Trusting Digital Records

Luciana Duranti, Professor
PI and Director, InterPARES Project
The University of British Columbia

Trusting Records

My presentation focuses on the key product of Digital Transformation, **Digital Records** and on ensuring their **Trustworthiness**

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 and dependence,** 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
- In the digital environment, the **standard of trustworthiness** is that of the ordinary marketplace, *caveat emptor*, or **buyer beware**

InterPARES TrustAI

Trustworthiness

Reliability

The trustworthiness of a record as a **statement of fact**,

based on:

- the competence (authority and capacity) of its author
- the controls on its creation

Accuracy

The **correctness and precision** of a record's data

based on:

- the competence of its author
- the controls on content recording and transmission

Authenticity

The trustworthiness of a record that **is what it purports to be**, untampered with and uncorrupted

based on:

- identity
- integrity

InterPARES TrustAI

Authenticity - Identity

Identity refers to the attributes of a record that uniquely characterize it and distinguish it from other records. These attributes include:

- the **names** of the persons concurring in its creation (i.e., author, addressee, writer, originator, creator);
- its **date(s)** of creation (i.e. making, receipt, filing) and transmission;
- the matter or **action** in which it participates;
- the expression of its **relationships** with other records (e.g. classification code); and
- an indication of any **attachment(s)**

Integrity refers to the quality of being complete and unaltered in all essential respects.

InterPARES TrustAI

Authenticity in the Digital Environment

- There was no question in the paper environment that the authenticity of a record could be assessed on its form, provenance and documentary context, but **this turned out to be linked to the immutability of a record affixed to a medium "permanently", that is to its integrity.**
- The problem is that **digital records cannot be kept, maintained or preserved**. **It is only possible to maintain our ability to re-produce or re-create them.** This is based on two considerations.

InterPARES TrustAI

Keeping Digital Records

• For **records that are the counterpart of traditional paper records**, the physical form in which they are stored is necessarily different from what is displayed on a screen or printed on paper, as well as from the physical form of the record in a computer processor.

- The entity we keep is the stored digital encoding of a record. This **stored record** is distinguished from the **manifested record**, which is either a copy of the stored record in a form suitable for human use or in a form suitable for use in an automated system designed to process such records.
- The general requirement for digital records, then, is that, regardless of how a record is represented in bits in digital storage, **it must be possible to generate a manifested record that has all the identifying attributes of the first manifestation of that record that was capable of fulfilling its purposes** (i.e. the original).

• For digital **records of which there are no analogue precedents**, such as records in dynamic, interactive and experiential systems, it is not possible to keep, maintain or preserve, in digital form, a record intended for human use.

InterPARES TrustAI

System Integrity?

- Thus, digital records are always new, **always green**. Given this, how do we know whether they are authentic… whether their identity and integrity is the same of when they were first produced?
- Some countries, like Canada, have established that records are authentic if we can demonstrate the integrity of the system(s) where the records have been stored at any given time (see CGSB 72.34 2022).
- Can we then consider **system integrity** another component of authenticity?
- **I do not think so**. The integrity of the system—as mentioned—supports authentication, which is a declaration of a record's authenticity at a specific moment, and is a deduction based on system/record security.

InterPARES TrustAI

Records Security

- It is protection of the system/records from **unauthorised access, use, alteration or destruction.** In a world where **integrity of a system is an inference** we make from policy, rules, audit trails and logs, from which **one infers integrity of the record**, from which **one infers its authenticity**, **security is the new authenticity**.
- Individuals enforce security with something they know (e.g. password), they own (e.g. tokens), or they are (e.g., biometrics of eyes, fingerprints, private keys in a PKI environment)
- Cloud providers enforce it through encryption, should **produce audit trails and access logs,** and should capture, maintain and make available **metadata** associated with access, retrieval, use and management of the data, in addition to those linked to the data themselves. If they do, **users do not have access to them, as anything produced by cloud providers belongs to them**.

InterPARES TrustAI

Cloud Computing Vulnerabilities

- Records can be in data centres anywhere in the world
- The location of the records is a criterion in **determining the law that applies** in case of litigation
- The **shared tenancy system** is an issue in the public or commercial cloud, but vulnerability also exists in the private cloud
- National strategies used to require that records resides within the boundaries of the country where they were created (very expensive for data centres, if Europe or North America).
- The international strategy no longer requires that, thereby underscoring the importance of **multilateral agreements** among countries for collaboration in security (new safe harbour)
- But is collaboration the key to safety? Clearly not. Hence…

InterPARES TrustAI

Technology Dependent Authenticity

Hence the attraction of **Blockchain** technology

- the underlying technology enabling Bitcoin
- a ledger, i.e. a distributed transactional database which keeps a final and definitive (immutable) trace of transaction records (their hash).

- relies upon a **distributed network** (all nodes—servers are equal) and **decentralized consensus** (no centre(s); no single point of control or attack)
- The confirmed and validated sets of transactions are held in blocks, which are linked (chained) in a chain that is tamper-resistant and append-only
- It starts with a **genesis block and each block contains, in addition to the hash of a predetermined number of documents, a hash of the prior block in the chain.**

InterPARES TrustAI

How is Blockchain used?

Blockchain can be used to confirm

- the **bitwise integrity** of a record kept elsewhere
- that a record **existed** or **was created** before (not at) a certain point in time (i.e. not after being timestamped and its hash registered in the blockchain)
- the **sequence** of uploading of the hash algorithm to the blockchain

Is it a **recordkeeping system**? No. It holds the hash of records, not records (*smart contracts*—i.e. agreements between parties directly written into lines of code—are not considered records). The records must still be stored and managed off chain. This is good, because, if they were in the blockchain, they would be **immutable.**

InterPARES TrustAI

Types of Blockchain

- **Public Blockchain**: anyone can participate in reading, writing and auditing the blockchain without permission. A public blockchain is **open and transparent**. Eg: Bitcoin, Ethereum, Litecoin and many others.
- **Private Blockchain**: the write permissions are limited to one organization. Read permissions may be public or restricted. They are a way of taking advantage of blockchain by setting up groups and participants who can verify transactions internally. **This creates the risk of security breaches like a centralized system**, as opposed to the public blockchain, which is secured by incentive mechanisms.
- **Consortium or federated blockchain**: a type of private Blockchain, which removes the individual autonomy that is responsible for bringing changes in the blockchain as in the private blockchain. They operate under the control of **a group of institutions that do not allow everyone to participate in the process of verifying transactions**. Eg: R3 (banks), EWF (Energy), and B3i (Insurance)

InterPARES TrustAI

Immutability/Integrity

- It is the attraction of blockchain: it is what ensures integrity, as nothing can be changed in a record or removed from a block
- It is also the key problem of blockchain:
  - with **current records**, any **updating or correction** of wrong data; any form of **privacy protection**; any exercise of the **right to be forgotten**; any **disposition** of no longer needed records; any **record making system upgrade**; in short, any change in the record/s would invalidate the blockchain
  - with **records identified for continuing preservation**, any **transfer** to a preservation system; any **migration**; any **addition** to the records aggregation would invalidate the blockchain

InterPARES TrustAI

Identity

- The hash on the blockchain does not allow for links to
  - the hash of related records, hence **no interrelationship among the records** is documented
  - the hash of metadata, hence **no context**
- If the metadata were embedded in the record at creation, the hash of such record would not allow for additions or changes

InterPARES TrustAI

Authenticity Problems

- Proving **authenticity at origin** is not possible
- Preserving the **contextual evidence** is not possible
- Handling the **decentralized** (and thus trans-jurisdictional) nature of the blockchain is difficult (who is the creator? the owner? What law applies?)
- Dealing with code in a situation where **the necessary components of the transaction are controlled by different actors in different jurisdictions** is a challenge; and,
- with **smart contracts**, lacking both the equivalent of a signature and the date of the completion of an agreement, it is impossible to authenticate the transaction.

InterPARES TrustAI

Decentralization Problems

- Information processing happens on a complex technological stack in which **different technical components may be in the custody of, and operated by, very different actors**.
- Some components may be under the control of a single organization, others under the control of business partners who are members of a blockchain consortium, and still others under control of unknown third-party actors.
- **An organization's records could be in the custody of thousands of independent actors over which records creators exercise little or no control.**

InterPARES TrustAI

Decentralization (cont.)

- The **consensus mechanism** and other protocols or standards determining how the blockchain operates, may not be within the decision-making purview of the records creator (or the creator's designated records professional)
- Instead, these may be decided by remote (and even unknown) third party developers. In many cases, these protocols and standards are still unstable, and thus the reliability of the upload of records to the blockchain could be very difficult to establish with any certainty

InterPARES TrustAI

Other Challenges and Limitations

This emerging technology is still in its early stage of development.

Despite the huge opportunities blockchain seems to offer, it suffers from challenges and limitation such as

privacy,

compliance,

governance,

scalability and

security issues

that have not yet been thoroughly explored and addressed.

Although there are some studies on the security and privacy issues of the blockchain, they lack a systematic examination of the security of blockchain systems.

InterPARES TrustAI

The Right to Trustworthy Records

- With the advent of the digital environment, the records have become **ubiquitous**, existing in multiple copies in different places, and **easier to access**.
- As a consequence of freedom of information laws, **records creators have become aware of their records in a supremely self-conscious manner**. They have figured that the records could not be left for others or for chance to determine their future: they are too significant politically, legally, and socially, and the time for accountability through them could come any moment, here and now, 4 days or 30 years from now, or 300. Historical accountability will also come.
- This is why it is vital to **focus not just on digital transformation but on the people's right to trustworthy records by ensuring not only their authenticity but the ability to verify and prove it.**

InterPARES TrustAI

Proving Authenticity of Digital Records

The **fundamental difference** between the authenticity of analogue and digital records is in the fact that, while the authenticity of analogue material can be proven and verified on its face and on the basis of its documentary context, and only exceptionally is circumstantial or extrinsic evidence necessary, the authenticity of digital material cannot.

The assessment of the **authenticity of digital material**

- **is always an inference** based on extrinsic elements such as their significant properties as shown by identity and integrity metadata, and
- **relies on circumstantial evidence** such as
  - the **integrity of the system** hosting it at any given time,
  - the **policies and procedures controlling its life**, and
  - the **technology encrypting or securing the access to it**.

InterPARES TrustAI

What About Artificial Intelligence?

Artificial Intelligence Systems (AIS) are computing systems using algorithms capable of carrying out complex tasks that were once believed to be the sole domain of natural intelligence:

processing large quantities of information,

calculating and predicting,

learning and adapting responses to changing situations,

recognizing and classifying objects.

Question:

Can we develop AIS for ensuring the continuing trustworthiness of the records?

InterPARES TrustAI

AIS Issues

Artificial Intelligence Systems provide

- Inconclusive Evidence (based on probabilities)
- Inscrutable Evidence (no interpretability or transparency)
- Misguided Evidence (as good as the data provided)
- Unfair Outcomes (disproportionate impact on one group of people)
- Transformative Effects (challenges for autonomy and privacy)
- Non Traceability (hard to assign responsibility)

Plus

- The decisions AIS make are based on past decisions, and
- when it comes to human affairs, tomorrow rarely resembles today, and data and numbers can't say what has a moral value, nor what is socially desirable

InterPARES TrustAI

Montreal Declaration Principles (2018)
Canada and the Association of Southeast Asian Nations

- Respect for Well-being principle
- Respect for Autonomy Principle
- Protection of Privacy Principle
- Solidarity Principle
- Democratic Participation Principle
- Equity Principle
- Diversity and Inclusion Principle
- Caution Principle
- Responsibility Principle
- Sustainable Development Principle

InterPARES TrustAI

Past Experience with AI

There have been several projects looking at AI for controlling and accessing records: they typically look at a particular tool in a specific context or even a single set of records.

- recurrent neural networks for classification of the content of large aggregations of records
- recommendation systems that connect relevant images to digitized letters, by using handwritten text recognition (HTR) to make digitized documents searchable
- chatbots that emulate human conversation through voice commands or text chats or both to help knowledge seekers find connected information
- a combination of Named Entity Recognition (NER), entity relations tools, and topic modeling to create visualization tools for the types of data stored on disk images

InterPARES TrustAI

The Record Problem

- Relying on existing off the shelf tools, as all the studies on AI for records have done, limits what challenges can be met, as it makes the needs of record governance subservient to the larger field of machine learning
- It may be practical, but many tangible instances of bias have been found in modern machine learning models, often driven by questionable data collection practices
- This raises the questions of a) whether off the shelf tools are the best solution for records and b) what AI could look like if this power relationship between AI and information governance were reversed, with the **theory of the record informing the creation of AI tools**

InterPARES TrustAI

I Trust AI Project Goal

The overall goal of the 5th phase of the InterPARES research project, I Trust AI, is to design, develop, and leverage Artificial Intelligence to support the ongoing availability and accessibility of trustworthy public records by forming a sustainable, ongoing partnership producing original research, training students and other highly qualified personnel (HQP), and generating a virtuous circle between academia, archival institutions, government records professionals, and industry, a feedback loop reinforcing the knowledge and capabilities of each party.

InterPARES TrustAI

Objectives

- Identify specific AI technologies that can address critical records challenges;
- Determine the benefits and risks of using AI technologies on records;
- Ensure that records concepts and principles inform the development of responsible AI; and
- Validate outcomes from Objective 3 through case studies and demonstrations.

The more than 40 studies in course relate to aspects of records creation and retention and disposition, to predicting the presence of types of PI in public records, to access etc. They are carried out by international (28 countries involved) multidisciplinary teams (AI and records and archives experts, forensic and legal experts, data science experts, etc.).

InterPARES TrustAI

Expected Outcomes

The project will improve upon existing tools and create new Machine Learning tools that will address records needs, such as

- machine translation,
- image recognition and description,
- optical character recognition (OCR) and handwritten text recognition,
- text summarization and classification, and
- text style transfer for language civilization (e.g., removal of bias, hate, and sexism)

InterPARES TrustAI

Trusting Records through AI?

- It might be possible to use **Artificial Intelligence based on records related concepts to classify, select and authenticate records**, and **to detect interference with them**, but we need to link AI tools to much more sophisticated **security** (remember: *security is the new authenticity*) than we have now to protect the records from both internal and external agents and **prove** that we have been successful in doing so.
- **Standards** may help us. Though they must be technologically neutral, they can state the principles to be respected, indirectly addressing issues presented by emerging technologies (such as immutability) and directly establishing requirements for record-making and recordkeeping systems.
- Alternatively, we can either keep the records permanently off line or maintain a complete identical reproduction of them in a secure physical off line location. It would not help privacy and confidentiality, but **at least we would have a set of verifiably authentic records.**

InterPARES TrustAI

Thank you!

luciana.duranti@ubc.ca

www.interparestrust.org

www.interparestrustai.org