

# OVERTURE

Manel Achichi<sup>1</sup>, Pasquale Lisena<sup>2</sup>, Eva Fernández<sup>2</sup>, Wafa Bouneb<sup>2</sup>,  
Konstantin Todorov<sup>1</sup>, Raphaël Troncy<sup>2</sup>

<sup>1</sup>University of Montpellier, France

<sup>2</sup>EURECOM, Sophia Antipolis, France

**Abstract.** In this paper, we introduce OVERTURE—an application allowing to explore the catalogs of major music bibliographic agencies, including the French National Library, Radio France and the Philharmonie de Paris. The application is based on the *marc2rdf* prototype, developed for this purpose and allowing for the conversion of bibliographical entries about music works, interpretations and expressions from their original MARC-format to RDF, following the DOREMUS model, an extension of the well-known FRBRoo model.

## 1 Introduction and Motivation

- The Doremus project: talk about the model ?? and the controlled vocabularies ?? because we need this info in next section
- the input data,
- why it is important to convert these data to RDF (impact, number of institutions using this format).

## 2 Converting Bibliographic Data to DOREMUS RDF

The data collected from the BnF and the Philharmonie de Paris describing music works are given in the MARC format, and more precisely, its variants UNIMARC and INTERMARC, respectively. A MARC file is a succession of fields, each carrying a label (a three-digit number). Each field, in turn, is a succession of subfields, delimited by the \$ symbol, as shown in Fig. ???. The semantics of the fields and subfields is described in various documents issued by The International Federation of Library Associations and Institutions (IFLA)<sup>1</sup>, according to the MARC variant. Note that a subfield tag can change its meaning depending on the field in which it is found. For instance, in the example in Fig. ??, the combination of the field 500 and the \$a tag stands for the musical genre (sonata in this case), 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 bibliographic records given in UNI- or INTERMARC to RDF<sup>2</sup>.

<sup>1</sup> <http://www.ifla.org/publications/ifla-series-on-bibliographic-control-36>.

<sup>2</sup> <https://github.com/DOREMUS-ANR/marc2rdf>

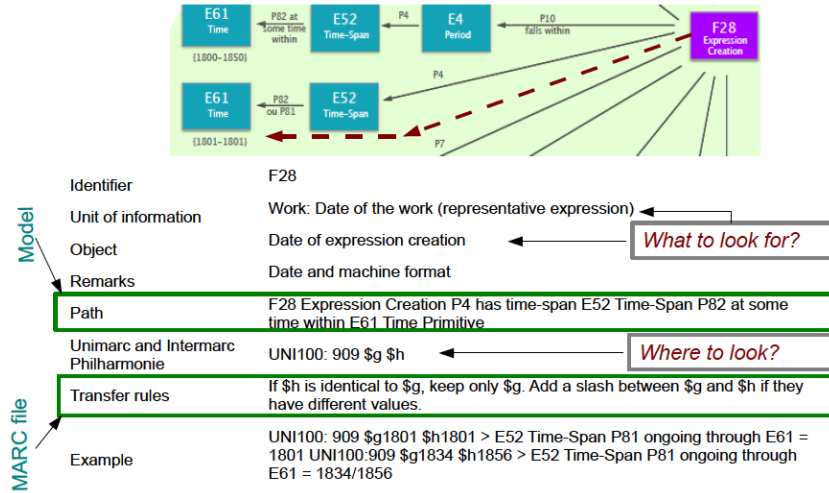
```

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.

One of the particularities of this converter is that it implements the DOREMUS model. 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 its corresponding property path in the model as well as useful examples that illustrate each transfer rule, as shown in Fig. ?? . The mapping rules reflect the practices of each institution and therefore a mapping table per institution has been provided by the experts. We have used the DOREMUS properties to name the extracted relations (as for example `mus:U12_has_genre`, using the “mus” prefix for `http://data.deremus.org/ontology/`, labeling the property describing the genre of a music work). The 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 in a unique manner an instance of the DOREMUS classe ”Expression”).



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

Often the RDF data that we extract will contain string literal values, such as genres (sonata), instruments (piano), tonalities (D-major) and other. As introduced above, these values are controlled by specific vocabulaires, such as the

MIMO vocabulary ?? describing musical instruments. Therefore, our data conversion methods include an automatic mapping of string literals to URIs coming from controlled vocabularies. We have used the string2uri mapper, relying on various string matching techniques ??.

An example of a converted MARC record is given in Fig. ??.

– Interconnexion: based on Silk

### 3 OVERTURE: an Exploratory Search Engine for Music

An interface screenshot(s), a link to a demo video + download

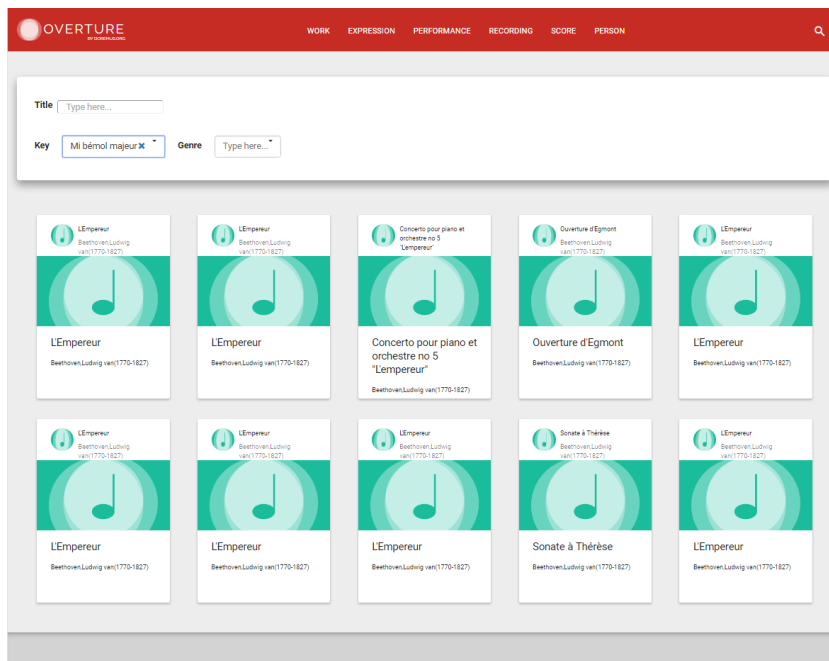
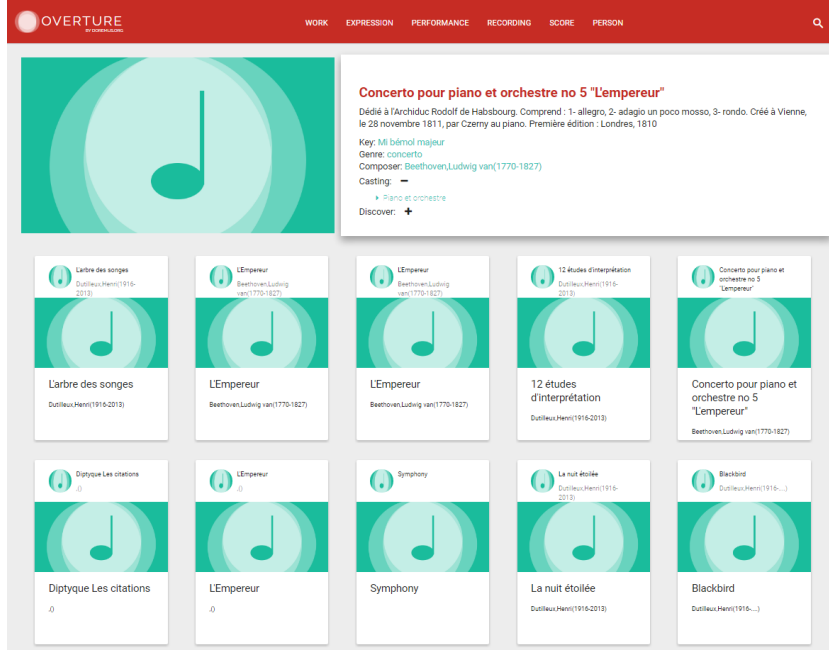


Fig. 3. Overture interface: the search form and results.

### 4 Conclusion and Future Work

We are currently working on the development of connectors for music works, allowing for the automatic interlinking of equivalent resources across datasets of different bibliographical agencies. Our first test with the SILK [] and LINES [] tools confirm our initial hypothesis that new methods, specific to the music field, need to be proposed, able to handle the high heterogeneity of these datasets. In



**Fig. 4.** Overture interface: the Expression.

particular, our first tests show that special attention has to be paid to multi-lingual descriptions of works, as well as to significant differences in lexical descriptions of music titles. Descriptive heterogeneities (level of detail, amount of information available) also appear to be hard to handle by the existing general purpose off-the-shelf linking tools. We plan to combine expert heuristics with an automatic heterogeneity centered data linking approach. Finally, the developed connectors will be integrated to the OVERTURE application, allowing for the automatic navigation from one dataset to another bringing to the user the richness of several connected datasets.

## 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.