ADVANCES IN KNOWLEDGE ORGANIZATION
Proceedings of International ISKO Conference - ISSN: 0938-5495 - 2026 - São Paulo

Transforming AI Transparency in Archival Contexts: Paradata as an Ethical Foundation for Knowledge Organization

Moises Rockembach

CEIS20/University of Coimbra, Portugal

Abstract

The integration of artificial intelligence (AI) in archival practices has created new opportunities and epistemological problems for Knowledge Organization (KO) studies. While AI has the potential to improve various archival practices, such as digitization, description, and access, it also creates problems related to the "black box" nature of AI, which is in conflict with archival principles such as provenance, authenticity, and accountability. This work aims to investigate the role of ethical theories, particularly those related to AI governance, in ensuring the development of transparent AI systems in memory institutions. This study uses a qualitative approach, involving literature review and documentary analysis, to examine the relationship between archival concepts, such as provenance, and the newly developed concepts of AI, such as paradata and accountability. This relationship is analyzed in the context of the "algorithmic condition," in which AI systems reflect and perpetuate historical biases. The results show that AI transparency is a documentary and ethical issue. The concept of paradata is introduced as a way of documenting AI algorithmic processes, ensuring auditability, detection of biases, and accountability.

1.0 Introduction

The concept of artificial intelligence (AI) has been increasingly incorporated into the workflows of memory institutions and the archival repositories. From the use of AI in the digitization of materials and the generation of description using algorithms, to the use of semantic search and AI-based appraisal and mediation, the potential for AI in the stewardship of cultural heritage has never been more promising. Unfortunately, the precipitous rollout of AI has also led to an unprecedented epistemological crisis for the field of Knowledge Organization (KO). Indeed, many of the more advanced machine learning-based approaches, such as deep learning and language models, are considered "black box" approaches that are antithetical to the traditional notions of provenance, authenticity, accountability, and transparency that have been the hallmarks of the archival field for centuries.

This study is guided by the following research question: How can ethical frameworks, grounded in both archival theory and contemporary AI governance, support the development of transparency in AI systems used in memory institutions? It adopts a qualitative approach based on literature review and documentary analysis. This study

argues that the development of transparent AI is not a technical problem but an ethical and documentary solution that the archival field has been familiar with for a long time.

2.0 Theoretical Background: Knowledge Organization and the Algorithmic Condition

Historical studies in the field of Library and Information Science (LIS) and KO have clearly established that traditional classification systems and metadata systems are, in fact, sociotechnical systems that promote particular worldviews. The dominant systems have faced significant scholarly criticism for promoting Western, patriarchal, and colonial epistemologies, which have marginalized alternate knowledge systems and identities (Shiozaki 2025).

However, the advent of AI-based KO systems poses new historical vulnerabilities in the field, which brings the entire discipline into the "algorithmic condition," as proposed by theorists in the recent past. In the algorithmic condition, the AI systems are not simply following the taxonomic systems devised by human beings; they are actually creating new patterns of meaning, clustering, and categorization based on the massive, undocumented datasets used for training these systems. However, these datasets are typically drawn from societies shaped by historical inequalities. As a result, such biases are reproduced and amplified by AI systems. These concerns are closely aligned with key principles in AI governance, particularly accountability, fairness, and human oversight.

Previous studies have examined the role of big data and artificial intelligence in information organization and mediation in archives and libraries (Melo and Rockembach 2019; Rockembach 2021; Rockembach 2024; Sousa and Rockembach 2025). In the case of appraisal of archives or subject indexing, this phenomenon leads to the creation of a new phenomenon called "Bias Squared," wherein the historical biases in the physical collection are multiplied manifold with the algorithmic bias of the AI systems used for appraisal, leading to the creation of a new level of epistemic violence in the digital systems of memory institutions (Jaillant and Rees 2023).

The emphasis on AI-based Knowledge Organization heightens the pre-existing weaknesses of the field, positioning it at the center of what has been named the “algorithmic condition.” In this framework, AI technologies do not merely execute pre-programmed schemes of categorization, as developed by humans, but instead develop their own schemes of categorization according to data that is, by and large, not transparent. In that these data sets are the products of pre-existing socially biased environments, AI technologies inevitably propagate these biases, as has been noted by Longpre et al. (2024), leading to what has been named bias squared when these technologies are utilized in the service of tasks such as appraisal or subject indexing, thereby heightening pre-existing archival biases that contribute to epistemic violence in digital memory infrastructures.

3.0 Methodology

This study adopts a qualitative research design based on literature review and documentary analysis. The research aims to examine how theoretical concepts from Knowledge Organization can be applied as ethical foundations for AI transparency. The literature was selected from key databases (Scopus and Web of Science) and other relevant publications and reports from 2018 to 2026. The search parameters included combinations of keywords such as 'artificial intelligence', 'transparency', 'ethics', 'paradata', and 'knowledge organization', filtered by peer-reviewed articles in English, resulting in a corpus of approximately 20 sources. The literature was reviewed based on its relevance to the field of archival science, knowledge organization, and AI ethics. This interdisciplinary review integrates literature from the fields of archival science, knowledge organization, and AI ethics. It highlights the conceptual relationships between existing concepts of archival science such as provenance, original order, authenticity, and evidentiality, and new concepts in AI governance such as paradata, accountability, fairness, and human oversight.

4.0 From Provenance to Paradata: Reclaiming Archival Accountability

The principle of provenance represents one of the most important theoretical milestones in the development of the theory of archival science. From the inductive approaches of the nineteenth century, which led archivists to respect the organic nature of document creation and production, the concept of provenance has developed through the French concept of respect des fonds and the Registraturprinzip of the Germans. Provenance allows for the understanding of the record as part of a network of actions and functions. This network represents the 'archival bond,' which offers the indispensable context for the authentication of historical actions (MacNeil 2004). Provenance represents an important concept in the contemporary KO theory as a domain analysis approach (Tognoli and Chaves-Guimarães 2019). The study of the organizational structures and discourse communities that generate knowledge is an important requirement for the application of the concept of provenance.

Nevertheless, the development of the digital environment and the emergence of AI have led to the requirement for a paradigm shift in the application of the concept of provenance (Collomosse et al. 2024). Digital records are no longer static physical objects with a static and linear provenance. They are dynamic and constantly changing as a result of the parallel and constant interventions of human and artificial agents. In this context, the study of the provenance of digital records requires the mapping of the interventions of the computer and the authentication of synthetic media. The development of the Coalition for Content Provenance and Authenticity (C2PA) has led to the integration of machine-readable credentials in the digital environment. Nevertheless, the technical application of the concept of provenance may authenticate the origin of the digital file but may not fully understand the complex reasoning and procedural nuances of the AI system for the entire archival collection.

The term paradata was first used in the statistical social sciences in the late 1990s to refer to data that is collected as a byproduct of the data collection process. Recently, a formal adaptation of this concept has been developed by the InterPARES Trust AI research group, which defined paradata as a set of information related to procedures, tools, and human

agents involved in the creation and processing of information resources (Cameron et al. 2023). If metadata represents a set of attributes related to a record, then paradata represents a set of attributes related to how that record has been manipulated, evaluated, or generated through an algorithmic process.

Paradata specifically tackles the "black box" problem by rendering invisible processes transparent. Paradata supplies archivists with not only the conceptual vocabulary but also the structural mechanism to describe AI systems, not only the result of a computerized appraisal or classification process, but also how that result was generated, including logic, parameters, and error rates (Davet et al. 2022). If paradata is incorporated as one of the fundamental KO principles, alongside "provenance," then processual transparency can be established, enabling accountability of nonhuman agency to public scrutiny. Recent literature corroborates that the ethical and accountability challenges brought by the use of Artificial Intelligence in archives are increasingly being addressed through the development and application of the concept of paradata (Arias Hernández and Rockembach 2025).

The findings also serve to illustrate the indispensable role that paradata plays in rendering the operation of AI systems auditable within memory institutions. Through the systematic collection of paradata, the operation of these otherwise opaque mechanisms can be made visible. Detailed paradata logs, which document the exact datasets employed within the training of the AI models, the parameters employed within the deployment of the system, and the statistical confidence thresholds employed within the automated classification of data, are key to identifying the sources of bias within the system (Davet et al. 2022). Paradata also offers the necessary operational knowledge to clarify the exact processes by which the output of the AI system is generated.

It is also important to highlight the need for the creation of “human-in-the-loop” or “computer-in-the-loop” logs (Shang et al. 2025). By documenting the exact circumstances by which the human archivist overrode the output of the AI system (for instance, within the automated review of classified government documents), the institution secures its intellectual accountability. This also confirms that the creation of transparent AI systems is ultimately a documentary challenge, requiring the documentation of the interplay between human and computer processes as opposed to the reverse-engineering of the underlying mathematics of the algorithm itself. To effectively manage this interplay, studies in Digital Humanities highlight that AI literacy is crucial for preparing cultural heritage professionals to navigate the ethical, legal, and technical challenges of digital curation (Melo and Rockembach 2023).

5.0 Conclusions

The integration of artificial intelligence into the archival workflow represents a turning point in the development of Knowledge Organization. As artificial intelligence increasingly performs tasks as complex as description generation and semantic access, the very opacity of these systems undermines the fundamental tenets of the archival profession. Yet, this epistemological crisis is not insurmountable. By breaking with the notion that transparency

is exclusively a technological problem, this research argues that memory institutions can achieve accountability by reclaiming and adapting their own rich documentary traditions.

The move from classical provenance to dynamic paradata represents a necessary development in KO methodology. While classical provenance remains a sine qua non for authenticating digital lineages and maintaining the archival bond, paradata offers the critical processual documentation necessary to shed light on the algorithmic black box. This suggests that practitioners may benefit more from workflow documentation than from direct access to source code when identifying biases, assessing system limitations, and promoting trust.

AI Usage Statement: No artificial intelligence tools were used in the preparation of this manuscript.

References

Arias Hernández, Richard, and Moisés Rockembach. 2025. “Building Trustworthy AI Solutions: Integrating Artificial Intelligence Literacy into Records Management and Archival Systems.” AI & SOCIETY 40: 4265–4282. https://doi.org/10.1007/s00146-025-02194-0.

Cameron, Scott, Patricia C. Franks, and Babak Hamidzadeh. 2023. “Positioning Paradata: A Conceptual Frame for AI Processual Documentation in Archives and Recordkeeping Contexts.” Journal on Computing and Cultural Heritage 16, no. 4: 75:1–75:19. https://doi.org/10.1145/3594728.

Collomosse, John, Andy Parsons, and Mike Potel. 2024. “To Authenticity, and Beyond! Building Safe and Fair Generative AI upon the Three Pillars of Provenance.” IEEE Computer Graphics and Applications 44, no. 3: 82–90. https://doi.org/10.1109/MCG.2024.3380168.

Davet, Jeremy, Babak Hamidzadeh, Patricia Franks, and Jenny Bunn. 2022. “Tracking the Functions of AI as Paradata & Pursuing Archival Accountability.” In Archiving 2022: Expanding Connections Across Digital Cultural Heritage – Final Program and Proceedings, 83–88. https://doi.org/10.2352/issn.2168-3204.2022.19.1.17.

Jaillant, Lise, and Arran Rees. 2023. “Applying AI to Digital Archives: Trust, Collaboration and Shared Professional Ethics.” Digital Scholarship in the Humanities 38: 571–585. https://doi.org/10.1093/llc/fqac073.

Longpre, Shayne, Robert Mahari, Naana Obeng-Marnu, William Brannon, Tobin South, Katy Gero, Sandy Pentland, and Jad Kabbara. 2024. “Data Authenticity, Consent, & Provenance for AI Are All Broken: What Will It Take to Fix Them?” arXiv:2404.12691. https://doi.org/10.48550/arXiv.2404.12691.

MacNeil, Heather. 2004. “Contemporary Archival Diplomatics as a Method of Inquiry: Lessons Learned from Two Research Projects.” Archival Science 4: 199–232. https://doi.org/10.1007/s10502-005-2592-x.

Melo, Jonas Ferrigolo, and Moisés Rockembach. 2019. “Arquivologia e Ciência da Informação na Era do Big Data: Perspectivas de Pesquisa e Atuação Profissional em Arquivos Digitais.” PRISMA.COM 39: 14–28. https://doi.org/10.21747/16463153/39a2.

Melo, Jonas Ferrigolo, and Moisés Rockembach. 2023. “Digital Humanities Curation to the Cultural Heritage: Building Competencies from AI and Digital Literacy.” Paper presented at the 10th DHBenelux Conference, Brussels. Zenodo. https://doi.org/10.5281/zenodo.7976087.

Rockembach, Moisés. 2021. “Ciência da informação e inteligência artificial: um caminho para arquivos e bibliotecas inteligentes.” In Organização do Conhecimento no Horizonte 2030: Desenvolvimento Sustentável e Saúde: Atas do V Congresso ISKO Espanha-Portugal, edited by Carlos Guardado da Silva, Jorge Revez, and Luís Corujo, 235–242. Lisboa: Centro de Estudos Clássicos and Edições Colibri.

Rockembach, Moisés. 2024. “Ciência aberta e inteligência artificial: desafios éticos e transparência em modelos generativos.” Ciência da Informação 53, no. 3: 165–170. https://doi.org/10.18225/ci.inf.v53i3.7227.

Shang, Nelson, Remi Prospero Fonka, and Jumbam Gideon. 2025. “The Epistemology of Artificial Intelligence: Understanding Knowledge Creation and Validation in the Digital Age.” Manuscripts on the Artificial Intelligence and Digital Research 2, no. 4: 39–52.

Shiozaki, Ryo. 2025. “Ethics of Knowledge Organization in the Algorithmic Age: Utilitarian Coordination and Constraining Frameworks.” Knowledge Organization 52, no. 6: 44431. https://doi.org/10.31083/KO44431.

Sousa, Nuno, and Moisés Rockembach. 2025. “Chatbots na Ciência da Informação: aplicações em arquivos e bibliotecas.” Páginas a&b: Arquivos e Bibliotecas, 3rd ser., no. 23: 314–330. https://doi.org/10.21747/21836671/pag23a18.

Tognoli, Natalia, and José Augusto Chaves Guimarães. 2019. “Provenance as a Knowledge Organization Principle.” Knowledge Organization 46: 558–568. https://doi.org/10.5771/0943-7444-2019-7-558.