InterPARES Trust

Luciana Duranti

4-5 June 2013

North America Team Research Workshop

InterPARES Trust

- Concept of Trust
- Records on the Internet (benefits and risks)
- Goal of InterPARES Trust
- Team Composition
- Theoretical Framework
- Methodology
- Outcomes
- Impact
- Ultimate purpose

Trust on the Internet

- Trust is 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.

What is involved in Trust?

- In business, trust involves confidence of one party in another, based on **alignment of value systems** with respect to **specific benefits**

- 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

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.

Trustworthy Trustees

Trustworthy trustees present the characteristics of:

- *reputation*, which results from an evaluation of the trustee's past actions and conduct;
- *performance*, which is the relationship between the trustee's present actions and the conduct required to fulfill his or her current responsibilities as specified by the truster;
- c*onfidence*, which is an assur-ance of expectation of action and conduct the truster has in the trustee; and
- c*ompe-tence*, which consists of having the knowledge, skills, talents, and traits required to be able to perform a task to any given standard
- But not always we have this information and this creates blind trust

Whom Do We Trust?

- We trust banks, phone companies, hospitals, government, etc. to keep and maintain digital data, records, archives about us or belonging to us on our behalf. However, where those records actually reside, how well they are being managed, how long they will be available to us...we have no idea!

- Nothing wrong with it. After all, we trust airplanes to fly us safely without any need to know the pilot, and we trust banks to manage our money, and hospitals to care for our health.

- What would be different in putting trust in the **Internet**?

Internet vs Cloud

Often the Internet is referred to as the Cloud. Technically this is a misuse of terms.

However, it conveys the nebulous nature of what happens on the Internet, and the fact that, differently from the other industries mentioned earlier, the services offered on the Internet are not much **regulated** nor are they **transparent.**

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.

Records on the Internet

Questions:

- How can confidentiality of organizational records and data privacy be protected?
- How can forensic readiness of an organization be maintained, compliance ensured, and e-discovery requests fully met?
- 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 an organization maintain governance upon the records entrusted to the Internet?

The Providers' Response

- Choosing the Internet is a **Risk Assessment** decision

  Risk = probability x impact. It is a question of comparison. If one cannot have everything, what does one give up?

- The first choice Providers offer us is **between Transparency and Security**: they offer "trust through technology." Security involves location independence: a core aspect of these services delivery models.

- The second choice Providers offer is **between Control and Economy**: they offer "trust through control on expenditures."

- But there is a tension between laws that protect records in a traditional way and the abdication of custody and process without responsibility

What is the Internet Used For?

- Backup
- Collaboration
- Distribution
- Recordkeeping
- Long-term storage
- Keeping Archives

- Email storage is number one.

Benefits

Reduced Costs

✓ No owning of hardware/software, so no huge upfront costs.

✓ Saving energy costs.

✓ Reducing IT personnel costs, as they don’t have to implement or maintain a Record Keeping System.

✓ Shared-tenant system allows pooling of resources to get more for less-better hardware/software and network.

Benefits

Scalability

✓ You can get whatever you need, and only pay for what you use.

✓ You can track use.

Benefits

Reliability

✓ Always there on demand, big or small.

✓ Available from anywhere, using a browser.

Benefits

Security

✓ Security can more robust than any one company 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.

Benefits

Collaboration

✓ Allows for easy collaboration as all files are in consistent format, viewed in web browser.

✓ Can collaborate and distribute information over geographic areas.

✓ Think Google Docs, Dropbox.

Risks

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.

Risks

Provider Reliability Issues

- ✓ Cloud can go bankrupt, disappear or be sold. Your records might be gone.

- ✓ Cloud can lose records, and sometimes can’t get them back or backups fail.

InterPARES Trust

Risks

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 last week 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 businesses used it, it may take them days to get access to their records.

Risks

Control

- ✓ You have no real control over the cloud.
- ✓ No control over who shares your cloud 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.

Risks

Control #2

✓ You do not know what happens when cloud hardware/software become obsolete

✓ You can’t always move or remove records (e.g. for transfer to archives).

✓ Audit is not allowed.

✓ Termination of contract: records portability and continuity

✓ Termination of provider: records sustainability

Risks

Transparency

- ✓ Chain of custody is not demonstrable
- ✓ Records reliability cannot be inferred from known processes
- ✓ Tampering possible in the cloud, so records authenticity cannot be inferred
- ✓ Records in the cloud cannot have forensic integrity (repeatability, verifiability, objectivity)
- ✓ Can then records be admissible as evidence in a court of law?

Risks

Privacy Risks

- ✓ EU Data Protection Directive deals with privacy. It regulates processing of personal data in EU. One can’t transfer personal information (or its processing) of EU residents to countries that don’t have similar privacy protection (like the US).

- ✓ EU is developing a right to be forgotten directive. Can le **droit** à **l'oublie** be protected?

Risks

Legal Risks

- ✓ Geographic location of information-jurisdiction issues.
- ✓ Trade secrets-are they still secret in cloud?
- ✓ Legal privilege-is it still applicable if cloud can access it?
- ✓ US Patriot Act-FBI gets court order under Section 215.
- ✓ Can you isolate documents for legal hold?
- ✓ If multiple copies exist in different locations, which is the authoritative one?
- ✓ How can its authority be certified?

InterPARES Trust

Risks

Legal Risks: Metadata

✓ how does metadata follow or trace records in the cloud?

✓ how is this metadata migrated as a recordkeeping activity over time?

✓ who owns the metadata, especially metadata created by the Cloud service providers related to their management of your records and data?

✓ Is metadata intellectual property? Whose?

✓ How can this metadata be accessed for court and what are the responsibilities of the CSP in cases of legal discovery or hold?

Metadata make assertions (intentional or otherwise) about records, information, and data, and their contexts. Issues must be addressed about agents and rights, what metadata should assert, how these assertions persist over time or are changed, and in what way they are changed.

InterPARES Trust

To answer these questions is only a prelude to addressing issues of 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.

What is New About IP Trust

- The objective of building the foundations for establishing a **relationship of trust** between the people and those organizations that hold the records and data related to and/or belonging to them on the Internet
- The focus on **data and records** created in the interaction of people and organizations
- The scope, i.e., **public and private** organizations and **all types of Internet service models**
- The composition of the research team, which involves **developed and developing countries in six continents**
- The projected final outcome, i.e. a **supra-national framework** capable of guiding the development of domestic legislation and regulatory instruments that are consistent across cultures and societies

InterPARES Trust Composition

The International Alliance comprises 7 Teams:

- North America – Barbara Endicott Popovsky
- South America – Juan Voutssas
- Europe – Karen Anderson
- Asia – Jian Wang
- Australasia – Gillian Oliver
- Africa – Thomas van der Walt
- International Organizations – Jens Boel

Supporting Partners

Pro-bono Consultants, among which the Terminology Expert

International Alliance Steering Committee

Project Coordinator

Project Administrator

Project Technology Expert

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 participants (cloud computing designers and policy developers), as a product of incessant judgment, subjected to critical argument"
- **human computer interaction,** with its knowledge of human cognition, technological capabilities, networking, human computer engagement and the importance of cultural contexts

Methodological Framework

Research data will result from

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

We will focus on gathering, analyzing and interpreting data from a wide cross-section of organizations and institutions in order to explore the nature of trust relationships on the Internet, and the risks, weaknesses, and fault-lines inherent in record management and storage in rapidly changing technologies where authorship, ownership, and jurisdiction may be questioned.

Methods (cont.)

At the conclusion of each study the results may be represented using **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.

We will use **diplomatic and archival analysis, digital records forensic analysis, and textual analysis**, as well as **visual analytics**.

We will employ **comparative analysis** to generate a theory of trust in cloud environments that transcends national and jurisdictional boundaries, and on that basis identify ways of addressing the challenges evidenced by modeling and visualization.

After having identified solutions, we will draft **model policies, procedures, and processes, and ask the test bed partners to test them**.

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 statute specific to the Internet or recommendations for each government’s continued development of its current fleet of uniform statutes.

InterPARES Trust

Impact

The outcomes will be central to

- proper authentication of identity on the Internet and protection against Internet fraud

- electronic commerce on the Internet, which is concerned with breach of contract, business and competitive intelligence, consumer behavior and protection, defamation, advertising and marketing, and e-signatures

- intellectual property: patents, copyrights, trademarks, trade secrets, digital rights management, file sharing, licensing, public domain and international conventions

- Trusted Computing, which seeks to maximize security and minimize threats from spam, computer viruses and phishing

Impact (cont.)

- behavioral targeting, data breaches, Global Positioning System (GPS) use, lawful access by government, National ID Cards, online anonymity and John Doe lawsuits, Public Video Surveillance, Radio Frequency Identification (RFID) uses, Recording Customer Telephone Calls, Social Networking, Street-level Imaging Technology, Transborder Data Flows, etc.

- the creation of policy models, and procedures and standards to manage them

- the development of functional requirements and specifications for secure online digital systems

- the design of analytic frameworks to evaluate innovative business models emerging from and only possible in the evolving Internet environment, and

- the formulation of education modules and training tools for professionals, and academic curricula for graduate programs

InterPARES Trust

Research Domains

- Infrastructure -- Victoria Lemieux chair
- Protection – Barbara Endicott-Popovsky chair
- Control – Giovanni Michetti chair
- Access – Jim Suderman chair
- Legal – Anthony Sheppard chair

Infrastructure

- Technology/Mechanisms/Services
- Types of clouds
- Reliability of infrastructure on the Internet (e.g. obsolescence, continuing access, sustainability)
- Types of contractual agreements
- Costs

InterPARES Trust

Protection

- Methods: Encryption, sharding, obfuscation, geographic location
- Breaches
- Cybercrime
- Servers sharing
- Information Assurance
- Governance
- Audit

Control

- Integrity Metadata
- Chain of custody
- Retention and disposition
- Transfer and acquisition
- Intellectual control
- Use control
- Preservation

Access

- Open data/big data/open government/FIPPA/etc.
- Searchability/Usability
- Traceability
- Transparency
- Accountability
- The right to remember
- Privacy
- The right to be forgotten

Legal

- Legal Privilege
- Intellectual rights
- Chain of evidence
- Admissibility/Weight
- Authentication
- Certification
- Contractual rules (e.g. safe harbour)

Research Cross-Domains

- Terminology – Richard Pearce Moses chair
- Resources – Luciana Duranti chair
- Policy – John McDonald chair
- Social/societal issues – Pat Franks chair
- Education –Joe Tennis and Carolyn Hank co-chairs

Terminology

- Multilingual glossary
- Multilingual dictionary with sources
- Ontologies as needed
- Essays explaining the use of terms and concepts within the project

Resources

- Annotated bibliographies:
  - – published articles, books, etc.
  - – case law
  - – policies
  - – statutes
  - – standards
  - – blogs and similar grey literature

Policy

- In depth analysis of existing policies relevant to all 5 domains, as well as regulations, procedures, standard agreements, etc.

Social Issues

- Analysis of social change consequent to the use of the Internet, including but not limited to
  - Use/misuse of social media of all types
  - trustworthiness of news
  - Data leaks (intentional or accidental/forza maggiore) consequences
  - development issues (power balance in a global perspective)
  - Organizational culture issues
  - Individual behaviour issues

Education

- Development of different models of curricula for transmitting the new knowledge produced by the project

Working Groups

- Researchers within regional teams will propose research projects (template to be distributed) associated with a domain and/or cross/domain and work together using wikis and the discussion forum
- Researchers across regional teams working within the same domain or cross-domain will share their work on a regular basis
- This will go on for 4 years
- The last year will be dedicated to building the final products

A Balance of Trust

In the last year, the activity with the greatest impact will be the development of trust relationships models, which will be iterative, as we will be working towards **resolution** of issues as they present themselves, with the aim of developing **solutions** framed as a balance of trust.

To establish a “balance of trust” requires **enabling** the development of trustworthy technologies, procedures, and contractual conditions. We will do so by

- **identifying the changes needed in our paradigms of trust in data, records and records systems, and**
- **developing an internationally shared trust framework** that both providers and users can live by, because the current framework within which law enforcement operates and security concerns are addressed is inconsistent within and across jurisdictional boundaries.

Only then we can require and expect transparency, compliance and accountability, in addition to security and economy, and develop **Trust in the Internet**

InterPARES
Trust