InterPARES Trust

Authentic Records Online: An Achievable Goal?

Luciana Duranti
InterPARES Trust Project Director
ITrust & CEDIF, Mid-Sweden University

National Library of Sweden, Stockholm 14 May 2014

Preservation Online

Preservation institutions are regarded as the **trusted custodians** of our documentary memory. Yet, they are beginning to entrust their holdings to Internet Providers because:

- Many of the materials they are mandated to preserve already **exist** online
- **Access** is possible from any location to anyone who can use a browser
- A trusted digital repository satisfying ISO standards as well as basic preservation requirements is not **affordable** and is often inadequate to the challenge
- The **knowledge** to deal with the digital products of complex technologies is not commonly available among information professionals and is very expensive
- Strong **protection** measures are often confused with preservation measures
- but mostly because archives are confronted with…

InterPARES Trust

A Creators’ Generational Change

Generation Y or Millennials -- Post-1981

- integration of private and public
- *produsing,* co-authoring, crowdsourcing
- co-owning, sharing
- distributed workforce vs. BYOD using multiple clouds
- media convergence
- constant connectivity
- visual language
- “liquid communication,” instantaneous impact,
- ephemeral output

InterPARES Trust

Characteristics of the Digital Output

- Digital materials are produced to be **viewed differently** based on choice of browser, application, and user preferences, or perceived preferences
- **Metadata may be constructed by any number of parties** to manipulate the behavior of retrieval systems that use it, rather than to describe the documents or other digital objects
- When the **goal is communication**, documents may exist in as many separate clouds as needed and may be scheduled to self-destruct (DSTRUX), or may never be destroyed (**involuntary permanence**)
- When the **goal is memory**, digital archives may be constructed on hard drives and scheduled to be regularly imaged (digital estate)
- The digital documents of persons, organizations and institutions may be mingled and **undistinguishable**
- Three primary challenges for preservation

Challenge 1: Reuse

While any research can be considered reuse, the reuse done before transfer to a designated preserver is different

- Reuse is often *remix*, a practice which results in derivative works that substantively change the intent and context of the appropriated material.

Social norms are emerging through

- **successive cycles of use** and **reuse,**
- **modification, repurposing,** and **take-down notices** (because people upload anything, asking for forgiveness rather than permission).

Challenge 2a: Sharing-Individual

Individuals share profiles for social media, dating, and shopping, and any posted material they consider funny or informative.

- They have a notion of the public web as a place that is conceptually in the **public domain**.

Our notion of a separation between public and private is challenged by movements such as

- *sousvellience* (individuals surveilling themselves and events around them from the bottom up),
- life logging,
- quantified individuals (self knowledge through numbers), and
- personal genotyping (tracks on one's own reading/research linked to regions/genes/variations)

InterPARES
Trust

Challenge 2b: Sharing-Group

- Social media platforms facilitate the **movement of material from one circle of people to another**, crossing the public-private lines.
- Ad hoc dynamic groups of **employees from public agencies collectively create bodies of interlinked material** related to work projects (e.g. gcpedia), or common interests (whose ownership?)
- Small groups assemble the stories of activities and events, and change them.
- Contributions to social media by people, programs, committees, or agencies now dead are linked to ongoing, active contributions of the living, disappear, or appear as created by their successors.
- Digital lives, activities, initiatives are linked to each other

InterPARES Trust

Challenge 3a:
Control and Access-Private Data

- Massive amounts of **data about individuals** are held online and controlled by corporations and governments
- **Medical records** including genotyping or gene sequencing data, medical history, prescription and insurance information, tests, and images are held online by the medical establishment.
- **Genealogies**, factual biographies, all sort of biographic data and personal images are available to the private platform hosting them.
- In addition to big data, open data and open government raise the issue of **traceability** to and **retention of the source records**

Challenge 3b:
Control and Access-Government Records

- **Government business on social media**
  - Customer service
  - Access to information
  - Direct community involvement
- **Government records in social media**
  - Public engaging with government re: decision making
  - Public consultation on policies, development proposals, etc.
  - Policy announcements
  - Social media as a primary means of communication/advice in emergencies (e.g. earthquakes, fires, etc.)

A Communication Paradigm Shift

Before

- Hierarchical
- Closed
- Systems
- Centralized
- Passive
- Static

After

- Collaborative
- Open
- Networks
- Decentralized
- Interactive
- Dynamic

What Does Online Mean?

Online generally refers to the Internet.

Budapest Convention on Cybercrime, 2001

Internet providers are "entities providing users the **ability to communicate** through a computer system **that processes or stores computer data** on behalf of such communication or users." There are three "actions" related to the definition of provider: **communication, data processing** and **data storage.**

A type of Internet Provider is a Cloud Provider

"Cloud computing is a model for enabling 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 provisioned and released with minimal management effort or service provider interaction." (NIST)

Can We Trust Internet Providers?

The **concepts of place, jurisdiction, legitimate custody, and stability** were recognised in the 1990s as **the precondition of records trustworthiness**.

The primary justification for them is **historical accountability**: the people have a right to access the "authentic" documentary evidence of how they were governed.

For this to happen, the records must be under the unbroken physical and intellectual control of a **trusted** third party ensuring that their **interrelationships** as well as those with their creator are and remain **stable**

With records stored online, responsibility for **legal custody and intellectual control** would be with the preserver (archives, library, museum…), while **physical custody and technological access** with the Internet Provider.

What is Trust?

- In business, trust involves confidence of one party in another, based on **alignment of value systems** with respect to **specific benefits**
- In legal theory, trust is defined as a relationship of **voluntary vulnerability, dependence and reliance,** based on **risk assessment**
- In everyday life, trust involves acting without the knowledge needed to act. It consists of **substituting the information that one does not have with other information**
- Trust is also a matter of **perception** and it is often **rooted in old mechanisms** which may lead us to trust untrustworthy entities
- On the Internet, the **standard of trustworthiness** is that of the ordinary marketplace, *caveat emptor*, or **buyer beware**
- This is because there is **no standard for a trustworthy trustee** on the Internet

Issues with Materials Online

- Transparency
  - Compliance, Audit, Chain of Evidence, Accountability
- Ownership, authorship, creatorship
- Authenticity, reliability, accuracy: trustworthiness
- Retention and disposition
- Preservation of context
- Privacy, confidentiality,
- Jurisdiction
- Ethical rules

Traditional tools

- Model policies/procedures
- Records management manuals
- Training and promotion for RM
- Model agreements – SLA
- Leadership from national/international organizations
- Understanding of organizational culture

Balance of Trust

We need to establish a **balance between trust and trustworthiness**

The **trustworthiness** we must be able to guarantee is not that of the Providers but **of the records** that are entrusted to them, keeping in mind that **trustworthiness must be protected from creation**

To do so we must overcome several issues. A key one relates in particular to the cloud model of online storage and is the Issue of **Location Independence**

Location Independence

A fundamental issue with keeping materials in the Cloud today is the distinction between the **entity responsible for their preservation and accessibility** (the Designated Preserver) and the **entity storing them** (the Provider), and the possibility that the **jurisdiction** under which either exists is different from that in which the materials physically reside.

India is planning to impose a ban on the use of foreign cloud-based email services to send official communications, before the end of the year. It would prevent civil servants from using Gmail, Yahoo! or Outlook.com. Instead they would be required to use a service provided by the country's own National Informatics Centre (NIC).

Brazil's president has confirmed her country plans to set up its own secure, encrypted email service to 'prevent possible espionage'.

Europe does not allow the data of European citizens to be stored outside Europe, but has not yet provided an alternative.

Models to Consider to Respect the Archival Right/Duty in the Cloud

Maritime rules of shipping center on the recognition of the authority of the **port state**, the **flag state** and the **coastal state**

Early international maritime agreements established that the nationality of the transport vessel (the **flag state**) would establish jurisdiction, and by extension, the laws that would be in effect

Following the abuse of such rule, the **port state** was given greater control to inspect vessels coming within its territorial waters by the Law of the Sea Convention in 1982

Similarly, **coastal states** through whose waters the flagged vessels transit, have authority over the safety and competency of the ship and its crews and are also allowed inspection and enforcement while the vessel is in the coastal state's waters regardless of the flag of either the vessel (flag state) or its destination (port state)

InterPARES Trust

Making an Analogy

A **Canadian university could place its archives** into the care of an **American CSP** which in turn maintains its **data centers in Brazil**. Following the maritime example, the American company would be the 'flag state' that would be 'moving the goods' through 'coastal states' to their ultimate destination in the 'port state' of Brazil.

This is problematic not only because the Canadian University owning the archives would have no jurisdiction, but also with regards to the rights of the coastal state, in that the 'pipe' used to move the records can transit through several countries (coastal states) as they are routed to Brazil.

Traditionally, 'coastal states' have not been granted access to inspecting packets of records as they move along the internet**.** The rules of conduct then become very difficult, if not impossible, to enforce by any of the parties involved.

As they are, **maritime rules would not work in the cloud environment**.

Alternatives

The **territoriality principle** is not applicable because it is not possible to know the location of the records at any given time

The **nationality principle** is not applicable because nationality is an attribute of persons, not records, and the principle cannot be used to connect persons to records

The **power of disposal** principle, which “connects any data to the person or persons that obtain sole or collaborative access and that hold the right to alter, delete, suppress or to render unusable as well as the right to exclude others from access and any usage whatsoever” can be considered

By analogy, it could be possible to consider a **power of preservation principle** that assigns jurisdiction to the institutions controlling the material as the trusted custodian and the place guaranteeing authenticity, if clauses to that effect are included in Service Licence Agreements

InterPARES Trust

A Preservation Cloud

- Creators would have the option of including in the **contract with providers** (digital estate contract) the transfer of selected materials to the preservation cloud (archives, library, museum…) maintaining the existing links (also if the provider becomes extinct)
- The preservation cloud would enable all the preservation functions in a framework of **distributed responsibility**
- Documentation of **context through description** would allow preservation of meaning, providing a collective authentication of the materials and of its relationships, and ensuring controlled and accurate access
- **Collaborative work with industry and businesses** to design contract models that support acquisition and preservation has begun: e.g. the business provides account administration, installation, server administration and user technical support; and the preservation network provides fee-based server hosting and digital object storage service.

A real life example

The Council of Prairie and Pacific University Libraries (COPPUL) is piloting a cloud-based preservation service using the Archivematica digital preservation system. The service is offered to COPPUL member institutions that wish to preserve digital holdings but are unable or unwilling to install and manage local Archivematica instances. This service is a joint effort of COPPUL, Artefactual Systems Inc. (Archivematica lead developers) and UBC Library (the cloud storage provider).

COPPUL is responsible for promoting the service, signing up new institutions and seeding the one-time set-up costs; Artefactual Systems provides account administration, installation, server administration and user technical support; and UBC Library provides fee-based server hosting and digital object storage service.

Transparency, Stability, Permanence

Transparency:

- An unbroken chain of legitimate custody would be demonstrable
- Records reliability could be inferred from known creation processes
- Records authenticity could be inferred from the documentary context and a known preservation process

Stability:

- An archival cloud would guarantee that each record’s context is defined and unchanged, with all its relationships intact.

Permanence:

- Retention and disposition plans would be integrated with migration plans
- There would be no risk of termination of contract or provider for the archival cloud

We do not have to renounce our preservation mission to protect today’ s digital heritage. We have to frame it differently, with the help of interdisciplinary and international research.

InterPARES Trust

The **goal of InterPARES Trust** is to generate the theoretical and methodological **frameworks** that will support the development of integrated and consistent local, national and international **networks of policies, procedures, regulations, standards and legislation concerning digital records entrusted to the Internet**, to ensure public trust grounded on evidence of good governance, a strong digital economy, and a persistent digital memory.

InterPARES Trust is funded by a 5-year SSHRC Partnership grant and matching funds from UBC and all the partners (who are in 6 continents and 35 countries)

InterPARES Trust Participants

- The International Alliance comprises 7 Teams:
  - North America
  - South America
  - Europe
  - Asia
  - Australasia
  - Africa
  - Transnational Organizations
- Supporting Partners
- Pro-bono Consultants

Total : 250+ researchers and growing

Theoretical Framework

- **archival and diplomatics theory**, in particular the ideas that are foundational to trusting records
- **resource-based theory**, which focuses on the importance of technical, managerial, and relational capabilities for leveraging resources to maximize competitive advantage
- **risk management theory** on “post-trust societies”, which represents an available body of knowledge for reflection and further investigation on the relationship between risk and trust, and risk management and trust management
- **design theory**, which adopts an “argumentative process where an image of the problem and of the solution emerges gradually among the parties, as a product of incessant judgment, subjected to critical argument”
- **human computer interaction,** with its knowledge of human cognition, technological capabilities, networking, and human computer engagement
- **digital records forensics theory**
- **theories** of **measurement** and **calculation**, and
- **psychology** of **symbology, presentation** and **interpretation of trust labels.**

InterPARES
Trust

Methods

Research data will result from

- a close **analysis** of the **services** offered on the Internet, as well as the **technology** that supports such services
- a study of **relevant law** and **case law**, **regulations** and **standards**,
- a combination of **surveys** and **interviews** of Providers and existing Users of Internet services; and
- **case studies** and **general studies**.

These data are analysed through

- **Activity and entity modeling**, an analytic tool that enables understanding of the situational realities and work processes before and after modifications have been introduced to address problems.
  - To prepare a reference baseline, we are working with the Object Management Group (OMG) to develop a UML international model of Preservation As a Service for Trust (PAST) detailing all the functional requirements that providers must respect.

Methods (cont.)

Diplomatic and archival analysis, digital records forensic analysis, and textual analysis, as well as **visual analytics**.

C**omparative analysis** to generate a theory of trust in cloud environments that transcends national and jurisdictional boundaries

After having identified solutions, we will draft **model law, policies, procedures, and processes** to establish an **international framework** that can be embedded in domestic legislation, policies and procedures by each country

InterPARES Trust

ITrust studies

To ensure coverage of all issues we have divided the research area into

- domains (**infrastructure, security, control, access, legal issues**) and
- cross-domains (**terminology, resources, policy, social issues, education**)

and designed specific <u>studies</u> within each.

Today you will hear about the work conducted to date in the context of a few of those studies.

To learn more, please go to the ITrust website

InterPARES Trust

www.interparestrust.org

InterPARES
Trust