

# The WOW Project: Bridging AI and Cultural Heritage for Actress Writings

Laura Pandolfo<sup>1,\*</sup>, Lucia Cardone<sup>1</sup>, Luisa Cutzu<sup>1</sup>, Raffaella Perna<sup>2</sup>, Beatrice Seligardi<sup>1</sup> and Giulia Simi<sup>1</sup>

<sup>1</sup>Università degli Studi di Sassari, DUMAS, via Roma 151, Sassari, 07100 Italia

<sup>2</sup>Sapienza Università di Roma, SARAS, Piazzale Aldo Moro 5, Roma, 00185, Italia

## Abstract

The Digital Humanities field has evolved significantly due to technological advancements and the availability of cultural heritage data. The project *WOMen Writing around the Camera* (WOW) leverages these advancements to create an ontology-based portal focusing on the noteworthy writings of Italian actresses, unveiling the dynamics between writing, acting, and image cultivation. This corpus represents a unique collection of cultural data for digital humanities scholars. In this paper, we present an overview of the project's methodology and its expected impacts in the research community and society.

## Keywords

Semantic Web, Computational Humanities, Digital Archive, Film and Literature Studies

## 1. Introduction

In recent decades, Digital Humanities (DH) [1] has experienced a significant transformation closely linked with rapid technological progress and the availability of cultural heritage data, giving rise to a prosperous field of research. The use of computer-based methods in humanities research has roots in the early days of DH, where pioneers employed emerging digital tools to examine and comprehend cultural objects. These initial efforts paved the way for a thriving area of investigation, as scholars employed technology to address intricate queries in literature, history, and the social sciences.

One of the main research challenges in the field of DH revolves around the analysis of extensive literary corpora. In fact, scholars usually deal with the intricacies of processing and deriving meaningful insights from vast collections of narrative and literary data. In response to this challenge, the so-called “distant reading” approach [2] is often employed, namely a set of different techniques that offers a method for analysing large number of texts in a more automated and quantitative manner. Instead of close, in-depth analysis of individual texts,

---

*2nd Italian Workshop on Artificial Intelligence for Cultural Heritage (IAI4CH 2023, <https://ai4ch.di.unito.it/>), co-located with the 22nd International Conference of the Italian Association for Artificial Intelligence (AIXIA 2023). 6-9 November 2023, Rome, Italy.*

\*Corresponding author.

✉ lpandolfo@uniss.it (L. Pandolfo); cardone@uniss.it (L. Cardone); lcutzu@uniss.it (L. Cutzu);

raffaella.perna@uniroma1.it (R. Perna); bseligardi@uniss.it (B. Seligardi); gsimi@uniss.it (G. Simi)

🆔 0000-0002-5785-5638 (L. Pandolfo); 0000-0003-0770-6324 (L. Cardone); 0000-0002-4679-3881 (R. Perna);

0000-0002-9400-0866 (B. Seligardi); 0000-0002-2206-2945 (G. Simi)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

distant reading looks at patterns, trends, and statistical data across a broad range of texts. This approach is particularly useful for uncovering overarching themes, identifying linguistic or stylistic trends, and making observations about literature on a larger scale. Moreover, it allows researchers to gain some insights that might be difficult and time-consuming to discern through traditional close reading methods. Nevertheless, this approach has its own set of shortcomings [3]. Distant reading often falls short when it comes to capturing the nuances and contextual complexity present in literary texts. It tends to prioritize quantitative patterns and statistical analysis over qualitative aspects, potentially missing out on the subtle meanings, interpretations and explanations that are fundamental in humanities research.

To address these limitations, the rapid proliferation of Semantic Web technologies and Artificial Intelligence (AI) techniques, has become increasingly prominent. In particular, Semantic Web and Linked Data facilitate the creation of structured, semantically rich data representations that enhance the contextual understanding of various types of cultural heritage texts [4, 5, 6, 7, 8]. This, in turn, allows for more nuanced and comprehensive analysis. Additionally, Machine Learning (ML) and Natural Language Processing (NLP) techniques can be employed to automate the extraction of meaning, sentiment, and relationships within literary corpora [9]. These techniques enable the identification of themes, sentiments, and even the recognition of subtle patterns within the text, offering a deeper layer of analysis beyond what distant reading alone can provide. By combining the strengths of Semantic Web technologies with AI methodologies, researchers in the DH domain can overcome the limitations of previous approach, allowing for more robust, context-aware, and insightful analyses of large literary texts. This integration opens up new possibilities for exploring and understanding humanities in an increasingly data-driven world. Also, bridging AI and cultural heritage objects, such as the literary texts, not only accelerates the processing of large volumes of data but also introduces avenues for nuanced comprehension and contextualization, signifying a fundamental shift in how we approach and engage with humanities research.

This paper aims to present the project "*Women Writing around the Camera*" (WOW) [10], funded by PRIN 2022, whose objective is to create an ontology-based portal in the field of humanities - following a similar approach to [11, 12, 13, 14, 15]. The WOW project will collect semantic data regarding noteworthy writings of Italian actresses, the so-called "*divagrafie*", with a specific focus on unveiling the intricate dynamics between writing, acting performance, and the cultivation of a star's image. This initiative builds upon the foundations established by the project "*Drawing a Map of Italian Actresses in Writing*" (DaMA), funded by PRIN 2017, which seeks to explore the scope, phenomenology, attributes, and historical trajectory of Italian actresses' writings (<https://www.damadivagrafie.org/>). The main goal of the WOW project is to develop a semantic portal containing various resource materials related to the visual and self-representative history of Italian actresses. This portal will facilitate a holistic examination, bringing together various forms of documentation (mainly photographs and texts) through a defined set of investigative methodologies, supported by AI techniques such as NLP and ML ones.

This paper is organised as follows: Section 2 presents the main aspects of the *divagrafie* phenomenon, which represents our application domain, while Section 3 outlines the methodology for developing the portal. Finally, some potential applications and expected impacts are given in Section 4.

## 2. The Divagrafie Phenomenon

The intersection of literature and cinema presents a unique field of cultural expression, one in which the narratives of actors and actresses transcend the silver screen to find a place within the written word. Renowned figures such as Franca Valeri and Monica Vitti exemplify this convergence, as their contributions span across the realms of acting, writing, and the introspective genres of autobiography and divagations.

The phenomenon of *divagrafie* stands as a captivating and multifaceted aspect of Italian cultural heritage, offering a unique view into the lives and experiences of Italian actresses who transcended the boundaries of their profession to become authors in their own right. Coined by Rizzarelli, the term "divagrafie" refers to the autobiographical writings produced by talented and versatile women [16]. While traditional academic discourse often overlooked these writings, they hold a deep significance for the study of both literary and cinematic history.

These texts are not mere chronicles of careers; they are literary artifacts that encapsulate the hopes, dreams, and ambitions of women who made significant contributions to Italian cinema and the broader world of entertainment. Through their writings, we not only gain insight into the struggles and triumphs that shaped their careers, but also the broader cultural, societal, and gender dynamics of the times in which these actresses thrived [17, 18].

Moreover, the study of *divagrafie* enriches our understanding of the intertwined relationship between literature and cinema. It provides a unique lens through which we can explore the convergence of artistic expression, offering a bridge between the worlds of storytelling on the stage or screen and within the pages of a book. These writings offer a backstage pass to the creative processes and personal reflections of these actresses, shedding light on the inspirations and motivations that fueled their performances and writings.

By exploring these writings, scholars not only uncover hidden gems of Italian literary history but also unravel the intricate connections between the art of acting and the art of storytelling. It provides a nuanced perspective on how these actresses navigated their dual roles as performers and authors, and it serves as a testament to the enduring impact of these women in Italian cultural and artistic history.

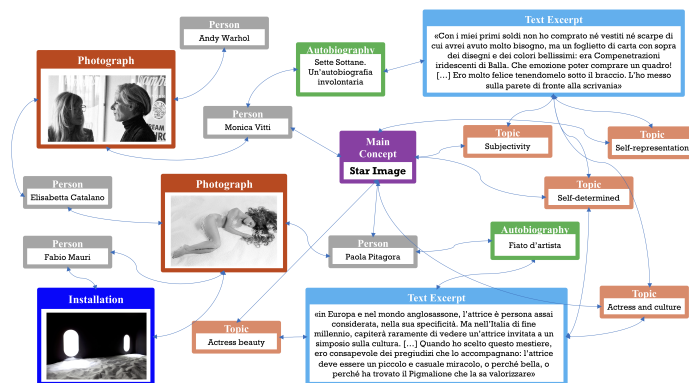
## 3. The WOW Project Overview

The primary goal of this project is to create the WOW semantic portal, which builds upon the ongoing research initiative known as DaMA. DaMA's mission is to delve into the expansion, phenomenology, characteristics, and historical progression of writings generated by Italian actresses, with a specific emphasis on the interplay between writing, acting performance, and the cultivation of their public image. These writings, referred to as *divagrafie*, constitute a diverse and intriguing body of texts that have traditionally been overlooked by the academic community. Despite their potential to significantly enhance our comprehension of stardom, as expressed through self-narratives or fictional writings, their value within the realm of star studies has only been partially recognized. Moreover, there exists a gap in the scholarship regarding the intricate relationship between the writings of a specific actress and the formation of her celebrity image.

This ongoing project is uncovering uncharted field and has thus far cataloged a substantial number of texts, including 80 autobiographies authored by 47 actresses. This literary collection, which is our primary focus, provides a distinctive testing ground for a convergent methodology that integrates AI techniques, Semantic Web technologies, and the latest advancements in Film and Media Studies, Literature, and Gender Studies. The DaMA team has thus far examined these volumes through a meticulous *close reading* approach, spotlighting paradigmatic examples and extracting various recurring themes. This analysis places particular emphasis on the intersections with perspectives from performance and actors' studies, as well as studies on stardom and celebrities, and additionally incorporates the use of videographic analysis tools.

Building on the multidisciplinary approach pioneered by the DaMA's research, current efforts are focused on expanding the existing corpus and applying on it AI techniques and semantic technologies. Below, we outline the key phases of our methodological approach:

1. **Recognition of DaMA Project's Outcomes and Data Integration.** This phase entails a thorough examination of all the materials collected by the DaMA research project team. Additionally, this stage involves the integration of new resources. In fact, our focus will be on retrieving visual materials – mostly photographs – housed in archives like the Elisabetta Catalano archive (Rome), Museo di Fotografia Contemporanea - MuFoCo archive (Cinisello Balsamo, MI), Museo del Cinema archive (Turin), Cinemazero archive, Luisa Di Gaetano archive (Rome), and Archivia - Casa Internazionale delle Donne (Rome). Furthermore, we intend to explore cinema and cultural periodicals to assess the circulation and impact of work by female photographers in the national press.
2. **Data Ingestion and Standardization.** Once collected, the materials need to be transformed into a standardized format, such as RDF (Resource Description Framework), as it enables the creation of Linked Data, establishing meaningful connections between different elements of the corpus. Moreover, the design of an ontology is crucial in this phase, since it acts as the structural backbone, providing a common language for organizing and interlinking the data [19]. Also, through the process of semantic annotation, each individual piece of data - whether a text or an image - is enriched with semantic identifiers. These annotations serve as the key to unlock the potential for meaningful queries and connections within the knowledge graph. As an example, the photographs that Elisabetta Catalano has taken of a variety of actors and actresses, writers and directors can be put in relationships with all the excerpts in the Italian actresses' autobiographies in which the relations with the cultural field strongly emerge as a symptom of overcoming the stereotype of actresses as a pure "not thinking" beautiful and fashioned body. As we can see in Figure 1, Elisabetta Catalano's photograph of Monica Vitti talking to Andy Warhol can be intertwined with passages in her autobiographies related to art as a means of self-expression and with other several entities and concepts in the corpus.
3. **Knowledge Graph Construction.** This phase relates the construction of the knowledge graph. This involves setting up an RDF triple store, which serves as the repository for our semantically annotated data. Each piece of information is represented in the form of triples, allowing for efficient storage and retrieval. Additionally, AI techniques are employed to further enhance and enrich the data. In particular, using ML algorithms for visual content and NLP for textual content, we aim at extracting additional features



**Figure 1:** Example of entities and relationships in the examined corpus.

or insights from the data. This may include identifying and uncovering objects, themes, sentiments, and more.

4. **Portal Development and User Interface.** The final step brings all the richly annotated and interconnected data to the forefront through the development of a user-friendly web portal. In this phase, we intend to investigate some automatic techniques for ontology population, such as those presented in [20, 21]. The interface is meticulously designed to ensure a seamless and intuitive user experience. By integrating a SPARQL endpoint, users have the power to craft advanced queries, enabling them to explore the knowledge graph in depth. Visualizations and exploration tools are implemented to present the data dynamically. Network graphs, timelines, and interactive maps are employed to convey the intricate relationships and contextual information encapsulated within the corpus.

By following these phases, our goal is to develop a comprehensive and accessible portal that unlocks the intricate network of knowledge within this domain, seamlessly blending the power of Semantic Web technologies and AI techniques.

## 4. Expected Impacts and Implications

The WOW project has significant potential applications in national and international research communities. It will provide access to a comprehensive collection of materials focused on Italian actresses, across various academic disciplines including Film Studies, Photography Studies, Women's Studies, and Literature Studies, among others. This initiative represents a pioneering effort in creating a corpus adhering to the rigorous practices of the DH approach. Additionally, the integration of semantic technologies will facilitate connections between data in the WOW portal and other international archives, enhancing its global impact.

Beyond addressing a research gap in the study area, specifically the examination of the relationship between women, photography, and cinema from a feminist perspective, this project aspires to a more profound and significant objective. It seeks to contribute to the cultural discourse in Italy, where gender issues, women's emancipation, feminist heritage, and women's

art are not yet adequately recognized as fundamental components of the country's cultural heritage. In terms of social impact, it will foster interest in cultural dynamics influenced by gender issues and promote a greater societal awareness of the cultural role of women in Italian society.

In alignment with the research objectives outlined in the National Research Program (PNR) 2021-2027 and the "Cultural Heritage" focus area of Horizon Europe, this project is poised to significantly impact the cluster "Humanistic culture, creativity, social transformation, society of inclusion" within the PNR, particularly the sub-category "Digital preservation and conservation of cultural heritage". All these research aspects will be considered within the development of the WOW portal that will have an expected impact in supporting the digital transformation of the cultural sector, but also strengthening the social inclusion, by reducing gender inequalities.

## Acknowledgments

This work has been supported by the project "WOMen Writing around the camera" (WOW) funded by Italian Ministry of University and Research (MUR) PRIN 2022, Grant 2022R43S2A.

## References

- [1] E. Gardiner, R. G. Musto, *The Digital Humanities: A Primer for Students and Scholars*, Cambridge University Press, 2015.
- [2] F. Moretti, *Distant Reading*, Verso Books, 2013.
- [3] F. Ciotti, *Distant Reading in Literary Studies: A Methodology in Quest of Theory*, *Testo e Senso* (2021) 195–213.
- [4] V. Bartalesi, C. Meghini, Using an Ontology for Representing the Knowledge on Literary Texts: The Dante Alighieri Case Study, *Semantic Web* 8 (2017) 385–394.
- [5] L. Pandolfo, L. Pulina, Building the Semantic Layer of the Józef Piłsudski Digital Archive with an Ontology-based Approach, *International Journal on Semantic Web and Information Systems (IJSWIS)* 17 (2021) 1–21.
- [6] E. Daga, L. Asprino, R. Damiano, M. Daquino, B. D. Agudo, A. Gangemi, T. Kuflik, A. Lieto, M. Maguire, A. M. Marras, et al., Integrating Citizen Experiences in Cultural Heritage Archives: Requirements, State of the Art, and Challenges, *ACM Journal on Computing and Cultural Heritage (JOCCH)* 15 (2022) 1–35.
- [7] 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 2021, October 26-28, 2021, SCITEPRESS, 2021*, pp. 341–345.
- [8] 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.
- [9] O. Suissa, A. Elmalech, M. Zhitomirsky-Geffet, Text Analysis Using Deep Neural Networks in Digital Humanities and Information Science, *Journal of the Association for Information Science and Technology* 73 (2022) 268–287.
- [10] L. Pandolfo, L. Cardone, L. Cutzu, B. Seligardi, G. Simi, Building the Semantic Portal of Italian Divagrafie (short paper)., in: *IPS/RiCeRcA/SPIRIT@ AI\* IA, 2022*.

- [11] G. Adorni, M. Maratea, L. Pandolfo, L. Pulina, An ontology for historical research documents, in: *Web Reasoning and Rule Systems - 9th International Conference, RR 2015*, Berlin, Germany, August 4-5, 2015, Proceedings, volume 9209 of *Lecture Notes in Computer Science*, Springer, 2015, pp. 11–18.
- [12] G. Adorni, M. Maratea, L. Pandolfo, L. Pulina, An Ontology-Based Archive for Historical Research, in: *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.
- [13] L. Pandolfo, L. Pulina, M. Zieliński, Towards an Ontology for Describing Archival Resources, in: *Proceedings of the Second Workshop on Humanities in the Semantic Web (WHiSe II) co-located with 16th International Semantic Web Conference (ISWC 2017)*, Vienna, Austria, October 22, 2017, volume 2014 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2017, pp. 111–116.
- [14] L. Pandolfo, L. Pulina, M. Zieliński, ARKIVO: an Ontology for Describing Archival Resources, in: P. Felli, M. Montali (Eds.), *Proceedings of the 33rd Italian Conference on Computational Logic*, Bolzano, Italy, September 20-22, 2018, volume 2214 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2018, pp. 112–116.
- [15] L. Pandolfo, L. Pulina, M. Zieliński, Exploring Semantic Archival Collections: The case of Pilsudski Institute of America, in: *Digital Libraries: Supporting Open Science - 15th Italian Research Conference on Digital Libraries, IRCDL 2019*, Pisa, Italy, January 31 - February 1, 2019, Proceedings, volume 988 of *Communications in Computer and Information Science*, Springer, 2019, pp. 107–121.
- [16] M. Rizzarelli, L'attrice che scrive, la scrittrice che recita. per una mappa della 'diva-grafia', *Vaghe stelle. Attrici del/nel cinema italiano*. Arabeschi, edited by Lucia Cardone, Giovanna Maina, Stefania Rimini, and Chiara Tognolotti 10 (2017) 366–371.
- [17] G. Simi, L'occhio che palpita. Monica Vitti e gli scritti sull'arte, *Cinergie—Il Cinema e le altre Arti* (2021) 153–166.
- [18] C. Tognolotti, Una diva fragrante. L'immagine divistica di Sophia Loren nei libri di ricette (2019).
- [19] N. Guarino, D. Oberle, S. Staab, What is an Ontology?, in: *Handbook on Ontologies*, Springer, 2009, pp. 1–17.
- [20] L. Pandolfo, L. Pulina, G. Adorni, A framework for automatic population of ontology-based digital libraries, in: *AI\*IA 2016: Advances in Artificial Intelligence - XVth International Conference of the Italian Association for Artificial Intelligence*, Genova, Italy, November 29 - December 1, 2016, Proceedings, volume 10037 of *Lecture Notes in Computer Science*, Springer, 2016, pp. 406–417.
- [21] ADnOTO: a self-adaptive system for automatic ontology-based annotation of unstructured documents, in: *Advances in Artificial Intelligence: From Theory to Practice - 30th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2017*, Arras, France, June 27-30, 2017, Proceedings, Part I, volume 10350 of *Lecture Notes in Computer Science*, Springer, 2017, pp. 495–501.