HAL
open science

Exploring Linked Classical Music Catalogs with OVERTURE

Pasquale Lisena, Manel Achichi, Eva Fernandez, Konstantin Todorov, Raphaël Troncy

To cite this version:

Pasquale Lisena, Manel Achichi, Eva Fernandez, Konstantin Todorov, Raphaël Troncy. Exploring Linked Classical Music Catalogs with OVERTURE. ISWC: International Semantic Web Conference, Oct 2016, Kobe, Japan. hal-01988237

HAL Id: hal-01988237

https://hal.science/hal-01988237v1

Submitted on 21 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exploring Linked Classical Music Catalogs with OVERTURE

Pasquale Lisena[1], Manel Achichi[2], Eva Fernández[1], Konstantin Todorov[2], Raphaël Troncy[1]

[1]EURECOM, Sophia Antipolis, France
[2]University of Montpellier, France

Abstract. In this paper, we introduce OVERTURE—a web application allowing to explore the interlinked catalogs of major music libraries including the French National Library, Radio France and the Philharmonie de Paris. We have first developed the DOREMUS ontology which is an extension of the well-known FRBRoo model for describing works and expressions as well as the creation process. We have implemented a so-called MARC2RDF tool allowing for the conversion and linking of bibliographical entries about music works, interpretations and expressions from their original MARC-format to RDF following this DOREMUS ontology. We present an exploratory search engine prototype that enables to browse through the reconciled collection of bibliographical records of classical music and to highlight the various interpretations of a work, its derivative, its performance casting as well as other rich metadata.

1 From Music Metadata to Linked RDF Graphs

The DOREMUS project[1] develops methods to describe, publish, connect and contextualize music catalogs on the web of data. Our focus is on the description of classical and traditional musical works as well as their interpretations (events) [1]. We rely therefore on the FRBRoo model, which combines the strength of the FRBR practices in distinguishing a work from its interpretation and derivatives and the CIDOC-CRM model that emphasizes the role of events for modeling the process of creation of creative works. We extend FRBRoo by adding classes and properties specific to the music domain (e.g. for representing the intended casting for interpreting a work). The research challenges we are tackling are related to the population of this rich ontological model from raw library metadata and on its visualization that should enable users to search for classical music but also see influencers and derivative works, or the many interpretations of a famous work.

Data Conversion. The data collected from the BnF and the Philharmonie de Paris (PP) describing music works are represented in the MARC format (UNI–

[1] `http://www.doremus.org/`

and INTERMARC variants). A MARC file is a succession of fields, each carrying a 3-digit label, and subfields, delimited by the $ symbol (Fig. 1). The semantics of these fields and subfields is described in documents issued by The International Federation of Library Associations and Institutions (IFLA)[2]. Note that a subfield tag can change its meaning depending on the field, under which it is found. In our example (Fig. 1), the $a tag under the field 500 stands for the musical genre (sonata), whereas the same tag under the field 700 stands for the composer name (Beethoven). We have developed an open source prototype that allows for the automatic conversion of UNI- or INTERMARC bibliographic records to RDF[3], implementing the DOREMUS model [2]. The conversion process relies on explicit expert-defined transfer rules (or mappings) that indicate where in the MARC file to look for what kind of information, providing the corresponding property path in the model as well as useful examples that illustrate each transfer rule, as shown in Fig. 2. We have used the DOREMUS properties to name the extracted relations (e.g., `mus:U12_has_genre`[4] is the property describing the genre of a work). Resources are identified by URIs that use the corresponding DOREMUS class labels in their names (as for example `http://data/doremus.org/expression/UUID` identifying an instance of the FRBRoo class `Expression`). Often, the RDF data that we extract will contain

```
001FRBNF428405510000000
50011$313908188$aSonates$rPiano$sOp. 27, no 2$uDo dièse mineur
700 |$313891295$aBeethoven$bLudwig van$f1770-1827$4230
```

Fig. 1. An excerpt of a UNIMARC record.

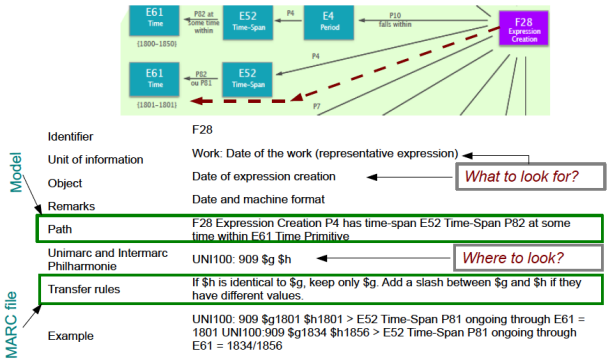

Fig. 2. Example of mapping rules describing the timespan of the composition of a work.

string literals, such as genres (*sonata*) or instruments (*piano*). These values are

[2] `http://www.ifla.org/publications/ifla-series-on-bibliographic-control-36`.
[3] `https://github.com/DOREMUS-ANR/marc2rdf`
[4] The `mus` prefix refers to `http://data.doremus.org/ontology/`

controlled by specific vocabularies, such as MIMO[5] describing musical instruments. Therefore, our data conversion methods include an automatic mapping of string literals to URIs coming from controlled vocabularies. Inspired by the Datalift platform [5], we have developed a generic `string2uri` component enabling to lookup strings in a SKOS controlled vocabulary and to propose best candidate match.

Among the several existing MARC to RDF conversion tools, such as the Catamandu toolkit[6] or MARCMODS2RDF[7], the one that is closest in spirit to ours is Marimba[8], developed in the context of the Linked Data at the BNE (Biblioteca Nacional de España) project. Just like our converter, and in contrast to Catmandu and MARCMODS2RDF, Marimba is based on the IFLA ontologies, including FRBR, and is designed by the help of librarians. However, Marimba focuses on literature library entries, while we implement the DOREMUS model, which is better suited to describe music.

Data Heterogeneities and Linking. Due to the high data heterogeneity, link discovery in the musical field becomes a challenging task. To train and test data linking tools, we have collected benchmark data from the BnF and the PP, now part of the OAEI instance matching evaluation campagne[9]. Our tests with SILK [3] and LIMES [4], achieving an average F-measure of around 55%, confirm that new methods, specific to the music field, need to be proposed, in order to handle specific heterogeneity types of these datasets. In particular, special attention has to be payed to multilingual descriptions of works, as well as to significant lexical differences in work titles. Descriptive heterogeneities (level of detail, number of properties) also hinder the performance of general purpose tools. Furthermore, certain properties, although mapped correctly according to the transfer rules, contain highly heterogeneous information, as for example `cidoc-crm:P3_has_note` – a field of free text in the form of comments. We are currently developing a linking approach that combines expert heuristics with indexing techniques that is able to propose a solution according to the heterogeneity type manifested by the data.

2 OVERTURE: an Exploratory Search Engine for Music

OVERTURE (Ontology-driVen Exploration and Recommendation of mUsical REcords) is a modern web application that relies on a Node.JS server sitting in front of a Virtuoso triple store. The web application performs SPARQL queries to access the data via the Node.JS server that could be a Linked Data Fragment server. The user interface is realized with the AngularJS2 library. At the top, a

[5] `http://www.mimo-international.com/vocabulary.html`
[6] `http://librecat.org`
[7] `http://simile.mit.edu/repository/RDFizers/marcmods2rdf/`
[8] `http://marimba4lib.com`
[9] `http://islab.di.unimi.it/content/im_oaei/2016/#doremus`

navigation bar directs the user towards one of the main concepts in the DOREMUS model (Work, Expression, Performance, Recording, Score, Person), that become distinct sections of the application. Inside each section, the user can filter the list of records shown in a class-specific and advanced search form. In this way, (s)he can search a performance by choosing a particular location, a performer, or an expression by selecting the musical key, the genre or by searching by title. From these parameters, the server generates the SPARQL query that matches the user information need.

```
SELECT ?expressions ?title
WHERE {
   ?expressions a frbroo:F22_Self-Contained_Expression ;
     cidoc:P102_has_title ?title .
     mus:U12_has_genre <http://data.doremus.org/vocabulary/genre/si>
}
```

Clicking on a result, more detailed information is shown such as the information about the composer, the casting, a longer descriptive note or different expressions. In the future, this selection will contain related expressions, selected through a recommendation algorithm. The application is available at `http://overture.doremus.org`[10].

Acknowledgments

This work has been partially supported by the French National Research Agency (ANR) within the DOREMUS Project, under grant number ANR-14-CE24-0020.

References

1. M. Achichi, R. Bailly, C. Cecconi, M. Destandau, K. Todorov, and R. Troncy. DOREMUS: Doing Reusable Musical Data. In *14th International Semantic Web Conference (ISWC)*, 2015.
2. P. Choffé and F. Leresche. DOREMUS: Connecting Sources, Enriching Catalogues and User Experience. In *24th IFLA World Library and Information Congress*, 2016.
3. A. Jentzsch, R. Isele, and C. Bizer. Silk: generating RDF links while publishing or consuming linked data. In *9th International Semantic Web Conference (ISWC)*, 2010.
4. A. N. Ngomo and S. Auer. LIMES: A Time-Efficient Approach for Large-Scale Link Discovery on the Web of Data. In *22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2312–2317, 2011.
5. F. Scharffe, G. Atemezing, R. Troncy, F. Gandon, S. Villata, B. Bucher, F. Hamdi, L. Bihanic, G. Kpklian, F. Cotton, J. Euzenat, Z. Fan, P.-Y. Vandenbussche, and B. Vatant. Enabling linked-data publication with the Datalift Platform. In *(AAAI) Workshop on Semantic Cities*, pages 25–30, 2012.

[10] A video is also available at `https://www.dropbox.com/s/4e7n5a4qgyzk6wu/OvertureDemo-v0.mp4?dl=0`