Big Data and the Afterlives of Medieval and Renaissance Manuscripts: the "Mapping Manuscript Migrations" Project

Abstract

Tens of thousands of European medieval and Renaissance manuscripts have survived until the present day. As the result of changes of ownership over the centuries, they are now spread all over the world, in collections across Europe, North America, Asia and Australasia. They often feature among the treasures of libraries, museums, galleries, and archives, and they are frequently the focus of exhibitions and events in these institutions. They provide crucial evidence for research in many disciplines, including textual and literary studies, history, cultural heritage, and the fine arts. They are also objects of research in their own right, with disciplines such as paleography and codicology examining the production, distribution, and history of manuscripts, together with the people and institutions who created, used, owned, and collected them.

Over the last twenty years there has been a proliferation of digital data relating to these manuscripts, not just in the form of catalogues, databases, and vocabularies, but also in digital editions and transcriptions and – especially – in digital images of manuscripts. Overall, however, there is a lack of coherent, interoperable infrastructure for the digital data relating to these manuscripts, and the evidence base remains fragmented and scattered across hundreds, if not thousands, of data sources.

The complexity of navigating multiple printed sources to carry out manuscript research has, if anything, been increased by this proliferation of digital sources of data. Large-scale analysis, for both quantitative and qualitative research questions, still requires very time-consuming exploration of numerous disparate sources and resources, including manuscript catalogues and databases of digitized manuscripts, as well as many forms of secondary literature. As a result, most large-scale research questions about medieval and Renaissance manuscripts remain very difficult, if not impossible, to answer.

The "Mapping Manuscript Migrations" project, funded by the Trans-Atlantic Platform under its Digging into Data Challenge for 2017-2019, aims to address these needs. It is led by the University of Oxford, in partnership with the University of Pennnsylvania, Aalto University in Helsinki, and the Institut de recherche et d'histoire des textes in Paris. The project is building a coherent framework to link manuscript data from various disparate sources, with the aim of enabling searchable and browsable semantic access to aggregated evidence about the history of medieval and Renaissance manuscripts.

This framework is being used as the basis for a large-scale analysis of the history and movement of these manuscripts over the centuries. The broad research questions being addressed include: how many manuscripts have survived;

where they are now; and which people and institutions have been involved in their history. More specific research focuses on particular collectors and countries.

The paper will report on the first six months of this project. The topics covered will include the sources of data which are being combined, and the data modeling being carried out to link disparate data sources within a Linked Data environment. Data from four different sources are being combined initially, including three relational databases – the Schoenberg Database of Manuscripts, Bibale, and Medium – and a catalogue built from TEI-encoded documents (Medieval Manuscripts in Oxford Libraries). Transforming these into RDF, mapping them to an ontology derived from CIDOC-CRM and FRBRoo, and linking them to external identifiers from VIAF, Wikidata, and GeoNames, raise various issues and challenges.

The paper will also report on the new digital platform being developed, and how it is being informed by specific research questions derived from an analysis of the literature and from discussions with a focus group of manuscript researchers. Requirements for visualizing and navigating the data will be discussed, drawing on work previously done by a project to analyse the history of the manuscript collection of Sir Thomas Phillipps.

Bibliography

Burrows, Toby (2017) "The History and Provenance of Manuscripts in the Collection of Sir Thomas Phillipps: New Approaches to Digital Representation," *Speculum* 92 S1, S39-S64

Da Rold, Orietta and Marilena Maniaci (2015) "Medieval Manuscript Studies: a European Perspective", in: *Writing Europe, 500-1450: Texts and Contexts* (Essays and Studies, 68), ed. Aidan Conti, Orietta Da Rold & Philip Shaw (Cambridge: D. S. Brewer, 2015), pp. 1-24

Hyvönen, Eero, Jouni Tuominen, Miika Alonen and Eetu Mäkelä (2014) "Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets", in: *The Semantic Web: ESWC 2014 Satellite Events* (Lecture Notes in Computer Science, 8798) (Berlin: Springer-Verlag, 2014), pp. 226-230

Scase, Wendy (2009) *Applying Semantic Web Technologies to Medieval Manuscript Research: European Science Foundation Exploratory Workshop Report*. Strasbourg: European Science Foundation, 2009.