Operationalizing Context: Contextual Integrity, Archival Diplomatics, and Knowledge Graphs

Jim Suderman
Independent Expert
Toronto, Canada

Frédéric Simard
School of Computer Science
Carleton University
Ottawa, Canada
0000-0001-6419-6899

Nicholas Rivard
School of Computer Science
Carleton University
Ottawa, Canada
0009-0002-3068-5496

Iori Khuhro
School of Information
The University of British Columbia
Vancouver, Canada
0009-0002-6403-4149

Erin Gilmore
S*chool of Information*
San Jose State University
San Jose, CA, USA
0009-0008-0249-1954

Michel Barbeau
School of Computer Science
Carleton University
Ottawa, Canada
0000-0003-3531-4926

Darra Hofman
School of Information
San Jose State University
San Jose, CA, USA
0000-0002-1772-6268

Mario Beauchamp
School of Computer Science
Carleton University
Ottawa, Canada
0009-0008-5636-5566

***Abstract*— Protecting privacy has become a pressing problem for archives, which are mandated with providing access to enormous volumes of records, both analogue and digital, with limited resources. Content-driven solutions, including anonymization and pseudonymization as well many automated solutions, have delivered unsatisfactory results, with many records being broadly restricted. In this paper, we lay out the theoretical framework for a context-based AI privacy solution for archival records, combining three approaches to "context" (contextual integrity, archival diplomatics, and knowledge graphs). This framework lays the groundwork for a GraphRAG workflow that identifies critical contextual information about privacy in records collections. By centering context and making it machine-legible, this solution operationalizes contextual integrity, allowing archivists and other records professionals to make informed decisions about privacy in a resource-efficient manner.**

Keywords—computational archival science, privacy, knowledge graph, machine learning, retrieval augmented generation, large language model

I. Introduction

Privacy protection is a core value according to the ethical codes of many professions, including the archival profession. However, despite their professional charge of caring for records of enduring value, both analogue and digital, archivists often have limited knowledge about the records that come into their custody. In the context of privacy protection, this means that they inherit considerable responsibilities without any detailed knowledge of the information that may have been available to the original creator of the records. Archival tools and practices are generally designed for administering records and making them accessible. There are few tools from which to develop strategic ways, proportionate to the resources available to archivists, to fulfil their responsibilities to protect the privacy of individuals and communities referenced within their collections.

Combining context-based approaches, including contextual integrity and archival diplomatics, with Artificial Intelligence (AI) tools and techniques to read documents, could change that. But to do so may require an adjustment of approaches and assumptions on which many privacy regimes are founded. Such change, in turn, would require advocacy to achieve. But like protecting privacy, advocacy is also a value well represented in the ethical principles of these communities.

However, meaningful advocacy requires a clear understanding of the problem and a vision of the solution. The problem of archival privacy management in an AI-age highlights the gap between data-centric, compliance-oriented approaches currently enshrined in privacy legislation, and context-dependent realities of effectively protecting privacy. Archivists must apply privacy-driven concepts of context *and* have detailed knowledge of document contents to effectively manage risks to the privacy of the data subjects. Integrating archival- and privacy-based concepts of context, as is proposed here, provides a framework for analyzing and establishing documentary privacy context. Applying knowledge graphs designed for that framework can allow Large Language Models (LLMs) to gather information about documents' contexts *and* contents to identify the range of privacy risks and their likelihood, enabling archivists to manage them more effectively. To address this problem, this paper poses two research questions:

1. Can context-driven theories of records and privacy - in particular, archival diplomatics and contextual integrity (CI) - be operationalized to protect privacy?
2. Can these frameworks be used to develop automated privacy-protection tools that can summarize the actors involved, the types of information being transmitted, and assess the norms of and exceptions to that transmission?

II. THE CLASSIC QUESTION: WHAT IS PRIVACY?

Adjusting the approaches and assumptions on which privacy regimes are founded inevitably leads to the question that seems to begin all privacy articles: “What is privacy?” The Glossary of the International Association of Privacy Professionals (IAPP) defines privacy as: “A nebulous philosophical, legal, social and technological concept which means different things to different observers[1].” The InterPARES Trust AI Project, an international and inter-disciplinary research project examining archival concepts in the light of changing technology, defines privacy as “1. A quality or state of seclusion, of keeping to one's self, and being free from intrusion or public scrutiny. 2. Control over access and use of one's personal information [2].” Both definitions are prone to being understood as: i) binary: one either has privacy or not; and ii) something that applies to individuals rather than contexts. These conceptions are problematic in that privacy exists on a continuum, qualitative and deeply contextual.

The nebulous nature of privacy as a concept is not simplified by varying definitions of what privacy is meant to protect. The general focus of privacy legislation is “personal information,” commonly referred to as personal data in a European context and personally identifiable information (PII) in the United States, with these terms used interchangeably herein. Furthermore, common to all the definitions of personal information is an understanding that the information that could potentially be PII is unbounded. Narayanan concluded that “[w]hile some data elements may be uniquely identifying on their own, any element can be identifying in combination with others[3, p. 20].” Their conclusion is supported by legal scholar Paul Ohm, who characterized PII-based approaches privacy protection as a game of whack-a-mole, concluding that “we must abandon the pervasively held idea that we can protect privacy by simply removing personally identifiable information (PII) [4, p. 1742].” Ohm wrote these words over a decade ago; in the intervening time, as yet more auxiliary data has become ever more accessible and analyzable about almost every person on earth[1], it has become increasingly clear that our current approach to privacy serves primarily to enrich data controllers and processors.

Even the widely accepted exception to privacy legislation - anonymized data - is problematic. While there are a number of approaches to anonymizing data - such as *k*-anonymity and differential privacy - their implementation in real life has proven to be problematic, with various attacks enabling data re-identification, or the noise used to anonymize the data negatively impacting analysis of said data [5], [6].
If protecting privacy is not possible by redacting or even anonymizing personal information, then two things are needed: 1) alternative ways of protecting privacy must be identified; and 2) existing regulations must be revised to reduce reliance on anonymization and strengthen alternative approaches.

If anonymization is not an effective means of protecting privacy, the idea of moving from a data-centric approach for privacy protection to a context-centric one holds promise. Although a context-centric approach may be more complex, it may also have the advantage of being more effective, particularly in terms of managing privacy-related risks. And AI tools may have the potential to make a context-centric approach practicable, even when resources are limited.

III. CONTEXT IN THE ARCHIVES

This study sought to operationalize contextual integrity for privacy protection of archival records by building on? the inherently context-bound nature of archival records, and the central role of context in archival science. Archival science is an academic and applied discipline involving the scientific study of process-bound information, both as product and as agent of human thoughts, emotions, and activities, in various *contexts*. The creation of records – archival science’s primary object of concern - relies on five contexts [7]:

1. *Juridical–administrative***:** the legal and organizational system in which the record creating body belongs, as indicated by laws, regulations, etc.;
2. *Provenancial*: the record creating body, its mandate, structure, and functions, as indicated in organizational charts, annual reports, the classification scheme, etc.;
3. *Procedural*: the business procedure in the course of which the record is created, which in the modern environment, is often integrated with documentary procedures, indicated by workflow rules, codes of administrative procedure, classification schema, etc.;
4. *Documentary*: the archival *fonds* to which the record belongs and its internal structure, as indicated by classification schemes, record inventories, indices, registers, etc.;

[1] Inhabitants of Sentinel Island, perhaps, excluded.

5. *Technological*: the characteristics of the technical components of the system in which the record is created.

These contexts are essential to preserving the authenticity and reliability of the records, that is, to demonstrate that records are what they purport to be and can be relied upon as evidence of past acts and facts. However, archival scholar Geoffrey Yeo observes that: "context is infinite, and every context has contexts of its own. In practice, if archivists and archival institutions seek to document context, they have to decide what levels of context are most relevant to their needs or the needs of their users [8, p. 223]." Decisions about what to emphasize and ignore vary across the archival domain, influenced by considerations such as the mandate of the organization holding the records and the nature of the users of those records. That said, archivists and archives have made such decisions, and the mechanisms for documenting contexts for centuries. It seems appropriate to consider archival contexts as a means of identifying and mapping privacy-related contexts. Privacy-related contexts may, like archival ones, be infinite in nature, leaving privacy specialists with the same kind of decisions as archivists: what to emphasize or to ignore.

IV. DIGITIZATION, THE NEED FOR ARCHIVAL PRIVACY, AND CONTEXTUAL INTEGRITY

Almost perversely, the proliferation of records available in the wake of our ever-expanding universe of digital information technologies seems to have led to an ever-shrinking portion of those records being accessible. Responsible decision-making regarding access to archival records has always been difficult. In addition to the ongoing tensions outlined above - between public and private interests, between dominant and non-dominant groups, between medium and message - archives have always sought to fulfill multiple, sometimes contradictory, roles. Terry Cook traces the evolution of archival paradigms, explaining that "[t]he focus of archival thinking has moved from evidence to memory to identity and community [9, p. 117]," with Devon Mordell noting that "To the four archival paradigms of evidence, memory, identity, and community [...] a fifth may now be poised to emerge: an archives-as-data paradigm [10, p. 140]." The appropriate balance between access and privacy may indeed differ amongst those paradigms.

Lemieux and Werner consider the possibilities of using "privacy-enhancing technologies (PETs) — a class of emerging technologies that rest on the assumption that a body of documents is confidential or private and must remain so [11, p. 83:2]" to allow analysis of archival records without opening access to those records and risking the leakage of personal information. These technologies include homomorphic encryption, trusted execution environments, secure multiparty computation, differential privacy, personal data stores and self-sovereign identity, privacy-preserving machine learning, and synthetic data. Indeed, Lemieux et al. have developed Clio-X, a "privacy-first platform" which provides access to, among other things, archival records, in part by computing and returning query responses based on archival data [12]. While archives-as-data has become a legitimate archival paradigm, the evidential, memory, identity, and community needs for archives and records remain, sometimes necessitating access to records *qua* records. Furthermore, while AI tools have automated a number of tasks, Lemieux and Werner note that "NLP models may not be capable of identifying more complex contextual privacy issues, such as statements that could affect the reputation of an individual [11, p. 83:4]."

In 2006, Barth et al. introduced a context-centric approach to the concept of privacy [13, pp. 2–3]. Essentially, it assumes: "[…]that people act and transact in society not simply as individuals in an undifferentiated social world, but as individuals in certain capacities (roles), in distinctive social contexts, such as health care, education, employment, the marketplace, and so on[…] [13, p. 3]." Key is the concept of "context": "One further feature is key to understanding what we mean here by 'contexts,' for not only are they characterized by roles and norms but also by certain ends, or values [13, p. 3]." Nissenbaum comments that "[c]ontexts are not formally defined constructs, but…are intended as abstract representations of social structures experienced in daily life [14, p. 134]." She goes on to state that "[c]ontexts may overlap and possibly conflict with one another [14, p. 136]." Even in our data-centric privacy regimes, the idea of context is implicitly assumed.

Sectoral privacy legislation, for example, assumes *a priori* that the contexts of different sectors (health, education, etc.) are foundationally different with regards to privacy. Even in omnibus privacy legislation, "purpose limitation" assumes that the *context* of the purpose determines whether the collection, use, and/or sharing of personal information is appropriate and, therefore, permissible. For example, Canada's Personal Information Protection and Electronic Documents Act permits collecting/sharing information in certain contexts, e.g., the "collection is clearly in the interests of the individual and consent cannot be obtained in a timely way [15]." Archival use is often protected as a special context. The General Data Protection Regulation (GDPR) affirms that "processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall […]not be considered to be incompatible with the initial purposes ('purpose limitation')," acknowledging the ways in which records change context when crossing the "archival threshold [16, art. 5.1(b)]." There is clear recognition even at a statutory level that privacy is contextual in nature.

The evidence of the contextuality of privacy becomes even more evident at the practical level of privacy impact assessments (PIAs). For example, the Digitial PIA template provided by the UK's Information Commissioner's Office includes many questions about context including: *Describe the context of the processing: what is the nature of your relationship with the individuals? How much control will they have? Would they expect you to use their data in this way? Do they include children or other vulnerable groups? Are there prior concerns over this type of processing or security flaws? Is it novel in any way? What is the current state of technology in this area?*

In both the archival and privacy domains, context is an element of a more foundational concept: the context of records creation in the archival domain and the context-relative information norms in the privacy domain. To cohere these understandings of context, we began mapping out concepts from the archival domain onto the notions of privacy as understood through Contextual Integrity (CI). CI, introduced by Helen Nissenbaum, is a means of understanding privacy within context by breaking it down to five parameters: sender, subject, information or data type, recipient, and transmission principle[14, pp. 140-147]. Hildebrandt summarizes: "The crucial 'constituents of a context where information is shared are defined as actors (sender, receiver, referent; which may overlap); attributes (types of information; noting that appropriateness of information flows is not one-dimensional, nor binary); and transmission principles (for instance, confidentiality, reciprocity, dessert, entitlement, compulsion, need; this entails a rejection of [simple] dichotomies such as those between access and control) [17, p. 46]."

CI posits that privacy violations can be identified by evaluating the context to determine the appropriateness of the sharing or transmission of information. Thus, it is acceptable for our doctor to send our medical information to an insurance provider so that our treatment can be covered. It is not acceptable for our doctor to share our medical information with our boss, however, as our boss is not a recipient with the appropriate role to accept such delicate information about us. In this example, while the sender, subject, and information type stayed the same, changing the recipient affected the transmission principle.

V. Making Meaning: A Context Ontology of Context

The CI conception of context consists of five parameters:

TABLE 1: PARAMETERS OF CONTEXTUAL INTEGRITY

Parameters of Contextual Integrity				
Sender	Subject	Info Type	Recipient	Transmission Principle

It becomes possible to operationalize CI in part by mapping concepts from the CI/Privacy domain, shown in *italics*, with those from the archival domain, shown in **bold**. Core to both archival and privacy understandings are "*sender"* and "*recipient*." CI describes the *recipient* as simply the entity that receives information. As such it is somewhat analogous to the archival concept of **addressee**, i.e., the recipient(s) intended by the **author** [18]. Similarly, the *sender* is simply the entity sending the information while in the archival domain the closest analogy would be the **author**, **writer***,* or **originator**, three roles that may be, but are not necessarily, combined in the same entity [18]. Regardless of whether the *sender* is the **author** or the **writer**, when dealing with the transmission of electronic records, the *sender* is always the **originator**, as they are "the person responsible for the electronic account or space in which the record is generated or from which the record is sent." The actors defined within the archival domain – **author**, **writer**, **originator**, **addressee** – "are the subjects of rights and duties, that is, they are entities recognized by the juridical system as capable of acting [18, p. 5]." In the CI context, there is no such assumption for the actors, which Kumar, Zimmer and Vitak assert can also be technologies [19, p. 21]. Archival theory certainly recognizes that information can be communicated outside of juridically relevant activities but does not provide specific definitions for the roles of such entities. There is no specific archival analogue to *subject*, a concept that is central to both CI and privacy generally.

As noted in Table 1, the concept of *attribute* is broken out into *sender*, *subject*, *recipient* and *information type*. The last, *information type*, is evident in existing privacy protection regulations, e.g., the distinction that some but not all personal information is "sensitive" as in Article 9 of the GDPR [16]. It does not have a close analogue in the archival domain, perhaps due to the fact that the business of archives is not traditionally concerned with the content of documents. That said, the concept of **act** recognizes different types of records, e.g., a contract, which when duly executed has juridical relevance, would identify the contracting parties, appropriate details of the arrangements between them, etc.

Nissenbaum describes the concept of *transmission principle* as the "terms and conditions under which such transfers [of information] ought (or ought not) to occur [14, p. 145]." Here the concept is broken out into *Aim*, *Condition*, *Modality*, and *Consequence*, with definitions from Crawford and Ostrom's "A Grammar of Institutions" provided in Table 2, below [20, p. 584]. The *Aim* of any transmission of information may be broad or specific. For example, Shvartzshnaider, et al., set three fairly specific *Aims* relating to smart devices to see how each influences consumer privacy expectations. The information is used for i) advertising; ii) legitimate business; and iii) fraud detection [21, p. 1301]. The *Conditions* used in the the Shvartzshnaider survey are if the subject has i) consented; ii) been notified; or iii) connected to a social media service [21, p. 1301]. The survey results noted that information "[f]lows conditioned on either consent or notification were viewed as more appropriate by a majority of respondents than identical flows without these conditions [21, p 1310]." Extrapolating from these sample *Conditions*, one can easily see situations where a *Sender* may be actively and intentionally transmitting information, e.g., using their credit card to buy a hat. In this scenario, most would be aware that they are also passively, but still intentionally, transmitting information, in this case to the credit card provider as a defense against fraud.

Many of us, however, become concerned when information about us is transmitted without our knowledge of the flow, the recipient or the purpose to which it will be put. While providing agency to the Subject is the focus of many approaches to privacy, legislation also exists to establish the *Modality*. For example, individuals in Ontario must, under s.125(1) of the *Child, Youth and Family Services Act, 2017*, report "any suspicions that a child is, or may be, in need of protection to a children's aid society." In such a situation, the Subject, the child, need not be aware of the transmission of information. In other situations, data recipients may simply rely on publishing a 'privacy policy' for individuals wanting to know what information they are receiving and the intended uses. The *Consequence* parameter expresses what, if any, sanctions may be imposed for violating the established *Aim(s)*, *Condition(s)*, and/or *Modality(ies)*. A person who did not report a child in need of protection could face fines or imprisonment.

TABLE 2. DEFINITION OF ELEMENTS FROM A GRAMMAR OF INSTITUTIONS THAT MAKE UP THE TRANSMISSION PRINCIPLE.

Aim	describes particular actions or outcomes to which the deontic is assigned
Condition	those variables which define when, where, how, and to what extent an *Aim* is permitted, obligatory, or forbidden
Modality [Deontic]	the three modal verbs using deontic logic: may (permitted), must (obliged), and must not (forbidden)
Consequence [Or Else]	those variables which define the sanctions to be imposed for not following a rule

It is important to note that not all four parameters may be in effect for any given *Transmission Principle*. Having only *Aim* and *Condition* present is designated by Crawford and Ostrom as a strategy [20, p. 584]. Shvartzshnaider, et al. quote Frischman, stating that a strategy can "'allow individuals and subgroups to ensure their practices are consistent with their values, even when overarching norms and rules conflict with their preferences[21, p. 1299], [22].'" A strategy becomes a *norm* with the addition of a modality. In effect, a norm is established when a community reaches a consensus regarding information flow [21, p. 1299]. When transgressing the norm results in some kind of consequence or penalty, the information flow becomes a rule.

In another modeling of privacy protection, the Interdisciplinary Privacy and Communication Model (IPCM) proposed by Bräunlich, et al. data flows are labeled "along the independent dimensions of *active versus passive* and *intentional versus unintentional* [23]." This labelling seems to tie in with Nissenbaum's categorization of Information and Communication Technologies (ICTs) into three categories: "(1) tracking and monitoring, (2) aggregation and analysis, and (3) dissemination and publication [14, p. 20] ." For example, posting a picture of the people at one's birthday party on social media is an active dissemination of personal information, although it may be passive and/or unintended by the data subjects in the image. Responses, e.g., emojis, posted in response are tracked by Facebook. Responses posted by attendees may indicate acceptance of the sharing of their personal information, while those posted by others provide additional data. Most social media users are probably aware of this, so responses might be considered intentional and passive. Such data is also aggregated, analyzed, and possibly disseminated to other agents, such as advertisers, but the subjects presumably did not intentionally transmit information for these purposes, so from their perspective these would be passive and unintentional. These labels would change when considering the same scenario from the perspective of Facebook, for example.

Well established *transmission principles* include "confidentiality, dessert (deserving to know), entitlement, compulsion, need, voluntary, notice, consent, exchange, reciprocity, anonymity, temporality, mutuality, requirement, and secrecy" but Nissenbaum notes that the "list [of principles] is probably indefinite, particularly if we allow for nuanced and complicating variations [14, p. 145]." Recognizing the ambiguity of the concept, Kumar, Zimmer and Vitak suggest approaching "the transmission principle as the sentiment that follows the expression, 'This information flow is fine IF...[19, p. 11].'" They also note that some "have linked medium with the transmission principle parameter, while others suggest it may need to be a separate parameter [19, pp. 21–22]."

Understanding that the "medium is the message[25]," we argue that it is necessary to evaluate the **medium**, i.e., the system or means in which information is shared, as an aspect of the transmission principle. There needs to be an assessment of what media are appropriate for what kinds of information. For urgent information, for example, is postal mail appropriate? Additionally, the archival concepts of **act** and **function** can enhance existing CI parameters. **Functions** are "[a]ny high level purpose, responsibility or task assigned to the accountability agenda of a corporate body by legislation, policy or mandate. Functions may be decomposed into sets of coordinated operations such as subfunctions, business processes, activities, tasks or transactions [7]." Records created in the course of completing activities in relation to a designated **function**, e.g., a press office, might be considered unlikely to have sensitive information (or sensitive only for a short period) and *norms* of transmission would likely be fairly expansive. By contrast, patient records held by a psychiatrist would likely be highly sensitive, and remain so over the long-term, with very different *norms* relating to sharing.

Duranti and Thibodeau identify six categories of records: dispositive, probative, supporting, narrative, instructive, and enabling records [26]. The first two categories include records like contracts and receipts that "are required by the juridical–administrative system within which they are created [26]." Supporting and narrative documents, which perhaps comprise the greatest volume of current records, simply support the action of which they are a part or are simply shared information. Instructive records include manuals, regulations, instructions, etc., while enabling records include software (patches to maintain security, analyze data, etc.), workflows, etc. These conceptions of human activity would contribute to identifying reasonable *norms* of transmission. For example, a user manual would likely have had a wide audience and as a result would be unlikely to violate a privacy norm, especially if no longer current – unless, perhaps, for a nuclear reactor.

Integrating CI and archival diplomatics yields an initial framework as shown in Table 3 below:

TABLE 3: CI AND ARCHIVAL DIPLOMATICS INITIAL FRAMEWORK

Primary Contextual Integrity Elements	**Secondary Contextual Integrity Elements**	**Primary Diplomatics Elements**	**Secondary Diplomatics Elements**
Data Type/ Attribute		Nature of Record	Private
Data Subject			Public
Sender		Persons	Originator/ Creator
			Writer
Recipient			Author
			Addressee
Transmission Principle		Acts	Dispositive Document
			Probative Document
			Narrative Document
			Supporting Document
			Instructive Records
			Enabling Records
		Types of Procedure	Executive Procedures
	Norms of Appropriateness		Instrumental Procedures
	Norms of Dissemination		Organizational Procedures
			Constitutive Procedures
	Aims (GKC-CI)	Form	Intellectual Form
			Physical Form
	Conditions (GKC-CI)	Facts	
	Consequences (GKC-CI)		Original document
			Authentic copy
			Draft
			Simple copy

VI. Applying Privacy Concepts to Archives

Where access must be provided to records, as distinct from data, application of a context-centric approach to privacy protection holds promise, particularly when one considers how privacy is not binary but rather a reflection of the nuances inherent in how individuals share personal information, depending upon transmission principles. For archives, the transmission principles are complex, as the ethical codes of the archival community incorporate a broad range of stakeholders. Archivists are expected to "have regard to the legitimate, but sometimes conflicting, rights and interests of employers, owners, data subjects and users, past, present and future [27]." Given the breadth of stakeholders, archivists must consider in the course of their work, and the fraught nature of archiving in the public interest, using AI sense-making tools and techniques that are informed by extensions of existing archival concepts may help custodians of unfamiliar document collections to make better-informed and thus more responsible decisions on how they are shared. To that end, we examined whether AI tools, specifically, LLMs informed by knowledge graphs, could help identify the contexts of archival records, allowing archivists to make more informed decisions about privacy and access. Specifically, we have created a context-centric pipeline using graph analytics, machine learning and retrieval augmented generation, using the underlying archival-privacy contexts as an ontology.

This pipeline addresses concerns of archivists such as context-dependent realities for effectively protecting privacy and integrating archival- and privacy-based concepts of context. As in archives and privacy, context is an active area of research in AI. Serafini and Bouquet noted in 2004 that "[t]he problem of context has a long tradition in different areas of artificial intelligence (AI) [28, p.41].” Originally pioneered by John McCarthy, who argued that “Whenever we write an axiom, a critic can say that the axiom is true only in a certain context, [29]” context formalization remains an active area of research in AI [30]. Indeed, work on the formalization of context in AI has led to the development of the sub-discipline of Context Engineering, which, like contextual integrity and archival contexts, provides means of identifying, making explicit, and making use of the contexts of information. A key component of context engineering, a “knowledge graph,” contextualizes information by representing the relationships between entities, providing semantic meaning to those relationships.

A knowledge graph contains entities (“nodes”), with lines between those nodes illustrating their relationships (“edges” or “links”). The nodes can stand in for any concept or entity that could coherently be put into relations with other entities in the same context. In Figure 4, we have an example knowledge graph. At the center, in pink, is Alice. The edges represent relationships, labeled by what they stand for. In this example, Alice is employed by Company in Role; because her Role needs access to System, Alice has read/write access to System. Alice and Bob are friends with each other, as represented by the relationship arrows going both ways. We can see different entities can have different relationships with the same entity, as Bob is a client of Alice’s employer, and is a data subject within the System that Alice accesses due to her role.

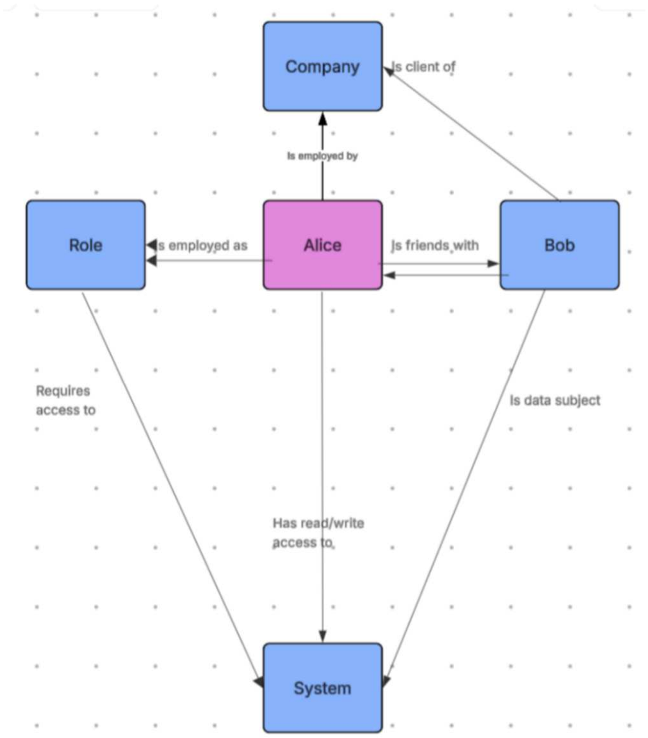

Fig. 4: A knowledge graph

Thus, knowledge graphs can be used to represent knowledge about a phenomenon in a graphical and succinct way. Structuring knowledge in this way can help in efficiently retrieving information and in disambiguating between multiple pieces of similar information. This is important in identifying potential privacy concerns based on the contexts of archival fonds. Knowledge graphs can be used with techniques such as retrieval augmented generation (RAG), wherein LLMs are combined with external knowledge sources to improve response quality and to overcome limitations of LLMs, including the “lost in the middle” problem, whereby information in the middle of long prompt gets lost, and the length limitations on prompts and the context that can be provided through the prompting process.

For this use case, a knowledge graph allows us to instantiate archival-privacy context concepts and relationships as an ontology. One particularly promising methodology for this use case is GraphRAG. Usually, when a question is asked of an LLM, it is asked directly. For example, if you ask “Where do llamas live?” an LLM would answer either based on its current knowledge base or would perform a search online to retrieve the answer. However, if an expert on Llamas knows the answer to that question is highly specific and wants to ensure the LLM returns it accurately, the expert can supply extra information in the form of a knowledge graph. Then, the LLM would search inside the knowledge graph for the answer.

In GraphRAG, when the question is asked, it is transferred first to another tool before being sent to an LLM. This tool performs a smart search within the knowledge graph for information relevant to the question. The information that it finds is sent, along with the question, to the LLM. The accompanying information guides the LLM in producing the final answer. The context provided by the knowledge graph makes the returned answer more likely to be correct and relevant. However, this still does not guarantee the LLM answers perfectly – this solution would still require an archivist-in-the-loop. Furthermore, as the purpose of this solution is to identify potential privacy concerns within record sets, it is important to continue to take appropriate precautions, such as working only on local machines and imposing access restrictions.

VII. Mapping context

Hereafter, a two-part architecture is outlined, see also Fig. 5. Firstly, a knowledge graph is built from user-provided files that contain contextual information. The knowledge graph begins top-down, with the ontology of archival-privacy contexts applicable to all collections, along with additional information (initial tests, for example, have shown the need to incorporate extensive information about the concept of provenance). It is further developed through the fonds under consideration and other relevant files. The knowledge graph constructed from the files represents the relevant information. Then, that knowledge graph is used in the second step (the GraphRAG) to answer questions.

In applying GraphRAG, the first step is to select which retriever(s) to use within the knowledge graph. The type of retriever(s) determine(s) how information inside knowledge graphs is searched. The primary choices are vector retrievers, which compute similarity in the vector representation of words based on similar meanings, and graph retrievers, which compute similarity by traversing the knowledge graph to find conceptually related words. In other words, a vector retriever might group monkeys, apes, and humans as having similar meanings; the same for apples, bananas, and oranges. A vector retriever would likely not associate monkeys with bananas. A graph retriever, however, may "find" a relationship of monkeys (node) like (relationship) bananas (node). It is also possible to do a hybrid retriever, wherein the results of a vector search of the files and a graph search of the knowledge graph are concatenated to provide the context to the LLM that will answer the query.

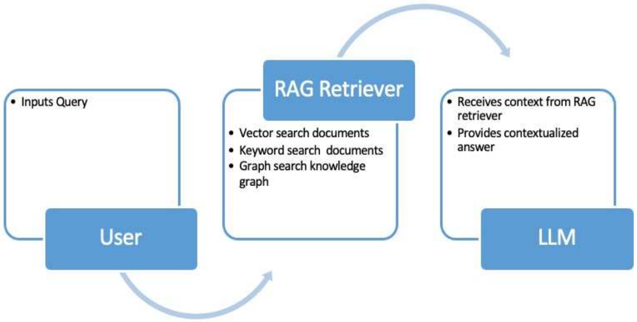
Inputs Query
User
RAG Retriever
Vector search documents
Keyword search documents
Graph search knowledge graph
Receives context from RAG retriever
Provides contextualized answer
LLM

Fig. 5: Two-part architecture

When put all together, the GraphRAG process looks like Fig. 6:

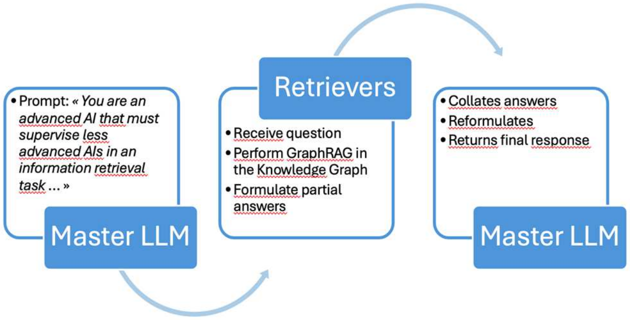
Prompt: « You are an advanced AI that must supervise less advanced AIs in an information retrieval task ... »
Master LLM
Retrievers
Receive question
Perform GraphRAG in the Knowledge Graph
Formulate partial answers
Collates answers
Reformulates
Returns final response
Master LLM

Fig. 6: GraphRAG process

There are two main elements. There is an input question. Every single LLM can be prompted. The prompting technique modifies its behavior. There is a straightforward prompt that is given to the last LLM that combines all the answers into a final one. It is a fundamental step. GraphRAG retrieves the information inside the knowledge graph. It first does a vector retrieval. Then it does a graph search. Finally, from the information that it finds, it formulates an answer, a text generation activity.

An initial experiment with the GraphRAG (using Neo4J and Llama 3.3) asked of a subset of publicly available emails regarding the Enron scandal: *What are the creating entities (creator) of the documents where the creator is determined by the concepts of provenance and fonds?* The knowledge graph included information about provenance and fonds. The GraphRAG returned:

"The documents were created by two types of entities: individuals and a corporate body. The individual sender creator is Phillip K Allen, while the corporate body is Enron, as indicated by the email addresses and references to Enron personnel and activities. Other Enron employees, such as Leah Van Arsdall, John J Lavorato, and Tim Belden, are also associated with the creation of these documents through their work at Enron."

As we can see, although the final answer is a bit vague (an issue that seems quite common from our experiments), the main elements of the answer were properly returned by the LLM. Despite the inexactness of the response, our imaginary archivist now has important information that a) would be hard to collect manually, and b) confirms (in this case) the archivist's supposition that the creator of the aggregation of documents is the Enron Corporation. The archivist can now proceed on that assumption, using further use cases to elaborate a picture of the persons (actors) represented and the activities being documented, essential information when considering privacy risks. Had the model returned a response that entities in addition to (or other than) Enron were responsible for creating the documents, our archivist would know to investigate further.

VIII. Limitations & Future Work

This paper presents the initial framework development for a context-based GraphRAG to identify potential privacy concerns in archival collections. While initial experimental outcomes have been encouraging, they are limited, as several conditions need to be further developed and tested, including optimizing the formalization of the ontology, evaluating the quality of different LLMs, and testing the most effective retrieval strategies within the RAG retriever. Future work also involves targeting agent architectures, aiming for a more sophisticated approach. For example, agents can be forced to follow a hard logical system or graph-based logic to review their answers, share them between agents or themselves, and provide feedback to one another. The agents can use an LLM as a judge to determine whether an answer is relevant in the context of privacy. It can also tie with the idea of putting the human in the loop. Meaning the humans can review the whole process so that they can help evaluate the quality of the answers before they are returned, hence improving the quality of the process. By definition, the capabilities of agents are bounded by the abilities of the LLMs they are based on, so the quality of the output of a sophisticated agent-based architecture still scales with the power of the server or machine used. .

IX. Conclusion

For archives to continue to serve their many purposes and users, solutions must be found that allow archivists to responsibly manage the tension between access and privacy to vast fonds that no human could ever hope to review manually. While innovative solutions are being pursued to provide access to archival holdings as data, providing access to records as records requires identifying privacy risks based not just on content, but on context. By combining three approaches to context – archival contexts, the contextual integrity theory of privacy, and knowledge graphs to provide context to LLMs – this work advances an automated approach to privacy that can support the archivist-in-the-loop, even in limited resource situations.

Acknowledgment

This paper draws on research conducted as part of InterPARES Trust AI and supported by the Social Sciences and Humanities Research Council of Canada.

References

[1] International Association of Privacy Professionals (IAPP), “Privacy,” *Glossary of Privacy Terms*. Accessed: Nov. 09, 2025. [Online]. Available: https://iapp.org/resources/glossary/

[2] InterPARES Trust AI, “Privacy,” *Terminology Database*. Accessed: Nov. 09, 2025. [Online]. Available: https://interparestrustai.org/terminology/term/privacy

[3] A. Narayanan, “Data Privacy: the Non-interactive Setting," the Graduate School, University of Texas at Austin, Austin, TX 2009. Available: https://repositories.lib.utexas.edu/server/api/core/bitstreams/b782b7e3-9ec5-47ab-b9c5-93cd35ebcbef/content

[4] P. Ohm, “Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization,” *UCLA LAW REVIEW*, vol. 57, no. 6, pp. 1701–1777, 2010.

[5] B. C. Kara, C. Eyupoglu, and O. Karakuş, “(r, k, ε)-Anonymization: Privacy-Preserving Data Publishing Algorithm Based on Multi-Dimensional Outlier Detection, k -Anonymity, and ε -Differential Privacy,” *IEEE Access*, vol. 13, pp. 70422–70435, 2025, doi: 10.1109/ACCESS.2025.3559410.

[6] M. Kobayashi, A. Fujioka, K. Chida, A. Nagai, and K. Yasuda, “k^m-anonymization Meets Differential Privacy under Sampling,” *Journal of Information Processing*, vol. 33, no. 0, pp. 646–656, 2025, doi: 10.2197/ipsjjip.33.646.

[7] M. J. Bates and M. N. Maack, Eds., “Diplomatics,” in *Encyclopedia of Library and Information Sciences, Third Edition*, 0 ed., CRC Press, 2009, pp. 1593–1601. doi: 10.1081/E-ELIS3-120043454.

[8] G. Yeo, “Trust and context in cyberspace,” *Archives and Records*, vol. 34, no. 2, pp. 214–234, Oct. 2013, doi: 10.1080/23257962.2013.825207.

[9] T. Cook, “Evidence, memory, identity, and community: four shifting archival paradigms,” *Archival Science*, vol. 13, no. 2, pp. 95–120, 2013, doi: 10.1007/s10502-012-9180-7.

[10] D. Mordell, “Critical Questions for Archives as (Big) Data,” *Archivaria*, vol. 87, no. 87, pp. 140–161, 2019.

[11] V. L. Lemieux and J. Werner, “Protecting Privacy in Digital Records: The Potential of Privacy-Enhancing Technologies,” *J. Comput. Cult. Herit.*, vol. 16, no. 4, pp. 1–18, Dec. 2023, doi: 10.1145/3633477.

[12] V. L. Lemieux *et al.*, “Clio-X: AWeb3 Solution for Privacy-Preserving AI Access to Digital Archives,” 2025, *arXiv*. doi: 10.48550/ARXIV.2507.08853.

[13] A. Barth, A. Datta, J. C. Mitchell, and H. Nissenbaum, “Privacy and contextual integrity: framework and applications,” in *2006 IEEE Symposium on Security and Privacy (S&P’06)*, Berkeley/Oakland, CA: IEEE, 2006, p. 15 pp. – 198. doi: 10.1109/SP.2006.32.

[14] H. F. Nissenbaum, *Privacy in context: technology, policy, and the integrity of social life*. Stanford, Calif: Stanford Law Books, 2010.

[15] Canada, *Personal Information Protection and Electronic Documents Act (PIPEDA)*, vol. c. 5. 2000.

[16] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*.

[17] M. Hildebrandt, “Location Data, Purpose Binding and Contextual Integrity: What’s the Message?,” in *Protection of Information and the Right to Privacy - A New Equilibrium?*, L. Floridi, Ed., Cham: Springer International Publishing, 2014, pp. 31–62. doi: 10.1007/978-3-319-05720-0_3.

[18] L. Duranti, *Diplomatics: new uses for an old science*. Lanham, Md: Scarecrow Press, 1998.

[19] P. C. Kumar, M. Zimmer, and J. Vitak, “A Roadmap for Applying the Contextual Integrity Framework in Qualitative Privacy Research,” *Proc. ACM Hum.-Comput. Interact.*, vol. 8, no. CSCW1, pp. 1–29, Apr. 2024, doi: 10.1145/3653710.

[20] S. E. S. Crawford and E. Ostrom, “A Grammar of Institutions,” *Am Polit Sci Rev*, vol. 89, no. 3, pp. 582–600, Sept. 1995, doi: 10.2307/2082975.

[21] Y. Shvartzshnaider, M. R. Sanfilippo, and N. Apthorpe, “GKC ‐ CI : A unifying framework for contextual norms and information governance,” *Assn for Info Science & Tech*, vol. 73, no. 9, pp. 1297–1313, Sept. 2022, doi: 10.1002/asi.24633.

[22] B. M. Frischmann, M. J. Madison, and K. J. Strandburg, Eds., *Governing Knowledge Commons*. Oxford University Press, 2014. doi: 10.1093/acprof:oso/9780199972036.001.0001.

[23] K. Bräunlich *et al.*, “Linking loose ends: An interdisciplinary privacy and communication model,” *New Media & Society*, vol. 23, no. 6, pp. 1443–1464, June 2021, doi: 10.1177/1461444820905045.

[24]

[25] M. McLuhan, *Understanding media: the extensions of man*, 10. print. in Media sociology. Cambridge, Mass.: MIT-Press, 2002.

[26] L. Duranti and K. Thibodeau, “The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES*,” *Arch Sci*, vol. 6, no. 1, pp. 13–68, Oct. 2006, doi: 10.1007/s10502-006-9021-7.

[27] “ICA Code of Ethics,” ICA. Accessed: Nov. 09, 2025. [Online]. Available: https://www.ica.org/resource/ica-code-of-ethics/

[28] L. Serafini and P. Bouquet, “Comparing formal theories of context in AI,” *Artificial Intelligence*, vol. 155, no. 1–2, pp. 41–67, May 2004, doi: 10.1016/j.artint.2003.11.001.

[29] J. McCarthy, “Generality in artificial intelligence,” *Commun. ACM*, vol. 30, no. 12, pp. 1030–1035, Dec. 1987, doi: 10.1145/33447.33448.

[30] L. Mei *et al.*, “A Survey of Context Engineering for Large Language Models,” 2025, *arXiv*. doi: 10.48550/ARXIV.2507.13334.