Missing Links: What Happens to the Chains of Custody and Preservation in the Cloud?

Kenneth Thibodeau

August 14, 2014

2014 Archives ★ Records
ENSURING ACCESS

Session 109: Ethics, Provenance, Metadata: Trust and Recordkeeping in the Cloud

Chain of Custody

- "1. Records · The succession of offices or persons who have held materials from the moment they were created. - 2. Law · The succession of officers or individuals who have held real evidence from the moment it is obtained until presented in court."

- "In both senses, the ability to demonstrate an unbroken chain of custody is an important test of the authenticity of records or evidence."

Richard Pearce-Moses, Glossary of Archival Terminology

InterPARES Trust

Chain of Custody

"An Internet Best Current Practices for the Internet Community"

- "You should be able to clearly describe how the evidence was found, how it was handled and everything that happened to it.

- "The following need to be documented

  - "Where, when, and by whom was the evidence discovered and collected.

  - "Where, when and by whom was the evidence handled or examined.

  - "Who had custody of the evidence, during what period. How was it stored.

- "When the evidence changed custody, when and how did the transfer occur (include shipping numbers, etc.)."

IETF, RfC 3227: Guidelines for Evidence Collection and Archiving. 2002.

InterPARES
Trust

Chain of Preservation

- "A system of controls that extends over the entire lifecycle of records and ensures their identity and integrity in any action that affects the way the records are represented in storage or presented for use." InterPARES Glossary

- Origin (disclosure):

  - Articulated by the Preservation Task Force in the InterPARES 1 project

  - Developed into an IDEF(0) model in InterPARES 2

- Asserts that in the case of electronic records, regardless of the chain of custody, either intentionally or inadvertently things may happen that result in corruption or loss of records.

- Therefore, there must be an unbroken process extending over the life of the records that, at the very least, enables an assessment of whether the records remain uncorrupted and ideally protects them against loss or corruption

- Minimum requirements for the Chain of Preservation are set out in the InterPARES 1 Baseline and Baseline Requirements

InterPARES Trust

How Should We Look at the Cloud?

Opportunities for Better Service Delivery and Improved Recordkeeping and Information Management via Cloud Computing

- cost savings

- reduced pressure on ICT departments to provide ever increasing storage capacity

- access to services outside normal office environments

- better opportunities for collaboration with geographically dispersed users

- potential opportunities for greater automation of recordkeeping as part of business processes, and

- more time for ICT personnel to devote to other issues where server maintenance and related tasks are lessened.

  - Council of Australasian Archives and Records Authorities. Advice on Managing The Recordkeeping Risks Associated With Cloud Computing. Version 1.0 29 July 2010.

Records Management Risks in Could Computing

- Legality of sending or storing of records outside the original jurisdiction

- Provider's compliance with laws or standards of the record-creating jurisdiction

- Imposition of laws or other requirements of the jurisdiction where records are stored

- Possibilities of unauthorised access to records

- Record destruction or loss

- Damage to the evidential value of records

  - Council of Australasian Archives and Records Authorities. Advice on Managing The Recordkeeping Risks Associated With Cloud Computing. Version 1.0 29 July 2010.

InterPARES Trust

Risks of Destruction or Loss of Records Specific to Cloud Computing

- The service provider may go out of business or be taken over by another company, which may not choose to honor the service agreement

- A person in another jurisdiction may access, claim ownership or take control of the records

- Records may not be returned upon request or at conclusion of the contract, or returned only on payment of a large fee

- Inadequate backup and restoration arrangements as a result of cost cutting by the service provider

- Changes in the provider's hardware or software that are not compatible with that of the records creator or archives

- Unauthorized disposal of records.

  - Council of Australasian Archives and Records Authorities. Advice on Managing The Recordkeeping Risks Associated With Cloud Computing. Version 1.0 29 July 2010.

InterPARES Trust

Risk That Evidential and informational Value of Records May Be Diminished

- Essential properties of records may be compromised by changes in

  - storage systems

  - software used to manage records

  - software used to retrieve or present records

  - management of record aggregates

  - data formats

  - metadata management

InterPARES Trust

How Can We Trust the Cloud to Preserve Records?

InterPARES Trust (ITrust) 2013-2018

a multi-national, interdisciplinary research project exploring issues concerning digital records and data entrusted to the Internet. Its goal is to generate theoretical and methodological frameworks to develop local, national and international policies, procedures, regulations, standards and legislation, in order to ensure public trust grounded on evidence of good governance, a strong digital economy, and a persistent digital memory.

URL: http://www.interparestrust.org/trust

InterPARES Trust

ITrust Participants

- Team North America

- Team Latin America

- Team Europe

- Team Africa

- Team Asia

- Team Australasia

- Transnational Team

Transnational Team

Alliance for Permanent Access to the Records of Science In Europe Network	International Federation of Red Cross and Red Crescent Societies

International Monetary Fund Archives	North Atlantic Treaty Organization	UNESCO
International Centre for the Preservation and Restoration of Cultural Property	The World Bank Archives	Barrick Gold Corporation

Team Latin America	Team Asia	
Universidad Nacional Autonoma de Mexico	State Archives Administration of the People's Republic of China	
Archivo General de la Nacion, Mexico	Renmin University of China	Anhui University
Universidad Nacional de Córdoba, Argentina	University of Gakushuin, Japan	National University of Malaysia
Arquivo Nacional do Brasil	National Archives of Malaysia	Centre for Development of Advanced Computing, India
Universidad de La Salle, Columbia	Sungkyunkwan University, Korea	Sarawak State Library
Team Africa	Team Australasia	
University of South Africa	Victoria University of Wellington, New Zealand	Queensland University of Technology, Australia

Team Europe

Mid-Sweden University	University College London
University of Amsterdam	British Library
Regione Toscana	Universita di Roma
Russian State University for the Humanities	University of Zagreb
Hacettepe University, Turkey	Israeli State Archives

Team North America

University of British Columbia	Artefactual Systems	British Columbia Information Access Operations Branch	
University of Washington	Library and Archives Canada	University of Victoria Libraries	City of Vancouver Archives
San Jose State University	City of Victoria, BC	University of Montreal	Webster University
University of Tennessee	Vancouver Police Department	Office of the Information and Privacy Commissioner for British Columbia	
McGill University	Clayton State University	University of Louisville	University of Toronto

ITrust Projects Related to Keeping Records in the Cloud

- **Retention & Disposition in a Cloud Environment**

  - 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?

- **Contract Terms for Cloud-based Record Keeping Services**

  - Contribute to understanding the degree to which contracts can be trusted to mitigate known record keeping risks.

  - Possibly develop new or revised "boiler-plate" contract terms.

- **Ensuring Trust in Storage in Infrastructure-as-a-Service (IaaS)**

  - Investigate national and international policies, standards and regulatory environments that influence trust in IaaS storage. Look into the ways the policies identify and address issues of risks of storage in IaaS.

  - Establish a cloud IaaS storage solution in the Croatian Financial Agency

InterPARES Trust

ITrust Projects Related to Keeping Records in the Cloud

- **Design Requirements for Authenticity in the Cloud**

  - What are the ideal and realized requirements for authenticity in systems with suspect or unclear controls?

  - What new metadata requirements surface from this new systems context?

  - How can we inform better systems design optimizing for the keeping and reproduction of authentic records?

- **Preservation as a Service for Trust**

  - Elaborate one or two models that will provide insight and guidance to both those who entrust records to the Internet and those who provide Internet services for the records.

  - Are requirements for the preservation of electronic records applicable to those entrusted to the Internet; do any of them need to be adapted? Are there other, special requirements for preservation of records entrusted to the Internet?

  - How can these requirements be satisfied when records are stored in cloud services?

  - Are there special requirements for records that are discovered and delivered via the Internet, even if they are not stored in a cloud? How can such requirements be implemented.

The Object Management Group Preservation Services Initiative

- The Object Management Group (OMG) is an international, open membership, not-for-profit technology standards consortium. Founded in 1989, OMG standards are driven by vendors, end-users, academic institutions and government agencies. OMG Task Forces develop enterprise integration standards for a wide range of technologies and an even wider range of industries.

    - Some OMG standards: Business Process Model and Notation, Common Object Request Broker Architecture, Knowledge Discovery Metamodel, Meta Object Facility Core, XML Metadata Interchange, Object Constraint Language, Model Driven Architecture, Unified Modeling Language

- In June 2014, ITrust proposed to the OMG the development of an OMG standard on digital preservation based on the functional requirements being defined in PaaST

- The proposal was accepted and is going forward under the OMG Government Domain Task Force

- The preservation standard will be aligned with the existing OMG Records Management Services Specification

InterPARES Trust

Challenges of Preservation as a Service for Trust

- The party responsible for preservation, aka the 'archives', lacks knowledge and control of how the party providing preservation services does things

- The archives will have difficulty, independently of the service provider, either to ascertain the state of the assets being preserved or to verify that the provider satisfies preservation requirements.

  - Information assets may go directly from the producer to the preservation service provider.

- The archives may lack relevant technological capabilities or expertise.

- As with all other options, preservation in the Cloud will suffer from the lack of proven preservation techniques for digital information.

  - Overcoming obsolescence

  - Taking advantage of progress

InterPARES Trust

Challenges of Modeling PaaST

- The model must be independent of any specific technology

- The model should accommodate different methods, not just technological solutions

- The model should be applicable across as broad a variety of contexts as possible, including:

  - Maximum heterogeneity in the types of content objects being preserved, specifically not limited to records.

  - Maximum variety in applicable directives, such as laws, regulations, standards, policies, business rules.

  - Including varying conditions of ownership, access, use, and exploitation

  - Broad variation in institutional arrangements, including preservation service agreements, and relationships between archives and both asset producers and users of preserved assets.

  - Across as broad a spectrum as possible from best practices to worst cases.

- The model must be demonstrably implementable

Reasons for using the Unified Modeling Language (UML)

- UML is a contemporary, widely used, proven method

- UML includes a family of techniques for multifaceted description of requirements and approaches, including

  - Use Cases

  - Class diagrams

  - State Transition Diagrams

  - Activity Diagrams

  - Functional requirements

  - et al.

- Requirements properly articulated in UML can be used in automatic generation of code

Potential for Implementation of PaaST

- The services approach provides broader options for implementation than specifying a system or application

- A Preservation Services specification would be a logical companion of the RMS spec

    - The OMG RMS spec addresses management of current records, but could be linked to specification of services for preservation

- The process of articulating the PaaST model in the context of developing an OMG specification will improve the model.

- OMG only undertakes development of specifications if industry members commit to implementing it.

PaaST Services

- Core Functions

  - Submission Receive Service

  - Object Characterization Service

  - Authenticity Assessment Service

  - Preservation Storage Service

  - Preservation Transformation Service

  - Dissemination Service

- Utilities

  - Notification Service

  - Rights and Privileges Service

Submission Receive Service

- The Submission Receive Service supports the transfer of information assets from producers or archives to an Internet based service provider and captures, reports and makes available attributes that determine whether assets have been successfully transferred to the party providing preservation services.

  - Package Acceptance Service

  - Package Contents Acceptance Service

Object Characterization Service

- The Object Characterization Service captures, reports and makes available attributes that characterize content objects included in a submission, including those mapping the digital objects in the submission to single items or aggregates, attributes that enable the assessment of the authenticity and the subsequent preservation of the assets at the level of authenticity assessed at the time of transfer. The services supports verification whether Content Objects correspond to related Preservation Description Information.

  - Item Characterization Service

  - Aggregate Characterization Service

Authenticity Assessment Service

- The Authenticity Assessment Service supports invocation of methods to assess authenticity of assets and captures, reports and makes attributes related to authenticity available.

  The service is neutral with respect to the authenticity assessment method or methods invoked.

  - Initial Authenticity Assessment Service

  - Preservation Authenticity Assessment Service

InterPARES Trust

Preservation Storage Service

- The Preservation Storage Service supports ensuring that assets are preserved and remain authentic. It captures, reports and makes available attributes related to the storage of records, movement in storage, and replacement and refreshment of storage media and systems, formats, and related software, such as software for metadata management.

  - Preservation Storage Integrity Service

Preservation Transformation Service

- The Preservation Transformation Service supports the invocation of transformation services on content objects and captures, reports and makes available attributes related to any transformation of the preserved content objects or of the software used to output copies of the records.

  - Data Normalization Service

  - Migration Service

  - Emulation Service

  - Change in Preservation Software Service

Dissemination Service

- The Dissemination Service supports access to and dissemination of preserved assets and captures, reports and makes available attributes about access to and dissemination of preserved records, supporting the production and certification of authentic copies.

  - ★ *Note: does not include the full range of services related to access and dissemination.*

  - Output Service

  - Authenticity Data Dissemination Service

PaaST Utilities

- <u>Notification Service</u>

  - Many activities within the scope of the model produce information that needs to be communicated to parties involved in or concerned with the preservation of the records. The Notification Service is a utility which supports such communications

- <u>Rights and Privileges Service</u>

  - Information being preserved may be subject to differential rights, privileges and restrictions. They are relevant at different times; e.g., the right of the service provider to receive restricted materials or execute transformations, the right of requestors to receive access, the right of the producer to withdraw materials, etc.