From Traditional Archival Knowledge to Computer Vision and Natural Language Processing The More it Changes…

Luciana Duranti, Professor Emerita
The University of British Columbia
Principal Investigator, InterPARES

Oslo, Norway. Norwegian Archival Association. 1 April 2025

Traditional Records Definition

In Archival Science, a Record (or Archival Document) is

"A document made or received by a physical (natural) or juridical (a collection or succession of natural persons) person in the course of a practical activity"

- The above definition, by itself, is not very helpful in the contemporary digital environment, especially when confronted with AI
- Whereas archival science focuses on aggregations of documents, we need to use the diplomatics focus on individual documents, and specifically on their form, in order to understand the facts or acts represented in them
- **Documentary form: the rules of representation which formalize the characteristics of a document that can be distinguished from the subject, the persons or the places it is about.**
- Form is **physical** and **intellectual**.

InterPARES Trust AI

Records Functions
(the way a record relates to an action)

- **Dispositive**: the record is the action (e.g., contracts)
- **Probative**: the record proves the action (e.g., marriage certificate)
- **Supporting**: the record generated to be used in the course of activity (ies) as a source of information, often by multiple users (e.g., GIS, lecture notes)
- **Narrative**: generated on a purely discretionary basis only as a means of communication (e.g., most e-mails, memos, some web sites)
- **Instructive**: the record provides guidance on the way in which actions external to the record are to be presented (e.g., music scores, scripts, regulations, manuals of procedure, instructions for filling out forms).

InterPARES TrustAI

In the Digital Environment

- **Same as in the analogue environment, plus:**
- **Enabling**: support the performance of artworks (software patches), the execution of business transactions (interacting business applications), the conduct of experiments (a workflow generated and used to carry out an experiment of which it is instrument, byproduct and residue), the analysis of observational data (interpreting software), etc.
- While we see the other types of records on a computer screen, enabling records only exists as stored encoding.

InterPARES TrustAI

Digital Record Necessary Components

- **Act**: an action in which the record participates or which the record supports
- **Persons Concurring to Its Creation**: author, writer, **originator**, addressee, and **creator**
- **Archival Bond**: explicit linkages to other records inside or outside the system
- **Identifiable Contexts**: **juridical-administrative**, **provenancial**, **procedural**, **documentary**, **technological**
  - **Medium**: necessary part of the technological context, not of the record
- **Fixed Form and Stable Content**

InterPARES TrustAI

Stored and Manifested Records

- **Stored record**: it is constituted of **the digital component(s) used in re-producing it**, which comprise
  - 1. the data to be processed in order to manifest the record (*content data* and *form data*) and
  - 2. the rules for processing the data, including those enabling variations (*composition data*)

  Sometimes a stored record does not have a corresponding manifested record (e.g. enabling records)
- **Manifested record**: **the visualization of the record in a form suitable for presentation to a person or a system**. Sometimes, it does not have a corresponding stored record, but it is re-created from fixed content data when a user's action associates them with specific form data and composition data (e.g. a record produced from a relational database)

InterPARES TrustAI

Fixed Form

- An entity has fixed form if its binary content is stored so that the message it conveys can be rendered with the **same documentary presentation** it had on the screen when first saved (different digital presentation: Word to .pdf)
- An entity has fixed form also if the same content can be presented on the screen in several different ways in a limited series of possibilities: we have **a different documentary presentation of the same stored record** having stable content and fixed form (e.g. statistical data viewed as a pie chart, a bar chart, or a table)

InterPARES TrustAI

Stable Content

- An entity has stable content if **the data and the message it conveys** are unchanged and unchangeable, meaning that data cannot be overwritten, altered, deleted or added to
- **Bounded Variability**: when changes to the documentary presentation of a determined stable content are limited and controlled by fixed rules, so that the same query, algorithm, or interaction **always generates the same result**, or so that we have different views of different subsets of content, due to the intention of the author or to different operating systems or applications

InterPARES TrustAI

Types of Digital Records

Static: They do not provide possibilities for changing their manifest content or form beyond opening, closing and navigating: e-mail, reports, sound recordings, motion video, snapshots of web pages

Interactive: They present variable content, form, or both, and the rules governing the content and form of presentation may be either fixed or variable

InterPARES TrustAI

Interactive Entities

- **Non-dynamic**: the rules governing the presentation of content and form do not vary, and the content presented each time is selected from a fixed store of data. Ex. Interactive web pages, online catalogs or inventories, records enabling performances—they are records
- **Dynamic**: the rules governing the presentation of content and form may vary—they are either information systems or potential records

Interactive Potential Records Systems

- Systems where the **variation is due to data that change frequently**, because the design permits updating, replacement or alterations (e.g. students register); or that **allow data collection** from users or about user interactions or actions (e.g. faculty self-service portal); or that **use the data input by users** to determine subsequent presentations (e.g. land registry)
- Systems where the variation is due to **data received from multiple external sources** at different times and **not stored within the system** (e.g. GIS)

They are presently not records systems but can and should be made into records systems if the sets of data they produce and hold fulfill one of the records functions (will return on this)

InterPARES TrustAI

Some Key Questions

- If it is not possible to have records in fluid form and with undetermined boundaries (i.e. an interactive dynamic record), should an entity with fixed form and stable content be generated for the purpose of making a record to be kept in a trusted recordkeeping system and perhaps preserved over the long term?
- If yes, **who** should make it?
- On the basis of which **criteria**?
- **When** in the entity's lifecycle?

InterPARES TrustAI

…and, if it were not possible to stabilize content and fix form…

- Could we trade stability of content and fixity of form with the ability to track changes?
  Record=latest manifestation + log of changes + metadata
- Could we think of the record as existing in two modes: a) in becoming, when the object is accessed to add information to it; and b) in being, when the object is accessed for use?
  Record=each manifestation accessed for use + metadata

InterPARES TrustAI

Case Study #1: the Alsace-Moselle Land Registry

The registry is required by the French real estate law, as the means to fulfill the requirement that the legal status of property (including the various forms of mortgages on the property) must be made publicly available to interested third parties by means of inscription within a land registry.

The Procedure of Inscription

- An electronic **request for inscription** is generated by a notary using custom software, which connects to the land registry in order to retrieve the information related to the parties or parcels of land
- Once the request is received at the land registry office, **it is dated**. This date determines the inception of the rights on the property.
- For each request, **a digital file is created containing all of the associated documents** (contract, cadastre, etc.), as scanned image files, unless they exist as digital data sets to which the request can be linked

InterPARES TrustAI

The Procedure of Inscription (cont.)

4. A **draft order of inscription** is prepared. Inscriptions are also drafted directly in the database, but are not visible to outside users of the database until a **judge** has signed them; **the draft order is transferred to the judge's "inbox" in the form of an XML document**
5. The judge is responsible for the required verifications: the custom software of the land registry office provides him/her with a "before" and "after" view of the inscription, that is, it shows the changes to the registry which the inscription will effect in the database;
6. After identifying himself through biometric (fingerprint) scan and inserting a smartcard with his private signature key, the judge signs the draft order. At that moment, **in a single step, the order is generated and signed, producing an inscription, and the relevant fields of the database are updated.**

InterPARES TrustAI

Digital Entities in the Registry

- **The order**, which becomes the inscription, listing the information relative to the land parcel, the parties to the transaction, and the nature of the transaction. It is delineated in fields, using XML tags, and may thus be readily processed. It is authored by the judge, who dates and signs it.
- **The tables of a relational database** (i.e., one table that records the characteristics of land **owners**, another of land **parcels**, another of the **charges**, another of the **mortgages**), with **links between the tables** that establish relationships between relevant data in the tables.
- The two most important **views** offered by the digital land registry are (a) the ownership history of a given land parcel and (b) the set of land parcels owned by a particular individual.

InterPARES TrustAI

Technological Structure

- An **Oracle database**, containing the land registry data;
- **Personal (Windows) computers**, for registry clerks, running web-based applications for consulting the registry and managing the inscription process;
- **Plugs-ins for commercial notarial software** for integration with the land registry;
- **Personal computers**, for land registry judges, running web-based applications for consulting the registry and for finalizing inscriptions to the registry and equipped with **biometric identification peripherals**, and **digital signature software**;
- A **PKI infrastructure**,* linking together all land registry offices and the central database, so that judges may sign orders and add inscriptions to the registry.

InterPARES TrustAI

* A Public Key Infrastructure is the technology that allows you to encrypt data, digitally sign documents, and authenticate yourself using certificates.

What Is Distinctive about the System?

- The system uses digital signatures to provide **continuous authentication services**, that is, regularly performed declarations of the integrity and identity of the data.
- Digital signatures provide an **extreme assessment of the integrity of data**: if even a single bit of the signed data is modified, the signature fails.
- They also **compare the orders with the inscriptions** every time their authenticity is questioned.

InterPARES TrustAI

Issues

- The District Archives **must receive registration records** 5 years after their date
- While the acquisition of the **orders** by the District Archives, as stand-alone documents, poses no particular problems, that of the **inscriptions** does
- **The digital inscriptions are not records, the land registry as a whole is.**
- As a record, **the land registry cannot be understood outside of its dynamic and interactive capabilities.**
- **The inscriptions cannot be authenticated outside the Public Key Infrastructure**
- Migration to overcome obsolescence risks loss of interoperability

InterPARES TrustAI

Proposed Solution

- The definition of an XML schema which may serve as a translation device between the complex data model used by the land registry, and a less complex model, to be defined, sufficient to satisfy the needs of future users.

- Inscriptions could then be exported to a file according to the XML schema and imported into a relational database sufficiently simple to be maintained by the designated preserver.

InterPARES TrustAI

Case Study #2: the VanMap x

- The cross-corporate GIS created by the City of Vancouver https://maps.vancouver.ca/portal/apps/sites/#/vanmap/.
- Used by city staff in

  - Engineering
  - Planning
  - Permits and Licenses
  - By-law Enforcement
  - Social Planning
  - Police
  - Fire and Rescue
  - Parks and Recreation

VanMap Technical Components

- **Oracle Spatial database**
- **Other databases linked to it,** existing in a variety of offices, of a variety of local authorities, and whose data flow continuously in the Oracle database
- **CAD drawings, satellite imagery, photographs, html pages**
- **Autodesk MapGuide**
- **Autodesk ActiveX Viewer**
- **Application servers**
- **Web server**

http://vanmappub.vancouver.ca - VanMap Public ActiveX Viewer version for Windows - Microsoft Internet Explorer provided by ICBC

VanMap
public edition

Contact Us

Address Search
Options | Reset | Help
Number Street
or select E 1st Av Go

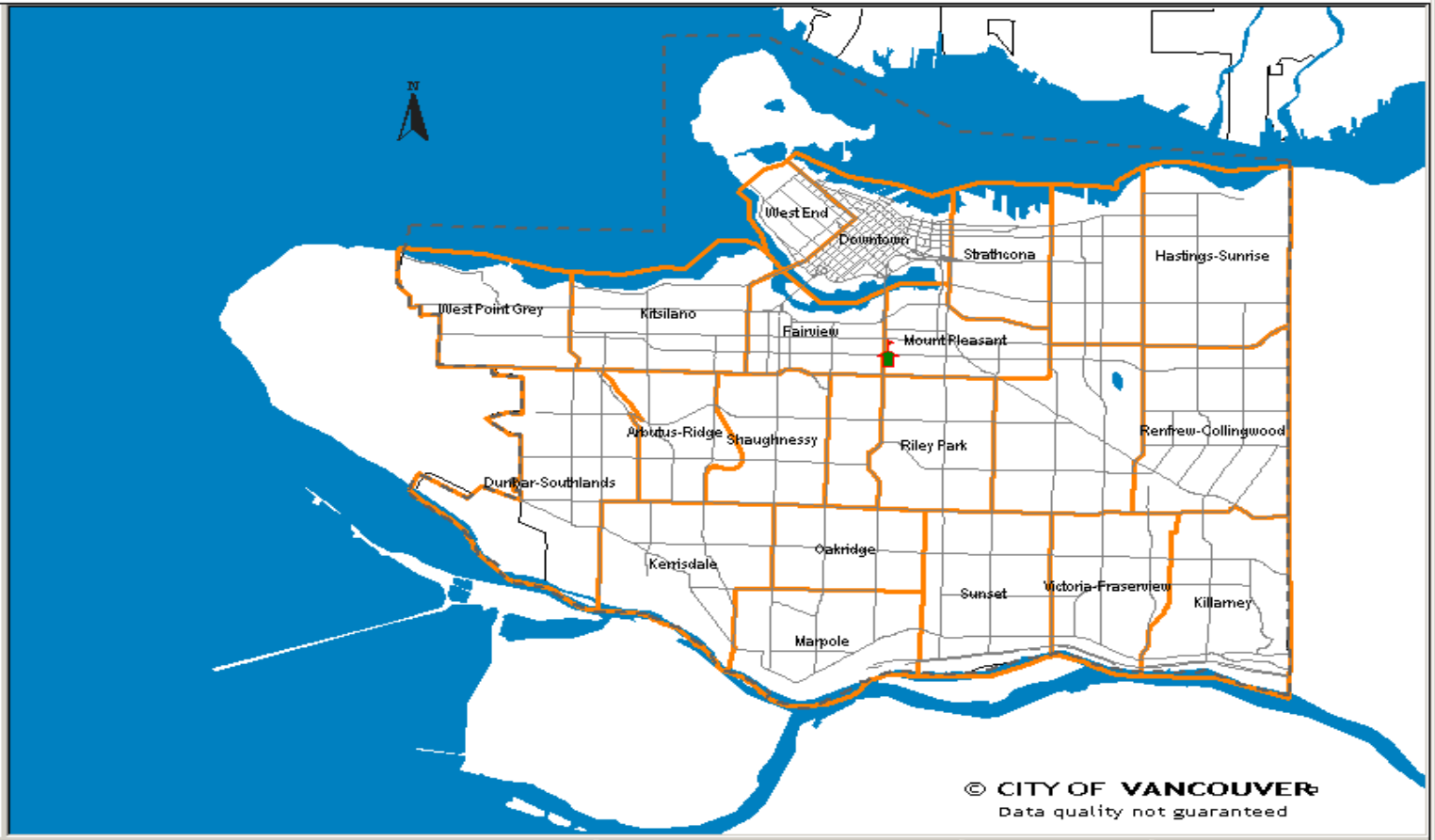
Toolbox
Right-Click Menu | Help

Help
About the Data

Applications
Select Application:

Legend Options

- Olympic Venues
- City Boundary
- Public Art
- Public Places
- The Road Ahead
- City Projects
- Child Care Facilities
- Web Cameras
- Non-Market Housing
- City-owned Properties
- Cemetery
- Urban Agriculture
- Contour Lines
- Truck Routes
- Canada Line Alignment
- Traffic Related
- Sewer
- Water
- City Streets Network
- View Cones
- Subdivision Categories
- DCL Areas
- Zoning Districts Types
- Zoning Districts
- Business Improvement Areas
- False Creek Navigable Chan
- Youth
- Property Information
- Facet Grids
- Administrative & Service Area
- Shore Lines (2002)
- Foreshore Contour Line
- Water Bodies



Water Bodies : Strait of Georgia | 0 feature(s) selected | 1 : 109,043 | 14.4 x 10.0 (mi)

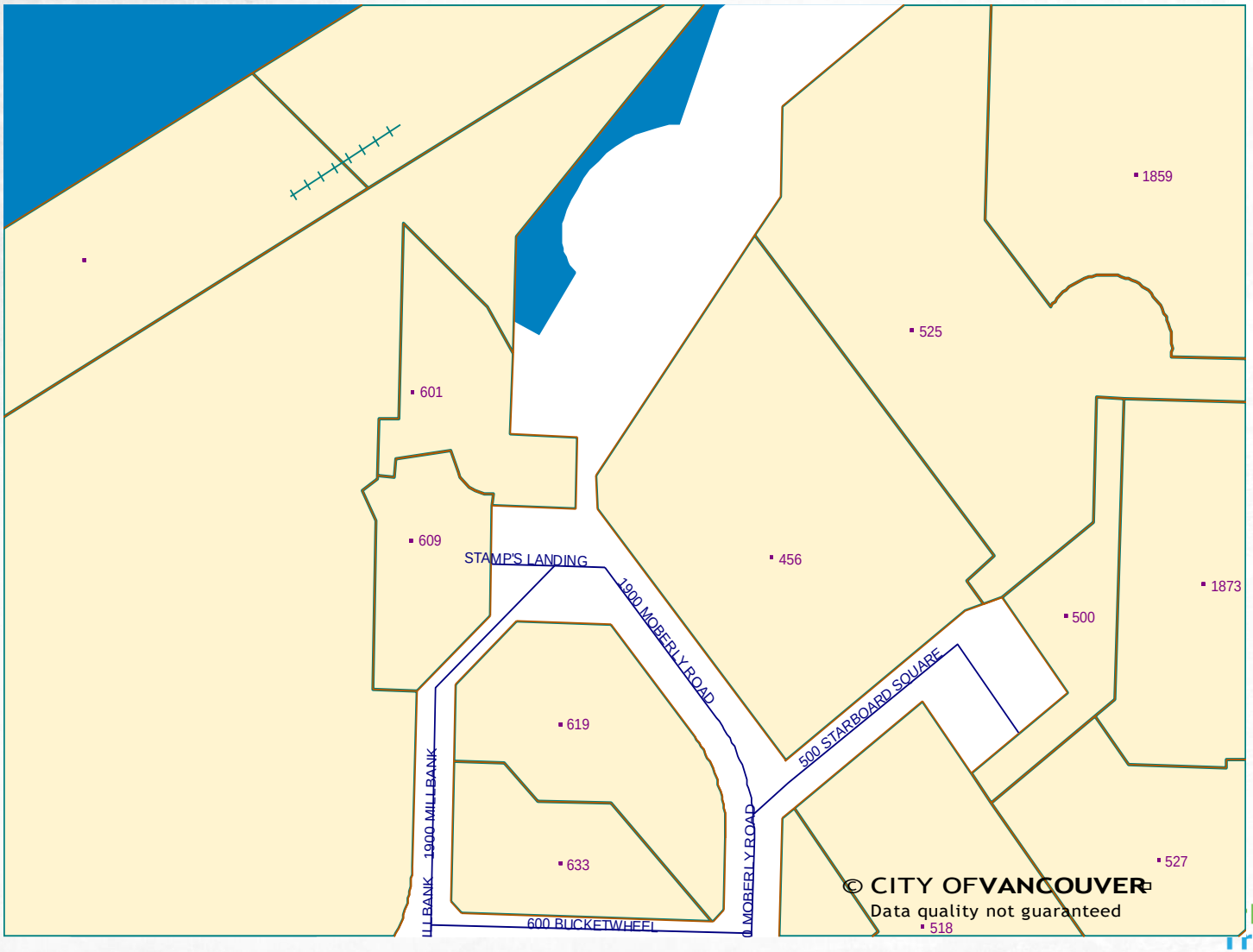
Internet

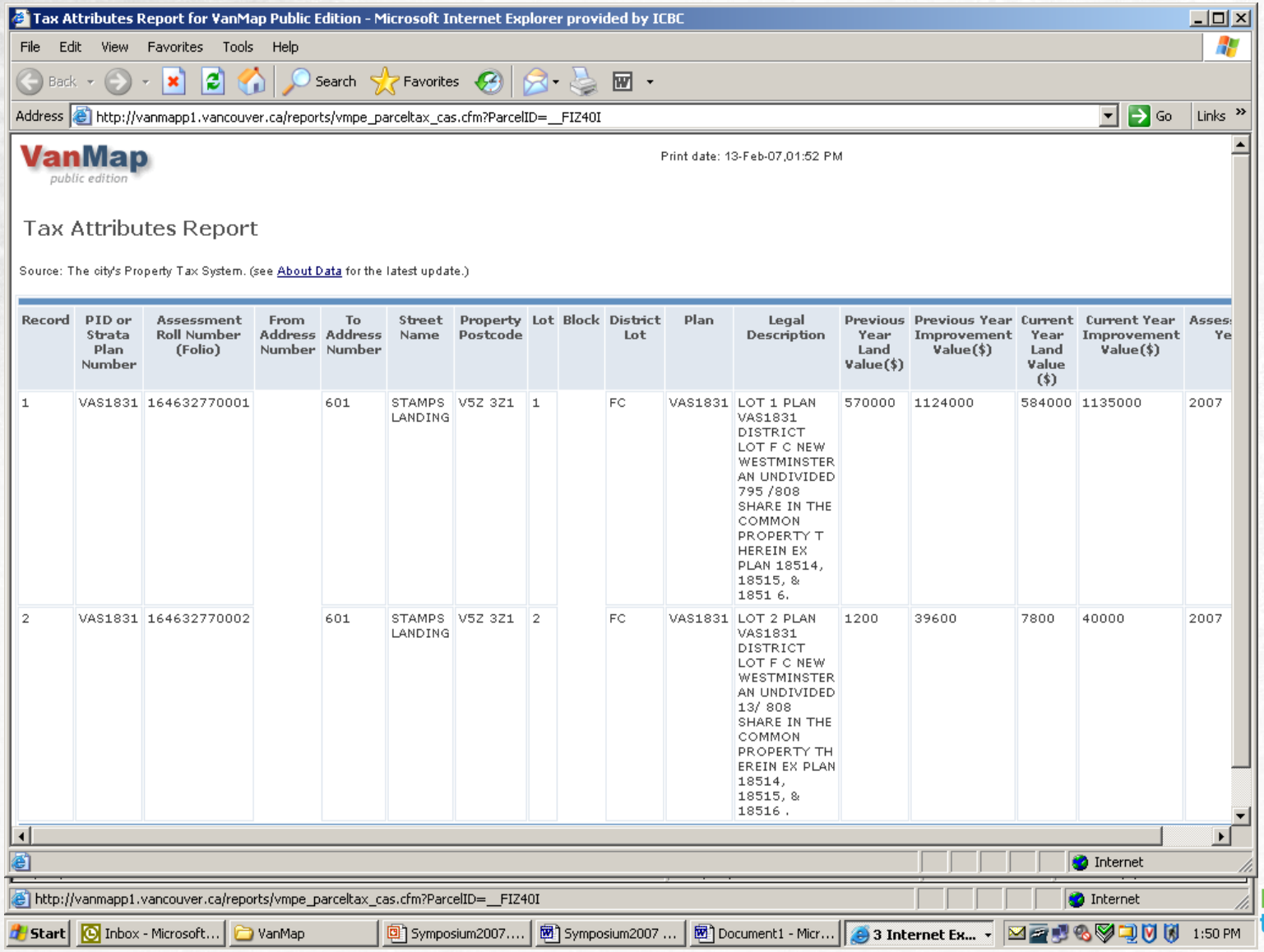
Start | Inbox - Micro... | 2 Windows ... | Microsoft Excel | Symposium 2... | Presentation1 | ACA presenta... | 2 Internet ... | 12:42 PM

InterPARES Trust AI


100 ATHLETES
100 WALTER HARI
100 W 1ST A
100 W 2ND
100 W 3RD
100 W 4TH
1700 COOK ST
1800 SPYGLASS PLACE
CAMBIE BRIDGE
CAMBIE BRIDGE
CAMBIE BRIDGE
200 W 1ST AV
300 W 1ST AV
200 W 2ND AV
300 W 4TH AV
400 W 2ND AV
1900 WYLIE ST
400 W 5TH AV
200 W 5TH AV
400 W 6TH AV
500 W 6TH AV
600 W 6TH AV
700 W 6TH AV
800 W 6TH AV
900 W 6TH AV
700 W 7TH AV
600 W 7TH AV
500 W 7TH AV
700 W 8TH AV
600 W 8TH AV
500 W 8TH AV
2100 CAMBIE ST
2200 CAMBIE ST
2200 CAMBIE ST
2200 ASH ST
2200 WILLOW ST
2200 LAUREL ST
2200 ALBERTA ST
2200 ALBERTA ST
2200 COLUMBIA ST
2300 YUKON ST
700 MILLBANK
700 MILLBANK
1900 MILLBANK
1900 MOBERLY ROAD
200 MILLYARD
1900 GREENCHAIN
SCHOOL GREEN
CHARLESON
800 CHARLESON
LAKE ST
© CITY OF VANCOUVER
Data quality not guaranteed


1859
525
601
609
456
1873
500
619
633
527
518
© CITY OF VANCOUVER
Data quality not guaranteed


1859
525
601
609
STAMP'S LANDING
456
1900 MOBERLY ROAD
1873
500
500 STARBOARD SQUARE
619
1900 MILL BANK
633
600 BUCKETWHEEL
0 MOBERLY ROAD
527
518
© CITY OF VANCOUVER
Data quality not guaranteed


Tax Attributes Report for VanMap Public Edition - Microsoft Internet Explorer provided by ICBC
File Edit View Favorites Tools Help
Back Search Favorites
Address http://vanmapp1.vancouver.ca/reports/vmpe_parceltax_cas.cfm?ParcelID=__FIZ40I
Go Links
VanMap
public edition
Print date: 13-Feb-07,01:52 PM
Tax Attributes Report
Source: The city's Property Tax System. (see About Data for the latest update.)
Record | PID or Strata Plan Number | Assessment Roll Number (Folio) | From Address Number | To Address Number | Street Name | Property Postcode | Lot | Block | District Lot | Plan | Legal Description | Previous Year Land Value($) | Previous Year Improvement Value($) | Current Year Land Value ($) | Current Year Improvement Value($) | Asses Ye
1 | VAS1831 | 164632770001 | | 601 | STAMPS LANDING | V5Z 3Z1 | 1 | | FC | VAS1831 | LOT 1 PLAN VAS1831 DISTRICT LOT F C NEW WESTMINSTER AN UNDIVIDED 795 /808 SHARE IN THE COMMON PROPERTY T HEREIN EX PLAN 18514, 18515, & 1851 6. | 570000 | 1124000 | 584000 | 1135000 | 2007
2 | VAS1831 | 164632770002 | | 601 | STAMPS LANDING | V5Z 3Z1 | 2 | | FC | VAS1831 | LOT 2 PLAN VAS1831 DISTRICT LOT F C NEW WESTMINSTER AN UNDIVIDED 13/ 808 SHARE IN THE COMMON PROPERTY TH EREIN EX PLAN 18514, 18515, & 18516 . | 1200 | 39600 | 7800 | 40000 | 2007
Internet
http://vanmapp1.vancouver.ca/reports/vmpe_parceltax_cas.cfm?ParcelID=__FIZ40I
Internet
Start Inbox - Microsoft... VanMap Symposium2007.... Symposium2007 ... Document1 - Micr... 3 Internet Ex... 1:50 PM

VanMap is a Dynamic Information System

- Contains data that often do not exist anywhere else, especially in the correlated form showed on the GIS layers
- Data are overwritten without being saved
- The data are viewed as maps but these views are not saved
- New layers are being added all the time
- VanMap does not contain records

Is VanMap as a Whole a Record?

Yes!

- It is made and kept in the course of a practical activity
- It is an instrument and a by-product of that activity
- It is an indivisible unit affixed to a medium
- It has all the diplomatic characteristics of a record in terms of persons and contexts

InterPARES TrustAI

Again: Is VanMap a Record?

No!

- It lacks fixed form and stable content
- It is not set aside for action or reference

Thus, **it cannot be used to render an account of the decisions made or as a memorial**.

Can VanMap Become a Record?

- Yes, **if we introduce fixed form and stable content**
- We need to configure the system so that, as each layer is updated, the data are saved rather than overwritten
- Then we need to develop a means of reproducing VanMap as it was on any given date

InterPARES TrustAI

What About Taking Map Views (snapshots)?

- The preserving body cannot do so because it would become the creator of digital objects never used by this person as a creator in the course of business
- It is not feasible to require City staff to save the map views in connection with the decisions based on them
- **We have to preserve not what the staff member *saw* at a given point in time but what s/he *should have been able to see***
- A detailed documentation of the business process would support this preservation activity

How to Build a GIS Preservation Environment

- Step 1: save the empty layers
- Step 2: add metadata to the layers
- Step 3: store time-stamped data in a secure environment
- Step 4: create infrastructure independence
- Step 5: migrate to new/neutral technology platforms
- Step 6: reproduce the system

Using Data Grid Technology

- Manages data and their associated metadata
- Separates the data from dependence on original creating infrastructure
- Maintains audit trails of all operations performed on the data
- Manages access and retrieval
- Supports migration of data to new platforms

InterPARES TrustAI

What Is Preserved?

- The data themselves
- The ability to see the data available on a given day and time as the data are time stamped
- The ability to render the data as interactive maps

InterPARES TrustAI

The Same Solution for All GIS?

- **No! Solutions are specific**
- What is identified as **the record to be generated and maintained over time depends on the use of the data by the creator and the reason for having records** rather than fluid information.
- The research GIS of the **Archaeological Society of Arizona** requires preservation of its ability to make the users detect underground materials from the layers that show vegetation and stratifications of the soil. Therefore it needs preservation of the records that suggest that excavations should be carried out.
- The **Canadian Atlas of Antartica** requires preservation of the content of external users interactions with specific layers and therefore preservation of the users' records that have changed the system output.
- **Solutions are also dynamic. Can AI help?**
- **What happens when records are AI generated?**

InterPARES TrustAI

I Trust AI Project Goal

The overall goal of I Trust AI is to design, develop, and leverage Artificial Intelligence to support the ongoing availability and accessibility of trustworthy 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 and archives challenges;
- Determine the benefits and risks of using AI technologies on records and archives;
- Ensure that archival and diplomatics concepts and principles inform the development of responsible AI; and
- Validate outcomes from Objective 3 through case studies and demonstrations.

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 a variety of disciplines: besides archival scientists, 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

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

- Our short-term approach focuses on **identifying high impact problems and limitations in records and archives functions, and applying AI to improve the situation**.
- Our long-term approach focuses on **identifying the tools that records and archives specialists will need in the future to flexibly address their ever-changing needs.** This includes decision support and, once decisions are made, rapid implementation of AI-based solutions to those needs.

InterPARES TrustAI

Approach (cont.)

- 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

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

Case Study: Deep Learning for parchment documents

Purpose: development of a tool to identify the "identity attributes" of thousands of digitized parchments issued by city notaries in medieval times. Study led by *Emanuele Frontoni and his Vision Robotic for AI team*

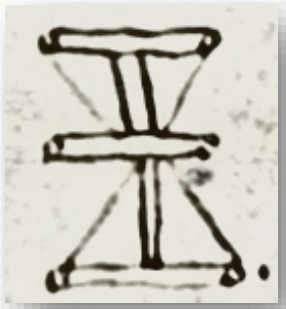
- It uses **computer vision**, a field of AI that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs, and take actions or make recommendations based on that information
- The basic feature chosen to be identified for the research is **the signum, an authentication element** affixed by notaries

InterPARES TrustAI

THE SIGNUM: A USEFUL MARKER

- The **signum or notarial sign** is a specific and personally drawn mark used by a single notary in the top part of the record and before his signature at the bottom of the record.
- Identifying the signum means that every notary can be **recognised and tracked** in a virtually infinite series of documents.
- The AI will contribute in creating both a **register of signa—a virtual registry of notaries,** and the basis for investigating the **less visible features of the parchments**



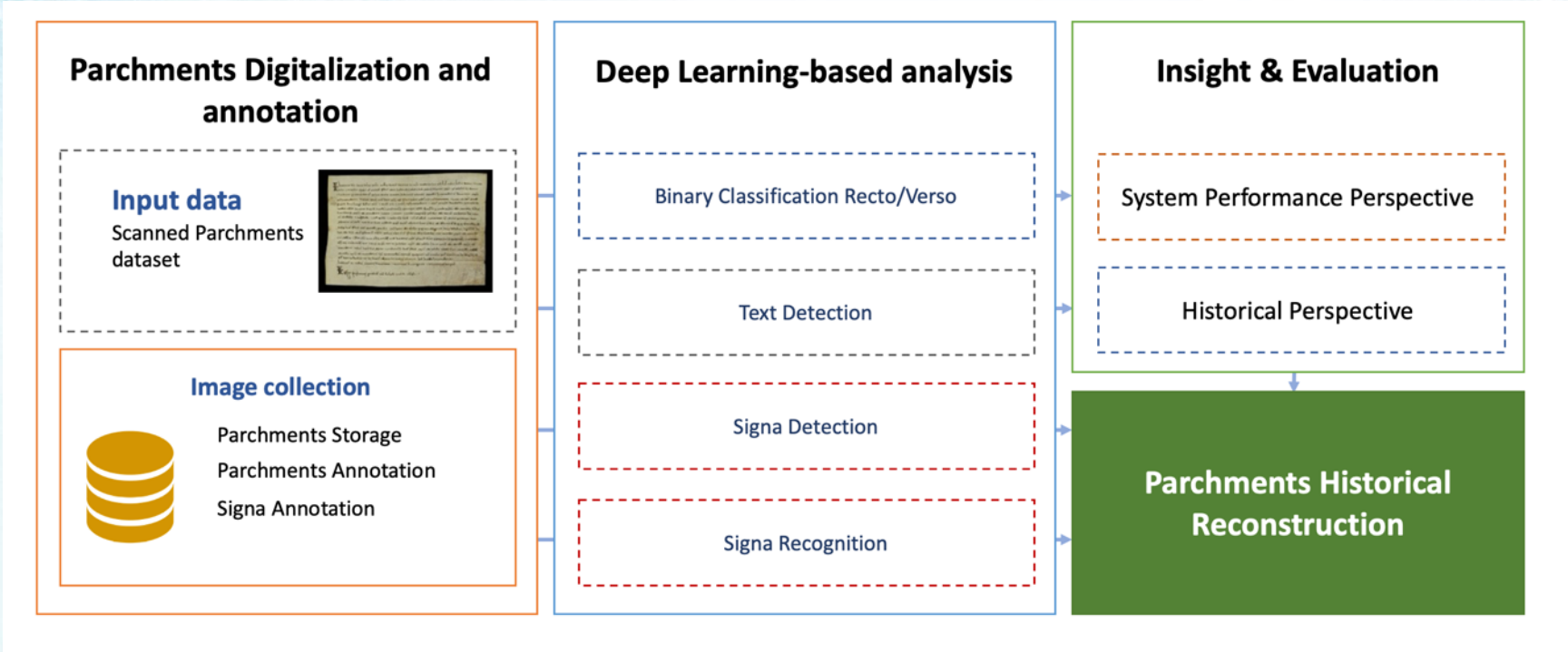



InterPARES TrustAI

PERGANET

A Deep Learning Framework for Automatic Appearance-Based Analysis

https://link.springer.com/chapter/10.1007/978-3-031-13324-4_25



InterPARES TrustAI

OTHER ARCHIVAL APPLICATIONS

Once refined and fully developed, this Deep Learning tool, Perganet, could be used in a **wide range of applications**:

- Recognize the **peculiar system of writing** of individual authors.
- Analyze **archival annotations on the back of the documents** and retrace previous archival arrangements or uses of groups of documents.
- Recognize **recurring images or other features** in huge series of documents.
- Identify **common patterns in manuscript maps or drawings**.
- Make publicly available **original & relevant AI datasets.**
- And many others… All objectives we can only achieve if we have an understanding of archival science and paleography.

InterPARES TrustAI

UNESCO Audio Archives

Interviews and reports 1950s-1980s.

~6,500 available on digital platform (16,000 total)

But only *~800* described so far

InterPARES TrustAI

UNESCO Archives Languages

70+ recognized languages

French

English

Spanish

Multilingual (4%)

InterPARES TrustAI

Metadata Scheme

57 elements total

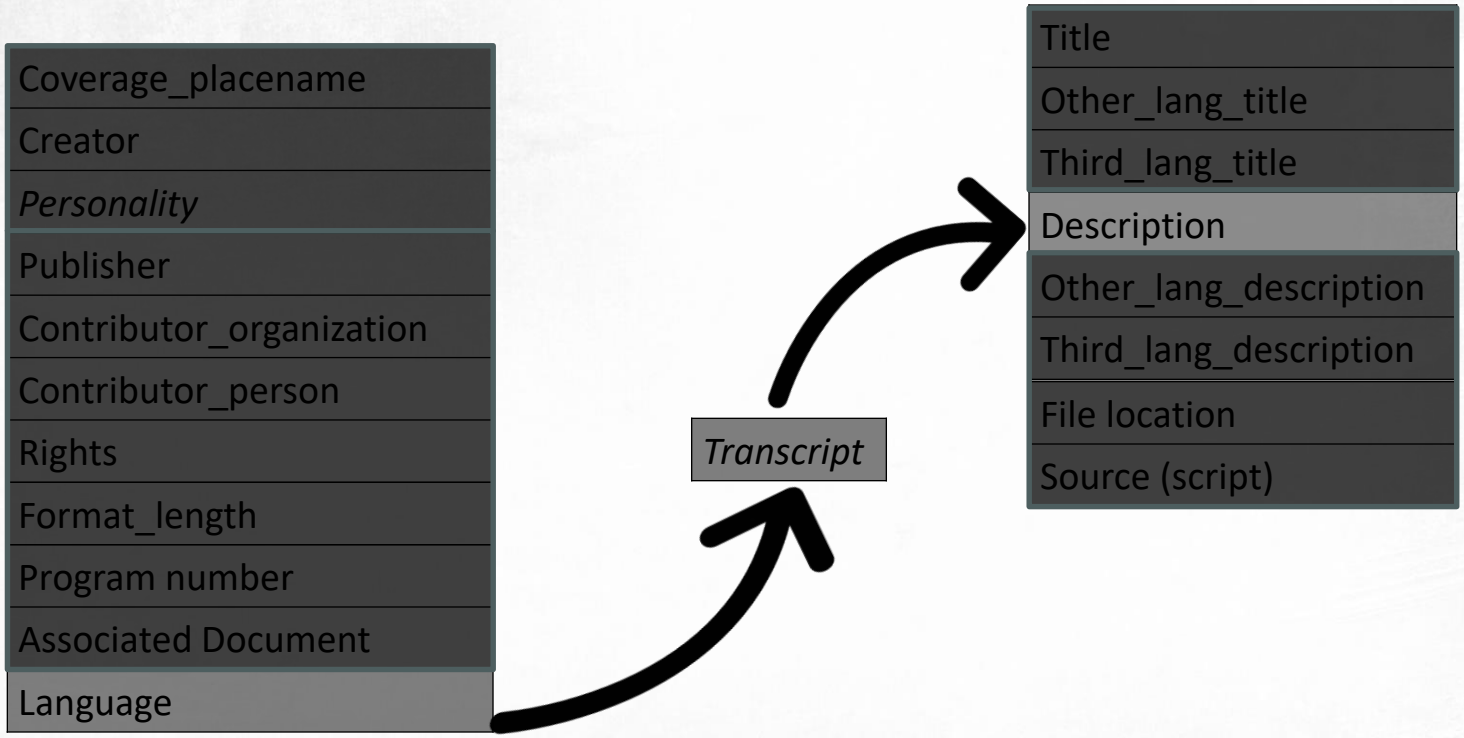
Vital for discoverability

Title
Other_lang_title
Third_lang_title
Description
Other_lang_description
Third_lang_description
File location
Source (script)

Coverage_placename
Creator
Personality
Publisher
Contributor_organization
Contributor_person
Rights
Program number
Associated Document
Format_length
Language
Access_category
Rightsholder

InterPARES TrustAI

UNESCO Archives Metadata Enrichment Plan



InterPARES TrustAI

Text vs. Speech Language ID

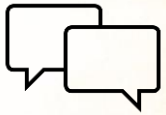
Text Based

- *Orthographic clues*
- *Full sentence used as input*

Models:

Transformer Model (BERT etc…)

Speech Based



- *Potential Confounders:*
  - *Speaker*
  - *Recording setup*
- *Short audio clip*

Models:

Convolutional Neural Network (x-vector) (Whisper etc.)

Fine-tuned Audio Transformer

InterPARES TrustAI

Steps in Metadata Creation

- **Speech Transcription**
- **Diplomatics Identification of genre/documentary form** (e.g. interview)
- **Diplomatics Labeling** according to: protocol (names of persons, dates, topic in interviews) or eschatocol (in reports), depending on genre of audio.

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 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

Research Questions

- Which questions need to be answered when designing, deploying and/or interpreting the results of the application of an AI tool? These are some of them:
- If error or even harm was a result—**what type of error or harm**?
- **Why** did it occur?
- **Who can or should be held responsible**? The **data curator** who provided the datasets? The **data scientist** who created the model? The **vendor** who supplied the AI tools? The **person or** the **organization** that implemented the AI tools? The **human users of the output of AI** who lacked sufficient training to understand the likelihood of error?

InterPARES TrustAI

Modes of Explanation

- **Causal** – How it functions
- **Epistemic** – How we know it functions
- **Justificatory** – On what grounds it functions

Justificatory -- Can refer to AI system properties (e.g., datasets and algorithms).

Must also reference **institutional and social facts** about the implementation of the system (e.g., regulations, standards, pertinent organizational processes).

ISO/IEC TR 24028 (2020-05) Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence

InterPARES TrustAI

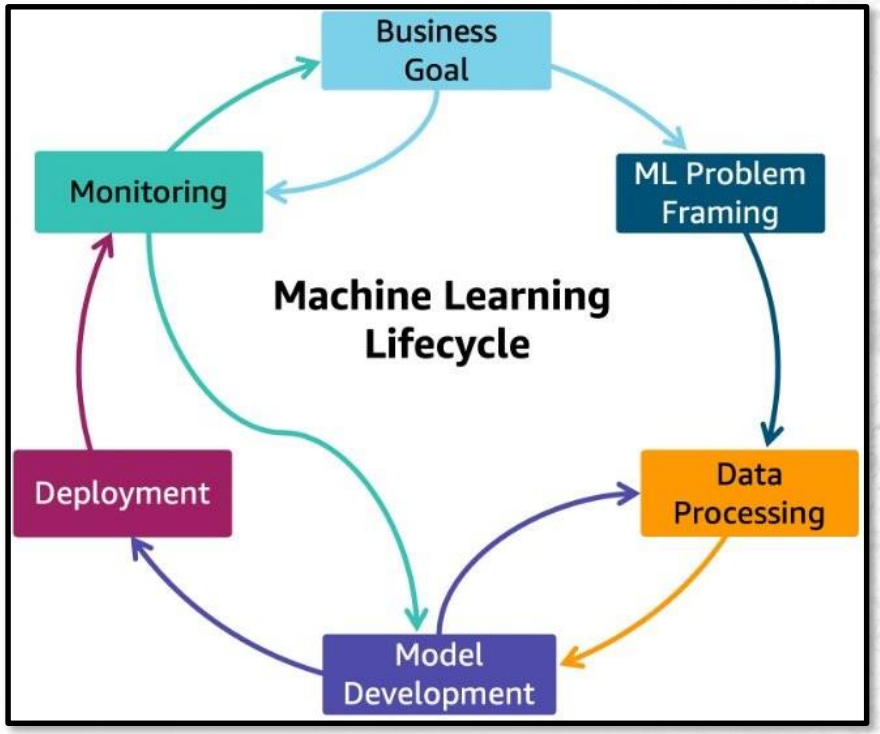
XAI vs. Accountable AI

- 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 in a particular context.**

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

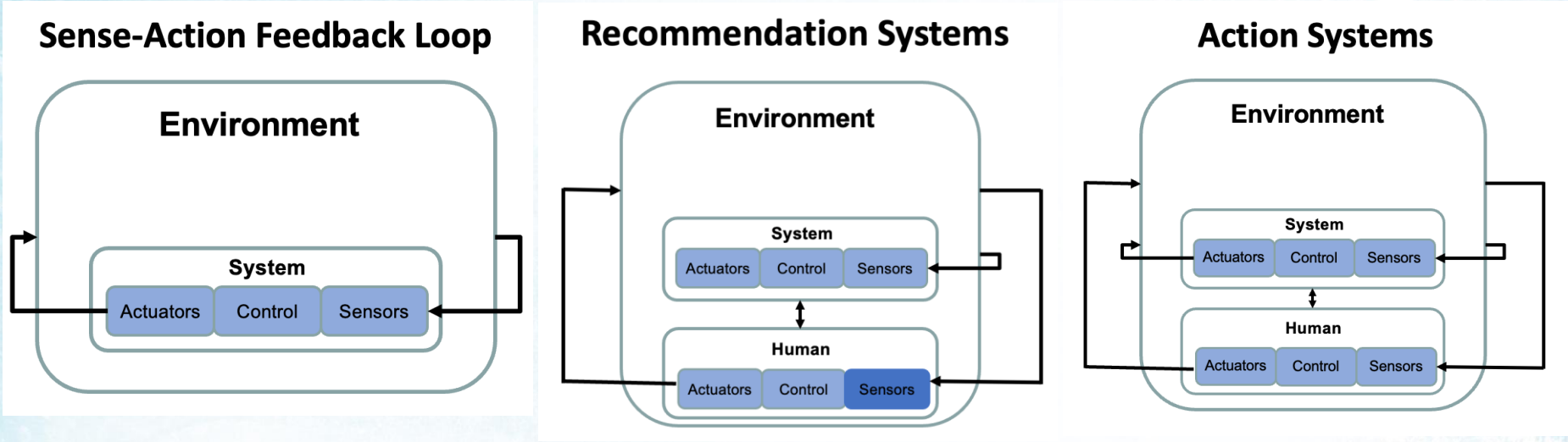
- AI Model (tested & selected)
- Evaluation & performance metrics
- Logs generated
- Model training data set
- Training parameters for model
- Vendor documentation
- Versioning information

Organizational Paradata

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

InterPARES TrustAI

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. Roomba vacuum cleaner (autonomous robot, no humans involved)
2. Buying recommendations (humans marginally involved)
3. Digital Twins (humans with shared agency)

InterPARES TrustAI

Shared Agency

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

- **Concurrent action (e.g. as airport traffic control & in building plans—digital twins)**
  - 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

Accountability for what action?

InterPARES TrustAI

Paradata Under Shared Agency

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

InterPARES TrustAI

Conclusion

- **Decisions made and actions taken by AI-enabled systems or tools must be documented.**
- Some of the documentation will be **automatic** as part of the AI system or tool; some will be **human-created prior to or after the creation and implementation of the AI system or tool**.
- **Paradata is necessary to promote transparency and accountability.**
- The capture and preservation of paradata ensures that the **AI process is documented in a way that preserves the authenticity of records but also supports their preservation. The more it changes…**

InterPARES TrustAI

Thank you!

www.interparestrustai.org

@itrustai

www.facebook.com/interparestrust