Retention & Disposition in the Cloud:

Mission Critical and/or Mission Impossible?

Patricia C. Franks, PhD, CA, CRM, IGP
Archives 2015
August 22, 2015

Tweet about session 707: #saa15

Sometimes it can seem like both!

Types of Clouds

Type	Purpose	Model	Typical User	Examples
Enterprise Clouds	Typically replace existing compute, network and storage systems	Public Private Hybrid	Large enterprises Mid-size companies	Amazon AWS Rackspace GoGrid Savvis Teremark
Overlap				
Business Service Clouds	Provide on-demand business services	Public	Large enterprises SMBs	Salesforce.com ZOHO Google Apps Netsuite DropBox
Overlap				
Consumer Clouds	Provide lifestyle convenience	Public	You and me	Flickr, Picassa Pandora, Spotify iTunes, iCloud OpenTable Netflix, Flixster YouTube, Vimeo

Robert Keahey (2012) - http://www.slideshare.net/rkeahey/cloud-computing-and-big-data

"Think simple" as my old master used to say – meaning reduce the whole of its parts into the simplest terms, getting back to first principles.

QUOTEHD.COM

Frank Lloyd Wright
American Architect

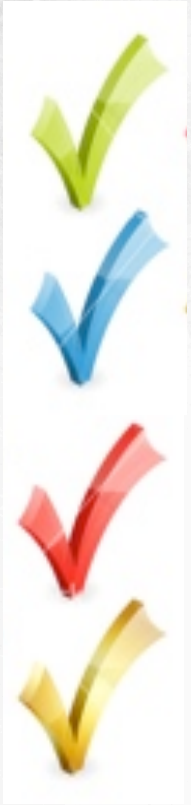
1. Core functional requirements necessary to implement retention and disposition actions

2. Functionality of several cloud models based on user experiences

3. Best practices for developing a defensible retention and disposition strategy for records residing in the clouds.

InterPARES Trust

Retention & Disposition System Requirements

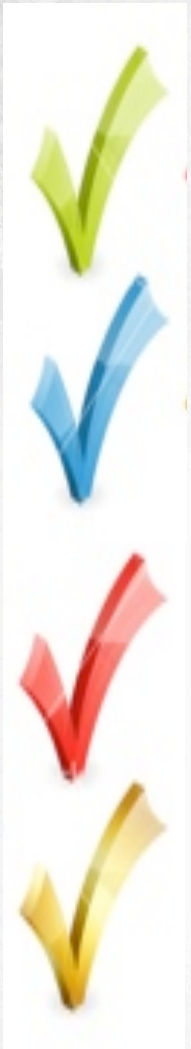


- Facilitate & Implement R&D Decisions
- Anytime in the existence of records
- Activated automatically
- Provide audit trails

~ISO 15489-1 2001, p. 10

InterPARES Trust

Defensible Disposition



Allow definition of retention periods (1 day to indefinite)

Allow definition of disposition classes

Include a disposition trigger, a retention period, and a disposition action

Support disposition actions: review, export, transfer, and destruction.

InterPARES Trust

~ISO 16175, Module 3, 53-54

Retention & Disposition Functional Requirements

Questionnaire for use when evaluating specific cloud products/services

No.	Questions	Yes	No	Don't 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 (ex. 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, 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?			

Cloud Vendors Under Review

1. Amazon Web Services
2. ArchiveSocial
3. CenturyLink Cloud/Tier 3
4. Cloud 9 Discovery
5. Crashplan
6. Egnyte
7. Dropbox
8. GoGrid
9. Google Apps for Business/ include Google Cloud
10. HP Records Manager (formerly HP Trim)
11. Microsoft One Drive
12. Office 365
13. SharePoint
14. MS SP Add-on Gimmal
15. MS SP Add-on Collabware
16. NextPoint
17. Preservica
18. Archivematica
19. Rackspace
20. Smarsh
21. Symantic

InterPARES Trust

Cloud Storage	**RM Software and add-ons**	**Infrastructure as a Service**	**Litigation Support & E-Discovery**
Amazon S3 & Glacier Dropbox for Business Egnyte One Drive for Business	Collabware Gimmal HP Records Manager (formerly HP Trim)	Amazon Web Services Century Link (Tier3) GoGrid Rackspace	NextPoint CloudNine
Archiving Solutions	**Collaboration / Content Mgmt.**	**Long-term Digital Preservation**	**Backup & Data Protection**
ArchiveSocial Google Vault (email & chats) Symantec Enterprise Vault Smarsh	SharePoint Online Office 365/ Exchange/Linc Online	Archivematica Preservica	CrashPlan HP Autonomy Live Vault

Oregon rides cloud to statewide records management system

HP TRIM (HP Records Manager)

Records Management Solution-as-a-Service (private, government cloud)

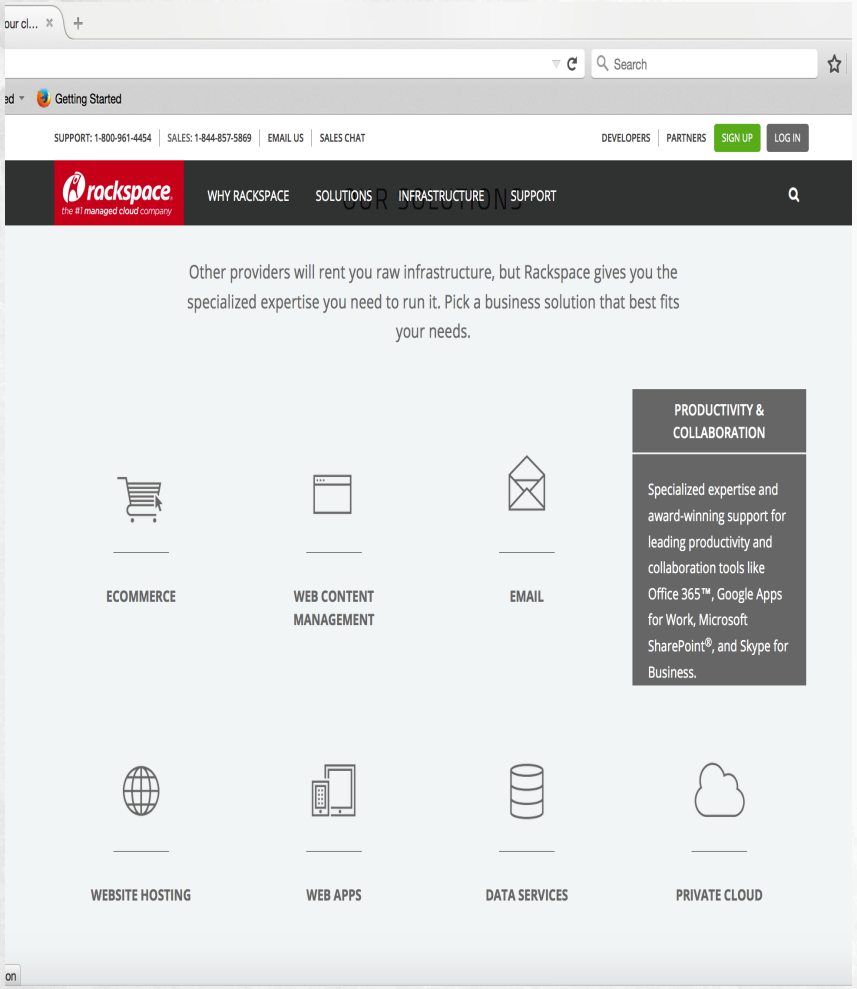
Designed to the international records management standard, ISO 15489:2001, and to elements of ISO 16175 - *Principles and Functional Requirements for Records in Electronic Office Environments;* DOD 5015.2 certified.

Hosted by Synergy Data Center and Services; integration services provided by technology integrator, Arikkan Incorporated.

InterPARES Trust

Rackspace

Infrastructure-as-a-Service
(host public & private cloud solutions)



Retention & and disposition functionality depends on cloud solution implemented. For example, an email hosting solution reviewed retains 9 copies of each message across multiple data centers. Users have access to archived email without having to ask the IT team. Users can locate and recover deleted emails.

InterPARES Trust

smarsh

Products Industries Resource Center About Us

OUR PRODUCT SUITE — **BUILT FOR COMPLIANCE AND E-DISCOVERY**

An Archiving Solution

WE ARCHIVE
EVERYTHING

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.

6 Essential Functions of Cloud Hosted Digital Preservation

Out-of-the-box Cloud hosted Digital Preservation offers great benefits, allowing you to get started quickly without needing to purchase and manage servers or have local IT resources. However, make sure any solution you choose is not just proven, well maintained, well supported and has an active user community – but also provides these 6 *essential* functions.

1. Active Preservation

✔ Preservation Planning and Action is a core requirement of OAIS. Being able to proactively manage and migrate content to newer file formats overtime ensures it has the highest chance of being readable and useable across multiple technology refresh cycles into the future. Simpler approaches such as only performing migration at ingest (normalization) forces file format choices to be made today: Active Preservation

2. Preservation & Access

✔ Choosing a solution that includes Digital Preservation and Public Access in one saves time, money and duplicated content. Having these two functions integrated is more important in the Cloud. Your Access experience should also ensure content can be viewed in an optimum way on multiple devices without needing to download across the internet: Universal Access

3. Upload of Large Files

✔ The easy and reliable upload and ingest of content, especially large files or collections of files (10s of GBs) is a critical function especially when working with a Cloud hosted Digital Preservation solution. Having a way to perform this as a background task from a network server also means that content contributors do not need to tie up their own PCs during upload and ingest.

4. Storage Flexibility

✔ Choosing a solution that allows you to configure (and automate) where different content types get stored enables you to optimize your storage needs and costs. For example, using fast Amazon S3 for frequently accessed presentation content and lower-cost Amazon Glacier for large infrequently accessed preservation content.

5. User Access Control

✔ Make sure any solution you choose has an easy way to control who gets to see what content, and what actions they can then perform. This is especially important if you plan to provide access to different internal and external communities, or are preserving confidential or sensitive information.

6. Lots of Connectors

✔ Choose a solution that has open interfaces and out-of-the-box connectors that will enable you to synchronize and automate the ingest of content from different systems across your organization. Preservica, for example, includes connectors for commonly used Catalogs, Digital Asset Libraries, Content/Records Management and eMail systems.

Two systems in one: a full suite of OAIS compliant workflows to actively preserve your digital content and a public access module to securely share and showcase your content and collections. (ISO 14721:2012)

Stores metadata in fast Amazon RDS and digital content in either low-latency Amazon S3 or lower-cost Amazon Glacier for infrequently accessed content.

Best Practices – A Strategic Approach

- Understand the use of cloud services within the organization.
- Become involved in the selection of a cloud provider that will help the organization/business unit achieve its goals.
- Identify content stored in the cloud that is evidence of an activity or transaction and not stored elsewhere.
- Apply retention and disposition requirements to that cloud content either within the cloud environment or by moving the content to an ECMS or ERMS or other repository.

InterPARES Trust

Best Practices – A Strategic Approach (cont.)

- Evaluate the cloud services employed using a tool such as the checklist illustrated earlier related to issues of privacy and security, retention and disposition, and integration.

- Identify the risks inherent in the choice of cloud provider(s) and take steps to mitigate, avoid, or accept those risks.

- Implement a review process to determine the effectiveness of the cloud initiative and revise it if/when necessary.

InterPARES Trust

THANK YOU!

Dr. Patricia C. Franks, CA, CRM, IGP
Professor & MARA Program Coordinator
School of Information
San Jose State University
patricia.franks@sjsu.edu

InterPARES Trust