Using LLMs for Semantic Alignment: A Study on Archival Metadata Description

Abstract

The advantages of aligning custom data schemas with standardised ontologies within their respective knowledge domain have long since been proven in practice. Sharing a common structural representation by mapping concepts and relationships between the schemas is essential to ensure data interoperability (especially on a semantic level), integration, reuse, and the ability to leverage machine-processable and advanced-search capabilities. Archival institutions preserve, manage, and provide access to large amounts of diverse cultural and historical data, demonstrating a high potential to be active contributors to a global knowledge network, should archival data be transformed and offered as linked (open) data. Based on the expert-validated dataset of the mapping (alignment) of the Swedish National Archives schema to the Records-in-Contexts (RiC-O) ontology, the purpose of this study is two-fold. First, to examine whether it is possible to automatically and effectively extend one case (Sweden) to other archival institutions and align new custom schemas to RiC-O, given an expert-curated dataset of this domain. Secondly, using the aforementioned dataset and one more of a few human-evaluated examples of mapping to other cultural heritage ontologies as input, to examine whether an LLM (e.g., GPT-4o) is capable of recommending meaningful alignments for enhanced metadata description to more ontologies within the same domain (CH and archives), but also across other domains. The experiments reveal several challenges and shortcomings of the LLM prompting approach for these tasks, but also possible opportunities to leverage towards this direction.

Keywords *Archival data, archives, cultural heritage, semantic interoperability, reusability, ontology mapping, ontology alignment, linked open data, linked data, digital humanities, digital archives, LLMs, GPT-4*

1. Introduction

The advantages of aligning custom metadata schemas with standardised ontologies within their respective knowledge domain have long since been emphasised [1]. Sharing a common structural representation by mapping concepts and relationships between the schemas is essential to ensure data interoperability (especially on a semantic level), integration, reuse, and the ability to leverage machine-processable and advanced search capabilities. Archival institutions preserve, manage, and provide access to large amounts of diverse cultural and historical data, demonstrating a high potential to be active contributors to a global knowledge network, should archival data be transformed and offered as linked (open) data. Digital archival collections are growing larger as mass digitisation takes place, accompanied by an equally increasing volume of born-digital archives over the past two decades [2]. However, mass digitisation often lacks the equivalent progress in digital humanities requirements for structured, discoverable, and interoperable data, bringing forth the fundamental necessity for Linked Data practices for this purpose [2].

The Swedish National Archives [3] (in Swedish: Riksarkivet), is the responsible institution for Sweden's documentary and historical heritage preservation, and one of the oldest archival institutions globally, tracing its foundation back in 1618. The archives cover a wide variety of documents, ranging from state, military, legal, genealogical, and regional documents, to royal charters, maps, letters and medieval scripts and books. Millions of archival records (amounting over 75 km of physical artifacts) are available and, although a small percentage of archives is digitised, due to the immense size of the archival collection, the digital archives still amount to over 100 million digital artifacts, available (some might fall under classification restrictions) through the Swedish National Archives' portal and search function. The digitisation process is still ongoing, aiming to make archives more accessible digitally, while taking into consideration the adoption and sharing of new methods and technologies for archival data management.

This research uses the outcome of the conceptual mapping between the Records-in-Context Ontology (RIC-O) and the schema in use at the Swedish National Archives as a basis for further experimentation on the potential of introducing automated work streams in the mapping process. The process of schema alignment may vary greatly from case to case and depending on the schema of both the institution and the ontology (or ontologies involved), some of them might reach too high of a level of complexity for human actors to process. Apart from the time constraints, human actors can easily get entangled in the complexity of a high interconnectedness between classes and attributes, leading to an incomplete or insufficient result. In fact, in many cases it could act as a deterring factor for even attempting this process. Existing tools and documentation might help in this regard, however, for most of them to function effectively and provide valuable mapping recommendations, still a lot of effort needs to be devoted by a human actor to input proper information, consider the peculiarities of each schema and case, and guide the tools towards the desired direction. So, the purpose of this study, apart from the alignment of a custom metadata schema to a standardised one, is to leverage a local case study towards experimentation which can potentially bring forth value in similar cases, as well as showcase how the results and produced data can further expand this research beyond the boundaries of a single case scenario. The experiment of this study is two-fold; the expert-validated mapping between the Swedish National Archives and Records-in-Contexts (RiC-O) is used as prior knowledge on GPT-4o in order to, first off, similarly perform the mapping given another custom archival institution schema and, secondly, to assist in recommending alignments with other schemas of the Cultural Heritage, but also other related, domains, such as Linked Open Vocabularies (LOV), in order to foster re-use and enhance semantic granularity for metadata description, wherever feasible.

This document is structured as follows: Section 2 (Background) includes the research basis and motivation of this study, as well as introductory information about the Swedish National Archives structural schema. Section 3 (Method) presents the methodological steps followed for the two-fold experiment, and Section 4 (Results) is where the outcomes of the experiments are presented and analysed. Section 5 (Discussion) includes the analysis and insights gained from the results, their potential use and re-use by practitioners of the field, and, finally, Section 6 refers to the limitations present in this work, along with future directions it can follow.

The present study is a practical application and continuation of the work presented at the Extended Semantic Web Conference - ESWC2025: "Empowering Knowledge Through

Semantics: From Knowledge Graphs to Neurosemantics" [4].

2. Background

Digital data is a core resource for digital humanities; however, digitalised archival data also needs to be integrated, interoperable, and interrogable (Bikakis, 2021) [5] and align with the FAIR guiding principles of data publication (Findable, Accessible, Interoperable and Re-usable) [6]. Linked data allows for the facilitation of FAIR data implementation, creating Archival Linked Data (ALD), which is machine-readable, contextual, and can be analysed using digital humanities (and beyond) methods for research and engagement, but despite the traction, linked data still remains under-examined in an archival data context [2]. Linked archival data shepherds numerous benefits; improvements in knowledge discovery, information retrieval, revealing unknown relationships across archival collections (even cross-domain navigation between cultural and non-cultural heritage data sources) [7], [8], improving data quality, and enabling semantic search queries (SPARQL) [9], are but a few of them, enabling digital humanists and users to more easily understand and use archival data in context [2].

The quest for making archival records' metadata available as Archival Linked Data (ALD) involves standardisation and alignment with archival descriptive standards and ontologies, such as the Records in Contexts Ontology (RiC-O) [10], unification with the global network of OpenGLAM for GLAM (Galleries, Libraries, Archives and Museums) [11] or Europeana's Linked Data web service [12]. Linked Open Data use in cultural heritage paves the way towards application also in the archival domain, with RDF and ontological approaches being sine qua non for semantic interoperability and better knowledge management for archival data [13]. RDF (Resource Description Framework) is the basis for Linked Data structure and expressed in triples: a semantic unit consisting of three components: subject -> predicate -> object. Archival resources may follow international standards, such as EAD (Encoded Archival Description), an XML standard for the encoding of finding aids for use in an online environment [14], the web indexer Schema.org model, or more specialised models such as LODE (Linking Open Descriptions of Events) [15].

However, to achieve contextual archival description in a semantically valuable manner, it is necessary to identify resources by the means of dereferenceable URIs, standardised descriptions and relations format, and linked descriptions to other information resources to the largest possible extent [13]. In general, the hierarchical structure of archival documents, according to the ISAD(G) - General International Standard Archival Description [16] follows the pattern shown in Fig. 1.

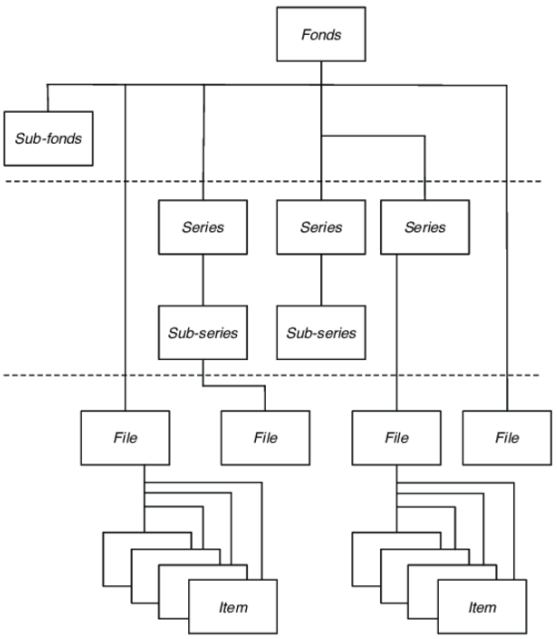


Figure 1. Archival document hierarchical structure according to ISAD(G) [16]

At the same time, archives usually contain diverse types of data and materials (letters, maps, paintings, books, photographs, sound recordings, etc.), a characteristic introducing challenges in archival description and the ability to sufficiently capture the richness and depth of each individual case, considering that existing ontological approaches and schemas differ in scope and descriptive granularity [15]. For instance, archival descriptions often offer less granularity compared to linked data vocabularies (e.g., vocabularies available in the Linked Open Vocabularies repository [17] or similar.). Modeling approaches and choosing among standards to form an appropriate metadata schema differ according to each institution's individual needs and data infrastructure. Chen (2019) [18] explored various methods of semantic enrichment for art-related archival resources and chose the Europeana Data Model (EDM) as the core data model design, while also following additive approaches such as direct reuse of other external vocabularies, local links to other data sources, introduction of contextual classes, and the utilisation of named entity extraction. Earlier studies, such as the research by Bountouri and Gergatsoulis (2011) [19] demonstrate different approaches in representing archival hierarchies and instead mapping the corresponding EAD (Encoded Archival Description) elements to the CIDOC Conceptual Reference Model (CIDOC CRM) for cultural heritage metadata integration. CIDOC CRM and additional ontologies were also encompassed in an archival context of a linked open data model for the Portuguese Archives in which case, due to the lack of deployment testing of the newest Records in Contexts Conceptual Model (RiC-CM) by ICA (International Council on Archives), it was deemed more suitable to turn to CIDOC-CRM's long maturity for element representation [20]. On the other hand, RiC-CM's most recent models (e.g., Ric-O), apart from the archival intrinsic structure, also feature a larger collection of properties to describe archival relations. RiC-O's adaptability to various archival contexts and connectability to other cultural heritage domains is also showcased by the several semantic modeling projects it is used in, including projects and initiatives led by the National Archives of France [21].

As far as the involvement of Large Language Models in semantics and ontology engineering is concerned, the symbiosis of humans and machines as a new domain to explore was brought up by Doumanas et al. (2024) [22], who performed LLM-based ontology engineering but at the same time observed that human contribution and involvement notably enhanced the process and the results. A similar statement was supported by Osman et al. (2024) [23], who emphasise the benefits and generalisability of having a well-designed, semi-supervised approach in ontology matching, after experimenting with automatic approaches and observing that fully automated solutions are still unreliable. Hoseini et al. (2024) [24] utilised the natural language processing capabilities of LLMs to label and model data semantically in the context of data spaces, while Pan et al. (2025) [25] presented an LLM-based retrieval-augmented generation (RAG) approach for automatic generation of competency questions in ontology engineering. Pan et al. (2025) [25] also observed that adding domain knowledge to their RAG process improved LLM performance in this task. Cigliano and Fallucchi (2025) [26] identified the potential of an intersection among open and linked data, ontologies, and LLMs, and support the statement that this combination may revolutionise how value from data and structured information can be derived. The potential of utilising LLMs for domain-specific (Cultural Heritage) but also cross-domain recommendations for alignment and evaluation was also studied by Maratsi et al. (2024) [27], and the potential of a proposed methodological framework combining automated means (LLMs), ontological foundation, and graph theory metrics for improving semantic interoperability and interdisciplinary discoverability of data as an enhanced semantic search capability, was presented by Maratsi et al. (2025) [28].

Following the latest advancements in LLM-enabled processes for semantic interoperability and ontology mapping, and taking into consideration the observed value of human-in-the-loop in such a process, the present work aims to combine human and machine expertise for ontology alignment and metadata enrichment in the archival and cultural heritage domain using LLMs (GPT-4o) by first exemplifying the mapping based on a priori human-validated manual mapping process. The ultimate goal of this venture is to pave the way towards technologically enabled ontology mapping, not only within the same domain, but also enabling cross-domain ontology and vocabulary re-use (such as vocabularies available at the Linked Open Vocabularies repository).

3. Method

3.1 Prior Knowledge - Groundwork

The groundwork for this study lies and builds on the conceptual alignment between the Swedish National Archives (abbreviated RA from “Riksarkivet”) and Records-in-Contexts (RiC-O) performed by Maratsi et al. (2025) [4]. The expert-validated dataset which resulted from this process is used to anchor a measurement of “truth” for the expected results and lead the LLM throughout the prompting interaction.

Before presenting the methodological process followed for the experiments of this study, a brief reference to the basics of the method followed for the prior semantic alignment between RA and RiC-O is made. As described in [4], to facilitate the conceptual mapping, the first step was to create a general taxonomy of archival data organisation in the Swedish National Archives, a

mental model to help visualise the concepts schematically. The RA taxonomy consists of the main archival document hierarchical structure, very closely following the ISAD(G) structure (see Fig. 1), with the differentiation, however, that RA includes the concept of "Volume (in Swedish: Volym)" which refers to a box of documents or items and is part of an archival Series (in Swedish: Serie), which is part of an Archival fond. Apart from the archive hierarchical structure, the taxonomy includes the types of archival resources (as these are organised in the RA search function), the archival institution responsible for the archive, the archivist (responsible for the insertion of archival entries), the archive contributors, and the places (and types of places) the archives derive from. For the archival record description there are several metadata groups, such as basic metadata (reference code, date, dimensions, access rules, archival institution etc.), metadata related to the content of the record, control and actions metadata (latest modification, source year, source, etc.), accessibility metadata, relevant records, and other (e.g., notes). The names of the items of the RA schema derive from their database schema (ARKIS) and they are denoted in both Swedish and English (the original fields are in the Swedish language).

The overall methodological process followed for the mapping is shown in Fig. 2. It is initiated by extracting all Classes, Data Type Properties and Object Properties of the Records-in-Contexts ontology (RiC-O), as well as the main Classes and Attributes of the RA schema. The classes and properties of RA were documented accompanied by a field description in English and Swedish. The list of fields from the RA schema involved in the mapping is not exhaustive but contains the necessary concepts to describe archival records and the most important entities and related actions to put them in context. The mapping matrix (set of spreadsheets) is then prepared, organised in a Class-to-Class mapping and an Object Properties-Relations mapping, the latter including Domain and Range to express the directive (or symmetric if this is the case) relations.

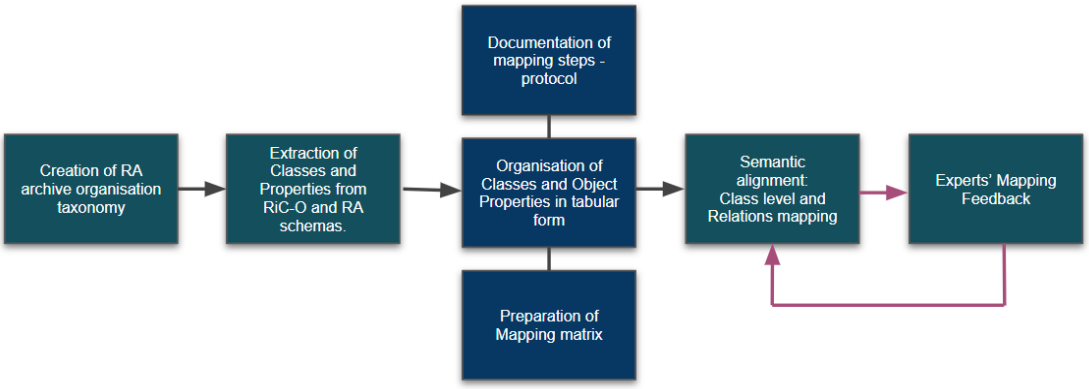


Figure 2. The mapping process to align RA and RiC-O [4]

The mapping between RA and RiC-O was performed by first identifying which RiC-O classes align well with the RA classes, so a mapping on Class level. After the Class level mapping, the identification of all related Object Properties and Relations was performed, keeping the relations in RiC-O that bear meaning for RA and can be reused to express context. This process leads to the creation of a set of class mappings and a set of object properties and relations mapping, the latter following an RDF triple structure in principle, e.g.,: **Document** *includesOrIncluded* in

Series, **ReferenceCode** *isOrWasIdentifierOf* **ArchiveFond**, or **Agent** *isOrWasManagerOf* **ArchiveFond**, where the first part of the triple is the Domain of the relation, the second part is the relation which connects the subject and the object of the triple, and the third part is the Range, so the passive part of the triple. The results of the mapping process described were evaluated by a group of experts during an organised Workshop session at the Swedish National Archives headquarters, where the produced mappings were shared with the group of experts in order to assess their suitability, level of alignment, and integrate feedback in the loop.

3.2 LLM (GPT-4o and GPT-4.5) Experiments

The two-fold experiment of this study includes the following two cases/scenarios where GPT-4o (and GPT-4.5 for the first one) was asked to perform.

i) Given the full, expert-validated dataset of the mapping between RA and RiC-O, and a custom archival institution schema, can the LLM perform the mapping between RiC-O and the new schema effectively?

ii) Given the full, expert-validated dataset of the mapping between RA and RiC-O and a few human-evaluated examples of recommended mapping between some RA elements and elements from other standard schemas (e.g., Linked Open Vocabularies - LOV), can the LLM similarly output meaningful recommendations to more vocabularies, according to a concept-in-word context?

The experiment set-up for the two scenarios is shown on Fig. 3.

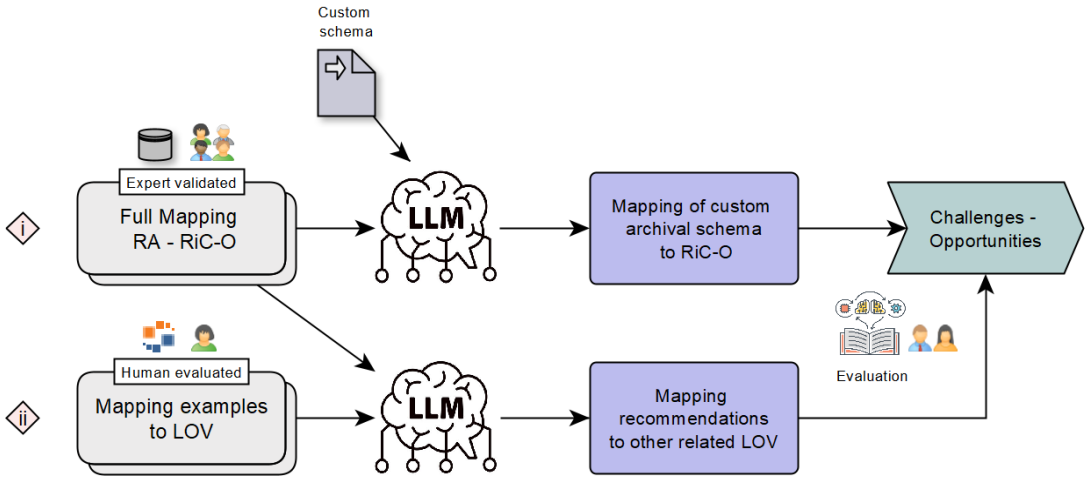


Figure 3. The methodological set-up for the two-fold LLM experiment

In the first case, the aim is to test the performance of the model in undertaking the role of a human expert in order to semantically align a given, custom archival institution schema with RiC-O without any prior knowledge (zero-shot) and with having the full expert-validated mapping (RA and RiC-O) as an example to anchor the process. In the second case, the aim is to test the ability of the model to identify relevant concepts in other schemas within proximal domains (e.g., Cultural Heritage, Archives) but also across other domains (e.g., Geospatial information, Arts, etc.) and recommend semantically interesting cases for re-use in metadata

description. Apart from aligning part of the RA schema (ARKIS) with the Records-in-Contexts Ontology (RiC-O), other ontologies in the Cultural Heritage (CH) domain could potentially be reused to enhance representation expressivity in cases where RiC-O classes do not live up to the same level of granularity. For the most generic descriptions, the Swedish National Archives already reuses fields from schema.org, such as schema:ArchiveComponent, schema:Creator, or schema:Identifier, however, for deeper level of granularity there is currently no reuse from other ontologies in the Archival or Cultural Heritage domain apart from RiC-O. RiC-O is an ontology which emphasises and offers a variety of relations and attributes for archival record description, but it does not primarily focus on a rich representation of an artifact itself. Other ontologies, for instance the Context Description Ontology (ArCo network) (cdesc), the Denotative Description Ontology (ArCo network) (ddesc), and the LinkedGeoData ontology (lgdo) could lend some of their classes to the RA schema. All these ontologies are also available in the Linked Open Vocabularies (LOV) repository. In other words, the model is expected to make recommendations on LOV re-use for concepts which can be more accurately expressed by borrowing concepts from other knowledge domain schemas, thus potentially facilitating an ontological multi-domain re-use.

Prompt preparation

The 1st experiment was performed on OpenAI's GTP-4o and GPT-4.5. The 2nd experiment was performed on GPT-4o. During the 1st experiment, the model was asked to conduct the mapping between a custom archival institution schema and RiC-O with and without prior knowledge. During the 2nd experiment, the model was asked to consider a dataset of validated mapping examples between some RA elements and ones borrowed from the Linked Open Vocabularies (LOV) ecosystem in order to similarly suggest more cases for re-use both within the CH domain and beyond. First the model was asked to make further LOV recommendations on the sample given (Appendix 3) and then extend the recommendations in a similar way to the full RA dataset (Appendix 1).

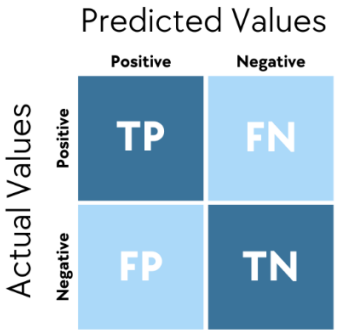
Zero-shot trial
For the first experiment, the model was initially asked if it is familiar with RiC-O and whether, given a custom data element set from an archival institution, it can map the concepts and make it compliant with RiC-O. In this case, no more contextual information was provided. This task was performed on both GPT-4o and GPT-4.5.

Informed trial
Following the zero-shot trial, the model was then asked to perform the same process, given prior knowledge to consider. In this case, the full, expert-validated mapping between RA and RiC-O was given as input, including both class–to-class and properties mapping. This task was performed on both GPT-4o and GPT-4.5.

Evaluation technique

The outputs of experiment 1 were initially human-evaluated by the authors, who were considered to be domain experts in this case, since they have manually performed this process before and have the experience necessary for a preliminary (not case-specific) judgement. The metrics used to evaluate the outputs were accuracy, precision, recall, and F1 score, which were separately

calculated for each experiment set-up (for the zero-shot and informed trial of experiment 1. In practice, how this task was performed was to add one more column in the output files, denoting for each mapping if it is acceptable or not. From the resulting confusion matrix, accuracy, precision, recall, and F1 score were calculated according to their formulas (Fig. 4).



$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Figure 4. Confusion matrix and the formulas [29]

The confusion matrix values in our experiment case are as follows:

- **TP:** A correct suggestion was given by the LLM
- **FP:** An incorrect suggestion was given by the LLM
- **TN:** No suggestion was given by the LLM, and no suggestion indeed could be valid
- **FN:** No suggestion was given by the LLM, but there is a match in reality

Apart from the semantic meaning of the mapping rows, the LLM-generated suggestion was also tested for "hallucinations", e.g., ensuring that the suggested concept truly exists and is part of the ontology or vocabulary, and not superficially manufactured to fit the match. For experiment 2, due to a more flexible nature of this task (recommending appropriate vocabulary fields within and across the CH domain for each given data element), the human evaluation of the output was performed by distinguishing the cases where the LLM proposed acceptable recommendations, and the cases for which it proposed non-acceptable recommendations (in which there is the distinction of True Negative and False Negative cases). Once more, the recommendations were all checked for "hallucinations", ensuring that each of them exists and belongs to the schema of a valid Linked Open Vocabulary.

In the next Section, the results of the described processes are presented. The purpose of the experiments is not to replace human involvement in such a process, but rather to augment it in a more time-efficient and effective way and help fine-tune domain-specific applications by providing insights and areas for improvement in this regard in the form of informed set-up and methodology. The datasets used as groundwork in the prompting interaction are shown in the Appendices, where Appendices 1 and 2 include the full RA-RiC-O mapping, and Appendix 3 includes the validated mapping examples (sample) between RA and other schemas in LOV. Appendix 4 includes the prompting template used to interact with GPT-4o in both scenarios and GPT-4.5 in the 1st scenario. The following Section includes the key prompting interaction results

and detailed process outcomes.

4. Results

The results of the two-fold experiment using the prompt interface of GPT-4o and GPT-4.5 and the human evaluation for each respectively are presented in this Section.

4.1 Experiment 1 - ALD: From one Archive to more

As described earlier, the first experiment concerns the mapping between Records-in-Contexts (RiC-O) and a given, custom data element list from an archival institution. The rationale is to first conduct the mapping without prior knowledge (zero-shot) and then ask the model to repeat the procedure by initially providing the full, expert-validated mapping between RA and RiC-O for context (RA 1 and RA 2). This trial was attempted at both GPT-4o and GPT-4.5. The promp interaction and the analysis were relatively fast, taking a few seconds to go through the input given in tabular form. The input tables were prepared accordingly prior to the experiments in order to present the information in a well-organised and clear form. In some cases, some intermediate clarifications were required but overall, the model seemed to clearly grasp the instructions given. The output seemed promising at first sight, but several shortcomings were revealed during the evaluation process. During the 0-shot trial in both GPT-4o and GTP-4.5, the model was provided only with the custom archival institution data elements list that were to be included in the mapping process to RiC-O. During the second phase of the experiment, the full, expert-validated mapping between RiC-O and RA was provided as well, to check whether the results of the mapping would be improved.

Table 1 shows an excerpt of the model output for the case with GPT-4o and taking into account the mapping with RA. The columns of the output include the custom archival institution schema element and their short description, a column with the initial mapping the model proposed, a column with Notes and Comments that the model assigned, and a column with the model's own

revised mapping based on the validated mapping as benchmark. The last column Evaluation (Human) is the extra column added by the human evaluators (authors) during the evaluation process. The evaluation result follows the rules described in the Method Section; TP for a correct suggestion, FP for an incorrect suggestion, FN for no given suggestion when one exists, and TN for no given suggestion when one does not exist.

Custom Archives Schema Element	**Description (Summary)**	**RiC-O Equivalent Class / Property**	**Notes / Comments**	**Revised RiC-O Mapping (based on RA) (GPT-4o)**	**Evaluation (Human)**
Level	The hierarchical level of the material	rico:RecordResource + rico:hasLevel	RiC-O uses rico:hasOrHadLevelOf Description as a datatype property.	rico:RecordResource + rico:hasLevel	TP
Legal Status	Public Record or not (under law)	rico:hasLegalStatus	A datatype or object property; Legal status can also be an instance of rico:LegalStatus.	rico:hasLegalStatus	TP
Language	Language(s) of the record	rico:hasLanguage	Object property linking to rico:Language.	rico:hasLanguage	TP
Reference	Unique identifier to link records	rico:hasIdentifier	Identifier is often modeled as a literal, possibly as an instance of rico:Identifier.	rico:hasIdentifier	TP
Former	Previous ID by	rico:hasOrHadIde	Used for historical	rico:hasOrHadIdentifi	TP

Reference (Dept)	originating body	ntifier	identifiers.	er	
Map Designation	Map series info for maps	rico:hasIdentifier / rico:hasTitle	If used for structured series title or map ID	rico:hasIdentifier / rico:hasTitle	FN
Place of Deposit	Repository where records are held	rico:heldBy	Links to rico:Agent (institution).	rico:Place	FP
Access Conditions	Conditions affecting access	rico:hasAccessRestriction	Could use rico:hasAccessCondition or link to rico:Rule.	rico:hasAccessRestriction	FP (Hallucination)
Restrictions on Use	Use or reproduction restrictions	rico:hasUsageRestriction	May be a note or Rule	rico:hasUsageRestriction	FP (Hallucination)
Custodial History	Description of prior custody	rico:hasCustodialHistory	Object or literal property.	rico:hasCustodialHistory	FP (Hallucination)
Map Designation	Map series info for maps	rico:hasIdentifier / rico:hasTitle	If used for structured series title or map ID	rico:hasIdentifier / rico:hasTitle	FN
Former Reference (Pro)	Former TNA identifier	rico:hasOrHadIdentifier	For previous/provenance-based identifiers	rico:hasOrHadIdentifier	TP
Publication Note	Reference to published finding aids	rico:hasBibliographicReference	Modeled with rico:Bibliography	rico:hasBibliographicReference	FP
Administrati	History of the	rico:hasHistory	Linked to rico:Agent	rico:hasHistory	FP

ve, Bibliographical Background	record creator				
Gender Indicator	Gender of the individual	rico:hasGender	Can use literal or vocabulary	rico:hasGender	FP (Hallucination)
Index Terms: Subjects	Subject terms	rico:hasSubject	Could be Topic or Thing	rico:hasSubject	TP

Table 1. An excerpt of the mapping output

Following the manual, human evaluation, the confusion matrix for each use case was calculated.

The use cases are as follows:

- **GPT-4o (0-shot):** No prior knowledge mapping to RiC-O
- **GPT-4o (RA 1)**: Starting from 0-shot and then directly inputting the RA and ask it to repeat
- **GPT-4o (RA 2)**: Start directly with RA- took a few extra steps to get the final mapping
- **GPT-4.5 (0-shot):** No prior knowledge mapping to RiC-O
- **GPT-4.5 (RA as validation):** In this case the mapping to RA was used as a strict validation anchor for the model

The results of the confusion matrix calculation for all case scenarios are shown in Table 2.

Confusion Matrix for all Use Cases	
TP	**FN**
GPT-4o (0-shot): 34	**GPT-4o (0-shot):** 6
GPT-4o (RA 1): 35	**GPT-4o (RA 1):** 7

GPT-4o (RA 2): 26	**GPT-4o (RA 2):** 1
GPT-4.5 (0-shot): 14	**GPT-4.5 (0-shot):** 35
GPT-4.5 (RA as validation): 13	**GPT-4.5 (RA as validation):** 27
FP	**TN**
GPT-4o (0-shot): 26	**GPT-4o (0-shot):** 0
GPT-4o (RA 1): 23	**GPT-4o (RA 1):** 0
GPT-4o (RA 2): 13	**GPT-4o (RA 2):** 0
GPT-4.5 (0-shot): 6	**GPT-4.5 (0-shot):** 13
GPT-4.5 (RA as validation): 19	**GPT-4.5 (RA as validation):** 8

Table 2. The confusion matrix for all use cases

At this point, having calculated the True Positive, True Negative, False Positive, and False Negative cases, the metrics for each use case were calculated and are presented in Table 3. The maximum accuracy (65%) was achieved with GPT-4o in the second informed trial with RA. The last column shows the percentage of hallucinations per trial, meaning the number of times out of the total that the model proposed a superficial (non-existent) class or property for the mapping.

Trial / Metrics	Accuracy	Precision	Recall	F1-Score	Hallucination Cases
GPT-4o (0-shot)	51.5%	56.66	85%	67.9%	31.8%
GPT-4o (RA 1)	53.84%	60.34%	83.33%	70%	26%
GPT-4o (RA 2)	65%	66.66%	96.29%	78.77%	30%

GPT-4.5 (0-shot)	39.7%	70%	28.5%	40%	16%
GPT-4.5 (RA as validation)	31.34%	40%	32.5%	35.86%	0

Table 3. The evaluation metrics for all cases

Similarly, the precision, recall, and F1-score percentages are presented, giving a better overview of the output score distribution and cases. The highest accuracy (65%) was reached by GPT-4o in the second trial with informed input from the RA mapping to RiC-O (RA 2), while the lowest was GPT-4.5's 31.34% when it used the RA-RiC-O mapping as validator for its own mapping. This locked the model's flexible potential to choose an appropriate mapping and forcefully tried to impose the input mapping on the new examples regardless of better, existing matches, which led to the lowest accuracy, although for the same reason the hallucination cases dropped drastically. All the average accuracies of this experiment had a hallucination percentage ranging between 26-31.8%.

4.2 Experiment 2 - LOV Ontological Multi- (or Inter-) Relation

The second experiment concerns the ability of the model to make recommendations for re-use from the Linked Open Vocabularies (LOV) ecosystem, considering both adjacent domains (e.g., CH, Archives), but also other domains, if these are useful for enriched metadata description.

The input given to GPT-4o was a table of human-validated examples of RA data elements and their extended mapping to several LOV vocabularies. The model was then asked to extend the mapping for the rest of the elements in the same way, given the full RA schema list. An excerpt of the output mapping by GPT-4o is shown in Table 4. The table includes the Swedish National Archives (RA) element list, description, and the equivalent mapping to RiC-O, followed by two columns generated by GPT-4o, which show its provided recommendations for LOV re-use both in the CH and other domains respectively. The last 3 columns of the table were added afterwards, and include the manual, human evaluation of these results.

RA Element	Description	RiC-O Mapping	LOV Recommendations (GPT-4o CH)	LOV Recommendations (GPT-4o cross-domain)	Hallucinations (Human Evaluation)	Acceptable Recommendations (Human Evaluation)	Non-valid Recommendations (Human Evaluation)
Arkiv	Archive (fond)	schema:ArchiveComponent, rico:RecordSet	crm:E78_Curated_Holding, edm:Aggregation	crm:E78_Curated_Holding, edm:Aggregation, prov:Entity, schema:CreativeWork, void:Dataset	0	crm:E78_Collection, edm:Aggregation, schema:CreativeWork, void:Dataset, prov:Entity, edm:Collection, frbr:Manifestation, bibo:Collection	0
Typ	Archive type	rico:RecordSetType	crm:E78_Curated_Holding, dc:type, edm:Aggregation, rdau:P60047,	crm:E78_Curated_Holding, dc:type, edm:Aggregation, prov:Entity, rdau:P60047, schema:CreativeWork,	0	rdau:P60047 (has type of agent), schema:additionalType	crm:E78_Curated_Holding, skos:Concept, prov:Entity, void:Dataset, schema:CreativeWork, edm:Aggregation

			skos:Concept	schema:additionalType, skos:Concept, void:Dataset			
Typ	Types of geographical units and divisions	rico:PlaceType	dc:type, rdau:P60047, skos:Concept	dc:type, rdau:P60047, schema:additionalType, skos:Concept	0	schema:additionalType, dc:type, rdau:P60047	skos:Concept
Arkivbildare/upphov	Archivist	schema:creator, rico:Agent	foaf:Person, schema:Person	foaf:Person, schema:Person, vcard:Individual	0	foaf:Person, vcard:Individual, schema:Person	0
Arkivinstitution	Archival institution	rico:CorporateBody	edm:Agent, foaf:Organization, schema:Organization	edm:Agent, foaf:Organization, gr:BusinessEntity, org:Organization, schema:Organization	0	edm:Agent, foaf:Organization, gr:BusinessEntity, org:Organization, schema:Organization	0
Dokument	Document	rico:Record	bibo:Document, ddesc:Document,	bibo:Document, ddesc:Document, foaf:Document, schema:Creative	ddesc:Document	bibo:Document, foaf:Document	schema:CreativeWork

			foaf:Document	Work			
Fotografi	Photograph	rico:Record	crm:E38_Image, edm:ProvidedCHO	crm:E38_Image, edm:ProvidedCHO, media:Image, schema:ImageObject, vcard:Photo	media:Image	crm:E38_Image, edm:ProvidedCHO, schema:ImageObject, vcard:Photo	0
Titel	Archive title	rico:Name	crm:E78_Curated_Holding, dcterms:title, edm:Aggregation, skos:prefLabel	crm:E78_Curated_Holding, dcterms:title, edm:Aggregation, prov:Entity, schema:CreativeWork, schema:name, skos:prefLabel, void:Dataset	0	dcterms:title, schema:name, skos:prefLabel	prov:Entity, void:Dataset, schema:CreativeWork, edm:Aggregation, crm:E78_Collection
Senast ändrad	Date and system time of latest modification	rico:isModificationDateOf	crm:P4_has_time-span, dcterms:date	crm:P4_has_time-span, dcterms:date, schema:dateCreated, time:TemporalEntity	0	0	crm:P4_has_time-span, dcterms:date, schema:dateCreated, time:TemporalEntity

Upphovsrätt	Copyright	rico:ConditionsOfUse	dcterms:isVersionOf, edm:WebResource, rico:Instantiation	dcterms:isVersionOf, edm:WebResource, rico:Instantiation, schema:isBasedOn	0	0	dcterms:isVersionOf, edm:WebResource, rico:Instantiation, schema:isBasedOn
Topografihänvisningar	Topography references	rico:IsAssociatedWithPlace			0	geonames:locatedIn	gn:Feature, crm:E53_Place, dc:spatial, gn:Place, schema:Place, wd:Q515 (City)

Table 4. An excerpt of the LOV recommendations output

As described earlier, the output was checked for hallucinations or random correspondence, and the output recommendations were judged as acceptable or non-valid ones. The model did not provide output for several examples, denoted as “no recommendations”, which were separated in True Negative and False Negative cases. The reason for this is that we would not like to forcefully receive recommendations if truly there are no good matches, so true negative cases support the model’s overall performance. It is meaningless to judge the result in the same way as Experiment 1 as, depending on the context, the choice of vocabulary can change and has many levels of flexibility. However, what is important is that the mapping recommendations make

sense and can be directly borrowed and trusted. The percentages of the evaluation of the results are presented in Table 5.

Total Cases	69	
Hallucinations	≈ 0	
Acceptable Recommendations	32 cases (46.37%)	**Both acceptable and non-acceptable recommendations:** 23 times (71.8%)
		Only acceptable recommendations (exclusively): 10 cases (31.25%)
		Only non-acceptable recommendations (exclusively): 8 times (25%)
No Recommendations	37 cases (53.62%)	**True Negative (TN) cases:** 16 (43.24%)
		False Negative (FN) cases: 21 (56.75%)

Table 5. The evaluation results

The total cases were 69, out of which 37 cases were given no recommendations, and 32 cases were given acceptable recommendations. The 32 acceptable recommendation cases are divided into cases where both acceptable and non-acceptable recommendations were given (71.8%), cases where exclusively acceptable recommendations were given (31.25 %), and cases where exclusively non-acceptable recommendations were given (25%). Similarly, the 37 cases where no recommendations were provided by the model are divided into TN cases (43.24%) and FN cases (56.75%).

Overall, the model recommended the re-use of 21 vocabularies, apart from RiC-O. The different domains in which the suggested vocabularies belong are shown in Fig. 5, including the vocabulary names in the label. These graphs were generated by the model, right after the

analysis, at the request of the authors. As can be seen in the Figure, apart from the Cultural Heritage and Archival domain, we have recommendations from Web Metadata, Libraries and Bibliography, Geodata, Provenance and Process, Media, and Organisations.

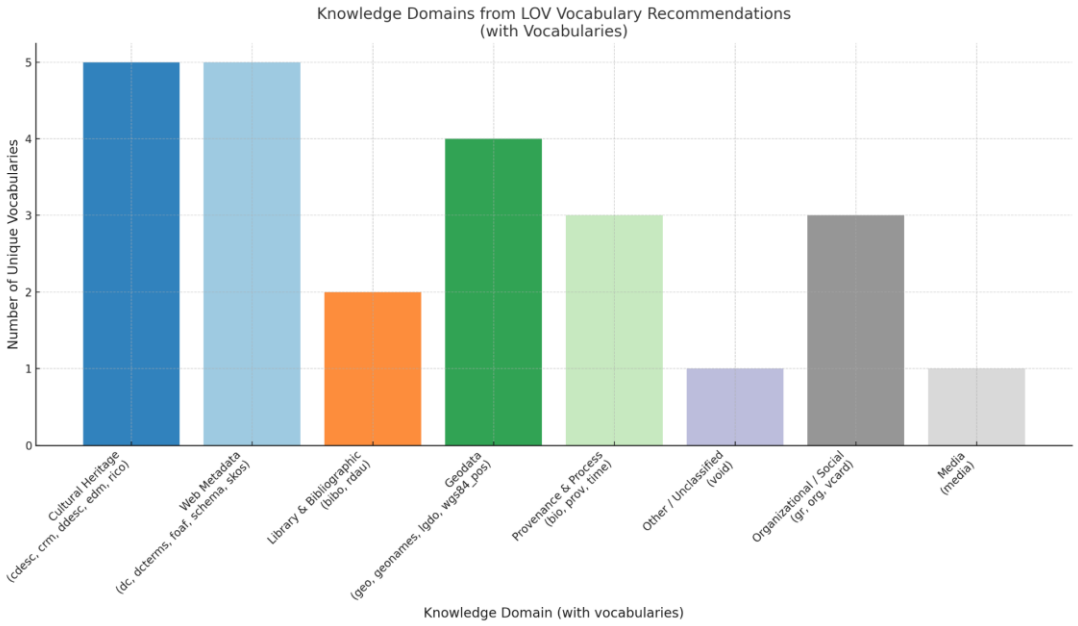


Figure 5. The domains (and LOV vocabulary per domain) included in the recommendations

Similarly, Fig. 6 shows the number of occurrences in the recommendations generated per vocabulary. For instance, classes and properties from cidoc-crm (crm) were suggested about 65 times, elements from bibo about 10 times, elements from the Europeana data model (edm) about 55 times, and ones from geo about 8 times.

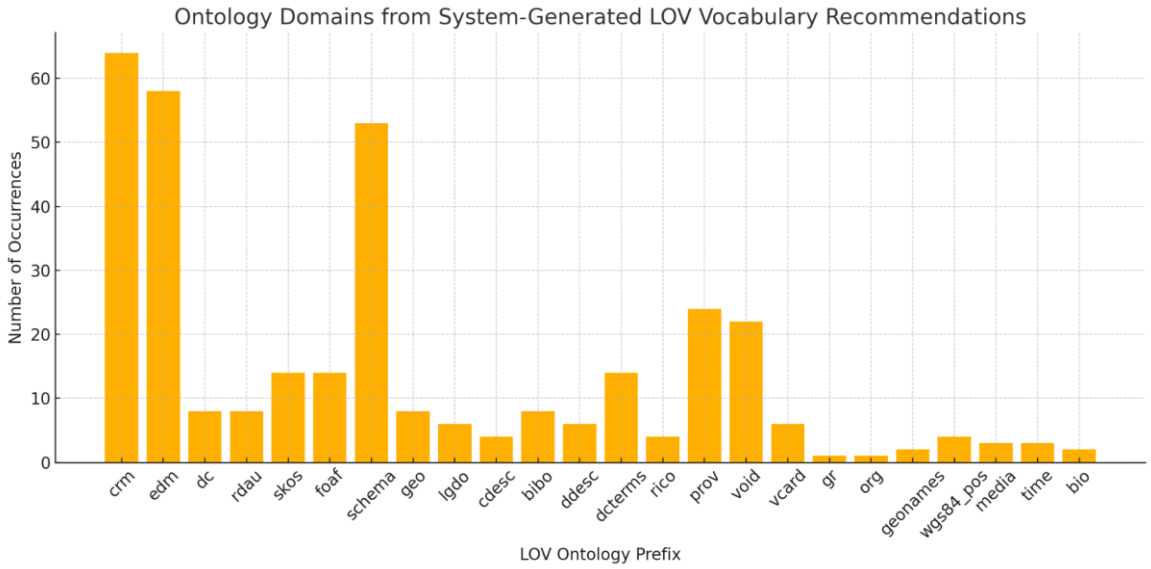


Figure 6. Number of occurrences for each LOV

The visualisation of Fig. 7 shows how each element of the RA schema is related to each recommended LOV, capturing also the most basic hierarchical relations among them. For instance, the class “Arkivinstitution” (Archival institution) is mapped to the LOV classes foaf:Organization, edm:Agent, org:Organization, gr:BusinessEntity, and schema:Organization. Similarly, “Fotografi” (Photograph) is mapped to vcard:Photo, media:Image, edm:ProvidedCHO, crm:E38_Image, and schema:ImageObject. “Serie” (Archive series) is mapped to bibo:Collection, schema:Collection, and cdesc: ArchivalSeries, while “Skapad” (Created) is mapped to schema:dateCreated, time:TemporalEntity, dcterms:date, crm:P4_has_time-span. “Koordinater” (Coordinates) are aligned with geo:lat, geo:long, wgs84:pos:lat, wgs84:pos:long, and “Arkivhistorik” (Archive history) is aligned with skos:note, bio:Event, prov:Activity, and schema:Event.

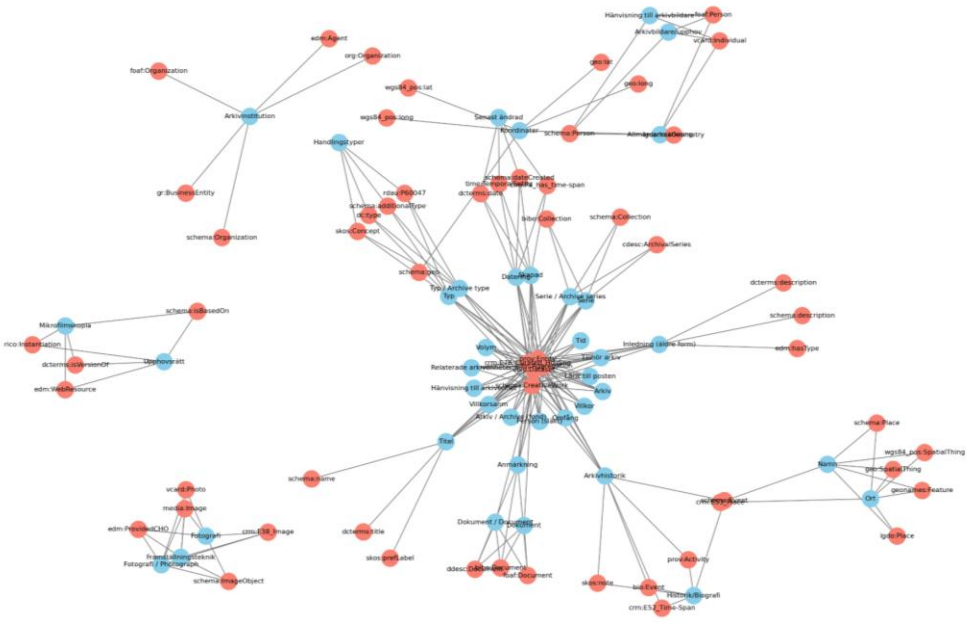


Figure 7. RA schema elements and their relation to the recommended LOV

In addition, a specific mapping to DCAT was requested, showing how the current schema could aim for compliance. A graph on how dcat as an overarching, higher-level metadata descriptor corresponds to RA is shown in Fig. 8, where (where meaningful) some RA elements are expressed in dcat terms.

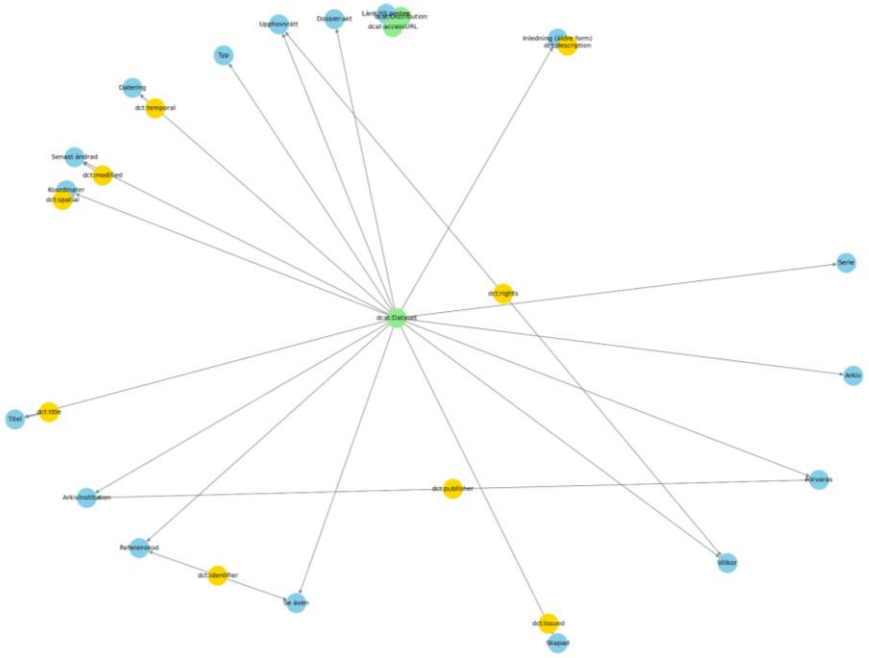


Figure 8. DCAT metadata descriptor correspondence to RA elements

In blue colour are the RA elements, and in green and yellow the dcat and dct terms respectively. Dcat:Dataset, dcat:Distribution, dcat:accessURL, dct:rights, dct: description, dct:publisher, dct:issued, dct:title, dct:identifier, dct:spatial, dct:modified and dct:temporal are used in this representation. The granularity level of this correspondence pertains to metadata which allows the data to be described appropriately through a catalog or data portal which uses a DCAT Application Profile (DCAT-AP).

5. Discussion

The purpose of this study was to test to what extent a user can talk an LLM, such as GPT, into performing (or rather assisting) the process of semantic alignment between a custom and a standardised schema of a given domain (in this case Archives), and to what extent can an LLM provide useful recommendations for vocabulary re-use (in this case LOV).

5.1 Experiment 1: ALD - From one Archive to more

As far as the alignment of a given custom archival institution schema to RiC-O is concerned, there is understandably not one single approach that is correct. The manual mapping process is an iterative and most often time-consuming process, which requires expert feedback and careful evaluation and alignment to each context. The process of manually mapping elements between two schemas remains highly subjective and is dependent on several factors which might affect the outcome. This is not necessarily a negative consequence as schema alignment needs to also be flexible and easily adaptable to each case scenario, but it introduces a number of parameters which need to be carefully assessed and considered. The organisation's mission, structural focus, legacy systems (if these exist), and infrastructure, are all factors which can force the mapping process to deviate significantly across applications of the same standard schema in various contexts or occasions. Different mapping approaches might work equally well for their intended purpose, provided that they are carefully corresponding to the institution's needs and infrastructural organisation. In this light, Experiment 1's intention is not to blindly trust an LLM to produce a well-tailored semantic alignment, but rather to make the process more time-efficient and, if possible, effective, if it is given enough context to analyse.

While from the results of the experiment it becomes obvious that, at least through prompting interaction, GPT is not ready to undertake this task as well as one would think, it is still an action worth investigating its limits and maximum performance boundaries should some guidelines allow for better results with time. It is noteworthy that there was no significant improvement from zero-shot to informed trial (with the RA-RiC-O mapping as input). The quite high percentage of hallucinations was reduced, but not to an extent that might strongly imply causality. GPT-4.5 generated considerably less hallucinations than GPT-4o but scored less in accuracy and all other metrics. A noticeable tendency was also that at zero-shot learning, the model tended to only use properties for the mapping (and making up quite some of them), but after the informed trial, it started (correctly so) identifying direct class-to-class mapping as well. What surprisingly did not work well either was that even when the RA-RiC-O mapping was used as an anchor of validation for the model, the model did not change its mind to give a better match alternative. In addition, in both set-ups, there were 0 True Negatives (TN), showing that GPT-4o

(same as for Experiment 2), was rather reluctant to not give any result, preferring a wrong answer instead.

Apart from the failures and pitfalls noticed, experiment 1 shows that at least from the viewpoint of saving time in the alignment process, the model can in very little time (few seconds) give over 60% good fits, which means the human might need to go through the output for validation and context-specific needs satisfaction but already having something to start from. Human intervention can definitely not be skipped and the results with high hallucination percentage are not to be trusted blindly, but allow for a first drafting of the mapping, providing also some explanation for each selection. Fine-tuning this process in an even more controlled environment could potentially bring forth better results.

5.2 Experiment 2: LOV Ontological Multi- (or Inter-) Relation

The initial alignment of the Swedish National Archives schema to Records-in-Contexts Ontology (RiC-O) [4] aimed to bring forth the advantages associated with standardisation and the transition to Linked Data and Archival Linked Data. RiC-O's main focus concept on archival resource description and relations to express actions performed by core entities in an archival setting make it a suitable candidate for standardised schema alignment. However, as far as describing the schema on more granularity levels is concerned, an alignment plan involving other ontologies in the archival or Cultural Heritage domain could be considered in order to achieve improved expressivity for the variety of different types of archival records offered by an archival institution (e.g., photographs, maps, paintings, video recordings, films, and other) which could require a more tailored approach, and this is the motivation behind Experiment 2. In a real scenario, the concerns around using multiple ontologies to describe one's schema, such as possible lack of ontology maintenance, resources, documentation, or technical support, are not to be neglected, however, to avoid customisation and keep it to a minimum, reuse of standard and well-mantained schemas is an ever-common practice.

It is worth mentioning that Experiment 2 had next-to-no hallucination cases. All recommendations from the vocabularies actually exist and are legitimate, meaning that, at least in the context of this small-scale study, the results for this experiment could be trusted from that aspect. However, the model produced repetitive results in some cases, possibly due to bias from the given input. The repetition for some cases is justifiable due to the lack of granularity of the given concepts, but other times the result is very poor when it is obvious that a match exists, the latter denoted as False Negatives (FN) during evaluation. For other cases and concepts, finding a direct match is not meaningful (e.g., mapping to more general purpose vocabularies), however, what was observed is that when the model does not find a match, in many cases, it simply repeats the same pattern as one of the previous examples, even if it is blatantly wrong, instead of not giving any results (True Negative -TN).

Overall, experiment 2 could be used to identify potential cross-domain use and give a good (also graphical) overview of the overlapping domains and intersections in a trusted way. The visuals generated are intuitive and worth consulting for a good overview of the output. The rest of the process is not yet very promising in the quality of results; even though in this experiment there are almost no hallucinations, which is good, the otherwise not so good performance is on one side expected as the model is faced with the challenging task of deciding a proper match and

granularity level for cases it does not have so much contextual information (except a few words description and title). On the other hand, all those False Negative (FN) cases could possibly be populated with maybe generic but correct recommendations, seeing as they clearly exist but the model did not find any.

5.3 Limitations and Future Directions

Some limitations of these experiments are due to limitations introduced by the model itself, and others fall under the responsibility of methodological dos and don'ts. Model limitations can be introduced due to short context windows, not allowing the model to sufficiently combine knowledge and input given all at the same time, without "forgetting" prior information, thus failing to consider all desired parameters to reach a decision. This entails the risk of requiring more correction steps during the prompting interaction, risking in its turn the soundness of the process, should one enter a loop of repetitive errors and propagation of misunderstanding. The input should balance clear instructions, well- structured input, and contextual information. As mentioned previously, the black-box effect is considerably high, causing insecurity when it comes to trusting the output. If a human is to spend as much time validating the output as it would take to perform the task themselves, the assistance loses any value. However, some cautious optimism might be allowed by the results, at least considering that the evaluation process for both experiments took very little time compared to the manual mapping process of [4] or similar.

In addition, methodological and experiment limitations include the flexibility in human interpretation of the results, which might affect the evaluation of the output as well. Semantic alignment in real-case applications is a process very highly dependent on case-specific context and mapping recommendations should be considered auxiliary and not written in stone. Moreover, different approaches in prompting techniques can most likely produce different results, the divergence of which is not easy to measure in this environment. Regardless of how cautious and methodologically sound the process might seem, the model's nature will always introduce entropy in the output, making it extremely challenging to set strict boundaries which allow for both improved performance and minimum intervention.

For the reasons described, future directions of this study may involve local, domain specific LLM training on clean data in order to repeat the experiments and see the difference. The potential and the outskirt limits of mere prompting interaction for this purpose is perhaps not exhausted and further experimentation may provide improved insights and guidelines to boost the process's performance. However, in order to define a highly configurable environment with more trusted output, training and fine-tuning a model in context and domain-specific setting would be a valuable pathway to walk and experiment with.

6. Conclusion

Some common pitfalls for all set-ups were identified. The output during the trials in some cases showed some predictability but many times the black box effect takes over. In addition, it is impossible to go one step back and receive exactly the same results as before. Even though this is expected from a highly probabilistic model, in cases with strict benchmarking and guidance, the result would be expected to be a bit more coherent and predictable. This is not a problem if it

only concerns the flexibility of vocabulary use or selection of mapping according to case-specific examples, and if the performance shows signs of normalised behaviour. However, there is still high uncertainty about trusting that "this is the best the model can do" or "this is the worst the model can do". This form of consistency could allow for a better quantification of the results and standardisation of the process to improve performance, and, regardless of the ability or not to reach a high accuracy level, it could produce much more consistently trusted outputs, maximising what this method can potentially offer.

References

[1] LinkedData - W3C Wiki. (n.d.). https://www.w3.org/wiki/LinkedData#:~:text=The%20term%20Linked%20Data%20refers,can%20look%20up%20those%20names.

[2] Hawkins, A. (2021). Archives, linked data and the digital humanities: increasing access to digitised and born-digital archives via the semantic web. Archival Science, 22(3), 319–344. https://doi.org/10.1007/s10502-021-09381-0

[3] Riksarkivet. (n.d.). Start – söktjänstens startsida - Riksarkivet - Sök i arkiven. https://sok.riksarkivet.se/

[4]

[5] Bikakis A, Hyvönen E, Jean S, Markhoff B, Mosca A (2021) Editorial: special issue on semantic web for cultural heritage. Semantic Web 12(2):163–167

[6] GO FAIR initiative. (2022, January 21). FAIR principles - GO FAIR. GO FAIR. https://www.go-fair.org/fair-principles/

[7] Gracy, K. F. (2017). Enriching and enhancing moving images with Linked Data. Journal of Documentation, 74(2), 354–371. https://doi.org/10.1108/jd-07-2017-0106

[8] Gracy, K. F. (2014). Archival description and linked data: a preliminary study of opportunities and implementation challenges. Archival Science, 15(3), 239–294. https://doi.org/10.1007/s10502-014-9216-2

[9] Hawkins, A. & University of Liverpool. (2021). Advocating for Linked Archives: the benefits to users of Archival Linked Data. In University of Liverpool [Journal-article].

[10] International Council on Archives Records in Contexts Ontology (ICA RiC-O) version 1.0.2. (n.d.). https://www.ica.org/standards/RiC/RiC-O_1-0-2.html

[11] A global network on sharing cultural heritage (no date) OpenGLAM. Available at: https://openglam.org/

[12] Overview for 'linked-open-data' | Europeana PRO (2023) Europeana PRO. Available at: https://pro.europeana.eu/itemtype/linked-open-data

[13] Mazzini, S., & Ricci, F. (2011). EAC-CPF Ontology and Linked Archival Data. (pp. 72–81). http://ceur-ws.org/Vol-801/paper6.pdf

[14] Encoded Archival Description (EAD) | Society of American Archivists. (n.d.). https://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-standards-ts-eas/encoded-archival-description-ead

[15] Dobreski, B., Park, J., Leathers, A., & Qin, J. (2020). Remodeling archival metadata descriptions for linked archives. International Conference on Dublin Core and Metadata Applications, 1–11. https://dcpapers.dublincore.org/pubs/article/download/4223/2417

[16] ICA. (2025, February 12). ISAD(G): General International Standard Archival Description - Second edition - ICA. https://www.ica.org/resource/isadg-general-international-standard-archival-description-second-edition/

[17] Linked Open Vocabularies (LOV). (n.d.). https://lov.linkeddata.es/dataset/lov/

[18] Chen, S. (2019). Semantic enrichment of linked archival materials. KNOWLEDGE ORGANIZATION, 46(7), 530–547. https://doi.org/10.5771/0943-7444-2019-7-530

[19] Bountouri, L., & Gergatsoulis, M. (2011). The semantic mapping of archival metadata to the CIDOC CRM Ontology. Journal of Archival Organization, 9(3–4), 174–207. https://doi.org/10.1080/15332748.2011.650124

[20] Koch, I., Lopes, C. T., & Ribeiro, C. (2023). Moving from ISAD(G) to a CIDOC CRM-based Linked Data Model in the Portuguese Archives. Journal on Computing and Cultural Heritage, 16(4), 1–21. https://doi.org/10.1145/3605910

[21] FranceArchives. (n.d.). FranceArchives. https://francearchives.gouv.fr/en

[22] Doumanas D, Bouchouras G, Soularidis A, Kotis K, Vouros G. (2025). From Human- to LLM-Centered Collaborative Ontology Engineering. Applied Ontology. 2025;0(0). doi:10.1177/15705838241305067.

[23] Osman, I., Pileggi, S. F., & Yahia, S. B. (2024). Uncertainty in Automated Ontology Matching: Lessons from an Empirical Evaluation. Applied Sciences, 14(11), 4679. https://doi.org/10.3390/app14114679

[24] Hoseini, S., Burgdorf, A., Paulus, A., Meisen, T., Quix, C., & Pomp, A. (2025). Challenges and Opportunities of LLM-Augmented Semantic Model Creation for Dataspaces. In Lecture notes in computer science (pp. 183–200). https://doi.org/10.1007/978-3-031-78955-7_17

[25] Pan, X., Van Ossenbruggen, J., De Boer, V., & Huang, Z. (2025). A RAG approach for generating competency questions in ontology Engineering. In Communications in computer and information science (pp. 70–81). https://doi.org/10.1007/978-3-031-81974-2_6

[26] Cigliano, A., & Fallucchi, F. (2025). The convergence of open data, linked data, ontologies, and large language models: enabling Next-Generation knowledge systems. In Communications in computer and information science (pp. 197–213). https://doi.org/10.1007/978-3-031-81974-2_17

[27] Maratsi, M. I., Gialoussi, N., Alexopoulos, C., & Charalabidis, Y. (2025). A proposed methodology for Sub-Ontology development in comprehensive scientific investigation methods and tooling. In Communications in computer and information science (pp. 28–43). https://doi.org/10.1007/978-3-031-81974-2_3

[28] Maratsi, M. I., Ahmed, U., Alexopoulos, C., Charalabidis, Y., & Polini, A. (2024). Towards Cross-Domain Linking of Data: A Semantic Mapping of Cultural Heritage Ontologies (pp. 165–176). https://doi.org/10.1145/3657054.3657077

[29] Soni, P. (2024, September 10). Confusion matrix, precision, and recall. Train in Data's Blog. https://www.blog.trainindata.com/confusion-matrix-precision-and-recall/

Appendix 1 - The Mapping Matrix (Classes and Properties)

RA Schema Classes and Properties	English Translation/Desc	Mapping to RiC-O
Arkiv	Archive (fond).	schema:ArchiveComponent, rico:RecordSet
Typ	Archive type	rico:RecordSetType
Arkivbildare/upphov	Archivist	schema:creator, rico:Agent
Arkivinstitution	Archival institution	rico:CorporateBody
Ort	Place	rico:Place
Serie	Archive series	rico:RecordSet
Volym	Archive volume	rico:Record
Dokument	Document	rico:Record
Dossier-akt	Folder	rico:RecordSet
Fotografi	Photograph	rico:Record
Karta/ritning	Map/Drawing	rico:Record
Rörlig bild	Video	rico:Record

Ljudupptagning	Sound recording	rico:Record
Mikrofilm	Microfilm	rico:Record
Mikrofilmskopia	Microfilm copy	rico:Instantiation
ADB-upptagning (fil)	A file born digital (produced directly by digital means)	rico:Record
ADB-upptagning (bildfil)	An image file born digital (produced directly by digital means)	rico:Record
Databärare för ADB	The digital means bearing the digital file	rico:CarrierType
Titel	Archive title	rico:Name
Tillhör arkiv	Where the archive belongs	rico:ProvenanceRelation
Förvaras	The entity that is currently preserving the file	rico:IsOrWasHolderOf
Tid	The time/dating of the original archive	Date, DateSet, SingleDate
Referenskod	Reference code	schema:identifier, rico:Identifier
Länk till posten	Link to the archive resource.	schema:identifier, RecordResource
Datering	The archive file's time range or date.	rico:Date

Villkor	Existence of archive usage terms	rico:ConditionsOfAccess
Villkorsanm	The specific terms of archive use	rico:ConditionsOfAccess
Omfång	Dimensions of the physical archive	rico:RecordResourceExtent
Sekretess	Secrecy	rico:ConditionsOfAccess
Allmän anmärkning	General note	rico:GeneralDescription
Inledning (äldre form)	A more detailed description of the archive and its content	rico:ScopeAndContent
Skapad	Date and system time of archive creation in the system	rico:isCreationDateOf
Senast ändrad	Date and system time of latest modification	rico:isModificationDateOf
Källa	Source	rico:IsOrWasDescribedBy
Källår	Source year	rico:IsDateAssociatedWith
Reproducerat	Reproduced	rico:hasOrHadDerivedInstantiation
Kopierat	Copied	rico:hasCopy
Relaterade arkivenheter	One or more relevant archives	rico:IsRecordResourceAssociatedWithRecordResource
Se även	Relevant reference codes	rico:IsRecordResourceAssociatedWithRecordResource

Anmärkning	Information about the archive documents' content	rico:GeneralDescription
Statlig myndighet	State authority	rico:CorporateBodyType
Kommunal myndighet	Community authority	rico:CorporateBodyType
Gård	Farm	rico:PlaceType
By	Village	rico:PlaceType
Förening	Association	rico:CorporateBodyType
Företag	Business company	rico:CorporateBodyType
Person (släkt)	A person or family who contributed an archive	rico:Person
Namn	Name of a place or a division	rico:Name
Typ	Types of geographical units and divisions	rico:PlaceType
Verksamhetstid	Start and end year for certain divisions (localities)	rico:IsAssociatedWithDate
Koordinater	Coordinates	rico:Coordinates
Höjd	Height	rico:CarrierExtent
Bredd	Width	rico:CarrierExtent
Restriktioner p.g.a. personuppgifter	Personal data restrictions	rico:ConditionsOfAccess
Rättighet för digital reproduktion	Right for digital material reproduction	rico:IntellectualPropertyRightsRelation

Omslagsformat	Picture or image format	rico:CarrierExtent
Fysisk form	The form of physical artifact	rico:CarrierType
Skala	Image scale	rico:CarrierExtent
Framställningsteknik	The technique used to produce the photograph or painting	rico:ProductionTechnique
Material	Material	rico:CarrierType
Handlingstyper	Action type	rico:ProductionTechnique
Arkivhistorik	Archive history	rico:History
Ordning och struktur	Order and structure	rico:Structure
Upphovsrätt	Copyright	rico:ConditionsOfUse
Rättigheter för metadata	Rights for metadata	rico:ConditionsOfUse
Hänvisning till arkivenhet	Reference to archive unit	rico:IsProvenanceOf
Topografihänvisningar	Topography references	rico:IsAssociatedWithPlace
Hänvisning till arkivbildare	Reference to archivist	rico:hasCreator
Överordnad namndel	“Parent” name	rico:NameOfAgent
Underordnad namndel	Subordinate name	rico:NameOfAgent
Giltighetstid start	Validity period start	rico:IsDateAssociatedWith
Giltighetstid slut	Validity period end	rico:IsDateAssociatedWith
Sektor (organisation)	Sector (organization)	rico:CorporateBodyType
Historik/Biografi	History/Biography	rico:History

Topografi	Topography	rico:HasOrHadLocation
Hänvisning till auktoritet	Reference to authority	rico:HasOrHadSubordinate, rico:HasOrHadSuccessor

Appendix 2 - RiC-O Object Properties Needed for Context Description in RA

RA Class	RICO Mapping	RiC-O Relation	Domain	Range
Arkiv	rico: RecordSet	describesOrDescribed	Arkiv	Dokument
Arkiv	rico: RecordSet	directlyIncludes	Arkiv	Arkiv
Arkiv	rico: RecordSet	followsInTime	Arkiv	Arkiv
Arkiv	rico: RecordSet	followsOrFollowed	Arkiv	Arkiv
Arkiv	rico: RecordSet	hadConstituent	Arkiv	Arkiv
Arkiv	rico: RecordSet	hasAccumulator	Arkiv	Arkivbildare/upphov
Arkiv	rico: RecordSet	hasAddressee	Arkiv	Arkivbildare/upphov
Arkiv	rico: RecordSet	hasAuthor	Arkiv	Person (släkt)
Arkiv	rico: RecordSet	hasCollector	Arkiv	Arkivbildare/upphov
Arkiv	rico: RecordSet	hasConstituentTransitive	Arkiv	Arkiv
Arkiv	rico: RecordSet	hasContentOfType	Arkiv	Typ
Arkiv	rico: RecordSet	hasCopy	Arkiv	Arkiv
Arkiv	rico: RecordSet	hasCreationDate	Arkiv	Tid
Arkiv	rico: RecordSet	hasCreator	Arkiv	Arkivbildare/upphov

Arkiv	rico: RecordSet	hasDirectConstituent	Arkiv	Arkiv
Arkiv	rico: RecordSet	hasDraft	Arkiv	Arkiv
Arkiv	rico: RecordSet	hasExtent	Arkiv	Omfång
Arkiv	rico: RecordSet	hasGeneticLinkToRecordResource	Arkiv	Arkiv
Arkiv	rico: RecordSet	hasOrHadConstituent	Arkiv	Arkiv
Arkiv	rico: RecordSet	hasOrHadDigitalInstantiation	Arkiv	Instantiation
Arkiv	rico: RecordSet	hasOrHadHolder	Arkiv	Arkivbildare/upphov
Arkiv	rico: RecordSet	hasOrHadInstantiation	Arkiv	Instantiation
Arkiv	rico: RecordSet	hasOrHadMainSubject	Arkiv	Allmän anmärkning
Arkiv	rico: RecordSet	hasOrHadManager	Arkiv	Arkivbildare/upphov
Arkiv	rico: RecordSet	hasOrHadSubject	Arkiv	Typ
Arkiv	rico: RecordSet	hasOrHadTitle	Arkiv	Titel
Arkiv	rico: RecordSet	hasOrganicOrFunctionalProvenance	Arkiv	Arkivbildare/upphov
Arkiv	rico: RecordSet	hasOrganicProvenance	Arkiv	Arkivbildare/upphov
Arkiv	rico: RecordSet	hasOriginal	Arkiv	Arkiv
Arkiv	rico: RecordSet	hasPublicationDate	Arkiv	Tid
Arkiv	rico: RecordSet	hasPublisher	Arkiv	Arkivbildare/upphov
Arkiv	rico: RecordSet	hasReceiver	Arkiv	Arkivbildare/upphov

Arkiv	rico: RecordSet	hasRecordSetType	Arkiv	Typ
Arkiv	rico: RecordSet	hasReply	Arkiv	Arkiv
Arkiv	rico: RecordSet	hasSender	Arkiv	Arkivbildare/upphov
Arkiv	rico: RecordSet	included	Arkiv	Arkiv
Arkiv	rico: RecordSet	includesOrIncluded	Arkiv	Arkiv
Arkiv	rico: RecordSet	includesTransitive	Arkiv	Arkiv
Arkiv	rico: RecordSet	isConstituentOfTransitive	Arkiv	Arkiv
Arkiv	rico: RecordSet	isCopyOf	Arkiv	Arkiv
Arkiv	rico: RecordSet	isDirectConstituentOf	Arkiv	Arkiv
Arkiv	rico: RecordSet	isDirectlyIncludedIn	Arkiv	Arkiv
Arkiv	rico: RecordSet	isDraftOf	Arkiv	Arkiv
Arkiv	rico: RecordSet	isIncludedInTransitive	Arkiv	Arkiv
Arkiv	rico: RecordSet	isOrWasConstituentOf	Arkiv	Arkiv
Arkiv	rico: RecordSet	isOrWasIncludedIn	Arkiv	Arkiv
Arkiv	rico: RecordSet	isOriginalOf	Arkiv	Arkiv
Arkiv	rico: RecordSet	isRecordResourceAssociatedWithRecordResource	Arkiv	Arkiv
Arkiv	rico: RecordSet	isReplyTo	Arkiv	Arkiv
Arkiv	rico: RecordSet	wasConstituentOf	Arkiv	Arkiv
Arkiv	rico: RecordSet	wasIncludedIn	Arkiv	Arkiv

Typ	rico:RecordSetType	isRecordSetTypeOf	Typ	Arkiv
Arkivbildare/upphov	rico:Agent	agentHasOrHadLocation	Arkivbildare/upphov	Ort
Arkivbildare/upphov	rico:Agent	authorizedBy	Arkivbildare/upphov	Villkor
Arkivbildare/upphov	rico:Agent	hadSubordinate	Arkivbildare/upphov	Arkivbildare/upphov
Arkivbildare/upphov	rico:Agent	hasDirectSubordinate	Arkivbildare/upphov	Arkivbildare/upphov
Arkivbildare/upphov	rico:Agent	hasOrHadAgentName	Arkivbildare/upphov	Name
Arkivbildare/upphov	rico:Agent	hasOrHadAuthorityOver	Arkivbildare/upphov	Arkiv
Arkivbildare/upphov	rico:Agent	hasOrHadController	Arkivbildare/upphov	Arkivbildare/upphov
Arkivbildare/upphov	rico:Agent	hasOrHadJurisdiction	Arkivbildare/upphov	Ort
Arkivbildare/upphov	rico:Agent	hasOrHadSubordinate	Arkivbildare/upphov	Arkivbildare/upphov
Arkivbildare/upphov	rico:Agent	hasOrHadWorkRelationWith	Arkivbildare/upphov	Arkivbildare/upphov
Arkivbildare/upphov	rico:Agent	hasSubordinateTransitive	Arkivbildare/upphov	Arkivbildare/upphov

Arkivbildare/upphov	rico:Agent	hasSuccessor	Arkivbildare/upphov	Arkivbildare/upphov
Arkivbildare/upphov	rico:Agent	isAccumulatorOf	Arkivbildare/upphov	Arkiv
Arkivbildare/upphov	rico:Agent	isAddresseeOf	Arkivbildare/upphov	Arkiv
Arkivbildare/upphov	rico:Agent	isAgentAssociatedWithAgent	Arkivbildare/upphov	Arkivbildare/upphov
Arkivbildare/upphov	rico:Agent	isAgentAssociatedWithPlace	Arkivbildare/upphov	Ort
Arkivbildare/upphov	rico:Agent	isCollectorOf	Arkivbildare/upphov	Arkiv
Arkivbildare/upphov	rico:Agent	isCreatorOf	Arkivbildare/upphov	Arkiv
Arkivbildare/upphov	rico:Agent	isDirectSubordinateTo	Arkivbildare/upphov	Arkivbildare/upphov
Arkivbildare/upphov	rico:Agent	isOrWasControllerOf	Arkivbildare/upphov	Arkivbildare/upphov
Arkivbildare/upphov	rico:Agent	isOrWasHolderOf	Arkivbildare/upphov	Arkiv, Allmän anmärkning, Inledning (äldre form)
Arkivbildare/upphov	rico:Agent	isOrWasManagerOf	Arkivbildare/upphov	Arkiv

Arkivbildare/upphov	rico:Agent	isOrWasResponsibleForEnforcing	Arkivbildare/upphov	Villkor
Arkivbildare/upphov	rico:Agent	isOrWasSubordinateTo	Arkivbildare/upphov	Arkivbildare/upphov
Arkivbildare/upphov	rico:Agent	isOrganicProvenanceOf	Arkivbildare/upphov	Arkiv, Allmän anmärkning, Inledning (äldre form)
Arkivbildare/upphov	rico:Agent	isPublisherOf	Arkivbildare/upphov	Arkiv
Arkivbildare/upphov	rico:Agent	isResponsibleForIssuing	Arkivbildare/upphov	Villkor
Arkivbildare/upphov	rico:Agent	isSubordinateToTransitive	Arkivbildare/upphov	Arkivbildare/upphov
Arkivbildare/upphov	rico:Agent	isSuccessorOf	Arkivbildare/upphov	Arkivbildare/upphov
Arkivbildare/upphov	rico:Agent	wasSubordinateTo	Arkivbildare/upphov	Arkivbildare/upphov
Arkivinstitution	rico:CorporateBody	hasOrHadCorporateBodyType	Arkivinstitution	Statlig myndighet
Arkivinstitution	rico:CorporateBody	isOrWasEmployerOf	Arkivinstitution	Person (släkt)
Arkivinstitution	rico:CorporateBody	resultedFromTheMergerOf	Arkivinstitution	Arkivinstitution
Arkivinstitution	rico:CorporateBody	resultedFromTheSplitOf	Arkivinstitution	Arkivinstitution
Arkivinstitution	rico:CorporateBody	wasMergedInto	Arkivinstitution	Arkivinstitution

Arkivinstitution	rico:CorporateBody	wasSplitInto	Arkivinstitution	Arkivinstitution
Ort	rico:Place	contained	Ort	Ort
Ort	rico:Place	containsOrContained	Ort	Ort
Ort	rico:Place	hasOrHadPlaceName	Ort	Name
Ort	rico:Place	isDeathPlaceOf	Ort	Person (släkt)
Ort	rico:Place	isOrWasAdjacentTo	Ort	Ort
Ort	rico:Place	isOrWasJurisdictionOf	Ort	Arkivbildare/upphov
Ort	rico:Place	isOrWasLocationOfAgent	Ort	Arkivbildare/upphov
Ort	rico:Place	isPlaceAssociatedWith	Ort	rico: Thing
Ort	rico:Place	isPlaceAssociatedWithAgent	Ort	Arkivbildare/upphov
Titel	rico:Name, rico:Title	isOrWasTitleOf	Titel	Arkiv
Tid	rico:Date, rico:DateSet, rico:SingleDate	hasDateType	Tid	rico:DateType
Tid	rico:Date, rico:DateSet, rico:SingleDate	hasWithin	Tid	Tid
Tid	rico:Date, rico:DateSet, rico:SingleDate	intersects	Tid	Tid
Tid	rico:Date, rico:DateSet,	isBeginningDateOf	Tid	rico:Thing

	rico:SingleDate			
Tid	rico:Date, rico:DateSet, rico:SingleDate	isBirthDateOf	Tid	Person (släkt)
Tid	rico:Date, rico:DateSet, rico:SingleDate	isCreationDateOf	Tid	Arkiv
Tid	rico:Date, rico:DateSet, rico:SingleDate	isDateAssociatedWith	Tid	rico:Thing
Tid	rico:Date, rico:DateSet, rico:SingleDate	isDateOfOccurrenceOf	Tid	rico:Event
Tid	rico:Date, rico:DateSet, rico:SingleDate	isDestructionDateOf	Tid	rico:Thing
Tid	rico:Date, rico:DateSet, rico:SingleDate	isEndDateOf	Tid	rico:Thing
Tid	rico:Date, rico:DateSet, rico:SingleDate	isFromUseDateOf	Tid	rico:Appellation
Tid	rico:Date, rico:DateSet,	isLastUpdateDateOf	Tid	rico:Thing

	rico:SingleDate			
Tid	rico:Date, rico:DateSet, rico:SingleDate	isModificationDateOf	Tid	rico:Thing
Tid	rico:Date, rico:DateSet, rico:SingleDate	isOrWasCreationDateOfAllMembersOf	Tid	Arkiv
Tid	rico:Date, rico:DateSet, rico:SingleDate	isOrWasCreationDateOfMostMembersOf	Tid	Arkiv
Tid	rico:Date, rico:DateSet, rico:SingleDate	isOrWasCreationDateOfSomeMembersOf	Tid	Arkiv
Tid	rico:Date, rico:DateSet, rico:SingleDate	isPublicationDateOf	Tid	Arkiv
Tid	rico:Date, rico:DateSet, rico:SingleDate	isToUseDateOf	Tid	rico:Appellation
Tid	rico:Date, rico:DateSet, rico:SingleDate	isWithin	Tid	Tid
Referenskod	rico:Identifier	hasIdentifierType	Referenskod	rico:IdentifierType
Referenskod	rico:Identifier	isIdentifierTypeOf	rico:IdentifierTy	Referenskod

			pe	
Referenskod	rico:Identifier	isOrWasIdentifierOf	Referenskod	rico:Thing
Villkor	rico:ConditionsOfAccess	isOrWasEnforcedBy	Villkor	Arkivbildare/upphov
Villkor	rico:ConditionsOfAccess	isOrWasExpressedBy	Villkor	Arkiv
Villkor	rico:ConditionsOfAccess	isRuleAssociatedWith	Villkor	rico:Thing
Villkor	rico:ConditionsOfAccess	issuedBy	Villkor	Arkivbildare/upphov
Villkor	rico:ConditionsOfAccess	regulatesOrRegulated	Villkor	rico:Thing
Villkorsanm	rico:ConditionsOfAccess	isOrWasRuleTypeOf	Villkorsanm	Villkor
Omfång	rico:Extent	isUnitOfMeasurementOf	Omfång	rico:Extent
Skapad	rico:isCreationDateOf	isCreationDateOf	Tid	Arkiv, Allmän anmärkning
Senast ändrad	rico:isModificationDateOf, rico:isLastUpdateDateOf	isModificationDateOf	Tid	rico:Thing
Senast ändrad	rico:isModificationDateOf,	isLastUpdateDateOf	Tid	rico:Thing

	rico:isLastUpdateDateOf			
Statlig myndighet	rico:CorporateBodyType	isOrWasCorporateBodyTypeOf	Statlig myndighet	Arkivinstitution
Kommunal myndighet	rico:CorporateBodyType	isOrWasCorporateBodyTypeOf	Kommunal myndighet	Arkivinstitution
Person (släkt)	rico:Person	hasAncestor	Person (släkt)	Person (släkt)
Person (släkt)	rico:Person	hasBirthDate	Person (släkt)	Tid
Person (släkt)	rico:Person	hasBirthPlace	Person (släkt)	Ort
Person (släkt)	rico:Person	hasChild	Person (släkt)	Person (släkt)
Person (släkt)	rico:Person	hasDescendant	Person (släkt)	Person (släkt)
Person (släkt)	rico:Person	hasFamilyAssociationWith	Person (släkt)	Person (släkt)
Person (släkt)	rico:Person	hasOrHadCorrespondent	Person (släkt)	Person (släkt)
Person (släkt)	rico:Person	hasOrHadDemographicGroup	Person (släkt)	Person (släkt)
Person (släkt)	rico:Person	hasOrHadSpouse	Person (släkt)	Person (släkt)
Person (släkt)	rico:Person	hasOrHadStudent	Person (släkt)	Person (släkt)
Person (släkt)	rico:Person	hasOrHadTeacher	Person (släkt)	Person (släkt)
Person (släkt)	rico:Person	hasSibling	Person (släkt)	Person (släkt)
Person (släkt)	rico:Person	isAuthorOf	Person (släkt)	Arkiv
Person (släkt)	rico:Person	isChildOf	Person (släkt)	Person (släkt)
Namn	rico:Name	isOrWasNameOf	Namn	rico:Thing

Typ	rico:PlaceType	isOrWasPlaceTypeOf	Typ	Ort
Koordinater	rico:Coordinates	isOrWasCoordinatesOf	Coordinates	rico:PhysicalLocation
Restriktioner p.g.a. personuppgifter	rico:ConditionsOfAccess	isDirectSubdivisionOf	Restriktioner p.g.a. personuppgifter	rico:Group
Restriktioner p.g.a. personuppgifter	rico:ConditionsOfAccess	isOrWasRuleTypeOf	Restriktioner p.g.a. personuppgifter	Villkor
Fysisk form	rico:CarrierType	isCarrierTypeOf	Fysisk form	Arkiv
Framställningsteknik	rico:ProductionTechnique	isProductionTechniqueOf	Framställningsteknik	Arkiv
Material	rico:CarrierType	isCarrierTypeOf	Material	Arkiv
Handlingstyper	rico:ProductionTechnique	isProductionTechniqueOf	Handlingstyper	Arkiv

Appendix 3 - Extended Mapping Examples RA - LOV

RA Class	**RA Description**	**RiC-O Mapping**	**Recommended Mapping to other LOV**
Arkiv	Ett arkiv. An archive (fond).	ric:RecordSet	cdesc:ArchivalFonds, cdesc:ArchivalRecordSet
Serie	En arkivserie. An archive series.	ric:RecordSet	cdesc:ArchivalSeries, cdesc:ArchivalSubseries

Typ	Allmän arkiv typ. General archive type.	ric:RecordSetType	cdesc:Subject
Dokument	Ett dokument. See earlier comment. "Dokument" usually refers to a paper document.	ric:Record	cdesc:SourceandDocument
Fotografi	Ett fotografi. Usually an analog photography on paper or glass etc.	ric:Record	cdesc:PhotographicSeries, cdesc:PhotographicDocumentation
Karta/ritning	Karta/ritning. Usually maps and drawings on paper.	ric:Record	cdesc:GraphicOrCartographicDocumentation
Rörlig bild	Film/Video. Usually analog film or video.	ric:Record	cdesc:FilmDocumentation
Ljudupptagning	Ljudinspelning. Usually analog recordings.	ric:Record	cdesc:AudioDocumentation
ADB-upptagning (fil)	ADB-upptagning. An archival information package (AIP) (containing files produced directly by digital means).	ric:Record or ric:Instantiation	cdesc:MultimediaDocumentation
Statlig myndighet	En statlig myndighet. A state	ric:CorporateBodyTy	lgdo:StateAgentOffice

	authority.	pe	
Kommunal myndighet	En kommunal myndighet. A municipality agency.	ric:CorporateBodyType	lgdo:CommunityCentre
Företag	Ett företag. A business company.	ric:CorporateBodyType	lgdo:Company
Omslagsformat	Bild- eller bildformat (t.ex. A1, A4). Picture or image format (e.g., A1, A4).	ric:CarrierExtent	cdesc:Format, ddesc:MeasurementType
Skala	Skala. Image scale.	ric:CarrierExtent	cdesc:Scale

Appendix 4 - Prompting Templates Used

Exp. 1

Hello. Are you aware of the RiC-O schema, its classes and properties?

If I give you the data schema/elements of an Archival institution

Can you do the mapping between the new schema and RiC-O?

Please map all remaining elements and then create a downloadable file for me.

I have produced a conceptual mapping between RiC-O and the Swedish National Archives Data Schema. Swedish National Archives Data Schema from now on I will abbreviate as RA.

If I give you:

1) the mapping between RA and RiC-O

2) the data schema/elements of another Archival institution

Can you see if you can do the mapping between the new schema and Ric-O based on my mapping?

Exp. 2

Hello. Are you aware of the Records-in-Contexts ontology?

I have produced a conceptual mapping between RiC-O and the Swedish National Archives Data Schema. Swedish National Archives Data Schema from now on I will abbreviate it as RA.

If I give you:

1) the mapping between RA and RiC-O

2) a mapping between RA and other vocabularies in the Cultural Heritage (CH) domain

Can you give me more recommendations for mapping with more Cultural Heritage domain ontologies?

I would like you to also make use of Linked Open Vocabularies, such as cdesc, ddesc, lgdo and more.
Based on the 2 files I provided you with and in a similar way, can you give me more recommendations for mapping with more Cultural Heritage domain ontologies (for example, more ontologies from LOV)?

Can you expand the mapping with more suggested alignments to other LOV? These can be CH- related but they could also belong to other knowledge domains, if that makes sense.

I would like you to be more context-specific. When we have an entity such as row 8, which refers to a sound recording, your recommendations are very generic. I want more context- tailored recommendations in order to enrich the metadata desciptiveness.

Yes, exactly. Please do the same for all rows and try to go as deep as possible in the contextual meaning of the fields in order to provide recommendations.

Now, what you did for the Extended CH Mapping file, I would like you to do for the RA RICO Class file. Do exactly the same process please.

Appendix 5 - A Custom Archival Institution Data Element Set

Data Elements	Description
Level	The hierarchical level of the material being described
Legal Status	A note as to whether the records being described are Public Records or not, as defined by the 1958 Act and its schedules (and other Public Records and FOI Acts)
Language	The language(s) in which the records are written
Reference	The unique identifier which links the catalogue to the records it represents and allows the user to order and refer to them
Former Reference (Department)	Unique identifier given to the material by the originating creator
Former Reference (Pro)	Former PRO/TNA identifier providing the link between the archival

	material and the description that represents it
Title	The name given to the unit of description. This should take the form, usually covered by standard or convention, appropriate to the level of description
Map Designation	Used for maps in clearly defined map series to provide that map series information
Creator(s)	Identifies the organisation(s) or individual(s) responsible for creating, accumulating or maintaining the records being described. Incorporates the authority controlled corporate or personal name(s) of the creator(s) of the records
Dates	Identifies and records the date(s) of creation of the records being described
Physical Description, Extent	The number of the records at the level being described
Physical Description, Form	The broad physical form (at department and division level 'form' is extended to mean series) of the records being described. The purpose is to give the user a basic idea of the type of document they will be faced with
Dimensions	The dimensions of the unit of description. This will usually only be used for individual, map(s), architectural drawing(s) or photograph(s)
Map Scale Number	Number providing the scale of a group of or individual map, plan, engineering/ technical/ architectural drawing
Physical Condition	A note concerning any aspect of the physical condition of the documents being described that may affect or limit their use. This will often, but not at all necessarily, be some form of damage to the document. This information

	should not be confused with that about documents which have not been made available for conservation reasons, which will be noted in 'Restrictions on Use'
Place of Deposit	The authority controlled name of the repository holding the records being described. This is a declaration of legal responsibility rather than a statement regarding the location of the physical repository. Records held by The National Archives may be held off site
Note	Specialised information that cannot be accommodated elsewhere (Archivist's note)
Access Conditions	Intended to give an indication of any conditions that restrict or affect access to the records at subsubseries level and above
Closure Type	Intended to identify conditions that restrict or affect access to pieces and items
Closure Code	A number code associated with the entry in the Access conditions element at piece and item level
Record Opening Date	The date a closed piece or item will be made available
Closure Status	Indicates the closure status of the description as well as that of the record/document itself
Restrictions on Use	A note to indicate restrictions to the use or reproduction of the material after access has been granted
Immediate Source of Acquisition	Records the authority controlled corporate or personal name of the depositor/donor from which the records were acquired and the year when

	they came to the PRO/TNA
Custodial History	Describes where and how records have been held from creation to transfer to PRO/TNA, giving those details of changes of ownership and/or custody that may be significant in terms of authority, integrity and interpretation
Accumulation Dates	The range of dates of accumulation of the records (series or subsubseries) being described. This cannot predate the existence of the creating body because these dates refer only to the record-keeping activities of the creator(s). The first date may be very different from the first date of creation of the records: the creator can only start accumulating when it/he/she comes into being; the records actually accumulated may well pre-date that coming into being
Appraisal, Destruction Information	Provides information on any appraisal, scheduling and destruction action taken; clarifies the process of determining the value, and thus the disposition of the records
Accruals	Intended to inform users of possible changes to the extent of the series. Indicates if future accruals are expected; can indicate quantity and frequency of transfers
Location of Originals	A note of the body, if known, that holds the originals if the records we hold are copies
Copies Information	Indicates existence, availability, location and format of copies of the unit of description, when known
Scope and Content	A prose statement giving information about the form and subject matter of the unit of description (often in terms of subjects, people and places)

Arrangement	Provides information on the arrangement of the unit of description; for example, the principle characteristics of the internal structure, the physical or logical ordering or filing sequence of the records, or how they have been treated by the PRO/TNA
Publication Note	A bibliographical reference to published (current, not obsolete) finding aids; published editions, transcripts and reproductions of records and significant publications based on the use, study or analysis of the records
Unpublished Finding Aids	A note with references to current, unpublished finding aids to the material being described
Related Material	A note identifying material that has a direct and significant connection to the unit of description. This will usually be in the form of references to material that dealt with similar functions at an earlier, or later, period to the unit of description
Separated Material	Identifies materials (and their location) that are organically part of the unit of description that have been physically separated or removed
Administrative, Bibliographical Background	Used to provide an administrative history, biographical details, or other historical statement about the corporate body/bodies or individual(s) responsible for creating and accumulating the records being described; placing the material in context and making it better understood
Index Terms: Corporate Names	The name(s) of corporate bodies whose activities are significant to the records being described
Index Terms: Personal Names	The name(s) of individuals whose activities are significant to the records being described

Index Terms: Places	The name(s) of places relevant to or significant to the level of the records under consideration. They will indicate any substantial set of records referring to a place, or to events which occurred there
Index Terms: Subjects	The subject terms which are significant to, or relevant to, or which summarise the records under consideration
Name	The legal or formal name by which a corporate body is commonly identified from items issued by the body, including its records, or from reference sources
Uncertain Start and End Date Codes	An indicator qualifying an uncertain date
Start Date	The date the corporate name came into use. This could be the date of a body's foundation/establishment or the date it changed its name
End Date	The date the corporate name ceased to be used. This could be the date of the body's dissolution or the date it changed its name
Remit and Function	The history of the corporate body
Variant	Versions of the corporate body's name in other languages or script forms
Validation	Sources used to validate the name and the rules used to construct it
Jurisdiction	The jurisdiction with which the corporate body is associated
National Place of Deposit Code	The place of deposit code
Surname	The last or most recent surname used by the individual, or, in instances where there is no surname, the forename, patronymic or toponym which stands as the main name

Gender Indicator	The gender of the individual
Title	Any forms of address attached to the individual in respect of the peerage or the ownership of, or association with, certain lands
Additional Elements of Name	Parts of surnames not used in the surname data element
Pre-Title	Any title that precedes the name
Forename	The forename (component of the name in addition to the surname) of an individual
Date of Birth	The date of birth of the individual
Bibliographical History	A brief description of the individual's life
Date of Death	The date of death of the individual
Name	The name of the place
Parish	The name of the civil parish
Town	The name of the town
County	The name of the county or other wider administrative unit
Country	The name of the country
Grid of Lat/Long	The grid reference or latitude/longitude of the place being described
Name Start Date	The date of a change of name/status
Name End Date	The date of a change of name/status

Place History	A brief history of the place