RePIM in LOD: semantic technologies to preserve knowledge about Italian secular music and lyric poetry from the 16th-17th centuries

Paolo Bonora[1] and Angelo Pompilio[2]

[1]Department of Classical Philology and Italian Studies, University of Bologna, Italy – paolo.bonora@unibo.it

[2] Department of Cultural Heritage, University of Bologna, Italy – angelo.pompilio@unibo.it

ABSTRACT

The *RePIM in LOD* project aimed to publish the "*Repertorio della Poesia Italiana in Musica, 1500-1700*" (RePIM) as Linked Open Data (LOD) dataset. For the extent and detail of its contents, RePIM is a reference archive for research on Italian secular music from the 16th-17th centuries. In recent years, scholars have been able to access it through a public web-based application. Due to the obsolescence of its information technology platform, the RePIM repository was set to be taken offline. To preserve this precious source, the project migrated its contents into a knowledge base (KB) adopting semantic technologies and designed an up-to-date end-user application. The paper illustrates this process as a use case of digital knowledge preservation of bibliographic and philological information in the field of Italian secular music and lyric poetry of the 16th-17th centuries.

KEYWORDS

Linked Open Data, Semantic Web, CIDOC-CRM, FRBRoo, digital knowledge preservation, Italian poetry.

TALK

1. INTRODUCTION

The "Repertorio della Poesia Italiana in Musica, 1500-1700" (RePIM), is the digital version of the incipit index of Italian poetry in music, a project conceived at the end of the 70s by Lorenzo Bianconi in collaboration with Angelo Pompilio and Antonio Vassalli. Since 1990, materials collected on paper cards during the 70s-80s, as well as bibliographic descriptions of poetic and musical sources, have been revised and widely integrated. As a rule, musical sources of secular music do not declare authorship of the intoned literary text. To identify the poet, it is necessary to collate the musical sources with coeval poetry books. The RePIM incipit index was designed to respond to this basic scientific need: to identify the greatest possible number of authors of poetic texts set to music in secular and spiritual compositions of the 16th-17th centuries through the systematic examination of poetry books. A minor subset of the archive was published online during in 2006 but it has never been fully published in any other analogic or digital form. Unfortunately, the technological platform has become obsolete since then and now must be discontinued. The *RePIM in LOD* project identified the publication as a Linked Open Data (LOD) dataset as the way to preserve it as a free and unlimited resource for the community of present and future scholars. Contents have been migrated in Resource Description Format (RDF) adopting two reference ontologies such as FRBR Object Oriented (now Library Reference Model) and CIDOC Conceptual Reference Model. The result is a KB relying on a set of open and well-established formal ontologies and then fully interoperable within the Linked Data ecosystem. Finally, to develop a new web-based application, the ResearchSpace platform ([5]) has been selected as an open source, cultural heritage oriented and scholar friendly tool which also aims to address digital knowledge preservation and maintainability over time.

2. FALLING IN LOD

RePIM starts from the consideration that, in the secular vocal music of the 16th-17th centuries, for the purposes of historical and critical understanding of the work, the lyric component has a structural and artistic equal to the musical one. Thanks to the systematic collection of textual incipits from both musical and poetry books from Italian and international libraries, about 30% of the musical compositions of the collected works have been identified.

The conceptual model of the repository has been designed adopting the Functional Requirements for Bibliographic Records (FRBR) guidelines. The consequent distinction between Work and Manifestation levels enables bibliographic,

musicological and literary analytical search. This gives scholars the opportunity to explore literary and musical production from a single perspective.

The repository is structured into three main sections: the repertoire of works, the bibliographic sources and the incipit index. The repertoire consists of the authority file of names, literary and musical works and the links identified between them. Musical and literary sources are described in an analytical form both in terms of bibliographic description and contents. Textual incipit, complete text (diplomatic and modern edition), title, dedication, caption, voices and articulation in parts or sections are available. The sources are also linked to their reprints or newer editions. In addition, bibliographic references used during the study of sources are recorded in the KB together with a complete authority file of owning libraries and a repository of digital copies, if available.

The current consistency is: a bibliography of literary sources of 1.500 titles; the incipit of over 66.000 musical compositions; a complete bibliography of musical books of secular and spiritual music from the 16th-17th centuries (ca. 3.500 items), references to bibliographic musical repertories (RISM A/I e B/I; Vogel-Einstein 1962 and Nuovo Vogel by Lesure and Sartori, 1977); an authority file with ca. 1.400 musicians and 3.200 poets; digital reproduction of approx. half of the recorded musical sources and literary books and full text transcriptions of more than 5.000 poetry works are also available.

For literary sources, the record contains: the complete transcription of the title page; the essential data about the dedicatory letter; the list of authors; the incipit (first two lines, or more) of the composition; any further information of a musical and literary nature such as: the complete transcription, incipits of texts marked as "*per musica*", references and citations relating to musical occasions, etc.

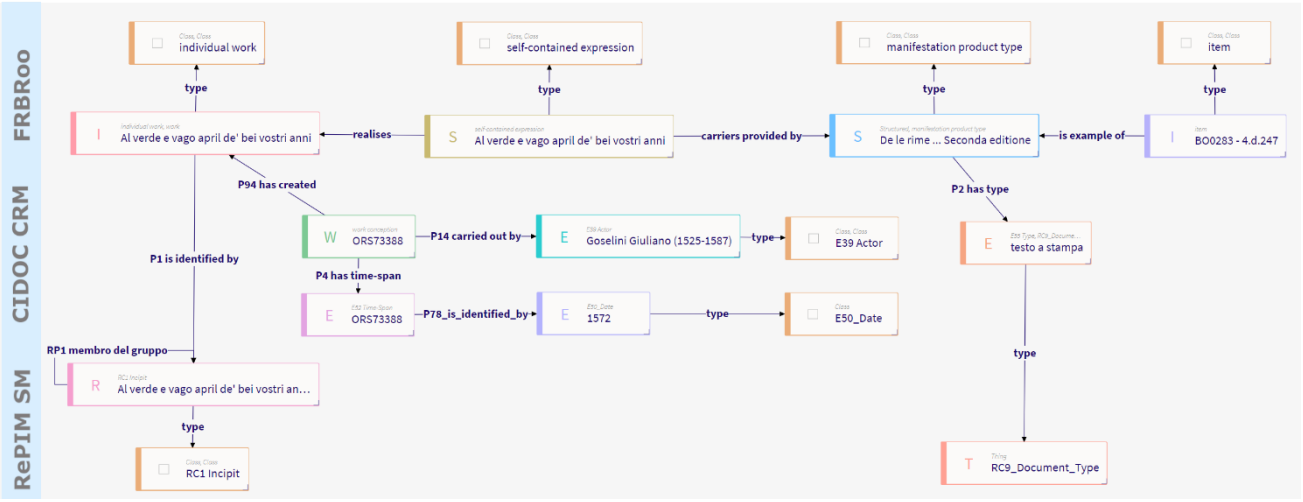

Figure 1. Part of the graph representing the work "Al verde e vago April de'bei vostri anni" in "Rime" by Paolo Gottardo Pontio, 1572.

In order to represent this structured knowledge, we adopted a set of formal ontologies integrated within the RePIM Semantic Model (RePIM SM)[1]. This introduces some domain specific concepts (such as the "incipit" as the key to identify works) and several specializations of properties from those defined by the two reference ontologies where a more specific semantics was required. Thus, the role of the RePIM SM is to extend the CIDOC CRM and FRBRoo to better represent domain-specific concepts while maintaining full alignment with them. This makes the dataset fully interoperable as data can be accessed through the standard CIDOC CRM or through the specialization proposed by the RePIM SM only when required. To express information about the document segmentation and bibliographic references, we adopted some other specialized ontologies such as DoCO[2] and BiRO[3]. The model relies on FRBRoo's concepts of work, expression and manifestation to describe musical and poetry works, their contents and books (fig.1). Starting from this core assertions, the other ontologies assert information about authors, primary and secondary sources, references, libraries and some domain specific topics as genres, voices, settings, etc.

The resulting KB's conceptual model consists of 30 classes (12 from CIDOC CRM, 11 from FRBRoo and 4 from DoCO, 2 from Repim SM and 1 from BiRO) and 49 properties (17 from CIDOC CRM, 5 from FRBRoo, 19 from Repim SM and others from DoCO, BiRO, DCTerms and RDFs).

[1] The RePIM SM is maintained at: https://github.com/paolobonora/RePIM-LOD/.
[2] DoCO: Document Components Ontology ([3]), http://www.sparontologies.net/ontologies/doco.
[3] BiRO: Bibliographic Reference Ontology ([4]), http://www.sparontologies.net/ontologies/biro.

Data were extracted from the legacy database management system with a Relational-to-RDF mapping tool following the semantic alignment between the FRBR-ER inspired proprietary relational model and the RePIM SM ([1]). The resulting knowledge graph contains 778.699 class instances and 5.324.973 property assertions (more than 5.5 million triples in total). We expect that the publication of RePIM's contents as a LOD resource through Open Science platforms as Zenodo[4] would lower the barriers of data access, facilitating both the reuse and extension of the KB by a wider community. We also expect that this should increase the resilience against the technological unavoidable obsolescence of any information system leveraging dissemination of knowledge among different projects and players.

To reach the wider community of users not directly interested in raw data reuse, we also developed a new web-based application that exploits potential of the RDF dataset. The ResearchSpace[5] platform has been selected to implement basic functionalities plus a set of requirements expressed over the years by the community of users but never fully implemented. The new application leverages both semantic and NLP technologies and introduces textual search and indexing features such as: full-text search, PoS and lemma based searching capabilities; graph-based browsing; NER (Named Entity Recognition) of names, toponyms, etc.; faceted browsing of topics[6].

The design of the web application followed a strictly functional approach trying to achieve an effective representation of knowledge to the end-user. This means reducing the complexity of the conceptual model with views that increase data readability. Although the ResearchSpace's templating mechanism allows to arrange an effective representation of resources as user's views, it does not support their explicit formalization within the semantic model as the introduction of a dedicated semantic layer would allow to ([2]). Besides the framework supports basic mechanism for expert-led refinements and expansions of the knowledge graph. This will be leveraged to introduce edit and annotation features into the web application.

3. FURTHER DEVELOPMENTS

The current implementation aims to support basic search and browsing functionalities of the current knowledge expressed within the KB. The next phase should aim to let user annotate further findings achieved through the exploitation of already available sources or newly added ones through the contribution by the community of users. The ResearchSpace platform supports the implementation of data editing procedures as well as creation of semantic narratives and visual representation of resources from the knowledge graph. This would lead to a collaborative approach to knowledge extension and maintenance. This will require a fine-grained information provenance tracing and a sharp profiling of contributors. Besides, the outcome would be a shared effort of knowledge preservation and extension among the wider community of users through the years to come. We believe that publishing the RePIM archive as LOD will preserve it and offers, both to music scholars and those interested in Renaissance and Baroque poetry, a useful tool and a valuable source of information, resulting from over 40 years of research, which otherwise would be lost in the digital mist.

4. ACKNOWLEDGEMENTS

The *RePIM in LOD* project was founded by I Tatti – The Harvard University Center For Italian Renaissance Studies with a Digital Humanities Fellowship grant.

REFERENCES

[1] Bizer, Christian, and Andy Seaborne. 2004. 'D2RQ-Treating Non-RDF Databases as Virtual RDF Graphs'. In *Proceedings of the 3rd International Semantic Web Conference (ISWC2004)*.

[2] Bonora, Paolo, and Angelo Pompilio. 2021. 'Corago in LOD. The Debut of an Opera Repository into the Linked Data Arena'. *JLIS.It* 12 (2): 54–72.

[3] Constantin, Alexandru, Silvio Peroni, Steve Pettifer, and et al. 2016. 'The Document Components Ontology (DoCO)'. *Semantic Web* 7 (2): 167–81.

[4] Di Iorio, Angelo, Andrea Giovanni Nuzzolese, Silvio Peroni, and et al. 2014. 'Describing Bibliographic References in RDF'. In *CEUR Workshop Proceedings*, 1155.

[5] Oldman, Dominic, and Tanase Diana. 2018. 'Reshaping the Knowledge Graph by Connecting Researchers, Data and Practices in ResearchSpace'. In *The Semantic Web – ISWC 2018, a Cura Di Denny Vrandečić, Kalina Bontcheva, Mari Carmen Suárez-Figueroa, Valentina Presutti, Irene Celino, Marta Sabou, Lucie-Aimée Kaffee, e Elena Simperl*, 325–40. Cham: Springer International Publishing.

[4] The dataset is available through the DOI: 10.5281/zenodo.5692109.

[5] https://researchspace.org/.

[6] The provisional URL for the web application is: https://repim.itatti.harvard.edu/.