

The History of Cultural Heritage Collections: Exploring New Approaches to Analysis and Visualization

Toby Burrows
King's College London
toby.burrows@kcl.ac.uk

In the nineteenth century, the English collector Sir Thomas Phillipps (1792-1872) assembled the largest private collection of European medieval and early modern manuscripts and documents. It is estimated to have contained more than 40,000 items, making it considerably larger than most of the collections in public institutions today, and included many manuscripts of considerable historical, textual and artistic significance. Their modern locations are spread across the globe – the dispersal of the Phillipps Collection took place gradually over more than one hundred years, and numerous institutions and collectors were involved. As a result, the Phillipps Collection provides a rich and varied set of data for tracking the history and provenance of cultural heritage collections.

In this paper, I will report on a project to reconstruct and analyse the history and provenance of the manuscripts which formed the Phillipps Collection. The scale of the Phillipps Collection has proved a significant challenge to traditional research methods in the past; the English librarian A.N.L. Munby spent more than a decade compiling an overview of Phillipps' collecting activities and of the dispersal of the collection up to the mid-1950s (Munby 1951-1960). In this project I am employing data modeling and analysis techniques to build a digital environment for tracing the entire history of these manuscripts, as far as it can be known. I am interested in mapping the provenance events and ownership networks which, taken together, constitute the history of these thousands of manuscripts over hundreds of years.

My paper will focus on four key technical aspects of the project.

- Frameworks for modeling and representing the data relating to ownership and provenance, using an event-based approach

Events are central to provenance research, but they have proved difficult to represent in existing ontologies and data models, with a variety of different approaches being used. I will discuss some of these – including CIDOC-CRM, the Europeana Data Model, and property graphs.

- Techniques for importing and combining existing data relating to manuscript histories

The existing data relating to the Phillipps manuscripts range from relational databases and MARC records to handwritten notes and card indexes. Capturing and cleaning these data and aligning them to a common data model are complex tasks which require multiple ingestion paths and crosswalks.

- The deployment of suitable software to manage the data and to support analysis and visualization

I will report on two specific platforms: the graph database software Neo4j (Van Bruggen 2014) and the nodegoat data management environment (Van Bree and Kessels 2015).

- Methods for visualizing and analyzing the data produced by the project, and for making them available for re-use by other researchers

I will look at a series of use cases and research questions related to the aggregated data, and will demonstrate how Neo4j and Nodegoat can be used to produce analyses and visualizations in response to these requirements. I will also discuss methods for linking the data produced by this project with the wider Linked Data cloud, in order to enable wider contextualization and analysis.

References

- Munby, A.N.L. (1951-60) *Phillipps Studies*, Cambridge: Cambridge University Press. 5 vols.
- Van Bree, P. and Kessels, G. (2015) "Mapping memory landscapes in nodegoat" in: *Social Informatics*, ed. L.M. Aiello and D. McFarland (Lecture Notes in Computer Science 8852), pp. 274-278, New York: Springer International.
- Van Bruggen, R. (2014) *Learning Neo4j*, Birmingham: Packt Publishing.