Brazilian Archives in InterPARES Trust AI: Towards Disruptive Technologies Supporting Twenty-First-Century Brazilian Archivists

Cláudio Gottschalg Duque
Faculty of Information Science, University of Brasília (UnB), Brasília, Brazil
R.E.G.I.I.M.E.N.T.O. Research Group · ORCID: 0000-0003-3558-466X

Abstract

Digital transformation has reshaped the production, management, preservation, and authentication of Brazilian archival records. The Sistema Eletrônico de Informações (SEI), now an official federal solution for electronic administrative records and processes, exemplifies this transformation. This chapter critically synthesizes two decades of Brazilian research related to three directions connected to the broader agenda of InterPARES Trust AI: artificial intelligence (AI), natural language processing (NLP), and ontologies for record classification; blockchain and smart contracts as complementary mechanisms for authenticity, integrity, and transaction verification; and AI-assisted organization of cultural and linguistic heritage. Through a narrative, citation-based review, the analysis distinguishes empirical findings, conceptual architectures, and prospective applications. The evidence indicates that AI can support human-validated classification and semantic representation, while blockchain can provide tamper-evident verification. Neither replaces archival appraisal, provenance, retention, or preservation governance. Trustworthy innovation therefore depends on archivists participating in system design, validation, and accountability.

Keywords: *Artificial Intelligence; Archives; Blockchain; InterPARES Trust AI; SEI; Natural Language Processing; Ontologies; Digital Preservation; Brazil.*

1. Introduction

Digital transformation has changed how public administration produces, manages, retrieves, and authenticates records. In Brazil, this change is particularly visible in the Sistema Eletrônico de Informações (SEI), developed by the Federal Regional Court of the 4th Region and made available to public institutions. The Brazilian Ministry of Management and Innovation in Public Services identifies SEI as the Federal Government's official solution for producing and managing electronic administrative records and processes (Brazil, Ministry of Management and Innovation in Public Services, n.d.). Its adoption improves efficiency and access, but it does not eliminate the archival challenges created by the scale, heterogeneity, and long-term preservation requirements of born-digital records.

These conditions make AI, NLP, ontology engineering, and distributed ledger technologies (DLTs), particularly blockchain, increasingly relevant to archival research. Their relevance,

however, cannot be assessed solely in terms of computational performance. Systems acting upon records must also be evaluated against provenance, context, authenticity, integrity, appraisal, retention, and accountability—requirements that predate and outlast any specific technology.

This chapter aims to critically synthesize a selected Brazilian research trajectory from 2005 to 2026 and relate it to Brazilian participation in InterPARES Trust AI. It adopts a narrative review of the cited studies and official project records rather than a systematic review. Accordingly, the analysis distinguishes three levels of evidence: empirical findings, conceptual or architectural contributions, and prospective research directions — a classification made explicit, study by study, in the synthesis table in Section 9.

The discussion follows three principal directions: AI, NLP, and ontologies for classification; blockchain and smart contracts for verification and transaction governance; and AI for cultural and linguistic heritage. A comparative extension to libraries situates these developments within Information Science, while the final sections examine the Brazilian contribution to InterPARES Trust AI and the archivist's role as a designer and auditor of documentary trust.

2. From Computational Linguistics and Ontologies to Archival AI

The intellectual foundations of this trajectory precede the current expansion of generative AI. Duque's (2005) doctoral research, SiRILiCO, investigated the indexing of electronic documents through computational linguistics and ontological representation. It applied syntactic and semantic analysis to Portuguese-language texts to derive labels supporting a lightweight, automatically generated ontology. This work provides a methodological antecedent for a problem now central to archival AI: representing relationships among concepts rather than merely extracting tokens.

Archival classification is not reducible to keyword extraction. A record's evidentiary meaning depends on relationships among the record, its creator, institutional functions, related records, and the context of production. Classical information-retrieval systems, designed primarily for topical relevance, do not necessarily preserve these relationships. AI-assisted archival classification can therefore combine statistical and linguistic methods with explicit knowledge structures: NLP identifies entities, subjects, and document types; ontologies represent the domain in which they acquire meaning; and machine learning identifies patterns across previously classified records.

The resulting architecture is hybrid by design: document → linguistic processing → semantic representation → ontology → recommended archival classification → human validation. The sequence positions the archivist inside, rather than outside, the intelligent system.

3. AI-Assisted Classification in Brazilian Government Records

SEI is a strategic environment for investigating this architecture because it supports the production and management of substantial volumes of born-digital administrative records. As production grows, exclusively manual classification can become costly and inconsistent. Within the broader Brazilian research trajectory examined here, this chapter proposes applying AI, NLP, and ontology engineering to assist document classification in SEI environments, including governmental contexts in the Federal District.

The proposed model is not fully autonomous. It processes document content and metadata, identifies entities and document types, relates them to classification schemes and domain ontologies, and produces recommendations accompanied by confidence information and an

explanation. The archivist validates, rejects, or corrects each recommendation. Validated corrections can subsequently inform model improvement under controlled procedures.

This human-in-the-loop architecture represents a conceptual and technical direction, not evidence of a validated production system. Its professional implication is nevertheless clear: archivists move from being exclusively manual classifiers to acting as knowledge engineers, curators, and auditors of intelligent recordkeeping systems. Without classification plans, provenance relationships, and retention requirements, computationally plausible classifications may remain archivally unfounded.

4. Blockchain, Authenticity, and Archival Trust

Digital records do not rely on the same visible physical cues—such as paper, ink, and continuous physical custody—by which authenticity was traditionally inferred. Their authenticity and integrity depend on metadata, digital signatures, audit trails, preservation controls, and, in some architectures, cryptographic verification mechanisms.

Mori et al. (2020) compared blockchain characteristics with requirements for trustworthy archival digital repositories (RDC-Arq), drawing on archivists' perceptions of authenticity, integrity, security, and chain of custody. The study identified potential affinities between blockchain properties and selected RDC-Arq requirements, while emphasizing the importance of following archival guidelines. Because its empirical component concerns professional perceptions and comparative analysis, the study supports a claim of potential compatibility rather than proof of operational preservation effectiveness.

Within this bounded role, blockchain can register hashes or other cryptographic evidence of documentary events while the original record remains in a governed preservation environment. The ledger then operates as a tamper-evident verification layer. It does not perform appraisal, classification, retention, or long-term preservation and should not be treated as an archive or as a substitute for an RDC-Arq.

This layered approach is relevant to electronically signed administrative records because authenticity cannot be reduced to the persistence of a digital signature. Chain of custody, provenance, institutional controls, and preservation policy remain indispensable. Duque and Alves (2020), examining archival policies for digital preservation in Federal District institutions, similarly emphasize long-term accessibility and the maintenance of unaltered records—functions that a verification layer may complement but cannot replace.

5. Smart Contracts and Electronic Health Records

The intersection of archival science and blockchain is especially relevant to electronic health records, which combine documentary evidence, sensitive personal data, multiple institutional actors, access-control requirements, and long retention periods. Xavier and Duque (2021) examined this domain through a systematic literature search and an exploratory, descriptive, and qualitative approach. The study identified gaps in legislation and scholarly production and considered possible technological contributions to record governance; it did not report a live implementation of blockchain or smart contracts in a healthcare institution.

The appropriate conclusion is therefore conditional. Smart contracts may formalize transactional rules, such as access permissions and consent conditions, while blockchain may

strengthen traceability and integrity verification. Neither mechanism determines archival value, establishes retention requirements, or replaces documentary governance. Their archival relevance depends on requirements defined and supervised by records professionals.

6. Prospective Direction: AI, Archives, and Cultural Memory

Beyond administrative records, this chapter proposes a prospective research direction for AI-assisted organization of cultural and linguistic heritage. Brazil's documentary heritage reflects Indigenous, African, European, and Asian traditions. Pomeranian-immigrant communities provide one possible setting in which languages, oral traditions, family records, and local knowledge could be organized and preserved through participatory digital initiatives. This example is presented as a research agenda, not as a validated system or completed case study.

Potential applications include NLP-assisted transcription, speech technologies for oral-history collections, computer vision for photographic description, and ontologies relating people, places, events, and documentary contexts. These possibilities require community-based validation, transparent documentation of model limitations, and safeguards against the erasure or homogenization of cultural meaning. Human and community interpretation is not an external correction to the system; it is a condition of representational legitimacy.

7. Comparative Extension to Libraries

Related research in libraries provides a comparative extension rather than a fourth archival application. Alentejo and Duque (2021) proposed an information architecture for a Brazilian union catalogue of libraries and information services on the Web. Flor (2026) subsequently developed a conceptual proposal combining blockchain and radio-frequency identification for the management and preservation of physical library collections.

These studies do not make archives and libraries institutionally interchangeable. They illustrate technological convergence at the infrastructure level: both domains address identification, provenance, access, integrity, and preservation, although they apply different professional principles and institutional responsibilities. This comparison contextualizes the broader Information Science trajectory but does not constitute evidence for the operational effectiveness of the archival applications examined above.

8. InterPARES Trust AI and the Brazilian Contribution

Brazilian participation in InterPARES Trust AI places these research directions within an international investigation of AI and trustworthy records. The Brazilian contribution includes the study “Inova HFA—Using AI to Better Manage Health Records and Patient Outcomes” (CU06), led at UnB by Cláudio Gottschalg-Duque. The official project record gives the study a timeline from September 2021 to December 2023 and defines an agenda involving AI, NLP, DLT, smart contracts, health-record management, patient outcomes, privacy, and compliance with Brazil's General Personal Data Protection Law (InterPARES Trust AI, n.d.).

A documented output associated with this agenda is the study by Kuroki Júnior and Gottschalg-Duque (2024), published shortly after the formal project period, which situates Multimodal Information Architecture within the socio-technical development of AI for the Inova HFA electronic health-record context. This publication provides a more direct connection between the

Brazilian InterPARES participation and the AI architecture discussed in this chapter. As of this writing, no further InterPARES Trust AI project record documents a continuation of Inova HFA beyond this publication; the study is treated here as concluded, with its findings feeding forward into the broader Brazilian research trajectory rather than into an active, ongoing project.

The earlier works reviewed here—particularly Duque (2005), Mori et al. (2020), and Xavier and Duque (2021)—should be understood as intellectual and methodological antecedents informing the Brazilian contribution, not automatically as outputs produced within InterPARES Trust AI. This distinction clarifies the chronology and prevents the broader Brazilian research trajectory from being conflated with the project's formal results.

9. Discussion: The Archivist as Designer of Trust

A central implication of these technologies concerns professional identity. The twenty-first-century archivist works at the intersection of archival theory, information architecture, data governance, and AI—not because every archivist must become a programmer, but because archivists must participate in the specification, governance, and validation of systems acting upon records.

In AI-assisted classification, archivists validate semantic and functional relationships; in blockchain architectures, they determine which documentary events require cryptographic evidence; and in cultural-memory environments, archivists and communities protect context and meaning from reductive algorithmic representation. Across these settings, the archivist mediates between computational capability and documentary trust as a condition of archival legitimacy.

The evidence reviewed is heterogeneous. It includes professional perceptions, literature-based analyses, conceptual architectures, and prospective applications, but it does not provide equivalent operational validation for every direction. Because the chapter adopts a selected narrative review centered on one Brazilian research trajectory, its conclusions should not be generalized to all Brazilian archival institutions. Future studies should report implementation context, classification accuracy, error patterns, bias, explainability, privacy protection, preservation continuity, and the effects of archivist and community oversight. Table 1 summarizes the evidentiary status of each research direction discussed.

Table 1. Evidentiary status of the Brazilian research directions reviewed

Research direction	Key citation(s)	Evidentiary tier	What is, and is not, established
AI/NLP-assisted classification (SEI)	Duque (2005); this chapter's SEI proposal	Conceptual / architectural	Methodological foundation (linguistic + ontological indexing) is established; the SEI human-in-the-loop model is a proposed architecture, not a deployed or evaluated system.
Blockchain vs. RDC-Arq requirements	Mori et al. (2020)	Empirical (perception-based)	Archivists' perceptions and comparative analysis indicate potential structural affinity; this does not constitute proof of operational preservation effectiveness.
Blockchain / Smart Contracts in health records	Xavier and Duque (2021)	Conceptual (literature-based)	Systematic literature review identifying gaps and possible contributions; no live implementation in a healthcare institution is reported.
Digital preservation policy (Federal District)	Duque and Alves (2020)	Empirical (policy analysis)	Documents governance and accessibility requirements in actual DF institutions; establishes that verification layers presuppose, but cannot supply, preservation policy.
AI for cultural and	Chapter proposal	Prospective	Presented explicitly as a research agenda; no

Research direction	Key citation(s)	Evidentiary tier	What is, and is not, established
linguistic heritage	(Pomeranian heritage example)		validated system or case study exists yet.
Semantic architecture for library catalogues	Alentejo and Duque (2021)	Conceptual / architectural	Proposed information architecture for a national union catalogue; comparative, not archival, evidence.
Blockchain + RFID for physical library collections	Flor (2026)	Conceptual	Conceptual proposal; not an operational deployment.
AI-assisted health-record architecture (Inova HFA)	Kuroki Júnior and Gottschalg-Duque (2024)	Conceptual / architectural	Documented, published output of the Brazilian InterPARES Trust AI contribution; situates Multimodal Information Architecture within Inova HFA's socio-technical design, concluded in scope with the project's formal period.

10. Conclusion: Contributions, Limits, and Research Directions

This chapter examined three directions in Brazilian archival research: AI, NLP, and ontologies for record classification; blockchain and smart contracts as complementary mechanisms for authenticity, integrity, and transaction verification; and AI-assisted organization of cultural and linguistic heritage. Taken together, the literature supports a clear but qualified conclusion: these technologies can strengthen specific archival functions only when they are designed and evaluated within requirements of provenance, context, appraisal, retention, preservation, and accountability.

In AI-assisted classification, Duque's (2005) SiRILiCO provides an early methodological foundation for combining computational linguistics with ontological representation. The model proposed for SEI environments extends this foundation into a human-in-the-loop architecture in which AI recommends classifications and archivists validate or correct them. Because this application has not been presented here as a validated production system, its contribution should be understood as conceptual and architectural rather than as evidence of operational effectiveness.

The blockchain studies indicate a similarly bounded contribution. Mori et al. (2020) identified, through archivists' perceptions and comparison with RDC-Arq requirements, the technology's potential to support integrity and authenticity verification. Xavier and Duque (2021), through exploratory qualitative analysis, identified possible uses of blockchain and smart contracts in electronic health-record governance. Neither study demonstrates that these technologies can replace archival custody, appraisal, classification, retention decisions, or long-term preservation. Their most defensible role is that of a complementary verification and transaction-governance layer.

In cultural and linguistic heritage, the applications discussed remain prospective. NLP, speech technologies, computer vision, and ontologies may assist transcription, description, and semantic organization, but their use requires validation by archivists and the communities represented in the records. Brazilian participation in InterPARES Trust AI connects these questions to an international agenda of trustworthy AI, while documented outputs such as the Inova HFA publication provide a basis for further technical and institutional evaluation.

The evidence reviewed therefore defines both possibilities and limits: AI can assist semantic and classificatory work, and blockchain can support tamper-evident verification, but neither can autonomously establish archival value or evidentiary context. The twenty-first-century archivist is not displaced by disruptive technologies; the archivist becomes essential to their design, validation, governance, and accountability.

References

Alentejo, E. S., & Duque, C. G. (2021). Arquitetura do Catálogo Coletivo Brasileiro de Bibliotecas e Serviços de Informação na Web: Uma proposta. Revista Ibero-Americana de Ciência da Informação, 14(3), 764–785. https://doi.org/10.26512/rici.v14.n3.2021.36559

Brazil, Ministry of Management and Innovation in Public Services. (n.d.). O que é o SEI? Retrieved August 25, 2026, from https://www.gov.br/gestao/pt-br/assuntos/gestaoeinovacao/informacoes-sistemas-e-servicos-de-gestao/processo-eletronico-nacional/conteudo/SEI

Duque, C. G. (2005). SiRILiCO: Uma proposta para um Sistema de Recuperação de Informação baseado em Teorias da Linguística Computacional e Ontologia [Doctoral dissertation, Universidade Federal de Minas Gerais].

Duque, C. G., & Alves, F. M. (2020). Políticas arquivísticas aplicadas à preservação de documentos digitais no Distrito Federal. Revista Brasileira de Biblioteconomia e Documentação, 16, 1–16.

Flor, A. R. B. (2026). Um estudo da blockchain e sua contribuição para a gestão e preservação de acervos em bibliotecas físicas [Master's dissertation, Universidade de Brasília]. https://repositorio.unb.br/handle/10482/55246

InterPARES Trust AI. (n.d.). Inova HFA—Using AI to better manage health records and patient outcomes (CU06). Retrieved August 25, 2026, from https://interparestrustai.org/trust/about_research/studies

Kuroki Júnior, G. H., & Gottschalg-Duque, C. (2024). Towards Multimodal Information Architecture: A social-technical approach of artificial intelligence development in Inova HFA's Electronic Health Record (EHR). In K. Arai (Ed.), Intelligent Systems and Applications (Lecture Notes in Networks and Systems, Vol. 825, pp. 792–806). Springer. https://doi.org/10.1007/978-3-031-47718-8_51

Mori, A., Duque, C. G., Orlandi, T. R. C., & Pequeno, W. de O. (2020). Análise comparativa entre os requisitos do RDC-Arq e a tecnologia blockchain: Uma perspectiva de profissionais arquivistas. Ibersid: Revista de Sistemas de Información y Documentación, 14(1), 29–40. https://doi.org/10.54886/ibersid.v14i1.4651

Xavier, A. C. C., & Duque, C. G. (2021). Prontuário eletrônico do paciente: Qual a contribuição da arquivística e do Smart Contracts para a sua gestão na Era da Saúde 4.0? AtoZ: Novas Práticas em Informação e Conhecimento, 10(3), 1–10. https://doi.org/10.5380/atoz.v10i3.81267