Records in the Cloud Documentary Truth in a Networked Environment

Luciana Duranti, Director, The InterPARES Trust Project

THE AUSTRALIAN AND NEW ZEALAND COLLEGE OF NOTARIES Conference "The Electronic Future is Now"

Singapore, 17-19 October 2018

Truth vs. Trust

Truth vs. **Trust**

- The **historical truth** is not directly accessible: facts and acts slide into the past as they happen
- There are two ways of indirectly accessing the past: witness testimony and the **documentary truth** represented by the written accounts of the facts and the material instruments of the acts, the **records**
- In both cases, what we will regard as truth will entirely depend on our trust in its source

InterPARES Trust

What is Trust?

- Some view it as a four-level progression: from **individual**, as a personality trait, to **interpersonal**, as a tie directed from one person to another (son to father); to **relational**, as a property of a mutual relationship (people doing business); and **societal**, as a feature of a community as a whole.
- InterPARES Trust defines it as confidence of one party in another, **based on alignment of value systems with respect to specific benefits**, and involving a relationship of voluntary vulnerability, dependence, and reliance, based on risk assessment
- Substantially, trust involves acting without the knowledge needed to act, by **substituting the information that one does not have with other information**, e.g. the competence of a notary, the report of a witnesses, oral tradition, documentary truth

InterPARES Trust

Documentary Truth

Records are instruments and by-products of action. They are the primary sources of evidence for any kind of research because they were not produced to answer the questions we ask of them, say, a century later, but as a means to act.

Records and archives form the infrastructure through which beliefs and values are upheld and understood;

This has been the case since antiquity.

The Born Digital Documentary Truth

The nature of born digital records:

- They are **vulnerable** (easy to destroy, lose, corrupt, tamper with, or become inaccessible if not protected) and **persistent** (forever there, if not purposefully destroyed)
- Their **content, structure, and form are no longer inextricably linked**
- The record as a stored entity is distinct from its manifestation on a computer screen, and its **digital components** have to be considered as well as its **documentary form**
- When we save a record, we take it apart in its digital components. When we retrieve it, we create a copy (re-production).
- **It is not possible to preserve born digital records, only the ability to reproduce or recreate it**
- Yet, we have been able to issue international standards for **trusted digital repositories,** but they are expensive

InterPARES Trust

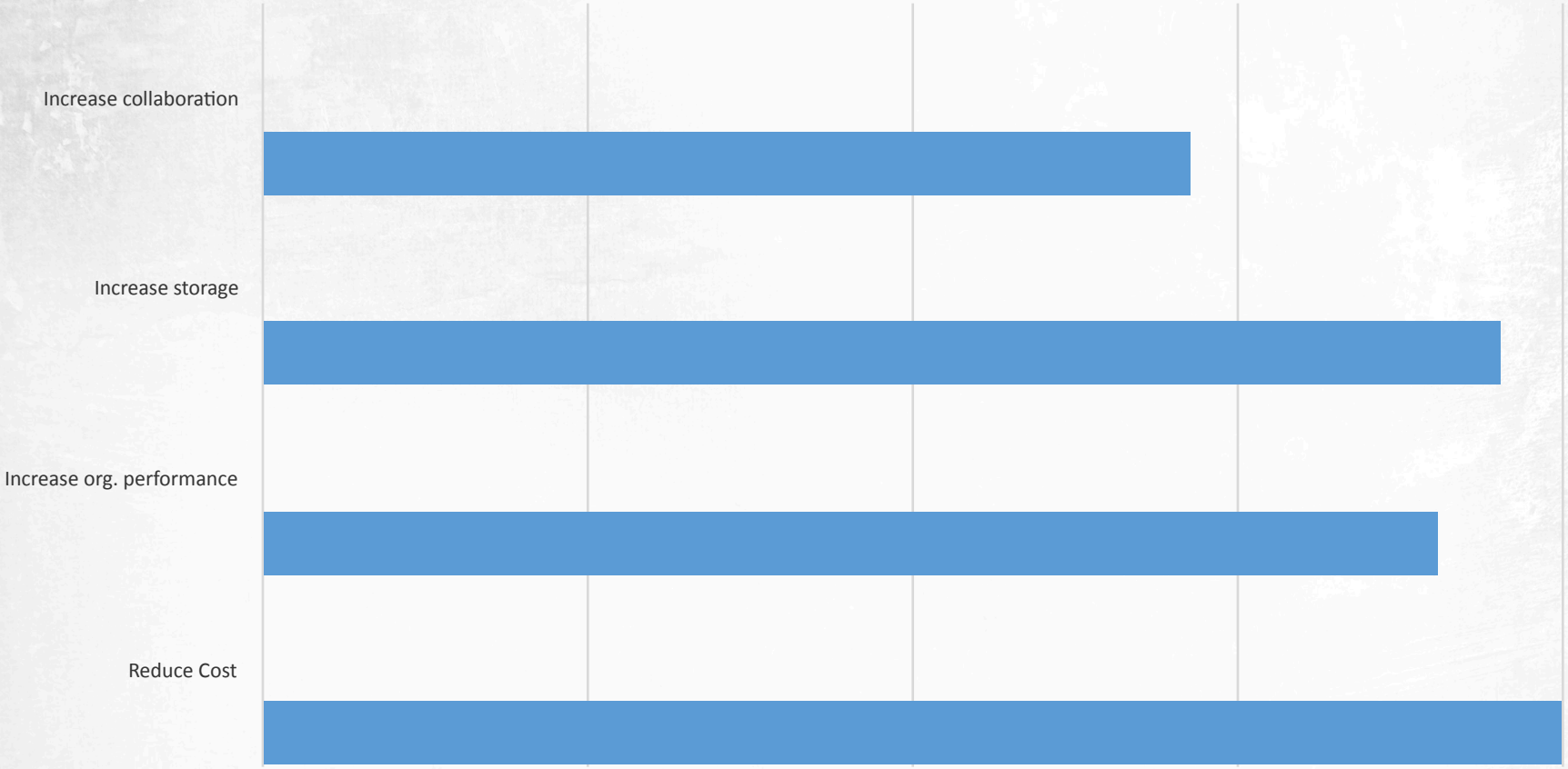
Records in the Cloud

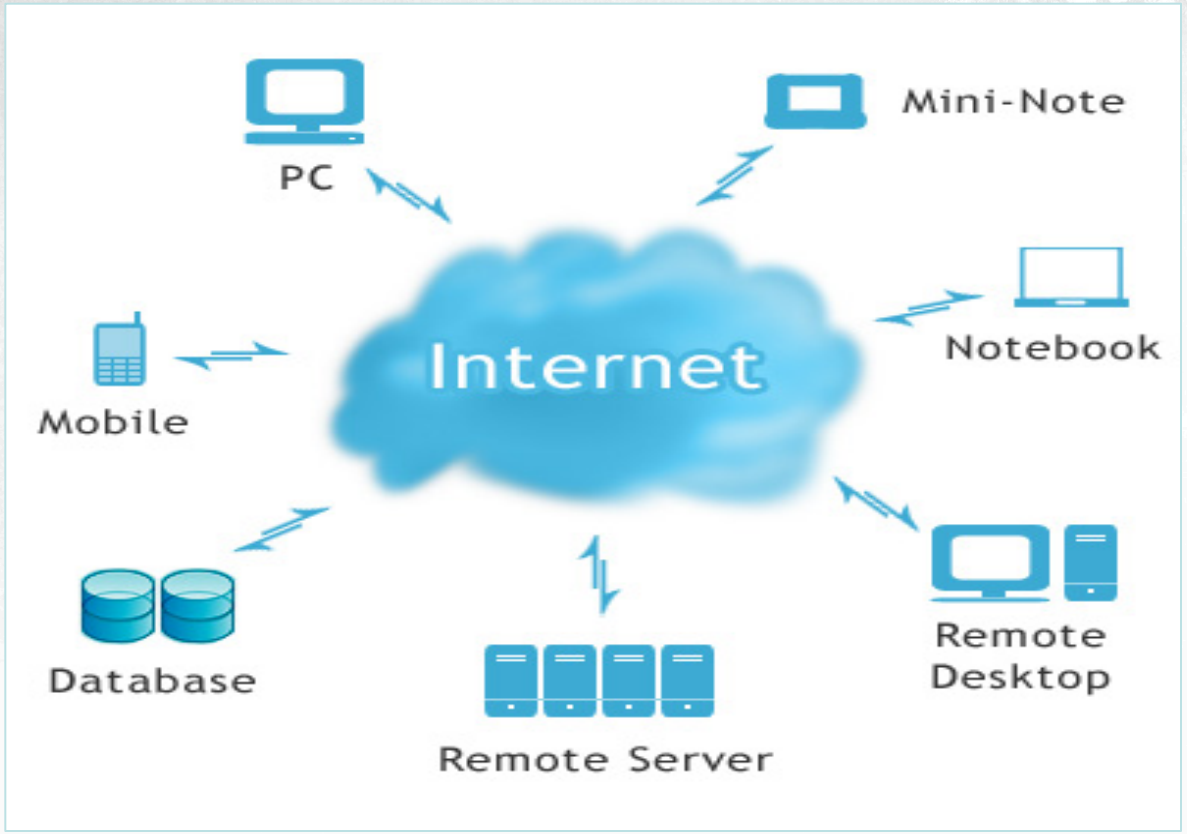
- Cloud computing is a model of services requiring a connecting **network** and **delivered ubiquitously to multiple users**, regardless of the location of the user and the provider's facilities, provisioned **on demand** and **paid proportionally** to usage.
- This model **can be modified as needed**, for example by offering the service only to a specific user, in one single location, off-net, by reservation, or for a flat rate. In fact, many believe that a **hybrid approach** is the best approach to the use of cloud services

InterPARES Trust

Motivations

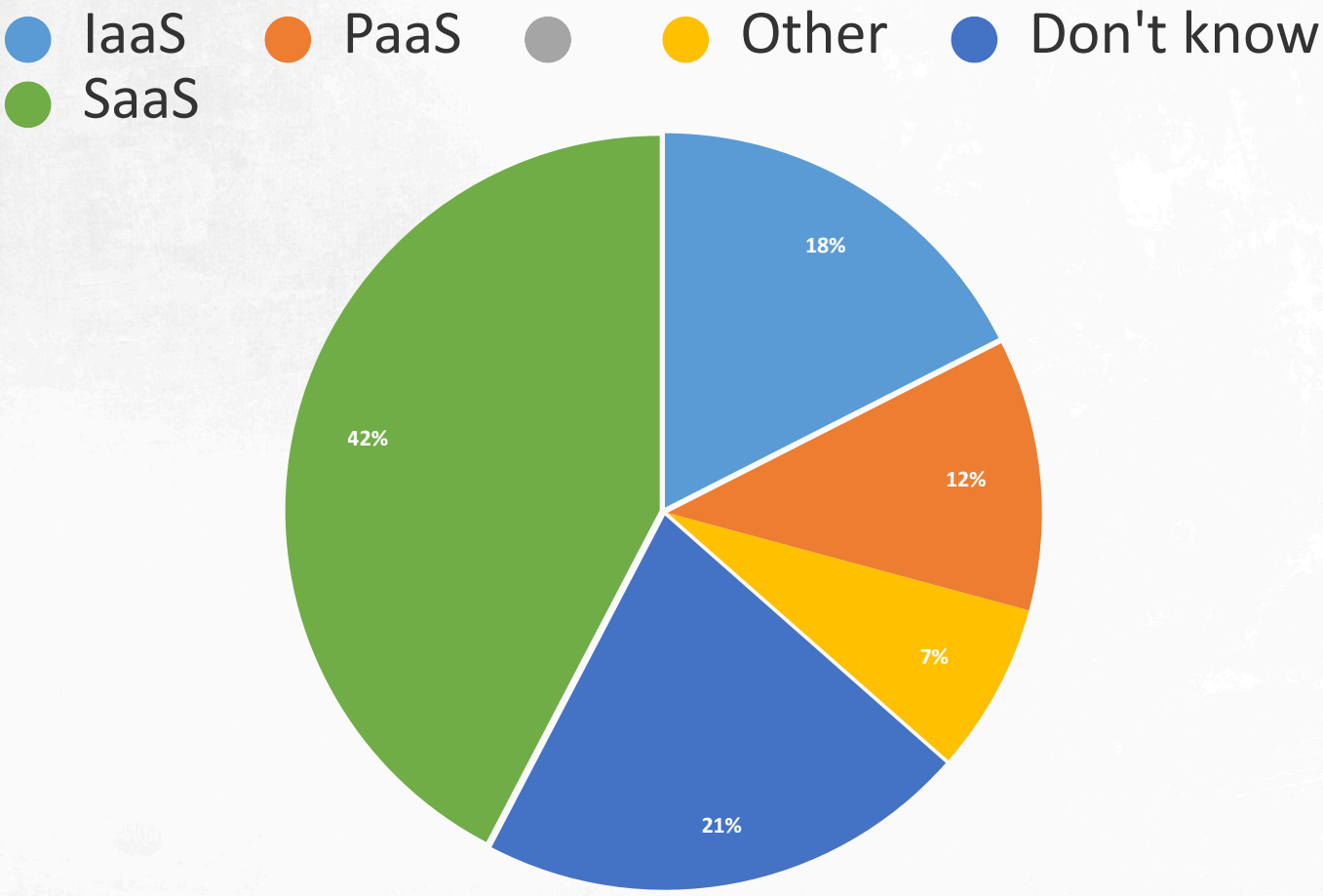
What are the motivations for keeping records online?





Through the internet, many services can be accessed, including Software, Platform, or Infrastructure

Type of Service in Use



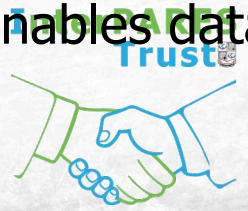
Levels of Deployment

Public cloud (the commercial cloud): The cloud infrastructure is available to the general public and is owned by an organization selling cloud services. Applications like Gmail and Dropbox are part of the public cloud.

Community cloud: The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). A city may use it.

Private cloud: The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise. Most Used (e.g. Preservica).

Hybrid cloud: The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability. Think Google and Amazon.



Issues

- Data ownership
- Availability, access, and reliability
- Storage and maintenance
- Security
- Location and transfer
- End of service
- Preservation

Cloud Computing – Data Ownership

- When a user entrusts its records to a provider and uses the latter's platform and application to generate additional data, the **provider will create data** related to actions about data processing, management, etc.
- While the content created and/or stored in the cloud by the user is owned by such user, **the metadata created by the provider are not**, and, as the user needs them to demonstrate the integrity of the records, contractual agreements should determine whether and how the user has the **right to access and use the provider's metadata**.

InterPARES Trust

Availability, Access, Reliability

- **Availability** is a fact, while **access** is a right, but the latter cannot be satisfied without the former
- In a cloud environment, **availability of the stored records** implies also the **availability of the infrastructure** (i.e. the amount of time that a system is expected to be in service is 100%), which facilitates the retrieval and readability of the data, because technical difficulties might slow a FOIA process and the owner of the data, being liable for providing access to them, may be sanctioned
- **Reliability** is the characteristic of behaving consistently with expectations: one must consider not only availability of the records through redundancy but also **consistency** and **accuracy** of access.

InterPARES Trust

Records Storage and Maintenance

- Storage and maintenance impact the quality of the records and they ability to serve as legal evidence, especially in legal jurisdictions where the **authenticity** of the record is an inference made from the **integrity** of the system where the data reside (CGSB 74:32 2017).
- Contractual agreements do not generally specify how records are maintained **across changing technologies and data formats**, and they generally say that users are responsible for backing up their data. All maintenance procedures, including proper storage, care, custody, and data control, are referred to by providers as "backup procedures."

InterPARES Trust

Records Security

- **Security is the new authenticity**. It has to do with the protection of the records from **unauthorised access, use, alteration or destruction**, which should allow to infer integrity, from which one infers authenticity.
- Individuals protect it 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)
- A cloud provider should **produce audit trails and access logs** and 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.
- **The security issue links directly to the matter of data location and cross-border data flow.**

InterPARES Trust

Records Location and Transfer

- The location of the records might be a criterion in **determining the law that applies** in case of litigation, though usually providers select a jurisdiction compatible with their own legal system (e.g. Patriot Act)
- The international strategy is moving away from requiring that data stay in the jurisdiction of creation, thereby underscoring the importance of **multilateral agreements** among countries for collaboration in security (new safe harbour)
- The cloud is the platform of choice for **mobile applications** and the data generated using them, as well as those created in **smart devices**

InterPARES Trust

End of Service – Contract Termination

- If the provider ceases to exist or terminates one or more of its services (for breach, inactivity, or convenience), the records will be **deleted** or **inaccessible**
- Contracts for paid services address their duration, but free services do not have an established duration and may close accounts **unilaterally**, requiring users to delete software and applications, and preventing them from accessing the data left with the provider
- When the data are given back to the user it is not certain that they will be in a **usable** and **interoperable** format
- If the contract is terminated by the user, the restitution of the data may be **expensive** and the data may not be in accessible formats. Also, the user may not have **the right to access the metadata** generated by the system for its recordkeeping or legal purposes, and may have no guarantee that the provider will **destroy** every copy of the data held in the data centers

InterPARES Trust

Records Preservation

- Preserving records in the cloud is a **black box process** in which you may know what goes in for preservation, and what you want to access and retrieve, but often do not know what technology is used by the providers to manage, store or process their data.
- Providers **may not know where the records are**, can and do **subcontract** some of their services to other providers, potentially maintaining servers or being registered as providers in different countries.
- One cannot expect that the same hardware and software will remain in service for as long as the records must be preserved, or that the technologies replacing them will be **compatible** with the previous ones.
- Standards give information about preservation formats but there is **no way of controlling compliance**

InterPARES Trust

Trustworthiness in Archival Science

Reliability

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

based on:

- the competence of its author
- the controls on its creation

Accuracy

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

based on:

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

Authenticity

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

based on:

- identity
- integrity

InterPARES Trust

Status of Transmission

Trustworthiness based on the **degree of perfection of a record**:

- **Draft** – a document prepared for purposes of correction and meant to be provisional, temporary
- **Original** – the first, complete document capable of reaching the purposes for which it was intended
  - Primitiveness, completeness, effectiveness
- **Copy** – a reproduction of another document, which may be an original, a draft or another copy
  - Authentic copy, facsimile (can be made by anyone), copy in the form of original, imitative, simple, insert, inspeximus (vidimus)

InterPARES Trust

…in the Digital Environment

- **Digitized originals** on other media correspond to a traditional authentic copy, are stable, and every instantiation that is authenticated has the force of original
- **Born digital originals** exist for a nano-second
- Synchronic and diachronic copies are generated for **redundancy** or **distribution** but they are not identical copies
- As mentioned earlier, keeping a born digital record involves **maintaining the ability to reproduce or re-create it in a trustworthy way**

InterPARES Trust

Digital Copies

- In the absence of originals, we need to either make or identify the most **authoritative** copy as the **trustworthy** copy for preservation and use as authentic source
- **Authenticity Metadata** are usually the primary means of providing or identifying the authority of a copy in terms of identity. But what about integrity?

InterPARES Trust

Data Integrity

Often identified with **Bitwise Integrity**:

- The data in the document are not modified either intentionally or accidentally
- The original **bits are in a complete and unaltered state** from the time of capture, that is, they have the exact and same order and value

A small change in a bit means a very different value presented on the screen or action taken in a program or database.

Duplication Integrity

Duplication integrity: it means that the process of creating a copy does not modify a document (either intentionally or accidentally) and the output is an exact bit copy of the original data set (form, content and composition data).

Duplication integrity is **linked to time** and one should consider the use of time stamps for that purpose.

But, in the digital environment, when we say duplication, we need to be explicit about what we mean...

InterPARES Trust

Documentary Duplication: Copy

Selective duplicate (e.g. PDF)

- You can only copy what you can see
- Rarely includes confirmation of completeness
- Provides incomplete picture of the digital environment

InterPARES Trust

Forensic Duplication: Image

A bit by bit reproduction of the storage medium and its content, including ambient data, swap space and slack space (and deleted material)

InterPARES Trust

Process Integrity: Principles

Principle of Non-interference: the method used to re-produce or re-create a digital document does not change the digital entities

Principle of Identifiable interference: if the method used does alter the entities, the changes are identifiable and identified (including para-data)

Authentication

Definition: A declaration of authenticity based on direct knowledge, material proof, inference, or deduction

Basis for authentication of digital records

- **A chain of legitimate custody** remains ground for inferring authenticity and authenticate a record.
- **Digital chain of custody:** the information preserved about the record and its changes that shows specific data was in a particular state at a given date and time.
- A **declaration** made by an expert who bases it on the trustworthiness of the system hosting the record and procedures and processes controlling its preservation and use

InterPARES Trust

Technology Dependent Authentication

Digital signature:

- protects **bitwise integrity**
- verifies a record's origin (part of its **identity**), makes record indisputable and incontestable (**non-repudiation**)
- has been given legal value by legislative acts (e.g., European Directive on electronic signatures) or regulatory bodies (Security Exchange Commission on hash functions)
- is enabled through complex and costly public-key infrastructures (PKI)
- ensures authenticity of information **across space**, not **time**!
- is subject to **obsolescence**, and compounds the problem of preservation, and the certificates have an expiration date

InterPARES Trust

Technology Dependent Authentication

Blockchain technology

- is the underlying technology enabling Bitcoin and many other applications
- is a ledger, i.e. an information store which keeps a final and definitive (immutable) trace of transactions.
- relies upon a **distributed network** and **decentralized consensus**
  - Distributed: all nodes (servers) are equal – no centre(s); no single point of control or attack

Blockchain is a type of distributed ledger technology in which confirmed and validated sets of transactions are held in blocks, which are linked (chained) in a tamper-resistant, append-only chain which starts with a genesis block and where each block contains a hash of the prior block in the chain.

InterPARES Trust

Advantage of a distributed network

Enables **decentralized consensus**

- every participant (node/server) includes every event (transaction) in its ledger ("main book"/database)
- consensus is used in order to
  - ensure that all ledgers are exact copies (i.e. are synchronised)
  - **to determine authenticity**
- An event is valid only if a qualified majority (50%+1 node) agrees upon it

IBM is pushing the British Columbia government to use it for all legal marijuana transactions to ensure shielding from illegal ones.

InterPARES Trust

How Can We Use Blockchain?

Blockchain can be used to confirm

- the **integrity** of a record
- that a record **existed** or **was created** at a certain point in time (i.e. not after being timestamped and registered in the blockchain)
- the **sequence** of records

Is it a **recordkeeping system**? No. It holds the hash of records, not records. The records must still be stored and managed off chain.

InterPARES Trust

Some of the Legal Problems With Blockchain Records

- Proving **reliability** (and therefore enforceability)
- Preserving the **archival bond** (and thus contextual evidence)
- Handling the **decentralized** (and thus trans-jurisdictional) nature of the blockchain
- When the records result from **smart contracts**, dealing with code

InterPARES Trust

Records/Archives in a Blockchain-based system

- InterPARES TRUSTER Preservation Model
  - Blockchain-based system called "**TrustChain**"
  - **VIP** (**V**alidity of **I**nformation **P**reservation) **solution**
  - Applies the concepts of
    - hash algorythms
    - Merkle tree
    - blockchain
    - distributed consensus
  - Presumptions:
    - private cloud blockchain
    - only approved nodes can write
    - everyone can read

InterPARES Trust

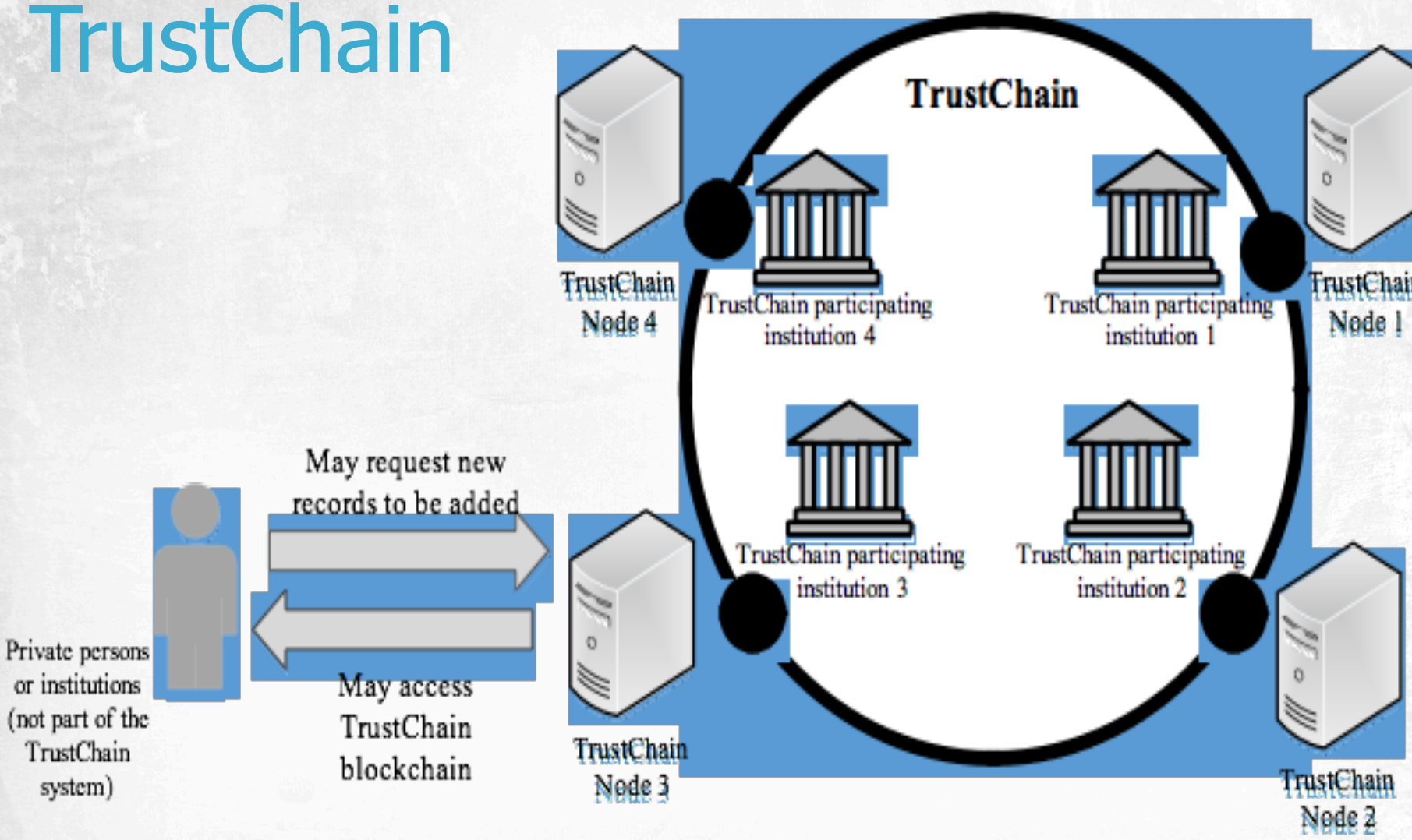
TrustChain

The proposed **TrustChain** system

- relies on the involvement of a **group of trusted institutions**
- **the recordkeeping in the creating office and the preservation systems** would work **in concert with** along the lifecycle of the records
- would provide confirmation of **integrity**, time of **creation/existence**, **sequence of records**, **non-repudiation**, **validity of e-signatures** whose certificate has not expired

InterPARES Trust

TrustChain



InterPARES Trust

Enigio Time "trace:original"

- It establishes a master document that exists as one, and only one, **original,** identifiable as such even where there may exist many copies of the same document.
- It uses Distributed Ledger Technology (DLT) components, and complies with the EU requirements for irrefutable evidence.
- It can be **transferred** from one holder to another, or placed in the custody of a third party
- Any holder can choose how to manage the storage; and only the holder can make **additions and changes** to the original;
- Anyone can **check the validity** of the original through an open interface, independent of their geographic location
- It does not require any central register of who is the holder of the digital originals.

InterPARES Trust

Getting to the Truth

One of the means to access historical truth is **documentary truth,** but understanding **whose truth** we are dealing with requires to

- use traditional principles, concepts and methods
- collaborate with technology experts while cultivating our disciplinary and professional knowledge
- produce functional requirements, tools, methods, and guidelines to ensure institutions' and people's **ability to access** complete factual information based on authentic, accurate and reliable contextualised records and archives

InterPARES Trust

THANK YOU

luciana.duranti@ubc.ca

www.interparestrust.org