



InterPARES Trust

Luciana Duranti
Project Director, InterPARES
ITrust International Symposium
Victoria, B.C. 17 October 2014

The Goal of InterPARES 1 and 2 (1999-2007)

To develop the body of **theory** and **methods** necessary to ensure that digital records produced in **databases** and **office systems** as well as in **dynamic, experiential and interactive systems** in the course of artistic, scientific and e-government activities can be created in **accurate** and **reliable** form and maintained and preserved in **authentic** form, both in the long and the short term, for the use of those who created them and of society at large, regardless of technology obsolescence and media fragility.

InterPARES
Trust



Goal of InterPARES 3 (2007-2012)

To **enable** public and private **archival organizations and programs** with limited resources **to preserve** over the long term **authentic records** that satisfy the requirements of their stakeholders and society's needs for an adequate record of its past.

It did so by building on the products of the first two phases of InterPARES (1998-2006)

InterPARES
Trust



Records Online

Increasingly individuals and organizations choose to keep their records on line. The primary uses of the online environment are:

- Backup
 - Collaboration
 - Distribution
 - Recordkeeping
 - Long-term storage
 - Keeping Archives
-
- Email storage is number one

So we decided to gain an understanding of the issues and possible solutions before things got out of hands



InterPARES Trust (2013-2019)

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.**

InterPARES Trust is funded by a 5-year SSHRC Partnership grant and matching funds from UBC and all the partners (in cash and/or in kind)

InterPARES
Trust



Internet vs Cloud

Internet provider refers to “entities providing users the **ability to communicate** through a computer system **that processes or stores computer data** on behalf of such communication or users.” (Budapest Convention on Cybercrime, 2001). Three “actions” relate to the definition of provider: **communication, data processing** and **data storage**.

“**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.” (National Institute of Standards and Technology, 2009)

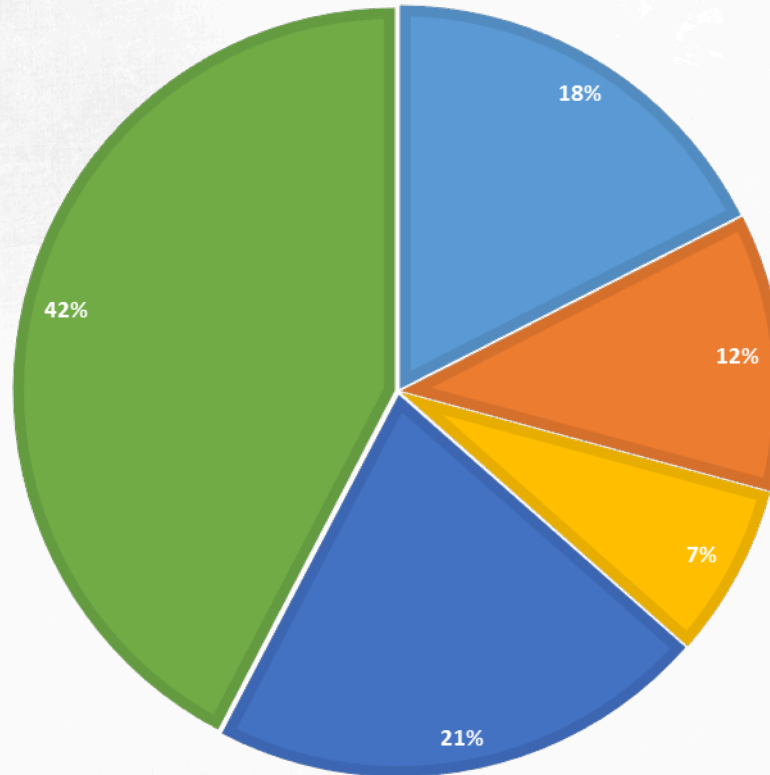
The term Cloud is useful as it conveys the nebulous nature of what happens on the Internet, and the fact that, differently from other industries presenting similar characteristics, like the aero-spatial one, the services offered on the Internet are not **regulated** nor are they **transparent**.

InterPARES
Trust



Type of Service in Use

■ IaaS ■ PaaS ■ Other ■ Don't know ■ SaaS



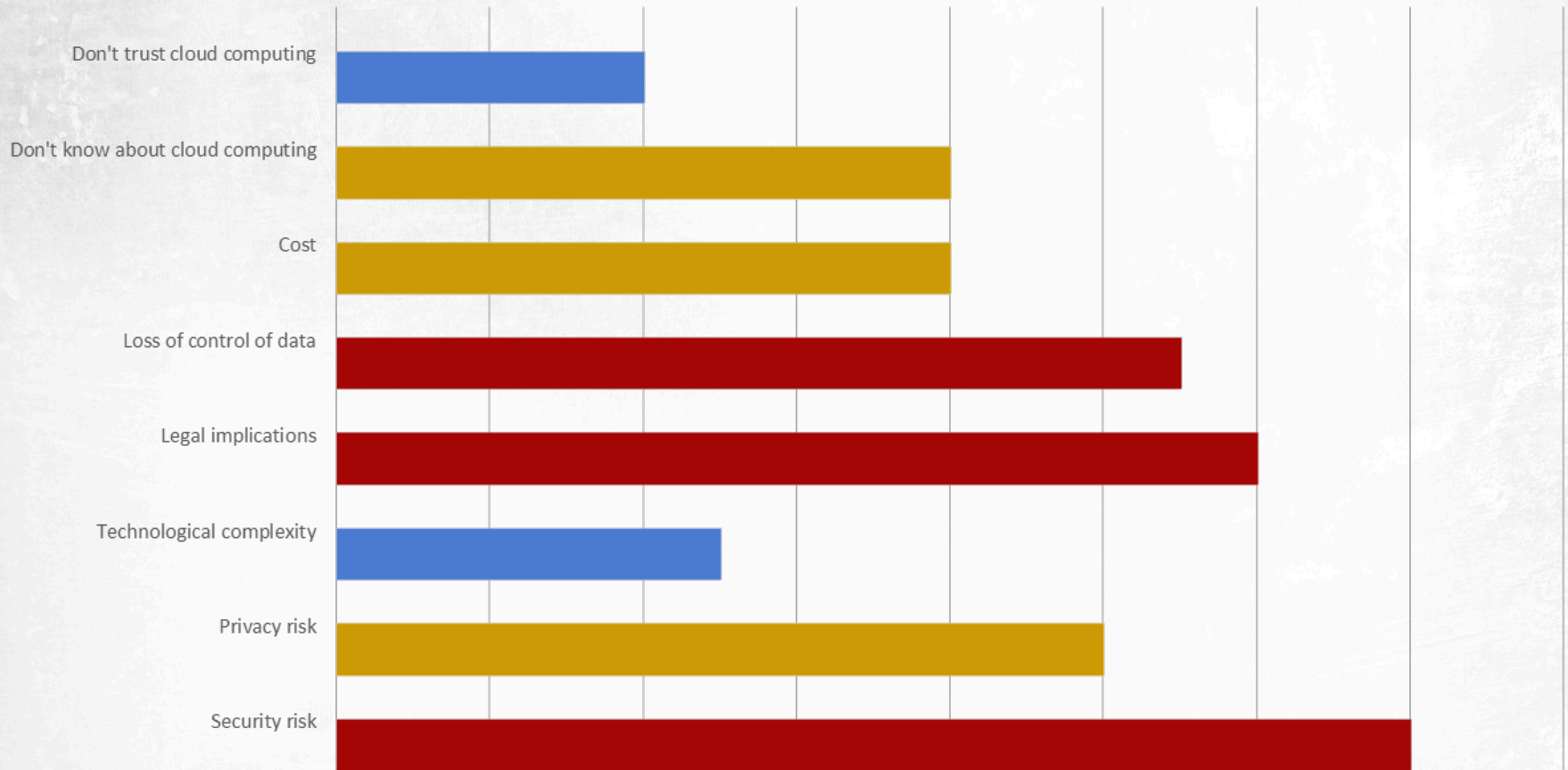
Motivations

What are the motivations for keeping records online?



Concerns

Why is your organization NOT considering cloud computing?



Trust on the Internet

- In fact we know very little about what happens on the Internet. The **standard of trustworthiness** for it is that of the ordinary marketplace, *caveat emptor*, or **buyer beware**
- Trust is defined in legal theory as a relationship of **voluntary vulnerability, dependence and reliance**, based on **risk assessment**
- The nature of trust relationships on the Internet is fraught with risks, weaknesses, and fault-lines inherent in the management of records and their storage in rapidly changing technologies where authorship, ownership, and jurisdiction may be questioned.



Benefit

Reduced Costs

- ✓ No owning of hardware/software, so no huge upfront costs.
- ✓ Lower energy costs.
- ✓ Reduced IT personnel costs, as they don't have to implement or maintain a Record Keeping or Preservation System.
- ✓ Even in a private cloud, shared-tenant system allows pooling of resources to get more for less-better hardware/software and network.
- ✓ You can get whatever you need, and only pay for what you use.
- ✓ You can track and measure use.



Risk

Cost Issues

- ✓ If you calculate transfer, implementation and subscription, costs are not insignificant. One can get unexpected license fees.
- ✓ Variability of costs-no set monthly fee.
- ✓ There is a significant per-request charge, to motivate access in large chunks.
- ✓ In Amazon, for example, although you are allowed to access 5% of your data each month with no per-byte charge, the details are complex and hard to model, and the cost of going above your allowance is high.

For **long-term storage**: a) it can be rented, as for example with Amazon's S3 which charges an amount per GB per month; b) It can be monetized, as with Google's Gmail, which sells ads against your accesses to your e-mail; c) it can be endowed, as with Princeton's DataSpace, which requires data to be deposited together with a capital sum thought to be enough to fund its storage "for ever".



Benefit

Reliability

- ✓ Always there on demand, big or small.
- ✓ Available from anywhere, using a browser.



Risk

Provider Reliability Issues

- ✓ Public providers can go bankrupt, disappear or be sold. Your records might be gone.
- ✓ Public and private providers can lose records, and sometimes can't get them back or backups fail.

The demise of Code Spaces is a cautionary tale, not just for services in the business of storing sensitive data, but also for end users who entrust their most valuable assets to such services. Within the span of 12 hours, the service experienced the permanent destruction of most Apache Subversion repositories

[http://en.wikipedia.org/wiki/Subversion_\(software\)](http://en.wikipedia.org/wiki/Subversion_(software)) and Elastic Block Store volumes <http://aws.amazon.com/ebs/> and all of the service's virtual machines. With no way to restore the data, Code Spaces officials said they were winding down the operation and helping customers migrate any remaining data to other services.

<http://arstechnica.com/security/2014/06/aws-console-breach-leads-to-demise-of-service-with-proven-backup-plan/>

InterPARES
Trust



Benefit

Security

- ✓ Security can be more robust than any one organization or unit could afford otherwise-both physical and virtual.
- ✓ Data sharding and data obfuscation requires a critical mass of data and complex technologies
- ✓ Centralized control on data easier to secure.



Risk

Security Issues

- ✓ Unauthorized access, sub contractors, hackers. It is not a matter of *if* but *when* a breach will occur. Are you told when it does?
- ✓ Documents can be stored anywhere and can be moved at any time-without you knowing.
- ✓ Encryption might not be done-in transit or in cloud. A security firm found recently that nearly 16% of the Amazon directories in which business customers store data could be perused by anyone online, revealing thousands of files containing sales records, passwords and personal data. It is a relatively new technology accessible to non-technical users.
- ✓ Shared servers could intermingle information.
- ✓ Law enforcement may seize servers for 1 person's actions. If 50 persons used it, it may take them days to get access to their records.



Benefit

Collaboration

- ✓ Allows for easy collaboration as all files are in consistent format, viewed in web browser.
- ✓ Can access and distribute information across distant geographic areas.
- ✓ Think Google Docs, Dropbox.



Risk

Control

- ✓ You have no real control over records online.
- ✓ No control over who shares your servers with you or to whom services are delegated.
- ✓ Terms of service or privacy policy may change.
- ✓ Backup may be done without you knowing and may not be disposed of as needed
- ✓ Records might be deleted without you knowing or may not be deleted according to the retention schedule.



The Trust Challenge

If we decide to carry out our activities online, we must find a balance between **trust** and **trustworthiness**, which is needed to ensure a balanced trust relationship.

Trust constitutes a risk which can only be mitigated by the establishment of a **trust balance**: we must trust trustworthy trustees and trustworthy records.

InterPARES
Trust



Issues linked to the Internet environment

- 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



Research Questions

- How can an organization's **records accuracy, reliability, and authenticity** be guaranteed and verifiable?
- How can an organization's records and information **security** be enforced?
- How can **forensic readiness** of an organization be maintained, **compliance** ensured, and e-discovery requests fully met?
- How can **confidentiality** of organizational records and data **privacy** be protected?
- How can an organization maintain **governance** upon the records entrusted to the Internet?
- How can **open government** and **open data** be guaranteed?
- How can records be **preserved over the long term**?



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 other, as required by that which we discover



Research 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.
 - You will hear today how, to prepare a reference baseline, we are working on the development of a UML international model of Preservation as a Service for Trust (PaaST) detailing all the functional requirements that providers must respect.



Research Methods (cont.)

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

Comparative 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, and get feedback from all interested parties.

InterPARES
Trust



Research Alliance

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

Total : 270+ researchers and growing

InterPARES
Trust



Partners Who Joined in the Past Year

- Universidade Federal Fluminense (Rio de Janeiro-Brazil)
- Estação Pinacoteca (Sao Paulo-Brazil)
- BC Information and Privacy Commissioner Office
- Queensland University of Technology (Australia)
- Atomic Energy International Agency (Vienna)
- Oslo University, Archives, Library & Info. Dept.
- National Archives of the UK (TNA)
- University of Kebangsaan (Malasia)
- University of Washington at Tacoma
- Object Management Group (we joined them)



Research Focuses

Domains

- Infrastructure
- Security
- Access
- Control
- Legal Issues

Cross-Domains

- Terminology
- Resources
- Policy
- Social/Societal Issues
- Education



Research Working Groups (RWG): Infrastructure

- **Infrastructure**

- Sensors in the Cloud (NA07)
- Contract Terms for Cloud-Based Record Keeping Services (NA10)
- International Reporting Program Creative Commons Content Archive (NA19)
- Ensuring Trust in Storage in Infrastructure-as-a-Service (IaaS) (EU08)



RWG Control

- The Calculus of Trust in Records (NA02/Transnational)
- **Retention and Disposition in a Cloud Environment (NA06)**
- **Modeling the Chain of Preservation of Records Entrusted to the Internet (NA12)**
- **Metadata, Mutatis Mutandis – Design Requirements for Authenticity in the Cloud and Across Contexts (NA16)**
- Research of Retention and Disposition Processes in an Internet Website of the Government of Israel: The Ministry of Foreign Affairs as a Case Study (EU01)
- Models for Monitoring and Auditing of Compliance in the Flow from Registration to Archive in e-Register (EU05)
- Evaluating Metadata Description Schemas of the Cultural Heritage Organizations in Electronic Environments for Interoperability (EU07)
- Comparative Analysis of Implemented Governmental e-Services (EU09)



RGW Security and Access

Security

- **Standard of Practice for Trust in Protection of Authoritative Records in Government Archives (NA03)**

Access

- The Implications of Open Government, Open Data, and Big Data on the Management of Digital Records in an Online Environment (NA08)
- Patents and Trust – From Traditional to Online Environments (NA13)
- The role of the Records Manager in an Open Government Environment in the UK (EU03)
- The Role of the Archivist and Records Manager in an Open Government Environment in Sweden (EU11)



RWG Legal Issues

- Developing Model Cloud Computing Contracts (NA14)
- Identifying Privacy Concerns and Attendant Issues in Cloud Computing (NA15)



RWG

Terminology

- **Core Terminology for InterPARES Trust (NA01/All Teams)**

Resources

- Open Government Data Literature Review (EU02)



RWG Policy

- Establishing Retention and Disposition Specifications and Schedules in a Digital Environment (NA09)
- Plurality and Policy: Designing for Trust in the Digital Age (NA17)
- Policies for Recordkeeping and Digital Preservation: Recommendations for Analysis and Assessment Services (EU04)
- Girona City Council in the Cloud: Do we really work in the Cloud? (EU06)



RWG Social Issues

- Putting the 'Fun' back in 'Functional' (NA04)
- **Social Media Use in Government (NA05)**
- **Historical Study of Cloud-Based Services (NA11)**
- Social and Ethical Foundations of Preservation of Digital Historical-Cultural Heritage (EU12)



Expected Outcomes

This project intends to generate

- **new knowledge** on digital records maintained online and accessed from all sorts of fix and mobile devices
- **shared methods** for identifying and protecting the balance between privacy and access, secrecy and transparency, the right to know and the right to be forgotten
- **legislative recommendations** related to e-evidence, cybercrime, identity, security, e-commerce, intellectual property, e-discovery and privacy
- **a model policy and a model international statute** specific to the Internet and recommendations for each government's continued development of its current fleet of uniform statutes.



Outcomes To Date

- 2 completed studies
- 37 publications
- 88 conference presentations (1 winner of best paper)
- 23 public relations publications, reviews, reports, interviews, videos (and a very popular Facebook page)



www.interparestrust.org

www.ciscra.org

InterPARES
Trust

