Umanistica Digitale – ISSN: 2532-8816 – n. 20, 2025: Special Issue
L. Giagnolini, A. Schimmenti, P. Bonora, F. Tomasi – *Expliciting Contexts: Semantic Knowledge Extraction from Traditional Archival Descriptions*
DOI: http://doi.org/10.6092/issn.2532-8816/21229

Expliciting Contexts: Semantic Knowledge Extraction from Traditional Archival Descriptions

Lucia Giagnolini

University of Bologna
lucia.giagnolini2@unibo.it

Andrea Schimmenti

University of Bologna
andrea.schimmenti2@unibo.it

Paolo Bonora

University of Bologna
paolo.bonora@unibo.it

Francesca Tomasi

University of Bologna
francesca.tomasi@unibo.it

Abstract

Archival finding aids are often only partially capable of fully expressing the informational potential of data due to the presence of numerous unstructured fields in the descriptions of documentary complexes. The prevalence of extensive literal sections, or full-text fields, limits both the possibility of semantic queries and the ability to uncover the latent contexts embedded in such unstructured text. This study proposes a methodology for the automatic extraction of knowledge (Knowledge Extraction, KE) from archival descriptions, aiming to enhance their structuring and semantic interoperability. Through a case study based on the Italian National Archival System (SAN) and leveraging ready-to-use tools such as Tint, FRED, and GPT-4o, we

conducted a preliminary evaluation of various morphosyntactic, lexical, and semantic analysis techniques. The most promising results highlighted the potential of Large Language Models (LLMs), leading to the development of a KE pipeline based on the open-source model Llama 3.3. The findings demonstrate a high capacity for extracting biographical events and relationships, achieving a good balance between precision and recall, thus confirming the validity of the approach. However, the need for a more robust software architecture emerges, as LLM-based pipelines must become truly scalable to enable effective integration into archival systems.

Keywords: Linked Open Data; information retrieval; knowledge extraction; knowledge representation; supervised annotation; archives; archival contexts; AIUCD2024

Gli strumenti di corredo archivistici sono spesso solo parzialmente capaci di esprimere il vero potenziale informativo dei dati, a causa della molteplicità di campi non strutturati presenti nelle descrizioni dei complessi documentari. La presenza di numerose sezioni *literal*, ovvero a testo pieno, limita, da un lato, la possibilità di interrogazioni a base semantica e, dall'altro, non consente l'apertura ai numerosi contesti latenti che tali porzioni di testo non strutturato veicolano. Questa ricerca propone una metodologia per l'estrazione automatica di conoscenza (Knowledge Extraction, KE) da descrizioni archivistiche, con l'obiettivo di migliorarne la strutturazione e l'interoperabilità semantica. Attraverso un caso di studio basato sul Sistema Archivistico Nazionale (SAN) e utilizzando strumenti *ready-to-use* come Tint, FRED e GPT-4o, si è valutata preliminarmente l'efficacia di diverse tecniche di analisi morfosintattica, lessicale e semantica. I risultati più promettenti hanno evidenziato il potenziale dei Large Language Model (LLM), portando allo sviluppo di una pipeline di estrazione della conoscenza basata sul modello open-source Llama 3.3. I risultati hanno dimostrato un'elevata capacità di estrazione di eventi biografici e relazioni, con un buon equilibrio tra precisione e recall, confermando la validità dell'approccio. Tuttavia, emerge l'esigenza di un'architettura software più robusta affinché le pipeline basate su LLM diventino davvero scalabili nell'ottica di un'integrazione nei sistemi archivistici.

Keywords: Linked Open Data; information retrieval; knowledge extraction; knowledge representation; supervised annotation; archivi; contesti archivistici; AIUCD2024

1. Introduction[1]

Archives are composed of elements characterized by unique characteristics. Together, these elements provide a stratified representation of the complex activities and entities involved in document creation. To make this representation explicit, it is essential to expose the network of relationships connecting individual components and linking the archive to its multiple contextual references [7].

[1] Author Contributions: L. Giagnolini developed Sections 1 and 2; Section 3 was developed jointly by L. Giagnolini and P. Bonora; A. Schimmenti developed Section 4; Section 5 was co-authored by L. Giagnolini and A. Schimmenti; F. Tomasi and P. Bonora supervised and revised the project and the manuscript. The conclusions reflect the collective insights of all authors. This article is an extended version of the following paper presented at the conference AIUCD2024: Giagnolini, Lucia, Bonora, Paolo and Francesca Tomasi, "Affinare il contesto: estrazione di informazioni strutturate per l'arricchimento dei contesti archivistici", in «Me.Te. Digitali. Mediterraneo in rete tra testi e contesti, Proceedings del XIII Convegno Annuale AIUCD2024», 2024, ISBN 978-88-942535-8-0, pp. 411-416. All URLs were last accessed on January 29, 2025.

While descriptions based on the ISAD(G)[2] have enabled a functional formalisation and structuring of the descriptive act, it is now widely acknowledged that applying this standard has led to representations primarily centred on hierarchical (thus strictly vertical) relationships. These are poorly permeable to contexts and only marginally enhance the portrayal of horizontal connections [7] [22]. For this reason, for over a decade, GLAM institutions have embraced the paradigm of Linked Open Data (LOD). This approach has required a systematic review of archival descriptions, deconstructing and restructuring their typology, granularity, and precision. The goal is to move beyond document-centric description schemas, adopting data-centric approaches that prioritize contextual relationships [10].

During the migration of traditional archival finding aids to LOD, highly significant informational blocks (such as historical and biographical notes) are often transposed merely as lengthy literal strings, i.e., plain text. These textual fields, while rich in information, could be structured in a more organized and functional manner, representing "il carburante indispensabile a far decollare il razzo dell'integrazione multicontestuale"[3] [21]. Indeed, while the Semantic Web has not fundamentally changed how institutions approach descriptive practices, it has emphasized the importance of explicit semantics, thereby facilitating interoperability and data reuse [19].

The textual contents of descriptive fields are expressed as aggregated string sequences. These could be made more explicit through the generation of additional triples, i.e. subject-predicate-object statements that describe relationships between entities in a structured, machine-readable way. Every new assertion expressed as a triple generates inference and new information: the more the contexts of these assertions grow and intersect, the richer the semantic network becomes, evolving into classified information [14]. Each new triple would convey a specific informational component from the descriptive text, such as references to institutions, people, events, places, and time coordinates. By systematically transforming existing data into semantic triples, we reuse trustworthy texts to extract both explicit and implicit relationships within textual content. This approach offers significant advantages: it enhances the depth of contextual knowledge, enables more sophisticated and precise searches, and provides support for disambiguating referenced entities.

This process is referred to as Knowledge Extraction (KE). KE typically encompasses various tasks that aim to automatically extract semantic information, both in vertical dimensions (e.g., taxonomies, classes, types, named entities) and horizontal dimensions (e.g., relations). It combines methodologies from Information Retrieval, Natural Language Processing (NLP), Symbolic Artificial Intelligence, and Machine Learning. By applying these methodologies, an unstructured information source, such as plain text, can be converted into structured, machine-readable formats.

The process of extracting entities from the text and assigning relational semantics among them constitutes, in essence, an interpretative act of the textual content [8]. Therefore, it is crucial that new triples, whether derived from supervised or unsupervised extraction processes, are explicitly identified as the result of a new analytical activity, distinct from the archival description that produced the original record. These additional triples should be accompanied by triples that explicitly declare their provenance, the methods of their production, and, ultimately, the attribution of responsibility. In other words, they should explicitly indicate their so-called provenance [19].

[2] Cfr. https://www.ica.org/resource/isadg-general-international-standard-archival-description-second-edition/

[3] "the fuel needed to launch the rocket of multi-contextual integration". Translation by the authors.

Therefore, to truly overcome hierarchical limitations and foster the creation of a semantically finer knowledge base, it is necessary to make better use of textual fields, structure their latent contexts, and provide appropriate documentation for the process of extracting new knowledge.

The remainder of this paper is organized as follows: Section 2 presents the proposed workflow for Knowledge Extraction (KE), Knowledge Representation (KR) and visualisation from preexisting archival descriptions. Section 3 describes our preliminary tests employing different ready-to-use KE tools. Section 4 details our proof of concept based on the proposed workflow and Llama 3.3. This includes the pipeline description, evaluation framework, and results. Section 5 discusses the findings and outlines directions for further developments of the pipeline. Finally, Section 6 concludes the paper, summarizing the main contributions and implications of the proposed approach for archival practice.

2. Workflow

To meet our objective, it is essential to clarify the steps of the KE process by designing a workflow that defines the type of analysis and the evaluation of the application output.

The approach we propose for implementing this process is structured as follows:

1. **Select the type of interpretative act delegated to the tool**, depending on the content to be analysed (e.g., morphosyntactic, lexical, or semantic analysis);
2. **Identify the technologies and corresponding implementations** based on the type of interpretative act expected (ranging from basic NLP techniques to Deep Learning (DL));
3. **Define a framework for evaluating the results and the quality of the automatic interpretative act**, where quality refers to the possibility of accessing reasonably reliable data, as they are part of a context that justifies and explains them [21]. The outputs of the interpretative act must be reviewed and validated by a domain expert to be considered reliable;
4. **Identify a model for consolidating the extracted knowledge** within the framework of a semantically controlled structure, ensuring interoperability for data access purposes (e.g., Entity-Relationship model in SQL; RDF with a Linked Open Data perspective, etc.);
5. **Model the criteria and methods for integrating the extracted knowledge** in alignment with the expressive capacity of the corresponding descriptive model (e.g., Dublin Core, RiC-O, SAN LOD) and editorial standards. This involves defining a provenance model that explicitly outlines the type of interpretative act, the tool and process used, the evaluation metrics (e.g., recall and precision), and the attribution of responsibility to the supervising expert;
6. **Assess strategies for linking the analysed data** in the native system with the resulting triples produced by the interpretative act;

7. **Model the user-system interaction** in terms of operational processes and interfaces, identifying information visualisation strategies that enable the extracted information content to be effectively presented and managed;
8. **Evaluate potential methods for enhancing the external tool** to improve its performance (e.g., producing annotated datasets, leveraging existing Pre-trained Models (PTLMs) or Large Language Models (LLMs)).

In this formulation, the process is abstract enough to be applied across various contexts and objectives of KE, operating at multiple levels – from surface lexical analysis to interpreting textual semantics. This means that the specific implementation may vary depending on several factors: the target document's structure and language style, the intended level of semantic granularity, the domain-specific ontological requirements, and the ultimate purpose of the extracted knowledge graph. On the other hand, existing KE tools are either extremely general or bespoke to specific domains and formats. There is no available tool designed for this specific task. Therefore, we selected general-purpose tools that we could adapt to this framework. Success in this task depends not only on technical implementation but also on contextual awareness to accurately preserve the nuanced relationships in the source materials.

3. Preliminary tests

Typically, the most extensive text fields in a traditional archival finding aid include the biographical note, the archival history, the system of arrangement, and the scope and content. To illustrate the potential outcomes of the methodological approach outlined above, we analyse the description of an archival creator within the Italian National Archiving System (SAN)[4]. This textual field corresponds to the value of "dc:description" and "abstract" properties in the SAN schema. As a sample text, we examine the "description" field in the SAN record dedicated to the biographical note on Andrea Costa (1851–1910)[5]:

> [Andrea Costa] was born in Imola on November 29, 1851, to Pietro and Rosa Tozzi in a practicing Catholic family of modest conditions. The following day he was baptised in St. Cassiano Cathedral with the names Andrea, Antonio, and Baldassarre and his godfather was Orso Orsini. He attended primary school run by a priest and in the school years 1866-1867 and 1867-1868 he attended the municipal technical school with Gaetano Darchini, Luigi Sassi, and Angelo Negri. In the school years 1868-1869 and 1869-1870 he attended the high school as an auditor for Italian and Latin literature lessons. On 15 December 1870, he enrolled in the Faculty of Philosophy and Fine Arts at the University of Bologna as an auditing student as he was unable to pay the regular admission fees, and to support himself, he worked as a scribe in an insurance agency in Imola. There a clerk, Paolo Renzi, associates him, or at least brings him close, to the International. He completed his novitiate in Imola and Bologna, in the atmosphere that soon became inflamed with enthusiasm for the

[4] http://san.beniculturali.it/

[5] http://dati.san.beniculturali.it/SAN/produttore_IT-ER-IBC_san.cat.sogP.66756

> Commune, and in contact with Carducci, who favoured him among his pupils[6].

Biographies can constitute a distinct literary genre with their own artistic conventions and narrative techniques. Literary biographies often contain rich layers of interpretive meaning, subjective opinions and analyses, and complex narrative devices. However, biographies can also be as straightforward as informational texts, such as biographical notes in archival finding aids. These notes follow an objective style and are usually linearly structured to report facts, though they can present variations in the depth of details. Informational texts typically use denotative language, emphasizing literal and precise meanings over connotative or figurative expressions. They are predominantly written in the past tense when describing historical events and figures and maintain an objective tone that prioritizes factual accuracy over artistic expressivity. These texts generally adhere to the Five Ws rule: who did what, when, where, why, and how. Their structure is usually made of a series of well-defined paragraphs, each containing one or more discrete events of the subject's life, often representing a distinct temporal unit or thematic cluster of biographical information.

The distinction between literary and informational biographies is crucial when applying NLP tools and KE techniques. This structural difference enables automated systems to bypass complex semantic parsing layers when processing informational biographies, as their content adheres to more predictable patterns. Given these characteristics, preliminary tests were made with a series of general-purpose tools to assess the general feasibility of the task. Given the lack of annotated training data in the archival domain, we propose leveraging existing off-the-shelf technologies rather than developing specialized models. Here, we document the results of three of them, Tint[7], FRED[8], and GPT-4o[9], in the analysis of the first paragraph of the target text (steps no. 1 and no. 2 of the workflow) [13]:

1. Tint provides robust linguistic analysis capabilities, including part-of-speech tagging, dependency parsing, and Named Entity Recognition (NER) specifically trained on Italian texts [1].
2. FRED is a machine reader tool capable of discourse representation. It is domain-independent and intended to be used as middleware, especially for KE tasks [12].
3. PTLMs and especially LLMs are valuable tools for any semantic-based text analysis and/or KE task. For this step, we chose, among available LLMs, GPT-4o through the ChatGPT interface [26].

[6] The translation from Italian to English has been provided by the authors to enhance accessibility for readers. However, all analyses were conducted on the original Italian text, which is available on the SAN page dedicated to this record: http://dati.san.beniculturali.it/SAN/produttore_IT-ER-IBC_san.cat.sogP.66756

[7] For the purposes of the paper presented at AIUCD2024, the "Online demo" version available at the link: https://dh.fbk.eu/Tint-demo/ was used.

[8] For the purposes of the paper presented at AIUCD2024, the "Online demo" version available at the link: http://wit.istc.cnr.it/stlab-tools/fred/demo/ was used.

[9] For the purposes of the paper presented at AIUCD2024, GPT-3.5 https://chat.openai.com/ was used.

Through the application of these tools to the biographical note from Andrea Costa's archival finding aid, we aimed to test their fitness for KE from archival descriptive texts.

Tint

The application of Tint for the analysis of the first paragraph of the biographical note enabled the identification of organisations, locations, and person names through NER (Figure 1), as well as the syntactic dependencies within the text. Additionally, it automatically classified the parts of speech[10] (Figure 2).

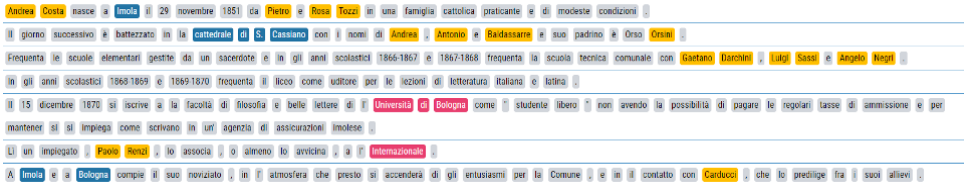

Figure 1. Entities recognized in the text and their classification.

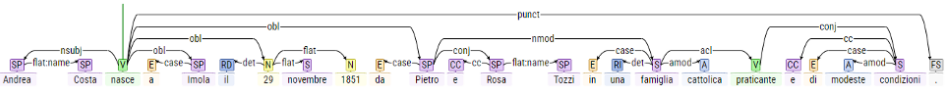

Figure 2. Graph of syntactic dependencies relating to the first sentence.

Tint demonstrates effectiveness in identifying the morphological structure of the text and basic named entities. Through additional rule-based extraction methods, it is possible to generate triples regarding specific events – for example, extracting birth information by targeting verbs like "*nascere*" (to be born) and their corresponding arguments. However, this approach, like many traditional NLP pipelines, presents several limitations:

- while robust in basic linguistic analysis, it provides limited semantic understanding;
- the necessity for manual rule creation for each type of information to be extracted reduces scalability;
- the output requires significant expert validation, highly limiting the benefits of automation;
- rule-based approaches struggle with handling variations in language and context.

While Tint offers the advantage of immediate deployment for basic linguistic analysis with minimal setup effort, its utility is primarily limited to preliminary text processing and basic entity recognition. This limitation is characteristic not just of Tint, but of traditional NLP approaches that rely heavily on explicit linguistic rules and pattern matching [25].

[10] To access the complete results of the analysis using Tint, see Giagnolini, Lucia, and Paolo Bonora. "Refining Context: Extracting Structured Information for Archival Context Enrichment. Results Of The Analysis Performed With Tint". figshare, January 31, 2024. https://doi.org/10.6084/m9.figshare.25119116.v4

FRED

FRED annotates semantic frames found within the text. For the sentence "Andrea Costa was born in Imola on November 29, 1851, to Pietro and Rosa Tozzi in a practicing Catholic family of modest means", FRED successfully identifies and represents several semantic relationships. It additionally performs NER and Entity Linking with DBpedia.

The application of FRED produced a unified and formalized graph representation of facts and concepts expressed by the text in natural language such as, for example, the interpretation of Andrea Costa's birth conditions (Figure 3).

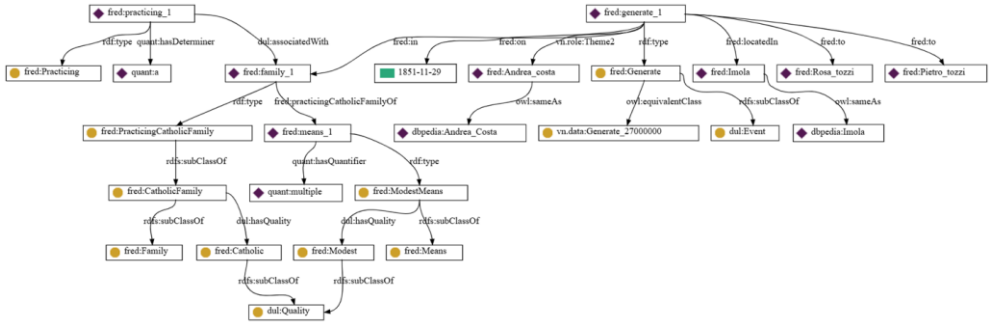

Figure 3. FRED graph output over the birth of Andrea Costa. The input text was translated in English.

The core event (fred:Generate) is linked to:

[1] Temporal information (fred:on 1851-11-29)

- Location (fred:locatedIn fred:Imola)
- The subject (fred:Andrea_costa, linked to dbpedia:Andrea_Costa)
- Parents (fred:to relationships to fred:Pietro_tozzi and fred:Rosa_tozzi)

The family context is represented through a complex hierarchical structure:

- fred:family_1 is categorized as fred:PracticingCatholicFamily
- This is further broken down through subclass relationships:
    - fred:CatholicFamily -> fred:Family
    - Properties fred:Catholic and fred:Modest are connected through quality relationships
- The economic status (ModestMeans) is represented with appropriate quantifiers

However, using the same sentence in the original Italian yields different results; classes are not identified correctly, the graph is more horizontally spread and there are many more unnecessary intermediate nodes[11] (Figure 4).

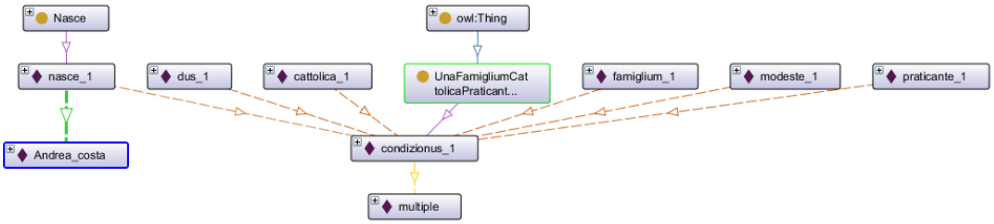

Figure 4. FRED graph output over the birth of Andrea Costa. The input text was kept in Italian.

While FRED demonstrates impressive quality across various inputs and shows to be promising in English, its characteristic production of highly variable graph morphologies means that the resulting structures typically require direct usage with domain-specific ontologies.

LLM

GPT-4 was used to explore information emergence through prompts adopting different strategies and addressing specific objectives. As an example, the prompt "List the entities that emerge from the text" provided the following results:

- **Place of birth**: Imola (29 November 1851)
- **Parents**: Pietro and Rosa Tozzi
- **Family background**: A practicing Catholic family of modest conditions
- **Religious acts**: Baptized in the cathedral of S. Cassiano with the names Andrea, Antonio, and Baldassarre; godfather Orso Orsini
- **Education**: Primary school run by a priest; attended municipal technical school with Gaetano Darchini, Luigi Sassi, and Angelo Negri; attended high school as an auditor for Italian and Latin literature lessons
- **University**: Enrolled in the Faculty of Philosophy and Fine Arts at the University of Bologna on 15 December 1870 as a 'auditing student'
- **Financial situation**: He was unable to pay the regular university admission fees; he supported himself by working as a scribe in an insurance agency in Imola
- **Political influence**: Associated or approached the International by Paolo Renzi, a clerk in the insurance agency

[11] To access the complete results of the analysis using FRED on the Italian text, see Giagnolini, Lucia, and Paolo Bonora. "Refining Context: Extracting Structured Information for Archival Context Enrichment. Results Of The Analysis Performed With Fred". figshare, April 3, 2024. https://doi.org/10.6084/m9.figshare.25534225.v1

- **Historical/political context:** Novitiate in Imola and Bologna in the atmosphere of enthusiasm for the Commune; contact with Carducci, who favored him among his pupils.

GPT-4o, being well versed in Natural Language Understanding such as question answering, disambiguation, and information retrieval, gave a comprehensive response as expected [17]. However, transforming this high-quality information into an equivalent knowledge graph is not straightforward. To this end, we performed additional tests. We asked the model to produce a Entity-Relationship (ER) diagram using the Mermaid syntax[12], producing encouraging results (Figure 5).

"You are an expert Knowledge Extraction model. Annotate all the entities, concepts, relationships, and other interesting information that can be structured. Be sure to annotate each relationship with its date, location, and participants. Return the output in Mermaid syntax. Text to annotate: [input text]"

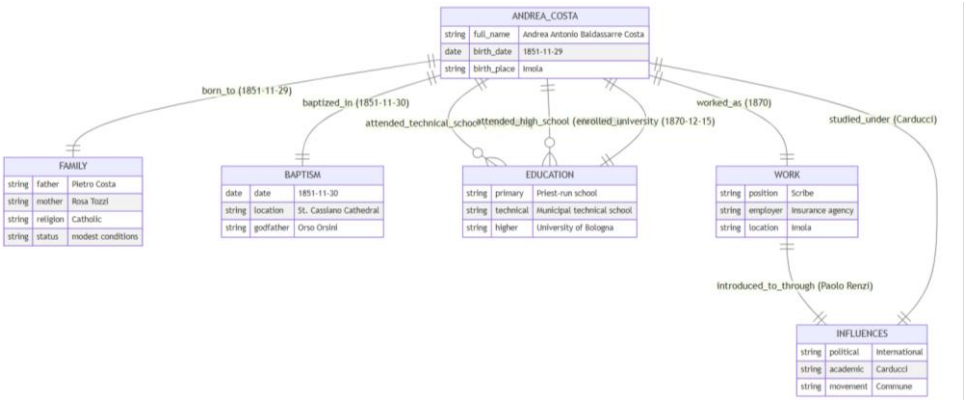

Figure 5. ER diagram of GPT-4o output of Andrea Costa's birth conditions (visualisation created by mermaid.live).

On balance, Tint, FRED and GPT-4o returned valid options for information structuring, but every output must be validated through supervision (step no. 3 of the workflow). This validation step is crucial, as automatic extraction systems, despite their sophistication, cannot fully replace human expertise in interpreting archival descriptions' contents. The quality of the extracted information must be assessed not just for factual accuracy, but also for its contextual relevance and semantic coherence within the broader archival framework.

Overall, the results provided by GPT-4o emerged as particularly promising. The results obtained through GPT-4o offered a range of information that extends beyond the identification of canonical entities. For instance, the model proved capable of extracting not only the individual name "Andrea Costa" but also his complete baptismal name "Andrea Antonio Baldassare Costa".
In general, transformer-based models have demonstrated strong performance in capturing complex semantic relationships and contextual dependencies through example-based learning

[12] Mermaid is a JavaScript-based markdown for defining diagrams. https://mermaid.js.org/

and reasoning elicitation [5][26]. Additionally, since LLMs have demonstrated what are usually called "emergent abilities", i.e., the capability of generalising to provide results even with previously unseen tasks and inputs [27], they can also be instructed without targeted training for new tasks.

However, using an open extraction approach, where the model is free to describe whatever relations it finds in the text, is considerably easier than adhering to a given schema. While these models perform very well in few-shot learning scenarios and handling domain-specific languages [5], they present notable limitations for our task, particularly in maintaining consistent output structure and adhering to strict KE schemas. LLMs, in general, suffer from a lack of inner structured control mechanisms, with inherent variability in their generative outputs [5]. While solutions exist (such as enforcing JSON object outputs in the latest OpenAI models or through function calling), these models require, at a certain scale, a significant investment for locally hosted models.

Additionally, ethical concerns about the openness of non-open-source models are significant, ranging from lack of transparency in training data and processes to limited accountability for bias and errors. Also, relying on a specific feature of a specific private company hinders reproducibility in the academic community, as researchers might not be able to verify or build upon published results due to changes in API features, pricing models, or access policies. This dependency on proprietary solutions poses challenges for long-term research sustainability and scientific validation.

4. Proof of concept

Given the previous exploration, LLMs seemed more adequate for the task. To solve some of the described criticalities, we opted for an extended proof of concept based on Llama, since the Llama family of models is open source and performs similarly to GPT-4o[13]. According to Meta's official model card, Llama-3.3-70B-Instruct achieves 86.0 on MMLU and 92.1 in IFEval, benchmarks that evaluate the model's NLU capabilities and instruction following, respectively (reiterating step no. 2 of the workflow)[14].

4.1 Pipeline Description

A mid-sized model such as Llama 70B requires fractioning the task into smaller steps, both for its smaller context size and its capabilities. This pipeline approach ensures the ability to check the process step-by-step, and to progressively add instructions or additional information. We modelled the pipeline around seven event types, each covering a possible important part of any biography (in our case, the biography of Andrea Costa). Starting from the plain text as input, the pipeline consists of three steps:

1. Identification of relevant instances of the given events and related contents;
2. extraction of information from related contents for each of the instances returned from the first step;

[13] Model card: https://huggingface.co/meta-llama/Llama-3.3-70B-Instruct

[14] The full code developed for this proof of concept is available on a dedicated GitHub repository https://github.com/aschimmenti/expliciting-context

3. representation of extracted information in the given output schema.

The general schema for events was loosely based on how DOLCE+DnS [4] represents events (i.e., as situations), where entities participate with specific roles during defined temporal and spatial contexts. The assumption is that following a foundational ontology simplifies any later process of reconciliation with a chosen model (step no. 4 of the framework).

The key assumptions when using DOLCE +DnS are:

- Framework:
    - An event is fundamentally a situation
    - It requires at least one participating entity
    - It occurs in a specific spatiotemporal context
- Participation Structure:
    - An entity is involved in an event
    - The entity can have a specific role when involved in it
- Temporal Dimension:
    - Events are anchored to points or intervals in time
    - They may have precise or approximate temporal boundaries
    - They may include duration or instantaneous occurrence
- Spatial Context:
    - Events occur in zero or more locations[15]
    - They may involve multiple spatial references
    - Location may have a situational role (e.g. "meeting location").

While this event schema could theoretically be applied to any occurrence, practical considerations led us to constrain it to specific biographical event classes, as described below. Our approach represents a hybrid solution between open and closed KE methodologies, employing a classification-based filter to maintain both flexibility and analytical precision.

The final schemata consider seven main biographical event classes, each structured to capture specific aspects of a historical figure's life:

1. **Birth and death events.** Birth events track three specific roles (parent, newborn, and circumstantial participant) along with birth date and location. This allows for complete family context reconstruction at a specific point in time and place. Death events follow a similar structure, focusing on the dying entity (dead, and eventual participant) and including the cause of death.

[15] Zero is accepted for instances in which the source does not mention a location.

2. **Education and employment events.** These events define roles for people (e.g., teacher, student) and organisations (e.g., school, university). The organisation is treated as a distinct entity rather than a location. Both event types include temporal context.
3. **Relationship events.** These events capture both personal and professional connections (e.g. friend, mentor, spouse), enabling the reconstruction of social networks over time. The schema includes temporal and spatial dimensions to track how relationships evolve and where they occur if the information is present.
4. **Political events.** This schema was the most complex to implement, given the general scope of the political phenomenology. It tracks an activity, entities with different types and roles, locations, and time frames.
5. **Record creation events.** This schema captures intellectual production, recording authorship, and publication details such as dates, genre and content.

Although the model's context window capacity (128,000 tokens for Llama 3.3) would technically allow simultaneous processing of both the input text and the output schema, such an approach could impair the model's disambiguation capabilities due to token density. Decomposing the extraction process into discrete steps enhances control over the pipeline. The key challenge lies in achieving optimal information density providing enough context for accurate extraction while avoiding cognitive overload that could compromise the model's performance. The steps of the pipeline from the selected paragraph to the output are shown in Figure 6.

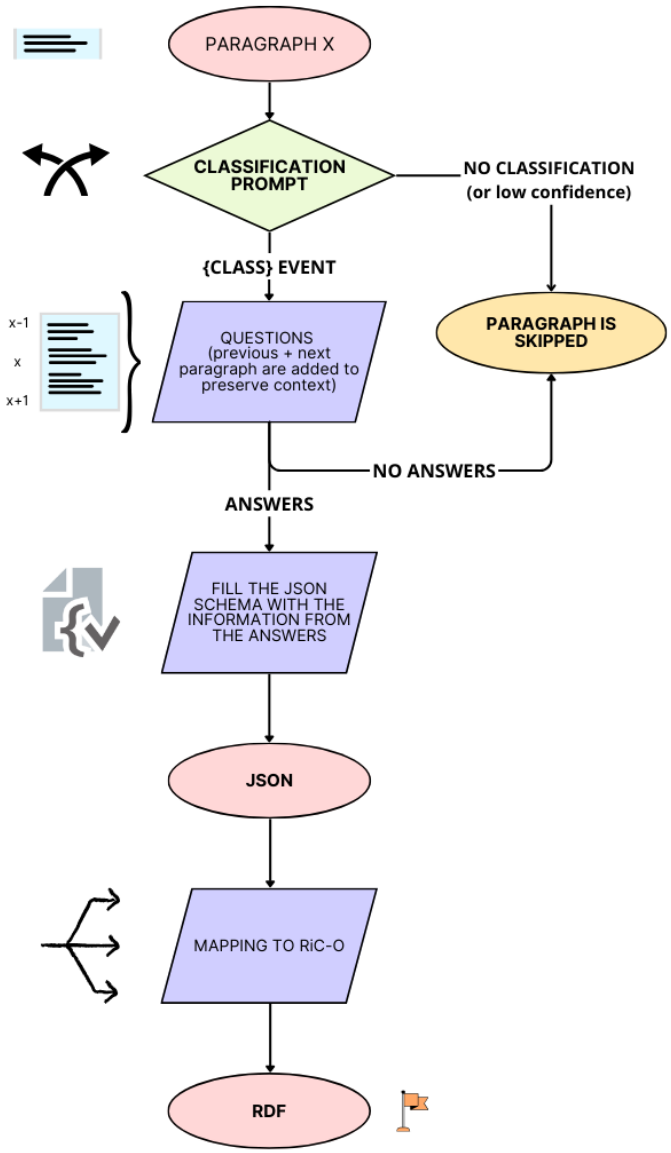

Figure 6. Flowchart of the KE pipeline.

The first part of the pipeline is shown in Listing 1. It instructs the model to classify the text into one or more of the classes mentioned above.

```
The following text contains a snippet of the biography of
{entity}. Classify the text depending on what is being
discussed. Use one or more of the following classes and
return the classification inside a JSON. The event must be
categorized independently of whether the event is
happening to {entity} or to someone mentioned in his
biography.
{classes_list}
Text: {text}
Return only a JSON array of classifications. If no proper
classification is possible, return any class with 0.0
confidence.
OUTPUT SCHEMA: [{{"type": "<EVENT_TYPE>", "confidence":
0.0-1.0, "reason": "<explanation>"}}]
```

Listing 1. First prompt of the pipeline. Graph parentheses represent injected variables.

Once the classification is done, each class detected with a confidence higher than 0.7 (empirical threshold) is retained and used for the second step. The second step injects a set of questions alongside the input text (Listing 2) with redundant instructions. We added additional context and examples to this step after multiple tests to improve the model performance. The target text contains both the preceding and following paragraphs as context.

```
The following text has been classified as describing a
'{event_type}' event in Andrea Costa's life.
### Context:
EVENT TYPE: {event_type}
**Previous context:** {prev_context or 'None'}
**Target text:** {text}
**Following context:** {next_context or 'None'}

### Instructions:
1. Read the questions carefully and only answer the questions
with information relevant to the {event_type} context.
2. Read the **target text** carefully, as it contains the
primary information you need.
3. Use the additional context (previous and following) only
as supplementary information when the target text alone does
not provide the full information about the {event_type}
event.
4. Assume the event involves {subject} if no explicit subject
is mentioned in the target text.
### Questions:
{questions}
### Examples:
{examples}
### Requirements:
- Provide concise, direct answers to each question in the
order listed.
- Focus on entities, dates, and their actions.
- Avoid speculation or assumptions not supported by the
provided text.
- Use dates in **DD/MM/YYYY** format or state the year if
precise dates are unavailable.
- Highlight the specific relations between entities,
institutions, and other places if any.
- Do not comment on any more information than asked.
- Keep the original language for entity labels.
- Return only the answers.
```

Listing 2. Second prompt of the pipeline. Graph parentheses represent injected variables.

While the second and third steps of the pipeline could be one, separating the questions' answering from the JSON output reduces the workload ensuring consistency and schema adherence. This output is then sent to the last step of the pipeline.

The last prompt merely instructs the model to return a JSON output from the answer's information following the given schema. Listing 3 shows the schema for the "Political event" class:

```
"POLITICS": {
  "instruction": "A political event encompasses any politically significant action or occurrence that involves participants with context-dependent roles in a defined spatiotemporal setting. Participants must be of type PERSON, ORGANISATION, or GROUP. Location and temporal data are captured as separate elements from participant information.",
  "description": "<Detailed description of the event>",
  "properties": {
   "actions": [
    {
     "action": "<specific event or action that the entities suffer or cause>",
     "participants":[
      {
       "name": "<Participant's name>",
       "type": "<person/organisation/group>",
       "role": "<role of the participant in the action>"
      }
     ],
     "date": {
      "startDate": "<Start date of subject's relation>",
      "endDate": "<End date of subject's relation>"
     },
     "location": [
      {
       "label": "<Location name>",
       "description": "<Detailed description of the location>"
      }
     ]
    }
   ]
  }
 }
```

Listing 3. Event schema of the pipeline for the class "POLITICS".

The output is additionally processed through simple rules, for example, to check whether the JSON output was indeed a valid JSON. The JSON output is ultimately processed and mapped to classes and properties defined by the Records in Contexts Ontology (RiC-O)[16], an OWL ontology for describing archival record resources and their contextual entities released by the International Council on Archives (ICA). This step ensures the extracted data adheres to a

16 https://github.com/ICA-EGAD/RiC-O/blob/master/ontology/previous-versions/RiC-O_v1-0_release/RiC-O_1-0.rdf

standardized semantic framework, facilitating interoperability and alignment with archival description standards[17].

While this modular approach introduces additional computational overhead compared to end-to-end extraction, it enables finer control over each step of the process and allows for targeted improvements where needed. The critical question, however, is how effectively this pipeline performs in the specific context of archival descriptions – a question that required a systematic evaluation framework.

4.2 Evaluation Framework

Generative KE outputs have a fundamental difference from rule-based or encoder-based models: while the labels and format of the output might differ from the ground truth or desired output, they can still be correct. This presents a well-known challenge when evaluating LLM performance. Therefore, we propose a three-level evaluation framework that operates at the structural, informational, and interpretational levels, employing both quantitative and qualitative metrics.

Given the dimension of our trial, the assessment of the model performance can be performed by hand. Human evaluation was also identified as a necessary step in the workflow to annotate any archival finding aid generated automatically (reiterating step no. 3 of the workflow).

4.2.1 Structural level

The structural evaluation assesses how well outputs adhere to the intended schema design, examining both overall performance and specific event types. We focus on two main criteria: *schema adherence* (checking if outputs strictly follow the predefined event schema format and requirements) and *consistency* (verifying that similar information is represented uniformly across different event types and that structural patterns are maintained). This evaluation precedes the deeper analysis of KE quality, counting the number of generated events that adhere to the JSON schema.

4.2.2 Information level

Information extraction performance is evaluated using three key metrics: Accuracy, Recall, and F1 score. These metrics are calculated using a 2x2 confusion matrix that categorizes results into four possible outcomes:

- True Positive (TP): Information is correctly identified and accurately reported in the output matching its presence in the input. The information must be both present and categorized correctly (e.g., if a date is present but incorrectly labeled as an endDate, it does not count as a true positive).
- True Negative (TN): Information is correctly identified as absent from both the input text and the output.

[17] The dataset containing the pipeline's output is available at: https://zenodo.org/records/14753707?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6ImRjZDQwZjZkLTI1NDUtNDM0OS04Y2Q0LTA0NGY5YzgxNjYyNCIsImRhdGEiOnt9LCJyYW5kb20iOiIzZjY4ZmYzMzU0OTAyOTk4NDJhMjBlYzA3NzM3ZWVmZCJ9.GlUZO4tReb n42QlSJap0XdAXOQdAXHgchdR2uqCapsosc9O0LffRG5yc0dnYjRgoNsSyGKVaLlMawZBhJWAaRQ

- False Positive (FP): Information is reported in the output but is either absent from or different in the input text.
- False Negative (FN): Information is present in the input text but not reported in the output (incorrectly marked as absent).

We will report overall and per-class scores. Accuracy will measure the proportion of correct predictions (both true positives and true negatives) out of all predictions; Recall will measure the proportion of actual positive cases that were correctly identified; while the F1 score will provide a balanced metric combining precision and recall. Event classification score will be reported as well.

4.2.3 Interpretative level

Since the process of KE can be described as an interpretation act, we also evaluate the accuracy of content interpretation by examining how correctly the system understands and represents information from the source. This includes assessing relationships' accuracy (verifying that connections between entities are properly identified), role attribution (confirming that roles are correctly assigned), and context preservation (ensuring the event context is preserved in the output). This evaluation is performed by scoring, from 1 to 10, how much information is correctly preserved by "verbalising" the extracted data and comparing it with the source.

4.3 Evaluation Results

Evaluating structured data output from LLMs is inherently complex, requiring a combination of precision, thoroughness, and consistency. In this section, we present both quantitative and qualitative evaluations to assess the reliability of extracted data and the schema adherence of outputs.

4.3.1 Quantitative evaluation

Structural level. Evaluating structured data is a long and tedious process. To ensure precision and agreement between the evaluators, a simple web app[18] was built. It compares the extracted events to the schema (Figure 7). Two checkboxes track whether the output is valid and whether the classification is correct. Four radio buttons (TP, FP, FN, TN) evaluate the per-field output. Once the evaluation is complete, a report can be downloaded.

[18] https://github.com/aschimmenti/expliciting-context/tree/main/evaluation-app

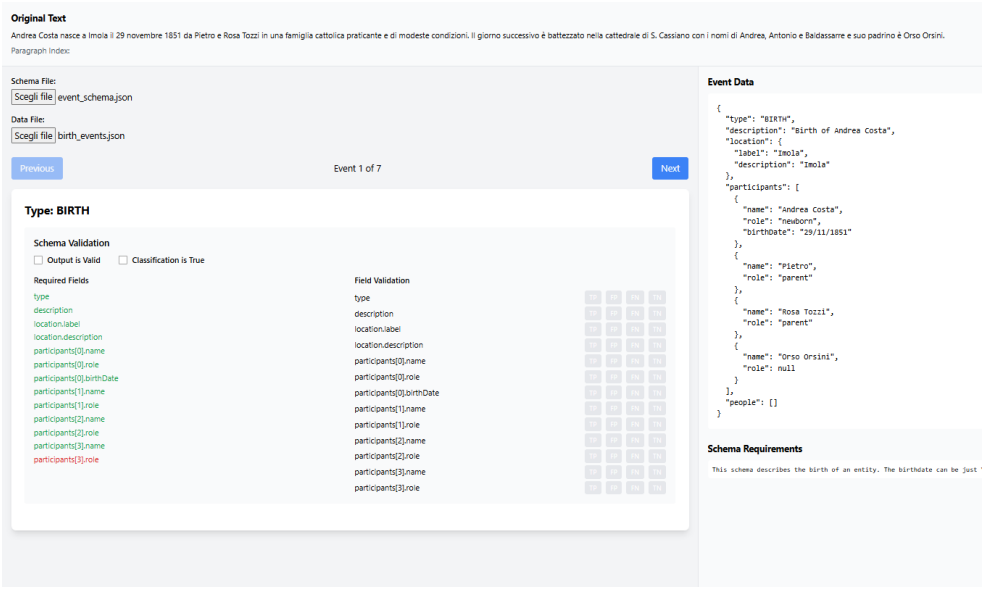

Figure 7. Output evaluation web application. In the example, the Birth Event of Andrea Costa is being evaluated.

One of the mentioned concerns using LLMs for structured data extraction was their capability to conform the output to a given schema. In our analysis, approximately 10% of outputs contained either minor errors (e.g., adding a "comment" in the schema or an additional field) or hallucinations. We classified both types as schema violations since even a 1% error rate means the JSON output becomes inconsistent and requires post-processing to ensure schema conformity. It must be noted that most of the hallucinations made sense (e.g., in one political event, the model inferred a correct date from the context but changed the key to "presumed_starting_date").

Valid Outputs	
Total Events	51
Schema Validity	46 (90.2%)

Table 1. Valid outputs percentage

To avoid this outcome, many libraries, tools, and functions have been proposed outside of the OpenAI models mentioned[19]. A fully developed LLM-based software could rely on open-source libraries such as LangChain[20] and Ollama[21] for structured output.

[19] For this proof of concept, we only used a simple rule-based library, json-repair (https://github.com/mangiucugna/json_repair).

[20] https://www.langchain.com/

Information level. The classification step was in 100% of the cases correct. Overlapping happened as well, with a single event (e.g. Andrea Costa's contacts with Jules Guesde, a French politician and journalist) being structured twice, as a relationship and as a political event.

The precision and recall metrics exceeded our expectations (Table 2). The high recall (0.982) indicates that when information is present in the input text, the model successfully extracts it, while the slightly lower precision (0.947) suggests the model occasionally produces incorrect or hallucinated information. This pattern - higher recall than precision - indicates the model prioritizes finding all relevant information even at the cost of occasional false positives, which is often desirable in KE tasks where missing data is typically more problematic than including extraneous information that can be filtered out during post-processing. It must also be kept in mind that the inferences were performed without injecting a controlled vocabulary (apart from the entity types) for fields such as the roles.

Macro-average metrics[22]	
Precision	0.947
Recall	0.982
F1	0.964

Table 2. Macro Precision, Recall, F1

Micro precision scores were generally above expectations except for Employment and Education events. The errors were caused by the model not identifying the correct dates and roles. It must be also noted that while the political events were generally correct, the roles were the most verbose.

Class	**Precision**	**Recall**	**F1**
BIRTH	0.999	0.960	0.979
DEATH	0.999	0.875	0.933
DOCUMENT	0.926	0.999	0.962
EDUCATION	0.902	0.841	0.871
EMPLOYMENT	0.840	0.971	0.901
POLITICS	0.952	0.999	0.975
RELATIONSHIP	0.999	0.955	0.977

Table 3. Micro Precision, Recall, F1

4.3.2 Qualitative evaluation

Interpretative level. Given the complexities of KE from archival records, qualitative evaluation is essential to complement quantitative metrics. While quantitative evaluation can identify structural and classification features, it does not capture the nuances of context, relationships, or

[22] The macro-average metrics are computed as the arithmetic means of Precision, Recall and F1 across all classes.

roles embedded within the data. For this reason, a qualitative assessment was performed to evaluate the interpretative accuracy and contextual fidelity of extracted events.

The qualitative evaluation was conducted on event-type basis (birth, death, education, employment, political, and record creation events), following an agreement between the evaluators. Each extracted event has been compared with the original textual paragraph (associated with a "paragraph index" number) focusing on the nuances of specific event types while applying a unified scoring framework. Starting from a maximum evaluation grade of 10 (meaning that all semantic content and contextual relationships from the source were maintained with full accuracy), points were subtracted when errors were identified.

1. Critical data missing: Missing essential data (e.g., a birth date in a birth event) incurred a flat penalty of -1 point.
2. Subevent omissions: Missing contextual details or subevents (e.g., the absence of a person's role within an event) were penalized as semi-errors, with a deduction of -0.5 points per missing element.
3. Annotation errors: Misclassification or annotation errors (e.g., confusing the role of a person or an institution within an event) were considered significant errors, and penalized by -0.8 points.
4. Incorrect or hallucinated information: Completely incorrect data due to excess (hallucinations) was penalized by -1 point per instance.

Each evaluation instance was documented through the indication of the paragraph index, the overall evaluation, and notes on the specific errors and their impact on the quality of the extraction. The reported scores are the result of a compromise between the two evaluators.

CLASS	**Mean Score**	**SUPPORT**	**Performance**	**Errors**
Birth	8.8	7	Generally accurate classifications	Missing dates with multiple concurrent birth events
Death	10	2	Well structured, dates and locations are correct.	
Relationship	9.5	4	Well structured, dates and locations are correct	Overstated relationship roles in some instances (e.g. "alliance" for a simple opinion agreement)
Education	8.4	3	Generally accurate	Errors in role descriptions and dates when processing multiple concurrent education events
Employment	8.8	6	Generally accurate	Errors in role descriptions and dates

				when processing multiple concurrent education events
Politics	8.3	26	Generally good grasp of context and events	Political roles and involvement are in some instances too generic or too context-dependent (e.g. the label "His followers" instead of "Andrea Costa's Followers")
Document	8.1	3	Generally accurate	Missing metadata (such as date and location)
Total	**8.8**	**51**	Good grasp of context, event succession, involved entities	Issues with role assignment and less explicit dates and locations

Table 4. Summary of the evaluations divided per-class.

5. Discussion e Future Works

The results obtained through the proof of concept are encouraging. The LLM-based pipeline demonstrated good performance in extracting structured information from archival descriptions, with high precision (0.947) and recall (0.982) metrics, while also providing a comprehensive contextualisation (8.8/10). Here we provide an example: the attempted insurrection of 1874 in Bologna and Romagna, when Costa tried to agitate multiple socialist groups alongside Antonio Cornacchia, with the anticipated presence of Michail Bakunin.

Input: "1874, a year of severe economic crisis, marked by widespread popular exasperation and numerous protests especially against the grain tax, was chosen by Italian internationalists for their first insurrectionary attempt. Costa was the main organizer of the movement, which was supposed to have Bologna as its epicentre and see the participation of Bakunin in person. However, inadequately organized, lacking the expected popular support, and preventively monitored by the police, the attempt failed miserably: Costa was arrested even before the attempt, and the small column of Romagna's Internationalists marching on Bologna under the command of Antonio Cornacchia was captured without a fight, while Bakunin managed to slip away, incognito as he had arrived[23]".

[23] The translation from Italian to English has been provided by the authors to enhance accessibility for readers. However, all analyses were conducted on the original Italian text, here specifically: "Il 1874, anno di grave crisi economica, contrassegnato da un diffuso stato di esasperazione popolare e da numerose agitazioni specie contro la tassa sul macinato, è prescelto dagli internazionalisti italiani per l'attuazione del loro primo tentativo insurrezionale. Costa è il principale organizzatore

```
Action: Attempted insurrection
      Time Period:
      - Begin: 1874

      Locations:
      - Bologna (intended epicenter of the insurrection)

      - Romagna (origin of the Romagna's internationalists)

      Key Participants
      - Andrea Costa
        Role: Primary organizer
        Type: Individual

      - Antonio Cornacchia
        Role: Commander of the Romagna's internationalists
        Type: Individual

      - Bakunin
        Role: Expected participant
        Type: Individual

      - Italian internationalists
        Role: Participants
        Type: Group

      - The Police
        Role: Opponent
        Type: Organisation
```

Listing 4. JSON output of the model from the text. It is represented as a list for readability.

As additional proof, the JSON output (Listing 4) can be converted by an LLM into a paragraph, for readability, during the qualitative evaluation step. The JSON content, elaborated by Llama 3.3 70B, provides the following reconstructed text:

> In 1874, an insurrectional plot began to take shape in Italy, concentrated mainly in Bologna with connections in the Romagna region. The main organiser was Andrea Costa, who worked together with Antonio Cornacchia, commander of the Romagna internationalists. The group of

del moto, che avrebbe dovuto avere come epicentro Bologna e vedere la partecipazione di Bakunin in persona. Ma, non adeguatamente organizzato, privo dell'atteso appoggio popolare, e preventivamente controllato dalla polizia, il tentativo fallisce miseramente: Costa viene arrestato ancor prima dell'inizio e la sparuta colonna degli internazionalisti romagnoli in marcia su Bologna al comando di Antonio Cornacchia viene catturata senza colpo ferire, mentre Bakunin riesce a dileguarsi, in incognito come era giunto".

> Italian internationalists acted as both organiser and participant, and Bakunin was expected to join as a participant and leader in the insurrection attempt[24].

While factually correct, an in-depth analysis of the output raises a few issues. Event representation is a known challenge, especially for complex historical events through semantic data structures. In this case, the problem is related to the definition and labelling of participants roles. The current implementation shows inconsistencies in labelling, moving between broad categorisations and highly detailed descriptions that embed temporal or conditional information directly in the role labels. The 1874 insurrection attempt provides a clear example: Bakunin is labelled as an *expected participant*, which creates two key problems. First, it places factual and anticipated participation on the same semantic level. Second, it affects data retrieval accuracy – a query for participants in the 1874 insurrection would include both actual participants and those who were merely expected to take part, reducing the truth value of the query. This inconsistency highlights the need for controlled vocabularies that can be suggested to the LLM to standardize role descriptions while maintaining the semantic richness of the original text.

The qualitative evaluation, while generally promising, also revealed concerns about knowledge loss during the extraction process. While the LLM-based approach showed significant accuracy in identifying and structuring explicit information, some nuanced contextual information expressed in the original narrative form may not be captured by the event schema. To assess the relevance of the lost information from the functional point of view the engagement of archivists and users is mandatory.

Several directions for future research emerge from these findings:

- The pipeline certainly needs to be tested on a broader corpus of case studies that should include a significant variety of biographical profiles, also pertaining to different eras, cultures, ethnicities and genders.
- The integration of additional event types and more expressive schemas can better represent the complex relationships available in archival descriptions, particularly focusing on temporal and contextual dimensions. Integrating Frame Semantics, partially following FRED's approach, could help create more dynamic templates while retaining flexibility.
- The integration of vocabularies for role attribution and event classification, potentially drawing from existing standards and practices.
- The exploration of hybrid approaches that combine LLMs capabilities to restrict outputs and evaluate automatically the distance between the output and the source text [11].

[24] The translation from Italian to English has been provided by the authors to enhance accessibility for readers. However, all analyses were conducted on the original Italian text, here specifically: "Nel 1874, in Italia iniziò a delinearsi un complotto insurrezionale, concentrato principalmente a Bologna con collegamenti nella regione Romagna. L'organizzatore principale era Andrea Costa, che lavorava insieme ad Antonio Cornacchia, comandante degli internazionalisti romagnoli. Il gruppo degli internazionalisti italiani fungeva sia da organizzatore che da partecipante, e si prevedeva che Bakunin si unisse come partecipante e leader nel tentativo di insurrezione".

- Research into ways to preserve and represent the narrative aspects of archival descriptions while maintaining structured data formats.
- Adding to the output graph additional data points such as sentiment.

The JSON outputs were then converted through a simple rule-based script into Turtle syntax, using RiC-O. Here is (a part of) the Turtle-format version of the same example:

```
@prefix ex: <http://example.org/#> .
@prefix rico: <https://www.ica.org/standards/RiC/ontology#>
.

ex:activity_failed_insurrection_attempt_by_italian_interna
tionalists_9
      a rico:Activity ;
      rico:hasActivityType
ex:activity_type_attempted_insurrection ;
      rico:name "Failed insurrection attempt by Italian
internationalists" ;
      rico:relationHasTarget ex:andrea_costa .
ex:andrea_costa
      a rico:Person ;
      rico:name "Andrea Costa" .
ex:andrea_costa_failed_insurrection_attempt_by_italian_int
ernationalists_9
      a     rico:PerformanceRelation ;
      rico:description     "principal     organizer"     ;
      rico:relationHasSource ex:andrea_costa
```

Listing 5. Part of the RDF version of the JSON in Listing 4.

At this point, it must be noted that the information extracted with this approach may not be directly reintegrated into the source knowledge base, due to its ontological limitations (step no. 6). Considering structured data extracted from Andrea Costa's biographical note, for example, within the model proposed by the SAN LOD ontology[25], it emerges that classes and properties capable to adequately representing the extracted information are not available. Therefore, since the expressiveness of the source model could become a further obstacle to the explication of latent contexts, it is more appropriate to opt for an approach that abstracts from a specific framework. The results of the KE should be represented as a stand-off graph adopting suitable extensions. For instance, the biographical note – or other descriptive fields of the archival record set – may become the object of an assertion that establishes the link between the source knowledge base and the graph resulting from the KE. The Web Annotation Ontology[26] (Figure 8) could be a suitable candidate.

[25] http://dati.san.beniculturali.it/lode/aggiornato.htm#d4e2193

[26] https://www.w3.org/ns/oa

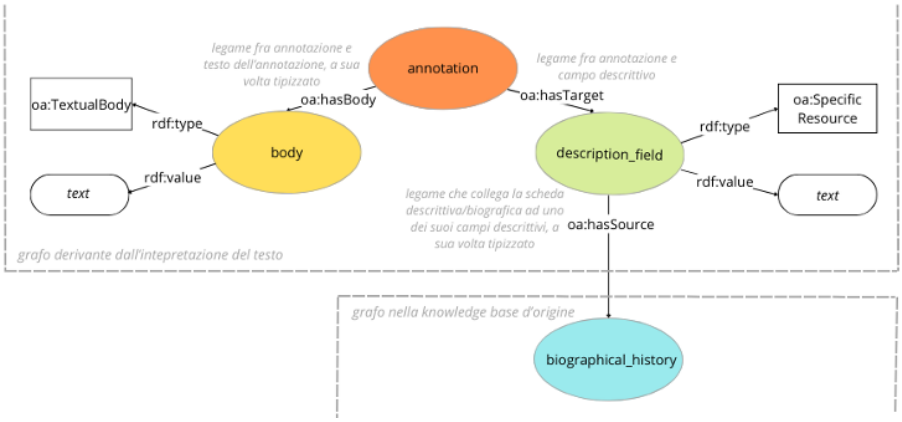

Figure 8. Hypothesis of representation of the link between the knowledge base of origin and the graph resulting from the interpretation of the text.

The Web Annotation Ontology enables the documentation of the provenance of interpretive acts and their inherent multiplicity – both in terms of tools and domain experts – thereby expanding the potential for explicit representation of contexts underpinning newly generated graphs (step no. 6) [8] [9].

However, the ontology should be adequately implemented to convey the extracted information, tailored to specific representation requirements. Furthermore, appropriate visualisation methods must be identified to present the information effectively. These methods should enable users to gain new insights by accessing the contexts that emerge from the KE processes (step no. 7).

Further work is also needed to improve the technical infrastructure supporting these systems. While open-source LLMs like Llama are promising, making these tools accessible to archivists and domain experts requires developing user-friendly interfaces and annotation tools. This democratisation of access is crucial for the widespread adoption and validation of automated extraction approaches in archival practice (step no. 8).

6. Conclusion

This contribution aims to highlight benefits deriving from the application of automated KE tools to textual metadata of archival records. The objective is not to evaluate the performance of individual tools but rather to propose a methodological approach for the automated extraction of structured information from archival descriptions. While we propose the application of certain tools, this is solely to demonstrate the approach's feasibility.

Nevertheless, from a wider perspective, a thorough comparative analysis of tools' effectiveness in terms of both the quantity and quality of extracted information will be indispensable. In this context, the LLM experimentation should be considered a complementary and documentation-enhancing function to support the human critical interpretation of data. Besides, the proposed approach to LLM-based KE applied to archival descriptions reveals both promising capabilities and important areas for development. These findings point to a crucial next step in the evolution of computational archival practices, such as the need to develop a comprehensive ecosystem of

KE tools and practices around LLM capabilities. This ecosystem must address several key requirements through the development of robust infrastructure. Critical components include user-friendly interfaces for archivists to review and correct LLM extractions, standardized metrics for assessing extraction quality, and automated workflows for converting extracted knowledge into linked data. These tools should seamlessly integrate with existing archival management systems while supporting collaborative annotation and validation workflows.

To ensure widespread adoption and sustainability, we advocate for solutions that prioritize accessibility and efficiency. This means focusing on open-source LLMs maintained and deployed by institutions with varying levels of technical resources. The development of modular, reusable components and optimization for limited computational resources will be crucial for smaller institutions and research centres. The ecosystem we envision must be flexible enough to accommodate different institutional needs while maintaining high standards of accuracy and reliability in KE.

7. Acknowledgments

Research partially funded by the European Union - Next Generation EU, investment I.4.1 PNRR Patrimonio Culturale, Decreto Ministeriale n. 351 del 9 aprile 2022.

Bibliography

[2] Palmero Aprosio Alessio, and Giovanni Moretti. "Tint 2.0: An All-Inclusive Suite for NLP in Italian." In *Proceedings of the Fifth Italian Conference on Computational Linguistics CLiC-It 2018*, edited by Elena Cabrio, Alessandro Mazzei, and Fabio Tamburini, 311–17. Torino: Accademia University Press, 2019.

[3] Babaei Giglou, Hamed, Jennifer D'Souza, and Sören Auer. "LLMs4OL: Large Language Models for Ontology Learning." In T*he Semantic Web – ISWC 2023*, edited by Terry R. Payne et al., 408–27. Cham: Springer Nature Switzerland, 2023.

[4] Bonora, Paolo, and Angelo Pompilio. "Automatic Extraction of Opera Character Characteristics through Lexical-Syntactic Patterns." *Umanistica Digitale* 5, no. 10 (January 2021): 193–210.

[5] Borgo, Stefano, Roberta Ferrario, Aldo Gangemi, Nicola Guarino, Claudio Masolo, Daniele Porello, Emilio M. Sanfilippo, and Laure Vieu. "DOLCE: A Descriptive Ontology for Linguistic and Cognitive Engineering." Special issue "Foundational Ontologies in Action," edited by Stefano Borgo, Antony Galton, and Oliver Kutz. Applied Ontology 17, no. 1 (March 2022): 45-69. https://doi.org/10.3233/AO-210259.

[6] Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack

Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. "Language Models Are Few-Shot Learners." In Advances in Neural Information Processing Systems 33, edited by H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, 1877-1901. Red Hook, NY: Curran Associates.

[7] Colavizza, Giovanni, Tobias Blanke, Charles Jeurgens, and Julia Noordegraaf. "Archives and AI: An Overview of Current Debates and Future Perspectives." *Journal on Computing and Cultural Heritage* 15, no. 1 (December 14, 2021): 4:1–4:15.

[8] Damiani, Concetta. "Archival Description and Conceptual Transversality." *JLIS.It* 13, no. 3 (September 15, 2022): 154–61.

[9] Daquino, Marilena, and Francesca Tomasi. "Historical Context Ontology (HiCO): A Conceptual Model for Describing Context Information of Cultural Heritage Objects." In *Metadata and Semantics Research*, edited by Emmanouel Garoufallou et al., 424–36. Cham: Springer International Publishing, 2015.

[10] Daquino, Marilena, Valentina Pasqual, and Francesca Tomasi. "Knowledge Representation of Digital Hermeneutics of Archival and Literary Sources." *JLIS.It* 11, no. 3 (September 15, 2020): 59–76.

[11] Daquino, Marilena. "Linked Open Data Native Cataloguing and Archival Description." *JLIS.It* 12, no. 3 (September 15, 2021): 91–104.

[12] Gangemi, A., Graciotti, A., Meloni, A., Marzi, E., Nuzzolese, A., Presutti, V., Recupero, D.R., Russo, A., & Tripodi, R. MusicBO, an application of Text2AMR2FRED to the Musical Heritage domain.

[13] Gangemi, Aldo, Valentina Presutti, Diego Reforgiato Recupero, Andrea Giovanni Nuzzolese, Francesco Draicchio, and Misael Mongiovì. "Semantic Web Machine Reading with FRED." *Semantic Web* 8, no. 6 (August 7, 2017): 873–93.

[14] Giagnolini, Lucia, Bonora, Paolo and Francesca Tomasi, "Affinare il contesto: estrazione di informazioni strutturate per l'arricchimento dei contesti archivistici", In *Me.Te. Digitali. Mediterraneo in rete tra testi e contesti*, Venezia, Associazione per l'Informatica Umanistica e la Cultura Digitale, 2024, pp. 411 – 416

[15] Guerrini, Mauro, and Tiziana Possemato. "Linked Data: Un Nuovo Alfabeto del Web Semantico." *Biblioteche Oggi* 30, no. 3 (2012): 7–15.

[16] Mihindukulasooriya, Nandana, Sanju Tiwari, Carlos F. Enguix, and Kusum Lata. "Text2KGBench: A Benchmark for Ontology-Driven Knowledge Graph Generation from Text." In *The Semantic Web – ISWC 2023*, edited by Terry R. Payne et al., 247–65. Cham: Springer Nature Switzerland, 2023.

[17] Polley, Katherine Louise, Vivian Teresa Tompkins, Brendan John Honick, and Jian Qin. "Named Entity Disambiguation for Archival Collections: Metadata, Wikidata, and Linked Data" *Proceedings of the Association for Information Science and Technology* 58, no. 1 (2021): 520–24.

[18] Shahriar, Sakib, Brady D. Lund, Nishith Reddy Mannuru, Muhammad Arbab Arshad, Kadhim Hayawi, Ravi Varma Kumar Bevara, Aashrith Mannuru, and Laiba Batool. 2024. "Putting GPT-4o to the Sword: A Comprehensive Evaluation of Language,

Vision, Speech, and Multimodal Proficiency" Applied Sciences 14, no. 17: 7782. https://doi.org/10.3390/app14177782

[19] Strötgen, Jannik, and Michael Gertz. "A Baseline Temporal Tagger for All Languages." In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 541–47. Lisbon, Portugal: Association for Computational Linguistics, 2015.

[20] Tomasi, Francesca. "Archival Finding Aids in Linked Open Data between Description and Interpretation." *JLIS.It* 14, no. 3 (September 15, 2023): 134–46.

[21] Valacchi, Federico. "The Parts and the Whole. Integrate Knowledge." *JLIS.It* 13, no. 3 (September 15, 2022): 1–11.

[22] Valacchi, Federico. "Not the Institutions but the Subjects Matter. Beyond the Necessary Approximation of Finding Aids?" *JLIS.It* 14, no. 3 (September 15, 2023): 1–14.

[23] Vitali, Stefano. "La Descrizione Degli Archivi Nell'Epoca Degli Standard e Dei Sistemi Informatici." In *Archivistica. Teorie, Metodi, Pratiche*, edited by Linda Giuva and Maria Guercio, 179–210. Roma: Carocci, 2014.

[24] Chen, Ruirui, Chengwei Qin, Weifeng Jiang, and Dongkyu Choi. 2024. "Is a Large Language Model a Good Annotator for Event Extraction?". Proceedings of the AAAI Conference on Artificial Intelligence 38 (16):17772-80. https://doi.org/10.1609/aaai.v38i16.29730.

[25] Shiri, Fatemeh, Van Nguyen, Farhad Moghimifar, John Yoo, Gholamreza Haffari, and Yuan-Fang Li. 2024. "Decompose, Enrich, and Extract! Schema-aware Event Extraction using LLMs." arXiv preprint arXiv:2406.01045. https://arxiv.org/abs/2406.01045.

[26] Waltl, B., Bonczek, G., & Matthes, F. (2018). Rule-based information extraction: Advantages, limitations, and perspectives. Jusletter IT (02 2018), 4.

[27] Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." In Advances in Neural Information Processing Systems 35, edited by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, 24824-24837. Red Hook, NY: Curran Associates.

[28] Wei, Jason, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. 2022. "Emergent Abilities of Large Language Models" Transactions on Machine Learning Research.