

# The History and Provenance of Manuscripts in the Collection of Sir Thomas Phillipps: New Approaches to Digital Representation

By Toby Burrows

## THE PHILLIPPS MANUSCRIPT COLLECTION

The manuscript collection of Sir Thomas Phillipps (1792–1872) was one of the largest private collections ever assembled. The illegitimate son of a wealthy Manchester industrialist, Phillipps was able to put together a vast collection of manuscripts, books, paintings, drawings, prints, photographs, and other materials. There were somewhere between forty thousand and sixty thousand manuscripts—bigger than most contemporary public collections.<sup>1</sup>

Phillipps was buying at a time when major private collections, like those of Johann Meerman, Leander van Ess, and Richard Heber, were coming onto the market. These collections, in their turn, had been formed from the dispersal of religious libraries in France and Italy in the late eighteenth and early nineteenth centuries. Phillipps was prepared to outbid institutions like the British Museum in what he said was a deliberate strategy to save vulnerable historical materials.<sup>2</sup> His collection included an extensive number of archival documents, especially relating to British regional and local history.

Negotiations in his lifetime with the British Museum and the Bodleian Library failed to find a home for the collection, and on his death it was inherited by his daughter Katherine Fenwick and her husband. After legal action to overturn Phillipps's will, the Fenwicks and their descendants spent decades gradually selling off the collection.<sup>3</sup> The first stage in this dispersal was a series of auctions at Sotheby's in London between the 1890s and the 1930s, accompanied by direct sales to various European governments and to a small number of private collectors. In 1946, the remainder of the collection was sold to the London booksellers W. H. Robinson Ltd., and the dispersal continued with further auctions at Sotheby's, catalog sales,

This research was carried out with the help of a European Union International Incoming Fellowship (Framework Programme 7, Marie Curie Actions, no. 626696) based in the Department of Digital Humanities at King's College London. I am very grateful to Professor Andrew Prescott and Dr. Peter Stokes for their guidance and support during this project.

<sup>1</sup> A. N. L. Munby, *The Formation of the Phillipps Library from 1841 to 1872*, Phillipps Studies 4 (Cambridge, UK, 1956), 166.

<sup>2</sup> *Ibid.*, 170.

<sup>3</sup> A. N. L. Munby, *The Dispersal of the Phillipps Library*, Phillipps Studies 5 (Cambridge, UK, 1960).

direct sales to specific institutions, and a large donation to the Bodleian Library. In the later 1970s, the so-called residue of the residue of the collection was sold to the New York firm of H. P. Kraus and appeared in several Kraus catalogs. Phillipps manuscripts have subsequently continued to appear in auctions, sale catalogs, and even AbeBooks. As a result of this lengthy process, the Phillipps collection has been scattered across the globe, and his manuscripts are now in numerous different public collections as well as—in some cases—in private hands.

A. N. L. Munby's account of the dispersal of the collection, up to 1956, takes the form of a narrative based mainly on the Phillipps papers in the Bodleian Library. It is a remarkable achievement, which is still of great value to researchers, but it is far from the final word on the Phillipps collection, especially after a further sixty years have elapsed. There have been many further sales and transfers of ownership during that period. More fundamentally, however, Munby does not present the evidence relating to the manuscripts in a systematic way, and his underlying data remain unpublished. They consisted primarily of annotations recorded in an interleaved copy of Phillipps's printed catalog; there are photocopies of this material in both the Cambridge University Library and the Bodleian Library, where the annotations were continued—at least for some years—by staff of those libraries. The original of Munby's annotated catalog is now in a private collection in North America.

A different solution to the problem of gathering the data relating to the history of the Phillipps manuscripts can be seen in the Manuscripts Reading Room of the British Library. It consists of a card index in which each of the many thousands of cards contains provenance information for a single Phillipps manuscript. This index was discontinued some years ago, and there are no plans to publish it or to make it available in digital form. The Institut de recherche et d'histoire des textes in Paris also has an unpublished list of current locations for Phillipps manuscripts, compiled by Edith Brayer.

These resources have obvious limitations: they are available only in a single physical location, they are no longer being updated, and none of them is comprehensive, though their content overlaps. More seriously, the evidence is in a form where it cannot readily be analyzed or used for any but the most specific research questions.

The potential for using new digital approaches to overcome these limitations has been the focus of a project funded by the European Union under its Marie Curie Fellowship scheme between 2014 and 2016. This project is a pilot study aimed at developing and testing a framework for aggregating provenance information and exploring its suitability for addressing research questions relating to the history of medieval manuscripts.

The work carried out in this project between 2014 and 2016 has encompassed three main areas: developing a model for recording provenance information in a digital environment, implementing this approach in a specific software environment, and testing it against specific research questions. The Phillipps collection provides a relatively well-documented case study on a large scale. It transcends individual present-day institutional collections—which are usually the focus of collection-oriented research—and provides a larger body of data with which to address broader questions about the movement of manuscripts over the centuries.

*Speculum* 92/S1 (October 2017)

TABLE 1  
Data sources relating to the Philipps manuscripts

Source	Format	Comments
Schoenberg Database of Manuscripts	Relational database	Incorporates other sources, esp. sales catalogs 6,000 Philipps MSS; 20,000 Philipps events
Library catalogs (BL, KB, etc.)	Relational databases	Generally MARC records Provenance in notes Export can be awkward
Union catalogs	Relational databases Printed bibliographies	Formats vary Coverage varies Export can be awkward
Sale catalogs	Printed books (some digitized) Online sources (PDFs, web sites)	Many included in Schoenberg MSS in ABE, eBay, etc.
Philipps catalogs and lists	Printed book; partly digitized Supplemented by hand-written notes	Partly included in Schoenberg Handwritten notes not digitized
Philipps provenance indexes (BL, IRHT)	Handwritten; not digitized	Arranged by Philipps number No longer updated
Annotated sales catalogs & printed catalogs	Handwritten; not digitized	Researchers (Munby), owners (Philipps), auctioneers (Sotheby's) Held in Cambridge UL, Bodleian, BL

#### SOURCES OF DATA

The annotated printed catalogs and card indexes already mentioned are only one of the numerous sources of data relating to the history of the Philipps manuscripts. These are summarized in Table 1. The most important digital source, which focuses on provenance events, is the Schoenberg Database of Manuscripts, developed and managed by the Schoenberg Institute for Manuscript Studies in the University of Pennsylvania Library: <https://sdbm.library.upenn.edu/>. It contains something like twenty thousand entries relating to about six thousand of the Philipps manuscripts. Most of its data consist of lot descriptions from sale and auction catalogs, which are each given individual records in the database. In the first iteration of the database, entries relating to the same manuscript could be manually linked, but there was no overarching “master record” for a manuscript. A new data model, currently in beta testing, does create such a record, to which each of the records that identifies provenance events can be linked.<sup>4</sup>

<sup>4</sup> Lynn Ransom, “The Schoenberg Database of Manuscripts Project: An Update,” paper presented at the Sixteenth Seminar on the Care and Conservation of Manuscripts, University of Copenhagen, Faculty of Humanities, 13–15 April 2016.

Apart from a specialist source like the Schoenberg Database of Manuscripts, the main sources of evidence are the many catalogs of individual institutions and some aggregated products derived from them, such as Digital Scriptorium, a North American consortial database with records from forty contributing partners: <http://bancroft.berkeley.edu/digitalscriptorium/>. Those libraries and museums with large manuscript holdings—and sufficient funds—have usually produced a printed catalog of their holdings.<sup>5</sup> These catalogs are normally very detailed, accurate, and thorough, though their coverage may be increasingly out of date.

In the digital environment, institutional cataloging practices vary greatly. Some institutions have transferred all the information from their printed manuscript catalogs or card indexes into their online catalogs; others still maintain separate manuscript databases or finding aids. Some—especially art museums—do not make their full collections database available over the web and only offer selected digital highlights of their collection.

Where manuscripts are described in catalogs and databases, the treatment of provenance can be inconsistent. Some catalogs have no provenance information at all, while other institutions provide detailed and thorough provenance histories. These are often located in a “notes” field, however, which may not be searchable. Only a few institutions provide a specific access point for former owners like Sir Thomas Phillipps, thereby enabling all former Phillipps manuscripts to be identified through an “author” search.

Saving and exporting catalog records can also be difficult, even with the most thorough and comprehensive online catalogs. For most library databases, relevant records can be exported only individually or in small batches, and often in a limited range of bibliographic formats, which may not include the crucial provenance notes. While this may partly reflect the limitations and inconsistencies endemic in manuscript cataloging practices, the main reasons are the limitations of the specific brand of software involved and the choices made in configuring that software for use. The Schoenberg Database of Manuscripts, in contrast, provides all its data as downloadable Excel or CSV files. These can then be analyzed, filtered, and imported into other environments.

Other sources of evidence include printed and handwritten lists of the Phillipps manuscripts, but using these to construct a master list against which to check institutional holdings is not a straightforward task. Phillipps’s own printed catalog has been the subject of extensive study and reconstruction, especially by Munby.<sup>6</sup> Published in stages between 1837 and 1871, it covers manuscript numbers up to 23,837. Subsequent numbers are covered to some extent by the probate inventory drawn up in 1872 by Edward A. Bond of the British Museum, after Phillipps’s death. There are two different handwritten, unpublished versions of this inventory

<sup>5</sup> See, for example, C. W. Dutschke with R. H. Rouse, *Guide to Medieval and Renaissance Manuscripts in the Huntington Library*, 2 vols. (San Marino, CA, 1989); Laura Light, *Catalogue of Medieval and Renaissance Manuscripts in the Houghton Library, Harvard University* (Binghamton, NY, 1995); Lilian M. C. Randall, *Medieval and Renaissance Manuscripts in the Walters Art Gallery* (Baltimore, MD, 1988–97); Barbara Shailor et al., *Catalogue of Medieval and Renaissance Manuscripts in the Beinecke Rare Book and Manuscript Library, Yale University* (Binghamton, NY, 1984–2004).

<sup>6</sup> A. N. L. Munby, *The Catalogues of Manuscripts and Printed Books of Sir Thomas Phillipps*, *Phillipps Studies* 1 (Cambridge, UK, 1951).

in the Harrison D. Horblit Phillipps Collection in the Grolier Club in New York (Phillipps Collection Cat. 13 and Cat. 14).

The first version, described by Munby, extends the list of Phillipps manuscripts from no. 23,838 to no. 26,179. Another copy of this version is in the Bodleian Library (MS Phillipps-Robinson e.466). Munby added the manuscript titles from this version to his annotated working copy of the printed catalog. The second version of the probate inventory in the Grolier Club is somewhat longer, finishing with no. 26,365. It seems to have been used as a working tool by Phillipps's grandson, Thomas FitzRoy Fenwick, during the decades he spent on the gradual dispersal of the manuscripts and includes numerous corrections and renumberings.

Both the printed and handwritten Phillipps catalogs suffer from several inherent problems. Some manuscripts have duplicate numbers, and some numbers refer to more than one manuscript. Phillipps was inconsistent in assigning numbers; in some cases, one number may cover several volumes or a whole collection of documents, while in other cases a number may simply refer to a single document or a single object.

#### MODELS FOR RECORDING PROVENANCE INFORMATION

Bringing together data from these disparate sources requires a framework or common model for recording provenance information to which the data can be mapped. But there is no agreed approach for doing this in a computational setting, which is why the initial stages of my project were occupied with examining alternatives and developing a workable solution.

Although libraries have had considerable experience in managing and cataloging manuscript collections, the development of the MARC format in the late 1960s did little for manuscripts. Libraries have tended to force them into the MARC record structure as if they were printed books. While manuscript-specific extensions of MARC and the Anglo-American Cataloguing Rules were eventually developed, they focused on elaborating the details of manuscript description.<sup>7</sup> Provenance was handled through the use of two unstructured "notes" fields: "541: Immediate source of acquisition" and "561: Provenance note." This approach was based on printed manuscript catalogs, which usually record provenance as a series of narrative statements about ownership. A typical example reads:

Ar. has a note "Boetii Musica, an Ancient MS. in fine Condition with diagrams. A MS. of a work of rare occurrence bound by C. Lewis. H. Drury 1824"; spine carries the small printed number "3345", being that of Sir Thomas Phillipps; Ev. in modern pencil "W. H. Robinson 5.9.1949. £LE-N/a/-" Ar. has the stock no. "587673"; Ev. bears the library's shelf-mark \*091/B63.

This account of a tenth-century copy of Boethius owned by the State Library of Victoria records the evidence relating to provenance in chronological order, assembled into a single narrative paragraph.<sup>8</sup>

<sup>7</sup> Gregory Pass, *Descriptive Cataloging of Ancient, Medieval, Renaissance, and Early Modern Manuscripts* (Chicago, 2003).

<sup>8</sup> K. V. Sinclair, *Descriptive Catalogue of Medieval and Renaissance Western Manuscripts in Australia* (Sydney, 1969), no. 223.

This narrative approach to recording manuscript provenance is reflected in the Text Encoding Initiative (TEI). The TEI's <msDesc> (Manuscript Description) section contains a <history> element for encoding information from a catalog entry about the life cycle of a manuscript.<sup>9</sup> Within this, the <acquisition> and <provenance> elements are similar to the note fields in a MARC record.

The problem with these approaches is that the provenance information is too loosely structured for computational purposes. The museum sector, which also has a strong interest in provenance (particularly in relation to authentication), is grappling with the same problem. In museum databases, provenance information for an individual object is usually given in the form of a chronological listing of owners. The American Association of Museums, in the *AAM Guide to Provenance Research*, provides a guideline for structuring these notes.<sup>10</sup>

The Carnegie Museum of Art has recently developed a draft *Provenance Standard*, which aims to make provenance information more suitable for machine-based parsing.<sup>11</sup> It uses a chronological data model, in which each object has a "timeline" divided into a number of "periods." Each of these includes a "party," an "acquisition method," a "location," and a "beginning and end."<sup>12</sup> This is still an experimental approach, which has not yet spread to the museum sector more generally.

From the point of view of manuscript provenance, this strictly chronological timeline is a useful starting point for thinking about data models. But it does not cover all the different aspects of manuscript history or the different types of evidence that need to be recorded. Many manuscripts contain various notes, numbers, foliations, and other possible evidence of ownership or history; these cannot always be fitted into a chronological model. Periods of ownership, perhaps inferred from other evidence, may also lack specific chronological boundaries. The changing nature of a manuscript over time (with different contents and versions) may also provide evidence for ownership without a definite chronology. The provenance of manuscripts is not limited to transfers of ownership, unlike the museum approach.

Another important contribution from the museum sector, this time with international endorsement, is the Conceptual Reference Model developed by CIDOC—the documentation committee of the International Council of Museums (ICOM). Known as CIDOC-CRM, this is "a formal ontology intended to facilitate the integration, mediation and interchange of heterogeneous cultural information," which is now an international standard (ISO 21127:2014).<sup>13</sup> CIDOC-CRM is a large and

<sup>9</sup> Text Encoding Initiative, *P5: Guidelines for Electronic Text Encoding and Interchange version 3.1.0*, updated 15 December 2016, section 10.8, "Manuscript Description: History," <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html#mshy>.

<sup>10</sup> Nancy H. Yeide, Amy Walsh, and Konstantin Akinsha, *AAM Guide to Provenance Research* (Washington, DC, 2001).

<sup>11</sup> Carnegie Museum of Art, "The CMOA Provenance Standard, Draft Version 0.2," 14 October 2016, <http://www.museumprovenance.org/reference/standard/>.

<sup>12</sup> Tracey Berg-Fulton, David Newbury, and Travis Snyder, "Art Tracks: Visualizing the Stories and Lifespan of an Artwork," paper presented at MW2015: Museums and the Web 2015, <http://mw2015.museumsandtheweb.com/paper/art-tracks-visualizing-the-stories-and-lifespan-of-an-artwork/>.

<sup>13</sup> Nick Crofts, Martin Doerr, Tony Gill, Stephen Stead, and Matthew Stiff, ed., *Definition of the CIDOC Conceptual Reference Model*, version 5.0.4 (CIDOC, 2011) [http://www.cidoc-crm.org/sites/default/files/cidoc\\_crm\\_version\\_5.0.4.pdf](http://www.cidoc-crm.org/sites/default/files/cidoc_crm_version_5.0.4.pdf).



complex model, but its focus on events is very relevant to the problem of representing manuscript provenance. Christian-Emil Ore has examined how manuscript provenance can be represented in CIDOC-CRM, using the Kringla Leaf as his example.<sup>14</sup> While this produces a dauntingly complicated series of diagrams, it helps to identify various types of provenance events that can serve as the basis for a new data model.

The data model developed for the Philipps Collection project draws together elements from CIDOC-CRM's event-based ontology, the Carnegie Museum's chronological framework for transfers of ownership, and the evidence-based approach of the Schoenberg Database of Manuscripts. It is very much a work in progress, sufficient to meet the initial requirements of that project; it does not aim to be exhaustive, but is designed to be hospitable to further expansion and elaboration. The model focuses on relationships between entities, conceptualized in the form of a property graph.<sup>15</sup>

Table 2 summarizes the initial structure of the model. Four basic nodes are identified: agent, object, publication, and work. "Agent" refers to the different types of actors, specifically persons and organizations, while "object" refers to a physical entity—the manuscript itself. "Work" is a conceptual entity, rather than a physical one, and may refer to the text(s) carried by a manuscript, the description of a manuscript contained in a catalog, or an exhibition described in a catalog. Relationships can include the various transactions relating to a manuscript—"bought," "sold," "produced"—as well as other types of connections between works, agents, objects, and publications.

#### A DIGITAL ENVIRONMENT FOR PROVENANCE RESEARCH

Once a data model has been developed, it needs to be implemented in a suitable software environment. For the Philipps Collection project, this implementation has been carried out using *nodegoat*.<sup>16</sup> Developed in the Netherlands by the Lab1100 group, *nodegoat* is described as a web-based data management, network analysis, and visualization environment. It was chosen for this project for its flexibility in data modeling, its ability to import data through CSV files, and its capacity to produce geographical and network visualizations of the aggregated data.

The data structure used in *nodegoat* consists of objects with subobjects attached to them. These subobjects can embed relationships to other objects, as well as containing properties like time and place. The generic data model for manuscript provenance described above was mapped onto this framework. A "manuscript" object can have subobjects for transactions, such as "sold" and "owned," and these subobjects can contain links to the "persons" and "organizations" participating in each transaction, as well as the times and places where the transaction occurred. Table 3 shows the basic form of the data model as mapped to the *nodegoat* structure.

<sup>14</sup> Christian-Emil Smith Ore, "CIDOC CRM and Object Provenance," 2009, [http://www.tei-c.org/SIG/Ontologies/presentations/CIDOC\\_CRM\\_provenance.pdf](http://www.tei-c.org/SIG/Ontologies/presentations/CIDOC_CRM_provenance.pdf).

<sup>15</sup> Ian Robinson, Jim Webber, and Emil Eifrem, *Graph Databases* (Sebastopol, CA, 2013).

<sup>16</sup> Pim Van Bree and Geert Kessels, "Mapping Memory Landscapes in Nodegoat," in *Social Informatics*, Proceedings of the 6th International Conference, Barcelona, November 11–13, 2014, ed. L.M. Aiello and D. McFarland, Lecture Notes in Computer Science 8852 (Berlin, 2015), 274–78.

TABLE 2  
Data model expressed as a property graph

Node	Relationship	Node
AGENT: Person	BOUGHT	OBJECT: Manuscript
AGENT: Organization	SOLD	OBJECT: Manuscript
OBJECT: Manuscript	CONTAINS	WORK: Text
AGENT: Person	COMPOSED	WORK: Text
PUBLICATION: Catalog	CONTAINS	WORK: Description
OBJECT: Manuscript	DESCRIBED_AS	WORK: Description
AGENT: Organization	PRODUCED	PUBLICATION: Catalog
WORK: Exhibition	DESCRIBED_IN	PUBLICATION: Catalog

The subobjects attached to each manuscript provide a summary of the known events in its provenance history. Fig. 1 shows a typical *nodegoat* record, for Philipps manuscript 12264. The “object” record, which contains a physical description of the manuscript, is accompanied by a series of “sub-objects” that show events in its history: “produced,” “owned,” and “sold.” Fig. 2 gives the details for two of the “sold” subobjects, showing how places, persons, and organizations are linked into each event in the history of this manuscript.

The vocabularies used for specific elements within the *nodegoat* database are partly imposed by *nodegoat* itself and partly chosen by the creator of the database. The Geonames ontology for place names (<http://www.geonames.org/ontology/documentation.html>) is incorporated into *nodegoat*; it makes geographical visualizations of the data relatively easy, but also causes some problems by recognizing only contemporary countries and cities. Personal names in *nodegoat* can be matched

TABLE 3  
Data model adapted for *nodegoat*

Object	Subobjects	Related to:
PERSON	Nationality (country)	Manuscript Text Catalog
ORGANIZATION	Location (city; country)	Manuscript Text Catalog
MANUSCRIPT	Sold Donated Owned Described In Produced Contents	Person/Organization: <i>Agent, Owner, Buyer, Donor, Recipient, Scribe, Artist, Producer</i> Location (city; country) Catalog Text
TEXT		Person: <i>Author</i> Manuscript
CATALOG		Organization: <i>Publisher</i> Person: <i>Compiler</i> Manuscript



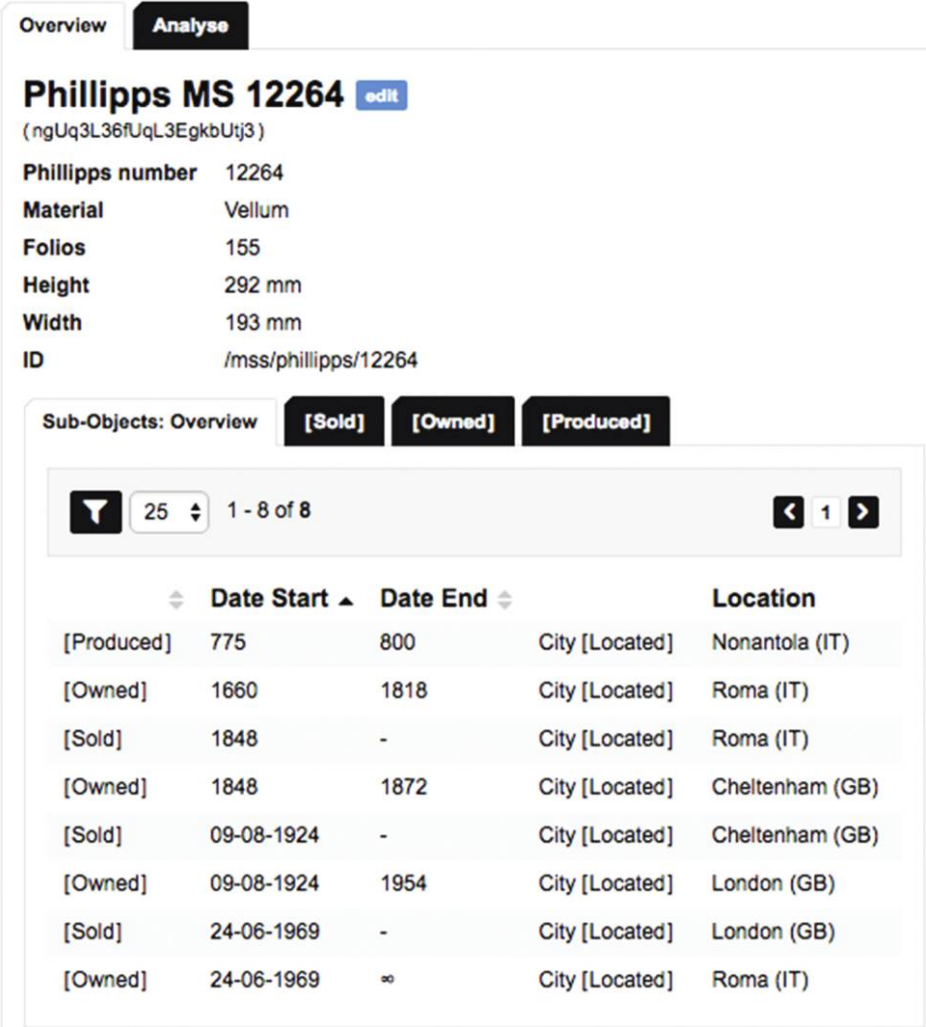


Fig. 1. nodegoat record for Phillipps MS 12264.

against DBPedia (<http://wiki.dbpedia.org/>) and the Virtual International Authority File (VIAF, <https://viaf.org/>). Other types of classifications can be created manually, including languages, currencies, and materials. The vocabulary used to describe subobjects can be derived from CIDOC-CRM properties, with some additions. The major gap is the lack of any standard method for identifying individual manuscripts, other than their Phillipps numbers or their current institutional shelf mark; for the purposes of the project, Phillipps numbers have been used as identifiers and as the basis for creating Uniform Resource Identifiers (URIs).

Data can be entered directly into *nodegoat* using web forms created from the data model designed by the user. But it is generally more efficient to import data in bulk, in the form of CSV (Comma Separated Values) files—simple text files used to store tabular data, which can be viewed and edited in standard spreadsheet software like

*Speculum* 92/S1 (October 2017)

Sub-Objects: Editor   **Sub-Objects: Overview**   [Sold]   [Owned]   [Produced]

[Sold] + ++   [Described In] + ++   [Donated] + ++   [Owned] + ++   [Produced] + ++   [Contents] +

[Contents 3] + ++   [Contents 4] + ++

[Sold]	del	ver
Date	09-08-1924	
Location	Reference	
Location Reference	City	Located
	Cheltenham (GB)	
Agent		ver
Owner	Fenwick, Thomas FitzF	ver
Schoenberg ID		ver
Price	800	ver
Currency		ver
Source of Data		ver
Lot number		ver
Notes	Private sale	ver
Buyer org		ver
Buyer indiv	Beatty, A. Chester	ver

[Sold]	del	ver
Date	1848	
Location	Reference	
Location Reference	City	Located
	Roma (IT)	
Agent		ver
Owner	Petrucchi, Giambattista	ver
Schoenberg ID		ver
Price		ver
Currency		ver
Source of Data		ver
Lot number		ver
Notes		ver
Buyer org	Thorpe, Thomas	ver
Buyer indiv	Phillipps, Thomas, Sir	ver

Fig. 2. “Sold” sub-objects for Phillipps MS 12264.

Excel. These files can be either keyed in or derived from library catalogs or other database exports. The Schoenberg Database is exemplary in making its data available for download as a CSV file, which can then be mapped to the *nodegoat* data model. An “import template” must be created for each unique CSV format by mapping the CSV columns to objects and subobjects in *nodegoat*. The import routine consists of loading the CSV file to *nodegoat*, choosing the appropriate import template, and running the import process. Various parameters can be set for this process, including whether to create new objects (such as manuscript records) automatically as part of the data load or to hold each object for manual review against possible matches. Fig. 3 shows the import template for a CSV file derived from the Bodleian Library’s list of numbered Phillipps manuscripts in the Phillipps-Robinson collection.

The major difficulties in the import process are likely to arise from mapping the source data to the *nodegoat* data model. This is especially so when multiple pieces of information are conflated into a single column in the CSV file, or into a single field in the source database. A person’s dates of birth and death may be included in the same column as his or her name, for example, or the names of multiple previous owners may be included in a single provenance field. Alternatively, a field may be given in a narrative or note format containing multiple pieces of information. In these cases, the source data are likely to require extensive editing with a tool like OpenRefine.<sup>17</sup>

<sup>17</sup> Ruben Verborgh and Max De Wilde, *Using OpenRefine* (Birmingham, UK, 2013).

Import Templates
Sources
String to Object Pairs

### Import Template: Bodleian PR Owned Location

Name

Bodleian PR Owned Locat

Description

Last Run

2015-05-06

Structure

del

add

Columns

{Name:""}	Manuscript	Object Name	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
{"Owner org":""}	Manuscript	[Owned] - Owner	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
{"Ownership number":""}	Manuscript	[Owned] - Owners	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
{"Date":""}	Manuscript	[Owned] - Date St	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
{"Notes":""}	Manuscript	[Owned] - Notes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
{"Location - city":""}	Manuscript	[Owned] - Location	<input type="checkbox"/>	<input type="checkbox"/>	City
{"Catalogue":""}	Manuscript	[Owned] - Source	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 3. Import Template for Bodleian Library CSV file.

Within *nodegoat*, the data can be searched and explored in various ways. All the records in the database can be browsed and listed, as shown in Fig. 4. The records can also be filtered to show only those which meet a single criterion or a combination of criteria. Fig. 5 shows the process for filtering records to show those where Alfred Chester Beatty appears as an owner. Fig. 6 shows the process for finding manuscripts that are or were located in Rome at some point in their history. These filters can easily be combined to show those manuscripts once owned by Beatty and also located in Rome.

A particularly valuable feature of *nodegoat* is its capacity to produce geographical and network visualizations. Fig. 7 shows a geographical visualization of the history of a single manuscript (Phillipps 12264). The visualization can be extended to a group of manuscripts; Fig. 8 shows the provenance of a sample of twenty-one manuscripts that were owned in the nineteenth century by Thomas Phillipps and in the twentieth century by Alfred Chester Beatty. There is a time slider attached to the geographical visualizations, so that they can be limited to specific periods of time. Fig. 9 shows the provenance of this group in the period up to 1781; the data are limited mostly to their places of origin, shown as purple dots, together with some earlier migrations shown in red. Fig. 10 shows the history of the same group up to 1927; by this stage they had passed through Phillipps's ownership and were now all owned by Beatty in London. Fig. 11 shows the results of Beatty's sales of most of these manuscripts from 1932; many of them are now owned by institutions in the United States.

*Speculum* 92/S1 (October 2017)

The screenshot displays the nodegoat web application interface. At the top, there are two tabs: 'Types' and 'Classifications'. Below these, a row of buttons represents different entity types: 'Organization', 'Text', 'Country', 'Source of Data', 'Person', 'Manuscript' (which is highlighted), and 'City'. A large 'Add Manuscript' button is positioned below the entity type buttons. Further down, there is a search bar with a magnifying glass icon, a filter icon, a dropdown menu set to '50', and a text indicating '51 - 100 of 1.466' records. The main content area is titled 'Name' and lists 18 manuscript records, each on a separate line: 'Phillipps MS 24025', 'Phillipps MS 11295', 'Phillipps MS 15439', 'Phillipps MS 14406', 'Phillipps MS 14405', 'Phillipps MS 14404', 'Phillipps MS 14403', 'Phillipps MS 14402', 'Phillipps MS 14401', 'Phillipps MS 14400', 'Phillipps MS 14399', 'Phillipps MS 14398', 'Phillipps MS 14397', 'Phillipps MS 14396', and 'Phillipps MS 14395'.

Fig. 4. *nodegoat* list of records.

The other kind of visualization is a network graph, showing the connections between manuscripts and persons, organizations, and places. Fig. 12 shows the network graph for the same group of Phillipps-Beatty manuscripts. The predominant elements are Beatty (who owned all these manuscripts), Sotheby's (which sold most of them), and London (where all of them were owned or sold). A time slider can be used here too, with Fig. 13 showing the network graph for these manuscripts since 1936. The network graph for a single manuscript can also be viewed, as in Fig. 14.

*Speculum* 92/S1 (October 2017)

**[Owned]**

**Amount**   del add

**Date**

**Date From**

**Date To**  del add

**Location**

**Location Reference**   del add filter

more

---

Owner - person - Owner - organization + Location - city + Location - country +

Ownership number + Source of Data +

del add filter

**Owner - person**  more

Fig. 5. Records filtered for Chester Beatty as owner.

Both the geographical visualizations and the network graphs can be explored for an entire data set. Fig. 15 shows the geographical visualization, and Fig. 16 the network graph, for a pilot data set of nearly 1,500 former Phillipps manuscripts. These visualizations are not explanations in themselves. Instead, they are helpful summaries of selected data, which may serve as useful pointers and diagnostic tools for thinking about the data set. The relative prominence of Quaritch in Fig. 15, for example, might prompt a more detailed study of the role of that book dealer in the dispersal of the Phillipps manuscripts. The visualizations are not substitutes for the data, however. They tend to gloss over uncertainties, ambiguities, and gaps in the evidence—and almost inevitably simplify the picture.<sup>18</sup>

The usefulness of *nodegoat* lies primarily in its capacity to gather and store large quantities of data in a consistent format and to present it for exploration and analysis by the researcher. The visualizations are a particularly helpful way into the data. *Nodegoat* is also valuable as a means of sharing and exporting data, either

<sup>18</sup> Johanna Drucker, *Graphesis: Visual Forms of Knowledge Production* (Cambridge, MA, 2014); Willard McCarty, *Humanities Computing* (Basingstoke, 2005), 29–34.

**Sub-Objects**

Sub-Objects: Any  [Sold]  [Described In]  [Donated ]

[Contents 2]  [Contents 3]  [Contents 4]

**Sub-Objects: Any**

**Amount**

**Date**

**Date From**

**Date To**

**Location**

**Location Reference**

Fig. 6. Records filtered for Rome as location.

by giving other researchers direct access to the *nodegoat* database or by making data available for reuse as CSV files.

## RESEARCH QUESTIONS

The *nodegoat* database assembled as part of this project is being used to address a series of provenance-oriented research questions. The first of these, illustrated in many of the figures above, relates to Phillipps manuscripts that were subsequently owned by Alfred Chester Beatty. Where were these manuscripts produced, what are their histories, and who has owned them? Where are they now? This research was presented at a symposium at Trinity College Dublin in 2015 and updates earlier work by Christopher De Hamel.<sup>19</sup>

Alfred Chester Beatty owned at least sixty manuscripts that had previously belonged to Phillipps. Most of them were bought directly from Phillipps's grandson, Thomas FitzRoy Fenwick, and were acquired in four groups between 1920 and 1925. Beatty also made ten other individual purchases of Phillipps manuscripts. His final purchase, in 1948, was one of Phillipps's great treasures: the Armenian Gospel Book (Phillipps 15364) with which he was photographed in 1860.

<sup>19</sup> Christopher de Hamel, "Chester Beatty and the Phillipps Manuscripts," *Book Collector* 40/3 (1991): 358–70.

*Speculum* 92/S1 (October 2017)



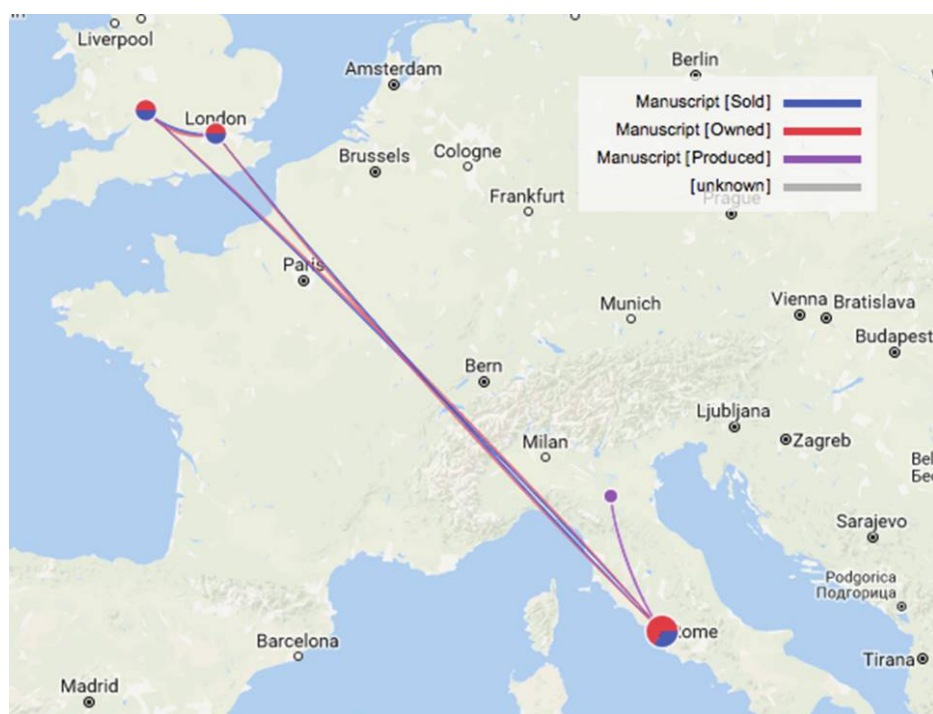


Fig. 7. Geographical visualization – history of Phillippus Manuscript 12264.

Only thirteen of these manuscripts now remain in the Chester Beatty Library in Dublin, Ireland. The others were sold, primarily in 1932–33 and 1968–69, and are now in collections around the world. The current locations of 34 of them are known: United States 18, Italy 6, United Kingdom 6, Switzerland 2, Germany 1, and Israel 1. One manuscript (Phillippus 2506) was broken up after it was sold in 1969; at least sixteen individual leaves from it have passed through the sale rooms in the last forty-five years. The current locations of the twelve remaining Phillippus-Beatty manuscripts remain unknown.

Recording and analyzing the provenance of these manuscripts in *nodegoat* helps to clarify and visualize their histories. The Phillippus-Beatty manuscripts have passed through the hands of most of the great private collectors of the nineteenth and twentieth centuries, for example, giving some indication of their significance and of the overlapping interests of these collectors. The pre-Phillippus provenance of these manuscripts included most of the major collectors of the nineteenth century: Leander van Ess, Luigi Celotti, Henry Drury, Gerard and Johann Meerman, Richard Heber, Joseph-Felix Allard, and Justin MacCarthy-Reagh. Their post-Beatty provenance includes many well-known collectors of the twentieth century: St. John Hornby, Major J. R. Abbey, Philip Hofer, Eric Millar, William Scheide, Martin Schøyen, Peter Ludwig, Sir Paul Getty, and A. S. Yahuda.

A subgroup with a particularly interesting provenance is the six manuscripts now in the Biblioteca Nazionale Centrale di Roma. They were all produced at the abbey

*Speculum* 92/S1 (October 2017)

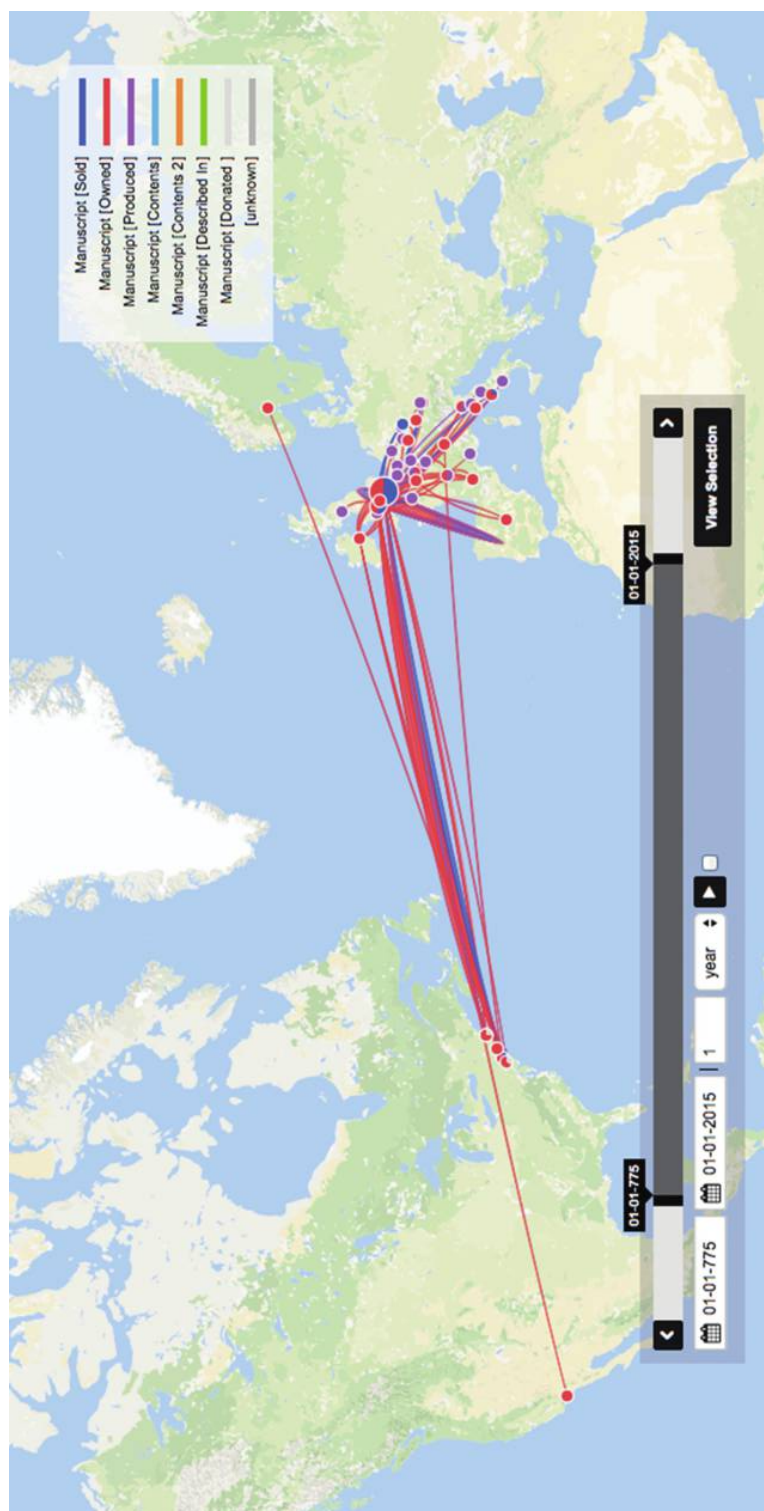


Fig. 8. Geographical visualization – Philipps-Beatty sample group.

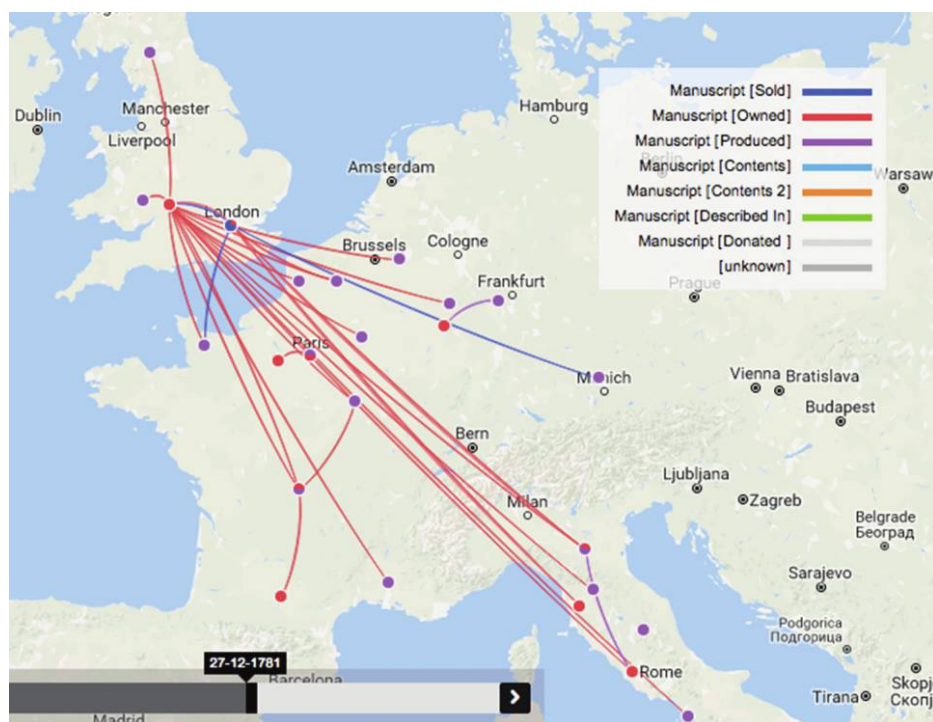


Fig. 9. Geographical visualization – Phillippus-Beatty sample group to 1781.

of Nonantola, near Bologna, mostly in the ninth century, so their history has taken them across Europe to Great Britain and eventually back to Italy. A seventh Nonantola manuscript was the one broken up after the 1969 Beatty sale (Phillipps 2506).

A second research question for which the *nodegoat* database has provided an analytical environment is an attempt to identify those Phillipps manuscripts which are now located in North American institutions.<sup>20</sup> A significant number of manuscripts moved from Great Britain to North America during the lengthy dispersal of the Phillipps collection. Some of these were acquired before Seymour De Ricci's *Census* of 1935, while others were acquired before Faye and Bond's *Supplement* in 1962.<sup>21</sup> Today, those collections contain more than two thousand three hundred Phillipps manuscripts.

The Phillipps manuscripts now in North America illustrate most of the main reasons for collecting medieval and early modern manuscripts. Some of them are beautiful and rare, especially illuminated manuscripts like the Crusader Bible now in the

<sup>20</sup> Toby Burrows, "Manuscripts of Sir Thomas Phillipps in North American Institutions," *Manuscript Studies* 1/2 (2016): 307–27.

<sup>21</sup> Seymour De Ricci, *Census of Medieval and Renaissance Manuscripts in the United States and Canada*, 3 vols. (New York, 1935–40), 209–11; C. U. Faye and W. H. Bond, *Supplement to the Census of Medieval and Renaissance Manuscripts in the United States and Canada* (New York, 1962), 619–21.

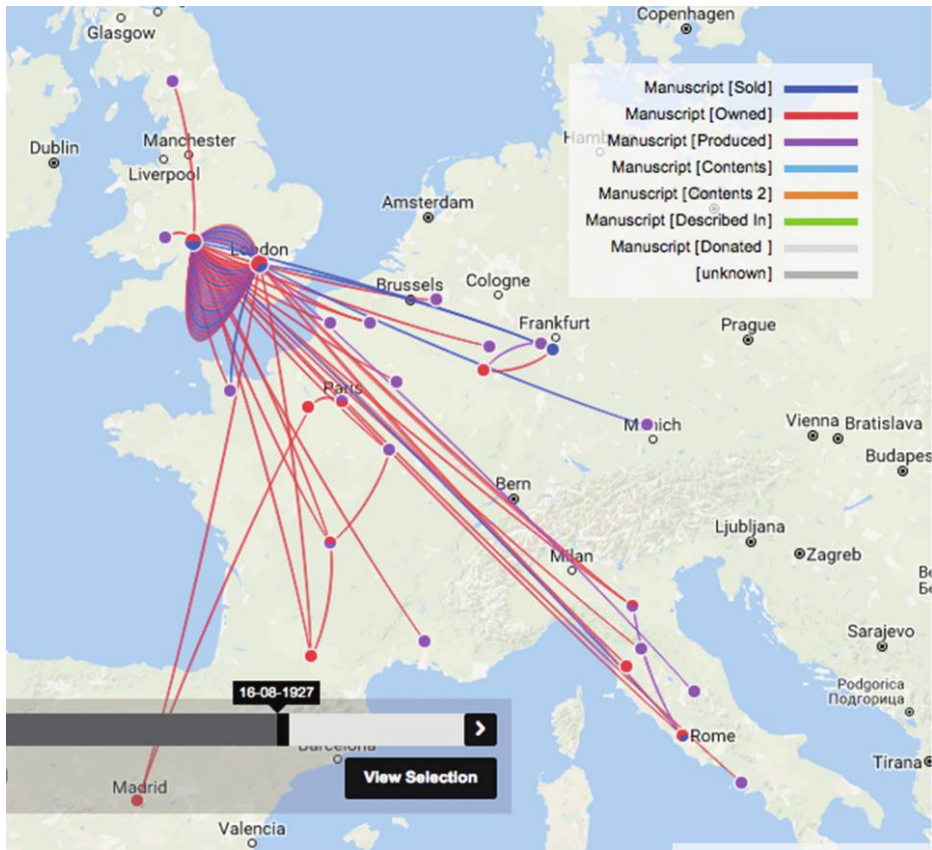


Fig. 10. Geographical visualization – Phillipps-Beatty sample group to 1927.

Morgan Library in New York. Some have considerable documentary significance for North American and European history, such as the Phillipps documents in the Library of Congress and the early modern materials in the Folger Library in Washington, D.C. Some, like the English common law manuscripts in the Harvard Law Library and in the Robbins Library at the University of California, Berkeley, are important for their specialist content. Some have considerable artistic value, such as the many paintings, drawings, and photographs in the collections of Harvard University, the Metropolitan Museum of Art, and the Gilcrease Museum of Western Art in Tulsa, Oklahoma.

The Phillipps manuscripts are something of a microcosm reflecting the history of manuscript collecting in North America, especially in the period between the 1890s and the 1970s. The libraries and museums that emerged during this period from the collections of well-known individuals, like Henry Huntington, Henry and Emily Folger, J. P. Morgan, and William and Henry Walters, have significant holdings of Phillipps manuscripts. The major private universities acquired a substantial number of Phillipps manuscripts, sometimes by purchase but more often as the result of donations and bequests by individual collectors. Phillipps manuscripts are also

*Speculum* 92/S1 (October 2017)



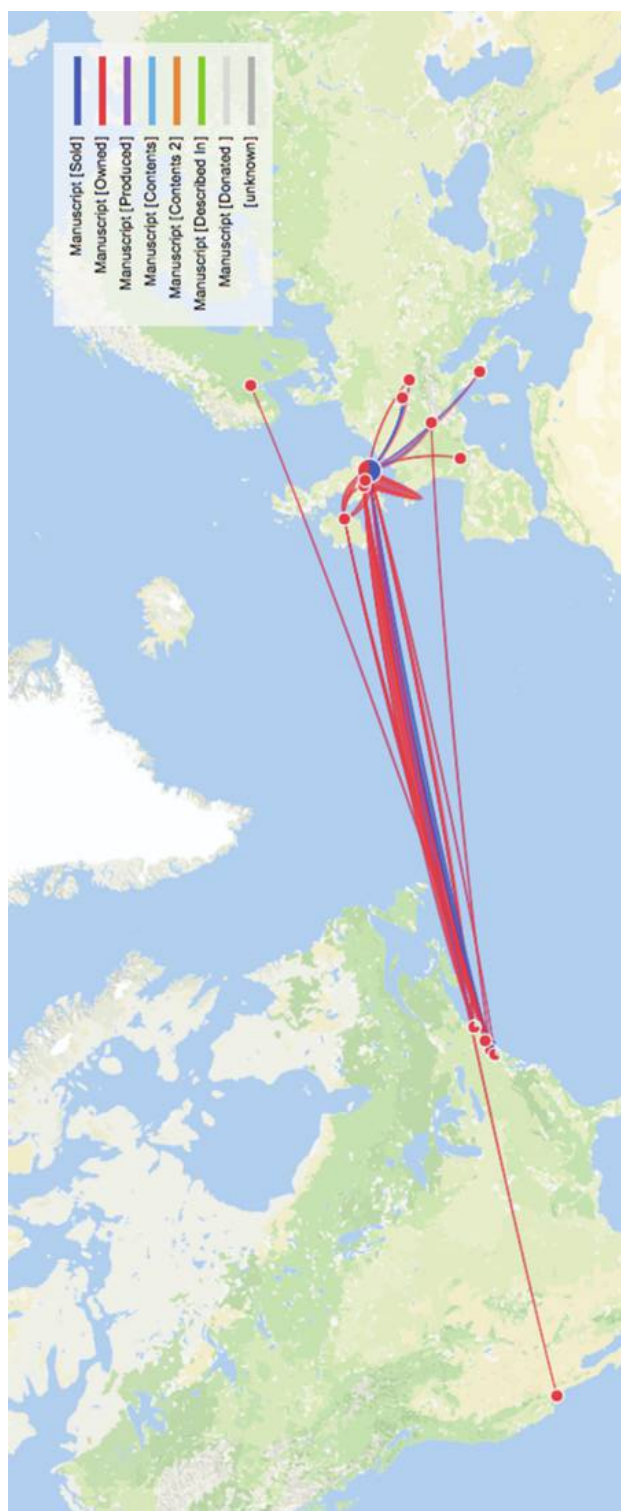


Fig. 11. Geographical visualization – Phillips-Beatty sample group since 1932.





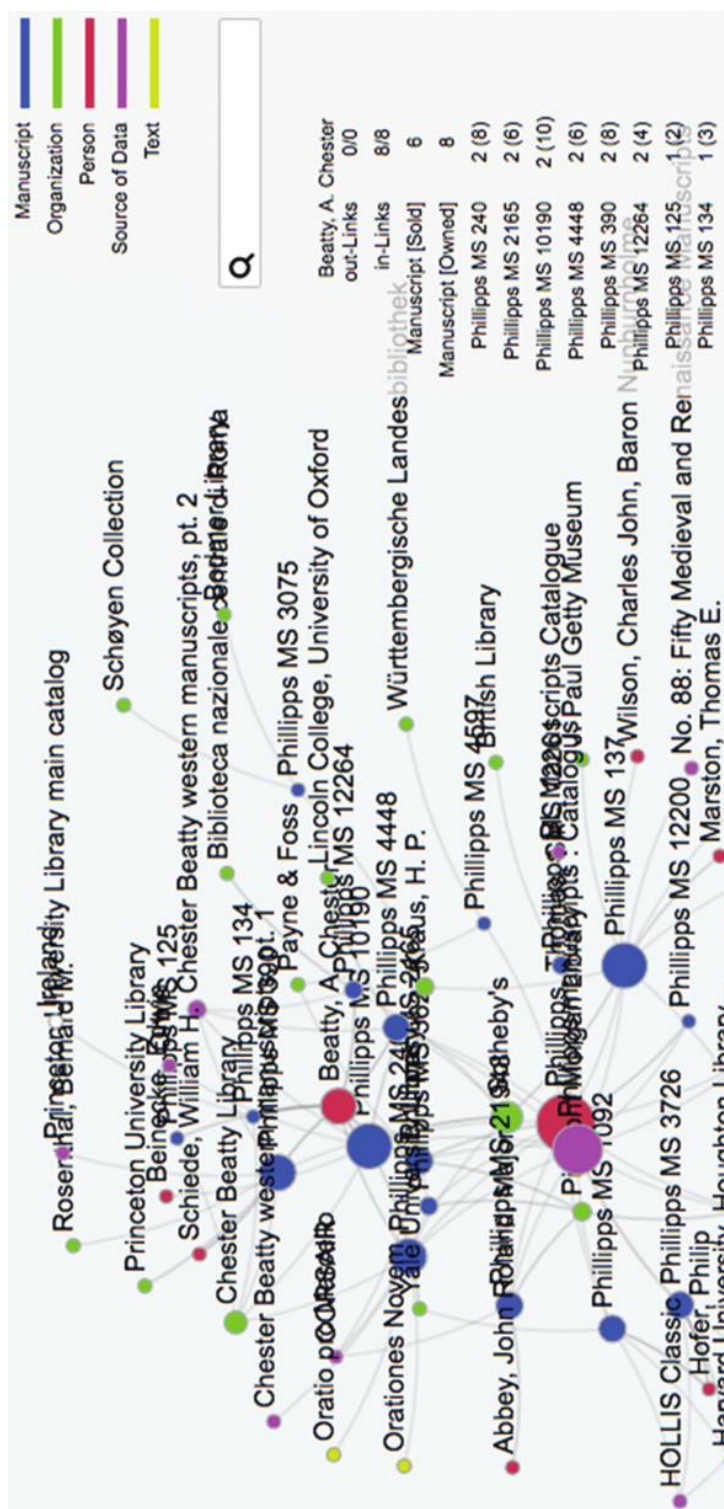


Fig. 13. Network graph – Phillipps-Beatty sample group since 1936.

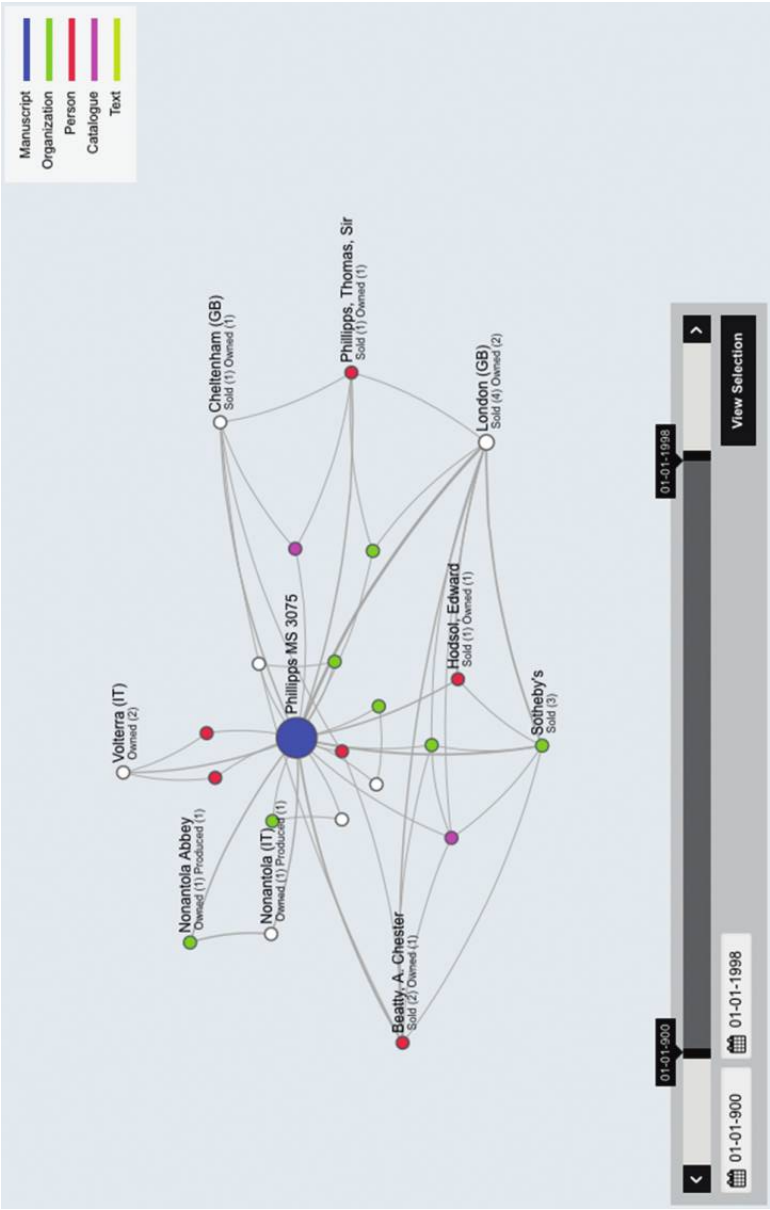


Fig. 14. Network graph – Phillipps Manuscript 3075.

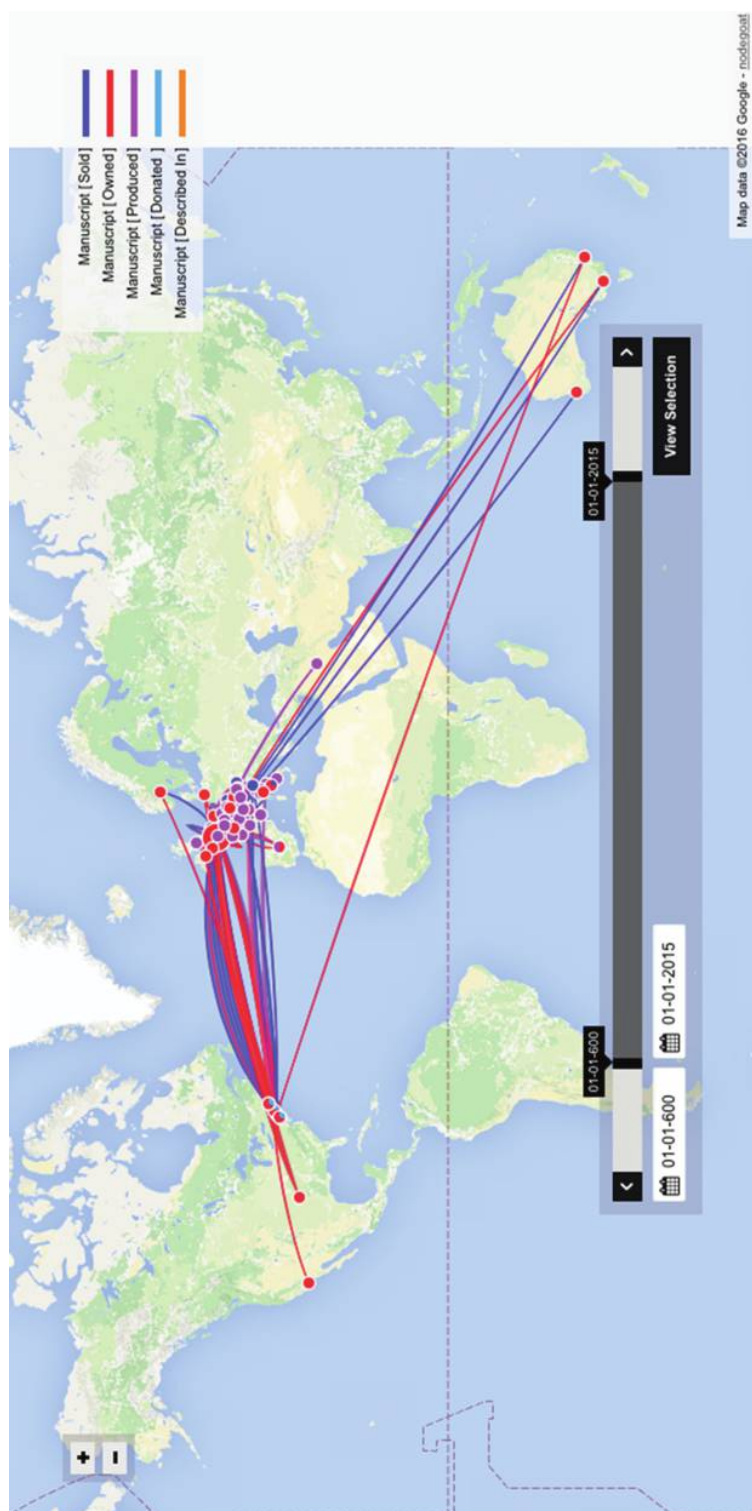


Fig. 15. Geographical visualization – pilot dataset.

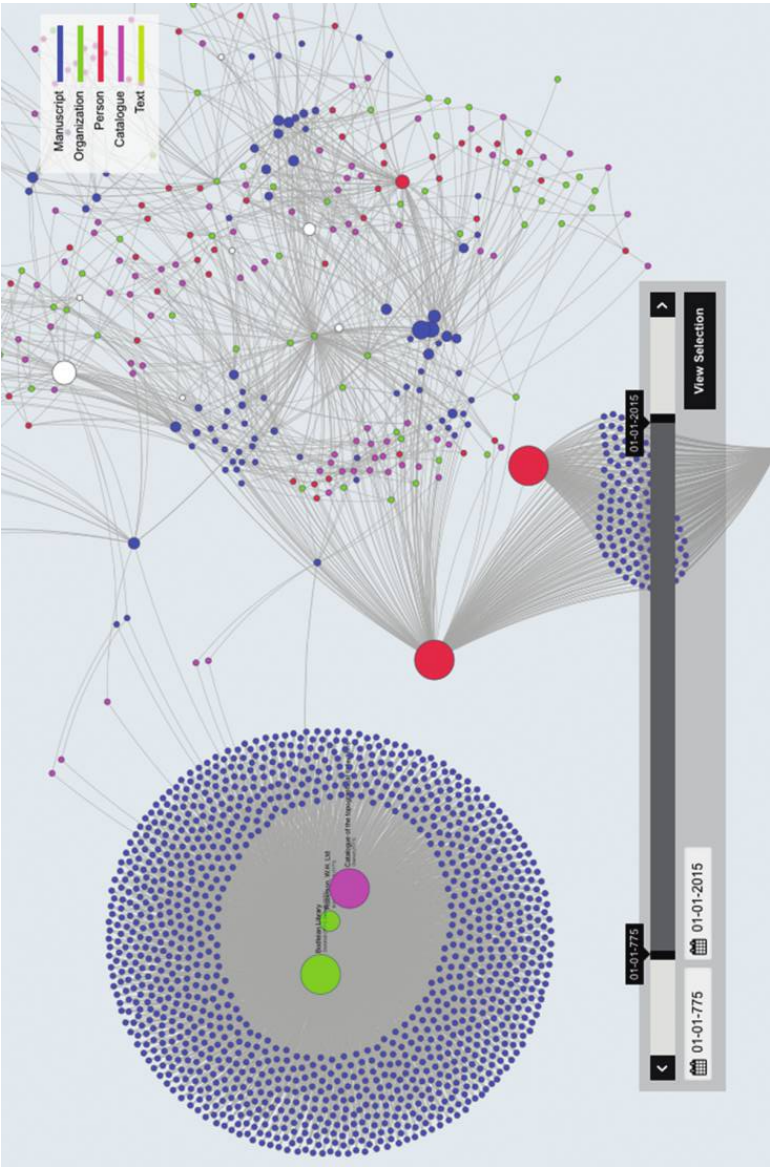


Fig. 16. Network graph – pilot dataset.

spread across many university and college libraries, as well as in a number of public-sector collecting institutions.

The *nodegoat* database makes it possible to identify manuscripts that fall into the category of being owned by American institutions and to analyze their characteristics and their relationships. This kind of research question is nevertheless a relatively small investigation into a specific grouping of manuscripts, rather than an analysis across the entire Philipps collection. This reflects the stage reached by the project, where only a small percentage of Philipps manuscripts have been covered and only the section of the network around these manuscripts has been documented. As the database expands its coverage, more quantitative questions should become feasible, particularly those connected to metrics for network analysis.<sup>22</sup>

#### THE FUTURE OF PROVENANCE DATA

This project was designed to test new digital approaches to aggregating evidence for the history of manuscripts. Its results have included a new data model for recording manuscript provenance, an assessment of the difficulties involved in combining data from a variety of heterogeneous sources, and an initial evaluation of the usefulness of geographical and network visualizations of the evidence relating to a pilot group of manuscripts. The project has also demonstrated how this kind of digital environment can be used as the basis for addressing research questions related to the history and provenance of specific groups of manuscripts.

The work done by this project is a first step towards a broader approach to the aggregation of data and the use of network analysis for large-scale research questions related to manuscript histories. The eventual goal should be the kind of linked data infrastructure that is being developed in classics and ancient history.<sup>23</sup> This “Graph of Ancient World Data” brings together information about people, places, texts, and objects from antiquity into an environment where large-scale questions can be asked and semantically complex topics can be explored. A similar framework for manuscript studies will require the following basic building blocks:

- Identifiers for medieval people, places, and organizations
- Identifiers for individual manuscripts—mapped to varying ways of citing institutional shelf marks
- Identifiers for the texts carried by manuscripts
- Linkable versions of specialist vocabularies for describing scripts, decoration and illumination, bindings, coats of arms, and bookplates

There is already a great deal of activity aimed at creating shareable digital materials derived from medieval manuscripts. This includes initiatives for the transcription and encoding of texts, especially using the TEI. There are also many libraries and projects creating and publishing digital images of manuscripts, and there is growing interest in enabling interoperability by sharing these images through the Inter-

<sup>22</sup> Robert A. Hanneman and Mark Riddle, *Introduction to Social Network Methods* (Riverside, CA, 2002), <http://faculty.ucr.edu/~hanneman/>.

<sup>23</sup> Leif Isaksen, Elton Barker, Rainer Simon, and Pau de Soto Cañamares, “Pelagios and the Emerging Graph of Ancient World Data,” in *WebSci '14: Proceedings of the 2014 ACM Conference on Web Science* (New York, 2014), 197–201.

national Image Interoperability Framework (IIIF).<sup>24</sup> A “Graph of Medieval Data” would sit as a unifying layer above all these digital resources and would provide a framework for cross-referencing and interlinking between existing services. One of its most important applications will be in the field of large-scale provenance research.

<sup>24</sup> Benjamin Albritton, “Fellow Travellers: The Canterbury Tales and IIIF,” on the blog *Digital Manuscripts at Stanford*, 2015, <http://web.stanford.edu/group/dmstech/cgi-bin/wordpress/fellow-travelers-the-canterbury-theses-and-iiif/>.

---

*Toby Burrows, School of Humanities, University of Western Australia (toby.burrows@uwa.edu.au)*

*Speculum* 92/S1 (October 2017)