AI for Archives, by Archives, in Archives
I Trust AI

Luciana Duranti, InterPARES Principal Investigator Professor Emerita, School of Information
The University of British Columbia, Vancouver, Canada

Marburg, Archivwissenschaftliches Kolloquium
“KI im Archiv – Steht uns eine Revolution bevor?”
20. – 21. Mai 2025

InterPARES Trust AI (2021-2026)

The goal of the fifth phase of InterPARES, I Trust AI, is to design, develop, and leverage Artificial Intelligence to support the ongoing availability and accessibility of trustworthy public records. We aim to

- Identify specific AI technologies that can address critical records challenges;
- Determine the benefits and risks of using AI technologies on records; and
- Ensure that records concepts and principles inform the development of responsible AI

InterPARES TrustAI

Participants

- 101 partner organizations in 42 countries (in 5 continents)
- 131 co-applicants (academics)
- 129 collaborators (professionals)
- 3 postdocs
- 60 Graduate Academic Assistants in any given year

InterPARES TrustAI

Approach

- The fact that the *I Trust AI* project is a **multinational interdisciplinary endeavour** means that our first effort had to be to **understand each other, starting with the language we use**. For example, archival professionals talk about **records**, while computer scientists and AI professionals talk about **data**. To archivists, data are the smallest meaningful unit of information in a record. To an AI specialist, data is (note: singular) organized information (possibly in a database), be it facts or not, regardless of size, nature and form.
- Thus, key to our work have been **AI tutorials and workshops** for non-AI researchers, and **archival and diplomatics theory tutorials** for non archival researchers. These educational endeavours are supported by the **Terminology Database** which is developed in collaboration by a multidisciplinary team.

InterPARES TrustAI

Indirect Outcomes

- New Professionals: by the end of the project, there will be well over 100 professionals who will have worked as student research assistants on case studies with test-bed organizations and who will spread the acquired knowledge, without counting all the future professionals taught such knowledge during their course of study
- Students from other disciplines: computer scientists, lawyers, etc. will understand and value the archival perspective in their work and the impact of records and recordkeeping on the broader society
- Knowledge co-creation: the project will enrich research in archival science, records management, AI, cybersecurity, information science, law, and ethics, through knowledge exchange and uptake between scholars and practitioners within and among those disciplines.
- Sensitizing AI developers, scholars, and other members of that community to the role of AI in record keeping and archival preservation and to the role of archival concepts and principles in AI design and development.

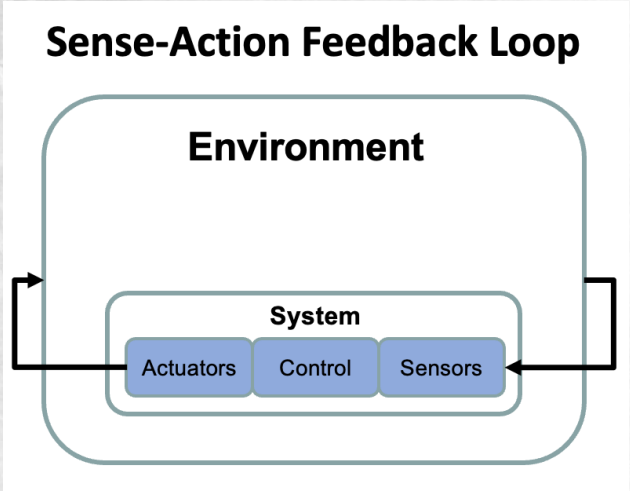
InterPARES TrustAI

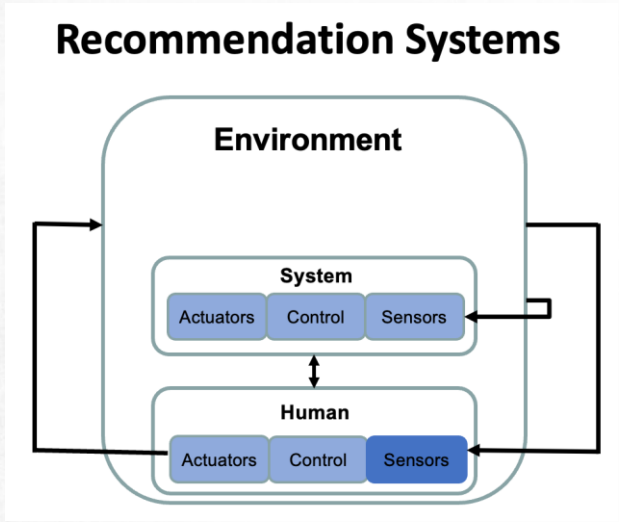
Studies

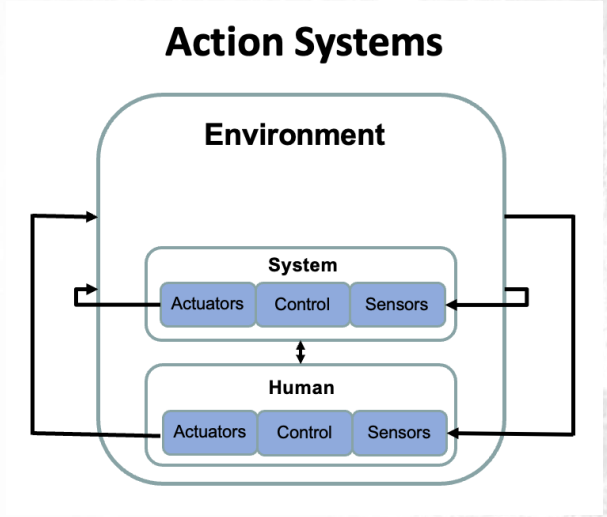
- Studies are **all international and interdisciplinary**
- Focus on all aspects of archival functions
  1. Creation and use of trustworthy records
  2. Appraisal and acquisition of archival material
  3. Arrangement and description
  4. Retention and preservation
  5. Management and administration of records and archives
  6. Reference and access

InterPARES TrustAI

Types of Environments







Continuous cycle:

- Sensors (HW or SW) measure the real world,
- Measurements are fed into & inform the control (AI) processes
  - Control processes determine responses to real-world stimuli
- Actuators (HW or SW) execute or effect the responses to real-world stimuli
- Consequence of system actions are measured by the system's sensors

1. Autonomous (no humans involved) – e.g. Perganet to identify notarial signa (for A)
2. Recommendation (humans marginally involved) – UNESCO study (by A)
3. Action System (humans with shared agency) – Digital Twins (in A)

InterPARES TrustAI

Shared Agency

Humans and the AI agent can make independent decisions and take independent actions.

- **Concurrent action**
  - Disjoint/Joint (overlapping)
  - Conflicts
  - Types: Correction, Enhancement
  - Overriding power
- **One at a time action**
  - Handoff
    - Boundaries
  - Emergencies
    - Who decides?

InterPARES TrustAI

Issues

Dynamic environments

- Time constraints; Real-time requirements
- Multi-Agent (collaborative, competitive)

Different degrees of autonomy

Who is Accountable for what action?

InterPARES TrustAI

Digital Twins

The National Academies Report provides a definition that abstracts the essential elements for all digital twins, namely, a DT

> "is a set of virtual information constructs [representation] that mimics the structure, context, and behavior of a natural, engineered, or social system (or system-of-systems), that is **dynamically updated with data from its physical twin**, has a predictive capability, and informs decisions that realize value. **The bidirectional interaction between the virtual and the physical is central to the digital twin**".*

* National Academies of Sciences, Engineering, and Medicine. 2024. Foundational Research Gaps and Future Directions for Digital Twins. Washington, DC: The National Academies Press. https://doi.org/10.17226/26894.

Application Areas: Engineering (factory, aerospace, defence); Bio Medical (Human twin, organs); Atmospheric & Climate Sciences (Simulation earth systems, modeling); Urban; 3d mapping

InterPARES TrustAI

Digital Twins Case Study

- The study is led by *Tracey Lauriault at Carlton University (edited slides from her presentation).*
- **Research question**: If digital twins intermediate and automate actions and decisions that affect social and material outcomes how should records about those actions, decisions and processes be managed and preserved in the long term?
- Can AI/ML enable that preservation of those records, and how would AI/ML processes in the Digital Twin be preserved?

D.T. Case Study Purposes

- Study the making of the Carleton University Immersive Media Studio (CIMS) Digital Twin
- Test the preservation of this Digital Twin
- Assess whether AI can automate the preservation of Digital Twins & their related technologies
- Explore how the AI/ML + sensor/metre + Automation + code in the Digital Twin can be preserved
- Create results that inform the making of Digital Twins w/preservation in mind

InterPARES TrustAI

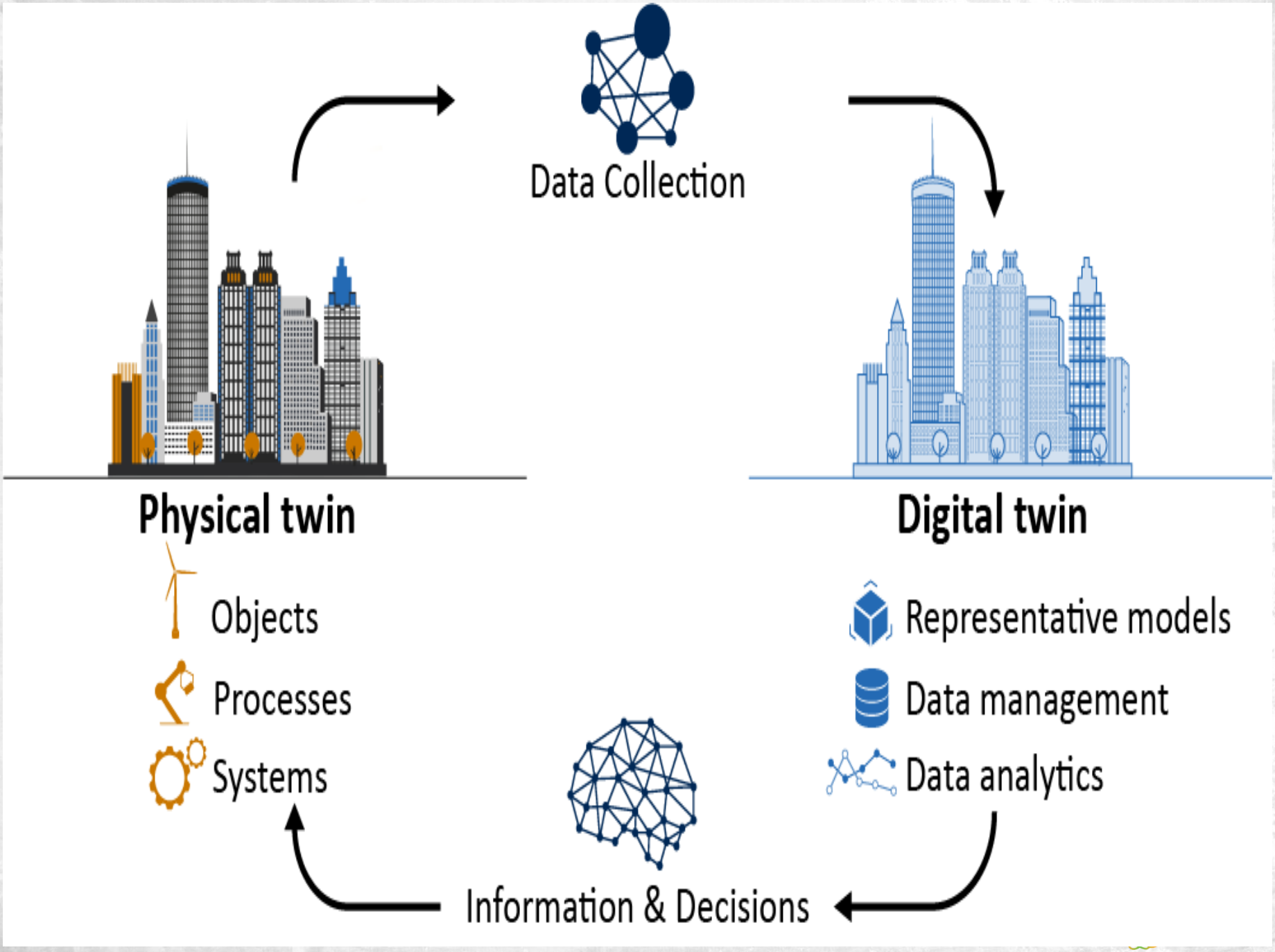
Object of the Case Study

“Spatial Digital Twins”:

They include a **specific spatial context** and **provide a holistic dimensional and location-based representation of assets, infrastructure and systems**. This refers to much more than the built-environment and exists at various levels of accuracy, detail and aggregation. **Spatial Digital Twins can cover buildings, clusters of buildings or other infrastructure, entire networks, cities, countries and even the globe”**

Spatial Digital Twins build **on the virtual representation of real-world entities and processes by using positioning and dimensions to uplift the value, insight and integrity of the virtual model which, in many instances, may be continuously updated at a synchronized frequency and fidelity.”**

InterPARES TrustAI


Data Collection
Physical twin
Digital twin
Objects
Processes
Systems
Representative models
Data management
Data analytics
Information & Decisions

Study Interrelated Frameworks

1. Critical Data Studies (Kitchin & Lauriault, 2018)
2. Digital diplomatics (InterPARES Project)
3. Social and Technological Data Assemblage (Lauriault 2022), and
4. Combination of technological Walkthrough (Light, Burgess & Duguay 2018) with Digital Record Forensics (Duranti 2009).

Carlton Immersive Data Studio

CIMS, in the Azrieli School of Architecture and Urbanism at Carleton University, is the Principal Investigator building a digital twin, funded by SSHRC.

Carleton Immersive Media Studio (CIMS)
Imagining Canada's Digital Twin (ICDT)
Carleton University Facilities and Plant Management (CU FMP)
Delta Controls

- Carleton University Building Performance Research Centre (CU BPRC)
- Carleton University Corporate Records and Archives (CU A)

City of Ottawa

- Building Code Services
- Geospatial Analytics, Technology and Solutions (GATS)
- **City Archives**

https://brookmcilroy.com/projects/carleton-university-campus-master-plan-update/

See Video - https://www.youtube.com/watch?v=_yFCl0MHUcQ&ab_channel=EsriCanada

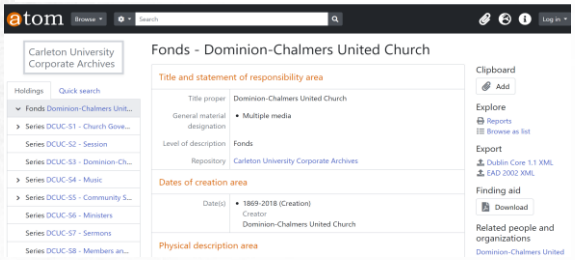
InterPARES TrustAI

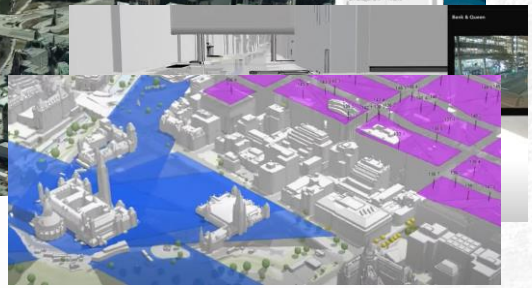
Carleton Immersive Media Studio

Delta CONTROLS
A Delta Group Company

Facilities Management and Planning

Corporate Records and Archives





Building Performance Research Centre

InterPARES TrustAI

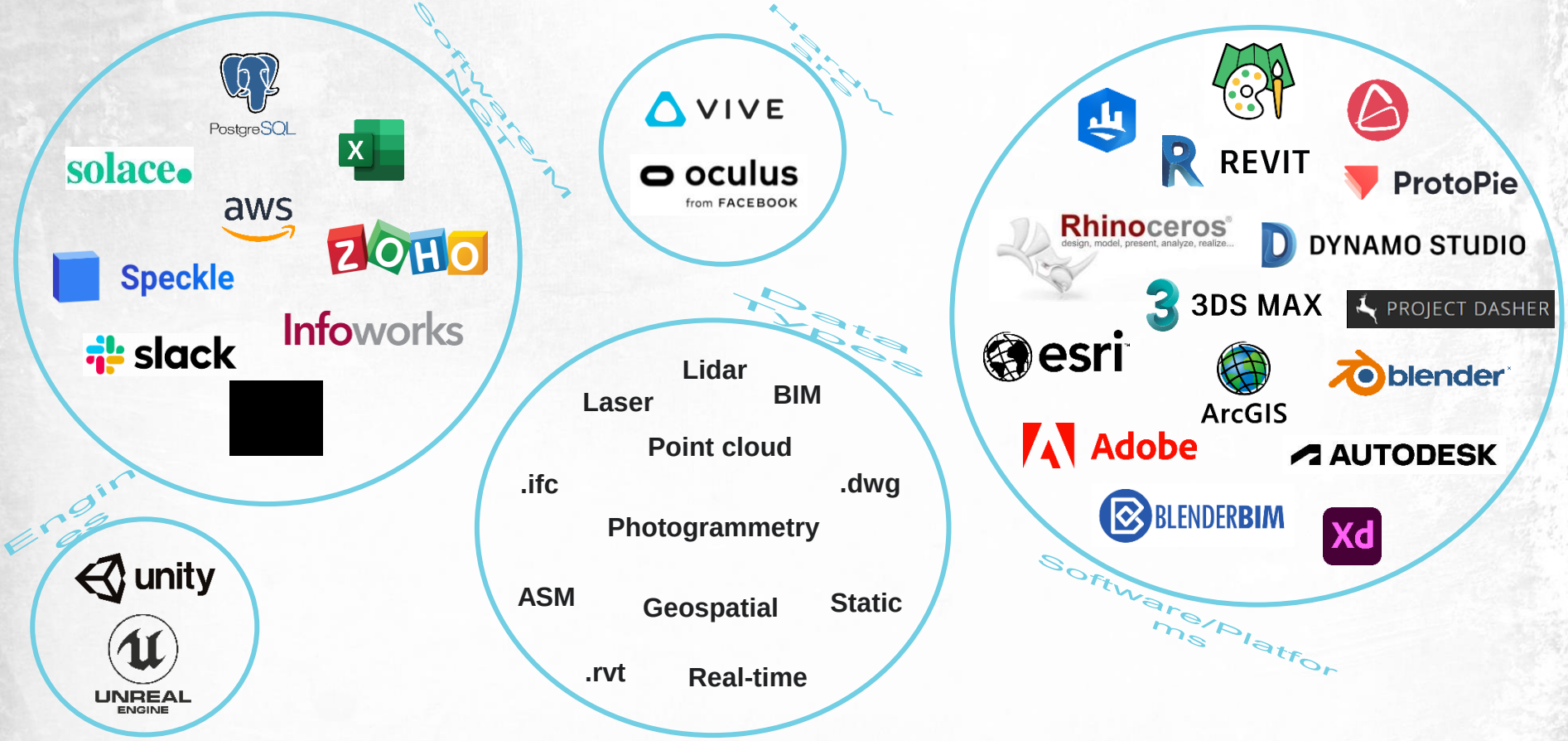
Findings

- The team identified **multiple platforms, hardware systems, software management systems, asset management databases, applications, rendering engines, data conversion systems, data types and interfaces including hardware such as Virtual Reality (VR) headsets, screens, and cameras**.
- They also discovered **practices** such as Building Information Modeling, architectural design and drawing, photogrammetry, database creation, interoperability, and open source and storage. Most of these practices are **to create the digital replicas of physical assets**, in this case buildings, **and develop ways to render the data viewable to users** – by video, via a game engine on a browser, in a VR system, on a screen.

InterPARES TrustAI

Preliminary Observations – 1st round of interviews

Identified software platforms, software management, VR hardware, VR and 3D creation engines, and data types

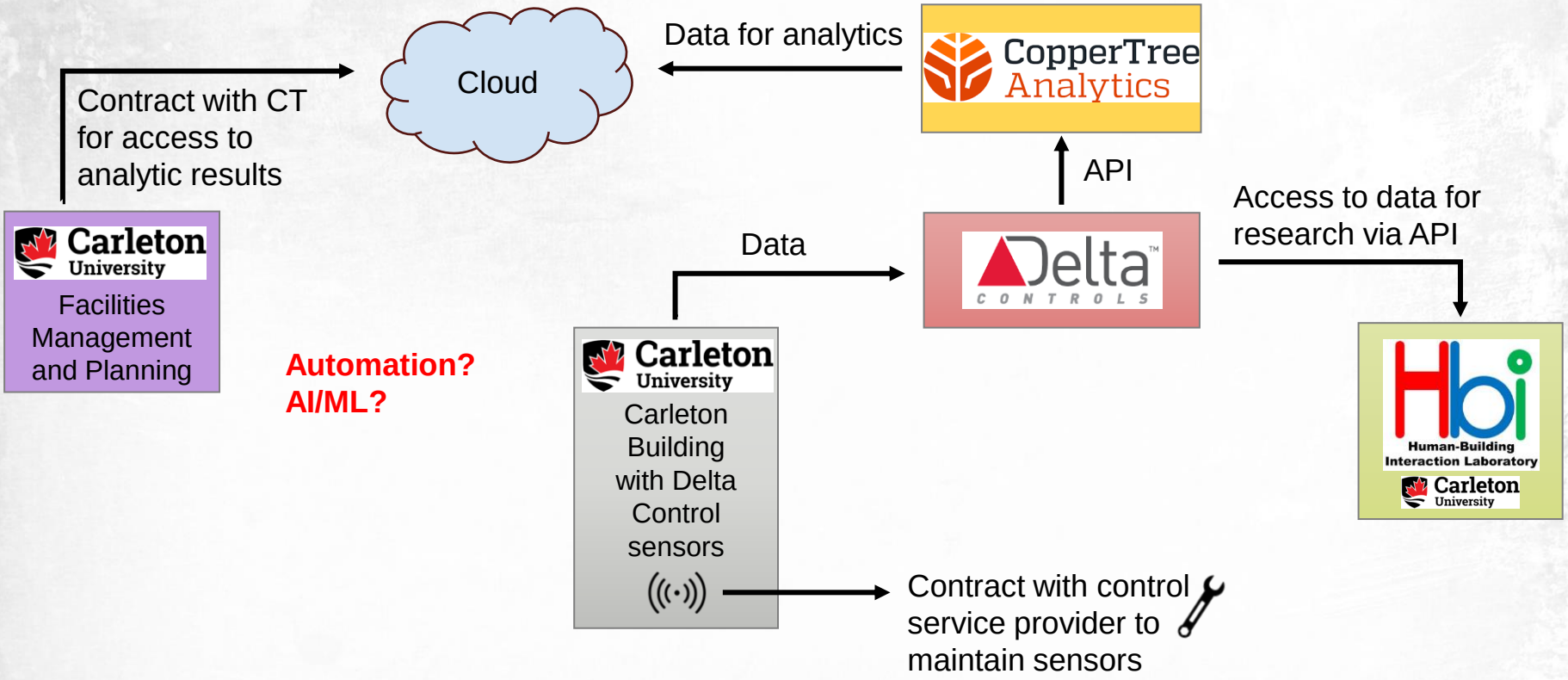


Findings

- After meeting with engineers and managers and toured facilities, plants, labs and control rooms, the team discovered that **the work for these actors is informed by operations data, the automation of data captured by sensing pre-programmed environmental parameters to operate the temperature and air circulation of environments** – thus, a building's circulatory system and not 3D rendering of the building's architectural features.
- The BLDG environmental view is the circulatory system of a building - such as **Heating Ventilation and Air Conditioning, electricity, lighting, heating/cooling as physical assets that are remotely operated by sensors and meters**.
- They also discovered **an assemblage of multiple sub-contractors**.
- https://deltacontrols.com/wp-content/uploads/cutsheets/enteliWEB%20Catalog%20Sheet.pdf
- https://deltacontrols.com/wp-content/uploads/enteliWEB_423_Catalog_Sheet.pdf

InterPARES TrustAI

Preliminary Observations
2nd round of interviews



Tracey P. Lauriault
Anna-Lena Theus

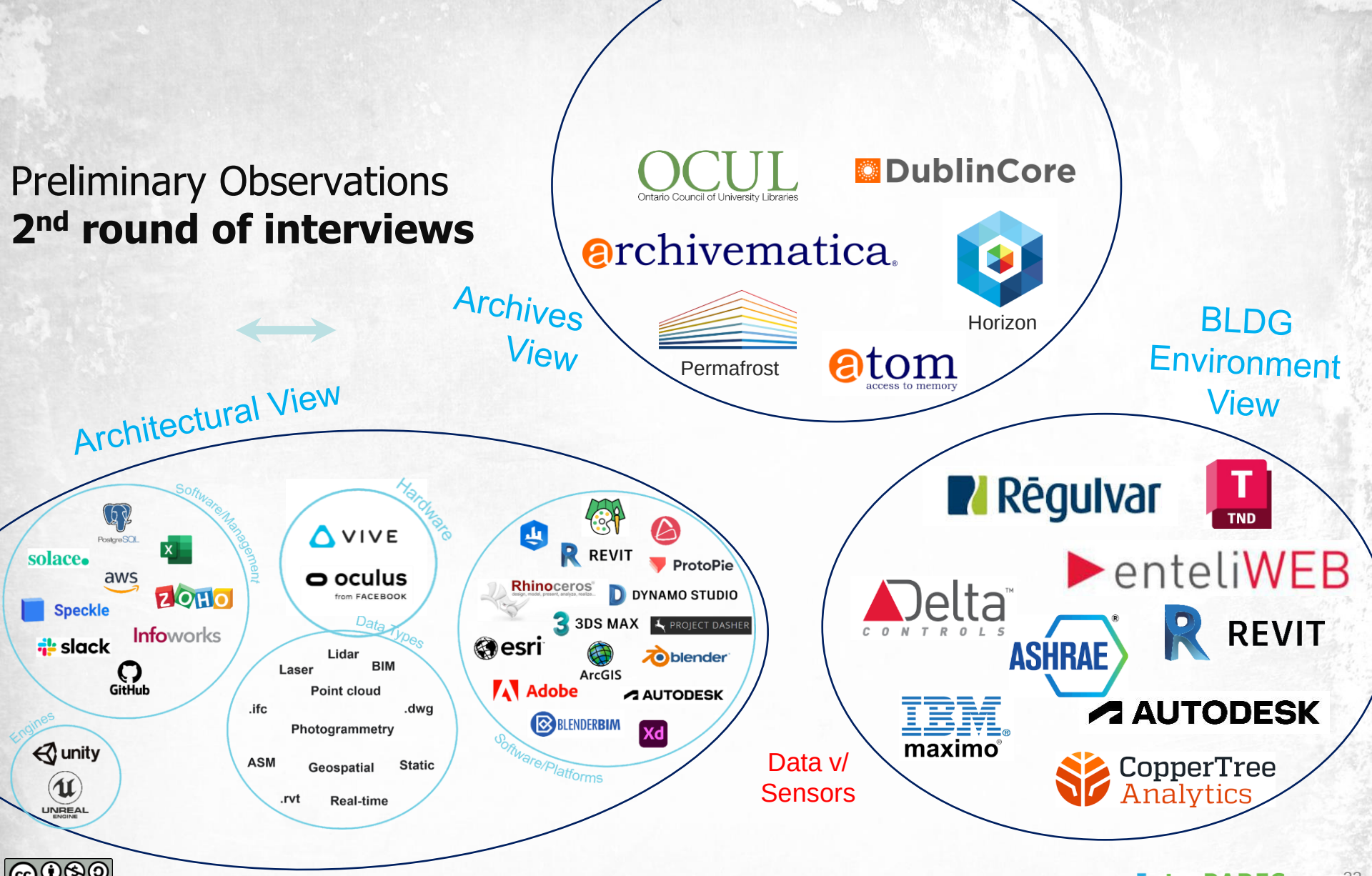
In more detail…

Delta Controls develops building automation systems according to building engineer schematics. A certified **Delta Controls** installer, **Regulvar,** installs Delta's sensors and systems using **Facility Management & Planning** (FMP) parameters and **Delta** code and connects them to an enterprise operating platform stored on Carleton's IT system. **enteliWEB,** an enterprise facility management platform has a database that includes each sensor/meter and its code and location, as well as a 3D image of where it is and what it operates, among other services and features.
Captured real time data are fed to **enteliWEB and** analyzed by **CopperTree Analytics** or signal activities to the plant and to actuators.
Alerts are received by Facility Management Planning via the **enteliWEB**, and staff can either remotely assess a problem and repair it, adjust the code remotely, or create a work order to send people to repair in situ.
FMP receives monthly reports from **enteliWeb** about operations.

There is another completely different assemblage of operations for the management of the gas heating plant, water facilities, and electricity.

InterPARES TrustAI

Preliminary Observations
2nd round of interviews



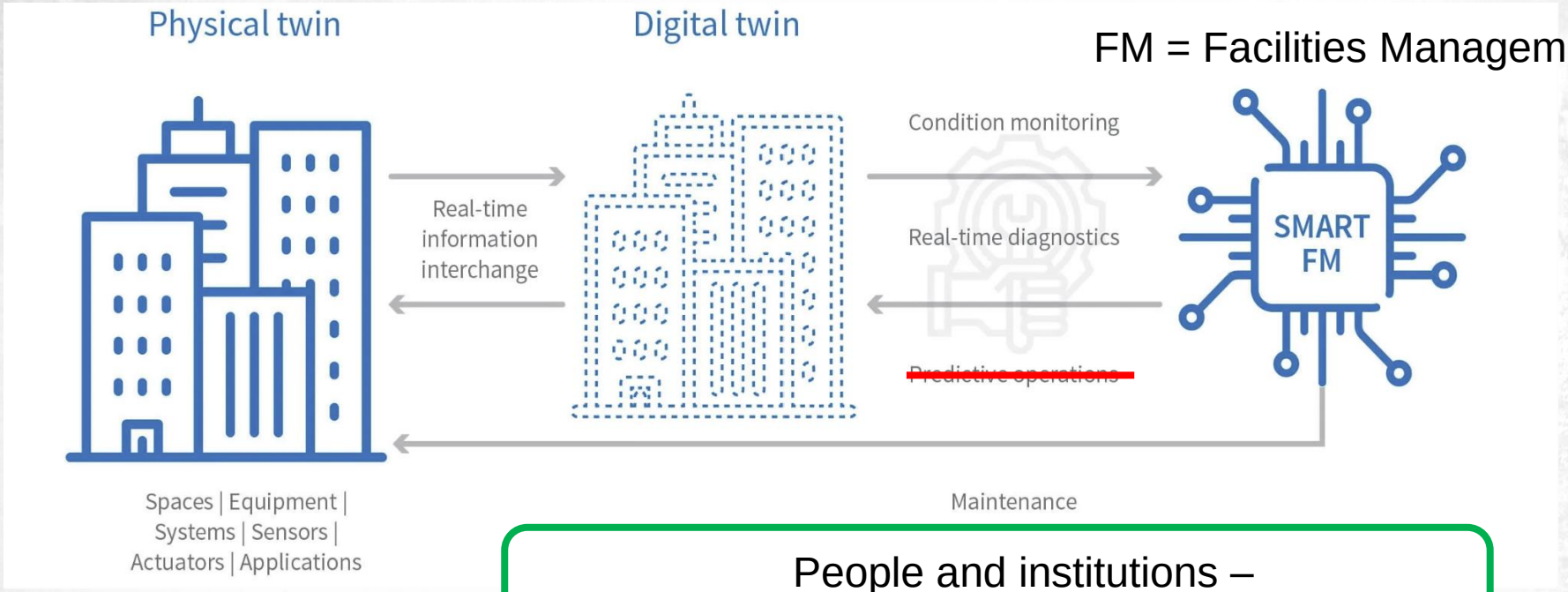
InterPARES TrustAI

Preliminary Observations: Records
2nd round of interviews

Types	Archives	FMP	CIMS	Research	City
Sensors	?	✓	✓	✓	-
Buildings	?	✓	✓	✓	?
Data (NRL, Sensors, Building)	?	?	✓	✓	?
BIMs	✓	✓	✓	-	✓
Seals / Certification	✓	✓	-	-	✓
Contracts	✓	✓	-	✓	✓
Permits	✓	✓	-	-	✓
Agreements / MOUs	✓	✓	✓	✓	?
Policies, Directives	✓	-	-	-	?
Blueprints	✓	✓	✓	-	
Standards	✓	✓	✓	✓	?
Software	✓	✓	✓	✓	?
Cloud	?	?	?	✓	?
Storage / Backup	✓	✓	✓	✓	?

Automation
AI

Is it a FM a Digital Twin?



People and institutions –
Juridical actors

Reflections on the Study

- The object of study includes **many autonomous systems** that are connected but not integrated
- There is **no governance structure** for the whole
- Many interactive **records are not static, but dynamic**, thus not preservable as they are
- **Organizations have responsibilities that are not coordinated**
- Those who operate systems in the Carlton study are, each of them, juridical actors that **have authority only on their own system** but not those of others
- **Should the Carlton Archives be the coordinator of the systems outputs?**
- The archives that has jurisdiction on the records of the university should, though buildings are on city land and so far the City of Ottawa archives has been following what goes on.
- The question is: **who is accountable for those records preservation**?

InterPARES TrustAI

I Trust AI Paradata Study

A General Study is one that concerns all archival functions. Purpose of the Paradata study:
Developing an approach for **documenting the AI process** to **fulfill archival accountability in the use and preservation of AI**, thereby also **supporting the authenticity of the outcome**.

Researchers: Pat Franks, Babak Hamidzadeh, Scott Cameron, Norman Mooradian, Alex Richmond, Mario Beauchamp

*The slides that follow are extracted from several of their presentations.

InterPARES TrustAI

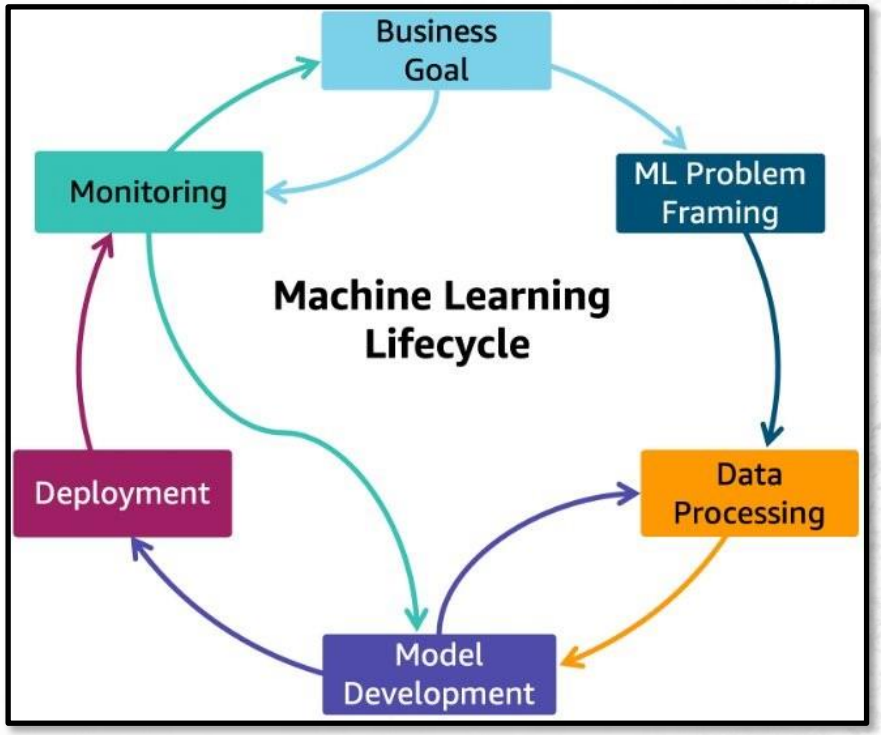
XAI vs. Accountable AI

- Explainable AI (XAI) has received a lot of attention. XAI focuses on **why a given tool produced a given output from a given set of inputs**.
- But building accountable AI **must also consider the individuals, organizations, and environment in which the AI tool operates**.
- **Paradata is necessary to explain why, how, by whom, and to what effect a given tool was used and kept in a particular context.**

InterPARES TrustAI

PARADATA & AI Process

Paradata is **information about the procedure(s) and tools** used to create and process information resources, along with **information** about the operation of the tools and the execution of the procedures, and **about the persons** carrying out those procedures and using the tools.



~ITrustAI working definition

InterPARES TrustAI

Metadata vs Paradata

Metadata is formalized data about a record needed to search for, display, and analyze that record

Paradata is formalized data on methodologies, processes, and persons associated with the production and assembly of records.

InterPARES TrustAI

Examples of Paradata

Technical Paradata

- AI Model (tested & selected)
- Evaluation & performance metrics
- Logs generated
- Model training data set
- Training parameters for model
- Vendor documentation
- Versioning information

Organizational Paradata

- AI policy
- Design plans
- Employee training
- Ethical consideration
- Impact assessments
- Implementing process
- Regulatory requirements

InterPARES TrustAI

Paradata Under Shared Agency

- Documentation and recording should be a mix of **continuous**, **sampled** & **event-based**
- Rely on **predefined trigger points**, control switches and reasons thereof
- Follow **feedback cycle model** (sensor, controller, actuator) to document each of these phases
- **Association between what is sensed, how it is acted on and control logic used must be documented.**
- **Temporal dimension** has to be captured and documented.

InterPARES TrustAI

General Paradata

What is documented	Types of documents/records
System by itself (independently of its specific uses and operation)	Preservation of the system itself and its versions; HW/SW architecture and design diagrams; Code, model, algorithms, logic and executables; Maintenance and upgrade documentation; Training data, test data and results, validation data and results; Means for running the system
Governance and compliance information	Organizational records documenting self-auditing processes, acceptance tests, change control; Sales history of models and configurations

InterPARES TrustAI

Operational Paradata

What is documented	Types of documents/records
Sensor input	Log of sensor data; Camera footage used for computer vision systems
Controller	Log of control directions; Relevant settings of control system; Intermediary process data leading up to a decision; Post-facto AI explanations of these processes; Log of warning notifications and control handover notifications
Actuators	Log of human control actions; Log of automated system's control actions; Log of messages communicated from system to human controller and external parties
Effects	Log of sensor data; Camera footage

InterPARES TrustAI

New Information Requirements

- Preservation of **evidence of intermediary processes**.
- Crucial information is not so much within the data point produced by the intermediary processes or the system's outputs, as it is about the ***relation between data points***.
- **Maintain interdependency between intermediary processes and the decisions formulated by the controllers.**
- Records must compile datasets in such a way as to **make the relationships between intermediary processes sensible & understandable.**

InterPARES TrustAI

Reflections

- Traditionally, a computer system is understood as a **tool** rather than an agent.
- AI systems challenge this by **increasing their exerted autonomy and introducing barriers to human comprehensibility**.
- If AI systems are recognized in a juridical context as **agents** rather than tools, record keepers will need to adjust their focuses accordingly.
- Under existing frameworks, AI systems do not constitute the author responsible for the actions documented. Without legislation recognizing the status of AI systems as juridical persons, AI systems would not constitute authors responsible for their own actions.
- Proposals exist for the implementation of **legal personhood** for AI systems in the European Union, making them at least partially responsible for the actions in which they participate. **AI systems would no longer be mere agent of human intentions and actions, but may comprise entities exerting agency, even if they may not be responsible as human agents are**. But then…
- **Whose accountability for their actions? Whose liability?**

InterPARES TrustAI

Conclusion

- **Decisions made and actions taken by AI-enabled systems must be documented.**
- Some of the documentation will be **automatic** as part of the AI system; some will be **human-created prior to or after the creation and implementation of the AI system**.
- **Paradata is recommended to document the AI process and promote the archives transparency and accountability.**
- The records perspective supports the capture and preservation of paradata and is necessary to ensure the **AI process is captured in a way that preserves the characteristics of authoritative records: reliability, accuracy, authenticity (identity and integrity), and usability**.
- Guidance in the form of laws, regulations, and frameworks must be monitored.

InterPARES TrustAI

Stay tuned!
Thank you!

www.interpares.org

www.interparestrustai.org

@itrustai

www.facebook.com/interparestrust

InterPARES TrustAI