

# A Multi-Modal Knowledge Graph for Mapping Narratives of Cinema's Divas

Giorgio Corona, Dario Guidotti and Laura Pandolfo\*

DUMAS, Università degli Studi di Sassari, via Roma 151, Sassari, 07100, Italy

## Abstract

Autobiographical writings by Italian cinema's divas offer rich insights into personal experiences, professional careers, and cultural contexts. Despite their value, these writings remain underexplored in academic research. This paper addresses this gap by presenting a multi-modal knowledge graph designed to analyse and organise the autobiographical works of Italian cinema actresses. The knowledge graph integrates multiple data sources, including textual narratives, biographical details, filmographies, and visual materials, creating a comprehensive framework that allows researchers to explore the personal and professional networks of these actresses. In this paper, we describe the development process of the knowledge graph, from defining the key requirements and competency questions to the conceptualization and evaluation stages.

## Keywords

Multi-Modal Knowledge Graph, Semantic Web, Linked Data, Digital Humanities

## 1. Introduction

Personal narratives, particularly autobiographies, offer deep insight into historical, social and cultural contexts, as well as the intricate relationships and events that shape an individual's life. Traditionally, autobiographies have been studied through close reading [1], a manual method that provides in-depth insights into personal narratives but lacks computational or automated analytical capabilities. In contrast, distant reading – also referred to as macro-analysis – employs computational tools to analyse large volumes of text, identifying patterns, trends, and relationships across many works without needing to engage with each text in detail [2]. While distant reading allows for the analysis of large sets of texts, it may overlook deeper contextual or semantic understanding. Recently, there has been a growing interest in publishing autobiographical data as Linked Open Data [3] and employing Knowledge Graphs (KGs)[4]. This approach not only facilitates computational analyses typical of distant reading, but also leverages Semantic Web tools and technologies, enabling a wider range of analytical and semantic methods [5, 6]. By structuring autobiographical data as Linked Data, researchers can uncover patterns, relationships, and insights that might be missed through close reading alone, thus enriching the understanding of the text and expanding data-driven exploration opportunities. However, converting autobiographical information into Linked Data presents significant challenges due to the unique and complex nature of such data, necessitating careful handling and representation.

In the case of Italian cinema divas, these autobiographies – often referred to as *divagrafie* – constitute a unique literary genre that blends personal experiences with the cultivation of a carefully curated public persona. *Divagrafie* go beyond simple life stories by engaging with both the private and professional spheres of the authors' lives, merging intimate reflections with the public image they have built. This, combined with a complex network of dates, names, and places, creates valuable and detailed data that help scholars understand both the individual and the larger cultural context. Often, the study of these autobiographical works is complemented by the analysis of other materials, such as photographs, audiovisual resources, and archival documents. To effectively represent the complexity and diversity of

---

IRCDL 2025: 21st Conference on Information and Research Science Connecting to Digital and Library Science, February 20-21 2025, Udine, Italy

\*Corresponding author.

✉ gcorona1@uniss.it (G. Corona); dguidotti@uniss.it (D. Guidotti); lpandolfo@uniss.it (L. Pandolfo)

🆔 0009-0008-6039-4933 (G. Corona); 0000-0001-8284-5266 (D. Guidotti); 0000-0002-5785-5638 (L. Pandolfo)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

autobiographical and visual data, a conceptual model that integrates both textual and visual elements is crucial for making sense of this data. Multi-Modal Knowledge Graph (MMKG) [7] provides an efficient solution, offering a unified framework that enables the analysis and extraction of insights across multiple data sources. This approach is particularly valuable within the *Women Writing around the Camera* (WOW) project [8, 9], which aims to create a semantic portal dedicated to the autobiographies of Italian actresses who have achieved a significant level of fame since the dawn of Italian cinema up to the present day. In this context, the multi-modal KG – which we will refer to as the WOW KG – serves as the central knowledge base, mapping the complex personal and professional networks associated with iconic figures such as Sophia Loren, Monica Vitti, and Franca Valeri. Through the integration of autobiographical texts and visual data, our work aims to uncover patterns and connections that deepen our understanding of these actresses’ lives and careers. The complexity of building a KG covering this domain led to the following Research Questions (RQs):

- RQ1: Which key areas of knowledge are necessary to represent the autobiographical, professional, and visual aspects of the lives of Italian cinema divas?
- RQ2: How should the knowledge represented in the KG be organised to ensure adaptability and reusability for diverse applications and research needs?
- RQ3: What are the most effective strategies for building, maintaining, and expanding a multimodal KG that integrates textual, visual, and external knowledge sources?

This contribution is organised as follows. Section 2 briefly presents relevant examples of datasets and structured knowledge in the cultural heritage domain concerning literary texts and biographical materials. Then, Section 3 presents an overview of the building process of WOW KG. Section 4 discusses the requirements and competency questions underlying the KG. Section 5 define the implementation details of the KG, by describing source datasets, data model and the population of the dataset. Section 7 presents a preliminary evaluation of the KG, focusing on how queries can be used to answer the Competency Questions (CQs). Finally, an analysis of the data quality and fairness is given in Section 6, while conclusion is provided in Section 8.

## 2. Related Work

The digitisation and semantic enrichment of corpora have significantly transformed the accessibility and analysis of biographical and literary data. Several initiatives have developed datasets to facilitate cataloguing and study of such data in a variety of research contexts. Some examples include the following:

- **BiographyNet** [10] is a project that enables semantic analysis of biographical dictionaries from the 18th century to today. Its main goals are to facilitate prosopographical analysis of social groups, examine biography topics, and provide insights into historiographical approaches across countries and languages.
- **BiographySampo** [11] is a semantic portal based on Finland’s National Biography Collection, providing an innovative platform for digital humanities research. The platform’s main strengths lie in its tools for data visualisation, which allow researchers to conduct in-depth analyses of biographical data and uncover new knowledge. BiographySampo thus offers valuable resources for humanities scholars interested in understanding the relationships and historical context of Finnish figures, particularly in the fields of history and sociology.
- **The World Literature Knowledge Graph** [12] features over 194,000 writers and 971,000 works, enabling exploration of global literature and authors. The graph integrates biographical and bibliographical information with data on the reception of literary works, allowing users to explore global literature and trace the histories of authors and their writings. By creating a unified

semantic model, it facilitates the analysis of literature from a global perspective, offering insights into the interconnections between writers, works, and reader communities across various time periods and cultures.

- **WarSampo** [13] represents a significant application of knowledge graphs in historical research. By integrating diverse sources like Finnish national archives, war diaries, veterans' magazines, photographs, and prisoner records, it offers a comprehensive view of World War II, facilitating both large-scale and micro-historical research using Semantic Web technologies.
- **arkivo** [14, 15, 16, 17, 18] catalogues the extensive collection of archives, manuscripts, photographs, and artefacts held by the Józef Piłsudski Institute of America. By linking these resources to Linked Open Data databases such as DBpedia [19] and Wikidata [20], the arkivo dataset enriches the available information, facilitating further research into Polish history.
- **Memorata Poetis** [21] focuses on preserving and exploring ancient poetic traditions from Greek, Latin, Italian, and Arabic literature. Its database incorporates semantic web features and is enriched with Linked Open Data resources, such as Pleiades for geographic data and DBpedia for various entities, providing enhanced tools for textual analysis.
- **Literary Theme Ontology** [22] proposes an ambitious attempt to create a comprehensive ontology for thematic analysis in digital literary studies, starting within the initially narrower framework of the science fiction genre and with a strong focus on the Star Trek saga.

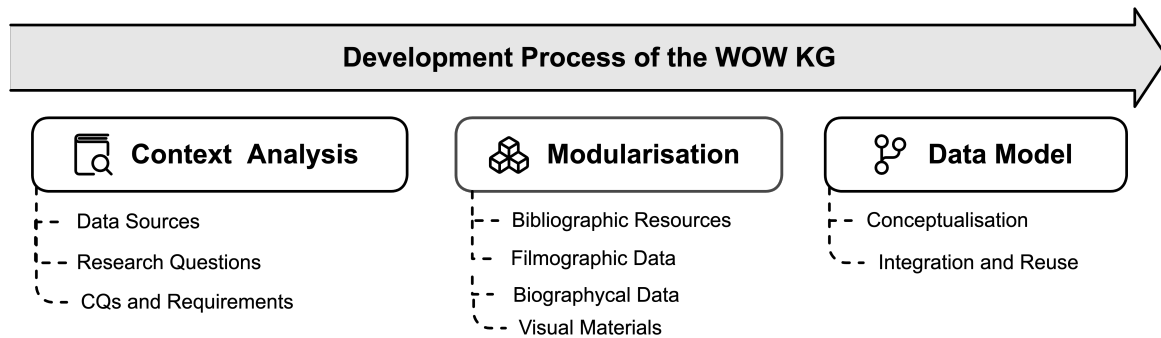
By combining different complex datasets, these initiatives offer valuable data for exploring biographical, historical and literary insights.

### 3. Methodological Approach to Building the KG

This section briefly describes the methodology adopted to construct the WOW knowledge graph. As illustrated in Figure 1, the development process consists of the following three stages:

1. **Contextual Analysis.** The first stage involves a comprehensive review of the available resources related to the autobiographies of Italian actresses. This analysis helps to identify data sources and the main entities that will form the foundation of the knowledge graph. The outcome of this phase is the formulation of the RQs presented in Section 1, which guide the investigation into how autobiographical and visual data can be effectively managed and integrated within the KG. Additionally, this stage focuses on identifying the specific requirements that the KG must satisfy. These requirements are then broken down into CQs, which help to refine the scope and ensure that it fully captures the personal, professional, and visual aspects of the actresses' lives.
2. **Modularisation.** The second stage focuses on splitting the knowledge into distinct modules. The data concerning various aspects of the actresses' lives, such as biographical info, filmography, and visual identity materials like photographs, is segmented into different "knowledge modules." This modular approach enables easier management of the knowledge, allowing for independent development and updates to each module while maintaining overall coherence across the system.
3. **Data Model.** The final stage involves the design of the data model, which serves as the conceptual framework for integrating and managing the knowledge. We follow a combination of multiple methodologies, i.e., Methontology [23] and MOMo [24]. Additionally, significant effort was dedicated to guaranteeing interoperability and compatibility with existing ontological schemas, enabling the knowledge graph to integrate seamlessly with broader semantic networks and established standards. The model is developed based on the requirements and competency questions identified in the previous stages, capturing the full complexity of the actresses' lives and their visual representation. This phase follows a structured methodology, ensuring that the

model is robust, flexible, and capable of integrating both textual and visual data in a coherent manner.



**Figure 1:** Development process of the knowledge graph.

In summary, this methodological approach provides a systematic framework for constructing the KG. By modularising the knowledge according to specific context analysis requirements, it facilitates the management and continuous updating of different aspects of the domain of interest.

## 4. Requirements and Competency Questions

The development of the knowledge graph is driven by a set of core requirements and competency questions that directly stem from the research questions introduced in Section 1. This section outlines these high-level requirements and identifies a set of CQs that will guide the design and implementation of the KG. The identified requirements for the KG include the following:

- Req1 The KG must provide a detailed representation of the bibliographic and thematic aspects of the divas' autobiographies, enabling to explore themes, keywords, and textual relationships.
- Req2 The KG must provide a comprehensive representation of the divas' professional lives, including their roles in films, collaborations with industry professionals, and temporal aspects of their careers.
- Req3 The KG must capture and model the personal lives of the divas, including their family relationships, romantic partners, and key life events.
- Req4 The KG must support the integration of multiple data sources, including biographical, filmographic, and visual data, such as photographs.
- Req5 The KG must be flexible and extensible, enabling future updates as new autobiographies, films, or photographs are added, and as the scope of the research expands.
- Req6 The KG must be compatible with existing ontological schemas, ensuring the ability to link the KG with broader digital humanities initiatives and datasets.

To further refine the scope and ensure the knowledge graph meets its goals, a set of competency questions were developed. These questions help define the types of queries that the graph must support and guide the structuring of the data model. Below are the CQs related to the first four requirements (Req1-Req4), which are a subset of the total set of competency questions. Req5 and Req6 focus more on the maintenance and integration of the graph rather than on querying and searching data, and therefore, no competency questions have been developed for these requirements.

**Table 1**

Competency questions and their relation to the requirements.

Requirement	Competency Questions
<b>Req1</b>	<b>CQ1:</b> What are the main themes and keywords discussed in a specific diva’s autobiography? <b>CQ2:</b> Which quotes or textual blocks in an autobiography address a specific theme or keyword?
<b>Req2</b>	<b>CQ3:</b> Which divas collaborated most frequently with specific directors, producers, or other industry professionals? <b>CQ4:</b> What genre of film did divas most often play in over the course of their careers, and how is this reflected in their autobiographies? <b>CQ5:</b> What roles (e.g., actress, director) did a diva have in a specific film?
<b>Req3</b>	<b>CQ6:</b> How are the divas’ family relationships, romantic partnerships, and life events represented in their autobiographies? <b>CQ7:</b> What biographical data is available for a specific diva (e.g., age, debut year, place of birth)?
<b>Req4</b>	<b>CQ8:</b> How do photographs and other visual data are associated with a diva? <b>CQ9:</b> What filmographic information is available for a specific film?

## 5. The WOW Knowledge Graph

The WOW knowledge graph captures the personal and professional relationships of Italian cinema divas from their autobiographies. The following sections detail its development, starting with the source datasets. We then describe the data model, which structures information to represent complex relationships. Finally, we explain the workflow for populating the model, detailing how raw data was transformed into RDF, linked to external datasets, and incorporated into the KG.

### 5.1. Source Datasets

We have used data from four primary sources: the divas’ autobiographies dataset, DBpedia, Wikidata, and photographs from specific archives. The divas’ autobiographies dataset has been developed from the annotations made by researchers and professors in the domain, which catalogued 102 autobiographies written by 59 actresses, identifying 425 quoted passages and 19 scholarly themes centred around 232 keywords. Additionally, 301 individual names mentioned in the texts were annotated, along with their occurrences. This data was provided in CSV format. From DBpedia and Wikidata, we obtained supplementary information on actress-writers and the individuals they mention, as well as details about the films and show in which these actresses participated, including 2,005 film and show entities. Additionally, we collected information on the professionals involved in these works, as well as the actresses’ relatives and romantic partners, adding 10,210 names to the dataset. Moreover, from Wikidata, we incorporated user IDs for the divas across major social networks.

In addition to these textual and filmographic sources, the KG also integrates visual materials, which are the subject of a specific research effort aimed at retrieving photographs from various archives. This includes notable collections, such as those stored in the Luisa Gaetano Archive, the Elisabetta Catalano Archive, and the J. Vodoz Archive. These photographs are annotated by experts, who not only catalogue the images but also establish connections between the photographs, the autobiographical texts, and other relevant materials. This expert annotation process ensures that the visual elements are meaningfully related to the thematic and narrative content, enhancing the semantic depth of the KG.

## 5.2. Data Model

The annotations were converted into RDF triples and integrated with other data sources into a cohesive data model <sup>1</sup>. Concepts from the source datasets are described using metadata schemas [25], such as DCMI Metadata Terms, and vocabulary models like SKOS and RDF Schema. Key entities and properties were extensively reused, applying established reuse strategies [26] to promote interoperability and reusability. Table 2 lists the imported ontologies and their namespace prefixes, while Figure 2 illustrates the schema of classes, subclasses, and main properties that comprise the data model.

**Table 2**

Details on the re-used ontologies in the data model. It includes the ontology name, the corresponding prefix used, and the expansion of each prefix into its full *Internationalized Resource Identifier* (IRI).

Ontology	Prefix Name	Expansion
arkivo	arkivo:	<a href="http://purl.org/arkivo/ontology#">http://purl.org/arkivo/ontology#</a>
DBpedia	dbo:	<a href="http://dbpedia.org/ontology/">http://dbpedia.org/ontology/</a>
	dbp:	<a href="http://dbpedia.org/property/">http://dbpedia.org/property/</a>
DCMI	dcterms:	<a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a>
DoCo	doco:	<a href="http://purl.org/spar/doco/">http://purl.org/spar/doco/</a>
FOAF	foaf:	<a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a>
Pattern	po:	<a href="http://www.essepuntato.it/2008/12/pattern#">http://www.essepuntato.it/2008/12/pattern#</a>
RDF Schema	rdfs:	<a href="http://www.w3.org/2000/01/rdf-schema#">http://www.w3.org/2000/01/rdf-schema#</a>
Schema	schema:	<a href="https://schema.org/">https://schema.org/</a>
SKOS	skos:	<a href="http://www.w3.org/2004/02/skos/core#">http://www.w3.org/2004/02/skos/core#</a>
Wikidata	wd:	<a href="https://www.wikidata.org/wiki/">https://www.wikidata.org/wiki/</a>
	wdt:	<a href="https://www.wikidata.org/wiki/Property:">https://www.wikidata.org/wiki/Property:</a>

For clarity, our data model can be categorised into four main areas. The first area, bibliographic and thematic, focuses on the autobiographies as literary works, incorporating bibliographic details and thematic content. The second area addresses the professional aspects of the actresses' lives, detailing their film roles and career-related information. The third area relates the personal aspects, including information about the actresses' family and sentimental lives. The fourth area focuses on the description of photographic resources, capturing visual materials and integrating them with the actresses' autobiographies and other related data. Each of these areas is described in detail below.

### 5.2.1. Bibliographic Data and Themes

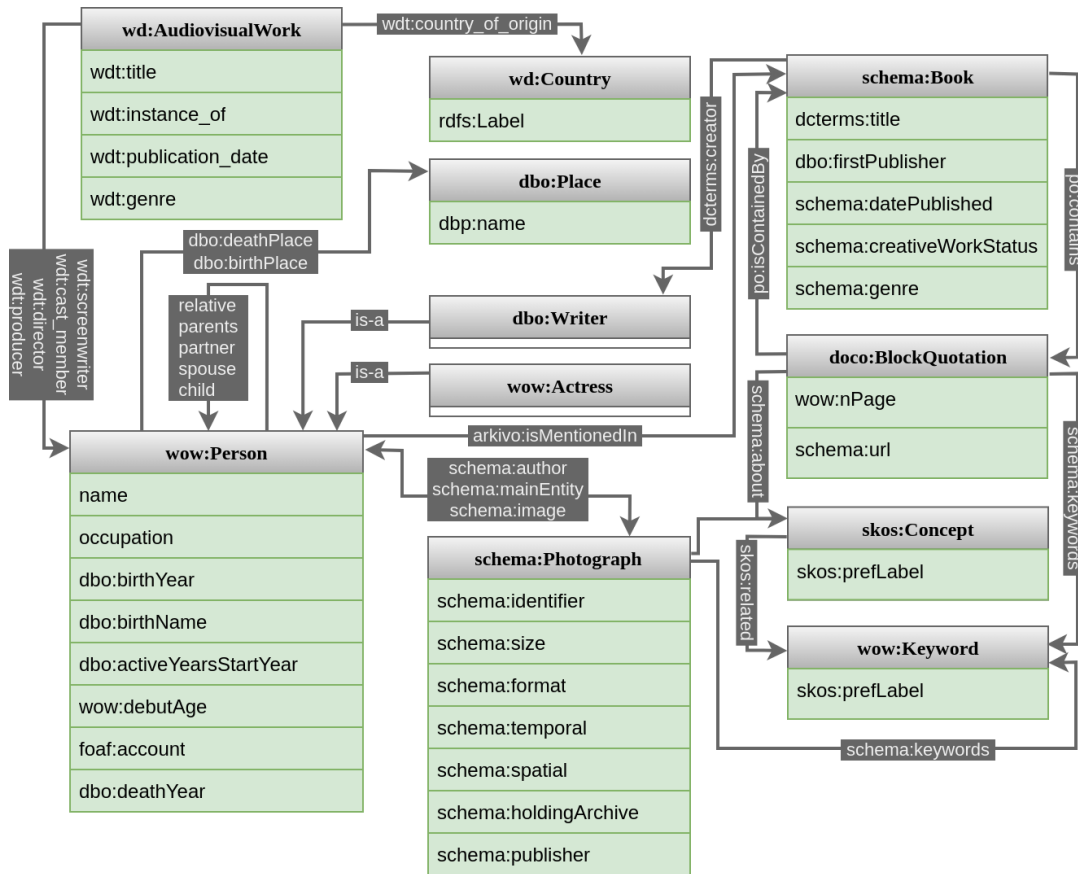
The bibliographic data and the classification system by subject and keyword form a key part of our dataset. Autobiographies are catalogued on the basis of their first edition, with parameters such as title, publisher, year of publication, and genre (including several subcategories of the autobiography genre). These data are supplemented by information on the publication status (whether still on the market or not), represented by the `schema:creativeWorkStatus` property.

Excerpts from autobiographies, selected and analysed by domain experts, are indexed by page number (`wow:nPage`), stored into the `doco:BlockQuotation` class (imported from the DoCO [27] ontology), and linked to the texts from which they were extracted through `po:contains` and `po:isContainedBy` properties from the Pattern ontology<sup>2</sup>. The content of each quoted passage is classified by one or more pairs of terms formed by a theme (`skos:Concept`) and a keyword, e.g. *Melancholy* (theme) and

<sup>1</sup>The data model is available at this link: [https://github.com/AIMet-Lab/PRIN-WOW/blob/main/wow\\_schema.ttl](https://github.com/AIMet-Lab/PRIN-WOW/blob/main/wow_schema.ttl). To produce a human-readable form of the schema, the pyLODE documentation can be found here: <https://aimet-lab.github.io/PRIN-WOW/>

<sup>2</sup><https://sparontologies.github.io/po/current/po.html>





**Figure 2:** Schema of the data model, illustrating the classes, main properties and relationships. Where the prefix is not specified, the corresponding properties were imported from both Wikidata and DBpedia.

*Loneliness* (keyword), or *Body* and *Mirror*. Usually each keyword belongs to only one theme, but rarely a keyword can belong to more than one theme: *Desire and attraction* can refer to both *Relationship with men* and *Relationship with women* themes (as well as *Disappointment*). The above mentioned non-hierarchical semantic relationship between themes and keywords was implemented through the use of the `skos:related` property.

Names referenced in autobiographies are linked to the texts in which they are mentioned through the `arkivo:isMentionedIn` property, which has been imported from the `arkivo` ontology. It is worth noting the implementation of two further annotation properties not present in the above schema's depiction. These are: `wow:nTimes`, which enumerates the number of mentions of names in autobiographies; and `wow:artistAge`, which notes the age of the divas at the time of the publication of their autobiographies.

### 5.2.2. Filmographic Data and Divas' Professional Lives

To gain insight into the professional lives of the actresses and delineate a timeline, we have started collecting information on their year of debut (`dbo:activeYearsStartYear`) and the age of debut (`wow:debutAge`). We then gathered their filmographies from Wikidata: these included not only performances as actresses (`wdt:P161`) but also roles as directors, producers, or screenwriters. After annotating these properties with the personal age of the divas at the time using the `wow:artistAge` annotation property, we provided each film and entity with information about genre and release year – using the same properties as for the autobiographies – and country of origin. For each movie, we then gathered information about cast, director, and producer: the retrieved names were incorporated into the dataset, along with the names of individuals referenced in the autobiographical texts. For each name in the dataset, we then collected information about their occupations through the `dbp:occupation`

and the `wdt:P106` properties. In the intersection between public and private, social networks provide valuable data, as noted by experts. For each diva in the dataset, efforts were made to retrieve their account IDs on major platforms (Facebook, Instagram, X). A search via Wikidata produced 36 results, recorded using the `foaf:account` property, opening opportunities for future research.

### 5.2.3. Divas' Private Lives

In regard to the personal histories of the writing divas, domain experts provided data on their dates of birth and death, which we supplemented with the corresponding geographical information (`dbo:birthPlace` and `dbo:deathPlace`), as well as with birth names. A further attempt was made to reconstruct the private lives of the divas by collecting data about parents, children, relatives, partners and husbands: 105 entities with different degrees of relatedness have been documented in this process. For each name collected, as before for names mentioned in texts or collected from filmographic research, professions were then recorded. This kind of research presents a significant challenge for domain experts, who are therefore highly interested in the results.

### 5.2.4. Divas' Visual Materials

A part of the WOW project is the retrieval from public and private archives of photographic documentation relating to the lives of the divas. The photographs share with the autobiographies the rich selection of themes and keywords identified by the domain experts, thus enhancing the semantic synergy between heterogeneous source materials. As presented in Figure 2, the data model foresees a class, `schema:Photograph`, to collect the recovered photographic material and to catalogue it with imported properties. The `schema:Photograph` class relates to the `wow:Person` class, which catalogue the authors of the photos, the divas portrayed and other recognisable persons. Furthermore, the `schema:Photograph` class can be related to the `dbo:Place` class, which contains the cities location of the shoots, and, if shot on set, to the `wd:AudiovisualWork` class. Each photograph be given an ID and information about its size, format, place and time of capture, as well as about the archive owner of the picture and, eventually, about the publisher.

## 5.3. Populating the Data Model

The names of individuals mentioned in the texts were extracted using Named Entity Recognition (NER) and stored in CSV files. We then assigned semantic meaning to these names, including those of the writer divas, through Named Entity Linking (NEL). Regular expressions were effective for NER, and NEL was executed through specific Python scripting. This process generated the IRIs necessary for subsequent searches. These IRIs were used to perform SPARQL queries against the DBpedia endpoint<sup>3</sup> and the Wikidata endpoint<sup>4</sup>. The individuals representing movies and theatre shows are imported from Wikidata, as well as the class that contains them `wd:AudiovisualWork` (`wd:Q2431196`). The `wow:Person` class contains both entities from Wikidata and DBpedia, and is defined by the `dbo:Person` class and the `wd:Person` class (`wd:Q215627`) through the `rdfs:isDefinedBy` property.

Actresses are further catalogued, according to their profession, into the `wow:Actress` subclass of the `wow:Person` class, and writer divas into both the `wow:Actress` and the `dbo:Writer` subclasses. The `schema:Book` class, the `doco:BlockQuotation` class and the `skos:Concept` class were imported and populated with individuals representing autobiographies, text excerpts and their themes, while the `wow:Keyword` class have been implemented.

Figure 3 illustrates a representation of the data transformation pipeline and the connections established through NEL, as described above. The cylinders represent RDF data organised by class, with solid arrows indicating the population of these classes and dashed arrows signifying NEL and data integration processes. In the initial phase, annotations stored in CSV format were converted into RDF

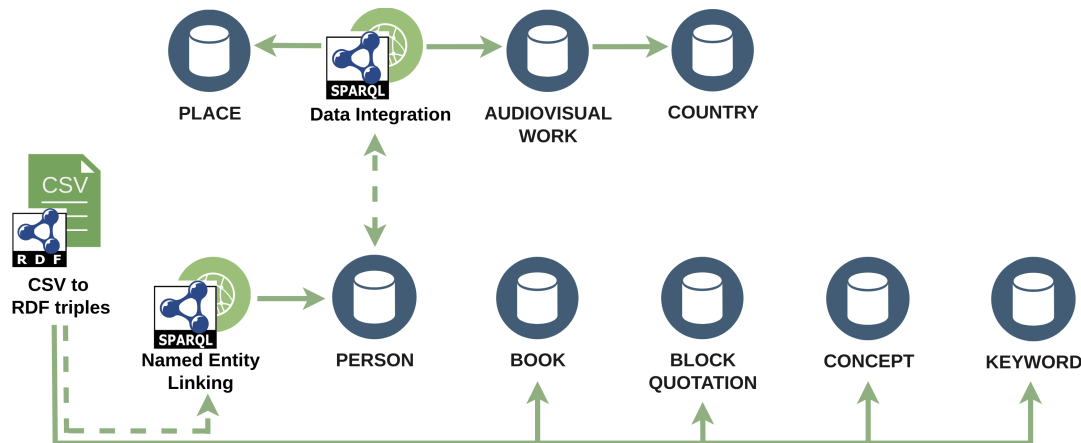
---

<sup>3</sup><https://dbpedia.org/sparql/>

<sup>4</sup><https://query.wikidata.org/sparql>



triples. Entities in the `wow:Person` class were processed using both NEL and data integration to enhance their quality and connectivity. In contrast, the classes `dbo:Place`, `wd:AudiovisualWork`, and `wd:Country` were populated with data sourced from external knowledge bases such as Wikidata and DBpedia. All other classes were populated exclusively with data derived from the original annotations. This approach combines manual annotation with automated linking and integration to construct a semantically rich and interconnected dataset.



**Figure 3:** Graphical representation of the data flow process.

The dataset currently comprises 13,199 individuals, 1,266 of which are derived from annotations by domain experts and 11,933 have been imported from external resources. Table 3 presents metrics related to the knowledge graph, while the most relevant properties and their frequency are reported in Table 4. The KG is available on Zenodo, with an associated citation [28], and is licensed with the open Creative Commons BY 4.0 license.

**Table 3**  
Metrics related to the WOW KG content.

Entity	Counter
Concepts	12
Object Properties	31
Data Properties	31
Annotations	11
Individuals	13,199
Axioms	114,169

**Table 4**  
Usage count of the main properties of the WOW KG.

Property Name	Usage
<code>wdt:occupation</code>	31,630
<code>wdt:cast_member</code>	22,689
<code>wow:artistAge</code>	1,834
<code>wdt:director</code>	1,764
<code>schema:keywords</code>	1,435
<code>schema:about</code>	1,014
<code>wdt:producer</code>	1,051

## 6. Data Quality & Fairness

The knowledge graph is based on heterogeneous data that are considered high quality, as they are ensured through multiple layers of validation and enrichment processes. First, domain experts meticulously annotate autobiographical texts, identifying key themes, quoted passages, and references, which are then transformed into RDF triples. This process leverages both manual and semi-automated techniques to enhance accuracy. In addition to the textual data, experts are also annotating photographs stored in archives, further enriching the graph with visual materials and their associated metadata. Moreover, integration with external datasets such as DBpedia and Wikidata enriches the KG by providing supplementary information and establishing robust links between entities.

In terms of fairness, the knowledge graph leverages publicly available information and datasets, including published autobiographies and data from Wikidata and DBpedia. These sources are open and

accessible, which supports transparency and reproducibility in research. The validation and enrichment processes ensure that the data is accurate and representative, while the use of open data adheres to principles of equitable access and responsible data usage.

An important part of the data enrichment process involves the collection of photographs that complement the autobiographical and professional data in the KG. We are currently gathering photographs from different archives, including many private ones. While some of these photographs can be scanned and made publicly available with the appropriate authorizations, others are restricted due to copyright or privacy concerns. For the photographs that cannot be publicly shared, we still collect and document their metadata, ensuring that this valuable information is preserved and accessible.

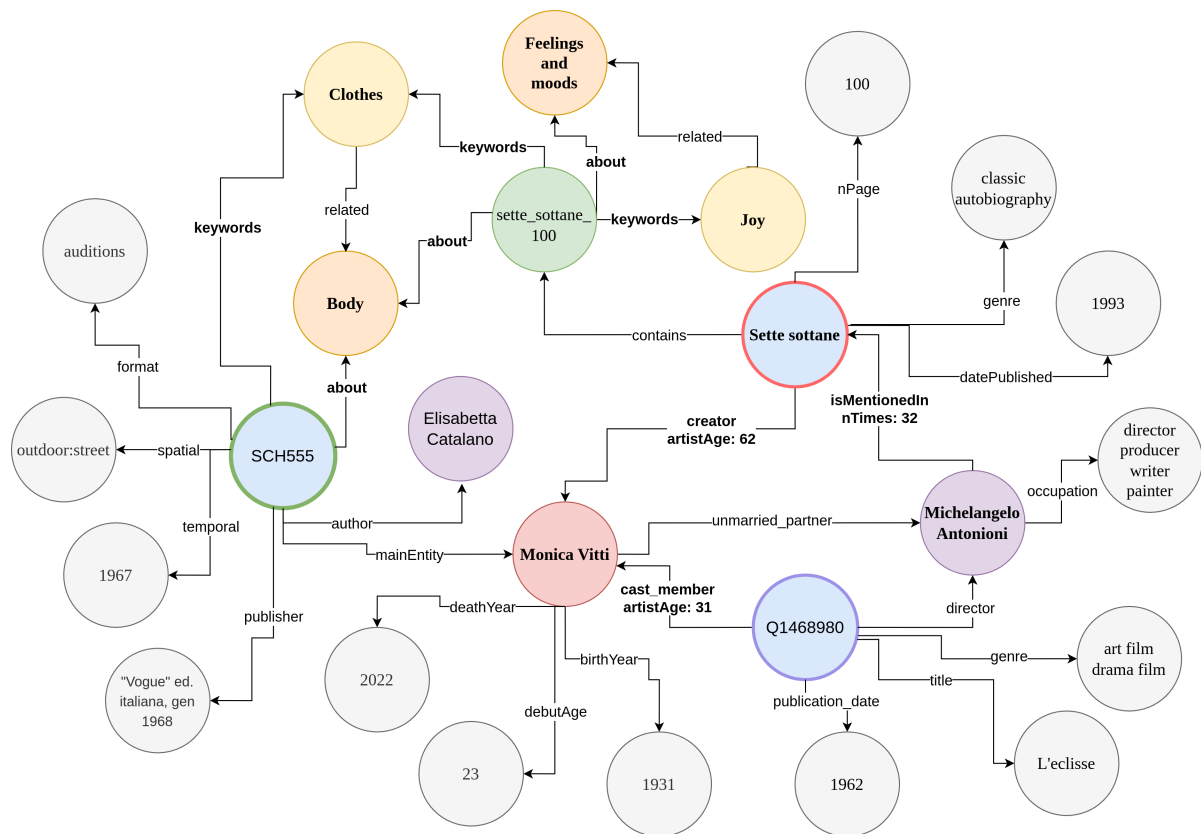
## 7. Evaluation

To evaluate the WOW KG, we adopted a structured approach aimed at verifying its ability to meet the objectives defined in the research questions, the requirements, and the competency questions. Specifically, the evaluation was conducted in three main phases: (i) determining whether the KG could provide meaningful answers to the research questions that guided its development, (ii) verifying that all stated requirements were satisfied, and (iii) demonstrating the KG's capacity to address the competency questions through targeted SPARQL queries. This process ensured that the KG adhered to its conceptual design and also delivered practical value for analysing the lives of Italian cinema divas.

**(i) To determine whether the KG could address the research questions.** To address the first research question (RQ1), we collaborated with domain experts to identify the conceptual domains necessary to comprehensively capture the multifaceted lives of Italian cinema divas. These domains, which include the autobiographical and personal, professional, and visual aspects, ensure a comprehensive representation of the divas' lives. Regarding the second research question (RQ2), the KG was carefully structured to balance specificity and flexibility. By organising the data into interrelated modules dedicated to specific domains, the KG enables diverse research applications across various contexts in the digital humanities. This modular approach ensures that the KG can support different types of analysis without compromising its coherence. In response to the third research question (RQ3), the construction and maintenance of the WOW KG required the combination of different methodologies to effectively address its multimodal nature. This approach combines strategies for modularisation and conceptual design.

**(ii) To verify whether the KG could address the requirements.** Firstly, the KG effectively represents bibliographic data, allowing users to explore autobiographies through themes, keywords, and the relationships between different texts (Req1). The comprehensive coverage of the divas' professional lives is also fully addressed, with detailed data on their roles in films, associated locations, and the timeframes of their careers, providing a complete overview of their professional trajectories (Req2). In terms of personal life, the KG offers rich insights into the divas' family relationships, romantic partnerships, and key life events, ensuring that their biographies are captured in a holistic way (Req3). Furthermore, the KG has been designed to accommodate the integration of multiple data sources, from biographical and filmographic data to visual materials such as photographs (Req4). Additionally, the KG is structured to be flexible, extensible, and modular, allowing for easy updates and the inclusion of new data such as autobiographies, films, and photographs as the research scope expands (Req5). Finally, to guarantee that the KG remains compatible with existing ontological schemas and can integrate seamlessly with broader digital humanities initiatives, it leverages best practices in aligning with major ontologies and semantic datasets (Req6).

**(iii) To demonstrate the KG's capacity to address competency questions.** In Figure 4 we report a portion of network data pertaining to the person of *Monica Vitti*. This data, retrieved through a SPARQL query about Vitti's network, provides different insights into her life and career. At the centre of the figure are key biographical details, including her birth year (1931), debut age (23 years), and death year (2022). These details address CQ7. Significant life events, such as her role in *Michelangelo Antonioni's* film *L'Eclisse* (1962), are also highlighted. Here, Vitti is explicitly identified as an actress, satisfying CQ5 and



**Figure 4:** Graphical representation of an example of data modelled within the knowledge graph. The coloured nodes correspond to individuals, where each colour reflects the class to which they belong. The gray nodes represent data properties, while the arrows denote object properties, illustrating the connections and relationships between entities.

CQ9. The graph explores Vitti’s autobiography, *Sette Sottane* (1993), where recurring themes such as *body* and *joy* are central. For instance, a specific textual block (nPage: 100) addresses the body, offering a deeper perspective on Vitti’s reflections and aligning with CQ1 and CQ2. Vitti’s professional collaborations are represented through her frequent work with Michelangelo Antonioni, who is identified both as the director of *L’Eclisse* and as Vitti’s romantic partner. This dual connection—professional and personal—addresses CQ3, and CQ6. The emotional and dramatic nature of their work together reflects the genre of drama films, a recurring theme in Vitti’s career and autobiography (CQ4). The example also incorporates visual data through the photograph *SCH555*, taken in 1967 by *Elisabetta Catalano*. This image, categorised under *body* and *clothes*, was later published in *Vogue Italia* in January 1968. Its metadata, including the spatial (outdoor:street) and temporal context, demonstrates how visual materials are associated with Vitti, effectively answering CQ8. In summary, this interconnected representation of textual, professional, and visual data offers a nuanced view of Monica Vitti’s persona. It not only captures biographical facts and artistic achievements but also contextualises her career within broader emotional and thematic narratives.

## 8. Conclusion

The development of the presented knowledge graph marks a significant advancement in representing the autobiographies of actresses within the humanities. By converting autobiographical texts into Linked Data and integrating them with external datasets, this knowledge graph offers a comprehensive, semantically rich resource for examining the lives and careers of Italian cinema divas. Furthermore, we are now incorporating visual materials, such as photographs, into the dataset. The work on integrating these photographs is still ongoing, and we are gradually adding these visual elements as they are

retrieved and annotated from various archives.

We are continually updating the knowledge graph, refining its conceptualisation as new data related to the divas' autobiographies, films, and visual materials becomes available. As the dataset evolves, we will focus on expanding its coverage and exploring new research applications. Moreover, we plan to make the entire WOW KG accessible through a SPARQL query endpoint to facilitate its use and enhance accessibility. We are also investigating the potential applications of Large Language Models for the automatic analysis of autobiographies, focusing on tasks such as sentiment analysis and keyword extraction. These tasks are critical for annotating texts, as they significantly enhance the efficiency of textual analysis, facilitate the exploration of thematic trends, and save time and effort compared to manual annotation processes.

## Acknowledgments

This work has been supported by the PRIN 2022 project "WOMen Writing around the camera (WOW)" funded by the European Union- Next Generation EU, Mission 4 Component C2, CUP: J53D23013480006.

## References

- [1] F. Moretti, *Distant Reading*, Verso Books, 2013.
- [2] K. Schulz, What is Distant Reading, *The New York Times* 24 (2011) 43–62.
- [3] E. Hyvönen, *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*, volume 3, Morgan & Claypool Publishers, 2012. URL: <https://doi.org/10.1007/978-3-031-79438-4>. doi:10.1007/978-3-031-79438-4.
- [4] C. Gutierrez, J. F. Sequeda, *Knowledge Graphs*, *Commun. ACM* 64 (2021) 96–104. URL: <https://doi.org/10.1145/3418294>. doi:10.1145/3418294.
- [5] E. Hyvönen, *Using the Semantic Web in Digital Humanities: Shift from Data Publishing to Data-analysis and Serendipitous Knowledge Discovery*, *Semantic Web* 11 (2020) 187–193. URL: <http://dx.doi.org/10.3233/SW-190386>. doi:10.3233/SW-190386.
- [6] G. Adorni, M. Maratea, L. Pandolfo, L. Pulina, An ontology-based archive for historical research, in: D. Calvanese, B. Konev (Eds.), *Proceedings of the 28th International Workshop on Description Logics*, Athens, Greece, June 7-10, 2015, volume 1350 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2015.
- [7] J. Peng, X. Hu, W. Huang, J. Yang, What is a Multi-Modal Knowledge Graph: A Survey, *Big Data Research* 32 (2023) 100380. URL: <https://www.sciencedirect.com/science/article/pii/S2214579623000138>. doi:<https://doi.org/10.1016/j.bdr.2023.100380>.
- [8] L. Pandolfo, L. Cardone, L. Cutzu, R. Perna, B. Seligardi, G. Simi, The WOW Project: Bridging AI and Cultural Heritage for Actress Writings, in: *Proceedings of the 2nd Workshop on Artificial Intelligence for Cultural Heritage (IAI4CH 2023) co-located with AIXIA 2023*, volume 3536 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023, pp. 34–41. URL: [https://ceur-ws.org/Vol-3536/04\\_paper.pdf](https://ceur-ws.org/Vol-3536/04_paper.pdf).
- [9] G. Corona, D. Guidotti, L. Pandolfo, Constructing a knowledge graph for italian cinema divas' autobiographies (short paper), in: *Proceedings of the 3rd Workshop on Artificial Intelligence for Cultural Heritage (IAI4CH 2024) co-located with the 23rd International Conference of the Italian Association for Artificial Intelligence (AIXIA 2024)*, Bolzano, Italy, November 28, 2024, volume 3865 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 22–29. URL: [https://ceur-ws.org/Vol-3865/03\\_paper.pdf](https://ceur-ws.org/Vol-3865/03_paper.pdf).
- [10] A. Fokkens, S. ter Braake, N. Ockeloen, P. Vossen, S. Legêne, G. Schreiber, V. de Boer, *BiographyNet: Extracting Relations Between People and Events*, arXiv e-prints (2018) arXiv–1801. URL: <https://doi.org/10.48550/arXiv.1801.07073>. doi:10.48550/arXiv.1801.07073.
- [11] M. Tamper, P. Leskinen, E. Hyvönen, R. Valjus, K. Keravuori, *Analyzing Biography Collections*

- Historiographically as Linked Data: Case National Biography of Finland, *Semantic Web* 14 (2023) 385–419. doi:10.3233/SW-222887.
- [12] M. A. Stranisci, E. Bernasconi, V. Patti, S. Ferilli, M. Ceriani, R. Damiano, The World Literature Knowledge Graph, in: T. R. Payne, V. Presutti, G. Qi, M. Poveda-Villalón, G. Stoilos, L. Hollink, Z. Kaoudi, G. Cheng, J. Li (Eds.), *The Semantic Web – ISWC 2023*, Springer Nature Switzerland, Cham, 2023, pp. 435–452. doi:10.1007/978-3-031-47243-5\_24.
- [13] M. Koho, E. Ikkala, P. Leskinen, M. Tamper, J. Tuominen, E. Hyvönen, Warsampo Knowledge Graph: Finland in the Second World War as Linked Open Data, *Semantic Web* 12 (2021) 265–278.
- [14] L. Pandolfo, L. Pulina, M. Zielinski, Towards an ontology for describing archival resources, in: *Proceedings of the Second Workshop on Humanities in the Semantic Web (WHiSe II) co-located with ISWC*, volume 2014 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2017, pp. 111–116. URL: <https://ceur-ws.org/Vol-2014/paper-12.pdf>.
- [15] L. Pandolfo, L. Pulina, M. Zielinski, ARKIVO: An Ontology for Describing Archival Resources, in: *Proceedings of the 33rd Italian Conference on Computational Logic*, 2018, volume 2214 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2018, pp. 112–116. URL: <https://ceur-ws.org/Vol-2214/paper12.pdf>.
- [16] L. Pandolfo, L. Pulina, Building the Semantic Layer of the Józef Piłsudski Digital Archive With an Ontology-Based Approach, *Int. J. Semantic Web Inf. Syst.* 17 (2021) 1–21. URL: <https://doi.org/10.4018/ijswis.2021100101>. doi:10.4018/IJSWIS.2021100101.
- [17] L. Pandolfo, L. Pulina, M. Zielinski, Exploring Semantic Archival Collections: The Case of Piłsudski Institute of America, in: *15th Italian Research Conference on Digital Libraries, IRCDL 2019*, volume 988 of *Communications in Computer and Information Science*, Springer, 2019, pp. 107–121. URL: [https://doi.org/10.1007/978-3-030-11226-4\\_9](https://doi.org/10.1007/978-3-030-11226-4_9). doi:10.1007/978-3-030-11226-4\_9.
- [18] L. Pandolfo, L. Pulina, ARKIVO dataset: A benchmark for ontology-based extraction tools, in: *Proceedings of the 17th International Conference on Web Information Systems and Technologies, WEBIST, SCITEPRESS*, 2021, pp. 341–345. URL: <https://doi.org/10.5220/0010677000003058>. doi:10.5220/0010677000003058.
- [19] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. Van Kleef, S. Auer, et al., DBpedia—A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia, *Semantic web* 6 (2015) 167–195. doi:10.3233/SW-140134.
- [20] D. Vrandečić, M. Krötzsch, Wikidata: a Free Collaborative Knowledgebase, *Communications of the ACM* 57 (2014) 78–85. URL: <http://dx.doi.org/10.1145/2629489>. doi:10.1145/2629489.
- [21] F. Khan, S. Arrigoni, F. Boschetti, F. Frontini, Restructuring a Taxonomy of Literary Themes and Motifs for More Efficient Querying, *MATLIT: Materialidades da Literatura* 4 (2016) 11–27. URL: [http://dx.doi.org/10.14195/2182-8830\\_4-2\\_1](http://dx.doi.org/10.14195/2182-8830_4-2_1). doi:10.14195/2182-8830\_4-2\_1.
- [22] P. Sheridan, M. Onsjö, J. Hastings, The Literary Theme Ontology for Media Annotation and Information Retrieval, *arXiv e-prints* (2019) arXiv-1905.
- [23] M. Fernández-López, A. Gómez-Pérez, N. Juristo, *Methontology: From Ontological Art Towards Ontological Engineering* (1997).
- [24] C. Shimizu, K. Hammar, P. Hitzler, Modular Ontology Modeling, *Semantic Web* 14 (2023) 459–489.
- [25] R. Gartner, *Metadata: Shaping Knowledge from Antiquity to the Semantic Web*, Cham, Springer International, 2017. doi:<https://doi.org/10.1007/978-3-319-40893-4>.
- [26] V. A. Carriero, M. Daquino, A. Gangemi, A. G. Nuzzolese, S. Peroni, V. Presutti, F. Tomasi, The Landscape of Ontology Reuse Approaches, in: *Applications and Practices in Ontology Design, Extraction, and Reasoning*, IOS Press, 2020, pp. 21–38. URL: <http://dx.doi.org/10.3233/ssw200033>. doi:10.3233/ssw200033.
- [27] A. Constantin, S. Peroni, S. Pettifer, D. M. Shotton, F. Vitali, The Document Components Ontology (DoCO), *Semantic Web* 7 (2016) 167–181. URL: <https://doi.org/10.3233/SW-150177>. doi:10.3233/SW-150177.
- [28] L. Pandolfo, G. Corona, D. Guidotti, Women Writings Around the Camera Knowledge Graph, [Dataset], 2025. doi:10.5281/zenodo.13784081.