Author accepted version of Cameron, S., Franks, P., Huvila, I., & Mooradian, N. (in print). Navigating Accountability: The Role of Paradata in AI Documentation and Governance. *Journal of Documentation*. DOI:10.1108/JD-01-2025-0009 (Accepted 19-March-2025)

Navigating Accountability: The Role of Paradata in AI Documentation and Governance

Abstract

Purpose: The increased use of Artificial Intelligence has prompted governments internationally to provide guidance and legislation to maximize the benefits of AI while minimizing the risks to humans and organizations. This paper explores how published requirements for documentation in a sampling of authoritative texts address the challenges of creating, capturing, and preserving records of the design, implementation, and use of AI tools for accountability and transparency, and how the analytical concept of paradata can help to meet the recordkeeping challenges presented by the design, development and implementation of AI systems.

Design/methodology/approach: Inductive reading and conceptual analysis of a set of AI laws, regulations and frameworks published by the EU, UK, US, Canada, and Singapore.

Findings: The authoritative texts reviewed clearly describe activities which imply the necessity of records creation and preservation. Identifying specific documents necessary to comprise a sufficient body of records to provide evidence of accountable AI implementation and operation can be difficult. Literature on paradata in archival applications of AI may prove productive in identifying relevant information artifacts for preservation in the AI process. Paradata is produced by those designing and implementing AI systems and by AI systems themselves.

Originality: No comparable comparative analyses have been published in the archives and information field.

Practical implications: Identifying relevant paradata produced by AI systems requires archivists to develop both the capacity to analyze and the vocabulary to discuss these systems in order to preserve evidence of their operation in compliance with legislation and international standards.

Introduction

The increased use of Artificial Intelligence has prompted governments internationally to provide guidance and legislation to maximize the benefits of AI while minimizing risks to humans and organizations. A major challenge arises from the need to provide evidence (e.g., a civil proceeding or an AI audit) that an AI tool or technique was designed, developed, and implemented in an ethical and responsible manner (Trace & Hodges, 2024). Similarly, evidence is necessary for critical understanding of the workings of AI and assessment of the credibility and reliability of information processed and produced. The extent of documentation necessary and available depends upon factors including the level of risk involved, the type of AI employed, and how it was designed, developed, and implemented. As emergent AI governance frameworks and other authoritative texts focus significantly on risk, the scope of necessary documentation will likely expand as risk increases.

Kommenterad [1]: Adding something like this could make our text more relatable for the readership I guess? Feel free to revise. It shifts a little the argumentation but hopefully not too much.

Kommenterad [2R1]: Looks good to me.

The topic of banning unacceptable uses of AI that may threaten safety, livelihood, and individual rights is being debated by governments worldwide. The use of AI that presents no or minimal risk, such as video games and spam filters, will be lightly governed in comparison. That leaves a wide area of AI development and use potentially harmful to individuals, organizations, and society that should be documented as part of an AI governance strategy.

A common objective in AI governance is to foster transparency and accountability through the creation of documentation and records addressing the complete AI development lifecycle, including planning, development, implementation, operation and use of AI systems. The creation of reliable records of the AI lifecycle is an area of active interest and research in the archives and records fields. Both fields support the objectives of transparency and accountability by creating reliable, authentic records using a lifecycle concept and a set of professional practices. Archival scholars are continually exploring new concepts and methods to extend and evolve archival practices and meet the challenges of trustworthy AI implementations and recordkeeping. One such concept is paradata, i.e. "information about the procedure(s) and tools used to create and process information resources, along with information about the persons carrying out those procedures" (Cameron et al, 2023). While the archival field has well-established definitions of reliability and authenticity as elements of a record's trustworthiness across diverse social and technological contexts, the question of what constitutes trustworthy recordkeeping for AI applications demands the development of new vocabularies and competencies on the parts of records and information professionals (Duranti, 1995).

Kommenterad [3]: I added the highlighted text.

Kommenterad [4R3]: I like this addition!

This paper analyzes the requirements for documentation specified in a sample of authoritative texts (e.g., AI laws, regulations, and frameworks) published by the EU, UK, US, Canada, and Singapore. So far no comparable analyses have been published in the archival field for AI. However, a recent study of Generative AI as addressed in these and similar documents explores AI governance in general (Luna, Tan, Xie, Jiang 2024). Further, studies of individual legal frameworks have emerged (e.g., Smuha et al., 2024; Ruschemeier & Bareis, 2024).

This paper addresses the following research questions across five authoritative texts:

- RQ1: In what ways do these instruments address the challenges of documenting the design, implementation, and use of AI tools, and the preservation of records to ensure accountability and transparency?
- RQ2: What prescriptions and provisions for record creation, capture and preservation are included in the examined authoritative texts?
- RQ3: How does paradata relate to the information artifacts identified in RQ2? Can the concept of paradata help to meet the recordkeeping challenges presented by the design, development and implementation of AI systems?

Related Work

In 2021, a team of InterPARES Trust AI researchers began to investigate the problem of documentation of AI processes beyond the traditional records and archival practices of retaining the output (records and archival materials) and metadata (data about the output) by turning attention to the concept of paradata (data about the process).

Origins and use of Paradata

The term “paradata” is attributed to sociologist Mick P. Couper, who in 1998 coined the term to describe data created as a byproduct of automated systems employed during the research process (Couper 1998, pp 41-49). Since then paradata has been applied in fields including cultural heritage, social science research, and archeology (Huvila, 2022). Statistical social science researchers view paradata as “data which fall outside the intentionally or purposefully collected data” but which nevertheless provide insight into the origins of a dataset (O’Connor and Goodwin 2020). In the heritage visualization domain, paradata is used to document “the decisions, selection processes and reasoning behind the interaction and combination of different data artifacts” (Baker, 2007). In the context of research data documentation, a recent study, the CAPTURE project, investigates what information about the creation and use of archeological and cultural heritage research data (paradata) is needed to make the data reusable in the future (Huvila & Ekman, 2024).

Paradata and the AI Process

The term paradata in relation to documenting the AI process in archival applications is a novel concept introduced by the InterPARES team. Davet, Hamidzadeh, Franks, and Bunn (2022) see the notion as critical for maintaining archival accountability to the public. Davet, Hamidzadeh, Franks (2023, pp. 277) write that “the archival definition of paradata is a reflection not only of a particular preoccupation with the functioning of automated systems within archives but also as an extension of evolving ideas around archival transparency, accountability, and record integrity”. Ongoing InterPARES research has applied the concept to describe the recordkeeping needs of organizations applying less-than-transparent AI tools in accountable contexts (Franks, 2023; Cameron et al., 2023).

AI and Recordkeeping

AI has emerged as a challenge for the records field in recent years, although the role of legislation in prescribing records needs has not yet been considered in depth within the literature. As early as 2019, Norman Mooradian (2019, pp. 13), advised that “defining an AI record and developing methods for capturing AI records is a project the [recordkeeping] profession should take on.” In particular, Mooradian was concerned with the creation of records that adequately document decisions impacting the interests of individuals.

In 2020, Jenny Bunn of The National Archives (UK) provided a recordkeeping view of XAI, concluding that documentation of the AI process is necessary to respond to questions including “What records are created within AI research teams to document their process? What records are created of the decisions to procure or deploy systems utilizing AI? What records are created of the decisions and impact of such systems?” (Bunn, 2020).

Records will be created to meet various needs. For example, Machine Learning (ML) developers may capture and preserve documentation to describe the actions and decisions made throughout the ML Lifecycle. The data captured would include artifacts of the ML process such as data and feature preparation, model development, and model operation tasks (Schlegel & Sattler, 2022).

The Public Record Office of Victoria (2024) specifies recordkeeping policies in relation to AI technologies. As the Office notes, "associated with those records of business, it is vital to also have records of the technologies and processes that produced them." In addition to outlining categories of AI and AI risk, they detail records that can be captured and preserved to illustrate risk mitigation strategies taken. For example, records documenting steps to reduce Human Rights Risks/Harm due to Bias (e.g., race, gender, economic, age, etc.) may include approval process decisions; notification that AI is being used,; and records of monitoring and addressing risks/areas of bias. (Public Records Office of Victoria, 2024). Detailed AI policies such as this will need to be developed in compliance with emergent AI legislation.

In the earlier work, it has been noted that AI regulation may vary, entailing technical requirements, moral requirements or de facto obligations (Veale et al., 2023). Some legislation specifies requirements (the what) but leaves the types of documentation (the how) to organizations employing AI technology.

To address the current lack of comparative overviews of AI legislation and guidance documents, this study conducted a review of AI legislation and guidance publications from the U.S. Canada, UK, Singapore, and the EU. The results of that review comprise the remainder of this article.

Research methodology

The frameworks and legislation selected for the present study (Table 1) represent current and prominent examples of authoritative texts in both the public and private realms. TThe resulting set of analyzed documents is a convenience sample, although it intends to provide a purposive selection of authoritative texts reflecting prominent approaches to AI governance in the Western world. Since this study is interested primarily in current AI legislation and published guidance, it has not engaged substantially with emergent legislation and governance schemes which explicitly target recent developments in generative AI tools. As the frameworks discussed in this paper are broad, generative AI will certainly fall within their scope, even though the frameworks were substantially devised or implemented before the recent prominence attained by the likes of OpenAI's ChatGPT.

Table 1. List of AI-related regulations, guidelines, and recommendations reviewed.

Authoritative texts reviewed	**Issuing body**	**Year**	**Status**	**Scope**	**Sector**	**Governance strategy**	**Character**
Artificial Intelligence Act AIACT–EU (European Union, 2024)	European Union	2024	Law	Supranational	Public/ Private	Legislation / Regulation	Risk-based with some rights-based elements
A pro-innovation approach to AI regulation PIAAIR–UK (Office for Artificial Intelligence, 2023)	Government of the United Kingdom	2023	Policy paper	National/federal	Private	Voluntary self-governance; co-operative regulation (Executive Summary 16)	Principles-based. Principles include: pro-innovation, proportionate, trustworthy, adaptable, clear, collaborative (36-37)

Model AI Governance Framework (2nd ed) MF–SG (IMDA and PDPC, Government of Singapore, 2020)	Singapore	2020	Voluntary framework	Unbounded scope; aiming for international compliance/interoperability	Unbounded	Voluntary self-governance	Principles based: AI intended to be explainable, transparent, fair, and human centric (i.e. safe and effective)
AI Risk Management Framework AI RMF–NIST (NIST, 2023a)	National Institute of Standards and Technology (United States)	2023	Voluntary framework	Unbounded; effectively federal	Unbounded	Voluntary self-governance	Risk-based
Artificial Intelligence and Data Act (Bill C-27 Digital Charter Implementation Act, Part III) AIDA–CA (House of Commons of Canada, 2022)	Government of Canada	2022	Proposed law	Federal	Private	Legislation/Regulation	Risk-based
Directive on Automated Decision Making DADM–CA (Government of Canada, 2019	Government of Canada	2019	Law	Federal	Public	Regulation	Risk-based with some rights-based elements

We employed inductive reasoning when reviewing the authoritative texts to move comparative observations into analysis. Table 1 categorizes the texts according to a variety of facets—their status, whether they apply to public or private sectors, and their governance strategies. While all the authoritative texts incorporated risk-based elements, the broader frameworks identified explicitly as 'principles-based' rather than risk-based were labeled as such.

Conceptual analysis

Throughout our analysis, the terminology used to describe information artifacts identified for preservation was flagged and key terms were noted, defined and sourced as in Table 2. These provided standard definitions to ground the comparisons of the terms used throughout the texts reviewed. The authoritative texts examined in this paper do not take the perspective of records professionals, and accordingly do not always use the terms as defined below.

Table 2: Definitions of key terms.

Term	Definition	References
Artificial Intelligence	A machine-based or engineered system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.	Russell, et al., 2023
Paradata	Information about the procedure(s) and tools used to create and process information resources, along with information about the persons carrying out those procedures.	InterPARES Terminology Database (IPTD). "Paradata." https://interparestrustai.org/terminology/term/paradata/en
Documentation (related term: technical documentation; document)	Information created or gathered in an intentional, formal way, that is meant to explain the origins and [or] operations of systems, programs, processes so that these can be operated or carried out, maintained, evaluated in relation to their intended functions, configurations, purposes, requirements, etc.	See IPTD. "Record" and "Document" as below.
Evidence	n. ~ 1. IP2 · All the means by which any alleged matter of fact, the truth of which is submitted to investigation, is established or disproved. – 2. Information, in any form, used to establish or disprove the validity of an assertion or fact.	IPTD. "Evidence" https://interparestrustai.org/terminology/term/evidence/en
Record	n. ~ 1. IP2 · A document made or received in the course of a practical activity as an instrument or a by-product of such activity, and set aside for action or reference. Syn.: archival document."	IPTD. "Record." http://interparestrust.org/terminology/term/record
Document	n. ~ 1. Archives, IP2 · An indivisible unit of information constituted by a message affixed to a medium (recorded) in a stable syntactic manner. A document has fixed form and stable content.	IPTD. "Document." http://interparestrust.org/termin

		ology/term/document/en

For InterPARES, documents are characterized by their indivisibility and contain information affixed to a physical or digital medium. Records are documents saved for use or reference in the future. They possess the characteristics of authenticity, reliability, integrity, and usability (Duranti & Thibodeau, 2006).

The term ‘documentation’ is less frequently used in the archival profession. However, “documentation” emerged as the most consistently used term to encompass a variety of the necessary information artifacts identified for preservation. The definition in Table 2 is descriptive rather than prescriptive or normative. As distinct from records or documents, documentation is created or preserved for a clear explanatory purpose and may include records as well as instruction manuals or technical reports used to explain the operations of a system.

The term paradata describes the information needed for archivists to document the AI process. Paradata entails both human and computerized processes related to the creation and curation of information resources. In archives, paradata is envisioned as addressing the transparency and accountability problems introduced by complex machine learning tools into archival practice (Davet et al., 2023). Since this encompasses both the human-led management and implementation of these tools and the machine-generated information about their use, in the contexts described by the texts in this paper, paradata frequently includes records, documentation, and other forms of information and data explicitly or implicitly called for. As information which documents processes of creation and management of other information resources, paradata is an encompassing term. While the term does not appear in AI legislation, this paper evaluates its relevance in relation to the documents analyzed.

The following sections provide an analysis of the six authoritative texts reviewed. Each section includes information on the purpose, scope, and nature of each AI governance document, followed by an analysis of the documents prescribed records practices and requirements. The comparative results of the study are also summarized in the table found in Supplementary Materials.

Results: Authoritative texts reviewed

A pro-innovation approach to AI regulation (PIAAIR-UK)

Overview

The UK policy paper “A pro-innovation approach to AI regulation” describes a general outline for the development of UK AI policy. It is empathetically a policy outline rather than a set of guidelines. AI is considered to have transformative power in “all areas of life” and to “stimulate the UK economy” and opportunities to deliver “societal benefits, from medical advances to mitigating climate change”.

Stated purpose

The policy outlined in PIAAIR-UK aims to “bring clarity and coherence to the AI regulatory landscape” with a goal of “to make responsible innovation easier”, to “strengthen the UK’s position as a global leader in AI, harness AI’s ability to drive growth and prosperity, and increase public trust in its use and application.”

Scope

The general approach is described as “pro-innovation” with the focus on enabling the AI industry to flourish and “principles-based” to allow “the framework to be agile and proportionate”. The policy paper refers to an urgency to regulate AI as other legislations have started to do the same. The risks are acknowledged especially from the perspective of the negative impact they may have on public trust in AI, which would have negative consequences for AI companies.

Key information artifacts identified for preservation

PIAAIR-UK makes no detailed stipulations about specific information artifacts but instead refers to “documentation” and “information” as quasi-artifacts that are suggested to be provided, to exist and be made accessible, also noting that “[c]onsumers, users, and regulators” require different information.

Concept of record, document(ation) and paradata in the document

The main concept used in the policy paper is documentation. The concept of record is not used explicitly, although keeping the expected documentation implies the necessity of record-keeping, e.g. “how failures and near misses can be recorded and used to inform good practice”. While the notion of paradata is absent, the paper discusses an “information” provision in the principle of “[a]ppropriate transparency and explainability” suggesting further that “[t]ransparency refers to the communication of appropriate information about an AI system to relevant people (for example, information on how, when, and for which purposes an AI system is being used)”. The policy anticipates that AI life-cycle actors need to provide paradata-like information on “the nature and purpose of AI” including information relating to specifics outcome, data and training data and “logic and process used” to support explainability of decision-making, outcomes and accountability.

Artificial Intelligence Act (AIACT-EU)

Overview

In 2024, the European Union enacted harmonized rules on AI (Artificial Intelligence Act) in a law regulating the “placing on the market, the putting into service and the use of AI systems in the” European Union (European Union, 2024b, hereafter AIACT-EU). The regulation follows a proportionate risk-based approach to ensure that the systems are safe and respect fundamental rights and the values of the EU. AIACT-EU is described in the proposal from 2021 as a “flagship legislative initiative” with a considerable political emphasis and aspirations to develop to a global standard. The general approach is to heavily regulate higher-risk systems and to allow systems with lower expected risk to be placed on the market and used without comparably strict rules. AIACT-EU has already been criticized for placing main regulatory tasks on AI providers, including ensuring adequacy of documentation, that may limit its effectiveness (Smuha & Yeung, 2024).

Stated purpose

The purpose of AIACT-EU is “in accordance with Union values, to promote the uptake of human centric and trustworthy [..] AI while ensuring a high level of protection of health, safety, fundamental rights [..], including democracy, the rule of law and environmental protection, against the harmful effects of AI systems [..], and to support innovation” (Introduction, paragraph 1).[1] In contrast to the prominence of ethics in the preparation of AIACT-EU, the final text is nearly void of explicit references to values

[1] On the European ambition to position itself as a global leader in ethical AI, see Krarup & Horst (2023).

(Ruschemeier & Bareis, 2024). The proposal underlined further supporting "EU's technological leadership" and achieving "a wide array of economic and societal benefits across the entire spectrum of industries and social activities," however, avoiding "risks or negative consequences for individuals or the society" (EU Permanent Representatives Committee, 2021, p. 1). Smuha and Yeung (2024) note that it is likely that AIACT-EU will be more successful in the latter i.e. supporting new markets whereas other existing legal frameworks are more effective in the protection of fundamental rights.

Scope

The focus of AIACT-EU is on "high-risk AI systems" whereas the transparency requirements, including those of "information" (specified in Annex IX), on non-high-risk systems are limited. AIACT-EU conceptualizes AI systems and models in terms of products (as per product safety), a perspective increasingly criticized as inadequate in capturing the complexity of how AI works (Ruschemeier & Bareis, 2024).

Key information artifacts identified for preservation

The key information artifact in AIACT-EU is *information and/or documentation* (e.g., Intr/par. 64, 164) where documentation appears to refer to outcomes of acts of documentation, and information to a resource provided for various audiences with an intention to inform. A parallel concept *comprehensible information* describes "how high-risk AI systems have been developed and how they perform throughout their lifetime is essential to enable traceability of those systems, verify compliance with the requirements under this [r]egulation, as well as monitoring of their operations and post market monitoring." (Intr/par 71).

The key part of the comprehensive information of AI systems is *technical documentation* (Annex IV). It is expected to contain "information which is necessary to assess the compliance of the AI system with the relevant requirements and facilitate post market monitoring", more specifically it "should include the general characteristics, capabilities and limitations of the system, algorithms, data, training, testing and validation processes used as well as documentation on the relevant risk-management system and drawn in a clear and comprehensive form" (Intr/par 71). The technical documentation should be "prepared and kept up to date" (Intr/par 101), it should be drawn before the system is put on to market or into service.

According to Article 53.1, parallel *technical documentation* relating to general-purpose AI models (cf. systems) is expected to include "its training and testing process and the results of its evaluation" with "information as appropriate to the size and risk profile of the model" consisting "at the minimum of" elements specified in the Annex XI of AIACT-EU. Similarly to technical information on AI systems, this information is to be provided, upon request, to the EU AI Office and the national competent authorities.

Transparency information in Article 53(1), refers to "technical information for providers of general-purpose AI models to downstream providers that integrate the model into their AI system" with a general description of the model including a series of obligatory pieces of information, and a description of the elements of the model and its development, similarly with a list of mandatory descriptions (Annex XII). Failing to provide information to regulatory bodies is subject to an administrative fine.

Concept of record, document(ation) and paradata in the text

As discussed above, AIACT-EU refers to documentation primarily as a top-level category consisting of technical documentation, and record keeping for regulatory purposes and information to users. Record keeping refers to keeping log data rather than records in an archives and records management sense. The

concept of paradata is not used as such even if the references to transparency in practice require it. Of specific artifacts, AIACT-EU mentions “model cards and data sheets” as examples of “widely adopted documentation practices”. Finally, AIACT-EU also makes references to “evidence” in juridical meaning, to “documentary evidence” in Article 29 relating to conformity assessment and in Article 30 in relation to notification procedure.

U.S. NIST (National Institute of Science & Technology) AI RISK MANAGEMENT FRAMEWORK (AI RMF-NIST-US)

Overview

The *AI Risk Management Framework (AIRMF-US)* released by the U.S. National Institute of Science and Technology (NIST) on January 26, 2023, is “intended for voluntary use and to improve the ability to incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products, services, and systems” NIST, 2023a).

An online companion resource, the *NIST AI RMF Playbook (AIRMF-PB-US)*, (NIST, 2023b) is a living resource that will evolve as AI technology advances. It will help organizations navigate the AI RMF and achieve their outcomes through suggested tactical actions they can apply within their own context.

Stated Purpose

The AI systems described generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments based on a given set of objectives, and do not address generative AI. They are designed to operate with varying levels of autonomy and pose varying levels of risk. Therefore, AI risk management is a key component of the responsible development and use of AI systems.

Scope

The AI RMF comprises two parts. Part 1 provides the foundational information related to understanding and addressing risks, and part 2 describes four AI RMF core functions of Govern, Map, Measure, and Manage.

Three categories of potential harm:

- Harm to people: Harm to their access to education or physical or psychological safety.
- Harm to an organization: Harm to business operations or reputation.
- Harm to an ecosystem: Harm to the global supply chain or damage to the environment.

Mitigating these risks will enhance trustworthiness of the AI system and, in turn, public trust. For AI systems to be trustworthy, the framework recommends balancing the following characteristics: safe; secure and resilient; explainable and interpretable; privacy-enhanced; fair with harmful bias managed; valid and reliable; and accountable and transparent.

Because AI risks may emerge during any stage of the AI lifecycle, activities that take place throughout the AI process as well as actors engaged in each stage and the tasks they perform are identified.

Key information artifacts identified for preservation

The four core functions identified help organizations address the risks AI systems pose in practice. The govern function must be addressed throughout the process, while map, measure, and manage apply at specific stages and in system-specific contexts. Certain actions are recommended throughout the AI system lifecycle; however, organizations should prioritize actions based on their own goals, objectives, and risk tolerance.

1. GOVERN. This function provides guidelines for the implementation of structures, systems, processes, and the human factor. Examples of the documentation recommended include AI policies, AI system documentation, applicable laws and regulations, data governance practices, impact assessments, and human factors such as training programs.
2. MAP. The Map function requires categorizing AI systems, understanding their capabilities and use, identifying risks and benefits, and characterizing impacts to individuals, groups, communities, organizations, and society. Examples include AI goals and mission, risk profile (tolerance), datasheets for datasets, and TEVV (testing, evaluation, validation, verification) metrics.
3. MEASURE. The Measure function assesses and monitors AI risk and related impacts. Examples include stakeholder engagement plans, algorithmic methodology, usability, and metrics developed and utilized to monitor, characterize and track external inputs, including any third-party tools.
4. MANAGE. Risks are mapped, measured and prioritized during the manage function based on their impact, likelihood of occurrence, and the resources or methods available for mitigation. Examples include AI system impact assessments related to data security and privacy, incident response plans, model cards and fact sheets, procedures for retiring the AI system, and third-party contracts/terms of service.

Concept of record, document(ation), and paradata in the document

The term “record” is not used in the *NIST AI RMF*, but the terms “document” or “documentation” are used throughout. Some of the documents or calls for documentation could be considered records that must be retained according to a retention and disposition schedule, such as an AI impact assessment. Other recommendations may result in a revision of existing records, such as ethics and compliance policies.

The *NIST AI RMF Playbook* provides guidance for documentation of actions recommended for each of the 72 specific recommendations. Understanding which documentation is retained as records is left to the implementing organization; the terms records or paradata are not used in this document.

Artificial Intelligence and Data Act, Canada (AIDA–CA)

Overview

In June of 2022, the governing Liberal Party of Canada introduced the Artificial Intelligence and Data Act (AIDA), packaged within Bill C-27 alongside updates to the existing federal Consumer Privacy Protection Act and the Personal Information and Data Protection Tribunal Act (House of Commons of Canada, 2022). While AIDA-CA has not passed into law, it would establish a regulatory structure for private sector AI tools distributed or implemented in Canada. The AIDA prescribes that AI designers and implementers maintain records to demonstrate appropriate risk mitigation practices throughout the AI

design, implementation, and operation processes, and empowers the Minister to levy appropriate penalties. The AIDA takes an agile risk-based regulatory approach to allow regulations to evolve with anticipated technological change.

Scope

AIDA-CA governs complete AI systems developed or implemented in the private sector (House of Commons of Canada, 2022, sec. 3). The AIDA-CA applies particularly to "high impact" AI systems, likely including employment screening services, biometric identification systems, online content recommendation systems, and safety critical systems such as automated vehicles (Innovation, Science and Economic Development Canada, 2023).

Stated purpose

The AIDA-CA's overall stated purposes include transparency, fairness and equity, safety, and accountability of AI systems. As per its own definitions, accountability "includes the proactive documentation of policies, processes, and measures implemented"; transparency provides information "sufficient to allow the public to understand the capabilities, limitations, and potential impacts of the systems" (Innovation, Science and Economic Development Canada, 2023).

Key information artifacts identified for preservation

AIDA-CA's Section 10 requires AI implementers to keep general records documenting each of these measures, which may be requested by the Minister to review AI applications and verify due diligence as per sections 13 and 14. Rather than prescribing precise records functions, AIDA-CA is written to allow regulators to mandate specific records practices as appropriate. It prescribes the retention of "general records" documenting four particular activities throughout an AI system's life-cycle:

1. Anonymization of data and data management (sec. 6)
2. Assessment of high-impact status (sec. 7)
3. Assessment of risks and mitigation processes for possible harms (sec. 8)
4. Ongoing risk mitigation measures (sec. 9)

As articulated in the Companion Document, AI users must engage in proactive "logging and monitoring the output of the system" to prevent harm. In addition to these records creation prescriptions, AIDA-CA mandates the availability of public-facing plain-language descriptions of the system, including an overview of the system, its scope and purpose, a risk assessment and mitigation report, and other relevant information (sec. 11). These prescriptions intend to provide public transparency and explanations, in addition to granular records requirements.

Concepts of record, document(ation), and paradata in the document

The AIDA-CA explicitly mandates records creation relating to specified activities throughout the AI lifecycle. Within these records, the Act mandates the logging and monitoring of system outputs, speaking to the preservation of what we describe as paradata. While the AIDA-CA does not use the terms "paradata," "document" or "documentation," its Companion Document calls for public-facing explanatory documentation to communicateAI practices to the public.

Directive on Automated Decision-Making, Government of Canada (DADM–CA)

Overview

Enacted in 2019 and effective in 2020, the DADM-CA applies a risk-based approach to AI regulation in the Canadian government. The Directive requires those employing automated decision-making systems to complete Algorithmic Impact Assessments, follow specific transparency requirements, assure the quality of automated systems before launch, and provide subjects of automated decisions with avenues for recourse.

Scope

The Directive applies to automated systems operated by the Canadian federal government.

Stated Purpose

The Directive intends reduce the risks which automated systems may introduce to federal government administration, and to make data and information about automated systems' use accessible to the public (4.1):

The objective of this directive is to ensure that automated decision systems are deployed in a manner that reduces risks to clients, federal institutions and Canadian society, and leads to more efficient, accurate, consistent and interpretable decisions made pursuant to Canadian law (4.1).

Key Information artifacts identified for preservation

The Directive requires those employing automated systems in government to classify applications by risk level before implementation. Systems are classified according to four risk levels, from "little to no impact" to "very high impacts" on individuals' and communities' rights, health, prosperity, and ecosystem sustainability. Higher-impact systems must clear a greater threshold of safety assurances and public notification. Mandated measures for high or very high-impact systems include gender-based impact analyses, expert review, plain language notification, assurances of humans-in-the-loop for impactful decisions, and the provision of meaningful decisions. Beyond these procedural prescriptions, the Directive identifies documentation requirements for automated systems' decisions. Documentation is intended to support ongoing system monitoring and quality assurance, responsible data management and use, and public information provision about automated systems' use (6.2.8).

Concepts of record, document(ation), and paradata in the document

The Directive uses the term "documentation" to describe multiple categories of information artifacts, including records of a system's implementation and public-facing explanations of operational automated systems. Records-creation is not explicitly identified within the Directive, although activities implicating records are discussed, such as risk assessments for specific implementations or ongoing quality assurance and monitoring.

Model Artificial Intelligence Governance Framework-Singapore. Second Edition (MF–SG)

Overview

The government of Singapore developed the Model Artificial Intelligence Governance Framework in 2020 through its InfoComm Media Development Authority (IMDA) and its Personal Data Protection Commission (PDPC).

Stated purpose

The Model Framework's goal is to provide a comprehensive governance framework to operationalize high level AI ethics principles into practice that can be adopted by organizations on a voluntary basis. The guiding principles of the Model Framework are explainability, transparency, fairness, human centricity, well-being, and safety. (p. 15)

Scope

The Model Framework is divided into four main areas:

a) Internal governance structures and measures
b) Determining the level of human involvement in AI-augmented decision-making
c) Operations management
d) Stakeholder interaction and communication

Key information artifacts identified for preservation

Of the four sections comprising the Model Framework, the Operations Management section is the most salient as regards requirements that imply a need for paradata and that provide evidence of its function within a governance context. The Operations Management section is divided into two sections based on an AI development lifecycle: (1) Data Preparation and (2) Algorithms and Chosen model.

Data Preparation

Data preparation includes the formatting and cleaning of data to be used to train a machine learning model. It includes three requirements areas: (a) lineage of data, (b) data quality, and (c) minimizing bias. The section on the lineage of data explicitly describes the need for "a data provenance record" that describes the data's origin and any transformations. The rationale is to trace back errors, support updates to the data, and to maintain knowledge of which data come from which sources (Operations\Data for Model Development\Sec.3.23, p. 37).

In the following section on data quality, there are seven criteria relating to the quality of the data set: accuracy, completeness, veracity, how the data set was compiled or updated, relevance and context, integrity, usability, and human interventions. While explicit calls for records preservation are not found in this section, these criteria imply a need for extensive documentation. (Operations\Data for Model Development\Sec.3.23, p. 38).

Finally, the section on minimizing bias describes types of bias, such as selection bias and measurement bias. As with the data quality criteria, this section implies activities documentation and data preservation activities to support bias mitigation measures. (Operations\Data for Model Development\Sec.3.23, p. 39).

Algorithm and Chosen Model

The section on algorithms and models provides extensive guidance on areas such as explainability, repeatability, robustness, regular tuning, reproducibility, traceability, and auditability. It also includes numerous, explicit requirements for documentation.

The section on explainability uses terms such as “documenting,” “descriptions,” and “specifications documentation” (Operations\Algorithm and Model\3.27|Explainability\a. & b., pp. 44-45). The sections on tuning and repeatability uses similar terms and adds “assessments” and “policies”. The section also describes performance tests such as counterfactual testing. implying the creation of documents to record results (Operations\Algorithm and Model\Repeatability\Regular Tuning\3.34, p. 48 Operations\Algorithm and Model\Repeatability\3.30,p. 46). The section on traceability uses terms such as “audit trail” and “black box recorder”, as well as “data” and its storage. The sections on Reproducibility and Auditability discuss documentation in the context of “independent verification” by “external auditors”, and sums up documentation supporting auditability (Operations\Algorithm and Model\Auditability\3.43., p. 51).

Concepts of record, document(ation), and paradata in the document

The Model AI Governance Framework explicitly references the concept of a record in places. Its uses are consistent with the InterPARES definition in Table 2. Reference to a “data provenance record” is consistent with a fuller sense of record as used in the records and archives fields where authenticity, reliability, integrity, and trustworthiness are key criteria of records. The term ‘documentation’ and variants of it are used in the Model Framework. Specific types of documentation are cited, such as ‘specification documentation’, as noted above.

In addition to reference to records and documentation, other types of documents are referred to that may fall in one or both categories. For example, the term 'assessment’ refers to a document or type of information that could easily count as a record under the InterPARES interpretation and under a fuller, evidential interpretation. The same holds for audit trails and other terms that explicitly refer to types of writings. Further, there are terms that describe entities or activities that imply some sort of documentation or record.

Findings

The authoritative texts reviewed have much to say about records requirements, but indirectly and in such a way as to make paradata a salient concept. Terms such as accountability, transparency, auditability, or explainability proliferate in authoritative texts and imply the necessity of records creation, capture, and preservation, even if they do not explicitly refer to recordkeeping directly. However, the task of identifying specific documents required for preservation remains one to be completed by AI implementers, archivists, and records managers. The texts reviewed illustrate a clear emergent consensus towards a risk-based approach to AI governance worldwide. We find that records provisions for AI implementation will be far more intensive and require the identification and preservation of far more granular paradata in high-risk applications than lower risk applications.

The AI risk governance framework comparison chart used to analyze the documents reviewed is available in the supplemental materials.[2] It contains the full chart and tabs that break the chart into four manageable sections: goals and approach, basic overview, records provisions, and documentation prescribed/terms.

Discussion

AI definition and stated goals

Across the authoritative texts surveyed, a common approach to defining and governing AI through risk-based frameworks emerges. Each definition recognizes AI as a machine-based or engineered system potentially applying diverse AI techniques including machine learning, neural networks, and decision tree algorithms in both static and adaptive models, focusing on its computational nature. The definitions typically draw comparisons based on AI's ability to simulate human cognitive functions such as reasoning, problem-solving, and learning for the purpose of generating outputs such as decisions, predictions, or recommendations, indicating an overlap in goals between AI and human intelligence. They highlight the core function of algorithmic decision making by emphasizing the role of AI in generating outputs such as decisions, predictions, or recommendations. Adaptiveness is a recurring theme, particularly where AI systems can learn from data inputs and evolve their behaviors, often through machine learning.

The core common approach to defining AI is in a functional definition focusing on AI systems' applications, rather than their nature. Of the six authoritative texts surveyed in this paper, five define AI based on a functional definition. With the exception of DADM-CA, the core definition offered by each document identifies AI as a non-human technological system which may be capable of making decisions with real-world consequences. Four of these documents describe the use of AI tools as being the generation of predictions, recommendations, or decisions, describing variable levels of AI systems' autonomy. While DADM-CA may appear to be an outlier, it is in fact more fine-grained in approaching variable AI autonomy: while an AI system in its definition only processes information, an automated decision system takes an active role in making real-world actions, whether that be through prediction, recommendation, or decision-making. Clearly, from a governance standpoint, the autonomy of AI systems poses a primary risk and emerges as the target definition of pertinent legislation. The nature of these systems – be they straightforward rules-based systems or opaque machine learning platforms – will influence the level of risk and the nature of regulation applied.

All the authoritative texts, to varying extents, embrace a risk-based approach to AI governance. This common risk-based approach informs the common functional AI definitions employed, as the documents aim to provide evergreen guidance in general terms rather than wedding themselves to particular technologies. Since the texts are interested in minimizing the real-world harms caused by automated systems, academic distinctions between the structure of AI models are less relevant than the fact that the systems introduce novel risks into the administration of business and government. Being applied across vastly different contexts, these frameworks and laws leave the assessment of risks to those applying AI tools, and generally do not forbid specific AI applications. Of the authoritative texts surveyed, the EU's AI Act is alone in explicitly forbidding categories of AI tools (for instance individual-based predictive policing); other guidance documents rely on existing, less technologically-specific legislation to permit or forbid given AI applications. As Ruschemeier and Bareis (2024) note, risk management frameworks leave

[2] This material will be available in Harvard Dataverse at https://doi.org/10.7910/DVN/JLWEE.

ethical considerations external to the legislation enacted. AIDA-CA, like the EU's AI Act as noted by Ruschemeier and Bareis, leaves ethical considerations in a companion document intended to interpret the legislation rather than in the law itself. Since risk-management approaches do not necessarily fully embody statements of AI ethics, records professionals working with AI tools will need to assess AI legislation in relation to their own professional ethics and field-specific legislation as well.

Each authoritative text reviewed articulates its goals as broader principles. Of primary importance across the texts is the accountable and responsible use of AI, particularly by government agencies. They reflect the common themes of the need for transparency and ethics in AI governance, balanced by a consistent desire to harness AI's potential for economic gains. For instance, the Canadian government's DADM public sector regulation mentions maximizing efficiency alongside minimizing risks. Putting similar goals in the terms of the private sector, the UK's PIAAIR explicitly courts private sector investment by stressing the flexibility of its regulatory scheme, much like Canada's AIDA.

Differences between the texts reviewed can be seen in the national focus taken to encourage in the US the responsible use of AI by federal agencies contrasted with the more general or global perspectives in other documents. Compliance with existing privacy legislation is a goal highlighted in the voluntary Singapore Model AI Governance Framework, but privacy as a goal is not specified in the other documents reviewed. And some texts specifically aim to build stakeholder confidence in AI, showing a focus on public perception and trust, which is not always explicitly mentioned in others. This requires a balanced approach that suggests a need for flexible and adaptive policies that can keep pace with the rapidly evolving AI landscape.

In summary, while there is a strong consensus on promoting AI innovation, protecting individuals, and ensuring agile regulation, the authoritative texts differ in their emphasis on public trust, compliance with privacy laws, and specific national concerns like the role of AI in US federal agencies.

Approach to risk assessment

All of the authoritative texts reviewed approach risk assessment by categorizing, with varying levels of formality, AI systems into levels of risk, typically distinguishing between high-risk systems and those with lower levels of risk. Differences appear in the approach to specifying risk levels: the AIACT-EU uses four levels (limited risk, minimal risk, high risk, and unacceptable risky, forbidden systems); Canada's DADM also uses four levels (e.g., little to no risk, moderate risk, high risk, and very high impact on the rights of individuals or communities, equality, health, economic interests, and ecosystem sustainability); Canada's AIDA recognizes two levels (e.g., high-impact vs. non-high-impact systems); and the EU's Model Rules specify three levels (no risk, high risk, and in-between).The remaining two texts do not prescribe formally tiered approaches to risk, but prescribe regulatory measures proportionate to the level of risk with higher-risk systems facing more stringent requirements. Among the high-impact or safety-critical systems mentioned are applications in healthcare, infrastructure, employment or government benefit eligibility screening, and law enforcement. Case studies, such as insurance pricing, chatbots, and autonomous vehicles, are often used to illustrate the types of systems in use in different risk categories. A few texts suggest involving humans in decision-making processes for higher-risk AI applications (human-in-the-loop); others do not. Certain statements prescribe ongoing and preparatory risk assessments, while others may not explicitly structure or formalize these assessments.

In summary, while there is broad agreement on categorizing AI systems by risk and applying proportionate regulatory measures, the specifics vary in the approach to risk assessment, the identification of high-risk systems, and the structure and formality of the risk management approaches.

Incorporating records creation into AI governance and AI system design for accountability and transparency

While risk-based approaches to AI governance clearly predominate alongside common foundational principles of transparency, accountability, and ethical considerations, necessary records considerations are not necessarily provided within the authoritative texts. What evidence can be provided as proof of the responsible governance of AI records? Our survey of the records requirements of existing AI legislation shows that purposeful records provisions designed to capture AI paradata will be a key aspect of AI governance to ensure compliance with existing laws and regulations. This necessitates creating/capturing, managing, and preserving the required records and providing relevant documentation mandated explicitly and implicitly to explain how the AI tool was developed and used. AI tools/technologies must be designed to provide explainable, transparent documentation for all stakeholders. This is where archivists and records and information managers can leverage their expertise to influence and assist those developing and deploying AI solutions to ensure the proper documentation is created, captured, and made accessible to stakeholders based upon their roles within the process (e.g., AI developer, business unit, external auditor, legal department, customer). Organizations must understand the legislation governing their responsibilities related to AI and extract the mandated requirements for documentation.

To address the challenges of preservation of records, we must first answer the question, *What records need to be preserved to enable accountability and transparency?* One way to provide context is by linking the requirements and recommendations from the authoritative texts reviewed to the stages of the AI lifecycle.

AI Lifecycle Approach

Some AI lifecycle models are specific to the type of AI technology employed, such as the Machine Learning Lifecycle. Others are model agnostic. The AI Process Lifecycle provides a starting point when considering the steps to be documented throughout the AI process. We can examine at each of the stages of the AI Process Lifecycle, the questions to be asked during each stage, and examples of the types of documentation implied or required as mandated or recommended by the documents reviewed areas outlined in Table 3.

Table 3. AI Process, Questions to be Answered, and Examples of Documentation.

Stage	**Examples of questions to be answered**	**Examples of evidence of actions taken/decisions made**
Planning & Design	What is the problem we are trying to solve and the desired outcome?	Use case: form describing problem, title, intended outcome; stakeholder analysis; IBM Fact Sheets
Data Collection & Processing	What data sets are needed, available, and what is their quality? Is it possible to transform the initial raw data into a format the model can use?	Microsoft Datasheets for Datasets

Modeling	What is the best existing model and settings to achieve the desired outcome? Would designing or adapting a model for this specific purpose be warranted?	Google Model Cards
Outcome Analysis	What are the results of tests run using the model with new data? Do the models generalize well and meet business goals?	Reports on integration with legacy systems, compliance testing, and validation
Deployment & Monitoring	Is the model effective when used with data that was not part of the design and development phase? Does the model continue to produce the desired result as it adapts to new data? Is there evidence of model drift?	Audit and Impact Assessments, ongoing system performance logs and monitoring

Challenges of Preservation for accountability and transparency

The volume and variety of documentation required and recommended demands a technological solution to preservation for accountability and transparency—one that facilitates access to the desired data depending upon the roles of stakeholders in the AI process. A scan of the commercial landscape shows that progress is being made. Some of the evidence of documentation in Table 3 is already part of the operations of the system. For example, IBM AI Factsheets track a machine learning model from request to production for defined business use cases in a model inventory. As the AI model advances in the life-cycle, the model use case and AI factsheet reflect all updates, including deployment and input data assets (IBM, 2024).

Future research into the development of a comprehensive *AI process management system*, perhaps based on Enterprise Knowledge Graphs, is recommended so that all data that may be required of any internal or external stakeholder could be aggregated and made available upon request. This would include not only datasheets, model cards, audits and impact assessments but also links to governance documents like privacy and ethics policies and environmental impact statements which would reflect the environment in which the AI system was developed, deployed, and operated.

Prescriptions and provisions for record creation, capture, and preservation

Our analysis shows that the authoritative texts sampled carry extensive records creation and preservation requirements that will provide ample opportunity for the records and archives professions to contribute significantly. Few of the texts analyzed use the term record; however, they frequently describe practices and documents that imply the creation of and need for (archival) records. In a similar manner, although none of the texts used the term paradata, they make indirect references to the need to record how AI

works and what was accomplished using it. On an overarching level the stipulations refer to transparency and explainability (e.g., PIAAIR-UK). This can be achieved by conveying evidence (e.g. AIACT-EU) and an understanding of the implementers' practices and the ways AI works for both legal and operational purposes. The more comprehensive texts also provide examples of processes such as risk and data management requiring special attention. Depending on their level of detail, the authoritative texts make occasional references to specific aspects of documentation and documents expected to provide aspired kind of transparency (e.g. AIACT-EU, RMF-PB-US). The findings echo Enqvist's (2024) remark that in a formal sense the concept of paradata does not exist in the legal domain. Another comparable aspect of paradata apparent in the present findings and the earlier literature is the diversity of documentary artifacts that can qualify as (carriers of) paradata (cf. e.g. texts in Huvila et al., 2024).

Explicit versus Implicative Terms

While the authoritative texts mention explicit information objects often, and records sometimes, there are many references to activities implying the creation or collection of information objects. We call terms denoting such activities 'implicative' terms contrasted with explicit terms that refer directly to information objects.

Starting with explicit terms, we can notice that certain general terms appear often. 'Documentation' appears in all of the authoritative texts. It also appears in qualified form, with 'technical documentation' being common. 'Record' appears infrequently, but it does appear in some documents. For example, the term 'data provenance record' is used in the Model AI Governance Framework in a manner consistent with the InterPARES definition and possibly with a fuller, evidential interpretation. Other explicit terms used include 'assessment', 'audit trails', 'data', 'internal policy', and 'Risk Management Plan', as well as AI specific terms such as 'Model Cards', 'Fact Sheets', 'Algorithmic Impact Assessments' and the like.

Such explicit terms point to information artifacts that are clearly potential records.

Implicative terms, by contrast, may be missed, even though their presence is indicative of records management requirements potentially more extensive than the explicit references to documentation. Implicative terms identified in the authoritative texts include (but are not limited to) phrases such as: "Accounting for changes," "Auditability," "Benchmarking," "Black box recorder," "Counterfactual fairness testing," "Traceable," and "Training, Validation and Testing."

These phrases describe activities, mechanisms, and capabilities. The terms all clearly imply that data and information will be created and need to be preserved. For example, auditability requires that documentation be in place and audits produce reports. Training requires documentation and often a record of training completed, while testing implies that results are recorded. The result of this is that the identification and interpretation of implicative terms will be as critical to defining records roles and responsibilities as the identification of terms that explicitly refer to information artifacts.

Paradata as Organizing Concept

The distinction between explicit and implicative terms is broad and encompasses many useful distinctions within it. For example, some terms imply automated processes and machine generated data, while others imply the intentional creation of documents. An example of the former is audit logs. While their parameters are set by persons, the logs are auto-generated by the system. Other terms indicating machine processes include "Traceable," and "Black box recorder." Terms referring to human-created information

include “internal policy,” “Risk Management Plan,” “Model Cards”, and “data provenance record.” The concept of paradata encompasses both types of terms. Cameron, Franks, and Hamidzadeh describe two senses of the term “paradata,” one that is statistical and one that involves the deliberate documentation of a process. (Cameron, et al., 2022) Cameron, et al. find evidence of the second sense in virtual heritage work. They quote Drew Baker who proposed using the concept “... as a descriptor of process documentation. (2007). That paradata as a concept includes both statistical, machine generated data and human information artifacts suggests its utility in the identification of diverse forms of data and information for purposes of records creation.

Among the human created information artifacts, some are clearly created with the intention that they serve record creation purposes, while others are created for independent purposes, but may be preserved as a record as a secondary function. Risk management plans and policies are two examples of the latter. While they serve specific purposes, they can be useful as artifacts that help us understand the behavior and context of an organization. In this way they can serve as records or sources of records. As such, they can be considered a kind of paradata according to the InterPARES definition cited in this paper and may even constitute a new category of paradata to be added to the nascent literature.

Conclusion: Paradata can address recordkeeping challenges with AI systems

As the review of the governance documents completed suggests, AI tools will likely emerge in both heavily regulated, high-risk contexts and lower risk applications with informal or self-governance structures regulating activity. None of these spheres will be exempt from recordkeeping requirements. Even in documents which make no explicit calls for records creation and preservation in terms recognizable to archivists and records managers, the authoritative texts reviewed clearly describe activities which imply the necessity of records creation and preservation. While specific calls for documentary records of AI activity are few and far between in law, the texts reviewed present a clear mandate for the development of approaches to AI recordkeeping which are scaled according to levels of risk as defined in legislation. Given the archives field’s long interest in assessing the nature of trustworthiness in records, the field must also engage with the documents produced by AI if these tools will be employed in responsible organizations.

While the need for AI records management programs is clearly emergent, ascribing formal and specific records requirements for AI is especially challenging. Technological specifics are often trade secrets and change rapidly. Accordingly, the reviewed texts tend to prescribe categories of activity aimed at reducing risks created by AI implementers or designers. For example, risk assessments and risk management plans tend to substitute for the identification of specific risk mitigation practices and prescriptions for the preservation of specific forms of information. Similarly, it is not a simple matter to identify the specific documents necessary to comprise a sufficient body of records to provide evidence of accountable AI implementation and operation. Illustrating the challenges with identifying specific information artifacts for preservation is Canada’s AIDA, which otherwise offers the most explicit and capacious calls for records preservation in the documents reviewed. Even so, AIDA-CA does not go beyond mandating the preservation of “general records” pertaining to relevant phases of the AI lifecycle. Clearly, identifying specific information artifacts for preservation will be a challenge for AI implementers and regulators in the months and years to come.

With this in mind, the burgeoning literature on paradata in archival applications of AI may prove productive in identifying relevant information artifacts for preservation in the AI process. Paradata is

produced by those designing and implementing AI systems and by AI systems themselves. Paradata as a category of information may be generated as an intentional or incidental product of ongoing human and technological processes (Huvila et al., 2024b). It is clear from the authoritative texts reviewed that necessary paradata will be both technical and organizational in nature. Technical paradata will be often machine-generated, but occasionally generated by people, and will record closely the details of the design and operation of technical systems; organizational paradata, more broadly speaking will be necessary to record the process of implementing and managing these systems within human organizations, and will comprise a much broader category of information artifacts. Since AI legislation and governance documents show clear interest in traditional bureaucratic documents (e.g. risk management plans), communicative documents (e.g. manuals for system operation) and technical, system-generated data (e.g. system output monitoring logs), AI recordkeeping will clearly require greater attention to information generated by processes which rely upon the independent and combined capacities of human and computerized agents. In this sense, a micro-level attention should be paid to informational output of these processes - and of its suitability to provide evidence of those processes and their responsible administration. Identifying relevant paradata produced by these systems and those responsible for them behooves information researchers and professionals, archivists and records managers to develop both the capacity to analyze and the vocabulary to discuss AI systems in order to identify and preserve evidence of their operation.

Acknowledgements

TBA after peer review.

References

Baker, D. (2012), “Defining paradata in heritage visualization”, in Bentkowska-Kafel, A., Denard, H. and Baker, D. (Eds.), *Paradata and Transparency in Virtual Heritage*, Ashgate, Farnham, pp. 163–175.

Bunn, J. (2020), “Working in contexts for which transparency is important: A recordkeeping view of explainable artificial intelligence (XAI)”, *Records Management Journal*, Vol. 30 No. 2, pp. 143–153.

Baker, D. (2007). “Towards Transparency in Visualization-based Research”, presented at the *VIZNET 2007 Conference “Intersections in Visualization Practices and Techniques,”* Loughborough University, UK, 19th April 2007.

Cameron, S., Franks, P. and Hamidzadeh, B. (2023). “Positioning Paradata: A Conceptual Frame for AI Processual Documentation in Archives and Recordkeeping Contexts”, *Journal on Computing and Cultural Heritage*, Vol. 16 No. 4, pp. 1–19.

Couper MP (1998). “Measuring survey quality in a CASIC environment”. in *Proceedings of the Section on Survey Research Methods of the American Statistical Association, 1998, Dallas*. American Statistical Association, Alexandria, VA, pp 41–49.

Davet, J.E., Hamidzadeh, B., Franks, P.C., Bunn, J., Hamidzadeh, B., Franks, P.C. and Bunn, J. (2022), “Tracking the Functions of AI as Paradata & Pursuing Archival Accountability”, *Archiving Conference*, Society for Imaging Science and Technology, Vol. 19, pp. 83–88.

Davet, J., Hamidzadeh, B. and Franks, P. (2023), "Archivist in the machine: paradata for AI-based automation in the archives", *Archival Science*, Vol. 23 No. 2, pp. 275–295.

Duranti, L. (2005), "Reliability and Authenticity: The Concepts and Their Implications," *Archivaria* 39, pp. 5-10.

Duranti, L. and Thibodeau, K. (2006), "The concept of record in interactive, experiential and dynamic environments: the view of InterPARES", *Archival Science*, Vol. 6 No. 1, pp. 13–68.

Enqvist, L. (2024), "Paradata as a Tool for Legal Analysis - Utilising Data-on-Data Related Processes", in Huvila, I., Andersson, L. and Sköld, O. (Eds.), *Perspectives on Paradata: Research and Practice of Documenting Data Processes*, Springer, Cham.

EU Permanent Representatives Committee. (2021), *Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts - General Approach, Interinstitutional File No. 2021/0106(COD) 14954/22*, Council of the European Union, Brussels.

European Union. (2024), Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Pub. L. No. 2024/1689, Document 32024R1689 (2026).

Franks, P.C. and Cameron, S. (2023), "Paradata: Documentation for Responsible Artificial Intelligence", *AIIM Blog*, 23 October, available at: https://info.aiim.org/aiim-blog/paradata-documentation-for-responsible-artificial-intelligence.

Government of Canada. (2019). Directive on Automated Decision-Making. Available at https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592.

Office for Artificial Intelligence. (2023). A pro-innovation approach to AI regulation (Policy Paper No. 815; Command Paper). HM Government, available at https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

House of Commons of Canada. (2022). "Bill C-27: An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts." Parliament of Canada, available at https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading.

Huvila I (2022). "Improving the usefulness of research data with better paradata". *Open Information Science* Vol 6 No. 1, pp. 28-48.

Huvila, I. and Ekman, S. (2024), "Documentation of data making, processing and use facilitates future reuse of research data: the CAPTURE project", in Elena Volodina, Gerlof Bouma, Markus Forsberg, Dimitrios Kokkinakis, David Alfter, Mats Fridlund, Christian Horn, et al. (Eds.), *Huminfra Conference*, Linköping University, Linköping, pp. 26–30.

Huvila, I., Andersson, L. and Sköld, O. (Eds.). (2024), *Perspectives on Paradata: Research and Practice of Documenting Data Processes*, Springer, Cham.

Huvila, I., Andersson, L. and Sköld, O. (2024b), "Concluding Discussion: Paradata for Information and Knowledge Management", in Huvila, I., Andersson, L. and Sköld, O. (Eds.), *Perspectives on Paradata: Research and Practice of Documenting Process Knowledge*, Springer International Publishing, Cham, pp. 249–264.

IBM (2024), "Using AI Factsheets for AI Governance", *Documentation for Cloud Pak for Data as a Service*, IBM, Armonk, available at https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/factsheets-model-inventory.html

Info-communications Media Development Authority (IMDA) and Personal Data Protection Commission (PDPC) Government of Singapore. (2020) Model Artificial Intelligence Governance Framework, 2nd Edition. Available at https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework

Innovation, Science and Economic Development Canada. (2023). "The Artificial Intelligence and Data Act (AIDA) – Companion document." Government of Canada. Available at https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document.

Krarup, T. and Horst, M. (2023), "European artificial intelligence policy as digital single market making", *Big Data & Society*, Vol. 10 No. 1, p. 20539517231153811.

Luna, J., Tau, I., Xie, X., and Jiang, L. "Navigating Governance Paradigms: A Cross-Regional Comparative Study of Generative AI Governance Processes & Principles", *Proceedings of the Seventh AAAI/ACM Conference on AI, Ethics, and Society (AIES 2024)*, pp. 917-931.

Mooradian, N. (2019), "AI, Records, and Accountability", *Information Management, ARMA-AIEF Special Edition 2019*, pp. 9–13.

NIST. (2023a). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, available at: https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

NIST. (2023b). Artificial Intelligence Risk Mangement Framework (AI RMF) Playbook, available at: https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook

O'Connor, H. and Goodwin, J. (2020). *Paradata*. SAGE Research Methods Foundations. SAGE: London.

Public Record Office Victoria. (2024), "Artificial Intelligence (AI)", *A-Z Topics*, Melbourne.

Ruschemeier, H., & Bareis, J. (2024). "Searching for harmonised rules: Understanding the paradigms, provisions and pressing issues in the final EU AI Act" , *SSRN Scholarly Paper*, Rochester, NY, available at https://papers.ssrn.com/abstract=4876206

Schlegel, M. and Sattler, K.-U. (2023), "Management of Machine Learning Lifecycle Artifacts: A Survey", *ACM SIGMOD Record*, Vol. 51 No. 4, pp. 18–35.

Russell, S., Perset, K. and Grobelnik, M. (2023), "Updates to the OECD's definition of an AI system explained", *AI Wonk*, 29 November, available at: https://oecd.ai/en/wonk/ai-system-definition-update (accessed 27 November 2024).

Schlegel, M. and Sattler, K.-U. (2023). “Management of Machine Learning Lifecycle Artifacts: A Survey”, *ACM SIGMOD Record*, Vol. 51 No. 4, pp. 18–35.

Smuha, N. A., & Yeung, K. (2024). “The European Union’s AI Act: beyond motherhood and apple pie?”, *SSRN Scholarly Paper*, Rochester, NY, available at https://papers.ssrn.com/abstract=4874852

Trace, C. B., & Hodges, J. A. (2024). “The Role of Paradata in Algorithmic Accountability”. In I. Huvila, L. Andersson, & O. Sköld (Eds.), *Perspectives on Paradata: Research and Practice of Documenting Process Knowledge* (pp. 197–213). Springer.

Veale, M., Matus, K. and Gorwa, R. (2023), “AI and Global Governance: Modalities, Rationales, Tensions”, *Annual Review of Law and Social Science, Annual Reviews*, Vol. 19 No. 19, 2023, pp. 255–275.