

The responsible use of Artificial Intelligence in archives through the use of paradata

Pierluigi Feliciati^(a)
Luciana Duranti^(b)

a) University of Macerata, <https://orcid.org/0000-0002-2499-8528>

b) University of British Columbia, Canada, <https://orcid.org/0000-0001-7895-1066>

Contact: Pierluigi Feliciati, pierluigi.feliciati@unimc.it; Luciana Duranti, Luciana.Duranti@ubc.ca

Received: 10 December 2024; **Accepted:** 07 March 2025; **First Published:** 15 May 2025

ABSTRACT

Using Artificial Intelligence in archives responsibly, clearly documenting why, how, by whom and with what results a specific system or tool was used in a given context, requires the conscious adoption of paradata throughout the entire period of use of such system or tool and not only at the start or end of a procedure. Paradata are structured and formalized data documenting the methodologies, processes (including technological ones) and persons (physical and juridical) associated with the production and management of records. This contribution focuses on paradata generated in connection with the use of AI systems or tools in archives to perform archival functions. Its content is primarily a result of the research conducted within the context of the international project InterPARES Trust AI.

KEYWORDS

Digital preservation process; Archival Records; Artificial Intelligence; Paradata.

L'uso responsabile dell'Intelligenza Artificiale negli archivi attraverso l'uso di paradati

ABSTRACT

Usare Intelligenza Artificiale responsabile, che documenti chiaramente perché, come, da chi e con quale risultato un certo sistema o strumento sia stato usato in un determinato contesto, richiede l'adozione consapevole dei paradati, necessari nell'intero ciclo di uso di uno strumento e non solo all'avvio o alla fine di una procedura. I paradati sono dati strutturati e formalizzati che documentano le metodologie, i processi (inclusi quelli tecnologici) e le persone (fisiche e giuridiche) associati alla produzione e alla gestione dei documenti. Il focus di questo contributo riguarda i paradati prodotti in connessione all'uso di sistemi o strumenti di Intelligenza Artificiale negli archivi per esercitare funzioni archivistiche. Il suo contenuto è principalmente il risultato parziale delle ricerche condotte nell'ambito del progetto internazionale InterPARES Trust AI.

PAROLE CHIAVE

Processi di conservazione digitale; Archivi digitali; Intelligenza Artificiale; Paradati.

Introduzione

“A pivotal question linked to data, information, and knowledge is what we need to know about data, information, and knowledge and how it has been managed in order to use it? Sometimes the answer is ‘very little.’ [...] This is where paradata, i.e., information on the processes of data creation, curation, and earlier and intended use, can be pivotal for the usefulness of data, information, and knowledge.” (Andersson et al. 2024, 1-2).

Nel contesto archivistico, mentre i metadati sono i dati strutturati e formalizzati relativi ai documenti archivistici, necessari per cercare, identificare, visualizzare, analizzare, valutare e gestire nel tempo quei documenti, i paradatai sono dati strutturati e formalizzati che documentano le metodologie, i processi (inclusi quelli tecnologici) e le persone (fisiche e giuridiche) associati alla produzione e alla gestione dei documenti. Come fenomeno informativo, i paradatai differiscono dai metadati anche qualitativamente, perché puntano a testimoniare il contesto di produzione e gestione piuttosto che l'identità e l'integrità dei documenti, e quantitativamente, in quanto comprendono informazioni più ampie e meno dettagliate rispetto ai metadati.

Il recentissimo volume curato da Lisa Andersson, Isto Huvila e Olle Sköld della Uppsala University (Andersson et al. 2024), nell'affrontare il tema dei paradatai, ha scelto di coinvolgere studiosi di diverse discipline, anche lontane dall'archivistica, che si confrontano con la gestione delle informazioni. Un punto di partenza per comprendere il ruolo dei paradatai può essere lo spinoso problema di come gestire gli aspetti taciti e impliciti della conoscenza e unirli con forme esplicite e formalizzate di informazioni e dati (Polanyi 1967), connettendo nei fatti le due principali prospettive sulla gestione della conoscenza, quella sociale e organizzativa di mediazione e traduzione tra le persone e quella tecnica di formalizzazione e gestione in sistemi, ambedue tendenzialmente soluzioni tecniche (Handzic 2004). I paradatai, applicati ad esempio nelle scienze statistiche per presentare i dati dei sondaggi, sono finalizzati alla descrizione del processo di elaborazione e conduzione del sondaggio (Couper 1998; 2000; Frauke Kreuter 2010). In modo simile, nel settore delle applicazioni tecnologiche al patrimonio culturale sono stati utilizzati per raccogliere sistematicamente informazioni sulla creazione e l'elaborazione dei dati, specie per le elaborazioni tridimensionali (Denard 2012). Infatti, la creazione di ricostruzioni 3D, un processo tutt'altro che lineare, richiede la registrazione dei dati utilizzati e la decodifica dei processi adottati per garantirne la documentazione (Papadopoulos 2024).

Va menzionato in particolare che, nel contesto archivistico, altri concetti e metodi non lontani dai paradatai sono già utilizzati: tra tutti, quello della *provenienza* è un concetto usato comunemente dalla professione archivistica per descrivere sia la creazione che la gestione di oggetti informativi e in particolare dei documenti archivistici. Rappresentare la provenienza può comprendere la descrizione sia delle persone giuridiche coinvolte, sia delle azioni che queste intraprendono per produrre (generare o ricevere), organizzare, gestire, valutare e scartare i documenti archivistici (Lemieux 2016). Il focus di questo articolo è la creazione di paradatai in connessione all'uso di strumenti di Intelligenza Artificiale (IA) negli archivi per esercitare funzioni archivistiche. Il suo contenuto è il risultato soprattutto della ricerca condotta nell'ambito del progetto internazionale InterPARES Trust AI (2021-2026).¹

¹ In breve, *I Trust AI*. Il sito web del progetto di ricerca è www.interparestrustai.org.

Lo studio *Preserving AI Techniques as Paradata*

I Trust AI ha lo scopo di concepire, disegnare, sviluppare e usare sistemi o strumenti di IA per garantire la conservazione e l'accessibilità di documenti archivistici affidabili, accurati, e autentici. A questo fine, il gruppo di ricerca porta avanti sia studi di casi, sia studi di carattere generale. Lo studio sui paradata fa parte di questi ultimi.² Il suo obiettivo è sviluppare un approccio per documentare l'uso di sistemi o strumenti di IA che soddisfi le responsabilità dell'archivio rispetto ai suoi utenti. L'IA responsabile (*accountable*) è cosa diversa rispetto all'IA spiegabile (*Explainable AI*), il cui scopo è solo quello di rendere uno strumento comprensibile. L'IA responsabile deve documentare chiaramente perché, come, da chi, e con quale risultato un certo sistema o strumento sia stato usato in un determinato contesto. Il tipo di spiegazione offerta può comunicare come lo strumento funziona, come sappiamo che funziona, perché funziona nel contesto in cui si intende usarlo, ma anche quali sono i riferimenti istituzionali nel cui ambito viene usato, ovvero le policy, i regolamenti, e le procedure pertinenti al suo uso specifico, insieme ai nomi delle persone responsabili per usarlo. Se in ambito informatico i termini IA responsabile e IA spiegabile sono considerati spesso pressoché intercambiabili³, uno dei risultati dei lavori interdisciplinari di InterPARES è stato chiarirne la differenza. *Accountable* non è solo *Explainable*, perché conoscere l'algoritmo e il modo in cui uno strumento funziona non attribuisce esplicitamente nessuna responsabilità, mentre le scelte di utilizzare un determinato strumento, istruirlo con determinati *data sets* creando un modello unico per uno scopo preciso sono attribuibili a una persona naturale o giuridica che le fa. Le persone, le procedure e il loro contesto sono componenti chiave del ciclo della responsabilità; quindi, devono essere documentate in connessione alle informazioni sulle componenti tecnologiche.

Conseguentemente a tali riflessioni, il gruppo di ricerca ha sviluppato prima di tutto una definizione operativa: "I paradata sono informazioni sulle procedure e sugli strumenti utilizzati per creare ed elaborare risorse informative, insieme a informazioni sul funzionamento degli strumenti e sull'esecuzione delle procedure, nonché sulle persone che eseguono tali procedure e utilizzano tali strumenti".⁴

Gli studi di InterPARES considerano inoltre i paradata necessari nell'intero ciclo di uso di uno strumento. Questo significa che i paradata non sono determinati solo all'inizio di una procedura o aggiunti alla fine della stessa, ma sono espressamente dichiarati dal momento della scelta del sistema o dello strumento per uno scopo specifico fino al raggiungimento di tale scopo, includendo la descrizione del suo risultato.

Quali paradata siano necessari e come raccogliarli e conservarli sono due questioni problematiche. Poiché l'implementazione di un processo di IA è unica per ogni soggetto archivistico e dipende dai suoi scopi, dalla sua tolleranza dei rischi e dalle sue risorse, la risposta alle due questioni sarà

² I ricercatori coinvolti nello studio *Research and Preservation 04, Preserving AI Techniques as Paradata*, sono: Patricia Franks, Babak Hamidzadeh, Scott Cameron, Norman Mooradian, Alex Richmond e Mario Beauchamp. Per approfondire cfr. (Franks et al. 2022; Cameron et al. 2023; Davet et al. 2023; Cameron e Hamidzadeh 2024).

³ Cfr. ad esempio Baker, Xiang 2023.

⁴ Traduzione italiana della definizione presente nell'InterPARES Terminology database: "Information about the procedure(s) and tools used to create and process information resources, along with information about the persons carrying out those procedures", <https://interparestrustai.org/terminology/term/paradata/en>.

necessariamente diversa da un contesto ad un altro. Al tempo stesso, lo scopo dello studio di I Trust AI è da un lato identificare elementi comuni che debbano essere documentati in tutti i contesti di uso archivistico di sistemi o strumenti di IA, e dall'altro identificare una tipologia di elementi unici che possono essere adottati da organizzazioni che intendano sviluppare una propria, specifica strategia.

A questo scopo, il gruppo di ricerca ha condotto uno studio di caso nell'archivio della Banca del Canada.⁵ I paradati identificati sono di due tipi: tecnici e organizzativi. Quelli tecnici includono il *data set* e i parametri usati per istruire il modello, i *log* generati durante l'uso dello strumento e, nel caso si tratti di un sistema o strumento commerciale, non sviluppato dall'archivio, la documentazione fornita dal venditore e le informazioni sulla versione usata. I paradati organizzativi consistono nella *policy* riguardante l'uso dell'IA, nei regolamenti e nelle considerazioni etiche, nella pianificazione progettuale, nella formazione del personale dell'archivio, nella valutazione dell'impatto dell'uso della IA, e nella procedura di implementazione.

Per quanto riguarda la conservazione, le domande che il gruppo di ricerca si è posto sono: in quale parte del sistema di tenuta dei documenti o di conservazione archivistica vanno mantenuti i sistemi o gli strumenti di intelligenza artificiale? Come possiamo conservare le loro componenti? In quali situazioni dobbiamo conservare i sistemi o gli strumenti di IA, o parti di questi? Fino a che punto ciò che conserviamo deve essere riproducibile? Se ci sono esseri umani coinvolti nel processo decisionale, come rappresentiamo e conserviamo le informazioni sui loro ruoli?

Dobbiamo infatti considerare che ci sono tre tipi di sistemi di IA. Il tipo più comune è quello che non prevede la partecipazione di essere umani. I sensori raccolgono dati dall'ambiente, gli strumenti di controllo li analizzano e decidono quali azioni compiere, le componenti esecutive agiscono e il risultato di tali azioni torna ai sensori, quindi il ciclo si ripete. Questo comportamento è quello tipico dei robot. Il secondo tipo di IA, con cui abbiamo sempre più familiarità è quello che raccomanda agli esseri umani, sulla base del loro comportamento online, cosa comprare o quale sito visitare. Questo tipo di sistema, in modo simile a quello precedente, non risulta difficile da documentare e da conservare. Gli esseri umani possono reagire o no, ma le loro attività si alternano, non sovrappongono, a quelle del sistema.

⁵ I professionisti responsabili per lo studio di caso sono Alex Richmond, Mario Beauchamp e John Cameron della Bank of Canada.

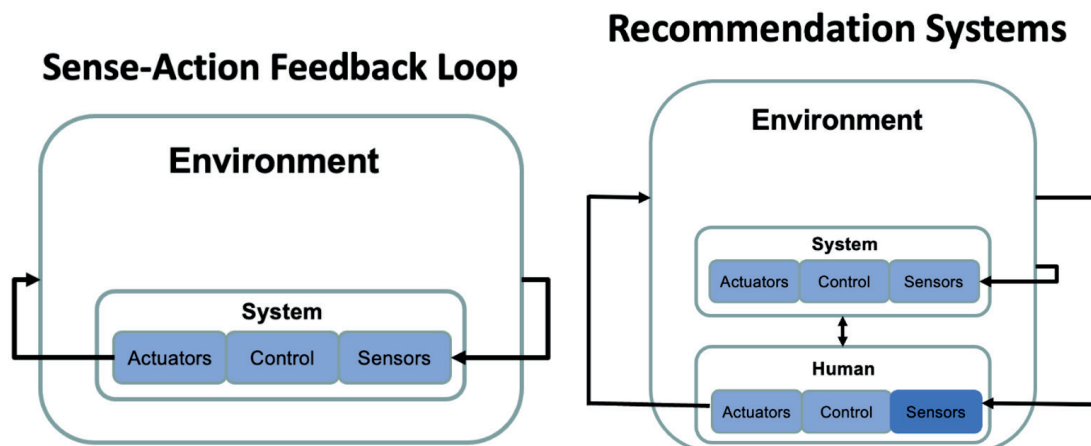


Figura 1. P. Franks, *Paradata - what is it and why do we care?* Presentazione in *Using AI To Interrogate Archives - iTrust AI Public Symposium*, 23 Febbraio 2024, Honolulu.⁶

Il terzo tipo di sistema, più problematico, è quello in cui l'IA e gli esseri umani interagiscono, ma possono farlo in modo indipendente, a volte a turno, altre in sovrapposizione o congiuntamente e altre volte per correggere, aggiungere, o sostituire l'azione l'uno dell'altro (come in caso di emergenze). Questi *action systems* sono dinamici, richiedono azioni in tempo reale e possono coinvolgere un numero variabile di agenti diversi.

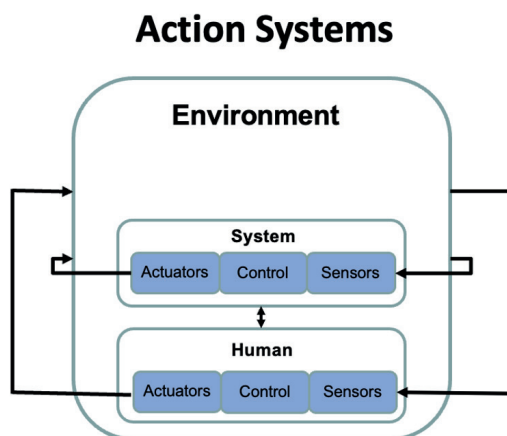


Figura 2. P. Franks, *Paradata - what is it and why do we care?* Presentazione in *Using AI To Interrogate Archives - iTrust AI Public Symposium*, 23 Febbraio 2024, Honolulu.

Cosa risulta preferibile e possibile conservare di questi sistemi o strumenti? Certo, prima di tutto informazioni sul sistema/strumento stesso: l'architettura e il *design* di hardware e software, il codice, il modello, gli algoritmi, la logica e i meccanismi di esecuzione. Inoltre, la documentazione

⁶ La registrazione del Symposium è disponibile in https://fb.watch/qEq_IVLAyu/.

relativa al mantenimento e all'aggiornamento, i dati usati per istruire il modello, testarlo e validarlo e infine i risultati di tutte queste azioni, più gli strumenti per il funzionamento del sistema/strumento. A tutte queste informazioni relative al sistema/strumento e al suo uso bisogna aggiungere i documenti prodotti dall'organizzazione che usa il sistema/strumento relativi al processo di auto-controllo, ai test che precedono l'adozione del sistema/strumento e, nel caso di sistemi/strumenti commerciali, alla storia delle vendite di modelli e configurazioni.

Tutti questi sono da considerarsi paradatai generali, che devono essere accompagnati da paradatai operativi. Questi ultimi includono:

- l'input fornito dai sensori (*log* dei dati e, nel caso di sistemi di *computer vision*, i video e/o le immagini);
- le decisioni prese dal *controller* (*log* delle direzioni e relativi parametri di *setting*, dati sugli intermediari che hanno supportato le decisioni prese, le spiegazioni *post factum* di questi processi, i *log* degli avvertimenti e delle notifiche di passaggio/affidamento di responsabilità);
- le azioni dell'attuatore (componente esecutiva) (*log* delle azioni controllate dagli esseri umani e di quelle controllate dal sistema IA, *log* dei messaggi del sistema per i controllori umani e per i soggetti esterni);
- e i risultati (*log* dei dati percepiti dai sensori dopo che le azioni degli attuatori sono state completate, o video delle conseguenze di tali azioni).

Per quanto riguarda i paradatai, come detto sopra, la documentazione dovrebbe essere continua, ovvero basata su un modello ciclico di *feedback* (sensori, controllori, attuatori) al fine di rispecchiare ogni fase, ma anche a campione e relativa a eventi specifici nel processo. Inoltre, è importante documentare la relazione tra i dati percepiti e registrati dai sensori, la logica usata dal controllo nel determinare l'azione che ne deve conseguire, l'attuazione di tale azione e le sue conseguenze, e infine la dimensione temporale.

Le decisioni prese dall'archivio e le azioni dei sistemi o degli strumenti di IA conseguentemente adottati devono insomma essere documentate. Parte di questa documentazione sarà automatica, come prodotto del sistema o strumento di IA, e un'altra parte sarà generata dall'archivio prima dello sviluppo o dell'adozione dello specifico sistema o strumento (ad esempio in forma di linee guida di natura etica, regolativa o procedurale), oppure durante e successivamente al loro uso, indicando le persone che hanno portato avanti le varie fasi della procedura, come la determinazione del *data set* per istruire il modello o il suo testing. Tutti questi paradatai sono necessari non solo a promuovere la piena trasparenza ma soprattutto a soddisfare i requisiti relativi alle attestazioni di responsabilità. I paradatai e la loro conservazione garantiscono che i processi portati avanti da sistemi o strumenti di IA siano documentati e l'autenticità dei documenti da loro generati sia garantita.

Conclusioni

Ciò detto, va tuttavia sottolineato che i profili applicativi e la necessità dei paradatai rimangono un campo di studio aperto nella maggior parte delle discipline interessate alla creazione e gestione di dati. Allo stesso tempo, le relazioni dei paradatai con i metadati e la prospettiva di incorporare le informazioni relative all'elaborazione dei dati in forma di descrizioni strutturate sono attualmente in fase di sperimentazione da parte di diversi gruppi di ricerca.

Esistono già, infatti, modelli ben definiti che coprono tipi specifici di paradati, come lo standard CIDOC-CRMdig,⁷ che ambisce a documentare le fasi e i metodi di produzione dei prodotti di digitalizzazione e delle rappresentazioni digitali, la specifica PROV del W3C per la modellazione della provenienza dei dati (Groth e Moreau 2013) e il *Common Workflow Language* per la formalizzazione dei flussi di lavoro analitici computazionali (Amstutz et al. 2016).

La inevitabile trasversalità interdisciplinare delle indagini sui paradati deve costituire l'occasione per attivare un effetto positivo, per espandere gli orizzonti della ricerca e delle pratiche di gestione delle informazioni e della conoscenza.

⁷ <https://cidoc-crm.org/crmvig/home-2>.

Riferimenti bibliografici*

Amstutz, Peter, Michael R. Crusoe, Nebojša Tijanić, Brad Chapman, John Chilton, Michael Heuer, Andrey Kartashov, Dan Leehr, Hervé Ménager, Maya Nedeljkovich, Matt Scales, Stian Soiland-Reyes, e Luka Stojanovic. 2016. *Common workflow language, v1.0*. <https://doi.org/10.6084/M9.FIGSHARE.3115156.V2>.

Andersson Lisa, Isto Huvila Isto, e Olle Sköld Olle. 2024. “An Introduction to Paradata.” In *Perspectives on Paradata. Research and Practice of Documenting Process Knowledge*, a cura di Lisa Andersson, Isto Huvila, e Olle Sköld. Cham, Switzerland: Springer, <https://doi.org/10.1007/978-3-031-53946-6>.

Baker Stephanie e Xiang Wei. *Explainable AI is Responsible AI: How Explainability Creates Trustworthy and Socially Responsible Artificial Intelligence*. Sottomesso il 4 dicembre 2023. <https://arxiv.org/abs/2312.01555v1>.

Bentkowska-Kafel, Anna, e Hugh Denard, a c. di. 2012. *Paradata and Transparency in Virtual Heritage*. London: Routledge. <https://doi.org/10.4324/9781315599366>.

Cameron, Scott, e Babak Hamidzadeh. 2024. “Preserving Paradata for Accountability of Semi-Autonomous AI Agents in Dynamic Environments: An Archival Perspective.” *Telematics and Informatics Reports* 14. <https://doi.org/10.1016/j.teler.2024.100135>.

Cameron, Scott, Patricia Franks, e Babak Hamidzadeh. 2023. “Positioning paradata: A conceptual frame for AI processual documentation in archives and recordkeeping contexts.” *ACM. Journal of Computing and Cultural Heritage* 16 (4): 1-19. <https://doi.org/10.1145/3594728>.

Couper, Mick P. 1998. “Measuring survey quality in a CASIC environment.” In *Proceedings of the Section on Survey Research Methods of the American Statistical Association* 6: 41-49. http://www.asarms.org/Proceedings/papers/1998_006.pdf.

Couper, Mick P. 2000. “Usability evaluation of computer-assisted survey instruments”. *Social Science Computer Review* 18(4): 384-396. <https://doi.org/10.1177/089443930001800402>.

Davet, Jeremy E., Babak Hamidzadeh, e Patricia Franks. 2023 “Archivist in the machine: paradata for AI-based automation in the archives.” *Archival Science* 23: 275-295. <https://doi.org/10.1007/s10502-023-09408-8>.

Franks, Patricia C., Jeremy E. Davet, Babak Jeremy E., Babak Hamidzadeh, e Jenny Bunn. 2022. “Tracking the functions of AI as paradata & pursuing archival accountability.” *Archiving 2022: Final Programs and Proceedings, Society for Imaging Science and Technology*, 19: 83–88. <http://dx.doi.org/10.2352/issn.2168-3204.2022.19.1.17>.

Groth, Paul, e Luc Moreau. 2013. *PROV-Overview: An Overview of the PROV Family of Documents*. <http://www.w3.org/TR/2013/NOTE-prov-overview-20130430/>.

Handzic, Meliha. 2004. *Knowledge management: Through the technology glass*. Singapore: World Scientific.

* Data di ultima consultazione dei siti web: 1 marzo 2025.

Kreuter, Frauke, Mick Couper, e Lars Lyberg. 2010. *The use of paradata to monitor and manage survey data collection*. In *Section on Survey Research Methods – JSM 2010*, 282-296. http://www.asarms.org/Proceedings/y2010/Files/306107_55863.pdf.

Lemieux, Victoria, a c. di. 2016. *Building Trust in Information*. Cham: Springer.

Papadopolous, Costas. (2024). “A Leap of Faith: Revisiting Paradata in 3D Scholarship.” In *Perspectives on Paradata. Research and Practice of Documenting Process Knowledge*, a cura di Lisa Andersson, Isto Huvila, e Olle Sköld, 61- 86. Cham: Springer. <https://doi.org/10.1007/978-3-031-53946-6>.

Polanyi, Michael. 1967. *The tacit dimension*. London: Routledge. Tradotto da Franco Voltaggio come *La conoscenza inespresa* (Roma: Armando. 1979).