

DOREMUS: un graphe d'œuvres musicales interconnectées

Manel Achichi,¹ Pasquale Lisena,² Konstantin Todorov,¹ Raphaël Troncy,²
et Jean Delahousse³

¹ LIRMM, University of Montpellier, CNRS, France
{achichi, todorov}@lirmm.fr

² EURECOM, Sophia Antipolis, France
{pasquale.lisena, raphael.troncy}@eurecom.fr

³ OUROUK, Paris, France
delahousse.jean@gmail.com

Mots-clés : Ontologie, humanités numériques, interconnexion des données, jeu de données.

Résumé de [Achichi *et al.* (2018)], publié à ISWC 2018.

Trois grandes institutions culturelles françaises, la Bibliothèque Nationale de France, Radio France et la Philharmonie de Paris, ont uni leurs efforts à ceux d'universitaires en informatique et en sciences sociales dans le cadre du projet DOREMUS¹. L'objectif principal était de développer des méthodes et outils pour décrire sémantiquement leurs catalogues d'œuvres et d'interprétations de musique classique et traditionnelle et de les ouvrir à la communauté Web en suivant les principes des données liées.

Une contribution majeure du projet est le développement de l'ontologie DOREMUS [Choffé & Leresche (2016)] qui étend les modèles CIDOC-CRM [Doerr (2003)] et FRBRoo [Doerr *et al.* (2008)] décrivant l'information bibliographique, en l'adaptant au domaine de la musique, comblant ainsi un important écart de représentation. Ainsi, nous avons défini un certain nombre de nouvelles classes et propriétés permettant de décrire les aspects d'une œuvre liés à la musique, tels que la clé musicale, le genre, le tempo, l'instrumentation tout en respectant les principes de modélisation de FRBRoo. Plusieurs vocabulaires contrôlés définissant des concepts propres à la musique (comme les genres musicaux ou les clés) ont été recueillis ou élaborés, interconnectés et publiés en utilisant le langage SKOS.

La construction du graphe de connaissances DOREMUS implique la mise en place d'un workflow qui inclut la conversion des données originales en RDF en suivant l'ontologie DOREMUS et l'interconnexion de ces données. Les données des catalogues des trois institutions partenaires sont disponibles au format MARC² ou XML. Des outils pour la conversion des données en RDF selon le modèle DOREMUS ont donc été développés : ils sont génériques comme `marc2rdf` (valable pour tous les formats MARC) ou ad-hoc et utilisés pour une sérialisation XML spécifique. Ce processus aboutit à la construction de plusieurs graphes de connaissances sur les œuvres et événements musicaux (un pour chaque source de données). Ces graphes ont été interconnectés à l'aide de *Legato*, un outil de liaison de données capable de gérer l'hétérogénéité des méta-données musicales. *Legato* implémente un algorithme basé sur la sélection et le classement des clefs afin d'améliorer considérablement la précision au niveau des liens générés. À des fins d'évaluation, un jeu de données de référence a été créé manuellement par les experts des bibliothèques et communiqué à la communauté du Web Sémantique dans le cadre de l'initiative OAEI³. Le processus de fusion de données aboutit à la construction d'un graphe pivot de référence des œuvres musicales communes aux trois institutions.

Enfin, un moteur de recherche exploratoire nommé *Overture*⁴ a été développé. L'application fait directement des requêtes au point d'accès SPARQL et fournit des informations issues du graphe de

1. <http://www.doremus.org>

2. <https://www.loc.gov/marc/>

3. <http://oaei.ontologymatching.org/>

4. <http://overture.doremus.org/>

connaissances dans une interface utilisateur web. *Overture* propose également un système de recommandation basé sur la similarité entre artistes et entre œuvres en s'appuyant sur des plongements calculés sur le graphe DOREMUS.

Le jeu de données DOREMUS complet ainsi que les vocabulaires contrôlés et quelques exemples de requêtes sont disponibles à <http://data.doremus.org>. Le tableau 1 récapitule le nombre d'entités pour les classes les plus représentatives et fournit des détails sur la présence d'informations spécifiques.

classe	BnF	Philharmonie	Itéma3	total
Oeuvre	135,818	9,005	8,319	153,142
- avec l'instrumentation	123,219	4,621	0	127,840
- avec la clé musicale	19,645	1,973	0	21,618
- avec le genre musical	128,497	3,820	8,071	140,388
- avec le compositeur	133,371	7,741	8,231	149,343
- avec la date de composition	91,566	5,712	4,856	102,134
- avec le numéro de catalogue	20,796	2,908	0	23,704
- avec le numéro d'ordre	11,598	1,612	0	13,210
- avec le numéro d'opus	21,836	1,985	0	23,821
Interprétation	15,784	784	1,531	18,099
- avec plus d'une oeuvre interprétée	0	713	1,277	1,990
Piste	0	6,538	18,273	24,811

TABLE 1 – Graphe de connaissances DOREMUS

Remerciements. Ce travail a été partiellement financé par l'Agence Nationale de la Recherche (ANR) dans le cadre du projet DOREMUS (ANR-14-CE24-0020).

Références

- ACHICHI M., LISENA P., TODOROV K., TRONCY R. & DELAHOUSSE J. (2018). DOREMUS : A Graph of Linked Musical Works. In *17th International Semantic Web Conference (ISWC)*, Monterey, USA.
- CHOFFÉ P. & LERESCHE F. (2016). DOREMUS : Connecting Sources, Enriching Catalogues and User Experience. In *24th IFLA World Library and Information Congress*, Columbus, OH, USA.
- DOERR M. (2003). The CIDOC Conceptual Reference Module : An Ontological Approach to Semantic Interoperability of Metadata. *AI magazine*, **24**(3), 75–75.
- DOERR M., BEKIARI C. & LEBŒUF P. (2008). FRBRoo : a conceptual model for performing arts. In *CIDOC Annual Conference*, p. 6–18.