InterPARES Trust

Luciana Duranti
Project Director
Oslo, 26 September 2013

The Overall Challenge

- The **nature of digital records**
- Establishing digital records **accuracy, reliability and authenticity** and maintaining it over time so that it can be proven
- Developing an infrastructure that ensures a **seamless controlled flow** of authentic data/documents/records from the creator to the preserver irrespective of changes in technology
- Providing **transparency** while protecting secrecy where warranted
- Ensuring that the **conflicting rights** of users, clients, employees, and future generations are protected
- Ensuring the **permanent preservation** of the documentary cultural heritage in digital form

InterPARES Trust

Digital vs. Traditional Records

In the digital environment:

- Record content, form and medium are no longer inextricably linked
- The stored entity is distinct from its manifestation and its "digital presentation" has to be considered as well as its "documentary one"
- When we save a record, we take it apart in its "digital components", and when we retrieve it, we reproduce it (*ergo*, it is not possible to preserve a digital record, only the ability to reproduce or recreate it)

Therefore, we can no longer determine authenticity on the object-record, which is composite (stored + manifested) and permanently new (re-production), but must make an **inference of authenticity from its environment of creation, maintenance & use and preservation**.

InterPARES Trust

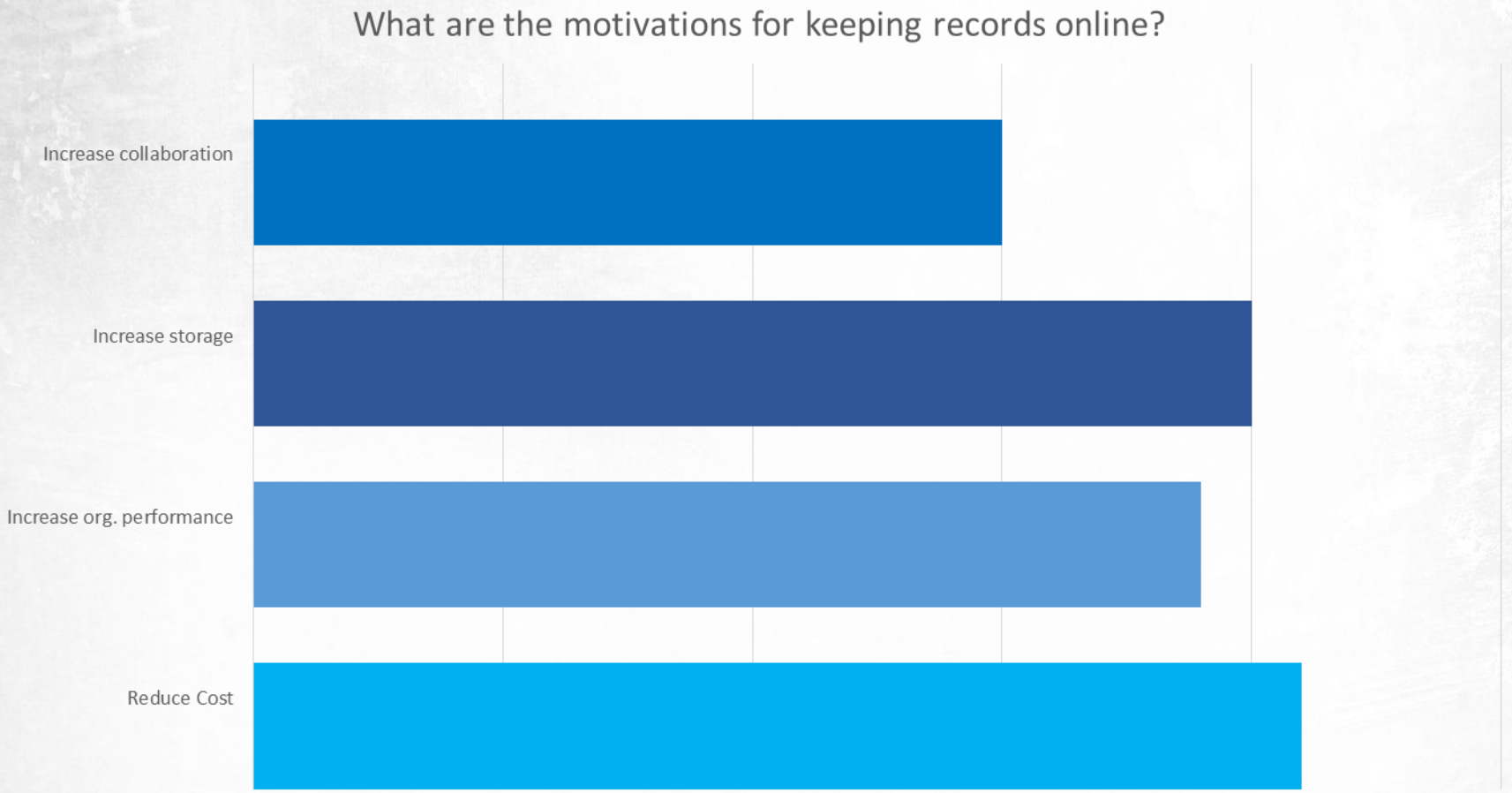
Records Online

Furthermore, 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.

Motivations



Internet vs Cloud

Often the Internet is referred to as the Cloud. Technically this is a misuse of terms. I will use the term Internet provider to refer 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). Therefore, there are three "actions" related to the definition of provider: **communication, data processing** and **data storage.**

However, the term Cloud is useful because 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 much **regulated** nor are they **transparent.**

InterPARES Trust

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.

InterPARES Trust

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

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**?

InterPARES Trust

Questions We Should Be Asking

- How can confidentiality of records and data privacy be protected in the Internet?
- 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?
- How can the preservation of records of permanent value be ensured?

InterPARES Trust

The Classic Response

- Choosing the Internet is a **Risk Assessment** decision where Risk = probability x impact. It is a question of comparison. If one cannot have everything, what does one give up?

- The first choice offered us is **between Transparency and Security**: the Internet offers "trust through technology." Security involves location independence: a core aspect of Internet services delivery models.

- The second choice offered us is **between Control and Economy**: the Internet offers "trust through control on expenditures."

- But there is a necessary tension between laws that protect records in a traditional way and the abdication of custody and process without responsibility. Many are aware of this tension.

InterPARES Trust

Benefits

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 System.
- ✓ Even in a private cloud, shared-tenant system allows pooling of resources to get more for less-better hardware/software and network.

InterPARES Trust

Benefits

Scalability

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

Benefits

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.

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.

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".

InterPARES Trust

Risks

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.

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 month 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.

InterPARES Trust

Risks

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.

InterPARES Trust

Risks

Control #2

- ✓ You do not know what happens when 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 on the Internet, so records authenticity cannot be inferred
- ✓ Records on the Internet cannot have forensic integrity (repeatability, verifiability, objectivity)
- ✓ Can then records be admissible as evidence in a court of law?

InterPARES Trust

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—regardless of the Safe Harbout clause).

- ✓ 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 (loss of location).
- ✓ Trade secrets-are they still secret when shared with a provider?
- ✓ Legal privilege-is it still applicable if a provider can access the records?
- ✓ US Patriot Act-FBI can get court orders 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 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 provider in cases of legal discovery or hold?

InterPARES Trust

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

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 (in cash and/or in kind)

InterPARES Trust Participants

- The International Alliance comprises 7 Teams:
  - North America
  - South America
  - Europe
  - Asia
  - Australasia
  - Africa
  - Transnational Organizations
- Supporting Partners
- Pro-bono Consultants
- International Alliance Steering Committee
- Project Coordinator
- Project Administrator
- Project Technology Expert
- Student Research Assistants

Total : 200+ members and growing

Research Questions: Same Mentioned Above

- 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?
- How can **open government** and **open data** be guaranteed?
- How can records be **preserved over the long term?**
- How can **a balance of trust and trustworthiness** be achieved?

InterPARES Trust

Research Objectives

- 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

- Ensuring the trustworthiness (reliability, authenticity, accuracy) of **data and records** created in the interaction of people and organizations

- Developing a **supra-national framework embracing both developed and developing countries and all sectors**, which is capable of guiding the development of domestic legislation and regulatory instruments that are consistent across cultures and societies

InterPARES Trust

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

In the first 4 years, 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.

InterPARES Trust

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**.

InterPARES Trust

Working Groups

Domains

- Infrastructure
- Protection
- Access
- Control
- Legal

Cross-Domains

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

Infrastructure Domain

- Technology/Mechanisms/Services
- Issues specific to types of infrastructure
- Reliability of infrastructure (e.g. obsolescence, continuing access, sustainability)
- Types of contractual agreements
- Costs

Infrastructure: Proposed Studies

- Contract Terms for Cloud-based Record Keeping Services

Cloud-based services (CBS) and the technological infrastructure (s/w, h/w) are primarily set by the vendors of these types of services and secondarily by the purchaser's needs/expectations. Terms of contracts for CBS thus represent interests from two perspectives: i) the service provider; ii) the purchaser. Through empirical analysis, the research will categorize: a) terms found in available contracts relating to record keeping requirements in terms of commonality or frequency of appearance; b) types of services purchased; c) types of technological infrastructure. It will determine, to the degree possible, whether the terms represent primarily the interests of the service provider or the purchaser. It will relate the terms, to the degree possible, to types of organizations, e.g., government, health sector, financial sector, etc.,

InterPARES Trust

Infrastructure: Proposed Studies

- Sensors in the Cloud

We intend to look at digital data provenance (Buneman, 2000) issues specific to mobile sensors to develop and carry out a risk assessment related to issues of interest to the InterPARES Trust project as they arise in a specific application of mobile sensing that is being developed at MAGIC. The questions we would like to address include the following:

- What are potential data provenance issues when dealing with mobile sensors?
- What are the ways we can ameliorate potential risks associated with mobile sensors to make them more trustworthy?

Protection Domain

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

Protection: Proposed Studies

- Standard of practice for trust in protection of authoritative records in government archives

Risk management decisions need to be made. For the protection arena, these are decisions of kind rather than amount, and have to last over long periods. As such they are architectural in nature. The objective of this research effort is to build a global consensus around a limited set of these decisions for government-level systems of records and archives. In essence, this will create a standard of practice for risk management in authoritative archives.

Control Domain

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

Control: Proposed Studies

- Model the Chain of Preservation for Records Entrusted to the Internet

The modeling project will address the following questions:

• 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?

InterPARES Trust

Control: Proposed Studies

- The calculus of trust in records

This study will identify a range of methods and limit cases for evaluating authenticity parameters based on authenticity parameters of inputs, examine the provenance issue, calculate how much metadata may be required to provide all of the relevant facts upon which calculation of authenticity of a record and associated claim may be done, augment these results to deal with changes in parameters based on later evaluations of information, and will look at theories of measurement and calculation methods for a small number of case studies, presentation methods, symbologies, and the psychology of presentation and interpretation of trust labels.

Control: Proposed Studies

- Retention & Disposition in a Cloud Environment

One approach under consideration is to develop a set of RM best practices (i.e., what we believe the answers will be) and then ask questions of the providers and measure their responses/knowns against the best practice. However, some of the questions for which we wish to find answers may not lend themselves to that format and would need to be included as open ended questions. For now we have a list of questions for which we would seek answers:

- What would we need to know if we moved to cloud in 3 years time?
- What makes it different from other types of remote storage/data base environments?
- What do organizations need to tell cloud service providers to do?
- What are the minimum standards for retention and disposition?

InterPARES Trust

Access Domain

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

Legal Domain

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

Terminology Cross-domain

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

InterPARES Trust

Terminology: Proposed Studies

- Big Data, Open Government, and Open Data - their evolution

Big Data, Open Data and Open Government are having a substantial impact on the online environment. The evolution and characteristics of these relatively recent themes are poorly understood, especially from a recordkeeping perspective. This lack of understanding will inhibit the effective undertaking of research projects that address the creation and management of digital records generated in these environments. Each of the themes is reflecting recordkeeping issues that need to be understood if ITrust research projects are to be relevant and effective. The inter-relationships among the three themes suggest that they may be experiencing the same or similar recordkeeping issues. Understanding the processes for establishing and managing Big Data, Open Data and Open Government initiatives will support ITrust research and help in the development of recordkeeping policies, standards, and practices for managing digital records in the online environment.

InterPARES Trust

Terminology: Proposed Studies

- Core Terminology for InterPARES Trust

How will InterPARES Trust define fundamental concepts, how are they understood in various contexts, and how do they relate to each other? Terms identifying such concepts, not defined in previous InterPARES terminology databases, have already surfaced at the initial meeting in Vancouver and in subsequent email, and include the following:

- *big data*
- *cloud (as distinct from the Internet), both public and private*
- *data sets*
- *Internet (as distinct from the cloud)*
- *open access*
- *open data*
- *open government*
- *platform as service*
- *trust*

Resources Cross-domain

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

Policy Cross-domain

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

Policy: Proposed Studies

- Establishing retention and disposition specifications and schedules in a digital environment

Issues being addressed: Impact of the digital environment on establishing retention and disposition specifications and schedules for digital records; and methods for developing and applying specifications and schedules. The objective is to develop recommendations on the establishment and implementation of retention and disposition specifications and schedules for digital records

Social Issues Cross Domain

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/*Force majeure*) consequences
- development issues (power balance in a global perspective)
- organizational culture issues
- individual behaviour issues

Social Issues: Proposed Studies

- **Historical Study of Cloud-based Services**

Identify, to the degree possible, those CBS that suffered significant loss of trust by the user community. From this subset of CBS, the research would assess the basis for that loss of trust and, where applicable, the service provider recovered/restored trust or why the user community renewed its trust in the service(s).

- **Social Media**

The first phase of the project will explore the types of social media initiatives undertaken by 5-10 government organizations (number TBD) in the US and an equal number in Canada to determine how they utilize social media to engage citizens and provide customer service, as well as how the public reacts to those initiatives. The ultimate goal of this research project is to develop two or more case studies that highlight the citizen experience with government social media tools, customer experience, and issues of trust.

InterPARES Trust

Social Issues: Proposed Studies

- **Putting the 'Fun' back in 'Functional'**

This project will explore some of the socio-technical factors that appear to affect the management of written and non-written information in organizations. It is based on the assumption that the social (i.e., cultural, historical, political, ideological, economic, ethical, linguistic, rhetorical, epistemological,… in one word, human) interactions that are involved in using available technologies shape and are shaped by the technologies used. In particular, we are interested in understanding how people *engage* with the information they create/use to accomplish their work in networked environments.

InterPARES Trust

Education Cross-domain

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

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 international statute specific to the Internet and recommendations for each government's continued development of its current fleet of uniform statutes.

InterPARES Trust

A Balance of Trust

In the last year of the project, 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

www.interparestrust.org

www.ciscra.org