JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

JLIS.it

Corago in LOD. The debut of an Opera repository into the Linked Data arena

Paolo Bonora[(a)], **Angelo Pompilio**[(b)]

a) Alma Mater Studiorum - Università di Bologna, https://orcid.org/0000-0001-8337-3379
b) Alma Mater Studiorum - Università di Bologna, https://orcid.org/0000-0003-3881-5312

Contact: Paolo Bonora, paolo.bonora@unibo.it; Angelo Pompilio, angelo.pompilio@unibo.it
Received: 16 December 2020; **Accepted:** 24 February 2021; **First Published:** 15 May 2021

ABSTRACT

The paper examines the adoption of the Semantic Web (SW) technologies and Linked Data (LD) principles to manage a knowledge base about opera. The Corago repository collects historical data and documentation about opera works, performances and librettos from the 16th to the 20th century. We experimented the use of semantic technologies to manage the repository's knowledge catalogued following the Functional Requirements for Bibliographic Records (FRBR) relational model. Cultural Heritage Knowledge Bases (CHKB) as Corago could leverage SW and LD to overcome proprietary models and to introduce new information to better satisfy user's requirements. Two well-established reference ontologies as CIDOC Conceptual Reference Model (CIDOC CRM) and FRBR Object Oriented (FRBRoo) are adopted to transpose contents form the legacy conceptual model to RDF. Through the process, we observed that formal semantics allow, not only to adequately represent the opera domain, but also enable to define the way information is being presented to users. This led to the definition of "reception pathways" which become themselves part of the knowledge about opera within the KB. This novel semantic approach is introduced with the Corago Semantic Model (Corago SM), a domain ontology dedicated to functional representation of opera's historical data. Advancements have been tested with an experimental system and assessed through a questionnaire submitted to a panel of users.

KEYWORDS

Opera; Performing arts; Knowledge base; Formal ontologies; FRBRoo; CIDOC CRM.

CITATION

Bonora, P., Pompilio, A. "Corago in LOD. The debut of an Opera repository into the Linked Data arena." *JLIS.it* 12, 2 (May 2021): 54–72. DOI: 10.4403/jlis.it-12699.

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

1. Introduction and background[1]

Since the publication of the Functional Requirements for Bibliographic Records by IFLA in late nineties, user requirements got a renewed attention in cataloguing applications design. The focus shifted from descriptive models itself to use cases within a wide range of Cultural Heritage (CH) domains, including performing arts. On the technological side, standard metadata formats were merged with Semantic Web technologies coupling data modelling and concept definitions in coherent semantic datasets. Also, some ontologies dedicated to music such as The Music Ontology emerged. Since then, following the introduction of Linked Data principles, within the musicological domain the alignment to generalist datasets, as DBPedia,[2] has been experimented (Raimond e Sandler 2008). These trials pursue the introduction of newer tools to support the dissemination of contents and digital assets. An example of the former is the project promoted by the British Broadcasting Corporation (BBC) Radio industry in 2009 looking for the alignment of its catalogue with external datasets (Kobilarov *et al.* 2009). The goal was to exploit the alignment to external resources to access to the BBC's repository. A similar case study is the MIMESIS project (Bartolini *et al.* 2012) that drafted an approach relying on SW tools for browsing the Directory of Italian Poetry in Music 1500-1700 (RePIM).[3] Other indexing services as the European Collected Library of Artistic Performance (ECLAP)[4] are a clear example of demanding requirements in terms of metadata interoperability in the performing arts domain. Aggregating contents from 35 institutions from 18 countries, ECLAP's information architecture adopted a semantic model to cope with the heterogeneity and limitations of metadata commonly adopted for metadata interoperability (Bellini e Nesi 2014).

In the musical domain, a peculiar role is played by repertoires which represent a specialized form of authority reference (Ceriani 2019). A controlled index of musical works, authors and performances is the key to retrieve, sort and understand information gathered from different sources. Such index would collect links between concepts and facts constituting the web of knowledge about a specific resource. To achieve this aim, its conceptual model must describe entities and their relations in a formal way. Formal ontologies are then the natural solution to manage the knowledge of musical repertoires in order to achieve better search, navigation and interoperability capabilities.

While the set of principles and practices from the Linked Open Data (LOD) paradigm are the way to realize a CH's web of data (Bizer, Heath, e Berners-Lee 2011), peculiarities of the domain has been taken into account. This led to the introduction of domain specific ontologies to properly express semantics. The aim is a seamless navigation between repertoires and related digital objects even beyond the "physical" boundaries of each repository. The repertoire's conceptual model must so fulfil both interoperability and functional requirements. The KB's ontology stack should then be selected to support a full LOD interoperability and high-level end-user's functional requirements.

[1] While both authors contributed to the research being illustrated here, P. Bonora is responsible for section 3,4,5; A. Pompilio is responsible for section 1 and 2; all authors collaborated in writing section 6.
[2] http://wiki.dbpedia.org/.
[3] Another interesting example is the alignment project between the Early Music Online (EMO) and the Electronic Corpus of Lute Music (ECOLM) (Crawford T. *et al.*, 2014).
[4] http://www.eclap.eu/.

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

These objectives guided the process that led to the release of the Corago LOD dataset,[5] the Corago Semantic Model and the implementation of a new web application that will replace the previous legacy one based on the traditional FRBR-ER model. We are presenting Corago LOD as a significant case study in the CH domain as it represents the latest evolution of a decades-long cataloguing activity based on an original descriptive model dedicated to opera history and documentation (Pompilio *et al.* 2005).

The traditional Corago repository is divided into three core areas: the repertoire of works and authors, the chronology of events and libretto's directory.

Starting from 1600 to 1900, a total of over 20,000 works, a nearly 40,000 names and more than 40,000 events and 57,000 librettos are listed. The Corago's conceptual model is the final evolution of the project RADAMES.[6] Over time, in its different implementations, the RADAMES descriptive model allowed to successfully describe different types of documents of the performing arts domain: opera librettos (Cipollone 2009), audio and video recordings, scenic arrangements, sketches and costumes (Bonora 2010). Originally designed as an implementation of the FRBR model, including elements derived from FRAD,[7] the Corago model represents a significant specialization of both. The relation between authors and their works (responsibility) and between the same works as well as in their internal structure (segmentation) is a core feature of the descriptive model. After the emerging of CIDOC CRM as a relevant standard in CH, some trials about the possible use to define entities from the RADAMES model has been made but did not reach a full implementation of the transposition process (Bonora, Ossicini, e Raffa 2006).[8] Since the introduction of the FRBRoo ontology (Doerr 2009) which is natively aligned to CIDOC CRM, the two ontologies become candidates as reference ontologies to express Corago's contents. During the transposition process from the original relational model to RDF we addressed both the semantics of the legacy model and adopted a user-oriented functional approach.

In paragraph 2 we introduce the approach of the experimentation and resulting requirements; paragraph 3 illustrates the experimental system's architecture; in paragraph 4 we present the ontology driven navigation paradigm and the resulting semantic model; paragraph 5 discusses evaluation trial and results; paragraph 7 draws conclusions.

2. Requirements: adopting the end user's perspective

We implemented an experimental system to perform a functional comparison with the legacy Corago's web application based on the legacy relational model in order to assess the effectiveness of the transposition. The objective was to point out benefits and drawbacks resulting from the adoption of a LOD KB as data source for the user application development. Success would indicate the feasibility of a reliable, efficient and effective representation of a deeply structured, and domain specific

[5] The Corago LOD dataset is available at: https://doi.org/10.5281/zenodo.3865867.

[6] For a complete discussion on the origins, the theoretical framework and the objectives of the project RADAMES see (Bianconi, Pompilio, e Pagannone 2004).

[7] FRAD: Functional Requirements for Authority Data is the conceptual reference model proposed by IFLA for description of the entities surveyed in authority files. It was born as a complement to FRBR for the definition of subject headings, thesauri, taxonomies and reports in support of the entities identified therein.

[8] Trials that are among the first experiences in the use of CRM in the field of performing arts (Le Boef 2012).

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

descriptive model, as required by historical data of opera. We also investigated if, thanks to DLs expressiveness, newer ways to exploit KB's contents become viable, such as different searching, browsing and data visualization modalities. We also expected positive results from the normalization of dimensions already present within the original database and the introduction of others from the alignment of the KB with external LOD resources.
Requirements for the experimental system were split into two main categories: requirements for a functional equivalence with a traditional web application and requirements that could take advantage of the LOD features. The new application should be able to perform both search for specific types of entities (i.e. consistent results for a single class) and generalized search (i.e. returning mixed types result sets) or search for entities using specific attributes or relations. We also considered some additional features previously not available in the legacy application: full text searches over all classes; search based on the relation types and space-time search.
The following diagram depicts the use cases implemented within the experimental system.

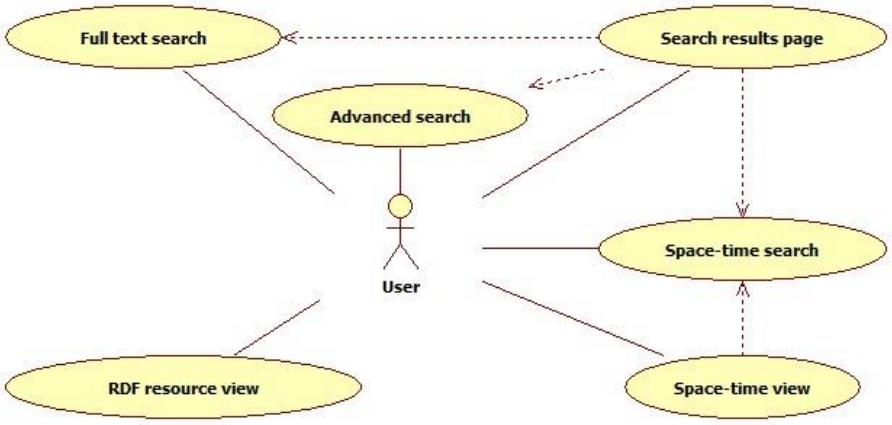

Diagram 1

Starting from the end user's point of view, the model design must be balanced between readability and analyticity of the description levels. This is mainly related to the characteristics of the two ontologies adopted as references for the dataset definition: the CIDOC CRM and FRBRoo. Nevertheless, there are some peculiar issues that arise in while expressing the performing arts domain trough the FRBR model's semantics. These spans from a proper identification of the "expression" level of a performance to a complete description of the author's role (Doty 2013). Depending on the characteristics of the resource being represented, implementation may lead to a very analytical graph to avoid information losses during the transposition to RDF. The resulting verbosity may be scarcely readable by the user while its complexity would still be required for proper definition of contents in the KB. The role that an author has played in the conception of a work is a typical example. Conceptually it represents the kind of relation linking an author and a man-made thing. The CIDOC CRM defines the crm:P14_carried_out_by property to link the author to authored entity and a "sub-property" crm:P14.1_in_the_role_of to specify in a crm:E55_Type instance the kind of contribution the author gave. This pattern, although conceptually correct, still being not implementable in RDF. This leads to some technically feasible solutions that may introduce long property chains between the

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

two major entities (Alexiev 2012). Then a bridge between the required analytical description in RDF and a reasonable compromise with usability is required. Other semantic models in the CH domain, as Europeana's EDM or musical oriented ontologies as DOREMUS,[9] try to set a balance between an analytical description of facts and data usability. We introduced instead an intermediate layer that decouples the application from its data layer and introducing specializations required to build user views. This layer also becomes a tool to formalize a competence on how to represent historical data about opera in a way that maximises its reception by users with different skills or interests. While an opera enthusiast may search for characters interpreted by a popular performer, a scholar could be interested to identify performers' movements through space and time to investigate the history of companies of actors through seasons and repertoire.

The adoption of a multi-layered semantics allows to manage different analytical levels in a single and coherent semantic model. This approach enables sharing definitions about domain knowledge representation between the community of experts. This multi-layered ontology design also supports the definition of specialized interfaces for interoperability with other CHKBs.

The resulting Corago Semantic Model (Corago SM)[10] was designed as a domain related ontology expressed in OWL. It is a natively aligned extension of both CIDOC CRM and FRBRoo. The conceptual model introduces an intermediate layer to support end user functional views over the KB as an implementation of the semantic lens pattern (Peroni *et al.*, 2014).

3. Architecture and methodology

The design started considering the granularity of the graph resulting from the highly analytical level and the strong characterization towards the temporal dynamics of the two reference ontologies. While this approach offers a considerable accuracy for a complete semantic description of resources, the resulting graph is slightly complex. While this guarantees a fine-grained semantic interoperability, which is good for LOD alignments, the usability of the KB is bound to a deep knowledge of these ontologies by users. Then, we should use fine-grained ontologies to describe domain entities at their maximum semantic granularity and derive from this a higher-level representation that could be easily understood both by domain experts and end users. This approach attempts to abstract the knowledge from the technical means that handles it and introduces an abstraction that fosters the access to information from a functional point of view.

The resulting architecture is represented in diagram 2.

[9] Refer to: http://www.doremus.org/.

[10] The Corago SM ontology is available at: https://github.com/paolobonora/Corago-SM/.

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

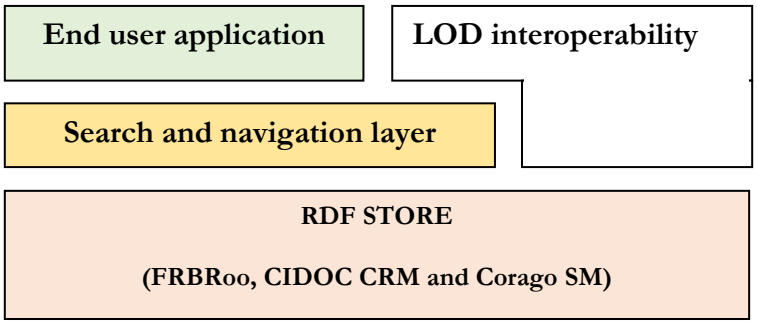

Diagram 2

In this design, the knowledge is defined through core reference ontologies and can be directly accessed by all applications that require the maximum level of descriptive granularity. These will typically be machine-to-machine applications or alignment activities from other KBs. In parallel, the intermediate layer is an interface to express semantics, of relationships and properties, as abstractions of the underlying analytical description. The goal is to create a semantic layer, in our case specialized for the opera domain, which will support the navigation of KB's contents.
This approach should be considered complementary to the definition of the KB using extensions of already existing domain ontologies. In this perspective, the adoption of multiple vocabularies and the introduction of specialization only where these are lacking expressivity is an approach followed by similar projects in the music domain (Pattuelli, Provo, e Thorsen 2015). Conversely, we based the dataset definition mostly on CIDOC CRM and FRBRoo and implemented the layer as part of a complementary semantic model within the Corago SM.
From previous experiences of adoptions of CIDOC CRM and FRBRoo in the CH field (Bellini e Nesi 2014) and (Byrne 2008) and guidelines for adopters of their promoting working group we derived some basic guidelines for data migration from the legacy relational data source. From URI formation to class selection following a principle of maximum generalization without loss of meaning, the alignment between the source model and the target semantic model has been carried out as in iterative process. During this process, the Corago SM was extended only when implementation constraints or emerged any lack of expressiveness in the two references ontologies.
Besides, this phase also requires a deep understanding of the cataloguing process that produced the original knowledge. Despite both source and target models shared the FRBR philosophy, extensions introduced in the original Corago model to support the documenting of the specific domain, might not fit completely within the FRBRoo design. Since, the extension of FRBRoo through Corago SM aimed to fully reflect the cataloguing principles and analyticity of the source without altering its semantics. When selecting the class definition from the reference ontologies to classify an entity, we tried to achieve the highest accuracy in terms of precision without forcing contents into too stringent classifications while avoiding loss of adherence to them.[11]

[11] We refer here to the specialization in the Corago SM ontology of crm:P14_carried_out_by property by defining two sub-properties: corago:CP2_carried_out_role and corago:CP3_carried_out_actor; the first of which is declared as a sub property of crm:P14_carried_out_by to allow its interpretation also according to the reference ontology.

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

Even if the conceptual alignment into RDF is mainly the result of a static analysis and an initial phase of the process, this has proved not to be free from maintenance during the implementation and even during the whole operational lifecycle. This requires maintaining the semantic alignment of the two models after the dataset publication as LOD. In addition, maintenance may also be influenced by software evolution induced by users' feedbacks or new requirements.

4. Modelling: towards an ontology driven navigation paradigm

The architecture described above, considers the triples belonging from the ABox and the TBox as both expressing domain knowledge. Besides, any application should rely on a self-consistent and coherent semantic infrastructure. Modelling activities must then produce a coherent and homogeneous set of classes that will constitute the core of knowledge base. To define the semantic alignment between Corago and FRBRoo types we proceeded splitting the relational model into thematic areas. Each area has a core entity (i.e. work, performance, etc.) represented by a table which is the source for instances of the corresponding RDF class (i.e. frbroo:F1_Work, frbroo:F31_Perfomance, etc.). In choosing target classes we considered, not only the semantics by their definition, but also the relative position that they would have in the final model resulting from the composition of all the reference ontologies of the KB.
This led to preferring a CIDOC CRM class, although there was a suitable corresponding FRBRoo specialization, opting instead for its generalized version. We tried to exploit the expressiveness of both ontologies to cover as much as possible of the semantic halo of the original Corago entities. Where they lack the required expressiveness, a new class or property has been introduced into the Corago SM as direct extension of a reference ontologies' ancestor.

4.1 The Corago Semantic Model

The Corago SM has been designed both as an aligned extension of CIDOC CRM and FRBRoo and an implementation of the semantic abstraction layer. Implemented in OWL, it extends the so called "Erlangen" version of CIDOC CRM and FRBRoo. Diagram 3 shows some core classes extending their direct ancestor in CIDOC CRM.

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

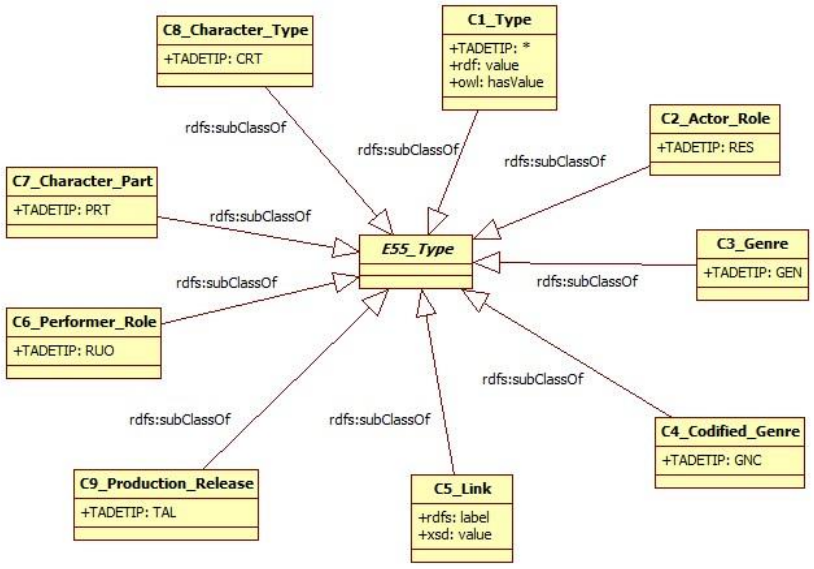

Diagram 3

Table 1 shows a sample of some classes introduced by Corago SM; table 2 illustrates some properties.

Class	Definition	URI
C2 Actor Role	Defines or specifies the role that an Author has had in a specific Activity. Should be used with CP2 and CP3 in order to link an E7_Activity to its E39_Actor with a specific Role. Could be considered the long path of P14_carried_out_by.	http://corago.unibo.it/sm/C2_Actor_Role
C3 Genre	Defines the Genre of the E1_CRM_Entity.	http://corago.unibo.it/sm/C3_Genre
C4 Codified Genre	Defines the Codified Genre of the E1_CRM_Entity.	http://corago.unibo.it/sm/C4_Codified_Genre
C5 Link	Defines the type of link between F1_Work or F2_Expression.	http://corago.unibo.it/sm/C5_Link
C6 Performer Role	Defines or specifies the role with a Character or Performance has to be performed with.	http://corago.unibo.it/sm/C6_Performer_Role

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

C7 Character Part	Defines the kind of part taken by the character in the Performance.	http://corago.unibo.it/sm/C7_Character_Part
C8 Character Type	Describes the stereotyped kind of Character.	http://corago.unibo.it/sm/C8_Character_Type
C9 Production Release	Defines the kind of production release in order of its public presentation as premiere or a reprise.	http://corago.unibo.it/sm/C9_Production_Release

Table 1

Property	**Description**	**URI**
CP2 carried out role	This property defines the relation between an E7 Activity and a C2 Actor Role which is the role an E39 Actor has carried it out with.	http://corago.unibo.it/sm/CP2_carried_out_role
CP3 carried out actor	This property defines the relation between a C2 Actor Role for a specific E7 Activity and the E39 Actor who has carried it out.	http://corago.unibo.it/sm/CP3_carried_out_actor
CP4 links[12]	Represent a direct (untyped) relation between two entity. Equivalent to P67 refers to, could be inferred by the property chain: CP5 refers with o CP6 refers to.	http://corago.unibo.it/sm/CP4_links
CP5 refers with	Relates the domain range entity with the instance of C5 Link which specifies the nature of the relation occurring with the target entity it is referring to through CP6 refers to.	http://corago.unibo.it/sm/CP5_refers_with
CP6 refers to	Relates the C5 Link which specifies the nature of the	http://corago.unibo.it/sm/CP6_refers_to

[12] The property is declared as equivalent to crm:P67_refers_to and inferred basing on the property chain: corago:CP5_refers_with and corago:CP6_refers_to through the class corago:C5_Link.

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

	relation occurring with the range entity it is referring to.	
CP7 performed part	Relates the C2 Actor Role with the part performed by the actor in the role.	http://corago.unibo.it/sm/CP7_performed_part
CP8 performed character	Specifies the Character performed with the Actor Role.	http://corago.unibo.it/sm/CP8_performed_character

Table 2

On top of asserted triples, the Corago SM implements the abstraction layer as a set of specialized families of OWL object properties. The layer is a collection of "semantic lens" formally defined as OWL assertions specifically designed to support users' interactions. We call this set of properties "navigation properties" or "NavProps". Each "NavProp" is defined as a chain of basic properties defined by the two reference ontologies[13]. A "NavProp" is intended to represent complex relationships between resources, often expressed by a deep graph counting several RDF triples, in a user-friendly way. These properties are classified following their main purpose: navigation, searching and georeferencing.

"Navigation properties" aim to give an immediate representation of resource's core attributes both simplifying the graph readability and selecting those attributes considered as basic features of each entity type. "Search properties" define those relations that express the basic relations between domain entities and should be considered as main entry points for exploring the KB. Finally, "georeferencing properties" have been introduced to identify property chains that connect a resource to related ones with geographical data; through these we can gather resources related to a specific geometry during space-time searching.

We illustrate here two cases: the date of birth or death of authors and the relation between a character and actors who played it. The diagram below illustrates the first one. Two "navigation" properties: corago:CPC7_actor_birth_date and corago:CPC8_actor_death_date, draw a direct relationship between the classes crm:E39_Actor and crm:E49_Time_Appellation whose rdfs:label contains the useful value for the property to be shown to the end user.

[13] For each NavProp the corresponding property path is defined as an owl:propertyChainAxiom. The formal definition of equivalence between the two predicates allows a direct translation into SPARQL as a property path predicate.

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

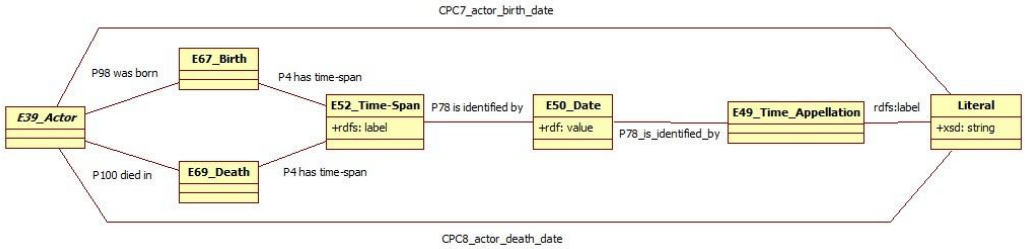

Diagram 4

Diagram 5 represents the property corago:CPC5_played_by that identifies the actor who played a specific part in a performance.

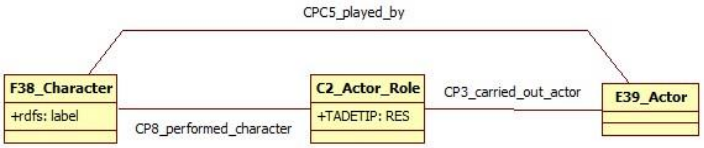

Diagram 5

These properties inherit the cardinality resulting from de aggregation of the triples of the underlying RDF graph and expose a semantics that summarizes those of their constituents. Through these, the domain expert can model the way each attribute of a class is being presented to users, trimming and specializing the analytical description prescribed by reference ontologies. Likewise, "search properties" specify search modalities selecting the most significant relationships between entities within the domain from a functional point of view.
Some examples of "search properties" defined as sub property (rdfs:subPropertyOf) of the ancestor corago:CPA1_searchable follows (table 3).

Concept	Property	Label	Property chain
F1_Work concepted by E39_Actor	CPC1_concepted_by	Concepted by	^R16_initiated / CP2_carried_out_role /CP3_carried_out_actor
F1_Work performed at F9_Place	CPC2_performed_at	Played at	R9_is_realised_in / ^R25_performed / P7_took_place_at
F1_Work performed in F31_Performance	CPC3_performed_in	Played in	R9_is_realised_in / ^R25_performed

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

F1_Work has character F38_Character	CPC4_has_character	Contains the character	P129_is_about
F38_Character played by E39_Actor	CPC5_played_by	Played by	^CP8_performed_character / CP3_carried_out_actor

Table 3

Navigation properties are defined as sub property of their ancestor corago:CPA2_navigable.

Concept	**Property**	**Label**	**Property chain**
E39_Actor was born in	CPC7_actor_birth_date	Date of birth	P98i_was_born / P4_has_time-span / P78_is_identified_by / P78_is_identified_by
F1_Work concepted by	CPC9_work_concepted_by	Conceived by	R16i_was_initiated_by / CP2_carried_out_role / CP3_carried_out_actor
P31_Performance took place at	CPC10_performance_place	Performace place	P7_took_place_at / P1_is_identified_by
F25_Performance_Plan author	CPC13_performance_plan_author	Author	^R17_created / CP2_carried_out_role / CP3_carried_out_actor
F3_Manifestation_Product_Type publisher	CPC21_published_by	Published by	frbroo:CLR6_should_carry/^ecrm:P94_has_created/ ecrm:P14_carried_out_by/ecrm:P131_is_identified_by

Table 4

Since these properties are defined as OWL object properties, they can be annotated to further specify their semantics. For instance, setting the domain and range for each property allows to restrict their usage. When dereferencing a resource, only those NavProps having it as domain class, will be retrieved from the graph and presented to users. This results in a set of semantically shaped and formally defined "views" for each class within the KB. Moreover, through the annotation mechanism, other auxiliary features can be specified. We can define the relevance of a property among all properties of a class. For instance, the title of an opera is more relevant than its first performance location is stated with the CPAN1_priority annotation property. This design allows to identify core attributes required to provide users with the "proper representation" for each entity type. In other words, the abstraction layer is designed to decouple the functional description of a resource from its analytical definition in RDF. An application will then use the layer to select contents to build the

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

dereferencing view of a resource. The user then sees KB's knowledge through a domain specific semantic lens designed to maximise accessibility and comprehension.
A similar approach was followed to give a "semantic retrieval context" for search results. When using full-text searches, we expect to get results that are both coherent with criteria and relevant entry point for start the navigation. Full-text capabilities of KBs normally relies on the basic pattern triple entity – rdfs:label – "some text" that states "the label for the resource". But relevant literal data for a given resource may be attributes stated by much more articulated graphs (i.e. the place of birth of a crm:E39_Author instance). In this case, matched literal will be many triples away from its "main" resource (i.e. our author) and the relation between the two could not be even self-evident to the user. If the system includes the "main" resource as result for the given match, then it has to be "motivated". For this purpose, we introduced a special subset of "search properties" to specify relevant literals for each core entity of our domain (works, events and documents). Only those literals, for whom a "search property" has been defined, are then included into the full-text index. We call this kind of properties "full-text breadcrumbs" as they have two main objectives: the first is to define the subset of literal data a resource is searchable by; the second is to specify the semantics of the relation linking matched text and the resource the system is picking up as result. For each "searchable" resource, the index will contain a set of qualified relations to significant literal attributes.
Diagram 6 depicts the relationship between the three different types of properties from the abstraction layer and their functional profile. Given "Venezia" as textual search parameter, the full-text index retrieves a set of resources matching the query parameter and their relation to results.

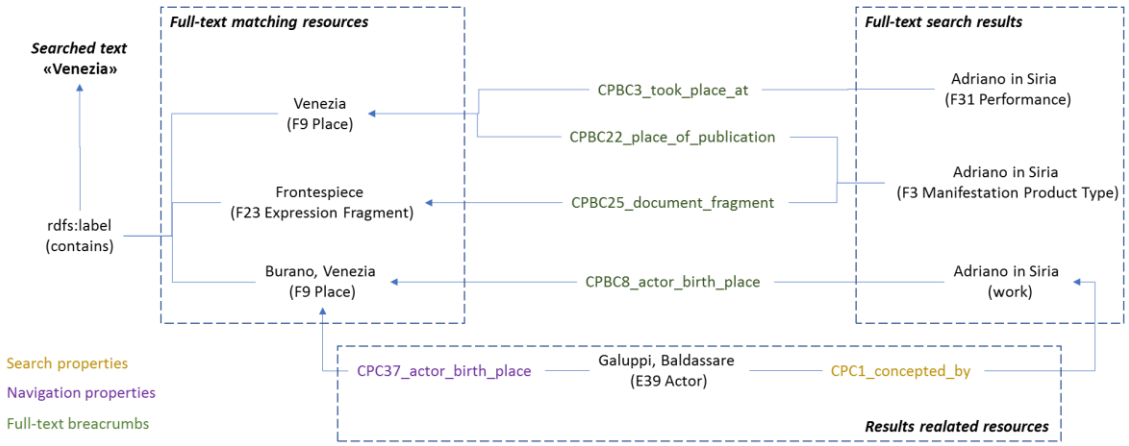

Diagram 6

This novel design provides two major benefits: through the Corago SM a domain expert can specify the subset of core attributes for each class and a "semantic retrieval context" for each result of a full-text search. The design also satisfies both a domain-oriented representation of data and the inclusion of this specific know-how as part of the KB itself. Our approach aimed to a sustainable balance between a full exploitation of SW technologies and LD principles for data interoperability while maintaining the control over data shaping to satisfy requirements of a specific cultural domain.
A similar approach has been followed in order to introduce space and time as search dimensions. The user can then investigate the co-occurrence of relationships between repertoire, chronology and

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

documentation. Moreover, cartography as visual metaphor, is an intuitive tool to quickly and accurately draft the geometry of the area of interest when searching and a synthetic and intuitive view of the output. Thanks to a simple alignment between toponyms, places and theatres from LinkedGeodata[14] through the rdfs:seeAlso statement, space-related resources were georeferenced. For instance, a performance is georeferenced by the property chain: crm:P7_took_place_at / crm:P1_is_identified_by / rdfs:seeAlso that relates it to the corresponding city in LinkedGeodata's dataset. Similar properties are defined within Corago SM as sub-properties of the ancestor property corago:CPA3_georeferenceable for all those classes with spatial references (i.e. scene settings).
Then a cartography-based view has been introduced to support querying ad results visualization. Also, a single timeline for the whole KB is obtained through the normalization of all time-related references. Search properties have been introduced in the navigation layer to let users query the KB by the two dimensions of space and time.
Then the user can choose to view results in a standard tabular way or to project them on a map. The map is connected to a timeline. The two views are synchronized in order to show the sequence of events occurring over the timeline on the map (see image 1).

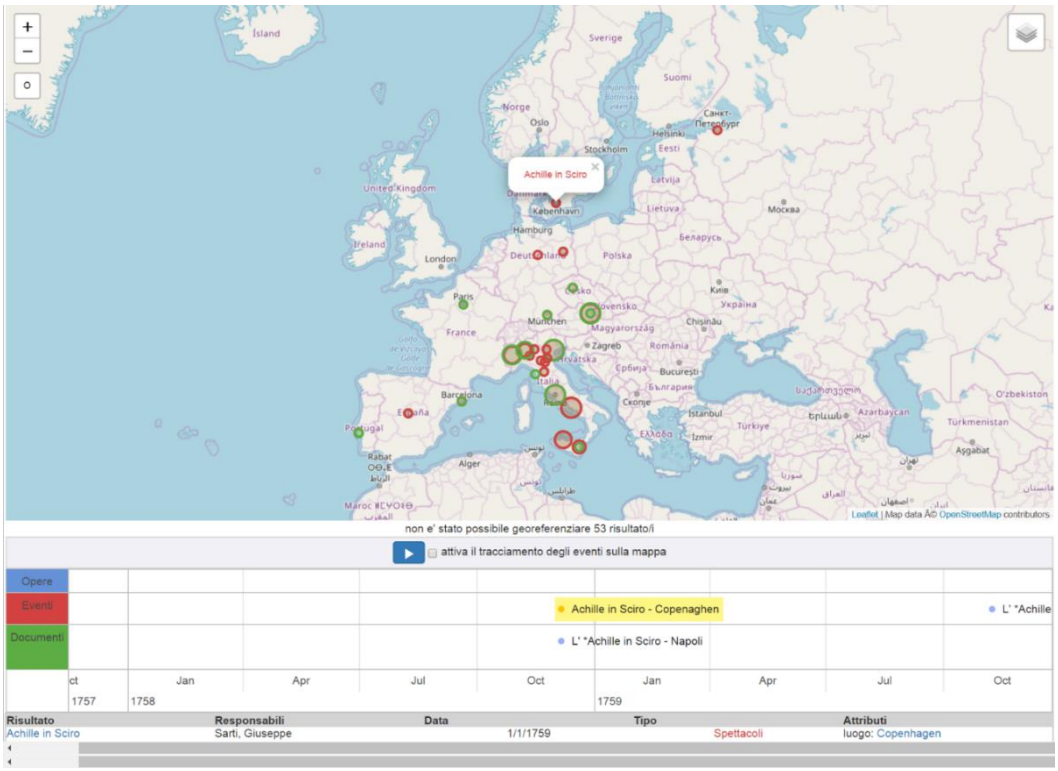

Image 1. The space-time viewer

The adoption of a time-synchronized cartographic viewer is not a novelty in CH domain.[15] But the full integration with search functions and the resulting dynamic representation of every KB's resource

[14] http://linkedgeodata.org.
[15] Good examples can be found at: https://knightlab.northwestern.edu/.

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

offer a plastic vision of events, highly effective and innovative for the representation of historical phenomena of opera.
We proposed here a new approach to formalize competence on a specific domain knowledge such as history of opera and how to put in a user-friendly way. The resulting conceptual model, implemented as a formal ontology, drives the application in rendering views to end-users. The experimental system was expected then to fulfil user requirements while fully supporting LOD interoperability.

5. Experimentation and results evaluation

The experimental system,[16] implemented as a Java based web application, has been evaluated with a survey submitted to different categories of users.
Four user profiles have been considered: domain experts (ESP), archivists/librarians (ARC), opera enthusiasts (MEL)[17] and generic users (GEN). The survey asked then panel to appraise the functionality and performances of the system compared to the legacy implementation[18] and evaluate the effectiveness of novelties introduced through the novel semantic approach. The panel consisted of eight members: 2 ESPs, 2 ARCs, 1 MEL, 2 GENs. The survey was anonymous, the panel was briefed about the questionnaire, the methodology and the systems being evaluated. Members could fill the form during following days and return it anonymous: only the type of user has been collected. No time limits were placed on the trial, experimentation was carried out independently by users in order to bring the experimenter's experience closer the real.
The questionnaire was arranged in comparative terms. The aim was to measure overall satisfaction of the new Corago LOD system compared to the traditional application. Where the functional coverage between the traditional and the new system overlaps, the questionnaire aims to detect satisfaction in comparative terms. New features are evaluated in absolute terms. The questionnaire was divided into three sections. The first focuses on the searching functionalities, the second on the perceived usability and finally an open question asks suggestions or complains about the experimental system. Each section consisted of five questions. Each question allowed a rating range from 1 to 5: 1 being "Low", "Poor" or "Unsatisfactory" and 5 "Completely", "Excellent" or "Very easy".[19]
Results are collected within a matrix, the scores of each question aggregated by profile. From the profile/score matrix the average score was calculated. The general average score for a question was then calculated as the averages per profile. For comparative questions, a matrix was calculated for each of the two systems.

[16] The experimental system is available at: http://corago.unibo.it/lod.
[17] The corresponding Italian term is "*melomane*", since the acronym "MEL".
[18] The traditional web application is available at: http://corago.unibo.it.
[19] The form and results are published through FigShare: https://doi.org/10.6084/m9.figshare.9466505.

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

Diagram 7 shows a direct comparation of results obtained by the two systems.

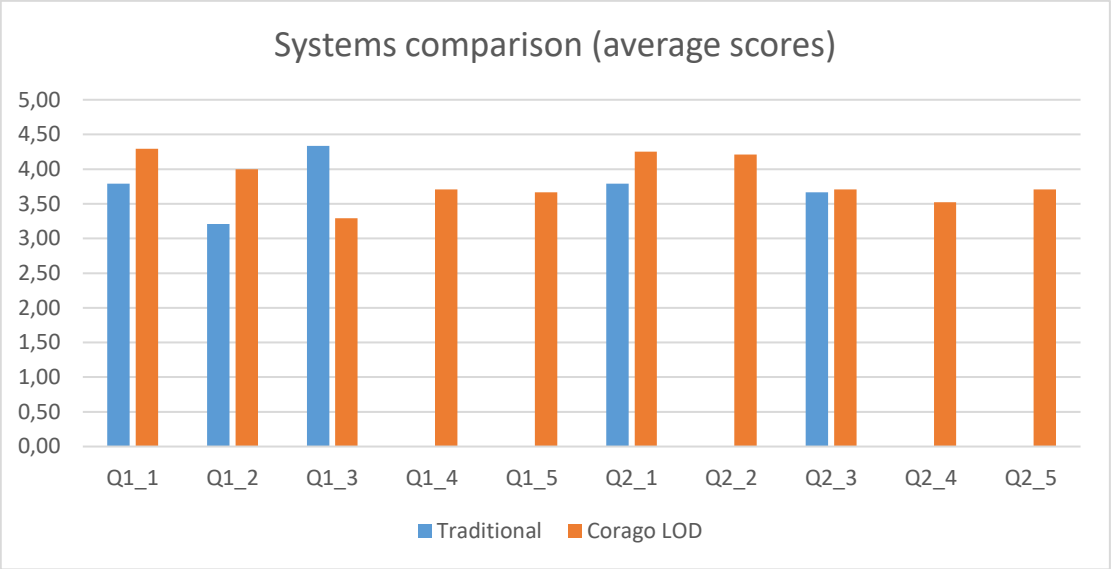

Diagram 7

The first comparative question (Q1_1) evaluated the consistency of results with search criteria. A slight prevalence of the new system (average score 4.29) compared to the traditional one (average score 3.79) emerged. Non-professional prized the newer system while archivists show a slight preference for the traditional one (average difference of 0.50). Overall, the two systems can provide results that are considered consistent with search criteria.

The second comparative question (Q1_2) addressed the effectiveness in information retrieval. A clear prevalence of the new system emerged: Corago LOD obtained an average score of 4.0 against 3.21 obtained by the traditional one. The preference for the new approach is expressed unanimously by all types of users with discards ranging from 0.50 for domain experts, up to 1.0 for archivists and music enthusiasts.

The third question (Q1_3) compared the performance of the two systems. It shows a gain in favour of the traditional system with an average score of 4.33 compared to 3.29 obtained by the experimental system. Therefore, whether the performance perceived by users is said to be "good" for the first, for the second they are only "sufficient". The legacy system, based on consolidated relational technologies, is still the benchmark while developing applications based on RDF graphs.

Question Q1_4 measured the capability to refine results using the ontology-based system. Overall, users found relatively "easy" to refine searches. However, domain experts marked as not "easy", others consider it more than "easy" to do: the average score assigned by GEN and MEL is 4.33 and 5.0 respectively. This could be explained by the much more intuitive approach of the semantic system resulting more effective to users with less prior knowledge about the domain.

The last question of part one (Q1_5) assessed the substantial approval of the space-time search function introduced by the new system with an average score of 3.67 with the professional users (ESP and ARC) rating it 4.0. This result underlines the effectiveness of the use of space and time as dimensions to query the knowledge base.

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

The second part of the survey investigated the readability of result sets (Q2_1). The data presentation of the Corago LOD was judged "very" readable with an average score of 4.25 compared to a score of 3.79 obtained by the traditional system. While expert users rate the two systems as equivalent, other users prefer the newer with an average plus of 0.60 points.
The next question (Q2_2) confirms the slight prevalence of the new application when evaluating the relevance of results according to search criteria. It is appreciated by non-professional users with an average score of 4.67. This seems to confirm that systems based on semantic models let users to better identify relation between results and search criteria, a feature that we called "semantic retrieval context".
Question Q2_2 asked users about the readability of RDF resource's data from system's views. Results showed a clear polarization between the two macro-categories of users. If the average result is comparable between the two systems: score of 3.67 for the traditional one and 3.71 for the newer, professional users preferred the first (with an average of 4.0) over the second (average ratings of 2.75). Non-professional users, on the other hand, rated the experimental system very positively (average ratings of 4.67), considering the traditional system to be "enough" readable (with an average of 3.33).
A similar polarization resulted from question Q2_4 which investigated the capability to navigate seamlessly through the graph offered by the experimental system. Non-professionals consider this as "very" useful (average score of 4.67); professionals consider it "not very useful", assigning an average score of 2.38. We can read the result as a different browsing approach: while the non-professional user explores KB's contents following relationships between resources; the professional one seems to value more timely access to the information sought.
Question Q2_5 evaluated the effectiveness of data representation through cartography. A substantial consensus emerged among the panel with an average score of 3.71. This is an innovative way to browse information within the musicological domain. This could justify the difference between the evaluation as "quite" useful by domain experts and a more positive evaluations of the other types of users who assigned an average of 3.94 corresponding to a judgment of "very" useful.
The third section of the questionnaire was meant to collect open feedback from users. It has been added for two main purposes: to motivate and extend the judgment expressed in the previous two closed-ended sections and to suggest improvements and new features. A summary of those considered to be the most significant of both types follows. Users N1_ARC and N6_ESP stressed the performance and usability of the new system. Both note the need to further refine the pages layout. N1_ARC and N3_ARC see the need to improve tools to refine search as faceted browsing. N2_MEL and N4_GEN ask to make the navigation logics between the different pages more evident. N2_MEL, N3_ARC and N4_GEN suggest improving the organization of contents and layout of the home page.
The survey, albeit circumscribed and based on a predominantly qualitative evaluation, shows a substantial effectiveness of the experimental system. The novel knowledge's RDF format seems able to fulfil user requirements in terms at least equivalent to the traditional solution. Results of questions Q 1_1, Q 1_2 and Q 1_4 seem to confirm that, in terms of effectiveness in searching, the new system can match the traditional approach. At the same time, elements of greater flexibility (question Q 2_4) and exhaustiveness (question Q 2_2) are clearly introduced. Performance issues (Q1_3) and content representation (Q2_3) need further development, new features deriving from the LOD paradigm led to a quite positive feedback (Q1_5 and Q2_5). Finally, feedbacks from section 3 suggest the interest

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

in further refinement of the experimental system to get the required maturity in order to replace the traditional solution.

6. Conclusions

The experimentation has demonstrated that the adoption of semantic technologies and LD principles allows the development of applications challenging systems based on traditional relational conceptual models. Moreover, it opens newer methodological perspectives for CH data management, at least in the peculiar domain of history of opera. The Corago KB is the result of the transposition of the Corago repository through the adoption of two reference ontologies as the CIDOC CRM and FRBRoo. We introduced a multi-layered semantic model to lead the end user through the knowledge graph. This led to an "ontology driven" navigation pattern, where the domain expert is entitled to guide the end user experiencing the domain. Defining both resource's representation and search criteria, the Corago SM drives the experimental system's data presentation layer. This results in a semantically driven user experience. Qualitative results confirmed both a similar expressiveness as traditional systems and the capability to introduce effective novel functionalities. The proposed approach frees the user from knowing the descriptive model behind the application while producing semantically coherent interpretation of data. We also expect that this should increase serendipity without reducing precision of information retrieval. Also, the adoption of standard ontologies to assert knowledge guarantees large LD interoperability. Besides, all the extensions and specializations, as well as the fruition best practices defined by domain experts, were formalized becoming a valuable part of the knowledge offered to the community of stakeholders. We expect that this approach could be shared with similar KBs in the performing arts domain, enabling entering the LOD arena with common reference ontologies while maintaining their peculiarities from a functional point of view.

References

Alexiev, Vladimir. 2012. "Types and Annotations for CIDOC CRM Properties". In Berlin-Heidelberg: Springer Verlag.

Bartolini, Ilaria, Cecilia Luzzi, M Patella, e Angelo Pompilio. 2012. "MIMESIS: A Semantic-Based Approach for a Digital Library Of". In *Poetry in Music. in Proceedings ECLAP 2012: Conference on Information Technologies for Performing Arts. Media Access and Entertainment: Florence 7-9*. Firenze University Press.

Bellini, Pierfrancesco, e Paolo Nesi. 2014. *Modeling Performing Arts Metadata and Relationships in Content Service for Institutions. Multimedia Systems*. Berlin-Heidelberg: Springer. https://doi.org/10.1007/s00530-014-0366-0.

Bianconi, Lorenzo, Angelo Pompilio, e Giorgio Pagannone. 2004. *RADAMES: prototipo d'un repertorio e archivio digitale per il melodramma. in Il Saggiatore Musicale. anno XI*. Olschki. Firenze.

Bizer, Christian, Tom Heath, e Tim Berners-Lee. 2011. "Linked Data: The Story so Far". In *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, 205–27. IGI Global.

JLIS.it 12, 2 (May 2021)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12699

Bonora, Paolo, a c. di. 2010. "Il computer in sartoria". In *Il quadrimensionale instabile*. ed. Bignami P. e Ossicini C. Strumenti del Dams. UTET Università.

Bonora, Paolo, Charlotte Ossicini, e Giuseppe Raffa. 2006. "From Relational Metadata Standards to CRM Ontology: a Case Study in Performing Arts". In *CIDOC2006 Proceedings*. Gothenburg.

Byrne, Kate. 2008. *Relational Database to RDF Translation in the Cultural Heritage Domain*.

Ceriani, Davide. 2019. "Internet Culturale: Catalogues and Digital Collections of Italian Libraries". *Nineteenth-Century Music Review* 16(01):161–71. https://doi.org/10.1017/S1479409817000805.

Cipollone, Barbara. 2009. "La catalogazione dei libretti nel sistema di archiviazione digitale Il Corago". *Quaderni Estensi Rivista* 1.

Doerr, Martin. 2009. "Ontologies for cultural heritage". In *Ontologies for Cultural Heritage*, Handbook on Ontologies: 463–86. Berlin-Heidelberg: Springer Verlag.

Doty, Colin. 2013. "The Difficulty of An Ontology of Live Performance. InterActions". *UCLA Journal of Education and Information Studies* 9 (1).

Kobilarov, Georgi, Tom Scott, Yves Raimond, Silver Oliver, Chris Sizemore, Michael Smethurst, Christian Bizer, e Robert Lee. 2009. "Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to Make Connections". In *The Semantic Web: Research and Applications*, ed. Lora Aroyo, Paolo Traverso, Fabio Ciravegna, Philipp Cimiano, Tom Heath, Eero Hyvönen, Riichiro Mizoguchi, Eyal Oren, Marta Sabou, e Elena Simperl, 723–37. Berlin-Heidelberg: Springer.

Le Boef, Patrick. 2012. "Towards Performing Arts Information As Linked Data?" In *SIBMAS 2012 Conference: Best Practice Innovative Techniques for Performing Arts Collections. Libraries and Museums*. France.

Pattuelli, M. Cristina, Alexandra Provo, e Hilary Thorsen. 2015. "Ontology Building for Linked Open Data: A Pragmatic Perspective". *Journal of Library Metadata* 15(3–4):265–94. https://doi.org/10.1080/19386389.2015.1099979.

Pompilio, Angelo, Lorenzo Bianconi, Fabio Regazzi, e Paolo Bonora. 2005. "RADAMES: A New Management Approach to Opera: Repertory. Archives and Related Documents". In *Proceedings - First International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution*. IEEE. https://doi.org/10.1109/AXMEDIS.2005.42.

Raimond, Yves, e Mark Sandler. 2008. "A Web of Musical Information". In *ISMIR 2008, 9th International Conference on Music Information Retrieval*. Lulu.