,000 employees.

Responses of current cloud service users

Of the 168 respondents who participated in the survey, ninety-seven (57.74 percent) said their organization used cloud services, forty (23.81) said it did not, twelve (7.14 percent) admitted they did not know, one (0.60 percent) declined to answer, and eighteen (10.71 percent) selected the "other" option. Other responses included statements that such use is not intentional, the cloud

is used on a limited basis, and the organization is currently considering cloud services.

The use of cloud products and services is a recent phenomenon according to the respondents, 56.82 percent of which indicated their organization had engaged in such use between one and three years, followed by 13.64 percent who have been employing cloud products and services less than one year. Only 25 percent have employed cloud services more than three years.

The types of cloud models in use vary. A private cloud is used by the majority of respondents (36.14 percent), followed by a hybrid model comprised of a private third-party hosted cloud and a public cloud (19.28 percent), a hybrid model comprised of a private enterprise hosted cloud and a public cloud (18.07 percent), a public cloud (12.05 percent), a government cloud (8.43 percent), and a community cloud (2.41 percent). The responses to questions related to general retention and disposition issues are illustrated in Figures 1–4.

When asked if the organization had performed any dispositions on its content in the cloud, the majority (53.75 percent) responded no, 27.5 percent did not know, 1.25 percent declined to answer, and only 17.5 percent stated yes. One reason that disposition may be problematic for these respondents is the fact that 49.37 percent did not include retention and disposition considerations in the initial decision to use specific cloud services and another 20.25 percent did not know if such considerations were made.

Figure 1: Percentage of respondents indicating their organization employs cloud services that have a retention and disposition policy in place

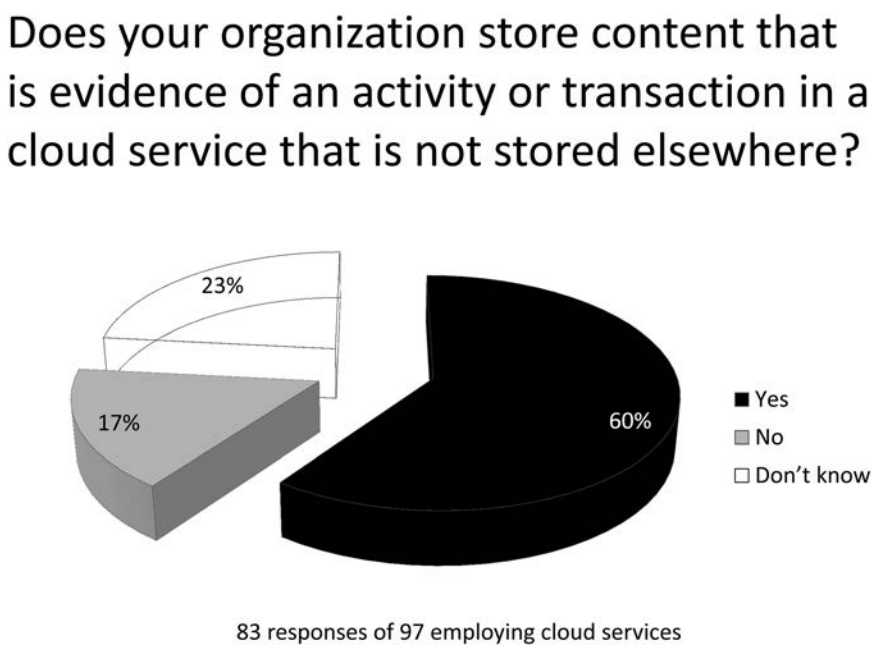


Figure 2: Percentage of respondents indicating their organization employs cloud services that store records in the cloud not stored elsewhere

Of the eighty individuals responding to the question whether vendor terms and conditions were consistent with their organization's goals and objectives for retention and disposition, 31.25 percent answered yes, 17.50 percent answered no, and 51.25 percent stated they did not know or declined to answer. The comments of two respondents indicate a position that could be taken to mitigate risk related to retention and disposition. Vendors that do not support retention and disposition are not considered.

Many of the survey questions related to disposition authorities and actions were replicas of those asked of cloud vendors on the questionnaire shown in Table 1. Responses to the questions are summarized in the following list.

- *Privacy and security considerations.* Almost 40 percent of the cloud vendors allow independent audits of their systems and processes. Over 49 percent state content is encrypted when in transit to the cloud, and over 40 percent state that content is encrypted when at rest in the cloud. Physical servers are located within a jurisdiction approved by the client for more than 53 percent of the organizations, and backup servers are located within an approved jurisdiction for more than 50 percent.
- *Establishing disposition authorities.* Respondents indicated the following indexing capabilities supported by the cloud services they employ: metadata schema (50 percent), document naming conventions (45.16 percent), classification codes (38.81 percent), taxonomies (29.03 percent), and retention periods (24.19 percent).

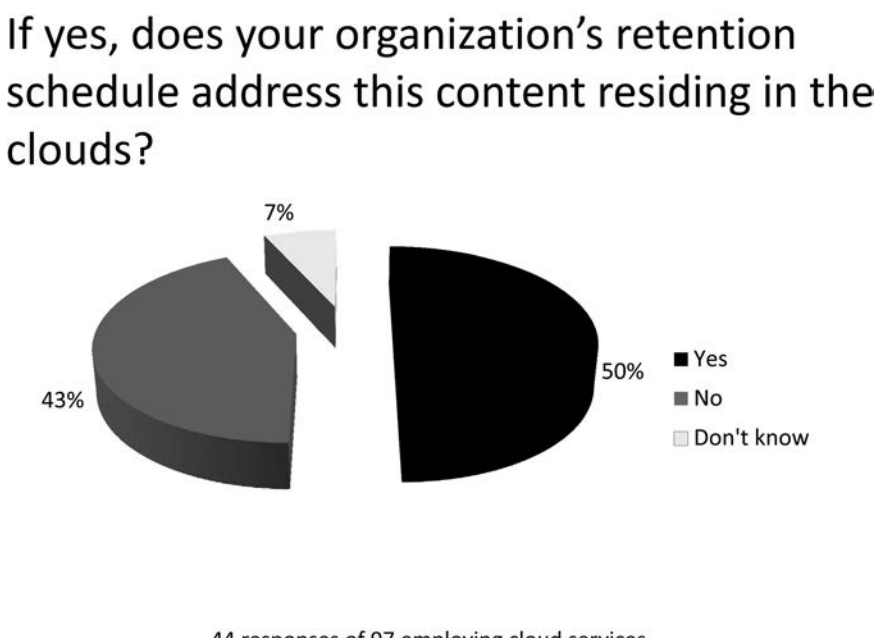


Figure 3: Percentage of respondents indicating their organization employs cloud services that address records in the cloud within their retention schedules

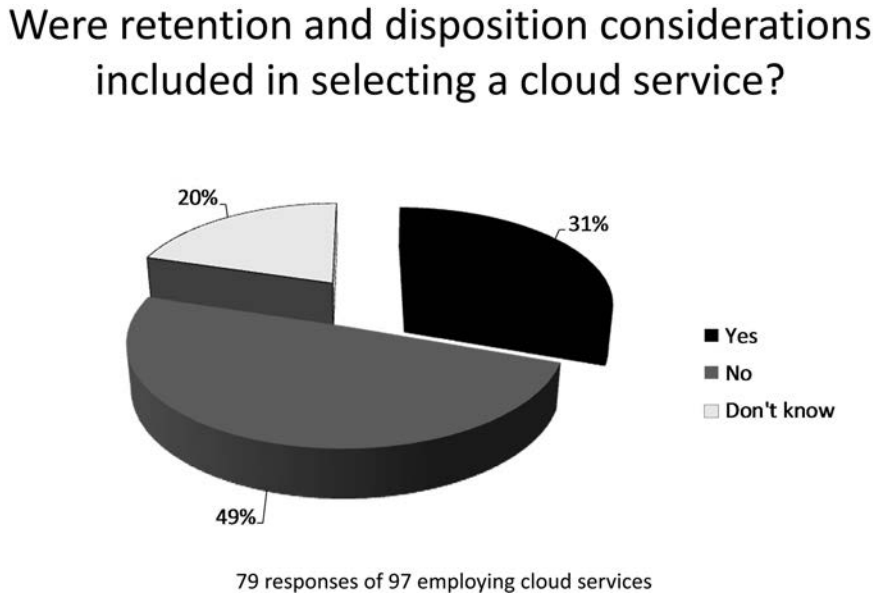


Figure 4: Percentage of respondents indicating their organization employs cloud services that include retention and disposition considerations when selecting cloud services

- *Applying disposition authorities*. Retention and disposition specifications can be applied to aggregations of records according to 25 percent of the respondents to this question, and records can be locked down for viewing according to 34.78 percent of the respondents. Records can be transferred to other systems from 56.16 percent of the cloud services in use.
- *Executing disposition authorities*. More than 45 percent of respondents indicated that records can be deleted according to a retention/disposition schedule, but only 24.29 percent indicated that backups could be deleted according to the same schedule. Twenty-eight percent stated that the administrator could change/override the disposition action, but another 59.70 percent did not know if this was possible for their cloud service provider.
- *Documenting disposal actions*. Almost 15 percent of respondents stated that disposal actions are documented in processes metadata; however, slightly over 66 percent did not know if this was the case in their organization. Disposal actions could be automatically recorded and reported to the administrator according to 21.74 percent of the respondents, but almost 58 percent did not know if this was the case for their organization.
- *Reviewing disposition*. When asked if disposition notifications are presented to the administrator to allow a review of content and records management metadata before disposition, 22 percent said yes, but 55.88 percent did not know. Almost 12 percent of the respondents stated that all decisions made during review are stored in metadata, but 73.91 percent did not know if this was the case within their organization. The system could generate reports on the disposition process for 18.84 percent of the respondents, but 60.87 percent of the respondents did not know if this was possible within the cloud services employed by their organization.
- *Integration*. When asked if the metadata schema was compatible with other systems, 29.17 percent said yes, but 57.75 percent did not know. In response to a similar question, 26.39 percent of respondents stated it was possible to integrate the cloud provider's system with other systems, such as an ECM or a records management system, but 63.89 percent did not know if this was possible for their organizations.

Conclusion

When considering the implications of the use of cloud services on the organization's retention and disposition process, the best approach is a strategic one. Begin the way you would if all records and information were stored on premise. Understand the business goals that can be achieved by using cloud services. Then consider the records and information generated. Develop a method to appraise the value of all information. Determine a process to classify information to assign retention periods and develop a disposition authority (retention and disposition schedule).

Once you understand the business goals and cloud services to be used (or already in use), investigate each of the cloud options using a questionnaire

similar to that introduced in Table 1. Analyse the data gathered. Consider carefully the potential risks and then decide if you will accept them, mitigate them, or avoid them completely. One way to mitigate risk presented by posting content to social media networks, for example, is to employ the services of a cloud-archiving solution to capture and manage the social media content in a way that enables your organization to comply with governing laws and regulations.

Records residing in a cloud environment must be captured, managed, preserved, and made accessible according to the organization's records management policy for "all" records. Functional requirements presented in de jure and de facto standards such as ISO 15489 and DoD 5015.2 apply to systems used to manage records whether they reside within the enterprise or in the clouds. However, cloud vendors may not meet all of the organization's retention and disposition requirements.

Even when records are under the control of a cloud vendor, the organization is ultimately responsible and accountable for managing its own records. Guidance for managing records is available from various sources, including professional associations. It is the client's responsibility to determine if a specific cloud provider meets their needs. This task is complicated by the services offered (for example, SaaS, IaaS, and PaaS), cloud environments (for example, public, private, and hybrid), and vendors (for example, IBM, Amazon, Microsoft, and Rackspace). When surrendering control of records and information to a cloud service provided by a third party, due diligence must be paid to identifying the appropriate mix of cloud services and providers.

Further research

The InterPARES Trust retention and disposition in a cloud environment research project is ongoing. Three separate research methods were employed to gather data related to retention and disposition functionality from existing standards and guidelines, from cloud vendor publications and interviews, and through a survey of records and information professionals. This study provides a valuable glimpse into the current landscape related to retention and disposition functional requirements offered within various cloud services. However, there are limitations to this study. The major issue is that the gap analysis is inconclusive for these reasons:

- *Cloud vendor questionnaire*. As a result of the reluctance or inability of cloud vendors to provide answers to questions about retention and disposition functionality within their offerings, capabilities may exist that have not yet been identified. In addition, since we first initiated this study, the vocabulary of many cloud vendors has broadened to include records management terms, retention and disposition features are now being offered by some cloud vendors, and technologies to integrate some cloud solutions with enterprise content management and/or records management systems are being developed. A similar study may reveal a very different landscape in the near future.

- *Records and information management user survey*. Another concern is the number of "don't know" responses to many of the questions in the records and information management user survey. For example, although 28.17 percent of the seventy-one respondents to the question about the existence of metadata schema compatible with other systems stated yes, 57.75 percent did not know. When asked if destruction can be automated, only 19.44 percent said yes, but 62.5 percent said they did not know. And when asked if backups could be deleted according to the retention/disposition schedule, 24.29 percent said yes, but 64.29 percent did not know. Several of the responses indicated a passive stance by records managers in that they stated they were not invited or did not have a seat at the table. With the current emphasis on information governance and the role records managers can play within the information governance process, a similar study in another year or two may provide fewer "don't know" responses.

Investigation into functionality built into cloud services is ongoing as new products and services are introduced and existing products and services are enhanced. In addition, case studies will be developed to describe successful cloud implementations models that enable organizations to retain and dispose of cloud-based records according to retention and disposition functional requirements.

Note

1. *In re Domestic Drywall Antitrust Litigation*, MDL No. 2437, 13-MD-2437 (E.D. Pa. 2014), http://www.paed.uscourts.gov/documents/opinions/14d0375p.pdf.

References

ARMA International. 2010. *ARMA International Maturity Model for Information Governance*. http://www.arma.org/r2/generally-accepted-br-recordkeeping-principles/metrics/metrics-retention.

Askhoj, Jan, Shigeo Sugimoto, and Mitsuharu Nagamori. 2011. "Preserving Records in the Cloud." *Records Management Journal* 21 (3): 175–87. http://dx.doi.org/10.1108/09565691111186858.

Austin, Doug. 2010. "eDiscovery Trends: What the Heck Is "Predictive Coding"?" eDiscovery Daily Blog, 15 December. http://www.ediscoverydaily.com/2010/12/ediscovery-trends-what-the-heck-is-predictive-coding.html.

Babcock, Charles. 2013. "Gartner: 50% of Enterprises Use Hybrid Cloud by 2017." Information Week, 1 October. http://www.networkcomputing.com/cloud-infrastructure/gartner-50–of-enterprises-use-hybrid-cloud-by-2017/d/d-id/1111769?

Bashir, Masooda N., and Jay P. Kesan. 2011. "Privacy in the Cloud: Going Beyond the Contractarian Paradigm." Annual Computer Security Applications Conference, Orlando, FL. https://acsac.org/2011/workshops/gtip/Bashir.pdf.

Blair, Barclay T. 2010. "Governance for Protecting Information in the Cloud." In *Making the Jump to the Cloud?: How to Manage Information Governance Challenges*, 1–4. Overland Park, KS: ARMA International. http://www.arma.org/docs/hot-topic/makingthejump.pdf.

Dutta, Amab, Chao Alex Peng Guo, and Alok Choudhary. 2013. ''Risks in Enterprise Cloud Computing: The Perspective of IT Experts.'' *Journal of Computer Information Systems* 53 (4): 39–48.

Favro, Philip. 2014. *Breaking News: Court Touts the Importance of Predictive Coding in Preparing for Trial*. Recommind, 20 May. http://www.recommind.com/blog/2014/05/20/breaking-news-court-touts-importance-predictive-coding-preparing-trial.

Forrester Consulting. 2013. *Building for the Future: What the New World of Cloud IT Means for the Network*. Cambridge, MA: Forrester Consulting.

Franks, Patricia C., and Alan Doyle. 2014. ''Retention and Disposition in the Cloud: Do You Really Have Control?'' *Proceedings of International Conference on Cloud Security Management ICCSM-2014*, The Cedars, University of Reading, Reading, UK, 52.

Géczy, Peter, Noriaki Izumi, and Kōiti Hasida. 2013. ''Hybrid Cloud Management: Foundations and Strategies.'' *Review of Business and Finance Studies* 4 (1): 37–50.

Grounds, Alison A., and Benjamin W. Cheesbro. 2013. ''Cloud Control: eDiscovery and Litigation Concerns with Cloud Computing.'' *Computer and Internet Lawyer* 30 (9): 23–31.

Hoke, Gordon E. J. 2011. *Challenges to Governing Remote Information*. Baseline, 4 October. http://www.baselinemag.com/c/a/IT-Management/Challenges-to-Governing-Remote-Information-709978/.

HP Autonomy. 2013. *Best Practices for Cloud-Based Information Governance*. HP Autonomy. http://www.informationweek.com/whitepaper/Infrastructure/Network-Systems-Management/making-the-move-to-the-cloud-best-practices-adv-wp1347981072?articleID = 191705703.

Ion, Iulia, Niharika Sachdeva, Ponnurangam Kumaraguru, and Srdjan Čapkun. 2011. ''Home is Safer Than the Cloud!: Privacy Concerns for Consumer Cloud Storage.'' Symposium on Usable Privacy and Security, Pittsburgh, PA, July 2011. https://www.vs.inf.ethz.ch/publ/papers/iion-cloud-2011.pdf. http://dx.doi.org/10.1145/2078827.2078845.

MoReq2010®. 2010. *Modular Requirements for Records Systems*, versioner 1.1., DLM Forum Foundation. http://moreq2010.eu/pdf/moreq2010_vol1_v1_1_en.pdf.

Ratner, Michael. 2013. *Introduction to Object Storage and Hitachi Content Platform*. Santa Clara, CA: Hitachi Data Systems. http://www.hds.com/assets/pdf/hitachi-white-paper-introduction-to-object-storage-and-hcp.pdf.

Skamser, Charles. 2013. *Predictive Coding is Expanding to Records Management and Information Governance*. http://ediscoverytimes.com/predictive-coding-is-expanding-to-records-management-and-information-governance/.

Supreme Court of the State of New York. 2013. *In Re Search Warrants Directed to Facebook, Inc*. http://s3.documentcloud.org/documents/1209711/court-order-on-facebook-search-warrants.pdf.

Tang, Yang, Patrick P. C. Lee, John C. S. Lui, and Radia Perlman. 2010. ''FADE: Secure Overlay Cloud Storage with File Assured Deletion.'' *Security and Privacy in Communication Networks* 50: 380–97. http://dx.doi.org/10.1007/978-3-642-16161-2_22.

Warland, Andrew, and Umi Asmá Mokhtar. 2012 ''Can Technology Classify Records Better Than a Human?'' Image and Data Manager, 19 December. http://idm.net.au/article/009392-can-technology-classify-records-better-human.

Weins, Kim. 2014. ''Cloud Computing Trends: 2014 State of the Cloud Survey.'' Rightscale, 2 April. http://www.rightscale.com/blog/cloud-industry-insights/cloud-computing-trends-2014-state-cloud-survey#Hybrid-Cloud-Is-the-Strategy-of-Choice.