**BURROWS — History**

<no images>

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

Burrows, T. N.

Each object in contemporary cultural heritage collections has its own history and its own historical significance, as Neil McGregor demonstrated so vividly with one hundred objects chosen from the collections of the British Museum (McGregor, 2010). An important part of that history is the process by which each object came to reside in its current location and its current collection. Each object has usually been part of a series of collections over its lifetime, and this movement of objects between collections has its own history. Similarly, each collection has its own history of formation and (usually) dispersal. These collections include personal and individual collections, private institutional collections, and modern public collections.

These relationships between cultural objects, collectors, and collections over time are an important example of what Alan Liu has described as ‘network archaeology’ (Liu, 2012) —the recovery and analysis of cultural, social, and artistic relationships at a particular period of time. As well as studying how and why some objects survived while others did not, and how and why the ownership of these objects changed, this ‘network archaeology’ can also address several larger research questions. Cultural collections can reflect broader historical trends and are shaped by them. In the European context, these include the dissolution of religious institutions, the decline of royal and aristocratic patronage, the rise of public cultural institutions (especially museums and libraries), the emergence of wealthy collectors in the industrial era, European global expansion and imperial power, and the repatriation of cultural objects. The network of relationships between people and institutions involved in the ownership and transmission of cultural collections can also reveal a good deal about the more general networks of cultural influence and social and political relationships in a particular society.

In the 19th 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. The manuscripts had very varied geographical origins across Western Europe, are written in various different European languages, and cover a wide range of different subjects and topics. 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 history of the Phillipps Collection provides a much richer and more varied set of data than a single contemporary institutional collection would provide.

In this paper, I will report on a project to reconstruct and analyse the history and provenance of the manuscripts that 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 a 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 innovative 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 that, taken together, constitute the history of these thousands of manuscripts over hundreds of years.

My paper will focus particularly 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 the various alternatives—including CIDOC-CRM and the Europeana Data Model—before presenting my own approach based on property graphs (Blanke et al., 2013).

• *Techniques for importing and combining existing data relating to manuscript histories*

The existing data relating to the Phillipps manuscripts are scattered across numerous digital and physical sources, in multiple languages. They are, inevitably, in a variety of different formats and schemas, ranging from relational databases and MARC records to handwritten notes and card indexes. Capturing these data and aligning them to a common data model are complex tasks that require multiple ingestion paths and crosswalks.

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

Suitable software is critical for a project of this kind. I will report on the options available, and discuss the reasons for my choice of the graph database software Neo4j to store, manage, and present the data (Van Bruggen, 2014). Like Blanke et al. (2013), I consider that graph databases are a good fit for ‘how humanities researchers think about their data and its relationships’. I will also discuss the implementation of the data model for aggregating the provenance data used in this project.

• *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 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. I will compare the results made possible by my software environment and data model with those produced by the Schoenberg Database of Manuscripts—a relational database that focuses on manuscript provenance. The graph database approach enables more sophisticated and complex pattern-matching across the full range of provenance data.

**References**

**Blanke, T., Bryant, M. and Hedges, M.** (2013). Back to Our Data—Experiments with NoSQL Technologies in the Humanities. Paper presented at *2013 IEEE International Conference on Big Data*, Santa Clara, CA, 6–9 October 2013.

**Liu, A.** (2012). Remembering Networks: Agrippa, RoSE, and Network Archaeology. Paper presented at *Network Archaeology Conference*, Miami University, Miami, OH, 21 April 2012.

**McGregor, N.** (2010). *A History of the World in 100 Objects.* Allen Lane, London.

**Munby, A. N. L.** (1951–1960). *Phillipps Studies*, 5 vols. Cambridge University Press, Cambridge.

**Van Bruggen, R.** (2014). *Learning Neo4j*. Packt Publishing, Birmingham, UK.