Ontofest: An Ontology to Integrate and Retrieve Data from the Locarno Film Festival Archives

Alessandro Cosentino[1], Webert Júnio Araújo[2] and Fabio Crestani[1]

[1]*Faculty of Informatics, Università della Svizzera italiana, Lugano, Switzerland*
[2]*Escola de Ciência da Informação, Universidade Federal de Minas Gerais*

Abstract
This work explores the transformation of the Locarno Film Festival (LFF) archives into dynamic, interconnected resources through the development of an ontology. The main aim of this work is to facilitate access to the information related to the LFF, thereby improving its discoverability and making it more user-friendly for research and preservation of cinematic heritage. Adopting an interdisciplinary approach, we developed OntoFest following the "Ontology Development 101" methodology. Thanks to the collaboration with LFF and RSI (Radiotelevisione Svizzera), it was possible to integrate data from diverse sources. Significant results of this work include the development of an ontology that facilitates semantic search and analysis, focusing on information related to the LFF. This work highlights the potential of digital archives in the cinematic field but also provides a foundation for future initiatives in the digitalization of cinematic heritage archives. In conclusion, OntoFest lays the foundation for the valorization and integration of cinematic heritage, which has profound implications for its preservation.

Keywords
Digital Archive, Semantic Web, Interoperability, Linked Open Data, Ontology

1. Introduction

In the era of cultural digitalization, archives are no longer passive repositories of defunct data. In the "Age of Big Data", where the previously overwhelming volume of data is now increasingly becoming more accessible and controllable, these archives must be reconceptualized as dynamic and significant repositories of knowledge. In this work, we focus on the Locarno Film Festival (LFF) archives as an outstanding case study of cinematographic archival collections, highlighting a largely dormant yet increasingly recognized legacy of which we are only now beginning to become aware. These archives, along with those of RSI Radiotelevisione Svizzera in Lugano and Cinémathèque Suisse in Lausanne, play a crucial role in preserving cinematic and cultural heritage. Spanning various locations and ranging from analog to digital formats, these archives are essential for historians and require proper management. Currently, the significant challenges include efficiently locating, accessing, and integrating these materials into new research. The OntoFest project aims to address a specific sector within the broad research area of humanities by developing standards and creating tools for the diffusion of Open Research Data (ORD) for

the analysis of cinematography. In particular, this project aims to: (i) create an ontology for semantically rich descriptions of the cataloged material; (ii) produce data following the Findable, Accessible, Interoperable, and Reusable (FAIR) principles [1]; and (iii) publish the data as Linked Open Data (LOD). The final goal of the project is to provide access to research material about films in various formats and supports. The main novelty introduced by OntoFest lies in its aim to transform traditional film archives into a dynamic and interconnected resource. This transformation is pivotal in enhancing the discoverability and usability of cinematic heritage, aiding both research and preservation initiatives. OntoFest distinguishes itself from other projects that emphasize the historical or cultural aspects of film, focusing on creating a versatile, semantically rich framework. This framework, initially tailored for the LFF, is designed with the goal to be adaptable to other film festivals, thereby improving discoverability and accessibility to this information.

This paper is organized as follows. Section 2 presents relevant projects and initiatives aimed at providing access to cinematographic research materials. Section 3 details the methodology used in the development of OntoFest, following the guidelines described in the Ontology Development 101 [2]. Section 4 discusses the specific characteristics of OntoFest, including its framework, data model and the structure of its ontology. Finally, section 5 offers final remarks and expectations for future works.

2. Related Work

The digitization of cultural heritage has emerged as a key area of interest in a wide range of research studies and projects. Among different frameworks, CIDOC-CRM [3] and its extensions, such as FRBRoo [4] and its reformulation LRMoo [5], hold particular importance. These models offer a comprehensive structure for cultural heritage data, yet they mainly focus on particular aspects that might not fully align to the particular needs of the seventh art domain. However, recent literature indicates numerous efforts to adapt these methodologies to the cinematic context, making significant progress in managing cinematic cultural heritage. For example, the Cinema Context [6] project investigates the history of Dutch cinema from the late 19th century, with a primary focus on distribution and dissemination, but it omits key figures such as directors and actors. OntoFest aims to fill this gap by defining an ontology that encompasses all aspects of cinema, with particular emphasis on central figures not fully explored in Cinema Context. Similarly, the CINECOS project [7] proposes to develop an open-access platform for exploring the history of cinema in Belgium and Flanders, integrating information on production, distribution, and other fundamental aspects. Its goal is to make data easily accessible and interoperable, thereby promoting research and preservation of cinematic heritage. While this project covers various cinema aspects, it does not particularly emphasize the development of a comprehensive ontology, a gap highlighted by the authors of Cinema Belgica [8], a spin-off of the project. Finally, the European Cinema Audiences (ECA) project [9] examines the film culture of the 1950s in seven European cities, collecting and harmonizing a standardized archive of materials related to the cinematic experience. The project focuses on audience experience and film popularity in a post-war context, providing valuable data on demographics, preferences, and audience behavior. While ECA aims to understand the diverse cinematic experiences

across Europe, OntoFest focuses on film production and presentation, effectively connecting the history of cinema to the structures and events of film festivals. In conclusion, projects like Cinema Context, CINECOS, and ECA significantly contribute to the understanding of cinema and its audience. However, there is a clear need for an integrated and comprehensive approach to representing the film industry, especially with respect to film festivals. OntoFest proposes to tackle this challenge, offering a holistic approach that connects the history, practice, and structural aspects of cinema, enriching the research and digitalization for this domain. This review has been limited to projects offering insights analogous to the goals of OntoFest, with a particular focus on those works that explicitly deal with cinematic archives. Further research in related areas might provide additional perspectives and insights beyond the scope of this study.

3. Methodology

The OntoFest ontology is developed using the Protégé editor and a knowledge management system, adhering to the guidelines outlined in Ontology Development 101 [2], chosen for its straightforward procedural steps and its alignment to the fundamental principles in ontology engineering. This methodology emphasizes best practices in ontology creation, including iterative development, collaboration, and integration with existing ontologies. The choice of this approach has ensured a robust and scalable ontological structure, facilitating knowledge representation and reasoning.

This approach involves seven steps:

Step 1: Determine the Domain and Scope of the Ontology At the first meeting with the project stakeholders, three key considerations for defining the scope of the OntoFest Ontology were discussed:

1. Domain covered by the OntoFest Ontology.

2. Potential users of the OntoFest Ontology.

3. Types of questions the OntoFest Ontology should answer.

The domain covered by the ontology is that of film festivals, covering various aspects of a film festival such as film screenings, festival events, directors and actors, critic reviews, and data on television programming regarding the film festival during its occurrence. This information, provided by RSI, may explore various elements of the festival, including the showcased films, details about the directors and actors involved, award ceremonies, and other significant aspects of the festival. Users will include researchers, film historians, students, and cinema enthusiasts. Some Examples of competency questions the OntoFest should be able to answer:

- **Film-Related Questions**: Film screenings, genres, directors, and actors involved in the festival.

- **Media Coverage Questions**: Coverage of films, actors, directors, and key events by television programs.

- **Awards Questions**: Information on awards given at the festival, including categories and winners.

Step 2: Consider reusing existing ontologies This step involved a comprehensive analysis of existing ontologies that fully or partially cover the domain of interest. We aim to assess and refine these ontologies to better align them with our specific requirements, ensuring they remain fully interoperable with the original frameworks. For this purpose, the Cinema Context [6] and the Cinema Belgica projects [8] have been selected as prominent examples. A detailed review of these projects was conducted to identify classes and properties that are pertinent to the domain our new ontology aims to cover.

Step 3: Enumerate important terms in the ontology The step involved in enumerating terms relevant to the domain of film festivals consisted of several phases. Initially, we collected and organized specific terms, starting with the identification of key components such as films, directors, actors, screenings, and festival events. Subsequently, we focused on terms describing the qualities or attributes of these entities. Finally, the process addressed terms that define the relationships between these entities, distinguishing between interactions or connections among entities and the inherent properties of each entity.

Step 4: Define the classes and the class hierarchy This step focused on grouping enumerated entities into classes and specifying their interrelations, adopting a combined approach (top-down and bottom up). We have started with some general classes like Film, Film Festivals, and some specific ones like TV Show, moving on to intermediate classes such as Film Festival Edition, Film Festival Section, Film Festival Award and so on. This led to the establishment of a taxonomy, which forms the hierarchical aspect of the ontology. We strongly emphasized the fundamental importance of each class, highlighting that classes should not merely serve as linguistic labels but encapsulate core concepts of the domain.

Steps 5 and 6: Define slots and their facets These were carried out together due to their mutual focus on class properties, covering both object and data properties. The focus was to define the crucial attributes for each class. Particular attention was given to delineating and establishing relationships among different classes. Finally, we specified the facets of properties and relationships, including cardinality, value type, domain, and range, and provided some axioms, such as class disjunction, inverse relations, and some restrictions (existential restriction and universal restriction).

Step 7: Create instances Finally, after the ontology schema was ready, we automatically populate it with data from the LFF and RSI archives.

4. OntoFest

The ontology is built upon five interconnected entities: Film, Film Festival, Person, TV Show, and Archive. These core entities have been meticulously selected to capture the essence of the modeled domain. Each of these entities is assigned a URI sourced from Wikidata [10], underlining the importance of leveraging shared vocabularies for the Semantic Web. Where the Wikidata vocabulary falls shorts, it is complemented by Dublin Core [11] and Schema.org [12]. This choice is crucial in improving the ontology's level of interoperability and reusability,

aligning it with the FAIR principles. Furthermore, we introduced custom classes and properties to depict terms not represented in the reference vocabularies. How we model these entities is described below (see Figure 1).

Figure 1: Classes and their interrelation in OntoFest. Note: The ontology includes inverse relationships and data properties in its representation that are not depicted in this figure.

Here is a short description of the classes of Ontofest:

Archive: represents the television and film archives. In our case, these are maintained and preserved by two organizations, LFF and RSI, both contributing to this project.

Film: provides information about its title, description, country of production, year of production, genre, and duration. These films represent the entire collection presented at the LFF. All this information is preserved at the LFF Archive.

Film Festival: comprises three subclasses: Film Festival Edition, Film Festival Section, and Film Festival Award, which offer a holistic view of the domain, encompassing the awards given out, detailed information on all past editions, and the various sections where films are screened.

Person: comprises two subclasses: Cast and Director, which play a crucial role in the domain by identifying the individuals behind the films. The key data properties include biographical information, providing a structured way to reference these individuals.

TV Show: provides details on the television coverage during the LFF, reflecting media perspectives on elements such as actors and directors participating in the festival, films showcased

in particular sections, and the awards received. All this information is archived at the RadioTelevisione Svizzera (RSI) Archive. The key data properties include broadcast date and time, program name, program description.

To validate the consistency of the ontology we employed HermiT reasoner [13]. This check guarantees that OntoFest infers knowledge accurately and delivers correct information. Table 1 shows some queries executed within the ontology. The results obtained from these queries align with reality, confirming that the model faithfully represents its domain.

Table 1
A list of SPARQL queries and the respective results retrieved from the OntoFest ontology. Note: Partial Results only contains a subset of the total results due to space constraints.

Q No.	Question	SPARQL Code	Partial Results
Q1	Which films, along with their production countries and directors, were featured in the 75th edition of the LFF?	SELECT ?films ?directors ?country WHERE { ?films schema:workPresented 75th_LFF_Edition ; of:isDirectedBy[1] ?director ; wiki:P495[2] ?country.}	[Film: Delta; Country: Italy; Director: Vannucci ... Film: Drii Winter; Country: Switzerland; Director: Koch]
Q2	What film from the "Concorso Internazionale" section was awarded the "Pardo d'Oro" at the 75th edition of the LFF?	SELECT ?films ?IMDB_id WHERE { ?films schema:workPresented of:75th_LFF_Edition; schema:workPresented of:Concorso Internazionale; wiki:P166[3] of:Pardo d'Oro; schema:sameAs ?IMDB_id. }	Film: Regra34; IMDB_id: tt15520368
Q3	Which television programs discussed the film that won the "Pardo d'Oro" in the "Concorso Internazionale" section at the 75th edition of the LFF?	SELECT ?films ?TVShow WHERE {?films schema:workPresented of:75th_LFF_Edition ; schema:workPresented of:Concorso Internazionale ; wiki:P166[3] of:Pardo d'Oro ; of:ismentionedBy ?TVShow.}	Film: Regra 34; TVShow: [Telegiornale, Il Quotidiano, Fuori Schema]

5. Conclusion and Future Work

OntoFest contributes to the broader effort of digitizing Film Festival archives for preserving cultural heritage and extracting value from information that might otherwise remain unused.

---

[1] of stands for Ontofest and isDirectedBy is the inverse relationship of hasDirected

[2] wiki:P495 stands for Country of Production

[3] wiki:P166 stands for awardReceived

[4] of:ismentionedBy is the inverse relationship of schema:mentions

This work aligns with similar European initiatives by adopting and promoting good practices in digital archive transformation and ontology development. However, we believe that OntoFest represents a step towards a more interconnected and accessible digital landscape for digitizing information in film festival archives. We believe that the ontology development process should be iterative and ongoing, since ontologies can only be improved as they are applied to actual instances of data and when these data are used to answer research questions. For this reason, we are collaborating with film experts to refine the ontology, ensuring it is best suited for their purposes. The work is at its early stage and this paper introduces our ontology. Although our initial results indicate that the developed ontology performs well in facilitating access to research material related to the LFF, future developments are crucial to further enhance the ontology's robustness. More specifically:

- **Ontology Enrichment**: we plan to integrate data from other essential sources, notably the Cinémathèque Suisse, a key player in LFF archives, to further enrich the ontology with greater value and diversity. To achieve this, we intend to use ontology enrichment methods [14] and leverage established shared vocabularies such as FOAF [15] and SKOS [16], as well as LOD dataset like DBpedia [17], to enhance the semantic content and knowledge representation within the ontology.

- **Automatic Analysis of Textual Data**: we aim to develop methods for the automatic analysis of collected textual data, such as film reviews and newspaper articles about film screenings. We intend to directly link these reviews to the corresponding films listed in OntoFest. This linkage will not only enrich the documents with relevant metadata but also enhance the ontology by adding contextual information to the film entries. For instance, we aim to automatically identify reviews pertaining to a specific film and then cross-link these reviews with the corresponding entries in OntoFest, as well as with other relevant Linked Open Data datasets.

- **Development of User Interfaces**: we plan to develop a unified user interface that integrates two complementary functionalities for data exploration. This interface will have a visual browsing feature, which will enable users to navigate and explore data intuitively without requiring technical expertise. Access to data using a direct query function will be facilitated through a SPARQL endpoint, accessible via an API, enabling complex queries in line with LOD standards and allowing for potential cross-references with other LOD datasets. These two functionalities are designed to work seamlessly together, allowing users to transition fluidly between them. For example, a user might start with a direct query to fetch specific data, then switch to browsing mode to explore related information, or vice versa.

- **Evaluation Methods**: we plan to conduct two different types of evaluation: a system evaluation [18] and a user evaluation. To start with, we will employ an automatic system to evaluate the ontology such as OntoEval [19]. Later, we will conduct a user-based evaluation, specifically through questionnaires. Based on the results of these evaluations, we will review and refine the ontology.

References

[1] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The FAIR guiding principles for scientific data management and stewardship, Scientific data 3 (2016) 1–9.

[2] N. F. Noy, D. L. McGuinness, Ontology Development 101: A Guide to Creating Your First Ontology, Technical Report KSL-01-05 and KSL Technical Report, Stanford Knowledge Systems Laboratory, Stanford, CA, USA, 2001. https://protege.stanford.edu/publications/ontology_development/ontology101.pdf.

[3] M. Doerr, The cidoc conceptual reference module: An ontological approach to semantic interoperability of metadata, AI Magazine 24 (2003) 75–75.

[4] M. Doerr, C. Bekiari, P. LeBoeuf, B. nationale de France, Frbroo, a conceptual model for performing arts, in: 2008 Annual Conference of CIDOC, Athens, Greece, 2008, pp. 15–18.

[5] P. Riva, M. Žumer, Frbroo, the ifla library reference model, and now lrmoo: A circle of development, in: IFLA WLIC 2018, Kuala Lumpur, Malaysia, Kuala Lumpur, Malaysia, 2018.

[6] L. van Wissen, T. van Oort, J. Noordegraaf, I. Kisjes, Cinema context as linked open data: Converting an online dutch film culture dataset to rdf, in: SEMANTiCS Posters & Demos, 2021.

[7] D. Biltereyst, P. de Potter, Cinema ecosystem (cinecos): A new cinema history inspired project aiming at building an open access data platform for cinema history in flanders and belgium, Presentation at CREATE Salon, Amsterdam (17 May 2018), 2018. URL: http://www.create.humanities.uva.nl/events/CREATE-Salon-Historical-Events-Data-Models-2/.

[8] V. Ducatteeuw, D. Biltereyst, P. Meers, C. Verbruggen, D. Moreels, J. Noordegraaf, S. Chambers, P. De Potter, T. Cachet, N. Franck, et al., Critical reflections on cinema belgica: The database for new cinema history in belgium, Journal of Open Humanities Data 9 (2023).

[9] T. Porubčanská, Building a digital archive for cross-national historical research, Iluminace 34 (2022) 101–112.

[10] D. Vrandečić, Wikidata: A new platform for collaborative data collection, in: Proceedings of the 21st International Conference on World Wide Web, 2012, pp. 1063–1064.

[11] S. Weibel, J. Kunze, C. Lagoze, M. Wolf, Dublin Core Metadata for Resource Discovery, Technical Report RFC 2413, Network Working Group, 1998. OCLC Online Computer Library Center, Inc., University of California, San Francisco, Cornell University, Reuters Limited.

[12] R. V. Guha, D. Brickley, S. Macbeth, Schema.org: Evolution of structured data on the web, Communications of the ACM 59 (2016) 44–51.

[13] B. Glimm, I. Horrocks, B. Motik, G. Stoilos, Z. Wang, (hermit): An owl 2 reasoner, Journal of Automated Reasoning 53 (2014) 245–269.

[14] W. J. Araújo, G. Â. de Lima, A methodological proposal towards domain ontology enrichment, in: Knowledge Organization at the Interface, Ergon-Verlag, 2020, pp. 23–30.

[15] M. Graves, A. Constabaris, D. Brickley, Foaf: Connecting people on the semantic web, in: Knitting the Semantic Web, Routledge, 2013, pp. 191–202.

[16] A. Miles, S. Bechhofer, Skos simple knowledge organization system reference, 2009. URL: http://www.w3.org/TR/skos-reference/.

[17] M. Booshehri, P. Luksch, An ontology enrichment approach by using dbpedia, in: Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics, Association for Computing Machinery (ACM), 2015, pp. 1–11.

[18] J. Völker, D. Vrandečić, Y. Sure, Automatic evaluation of ontologies (aeon), in: International Semantic Web Conference, Springer, 2005, pp. 716–731.

[19] A. Zaitoun, T. Sagi, K. Hose, Ontoeval: an automated ontology evaluation system, in: Companion Proceedings of the ACM Web Conference 2023, 2023, pp. 82–85.