



Jenkinson Disrupted?

The InterPARES Journey

Luciana Duranti
Director, InterPARES Project
The University of British Columbia
Jenkinson Lecture
London, UK 16 May 2019

Why InterPARES (International research on Permanent Authentic Records in Electronic Systems)?

Characteristics of born digital records:

- Their **content, structure, and form are not 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**
- They are **vulnerable** (easy to destroy, lose, corrupt, tamper with, or become inaccessible if not protected) yet **persistent** (forever there, if not purposefully destroyed)
- When we save a record, we take it apart in its digital components. When we retrieve it, we create a copy: there are **no originals** in the digital environment
- **Hence, it is not possible to preserve born digital records: we can only preserve the ability to re-produce or re-create it**



The Goal of InterPARES 1 and 2 (1998-2006)

To develop the body of **theory** and **methods** necessary to ensure that 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 **reliable** and maintained and preserved **authentic**, 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.



Need for Shared Concepts and Terminology

Team members:

- **From countries in 5 continents** (Africa joined in 2007):
 - different cultures, languages, traditions, also within the same country
- From different fields:
 - **30% from records and archives disciplines/professions** (among them 25% academics and 75% professionals)
 - **10% from IT, computer science and engineering**
 - **60% were records creators:** from business (e.g. pharmaceutical and chemical industry), arts (e.g. music, dance), sciences (e.g. physics), etc.

We had to agree on the basics for research and communication purposes: what is a record and how we recognize it; what is an authentic record.



Basic Concepts

Based on Jenkinson:

- A **record** – or **archival document** – is any document made or received in the course of activity, and kept for further action or reference (i.e. created)
 - Being a document (i.e. information affixed to a medium), a record has stable content and fixed form
- Because of the circumstances of its creation a record is **natural**, (a by-product of activity), **interrelated** (linked by an archival bond) **impartial** (not created to answer questions researchers may ask of it in the future), and **authentic** (with respect to the creator, if used as an instrument of activity)

To **preserve** a record means to ensure its physical and/or technological stabilization (for the purpose of extending its life indefinitely) and the protection of its intellectual content and relationships

- Digital preservation is the process of maintaining digital materials authentic and accessible during and across different generations of technology over time, irrespective of where they are stored



Did These Concepts Work?

- Like a charm! Why?
- By making archival document and record coincide we were able to move archival preservation (and other archival functions linked to it, such as appraisal – pretty much as Jenkinson wanted) up **to the time of creation**
- They were robust enough to allow us not only to distinguish records from other digital entities—and find ways of protecting their ability to serve as evidence, but also to identify a typology of digital records, from non-dynamic to interactive
- They enabled us to elaborate on the concepts of impartiality and authenticity and define the concept of **trustworthiness**



Trustworthiness

Reliability

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

based on:

- the competence 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



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

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

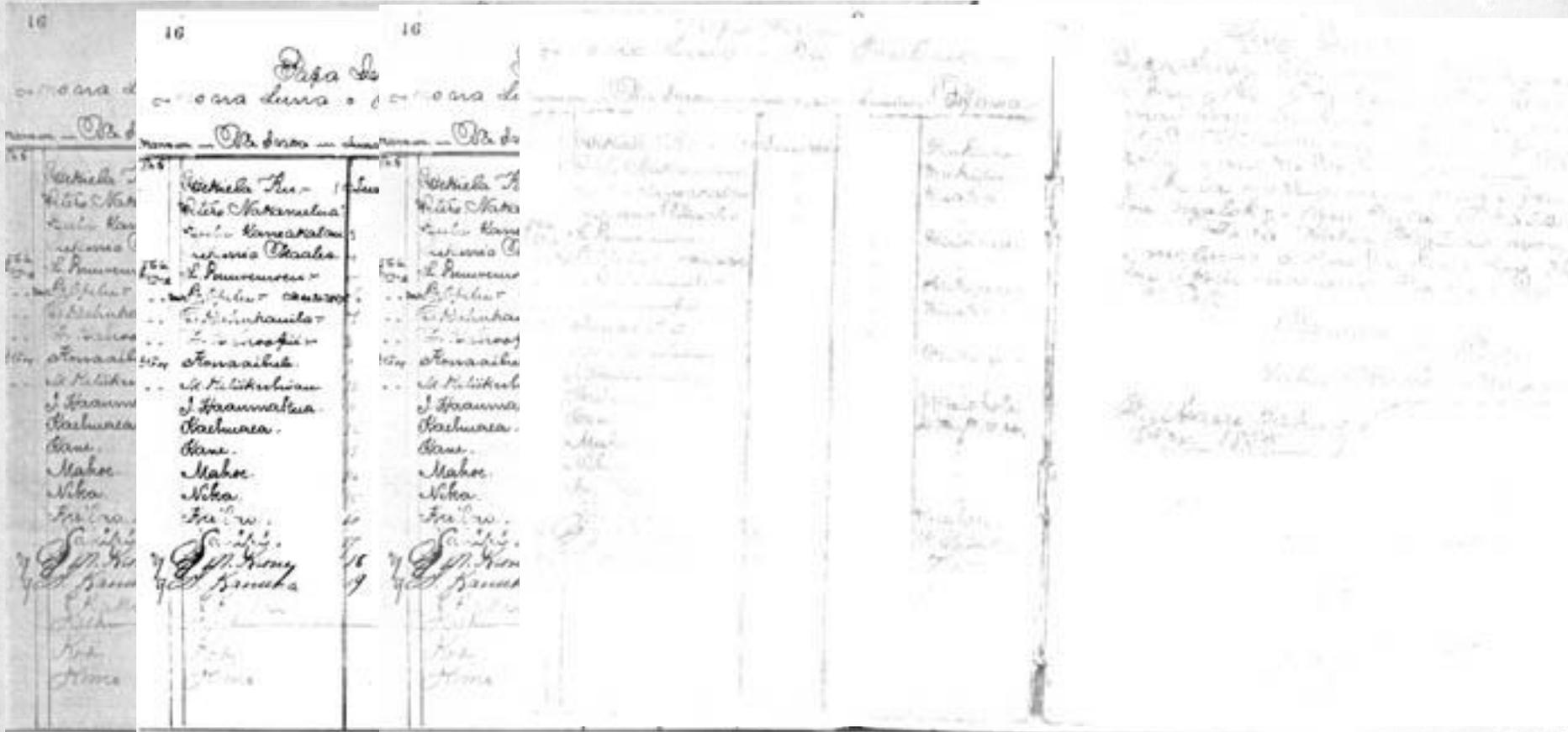
We were never fussy about it. What if a document had holes, was burned on a side or the ink passed through?

The same definition of integrity was used with respect to data, documents, records, copies, records systems

As long as it was good enough to understand it, ...but how good is good enough in the digital environment?



Loss of Integrity: Analog Document



Loss of Integrity: Digital Document

- If Original Bits 101
- Change state to 110
- Continues to a 011

- Same bits, but
Different value



Bitwise Integrity

Often identified with **Data 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

The process of creating a copy does not modify a record (either intentionally or accidentally) as 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, copy or image?



Duplication: Copy

Selective duplicate (e.g. PDF)

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



Duplication: Image

Forensic duplicate:

A bit by bit reproduction of the storage medium and its content, including ambient data (e.g. snapshots of each open file), swap space (virtual memory, with passwords and encryption keys) and slack space (with deleted material)



Duplication 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 either direct knowledge, material proof, inference, or deduction

Basis for authentication of digital records

- A **chain of legitimate custody** remains ground (an increasingly significant ground!) for inferring authenticity and authenticate a record (**Jenkinson!** unbroken chain of custody).
- **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



IP 1 & 2 Products

Policy Framework

A framework of principles guiding the development of policies for records creating and preserving organizations

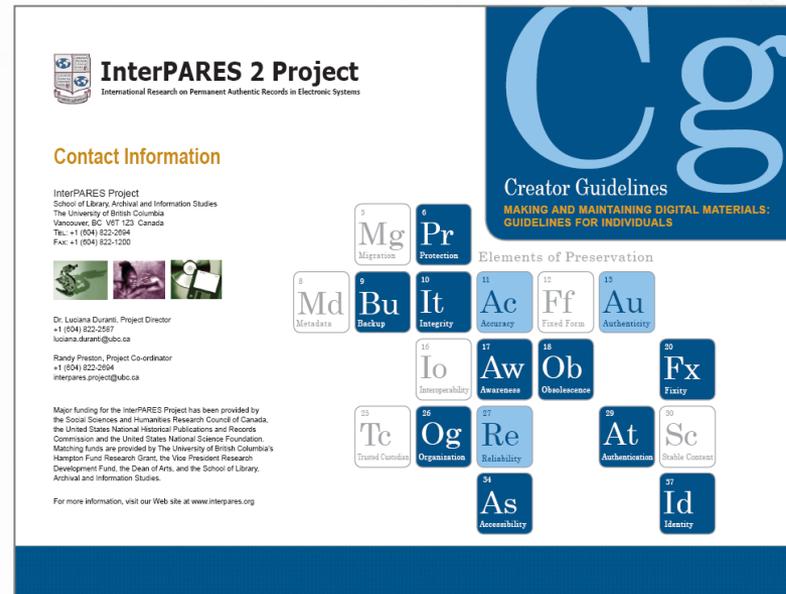
 InterPARES 2 Project International Research on Permanent Authentic Records in Electronic Systems	<p>Policy Framework, v.1.2 (March 2005) L. Duranti, J. Suderman and M. Todd</p> <p>Table of Contents</p> <p>INTRODUCTION 1</p> <p>STRUCTURE OF THE PRINCIPLES 3</p> <p>PRINCIPLES FOR RECORDS CREATORS 4</p> <p>(C1) Digital objects must have a stable content and a fixed documentary form to be considered records and to be capable of being preserved over time. (P1) 4</p> <p>(C2) Record creation procedures should ensure that digital components of records can be separately maintained and reassembled over time. (P4) 5</p> <p>(C3) Record creation and maintenance requirements should be formulated in terms of the purposes the records are to fulfil, rather than in terms of the available or chosen record-keeping or recordkeeping technologies. (P1) 5</p> <p>(C4) Record creation and maintenance policies, strategies and standards should address the issues of record reliability, accuracy and authenticity expressly and separately. (P2) 6</p> <p>(C5) A trusted record-keeping system should be used to generate records that can be presumed reliable. (P3) 7</p> <p>(C6) A trusted recordkeeping system should be used to maintain records that can be presumed accurate and authentic. (P1) 8</p> <p>(C7) Preservation considerations should be embedded in all activities involved in record creation and maintenance if a creator wishes to maintain and preserve accurate and authentic records beyond its operational business needs. (P7) 9</p> <p>(C8) A trusted custodian should be designated as the preserver of the creator's records. (P1) 9</p> <p>(C9) All business processes that contribute to the creation and/or use of the same records should be explicitly documented. (P10) 10</p> <p>(C10) Third-party intellectual property rights attached to the creator's records should be explicitly identified and managed in the record-keeping and recordkeeping systems. (P9) 11</p> <p>(C11) Privacy rights and obligations attached to the creator's records should be explicitly identified and provided in the record-keeping and recordkeeping systems. (P8) 11</p> <p>(C12) Procedures for naming records across different jurisdictions should be established on the basis of the legal requirements under which the records are created. (P11) 12</p> <p>(C13) Reproductions of a record made by the creator in its usual and ordinary course of business and for its purposes and use, as part of its recordkeeping activities, have the same effects as the first submission, and each is to be considered as any given time the record of the creator. (P3) 12</p> <p>PRINCIPLES FOR RECORDS PRESERVERS 13</p> <p>(P1) A designated records preserver fulfils the role of trusted custodian. (C8) 13</p> <p>(P2) Records preservation policies, strategies and standards should address the issues of record accuracy and authenticity expressly and separately. (C4) 14</p> <p>(P3) Reproductions of a creator's records made for purposes of preservation by their trusted custodian are to be considered authentic copies of the creator's records. (C13) 14</p> <p>(P4) Records preservation procedures should ensure that the digital components of records can be separately preserved and reassembled over time. (C2) 15</p> <p>(P5) Authentic copies should be made for preservation purposes only from the creator's records that is from digital objects that have a stable content and a fixed documentary form. (C1) 15</p> <p>(P6) Preservation requirements should be formulated in terms of the purposes or desired outcome of preservation, rather than in terms of the specific technologies available. (C3) 17</p> <p>(P7) Preservation considerations should be embedded in all activities involved in each phase of the records lifecycle in their continuing authentic existence over the long term to be assumed. (C7) 17</p> <p>(P8) Third-party intellectual property rights attached to the creator's records should be explicitly identified and managed in the preservation system. (C10) 19</p> <p>(P9) Privacy rights and obligations attached to the creator's records should be explicitly identified and managed in the preservation system. (C11) 19</p> <p>(P10) Archival appraisal should identify and analyze all the business processes that contribute to the creation and/or use of the same records. (C9) 20</p> <p>(P11) Archival appraisal should assess the authenticity of the records. (C6) 20</p> <p>(P12) Archival appraisal should be used as a proactive authentication of the records in archival fonds. (C5) 20</p> <p>(P13) Procedures for providing access to records created in one jurisdiction to users in other jurisdictions should be established on the basis of the legal environment in which the records were created. (C12) 21</p> <p>InterPARES 2 Project, Policy Cross-domain 1</p>
--	--



IP 1 & 2 Products

Creator Guidelines

Recommendations for making and maintaining digital materials for individuals and small communities of practice



InterPARES 2 Project
International Research on Permanent Authentic Records in Electronic Systems

Contact Information

InterPARES Project
School of Library, Archival and Information Studies
The University of British Columbia
Vancouver, BC V6T 1Z3 Canada
Tel: +1 (604) 822-2604
Fax: +1 (604) 822-1200

Dr. Luciana Duranti, Project Director
+1 (604) 822-2637
luciana.duranti@ubc.ca

Randy Preston, Project Co-ordinator
+1 (604) 822-2604
inter pares.project@ubc.ca

Major funding for the InterPARES Project has been provided by the Social Sciences and Humanities Research Council of Canada, the United States National Historical Publications and Records Commission and the United States National Science Foundation. Matching funds are provided by The University of British Columbia's Hargrave Fund Research Grant, the Vice President Research Development Fund, the Dean of Arts, and the School of Library, Archival and Information Studies.

For more information, visit our Web site at www.interpares.org

Creator Guidelines
MAKING AND MAINTAINING DIGITAL MATERIALS:
GUIDELINES FOR INDIVIDUALS

Elements of Preservation

5 Mg Migration	6 Pr Protection			
8 Md Metadata	9 Bu Backup	10 It Integrity	11 Ac Accuracy	12 Ff Fixed Form
	13 Io Interoperability	17 Aw Awareness	18 Ob Obsolescence	20 Fx Fixity
25 Tc Trusted Custody	26 Og Organization	27 Re Reliability	29 At Authentication	30 Sc Stable Content
		34 As Accessibility		37 Id Identity



IP 1 & 2 Products

Preserver Guidelines

Recommendations for digital preservation for archival institutions



InterPARES 2 Project
International Research on Permanent Authentic Records in Electronic Systems

Contact Information

InterPARES Project
School of Library, Archival and Information Studies
The University of British Columbia
Vancouver, BC V1T 1Z3, Canada
Tel: +1 (604) 822-2694
Fax: +1 (604) 822-1200



Dr. Ludiana Duranti, Project Director
+1 (604) 822-2687
ludiana.duranti@ubc.ca

Randy Preston, Project Co-ordinator
+1 (604) 822-2694
interpares.project@ubc.ca

Major funding for the InterPARES Project has been provided by the Social Sciences and Humanities Research Council of Canada, the United States National Historical Publications and Records Commission and the United States National Science Foundation. Matching funds are provided by The University of British Columbia's Hamilton Fund Research Grant, the Vice-President Research Development Fund, the Dean of Arts, and the School of Library, Archival and Information Studies.

For more information, visit our Web site at www.interpares.org



Pg
Preserver Guidelines
PRESERVING DIGITAL RECORDS:
GUIDELINES FOR ORGANIZATIONS

Elements of Preservation

5 Mg Migrating	6 Fe Feasibility				
8 Be Baseline Requirements	9 Id Identifying	10 Pr Preserving	11 Ac Accuracy	12 St Storing	13 Au Authenticity
		16 De Describing	17 Ma Managing	18 Ob Obsolescence	20 Mo Monitoring
26 Tc Trustworthy	29 Op Outputting	31 Ba Baseline Requirements	33 Ap Appraising	35 Tr Transferring	
		34 Ac Acquiring			37 Do Documenting



IP 1 & 2 Products

Benchmark and Baseline Authenticity Requirements

Requirements for assessing and maintaining the authenticity of digital records

<< REQUIREMENT SET A >>

To support a presumption of authenticity the preserver must obtain evidence that:

REQUIREMENT A.1: Expression of Record Attributes and Linkage to Record
The value of the following attributes are explicitly expressed and inextricably linked to every record. These attributes can be distinguished into categories, the first concerning the content of the record, and the second concerning the integrity of records.

A.1.a Identity of the record:

- A.1.a.i** Names of the persons concurring in the formation of the record,
- name of author^a
- name of writer^b (if different from the author)
- name of originator^c (if different from name of author or writer)
- name of addressee^d

A.1.a.ii Name of action or matter

A.1.a.iii Date(s) of creation and transmission, that is:

- chronological date^e
- received date^f
- archival date^g
- transmission date(s)^h

A.1.a.iv Expression of archival bondⁱ (e.g., classification code, file identifier)

A.1.a.v Indication of attachments

A.1.b Integrity of the record:

A.1.b.i Name of handling office^j

A.1.b.ii Name of office of primary responsibility^k (if different from handling office)

A.1.b.iii Indication of types of annotations added to the record^l

A.1.b.iv Indication of technical modifications^m

REQUIREMENT A.2: Access Privileges

The creator has defined and effectively implemented access privileges concerning modification, annotation, relocation, and destruction of records.

<< REQUIREMENT SET A (cont) >>

REQUIREMENT A.3: Protective Procedures: Loss and Corruption of Records

The creator has established and effectively implemented procedures to prevent, discover, correct loss or corruption of records.

REQUIREMENT A.4: Protective Procedures: Media and Technology

The creator has established and effectively implemented procedures to guarantee the content identity and integrity of records against media deterioration and across technological change.

REQUIREMENT A.5: Establishment of Documentary Forms

The creator has established the documentary forms of records associated with each procedure either according to the requirements of the juridical system or those of the creator.

REQUIREMENT A.6: Authentication of Records

If authentication is required by the juridical system or the needs of the organization, the creator has established specific rules regarding which records must be authenticated, by whom, and the means of authentication.

REQUIREMENT A.7: Identification of Authoritative Record

If multiple copies of the same record exist, the creator has established procedures that identify which record is authoritative.

REQUIREMENT A.8: Removal and Transfer of Relevant Documentation

If there is a transition of records from active status to semi-active and inactive status, it involves the removal of records from the electronic system, the creator has established effectively implemented procedures determining what documentation has to be removed and transferred to the preserver along with the records.

<< REQUIREMENT SET B >>

The preserver should be able to demonstrate that:

REQUIREMENT B.1: Controls over Records Transfer, Maintenance, and Reproduction
The procedures and system(s) used to transfer records to the archival institution or program; maintain them; and reproduce them embody adequate and effective controls to guarantee the records' identity and integrity, and specifically that:

B.1.a Unbroken custody of the records is maintained;

B.1.b Security and control procedures are implemented and monitored; and

B.1.c The content of the record and any required annotations and elements of documentary form remain unchanged after reproduction.

REQUIREMENT B.2: Documentation of Reproduction Process and its Effects

The activity of reproduction has been documented, and this documentation includes:

B.2.a The date of the records' reproduction and the name of the responsible person;

B.2.b The relationship between the records acquired from the creator and the copies produced by the preserver;

B.2.c The impact of the reproduction process on their form, content, accessibility and use; and

B.2.d In those cases where a copy of a record is known not to fully and faithfully reproduce the elements expressing its identity and integrity, such information has been documented by the preserver, and this documentation is readily accessible to the user.

REQUIREMENT B.3: Archival Description

The archival description of the fonds containing the electronic records includes—in addition to information about the records' juridical-administrative, provenancial, procedural, and documentary contexts—information about changes the electronic records of the creator have undergone since they were first created.



IP 1 & 2 Products

File Format Selection Guidelines

Principles and criteria for adoption of file formats, wrappers and encoding schemes

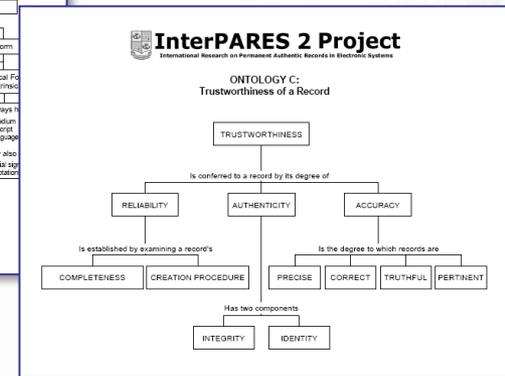
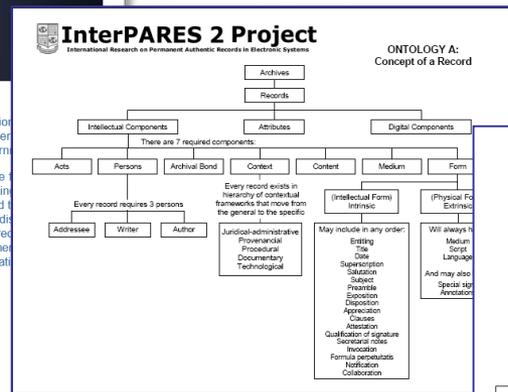
<p> InterPARES 2 Project International Research on Permanent Authentic Records in Electronic Systems</p> <p>Title: General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation</p> <p>Status: Final (public) Version: 1.1 Release: March 2007 Author: The InterPARES 2 Project Writer(s): Evelyn Peters McLellan Project Unit: Domain 3 (Methods of Appraisal & Preservation) URL: http://www.interpares.org/display_file.cfm?doc=ip2_file_formats(complete).pdf [English] http://www.interpares.org/display_file.cfm?doc=ip2_file_formats_fichiers_numeriques.pdf [French]</p>	<p>Selecting Digital File Formats for Long-Term Preservation E. McLellan</p> <p style="text-align: center;">Table of Contents</p> <table><tr><td>Introduction</td><td>1</td></tr><tr><td>1. Terminology</td><td>1</td></tr><tr><td>1.1 What is a file format?</td><td>1</td></tr><tr><td>1.2 "Open" file formats</td><td>3</td></tr><tr><td>1.3 "Standard" file formats</td><td>4</td></tr><tr><td>1.4 "Stable" file formats</td><td>6</td></tr><tr><td>1.5 Standardizing terms</td><td>6</td></tr><tr><td>2. Selection criteria</td><td>6</td></tr><tr><td>2.1 Widespread use</td><td>6</td></tr><tr><td>2.2 Non-proprietary origin</td><td>7</td></tr><tr><td>2.3 Availability of specifications</td><td>8</td></tr><tr><td>2.4 Platform independence (interoperability)</td><td>9</td></tr><tr><td>2.5 Compression</td><td>10</td></tr><tr><td>2.6 Discussion of criteria</td><td>11</td></tr><tr><td>3. Policy implications</td><td>13</td></tr><tr><td>4. Recommendations for developing and implementing policies</td><td>16</td></tr><tr><td>Appendix A: list of repositories reviewed</td><td>18</td></tr><tr><td>Appendix B: URLs of documents reviewed</td><td>19</td></tr><tr><td>Bibliography</td><td>21</td></tr></table> <p style="text-align: center;">InterPARES 2 Project, Domain 3 v1.1 (March 2007)</p>	Introduction	1	1. Terminology	1	1.1 What is a file format?	1	1.2 "Open" file formats	3	1.3 "Standard" file formats	4	1.4 "Stable" file formats	6	1.5 Standardizing terms	6	2. Selection criteria	6	2.1 Widespread use	6	2.2 Non-proprietary origin	7	2.3 Availability of specifications	8	2.4 Platform independence (interoperability)	9	2.5 Compression	10	2.6 Discussion of criteria	11	3. Policy implications	13	4. Recommendations for developing and implementing policies	16	Appendix A: list of repositories reviewed	18	Appendix B: URLs of documents reviewed	19	Bibliography	21
Introduction	1																																						
1. Terminology	1																																						
1.1 What is a file format?	1																																						
1.2 "Open" file formats	3																																						
1.3 "Standard" file formats	4																																						
1.4 "Stable" file formats	6																																						
1.5 Standardizing terms	6																																						
2. Selection criteria	6																																						
2.1 Widespread use	6																																						
2.2 Non-proprietary origin	7																																						
2.3 Availability of specifications	8																																						
2.4 Platform independence (interoperability)	9																																						
2.5 Compression	10																																						
2.6 Discussion of criteria	11																																						
3. Policy implications	13																																						
4. Recommendations for developing and implementing policies	16																																						
Appendix A: list of repositories reviewed	18																																						
Appendix B: URLs of documents reviewed	19																																						
Bibliography	21																																						

IP 1 & 2 Products

Terminology Database

Including a glossary, a dictionary and ontologies. The glossary became the ICA Multilanguage Terminology Database

The screenshot shows the website interface with a navigation menu (Home, About Us, etc.), a search bar, and a list of letters (A through W) for navigating the terminology database. The main content area includes a description of the database's purpose and links to the Glossary, Dictionary, and Ontologies.

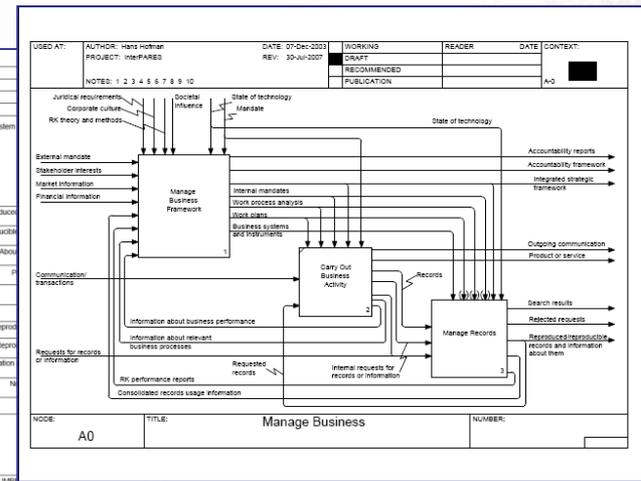
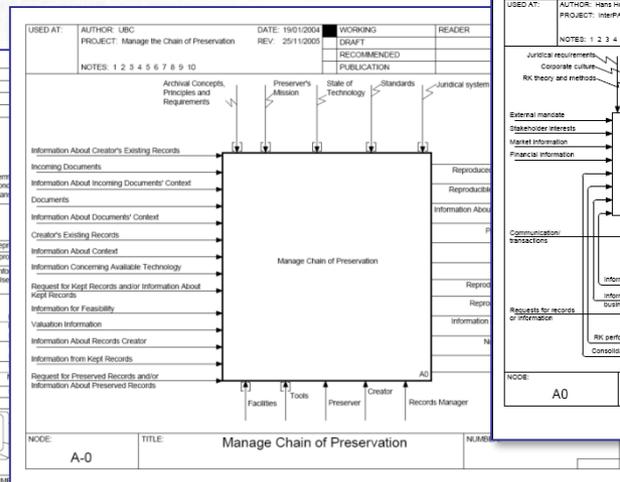
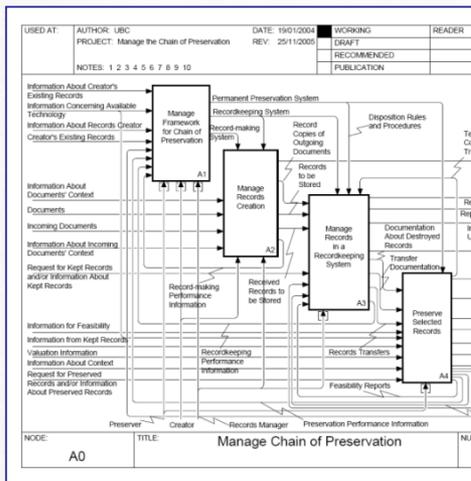


IP 1 & 2 Products

Two Records Management Models

Chain of Preservation (COP) Model (lifecycle)

Business-driven Recordkeeping (BDR) Model (continuum)



IP 1 & 2 Final Products

Two books:

Luciana Duranti, ed. *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (San Miniato: Archilab, 2005).

Available on line at

<http://www.interpares.org/book/index.cfm>

Luciana Duranti and Randy Preston, eds. *InterPARES 2: Interactive, Dynamic and Experiential Records* (Roma: ANAI, 2008). Available on line at

<http://www.interpares.org/ip2/book.cfm>.

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 1 & 2

InterPARES
Trust



Some InterPARES 3 Studies

- Community Archives e-Records Assessment
- Public Sector Audit Report for Digital Recordkeeping
- Open Source Records Management Software
- **Metadata Applications Profiles**
- Organizational Culture & Risk Assessment
- **E-mail Preservation**
- Preservation of Registries
- Education Modules on Digital Preservation



InterPARES Impact

- Legislation: Italy, China
- Standards: DOD 5015.2 (2007), MoReq 2 (2008), OAIS (2009), ARMA (e.g. electronic messages 2012 & 2018), CGSB (72-34: 2005 and 2017), ISO 30300 series
- Policies and Procedures: participating countries in both public and private sectors
- ICA Education Modules and Multilingual Archival Database

All InterPARES products are available at www.interpares.org



Records in the Cloud

What are the motivations for keeping records online?



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**, to ensure public trust grounded on evidence of good governance, a strong digital economy, and a persistent digital memory.

InterPARES
Trust



Issues

- Data ownership
- Availability, access, and reliability
- Retention and disposition
- Storage and maintenance
- Security
- Location and transfer
- End of service
- Preservation
- Users Behaviour
- Trustworthiness

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**.



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.



Records Retention and Disposition

- **Compliance** is difficult to verify.
 - **transfer** from a system to another for retention might involve loss of authenticity
 - **destruction** might involve
 - a breach of confidentiality or privacy,
 - persistence of some of the copies and related metadata, and
 - persistence of the metadata generated by the provider about the user's data.



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 (Canadian Government Standard Board 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.”



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 from which one infers integrity of the record, from which one infers its authenticity and then trustworthiness, **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)
- A cloud provider enforces it through encryption and 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.**



Records Location and Transfer

- The cloud is the platform of choice for **mobile applications** and the data generated using them, as well as those created in **smart devices**. 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
- 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)



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**
- 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



Records Preservation

- Preserving records in the cloud is a **black box process**
- 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**



Users Behaviour: Reuse

- Reuse is often *remix*, which results in **derivative works** that may substantively change the intent and context of the appropriated material.
- Reuse involves **successive cycles of use, modification, repurposing**, and of course **take-down notices**.



User Behaviour: Sharing

- People **share profiles** for a variety of social media and web sites, and **post material** they consider informative.
- Social media platforms facilitate the movement of material **from one circle of people to another**.
- Groups assemble and change **stories of activities and events**.



Sharing (cont.)

- Dynamic groups of employees from public agencies collectively create a body of **interlinked material** related to a work project (e.g. gcpedia), a common interest, putting into question who owns it and what the context is.
- Contributions to social media by people, programs, committees, or agencies **now dead are linked to ongoing, active contributions** of the living, or disappear, or appear as created by their successors.
- **Digital lives, activities, initiatives are linked to each other.**



User Behaviour: Control and Access

- Massive amounts of **data about individuals** in the cloud are controlled by corporations and governments (**big data**)
- **Medical records** including genotyping or gene sequencing data, medical history, prescription and insurance information, tests, and images are **held by the medical establishment in community clouds** which makes them accessible.
- **Genealogies**, factual biographies, biographic data and personal images available on the web **belong to the private platform hosting them** (Instagram, Facebook, Google).
- In addition to big data, open data and open government raise the issue of **traceability to and identification of the source records (lineage)**



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 as it cannot be migrated with the record it is attached to, and the certificates have an expiration date
- Theory tells us that it has the function of a **seal**, rather than of a signature, so it can be removed and substituted with metadata



Technology Dependent Authentication

Blockchain technology

- the underlying technology enabling Bitcoin
- a ledger, i.e. an information store which keeps a final and definitive (immutable) trace of transactions (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.



How is Blockchain used?

Blockchain can be used to confirm

- the **integrity** of a record kept elsewhere
- 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 uploading of records to the blockchain

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



Legal Problems with Blockchain Authenticated Records

- Proving **reliability, accuracy** and **authenticity at origin** (impartiality and authenticity)
- Preserving the **archival bond** and contextual evidence (naturalness and interrelatedness)
- Handling the **decentralized** (and thus trans-jurisdictional) nature of the blockchain (who is the creator?)
- When the records result from **smart contracts**, dealing with code in a situation where the necessary components of a record are controlled by different actors in different jurisdictions. Plus, no required elements like a signature and the date of the completion of the agreement.



Jenkinson Disrupted

- The record has disintegrated, pulverized, in a myriad data: if data are the new record, what about context? **Interrelatedness** among what?
- Trustworthiness is an inference and not verifiable: if security and immutability are the new trustworthiness, what about reliability, identity, and integrity? **Impartiality** and **authenticity** on the basis of what?
- Technological control is the new panacea: if the creator and the preserver delegate responsibility to technology and their providers, what about the usual and ordinary course of business? **Naturalness** of what?
- **Can the record as we knew it continue to exist?** Can legislation on the duty to document resurrect it? Should we even wish for it? Shouldn't we go our merry way trusting the "information" that "we feel like" trusting and dispense with the idea of the trustworthy record altogether?



The InterPARES Reply

Context:

- ISO has embedded in its latest records management standard (2016) the most controversial of Jenkinson ideas, appraisal by the creator, and its appraisal technical report (2018) supports the idea that long term value to the creator coincides with historical value (as in the Grigg report and the Public Records Act of 1958)
- the information environment in which we operate agrees that preservation consists of protecting the nature of the record as intended by its creator (the moral defence of archives), as Jenkinson believed
- the need for the evidentiary capacity of information is made compelling and non renounceable by the misinformation and disinformation circulating in society: Jenkinson's sanctity of evidence



The InterPARES Reply

Conclusion:

The archival document, or record – natural, interrelated, impartial and authentic, together with the archives in which it belongs, must continue to form the **infrastructure through which beliefs and values are upheld and understood by society.**

We have no choice but to hold on to it!

InterPARES
Trust



Records/Archives in a Blockchain-based system

- InterPARES TRUSTER Preservation Model
 - Blockchain-based system called “**TrustChain**”
 - Applies the concepts of
 - hash algorithms
 - blockchain
 - distributed consensus
 - Presumptions:
 - private cloud blockchain
 - only approved nodes can write
 - everyone can read



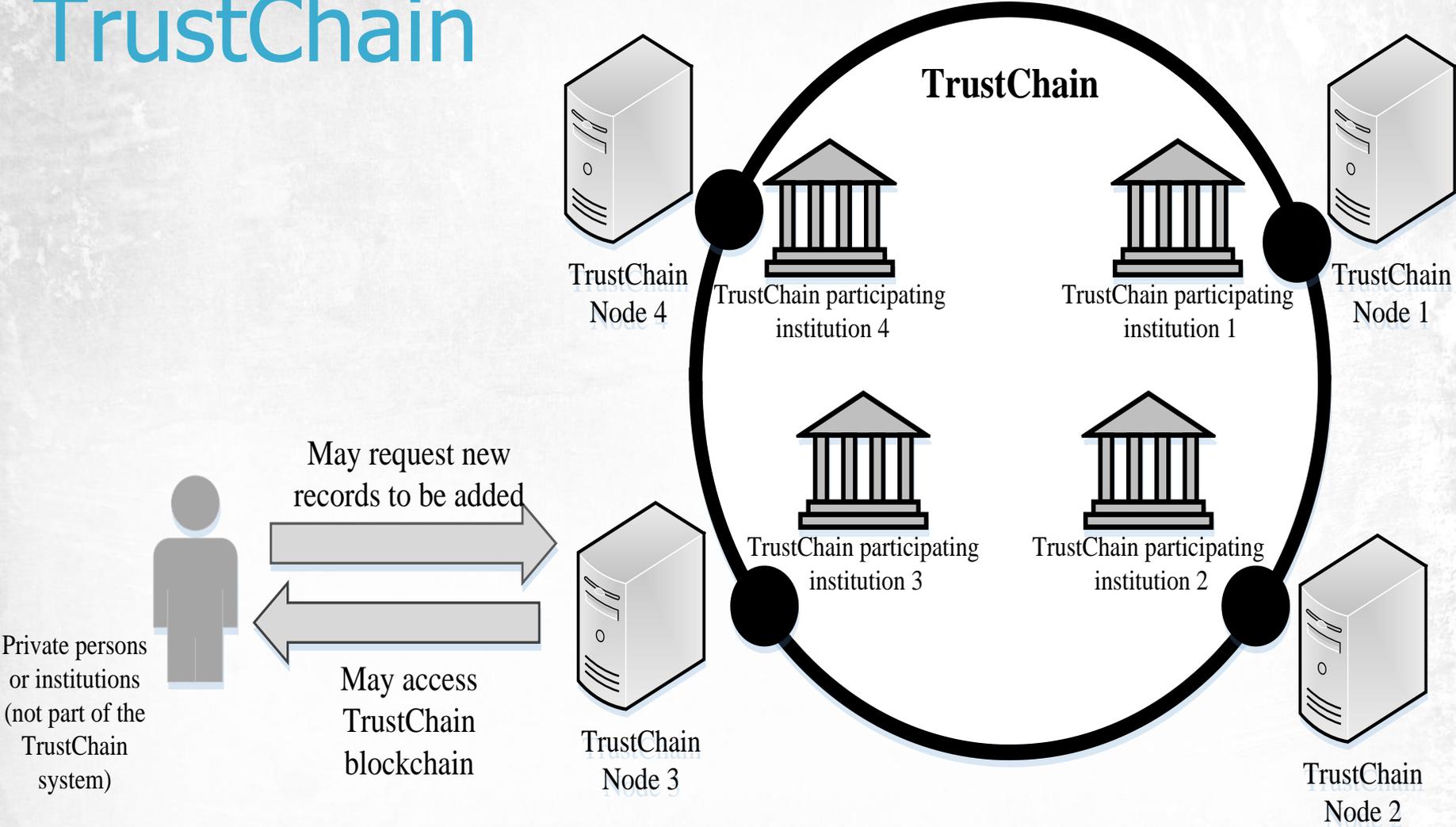
TrustChain

The proposed **TrustChain** system

- relies on the involvement of a **group of trusted institutions**
- **the recordkeeping system in the creating office and the preservation system in the archives** would work **in concert** 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



TrustChain



InterPARES Products Supporting Trustworthines

Security Classification:

https://interparestrust.org/assets/public/dissemination/TR03-Checklist-2_final.pdf

Cloud Services Contracts:

https://interparestrust.org/assets/public/dissemination/NA14_20160226_CloudServiceProviderContracts_Checklist_Final.pdf

https://interparestrust.org/assets/public/dissemination/NA14_final_report_v5-1.pdf

Preservation as a Service for Trust:

https://interparestrust.org/assets/public/dissemination/PreservationasaServiceforTrust1_0-FINAL1.pdf

InterPARES
Trust



Jenkinson Disrupted?

The InterPARES research findings show that, far from being disrupted by emerging technologies, **Jenkinson's ideas are alive and well and more useful in the digital environment than they ever were in the analogue one.**

www.interparestrust.org

InterPARES
Trust



THANK YOU

luciana.duranti@ubc.ca

InterPARES
Trust

