



Open Library of Humanities



Part of the Ubiquity
Partner Network



Open Library of Humanities

Remaking Collections

How to Cite: Burrows, T 2018 Connecting Medieval and Renaissance Manuscript Collections. *Open Library of Humanities*, 4(2): 32, pp.1–23. DOI: <https://doi.org/10.16995/olh.269>

Published: 02 November 2018

Peer Review:

This article has been peer reviewed through the double-blind process of *Open Library of Humanities*, which is a journal published by the Open Library of Humanities.

Copyright:

© 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Open Access:

Open Library of Humanities is a peer-reviewed open access journal.

Digital Preservation:

The Open Library of Humanities and all its journals are digitally preserved in the CLOCKSS scholarly archive service.

REMAKING COLLECTIONS

Connecting Medieval and Renaissance Manuscript Collections

Toby Burrows^{1,2}

¹ University of Oxford, UK

² University of Western Australia, AU

toby.burrows@uwa.edu.au

This article examines issues affecting the reuse of data relating to collections of medieval and Renaissance manuscripts in libraries, museums and archives. These manuscripts are increasingly being made available in digital formats, although the extent is perhaps less than expected; a recent report on the situation in Germany estimated that only about 7.5% of the country's 60,000 manuscripts had been digitized. Discovering these manuscripts is heavily dependent on the quality and consistency of descriptive data about them, but the current situation is very mixed and inconsistent, despite several national programmes (such as *Manuscriptorium* in the Czech Republic, *e-codices* in Switzerland, and *Biblissima* in France) and a major international effort by *Europeana*. This article reports on recent work in manuscript studies, drawing on two major international projects. The first, funded by the European Union between 2014 and 2016, focused on the manuscript collection assembled in the nineteenth century by Sir Thomas Phillipps. It investigated ways of reconstructing the history of this vast, but now-dispersed collection, by bringing together data from a range of digital and non-digital sources. The second project, Mapping Manuscript Migrations, funded by the Trans-Atlantic Platform under its Digging into Data Challenge from 2017 to 2019, extends the earlier work to a much larger scale, and implements a Linked Open Data framework for combining and managing data related to medieval and Renaissance manuscripts. It enables large-scale analysis and visualization of their history and movements over the centuries.

Introduction

Hundreds of thousands of European medieval and Renaissance manuscripts have survived until the present day. Ranging from elaborately decorated volumes to individual documents and single pages, they often feature among the treasures of libraries, museums, galleries, and archives, and 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 provide inspiration and illustrations for films, novels, and artworks, and increasingly for sharing on social media.

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. These manuscripts form a crucial evidence base for the humanities, and research into their histories has important benefits for a wide range of disciplines.

As the result of changes of ownership over the centuries, medieval and Renaissance manuscripts are now spread all over the world, in collections across Europe, North America, South America, Asia, and Australasia. This means that researchers are heavily dependent on local cataloguing and digitization practices in order to discover and get access to manuscripts which may be relevant for their research. In general, there is a lack of coherent, interoperable infrastructure for the data relating to these manuscripts, and the evidence base remains scattered across hundreds of resources.

Overcoming these difficulties and connecting up fragmented data and collections are crucial for enabling large-scale research. This article reports on some recent initiatives aimed at achieving these goals. It begins with a consideration of the kinds of research questions which are important in this field but which are difficult, if not impossible, to answer with the current scattered data. These are not limited to quantitative studies; they also encompass complex queries aimed at combining a range of different criteria and characteristics.

The following section examines the current plethora of digital resources and initiatives in the field of manuscript studies. Though there are applications for

working with dispersed digital images, there is a lack of interoperability more generally between different sources of data. The focus has been on library and museum collection descriptions and digitization programmes rather than developing interactive research platforms and services.

In the next section, I discuss a recent project designed to test some new approaches for bringing together collections data relating to medieval and Renaissance manuscripts. This project focused specifically on manuscripts formerly owned by Sir Thomas Phillipps but subsequently dispersed around the world, and examined ways of structuring and visualizing the evidence for the history and provenance of these manuscripts.

It became clear from the Phillipps project that an approach to combining data based more formally on Linked Open Data principles was required for larger-scale initiatives. The next section of the article looks at the potential of Linked Open Data for medieval studies and manuscript resources. There has been relatively little specific work to date, but a comparison with the field of classics and ancient history – and more generally, the activities of the Linked Open Data in Libraries Archives and Museums (LODLAM) network – shows what can be achieved.

I then look at a new project which is taking the lessons learned in the Phillipps project and applying them on a much larger scale in a Linked Open Data environment, with the goal of connecting up large amounts of data about medieval and Renaissance manuscripts and their histories. Projects like these depend heavily on the availability of suitable collections data from institutions which own such manuscripts. The final section of the article discusses the extent to which data are currently available for reuse, and makes recommendations for future action in this area.

Medieval and Renaissance Manuscripts Research Questions

There has been a significant increase in medieval and Renaissance manuscript research over the last twenty years, as documented in a recent survey by Da Rold and Maniaci (2015). Much of this work has focused on specific libraries, collectors, centers of scribal production, and scripts, as well as on specific authors and texts. There has

also been a continuing interest in more thematic research topics, concentrating – for example – on regional histories of manuscript production, social phenomena connected with manuscripts (such as rates of literacy), or artistic developments within illuminated manuscripts. But the many publications arising from this research – journal articles, monographs, and catalogues especially – are difficult to match with online data or digital versions of the specific manuscripts they discuss, and the data contained in them are difficult to aggregate. One important exception is the database maintained by the specialist journal *Scriptorium*, which indexes the manuscript shelf-marks referred to in articles.

There has been comparatively little large-scale research based on aggregated digital data relating to medieval and Renaissance manuscripts. An exception is Eltjo Buringh's work on global manuscript production (Buringh, 2011), which addresses several quantitative questions: how many medieval and Renaissance manuscripts survive worldwide? How many manuscripts were originally produced? What fraction of manuscripts has been lost? Buringh's methodology involves statistical extrapolations from a data sample of nearly thirty thousand manuscripts, drawn from information in printed catalogues. His database, which has not been made public, is global in scope and covers the period from the first to the nineteenth-centuries A.D.

Buringh's work – like earlier production estimates focused on specific regions (Neddermeyer 1998), specific languages (Bloom, 2001; Posner and Ta-Shema, 1975) and on codicology (Bozzolo and Ornato, 1983) – concentrates on purely quantitative calculations. Individual manuscripts are not of interest in themselves. Questions of ownership and provenance are not addressed. Nor is there any provision for new methods of exploring and reusing the aggregated data, especially through visualizations.

For most manuscript researchers, purely quantitative research is not their primary interest. Their focus is more likely to be on identifying and studying a group of manuscripts which fit a complex pattern of specific characteristics and properties. While they will want to know how many manuscripts fall into such a group, they are more interested in the specific manuscripts. This process of identifying a relevant group of manuscripts may involve several steps, such as these:

- A group of manuscripts in the same language, from the same region, produced in the same time period;
- Within this group, those which were produced in a particular type of institution or for a specific type of user;
- Within this narrower group, those which also have a similar script or similar illumination and decoration.

These selection criteria will result in a list of specific manuscripts which can be subjected to further exploration and analysis. A typical example is the work of Hanno Wijsman on illuminated manuscripts from the Burgundian Netherlands in the fifteenth and early sixteenth centuries (Wijsman, 2003 and 2010). This focuses on a particular type of manuscript from a specific region in a specific period, but aims at a comprehensive presentation of the evidence, both quantitative and individual. For most manuscript research, researchers must be able to move between an individual manuscript and more generalized and aggregated data. This reflects the wide range of research questions asked in the field of manuscript studies, but framed in the context of systematically aggregated data.

A research area of growing importance focuses on the histories of manuscripts as cultural objects in cultural collections. 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, depending on whether the collectors were individuals, private institutions, or modern public institutions. 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.

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 influence and social and political relationships in a particular society.

For manuscripts, this kind of research relies very heavily on evidence about their provenance – who owned them, in which places, and at which times (Pearson, 1998). This kind of data lends itself very well to visualization, as demonstrated by Mitch Fraas in his blog postings on the dispersal of the medieval libraries of Great Britain (Fraas, 2013) and the former owners of manuscripts now held by the University of Pennsylvania (Fraas, 2014).

Digital Resources for Manuscript Research

The growth in manuscript research has been accompanied by a proliferation of digital data relating to medieval and Renaissance manuscripts, not just in the form of catalogues, databases, and vocabularies, but also in digital editions, transcriptions, and – especially – in digital images. These developments have come on top of what was already a complex landscape of multiple printed sources, and have only increased that complexity. Large-scale analysis, for both quantitative and qualitative research questions, still requires time-consuming exploration of disparate sources and resources, including manuscript catalogues and databases of digitized manuscripts, as well as many forms of secondary literature.

As a result of this situation, research into the history and transmission of medieval and Renaissance manuscripts still involves a painstaking search through numerous printed and online catalogues and the associated secondary literature. There have been initiatives aimed at aggregating catalogue descriptions for manuscripts in Europe and North America, but these are mostly based on library catalogue records in Machine-Readable Catalogue (MARC) or Dublin Core format, which lack the data essential to address more sophisticated research questions. The CERL (Council for European Research Libraries) Portal, for example, provides a search across fourteen

different manuscript databases from Europe and North America with a combined total of about 380,000 records. But its search functionality is limited to keyword, person, subject, place, shelf-mark, and year.

Other aggregated databases are limited to digitized manuscripts. *Manuscriptorium*, hosted by the Národní knihovna České republiky (National Library of the Czech Republic), provides records for more than 4,000 medieval and Renaissance manuscripts from Eastern European libraries. *Digital Scriptorium* provides records for about 2,000 digitized manuscripts in North American libraries. The *e-codices* service offers digital copies of 1,885 manuscripts from 76 different libraries. The *Europeana* digital library includes at least 250,000 records for digitized manuscripts among its 53 million items, but medieval and Renaissance manuscripts cannot easily be separated out. The Europeana Data Model – unlike most other services – is based on Linked Data principles, and includes semantic enrichment from external Linked Data sources (Charles and Isaac, 2015). This resource is used mainly for refining searches within very broad categories.

It is important to note that only a small proportion of the surviving medieval and Renaissance manuscripts have actually been digitized, which significantly limits the scope and coverage of datasets restricted to digitized manuscripts. Fabian and Schreiber (2014: 3) estimate that only 7.5% of Germany's 60,000 medieval manuscripts have been digitized. Most researchers are interested in all relevant manuscripts, however, not just those which have already been digitized.

Other services provide a cross-search of manuscript-related data from various digital sources. *Manuscripts Online*, developed at the University of Sheffield, searches a mixture of descriptive and full-text resources, mostly from English sources. The *Medieval Electronic Scholarly Alliance* (MESA) is similar, enabling researchers to search across more than twenty catalogues and text collections from Europe and North America. Neither of these services enables researchers to address complex research questions nor to pursue large-scale quantitative investigations.

Other digital initiatives in the field of medieval and Renaissance manuscript studies have concentrated on developing tools for specific scholarly purposes.

DigiPal, developed at King's College London, combines digital images of medieval handwriting with detailed descriptions and characterizations of the script, and the associated text and manuscript, together with tools for annotating and manipulating the images (Brookes et al., 2015). Various tools for transcribing and annotating the texts of medieval and Renaissance manuscripts have also been developed, reflecting the high level of interest in transcriptions and editions generally; these are often encoded using the framework of the Text Encoding Initiative (Pierazzo, 2015). None of this work addresses comparative manuscript history and provenance, or supports large-analysis of manuscript data across multiple collections.

Another important recent development is the International Image Interoperability Framework (IIIF). Its potential applicability to manuscript research was noted at the Stanford Linked Data workshop in 2011, which referred to the need to 'develop the tools and agreements to support interoperability for scholarly functions across silos of digitized manuscripts' in the specific domain of digitized ancient, medieval, and early modern manuscripts (Keller et al., 2011: 46). The interoperability of digital images from different sites through an IIIF viewer, like *Mirador*, has considerable potential for manuscript studies. One important use case is to re-unite, in the correct order, images of manuscript leaves which have been dispersed in the course of their history and are now held in multiple institutions (Albritton, 2015; Davis, 2015). Another important use of IIIF is to compare different manuscripts which have similar characteristics, such as their scripts, their decoration, or their text.

The Shared Canvas data model, which provides the basis for the IIIF APIs (Application Programming Interfaces), focuses on describing the digital image. It 'does not directly address search services, or full bibliographic descriptions of the objects that are being rendered by the digital facsimiles', though it does provide hooks for pointing to related descriptions and may include summary metadata for agents, dates, and locations associated with the physical object (Sanderson and Albritton, 2013: section 5.5). For all its undoubted value in working with digital images, IIIF does not aggregate manuscript collections in a way which can support systematic large-scale research into their contents.

Combining Manuscript Data

In a research project funded by the European Union between 2014 and 2016, I investigated new ways of combining data relating to the history and provenance of medieval and Renaissance manuscripts (Burrows, 2017). I drew my sample from the manuscripts formerly owned by Sir Thomas Phillipps (1792–1872), a prominent English collector who owned the largest private collection ever assembled (Munby, 1951–1960). After his death, the collection was gradually dispersed, initially by his grandson Thomas Fitzroy Fenwick, who died in 1938, and subsequently by the booksellers W.H. Robinson Ltd. The final auctions devoted to Phillipps's manuscripts occurred at Sotheby's in the 1970s. The Phillipps manuscripts are now scattered across the world, in both institutional and private collections.

The Phillipps collection was notable not only for its sheer size (more than 40,000 items), but also because it contained numerous significant and valuable individual manuscripts and documents. Tracing the subsequent history and current locations of these objects is often important for specific research questions. The pattern of dispersal of the whole collection is also very important for analyzing the trade in manuscripts since the nineteenth century, and for what it reveals of the history of a major branch of collecting, which occupied many of the wealthiest North American collectors, including the Morgan and Getty families, as well as Henry Huntington and Henry Clay Folger (Burrows, 2016).

There are numerous sources of evidence about these manuscripts, ranging from handwritten inventories to Phillipps's own printed catalogue to current library catalogues and provenance databases, as shown in **Table 1**. The data models or schemas used to describe the manuscripts vary enormously; some contain no more than a number and a short title, while others consist of sophisticated structures which provide a detailed description and history of the manuscript. The *Schoenberg Database of Manuscripts* is the best example of the latter approach, designed as it is to record 'observations' of manuscripts at specific points in their history, taken from auction and sale catalogues and other sources.

Connecting up the data involved mapping these disparate models and schemas to a common data model. For the purposes of the project, I constructed my own

Table 1: Sources of evidence for the Phillipps collection.

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

data model which can readily be mapped to a larger ontology like CIDOC-CRM (Conceptual Reference Model) (Le Boeuf et al., 2015). My model focuses on recording events relating to manuscript histories and provenance: production, ownership, sale, donation and so on.

The second major step was then to transform the source data for ingesting to a new database. I used the *nodegoat* software developed by Lab1100 in the Netherlands, which enables data to be imported as CSV files or by manual entry into Web forms (Van Bree and Kessels, 2015). Data not already in digital form could

be input directly to *nodegoat* or captured in a CSV file. Data already in digital form needed to be exported from the data source, either as a CSV file or for subsequent transformation into a CSV format.

Some data sources were exemplary in enabling data export. The *Schoenberg Database* provides a daily CSV dump of the entire database, which can then be explored for Phillipps-related entries. Other sources were much more difficult to use (Burrows, 2016). Many library catalogues, in particular, make it difficult or even impossible to export groups of specific records in a suitable format. There often seems to be an assumption that the only reuse of catalogue records should be in the context of bibliographic management software like EndNote. The widespread adoption of software like Primo from Ex Libris has shifted the focus from searching the institutional collection to a much broader 'resource discovery' environment, with a concomitant emphasis on bibliographic referencing rather than object description.

As well as allowing a customized data model to be implemented, *nodegoat* also enables the assembled data to be visualized in different ways. The map-based visualization shows the geographical locations of various events in the history of a manuscript, and links those relating to the same manuscript. The time-slider makes it possible to view only those events occurring in a specific chronological period. The aggregated data for nearly 1,500 Phillipps manuscripts illustrates larger patterns in their combined histories, notably their movement from continental Europe to Great Britain in the nineteenth century, and their subsequent journey to North America in the twentieth century. The extent of their current dispersal is also revealed (**Figure 1**).

The other main *nodegoat* visualization shows the network graph of people, places, and sources of evidence relating to the manuscripts. This also has a time-slider. In the pilot version of the database, the Bodleian Library (at the University of Oxford) is disproportionately represented, skewing the network graph accordingly (**Figure 2**). Only relationships to manuscripts are shown; no attempt was made to incorporate relationships directly between people and institutions, though this would significantly enhance the analytical value of the graph if implemented.

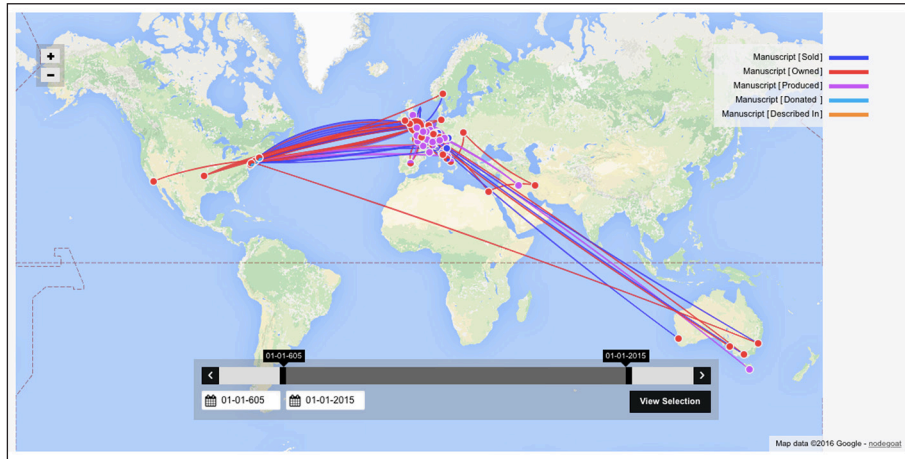


Figure 1: Nodegoat map-based visualization of the Phillipps dataset.

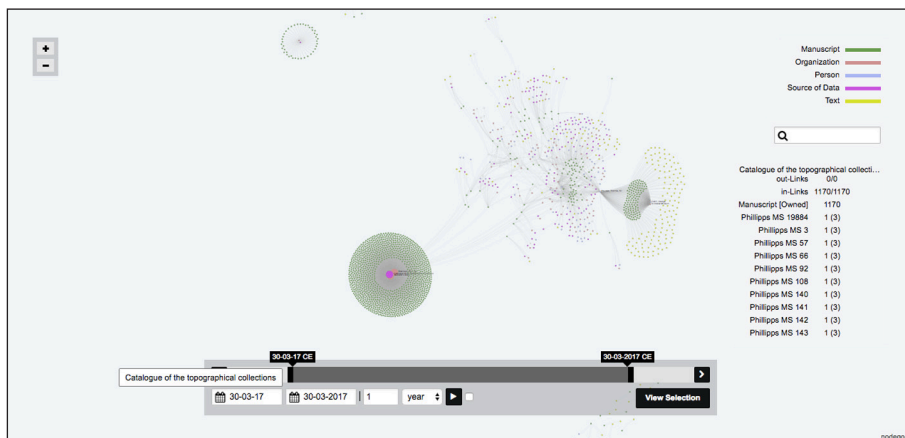


Figure 2: Nodegoat network visualization of the Phillipps dataset.

The Phillipps project served as a proof-of-concept demonstration of the value of modeling and combining data relating to the history and provenance of manuscripts. As well as showing the possibilities for visualizing the combined data, it demonstrated how sub-sets of the data could be selected and analyzed, such as the group of manuscripts which had belonged to Phillipps and to another important collector such as Alfred Chester Beatty, as this example shows (**Figure 3**). It went some way towards reconstructing the now-dispersed Phillipps collection in a virtual,

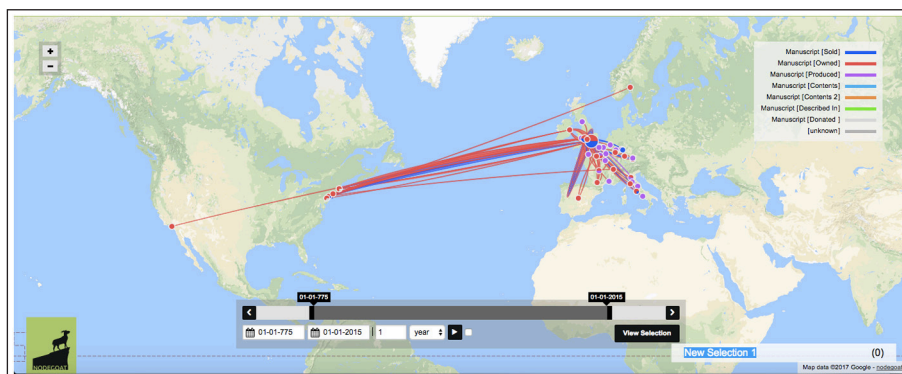


Figure 3: Nodegoat map-based visualization of Phillipps-Beatty manuscripts.

digital setting. But it also revealed that a more formal Linked Open Data framework would be required if manuscript-related data were to be combined on a larger scale and in a more sustainable way.

Linked Open Data for Medieval and Renaissance Studies

Despite all the activity in developing digital resources and services, there is little evidence of the use of Linked Open Data applications for combining data in medieval and Renaissance studies. This is despite considerable work in this area from the libraries, archives and museums community more generally, often gathered under the rubric of LODLAM (EuropeanaTech, 2017). A tool like DIVE+ connects four Dutch cultural heritage collections with a total of 350,000 objects, resulting in a knowledge graph of more than 15 million RDF (Resource Description Framework) triples (de Boer, 2017). The Linked Open Data initiative of the American Art Collective has published RDF versions of more than 230,000 museum objects from the collections of fourteen United States institutions, accompanied by recommendations for best practice in museums (Fink, 2018). There have also been exemplary projects to transform specific canonical catalogues into Linked Open Data, such as Bernard Berenson's *Drawings of the Florentine Painters*, from Harvard University (Klic, Miller, Nelson, Pattuelli and Provo, 2017).

Linked Open Data approaches have also been developed and implemented systematically in humanities fields like Classics and Ancient History (Isaksen et al.,

2014), where the Pleiades community-built gazetteer of ancient places forms the basis for a collaborative annotation environment – known as Pelagios – for linking place references to texts and images. This ‘Graph of the Ancient World’ is built on Uniform Resource Identifiers (URIs) for ancient places and the Open Annotation ontology. The Perseus Digital Library is a long-standing effort in the same field to bring together a digital library of classical texts, using unique identifiers for texts and sections within them (Almas, Babeu, and Krohn, 2014). A third project, “Integrating Digital Epigraphies”, is using shared identifiers to link a variety of different resources relating to Greek epigraphy (Löser, 2015). Researchers are able to find related resources linked by the same identifier, including secondary literature published in journals available through JSTOR. In the field of numismatics, the Nomisma service now provides Linked Open Data digital representations of numismatic concepts, drawing on 33 different datasets (Wigg-Wolf and Duyrat, 2017).

This kind of framework for linking disparate resources is lacking for medieval and Renaissance studies. In particular, there are no unique identifiers for manuscripts and no reliable way to link resources relating to the same manuscript. Library shelf-marks for manuscripts are not an adequate substitute, since they are often cited inconsistently in different sources and contexts. The importance of this need for manuscript identifiers has been noted in a recent report by Fabian and Schreiber (2014: 12):

Another prerequisite will be to establish manuscripts themselves, that is, unique artefacts, as clearly identifiable entities, namely individual works of art, by assigning them a unique identifier or standard number useful in the semantic web environment.

In this context, the International Standard Manuscript Identifier (ISMI) initiative is particularly promising. First discussed at a meeting of manuscript librarians and researchers in April 2017, this proposal envisages the creation of a prototype for assigning unique identifying numbers for medieval and Renaissance manuscripts, independent of library shelf-marks. The initial idea is that the Institut de recherche

et d'histoire des textes (IRHT) in Paris will explore the possibility of creating and recording these identifiers in the *Medium* database, which contains brief records for more than 109,000 manuscripts.

Fabian and Schreiber (2014) also point to the difficulties caused by variant titles for medieval works, previously documented in detail by Sharpe (2003). Their report on plans for a national digitization program for Germany calls for the standardization of subject terminology (Fabian and Schreiber, 2014: 12):

While persons' names have long been subject to and integrated into national and international authority files, work on the standardization of other entities has only just begun. Thesauri or classified and ideally hierarchically grouped lists of terms should be developed for the most important concepts or subjects in the disciplines of paleography, codicology and art history.

They note the 'current fragmentation of information in a plethora of local digitization projects' and call for 'a unitary and complete virtual research environment for all those interested in medieval manuscripts in German collections and beyond' (Fabian and Schreiber, 2014: 3).

A Linked Open Data framework offers an alternative to standardization of the kind envisaged by Fabian and Schreiber. This alternative approach was proposed in a Road Map developed as a result of a European Science Foundation workshop organized by the CARMEN Medieval Manuscripts Research Group in 2009 (Scase, 2009). This Road Map put forward ways of identifying and linking entities of various kinds occurring in or connected with medieval manuscripts, and mapping between the different frameworks for describing and representing manuscripts. At a more recent meeting, the 'Linking the Middle Ages' workshop held at the University of Texas Austin in May 2015, the work being done in Classics was also discussed but no specific proposals for linking data about manuscripts were put forward (Turnator et al., 2015).

The *Biblissima* observatory in France experimented with a Linked Open Data approach for aggregating data relating to manuscripts and early printed books from

a range of French databases (Gehrke et al., 2015). An initial prototype provided federated access to a subset of data from two iconographic datasets: *Mandragore* (from the Bibliothèque nationale de France) and *Initiale* (from the Institut de recherche et d'histoire des textes). The ontological structure was provided by a combination of CIDOC-CRM and FRBRoo (Functional Requirements for Bibliographic Records – object-oriented). While the subset of data was relatively small (5,000 descriptors for 20,000 manuscript illuminations), the prototype gave some idea of what is possible by way of searching and visualization. The full *Biblissima* service took a different approach, using Pivot-XML to aggregate data for searching – rather than a large-scale Linked Open Data framework.

Next Steps and Future Directions

Taking the lessons from the *Phillipps* project and applying them in a Linked Open Data environment on a much broader scale is the aim of the *Mapping Manuscript Migrations* project, funded by the Trans-Atlantic Platform under its *Digging into Data Challenge* from 2017 to 2019. Led by the University of Oxford through the Oxford e-Research Centre and the Bodleian Libraries, this new project includes partners at Aalto University (Semantic Computing Group), the University of Pennsylvania (Schoenberg Institute for Manuscript Studies), and the Institut de recherche et d'histoire des textes (IRHT) in Paris.

Mapping Manuscript Migrations is building a coherent framework for linking manuscript-related data from different sources, using Linked Data principles. It began with the *Schoenberg Database of Manuscripts* and the new TEI (Text Encoding Initiative)-based *Medieval Manuscripts in Oxford Libraries* catalogue of the Bodleian Libraries and will be adding the *Bibale* and *Medium* databases of the IRHT. These are being mapped to a common data model, which focuses on event-based data for manuscript histories and provenance. A software platform for enabling searchable and browsable semantic access is being implemented across the linked data. The project is drawing on previous work at Aalto University on publishing and brokering Linked Open Data for cultural heritage datasets (Hyvönen et al., 2014 and 2016).

The purpose of linking and surfacing the disparate data in this environment is to make possible large-scale analysis and visualization of the history and movement

of these manuscripts over the centuries. This includes such research questions as: how many manuscripts have survived?; where they are now?; and, which people and institutions have been involved in their history? More specific analyses focused on manuscripts of particular types or from certain places, or on specific collectors and collections, will also be possible. These are very difficult – if not impossible – to carry out in the present fragmented landscape of manuscript-related data. By combining collections data in this way, the project is aiming to overcome these existing limitations.

Mapping Manuscript Migrations is built on Linked Open Data principles. The data model draws on the CIDOC-CRM and FRBRoo ontologies, and the source datasets are transformed into RDF and housed in a triple store. Entities such as persons and places are linked to their occurrences in other Linked Open Data sources, including Virtual International Authority File (VIAF), GeoNames and the Union List of Artists' Names (ULAN). The software used to explore the combined data are capable of constructing complex queries based on SPARQL (SPARQL Protocol and RDF Query Language). The aim is to demonstrate the effectiveness of this kind of approach for combining diverse data relating to medieval and Renaissance manuscripts without losing their inherent richness.

Making Data Available

Manuscripts are an important form of evidence for medieval and Renaissance studies, and there are numerous sources of data relating to them. But the present, highly fragmented, environment makes large-scale, comprehensive research extremely difficult. Combining collections of data about manuscripts has significant value for researchers, and the projects described here are designed to implement and evaluate methods for putting this into practice.

Projects like these rely on the availability of data relating to the history and provenance of medieval and Renaissance manuscripts. Surprisingly few cultural heritage institutions make this kind of data readily available for computational reuse. This is partly because of the limitations of metadata schemas like MARC and Dublin Core, but also because few of the relevant databases are set up to enable systematic downloads or programmatic access using APIs.

Institutional commitment to sharing data about manuscript collections in a reusable form is crucial. This kind of commitment has already been demonstrated in sharing digital images of manuscripts through the IIIF approach. What is needed now, as a first step, is the availability of manuscript records as discrete datasets, exportable in standard formats like CSV and XML. Providing Linked Data versions of manuscript data for reuse, though more ambitious, would also be very valuable. Implementing a unique, shareable International Standard Manuscript Identifier will be an important component.

Above all, institutions should recognize that collections data relating to medieval and Renaissance manuscripts are of the highest value to researchers across many disciplines. Re-thinking the ways in which manuscript data are structured and disseminated will make a huge contribution to enabling new kinds of research and taking research to new levels. It will also contribute significantly to demonstrating the value of these unique assets within the broader landscape of library and museum collections. Perhaps Ernest C. Richardson's dream (1937: 8) of a 'union world catalog' of manuscript books will finally be within reach, after more than eighty years.

Competing Interests

The author has no competing interests to declare.

Author Informations

Dr Toby Burrows is a Senior Researcher in the Oxford e-Research Centre at Oxford University and a Senior Honorary Research Fellow in the School of Humanities at the University of Western Australia. His research interests focus on the history of collections and collecting, especially in relation to medieval and Renaissance manuscripts. He is particularly interested in the digital humanities and the application of digital methodologies to historical research.

He is a Principal Investigator on the project, Mapping Manuscript Migrations funded by the Trans-Atlantic Platform under its Digging into Data Challenge, as well as on the project Collecting the West, funded by the Australian Research Council. Between 2014 and 2016, he held a Marie Curie International Incoming Fellowship in the Department of Digital Humanities at King's College London. He has also held

visiting fellowships at the University of Pennsylvania, the Royal Netherlands Academy of Arts and Sciences, the Grolier Club of New York, and Churchill College, Cambridge.

References

- Albritton, B** 2015 Fellow Travellers: The Canterbury Tales and IIF. *Digital Manuscripts at Stanford* (blog). Available at: <http://web.stanford.edu/group/dmstech/cgi-bin/wordpress/fellow-travelers-the-canterbury-tales-and-iif/> [Last accessed 15 October 2018].
- Almas, B, Babeu, A and Krohn, A** 2014 Linked Data in the Perseus Digital Library. In: *ISAW Papers 7: Current Practice in Linked Open Data for the Ancient World*. New York: Institute for the Study of the Ancient World, New York University. Available at: <http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/almas-babeu-krohn/> [Last accessed 15 October 2018].
- Bloom, J M** 2001 *Paper before Print: The History and Impact of Paper in the Islamic World*. New Haven: Yale University Press.
- Bozzolo, C and Ornato, E** 1983 *Pour une histoire du livre manuscrit au Moyen Age: Trois essais de codicologie quantitative*. Paris: Éditions du Centre national de la Recherche scientifique.
- Brookes, S, Matos, D, Watson, M and Stokes, P A** 2015 The DigiPal Project for European Scripts and Decorations. In: *Writing Europe, 500–1450: Texts and Contexts*, Conti, A, Da Rold, O and Shaw, P (Eds.), pp. 25–58. Cambridge: D. S. Brewer.
- Buringh, E** 2011 *Medieval Manuscript Production in the Latin West: Explorations with a Global Database*. Leiden: Brill.
- Burrows, T** 2016 Manuscripts of Sir Thomas Phillipps in North American Institutions. *Manuscript Studies*, 2: 307–327. DOI: <https://doi.org/10.1353/mns.2016.0019>
- Burrows, T** 2017 The History and Provenance of Manuscripts in the Collection of Sir Thomas Phillipps: New Approaches to Digital Representation. *Speculum*, 92(S1): S39–S64. DOI: <https://doi.org/10.1086/693438>
- Charles, V and Isaac, A** 2015 *Enhancing the Europeana Data Model (EDM)*. The Hague: Europeana Foundation.

- Da Rold, O** and **Maniaci, M** 2015 Medieval Manuscript Studies: A European Perspective. In: *Writing Europe, 500–1450: Texts and Contexts*, Conti, A, Da Rold, O and Shaw, O (Eds.), pp. 1–24. Cambridge: D. S. Brewer.
- Davis, L** 2015 The Promise of Digital Fragmentology. *Manuscript Road Trip* (blog). Available at: <https://manuscriptroadtrip.wordpress.com/2015/07/13/manuscript-road-trip-the-promise-of-digital-fragmentology/> [Last accessed 15 October 2018].
- de Boer, V, Inel, O, Aroyo, L, van den Akker, C, Legene, S, Martinez, C, Helmich, W, Hagendoorn, B, Sauer, S, Blom, J, Melgar, L and Oomen, J** 2017 DIVE+: Exploring Integrated Linked Media. *EuropeanaTech Insight* 7. Available at: <https://pro.europeana.eu/page/issue-7-lodlam> [Last accessed 15 October 2018].
- EuropeanaTech** 2017 LODLAM. *EuropeanaTech Insight* 7. Available at: <https://pro.europeana.eu/page/issue-7-lodlam> [Last accessed 15 October 2018].
- Fabian, C** and **Schreiber, C** 2014 Piloting a National Programme for the Digitization of Medieval Manuscripts in Germany. *LIBER Quarterly*, 24(1): 2–16. DOI: <https://doi.org/10.18352/lq.9265>
- Fink, E** 2018 *American Art Collaborative (AAC) Linked Open Data (LOD) Initiative: Overview and Recommendations for Good Practices*. Available at: http://americanartcollaborative.org/wp-content/uploads/2018/03/AAC_LOD_Overview_Recommendations.pdf [Last accessed 15 October 2018].
- Fraas, M** 2013 The dispersal of the medieval libraries of Great Britain. *Mapping Books* (blog). Available at: <http://mappingbooks.blogspot.com/2013/11/the-dispersal-of-medieval-libraries-of.html> [Last accessed 15 October 2018].
- Fraas, M** 2014 Charting former owners of Penn's codex manuscripts. *Mapping Books* (blog). Available at: <http://mappingbooks.blogspot.com/2014/01/charting-former-owners-of-penns-codex.html> [Last accessed 15 October 2018].
- Gehrke, S, Frunzeanu, E, Charbonnier, P and Muffat, M** 2015 Biblissima's prototype on medieval manuscript illuminations and their context. In: *SW4SH 2015: Semantic Web for Scientific Heritage 2015: Proceedings of the First International Workshop Semantic Web for Scientific Heritage at the 12th ESWC*

2015 Conference, Portorož, Slovenia, June 1st, 2015, Zucker, A, Draelants, Zucker, CF and Monnin, A (Eds.). Available at: <http://ceur-ws.org/Vol-1364/paper5.pdf> [Last accessed 15 October 2018].

Hyvönen, E, Ikkala, E and Tuominen, J 2016 Linked Data Brokering Service for Historical Places and Maps. In: *Proceedings of the 1st Workshop on Humanities in the Semantic Web – WHiSe, Heraklion, Crete, Greece, May, 2016*. <http://seco.cs.aalto.fi/publications/2016/hyvonen-et-al-hipla.pdf> [Last accessed 15 October 2018].

Hyvönen, E, Tuominen, J, Alonen, M and Mäkelä, E 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*, pp. 226–230. Berlin: Springer-Verlag.

Isaksen, L, Barker, E, Simon, R and de Soto Cañamares, P 2014 Pelagios and the Emerging Graph of Ancient World Data. In: *WebSci '14: Proceedings of the 2014 ACM Conference on Web Science* (New York), pp. 197–201. DOI: <https://doi.org/10.1145/2615569.2615693>

Keller, M A, Persons, J, Glaser, H and Calte, M 2011 *Report on the Stanford Linked Data Workshop* (Washington, DC: Council on Library and Information Resources). <https://www.clir.org/pubs/reports/pub152/stanford-linked-data-workshop/> [Last accessed 15 October 2018].

Klic, L, Miller, M, Nelson, J K, Pattuelli, C and Provo, A 2017 The Drawings of the Florentine Painters: From Print Catalog to Linked Open Data. *code4lib journal*, 38. Available at: <http://journal.code4lib.org/articles/12902> [Last accessed 15 October 2018].

Le Boeuf, P, Doerr, M, Ore, C E and Stead, S (Eds.) 2015 *Definition of the CIDOC Conceptual Reference Model*, version 6.2.1. Available at: http://www.cidoc-crm.org/sites/default/files/cidoc_crm_version_6.2.1.pdf [Last accessed 15 October 2018].

Liu, A 2012 Remembering Networks: Agrippa, RoSE, and Network Archaeology. *Paper presented at Network Archaeology conference*. Miami University, Ohio, 21 April.

- Löser, L** 2015 *The March of IDEs: How Integrating Digital Epigraphies Could Facilitate Scholarly Use of Epigraphic Material*. Unpublished dissertation, University of St Andrews.
- Munby, A N L** 1951–60 *Phillipps Studies*, 5. Cambridge: Cambridge University Press.
- Neddermeyer, U** 1998 *Von der Handschrift zum gedruckten Buch: Schriftlichkeit und Leseinteresse im Mittelalter und in der frühen Neuzeit: Quantitative und qualitative Aspekte*. Weisbaden: Harrassowitz.
- Pearson, D** 1998 *Provenance Research in Book History: A Handbook*. London: British Library.
- Pierazzo, E** 2015 *Digital Scholarly Editing: Theories, Models and Methods*. London: Routledge.
- Posner, R** and **Ta-Shema, I M** 1975 *The Hebrew Book: An Historical Survey*. Jerusalem: Keter.
- Richardson, E C** 1937 *A union world catalog of manuscript books, preliminary studies in method. VI, Summary of method*. New York: H. W. Wilson.
- Sanderson, R** and **Albritton, B** 2013 *Shared Canvas Data Model*. Document prepared for the International Image Interoperability Framework group. Available at: <http://iiif.io/model/shared-canvas/1.0/> [Last accessed 15 October 2018].
- Scase, W** 2009 *Applying Semantic Web Technologies to Medieval Manuscript Research: European Science Foundation Exploratory Workshop Report*. Strasbourg: European Science Foundation.
- Sharpe, R** 2003 *Titulus: Identifying Medieval Latin Texts: An Evidence-based Approach*. Turnhout: Brepols.
- Turnator, E**, **Bolinteanu, A**, **Rose-Steel, T**, **Whearty, B** and **Widner, M** 2015 *Summary of Proceedings of the "Linking the Middle Ages" Workshop (May 11–12, 2015) at the University of Texas at Austin*. Available at: <https://repositories.lib.utexas.edu/handle/2152/30220> [Last accessed 15 October 2018].
- Van Bree, P** and **Kessels, G** 2015 Mapping memory landscapes in nodegoat. In: *Social Informatics*, Aiello, L M and McFarland, D (Eds.), pp. 274–278. Berlin: Springer International Publishing.

Wigg-Wolf, D and **Duyrat, F** La révolution des Linked Open Data en numismatique:

Les exemples de nomisma.org et Online Greek Coinage. *Archéologies numériques*, 1(1). [Last accessed 15 October 2018]. DOI: <https://doi.org/10.21494/ISTE.OP.2017.0171>

Wijsman, H 2003 Manuscrits illustrés dans les Pays-Bas bourguignons: Quelques remarques quantitatives. *Gazette du livre médiéval*, 43: 23–33.

Wijsman, H 2010 *Luxury Bound: Illustrated Manuscript Production and Noble and Princely Book Ownership in the Burgundian Netherlands (1400–1550)*. Turnhout: Brepols. DOI: <https://doi.org/10.1484/M.BURG-EB.5.105851>

Websites

Bibale <http://bibale.irht.cnrs.fr/>.

CERL Portal https://www.cerl.org/resources/cerl_portal.

Digital Scriptorium <http://www.digital-scriptorium.org/>.

e-codices <http://www.e-codices.unifr.ch/en>.

Europeana <http://www.europeana.eu/portal/en>.

International Image Interoperability Framework (IIIF) <http://iiif.io/>.

Manuscriptorium <http://www.manuscriptorium.com/>.

Manuscripts Online <https://www.manuscriptsonline.org/>.

MESA (Medieval Electronic Scholarly Alliance) <http://www.mesa-medieval.org/>.

Medieval Manuscripts in Oxford Libraries <https://medieval.bodleian.ox.ac.uk>.

Medium <http://medium-avance.irht.cnrs.fr/>.

Schoenberg Database of Manuscripts <https://sdbm.library.upenn.edu/>.

How to cite this article: Burrows, T 2018 Connecting Medieval and Renaissance Manuscript Collections. *Open Library of Humanities*, 4(2): 32, pp.1–23. DOI: <https://doi.org/10.16995/olh.269>

Published: 02 November 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Open Library of Humanities is a peer-reviewed open access journal published by Open Library of Humanities.

OPEN ACCESS