

Credits: Image generated by ChatGPT

Conceptualizing Access to Archives in the AI era. A study within Itrust AI.

Pierluigi Feliciati

Giorgia Di Marcantonio

University of Macerata



InterPARES Trust AI (2021-2026) is a five-year multi-national project funded by the Canadian SSHRC that aims to design, develop, and leverage Artificial Intelligence to support the ongoing availability and accessibility of trustworthy public records.

<https://interparestrustai.org/trust>

I Trust AI WG 6

REFERENCE AND ACCESS

RA05 - Users approaches & behaviors in accessing records
& archives in the perspective of AI

RA08 - Trusted access/use of archives and AI: a
conceptual model

InterPARES
TrustAI



RA05 - Users approaches & behaviors in accessing records & archives in the perspective of AI

TEAM: Pierluigi Feliciati (Coord.), Jessica Bushey, Thsepho Mosweu, Darra Hoffmann , Adele Torrance, Sindiso Bhebhe, Giorgia Di Marcantonio

Research questions :

1. How much do we know about how users access digital archives?
2. Are they aware of what AI is?
3. Would they share their research data to improve archival services by adopting AI tools?
4. Are they aware of what this could mean?

Activities

- a) Literature review of digital archives users
- b) A **user study protocol** (in English) focus groups + interviews
- c) Because of resource intensity & to start the study, the protocol was **translated in Italian**, adapted into a questionnaire **for an online survey**
- d) The survey was shared w/ **4 Italian State Archives** with digital services. Launched 10 October (2022) & open until the end of November (2022)

Results

Report, April 2023: <https://doi.org/10.5281/zenodo.7852020>

- **Next to none have** used AI tools w/ awareness
- **50%** are open to providing their research data to support AI tools to improve online archival services
- **75%** are ready to share their research data
- **+70%** are favorable to receiving support w/ automated suggestions
- **66%** want to use natural language queries based on AI
- **83%** are favorable to the use of AI tools to analyze & extract text elements in digitized documents

RA08 - Trusted access/use of archives and AI: a conceptual model

TEAM: Pierluigi Feliciati (coord.), Giorgia Di Marcantonio, Jim Suderman, Kenneth Thibodeau, Jessica Bushey, Pat Moore, Michel Barbeau, Michele A. L. Villagran, Hugo Carrion, with the support of Basma Makhoul and Jenny

Focus on:

- Starting from the existing standards, models, and reference documents on resource and record creation, management, appraisal, preservation, and description
- reflecting, refining, and releasing a **robust conceptual model of access/use to records/archives** to support the development of systems and tools.

RA08 Study steps

Analysis of existing conceptual models and glossaries: to identify their strengths and weaknesses and consider their applicability and improvements

Initial analysis of requirements and concepts: essential requirements, entities and relationships, to establish the core of the conceptual model

Draw a draft Conceptual Schema: an initial, comprehensive schema, to be refined over time to eliminate its redundancies and optimize structure

Classes and Subclasses: organize and categorize the entities and concepts identified during the initial analysis.

Model optimization: Refine and streamline the model to ensure efficiency, effectiveness, and ease of use, by removing redundancies and enhancing structure.

Applicable Case Studies: Find specific case studies relevant to test and present the validity, applicability, and effectiveness of the conceptual model.

Ongoing Updates: Plan updates to the model, considering the rapid evolution of Artificial Intelligence solutions, to ensure the model relevancy, and effectiveness

DONE

**TO BE
DONE**

E.g.

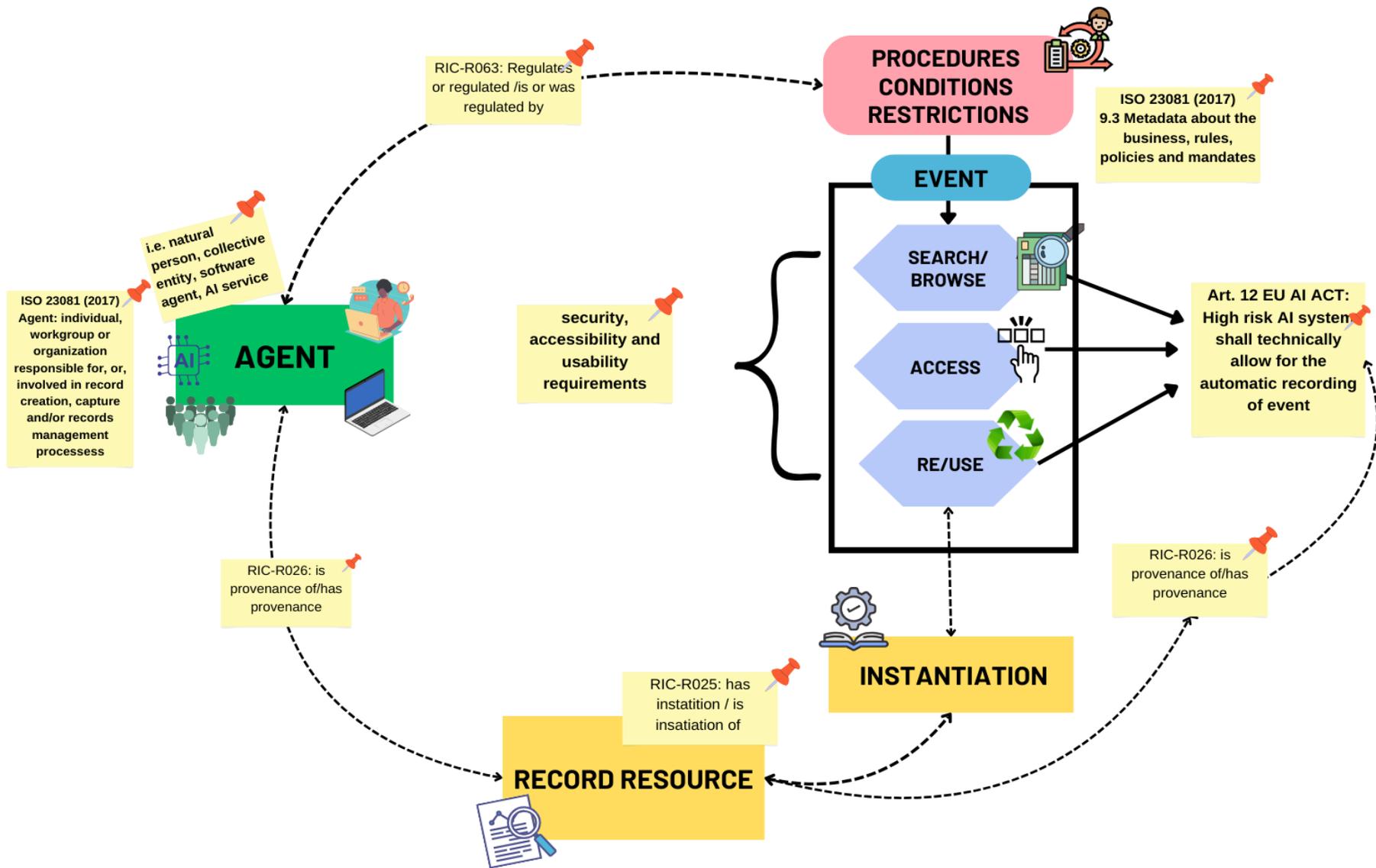
AGENT	Definition	Source
Agent	Agents may be human or machine. A category of agents, those responsible for the creation, compilation, and maintenance of the records is particularly important and must be described. Describing these agents requires archivists to document agents' roles, functions, occupations, and activities.	Describing Archives: A Content Standard DACS 2019.0.3. by Society of American Archivists
Actor	Entity that communicates and interacts	Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence. ISO/IEC TR 24028
Agent	Agent could be User, Archivist and Institutions (no machine)	Principles of Access to Archives. ICA 2012
Agent	"Agent" is defined as an actor (human, machine, or software) associated with one or more events and/or rights related to a digital object. Additionally, machine or software agents can themselves be captured as a digital object	PREMIS
Agent	An agent is a grouping or an organization of two or more people that carries a particular name and is capable of acting as a unit. A collective agent includes a legal entity and a family	RDA
Agent	Agents can be entities such as corporate bodies, persons, families, groups, positions, and mechanisms that perform activities, generate, or use record resources. The concept of an agent includes entities that are created by a person or group that act on behalf of the creating agent in an autonomous or semi-autonomous manner, such as software agents, robots, and probes that generate data (records) during their activities	RIC-CM

E.g.

ACCESS	Definition	Source
Access	is considered as a material and physical access to the archival material.	Describing Archives: A Content Standard DACS 2019.0.3. by Society of American Archivists
Access	right, opportunity, means of finding, using or retrieving information	Information and documentation — Records management. Part. 1 Concepts and principles. ISO/TR 15489-1
Access control	determining the behaviour or supervising the use of records through access and permission rules (3.5.1)	Information and documentation — Records management — Core concepts and vocabulary. ISO 30300
Access	The access point is an expression (name, term, code, etc.) that represents a specific entity; the authorized access point is the standardized access point that represents an entity; the variant access point is an alternative to the authorized access point that represents an entity. Through the creation of access points and the significant enhancement of the RDA relational network, users are enabled to find what they desire and to perform the function of 'navigating' among resources	RDA

E.g.

USE	Definition	Source
Use	when automation enables humans to achieve their goals;	Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence. ISO/IEC TR 24028
Use	a) identifying the records system user permissions associated with individuals and their positions within the organization, b) identifying the access and security status of records , c) identifying the access rights for people external to the organization, d) ensuring that only individuals with the appropriate user classification or security rights have access to records with restricted status, e) tracking the movement of the record to identify those who have or have had custody of it, f) ensuring that all use of the records is recorded to an appropriate level of detail, and g) reviewing the access classifications of records to ensure that they are current and still applicable.	Information and documentation — Records management. Part. 2 Guidelines. ISO/TR 15489-2
Condition of Use	These refer to terms and circumstances affecting the use of a record resource or an instantiation after access has been provided. This includes conditions governing reproduction of the record or record part under applicable copyright (intellectual property) and/or property legislation or due to conservation status. For example, permission from the copyright owner might be required before use, or there could be restrictions on copying using specific types of copying machines or photography	RIC-CM
Conditions Governing Reproduction and Use	This element identifies any restrictions on reproduction due to copyright or other reasons , as well as restrictions on further use of the materials being described, such as publication, after access has been provided.	Describing Archives: A Content Standard DACS 2019.0.3. by Society of American Archivists



Pillars (under discussion)

Assertion form

- An agent under specific conditions is allowed to search, access, and reuse a record resource (and this is an event to be recorded).
- The agent can use the instantiation of the resource, i.e., its version prepared (delivered) as an ad hoc version (see DIP in OAIS).
- The event of searching, accessing, and reusing generates a new record and is operated under conditions of accessibility, usability, security, and reliability.

Open questions

- Define classes and **subclasses**
- Formalize the **time** implications of the process
- Better definition of the differences among access, use, and reuse



**THANK
YOU!**

g.dimarcantonio@unimc.it
pierluigi.feliciati@unimc.it