

---

# Modelling the History of Medieval and Renaissance Manuscripts for the Mapping Manuscript Migrations Portal

Toby Burrows\*<sup>1</sup>, Mikko Koho<sup>2</sup>, Jouni Tuominen<sup>2</sup>, Eero Hyvönen<sup>2</sup>, Kevin Page<sup>1</sup>, David Lewis<sup>1</sup>, Doug Emery<sup>3</sup>, Hanno Wijsman<sup>4</sup>, Lynn Ransom<sup>3</sup>, and Emma Thomson<sup>3</sup>

<sup>1</sup>Oxford e-Research Center – United Kingdom

<sup>2</sup>Aalto University, Semantic Computing Research Group – Finland

<sup>3</sup>University of Pennsylvania, Schoenberg Institute for Manuscript Studies – United States

<sup>4</sup>Institut de recherche et d'histoire des textes – Centre National de la Recherche Scientifique : UPR841 – France

## Abstract

The Mapping Manuscript Migrations (MMM) project, funded from 2017 to 2020 by the Trans-Atlantic Platform's Digging into Data Challenge, implements a Linked Open Data (LOD) framework to aggregate major datasets relating to the history and provenance of medieval and Renaissance manuscripts. Its main product is the MMM Portal, launched in January 2020, which enables browsing and searching across the histories of more than 216,000 manuscripts, as well as providing map-based visualizations of the data. (1)

The Data Model developed for the MMM project combines elements from the CIDOC-CRM and FRBROO ontologies, together with some additional entity classes and relationships specific to MMM. A "modelling group" within the project worked for more than twelve months to analyse and compare the data models of three source datasets. Two of these (Bibale and the Schoenberg Database of Manuscripts) are relational databases, while the third (the Bodleian Library's Medieval Manuscripts in Oxford Libraries) is a collection of TEI-encoded XML documents. (2) Each of these sources includes references to major LOD vocabularies, which can be used to reconcile the data.

The MMM Data Model focuses on five entity classes: Manuscripts, Agents, Works, Places, and Events. (3) Provenance events provide the main modelling challenge, but the other types of entities raise their own issues. Places are reconciled using the Getty Thesaurus of Geographic Names (TGN), which provides a good basis for hierarchical reasoning when exploring the data geographically. But manuscript production is frequently ascribed to a region rather than a specific place, which causes problems when regions are not included in the TGN hierarchy (as is the case with Northern Italy). There are also obvious problems with medieval and early modern kingdoms and provinces which cannot be directly mapped to a current area (such as the Kingdom of Castile and Aragon).

Agents, in comparison, have been reconciled against the Virtual International Authority File (VIAF) wherever possible. They include medieval and Renaissance authors as well as more recent collectors and dealers, not all of whom are listed in VIAF. A recurrent problem

---

\*Speaker

arises when a dealer and their firm share the same name: is H. P. Kraus a person or an organization? Another issue relates to religious houses (e.g., Fountains Abbey) and feudal groupings (e.g., the Earldom of Chester): are they places or organizations? The data sources are inconsistent in their treatment of these questions.

The three source datasets take different approaches to modelling the provenance events connected with the history of manuscripts. In Bibale, the basic record is for a manuscript, containing links to persons, organizations, collections, and places relevant to the production and ownership of that manuscript over the centuries. In the Schoenberg Database, the basic record is for an observation of a manuscript in a catalogue, with the names of former owners listed in chronological sequence. For the Oxford catalogue, the basic record describes a manuscript; history and provenance information is given in a free-text narrative field, with personal names encoded as entities.

Provenance events are at the heart of the MMM Data Model, but the source data do not usually support a specific typology of different events beyond a manuscript's production. The Schoenberg Database records a list of owners, with associated dates (where known,) but no types of relationships between the actor and the manuscript are specified, except for a direct transfer to the next owner. The Oxford record usually provides a narrative of the manuscript's history and ownership; sometimes this information is given in the form of a list of owners instead. Personal names are marked up with the attribute role="fmo" for former owners, but there is no wider vocabulary for ownership or provenance events. Bibale is the only data source which offers such a vocabulary, including the following relationships for manuscripts:

*a été prêté par / à*

*a été possédée par*

*a été légué par / à*

and the following for actors:

*a légué*

*transmet*

*est le commanditaire de*

Because of this, the MMM Data Model has had to remain quite generic in its coverage of provenance events, as the entity counts for the "Events" category illustrate:

*E10 Transfer of Custody [25,871]*

Bibale Database [13,465]

Schoenberg Database of Manuscripts [12,406]

*E12 Production [219,761]*

Schoenberg Database of Manuscripts [194,150]

Medieval Manuscripts in Oxford Libraries [13,347]

Bibale Database [12,264]

*Manuscript related activity [658,617]*

Schoenberg Database of Manuscripts [228,167]

Medieval Manuscripts in Oxford Libraries [7,153]

Bibale Database [106]

Rest = unknown

Bibale is the richest source for "Transfer of Custody" statements, while all three datasets provide "Production" events for every manuscript (except seven in the Schoenberg Database and six in Bibale). But for other events both the Schoenberg Database and the Oxford catalogue can only default to generic "manuscript related activity" statements. These use the MMM schema's "ManuscriptActivity" entity class, which acts as a bridge between CIDOC-CRM's E7\_Activity and more specific types of transaction events.

These are supplemented by the "last known location" property, calculated from the "has former or current owner" relationship. This covers all the Oxford manuscripts, more than 80% of the Schoenberg manuscripts, and 50% of the Bibale manuscripts:

*Last known locations [178,497]*

Bibale Database [6,531]

Medieval Manuscripts in Oxford Libraries [13,347]

Schoenberg Database of Manuscripts [158,219]

This level of specificity is sufficient for navigation across places of production, last known locations, and the movement from one to the other over time, and is used as the basis for map-based visualizations. More detailed visualizations of specific steps in the history of individual manuscripts would require a more detailed modelling of different types of events, such as that which the Linked Art initiative is working on (4), or a vocabulary of acquisition similar to that developed by the Art Tracks project. (5) MMM shows that existing data sources are unlikely to have the necessary level of granularity and specificity to support such a detailed approach, but it also demonstrates that this level of detail is not essential for supporting in-depth browsing, searching, and visualizations.

## References

1. Mapping Manuscript Migrations Portal: <https://mappingmanuscriptmigrations.org/>
2. Bibale: <http://bibale.irht.cnrs.fr>; Schoenberg Database of Manuscripts: <https://sdbm.library.upenn.edu>; Medieval Manuscripts in Oxford Libraries: <https://medieval.bodleian.ox.ac.uk>
3. Mapping Manuscript Migrations, "Data Model": [https://github.com/mapping-manuscript-migrations/mapping-manuscript-migrations.github.io/tree/master/data\\_model](https://github.com/mapping-manuscript-migrations/mapping-manuscript-migrations.github.io/tree/master/data_model)
4. Linked Art, "Object Provenance": <https://linked.art/model/provenance/>
5. Art Tracks, "Acquisition Method Vocabulary": [http://www.museumprovenance.org/reference/acquisition\\_method](http://www.museumprovenance.org/reference/acquisition_method)

**Keywords:** Manuscripts, Provenance, Events