

Modélisation des métadonnées des correspondances de Pablo Picasso

Ibtissam Slalmi*, Solène Labbé**, Pauline Gérôme***, Jérôme Darmont*, Fatiha Idmhand**, Anne-Emmanuelle Kahn***, Sabine Loudcher*, Alexandre Quiquerez***, Juliette Pozzo****, Charlotte Quiblier****

*Univ Lyon, Univ Lyon 2, UR ERIC – **Université de Poitiers, UMR ITEM

Univ Lyon, Univ Lyon 2, UR TRANSVERSALES – *Musée national Picasso de Paris

Avec près de 50 000 œuvres sur l'ensemble de sa carrière, Pablo Picasso est considéré comme l'un des artistes les plus productifs et l'un des plus influents du XX^e siècle. Si le talent de l'artiste ne fait aucun doute, savons-nous réellement quelles ont été les étapes de la construction de cette renommée planétaire ? Qui a contribué à la transnationalisation de Pablo Picasso ? Quels sont les acteurs et les œuvres qui ont contribué à cette renommée planétaire ? L'histoire de l'art, les expositions et la sociologie de l'art ont permis d'ouvrir de nombreux champs de recherche liés à la trajectoire exceptionnelle de Pablo Picasso. Rares demeurent cependant les études quantitatives portant sur ces sources historiques et notamment sur ses archives personnelles. L'objectif du projet PICASSOLETTERS¹, financé par la Maison des Sciences de l'Homme Lyon Saint-Étienne, est de contribuer à combler ce manque à travers une étude originale de sa correspondance pour permettre de cartographier le réseau de Pablo Picasso, d'en identifier les membres et de repérer celles et ceux qui ont agi pour promouvoir et diffuser ses œuvres à l'aide de méthodes et d'outils informatiques. Le corpus épistolaire que ce projet explore est une partie du courrier envoyé par des contemporains de Pablo Picasso, connus ou non, ainsi que des sociétés, à Pablo Picasso. Il est composé de plus de 20 000 lettres qui ont été adressées à Pablo Picasso par 4200 correspondants et qui n'ont jamais été analysées de façon globale et détaillées à ce jour ; le projet PICASSOLETTERS entend contribuer à combler deux manques : un premier concerne le déficit d'études quantitatives sur le peintre, un second concerne l'absence de base de métadonnées accessibles en ligne et ouvertes à tous et toutes. Les métadonnées au cœur du projet PICASSOLETTERS sont issues de trois sources principales : celles recueillies par le Musée national Picasso de Paris, par les chercheurs en études hispaniques du projet et celles qui sont disponibles sur le Web sémantique.

Pour atteindre les objectifs du projet, nous avons constitué trois séries de métadonnées pour décrire les correspondants, les correspondances (ainsi que les personnes ou organismes cités dans les correspondances) et les œuvres. Les métadonnées incluent par exemple les noms des correspondants, leurs rôles, dates de naissance et de décès. Les métadonnées des correspondances comprennent les adresses et dates d'expédition, les personnes ou les œuvres mentionnées. Celles des œuvres incluent les titres, types d'œuvres (peinture, sculpture...), dates d'achèvement, etc. Par ailleurs, les différentes sources de données engendrent une hétérogénéité importante, notamment en termes de format, de type (fichiers CSV, XSLX, XML, HTML,

1. <https://picassoltrs.hypotheses.org>

RDF, JSON ou Turtle), de dénomination, de codage ou encore de vocabulaire utilisé. Face à cette hétérogénéité, nous avons réalisé un dictionnaire pour recenser les métadonnées disponibles, les redondances, les contradictions ou les complémentarités. Le projet PICASSOLETTERS étant en phase de démarrage, nous avons décidé de construire dans un premier temps une base de métadonnées relationnelle pour permettre rapidement aux chercheurs du projet de saisir, stocker, interroger et analyser les métadonnées. Notre modèle conceptuel UML des métadonnées est organisé en trois ensembles de classes et regroupe en un seul modèle les trois séries de métadonnées citées précédemment. À ce modèle est associé un vocabulaire contrôlé pour décrire les métadonnées, les homogénéiser et alimenter la base de données. Le modèle conceptuel est traduit en un modèle logique relationnel, puis au niveau physique dans le système de gestion de bases de données MariaDB². La base est stockée sur les serveurs de la TGIR Huma-Num³. Pour l'alimenter, nous avons mis en œuvre un travail conséquent de pré-traitement, de nettoyage, de restructuration, de normalisation, de renommage et d'indexation à l'aide de scripts en Python dédiés.

L'objet du poster associé à cet article est de présenter les métadonnées du projet ainsi que leur modélisation conceptuelle. À court-terme, la mise en place de la base de métadonnées doit s'accompagner d'une interface graphique conviviale pour permettre, dans un premier temps, aux membres du projet de saisir, interroger et analyser les métadonnées afin de construire le réseau de relations de Pablo Picasso. Le projet a également pour ambition d'enrichir les métadonnées actuelles par des métadonnées provenant de référentiels du Web des données (Bibliothèque nationale de France, Musée Getty de Los Angeles, Wikidata⁴) tout en traitant les problèmes de cohérence, de redondance et d'hétérogénéité que cela va poser à nouveau. La constitution d'une base de métadonnées des correspondances de l'artiste pose également de nombreuses questions juridiques. Il en va ainsi notamment en droit d'auteur, en droit *sui generis* des bases de données, en droit des données personnelles et en droit pénal (secret des correspondances, atteinte au système de traitement automatisé de données), droit des contrats (conditions générales d'utilisation). Une analyse inédite de la technique de *Web scrapping* devra aussi être menée. Le caractère ouvert des données devra être pris en compte dans cette analyse juridique. Enfin, le projet prévoit la mise en place d'un lac de données dont les métadonnées décrites dans le poster vont être la colonne vertébrale. Les métadonnées vont en effet être interfacées avec des données non structurées : « scans » des correspondances adressées à Pablo Picasso, textes intégraux et résumés, croquis et possiblement des photos des œuvres du peintre qui sont mentionnées dans ses correspondances. L'objectif est de pouvoir requêter tous ces types de données et leurs liaisons depuis une interface unique.

Summary

This article describes the context of the PICASSOLETTERS project, funded by the Maison des Sciences de l'Homme Lyon Saint-Étienne. This project aims at modeling and enriching the metadata of Picasso's correspondences. We design and implement a metadata model for the epistolary corpus in three main series of metadata: correspondences, authors and artworks.

2. <https://mariadb.org>

3. <https://www.huma-num.fr>

4. <https://www.wikidata.org>