Schriften des Instituts für Dokumentologie und Editorik — Band 3

Kodikologie und Paläographie im digitalen Zeitalter 2

Codicology and Palaeography in the Digital Age 2

herausgegeben von | edited by

Franz Fischer, Christiane Fritze, Georg Vogeler

unter Mitarbeit von | in collaboration with

Bernhard Assmann, Malte Rehbein, Patrick Sahle

2010
BoD, Norderstedt

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über **http://dnb.d-nb.de/** abrufbar.

Online-Fassung
Herstellung und Verlag der Druckfassung: Books on Demand GmbH, Norderstedt 2010
ISBN: 978-3-8423-5032-8
Einbandgestaltung: Johanna Puhl, basierend auf dem Entwurf von Katharina Weber
Satz: Stefanie Mayer und LaTeX

Applying Semantic Web Technologies to Medieval Manuscript Research

Toby Burrows

Abstract

Medieval manuscript research is a complex, fragmented, multilingual field of knowledge, which is difficult to navigate, analyse and exploit. Though printed sources are still of great importance and value to researchers, there are now many services on the Web, some commercial and many in the public domain. At present, these services have to be consulted separately and individually. They employ a range of different descriptive standards and vocabularies, and use a variety of technologies to make their information available on the Web. This chapter proposes a new approach to organizing the international collaborative infrastructure for interlinking knowledge and research about medieval European manuscripts, based on technologies associated with the Semantic Web and the Linked Data movement. This collaborative infrastructure will be an open space on the Web where information about medieval manuscripts can be shared, stored, exchanged and updated for research purposes. It will be possible to ask large-scale research questions across the virtual global manuscript collection, in a quicker and more effective way than has ever been feasible in the past. The proposed infrastructure will focus on building links between data and will provide the basis for new kinds of services which exploit these data. It will not aim to impose a single metadata standard on existing manuscript services, but will build on existing databases and vocabularies. The article describes the architecture, services and data which will comprise this infrastructure, and discusses strategies for making th challenging and exciting goal a reality.

Zusammenfassung

Die Handschriftenforschung ist ein umfangreiches, fragmentiertes und multilinguales Wissensgebiet, das nicht leicht zu überblicken, zu analysieren oder auszuschöpfen ist. Obwohl gedruckte Quellen für die Forschung weiterhin von großer Bedeutung und Nutzen sind, gibt es inzwischen viele Angebote im Web, teils kommerzieller Natur, teils frei zugänglich. Im Augenblick können diese Angebote jedoch nur isoliert voneinander in Anspruch genommen werden. Sie fußen auf einer Reihe verschiedener Beschreibungsstandards und Vokabularien und nutzen eine Bandbreite

Kodikologie und Paläographie im Digitalen Zeitalter 2 – Codicology and Palaeography in the Digital Age 2. Hrsg. Franz Fischer, Christiane Fritze, Georg Vogeler, unter Mitarbeit von Bernhard Assmann, Malte Rehbein, Patrick Sahle. Schriften des Instituts für Dokumentologie und Editorik 3. Norderstedt: Books on Demand, 2010. 117–131.

unterschiedlicher Technologien, um ihre Informationen im Web zur Verfügung zu stellen. Dieser Beitrag beschreibt einen neuen Ansatz, um Wissen und Forschung über mittelalterliche europäische Handschriften zusammenzuführen: eine internationale, kollaborative Infrastruktur, die auf den Technologien aus den Bereichen "Semantic Web" und "Linked Data" aufbaut. Daraus entsteht im Web ein "open space", in dem Informationen zu mittelalterlichen Handschriften gespeichert, bereitgestellt, ausgetauscht und aktualisiert werden können. Damit wird systematische Forschung auf Grundlage einer globalen, virtuellen Handschriftensammlung schneller und zielführender als bislang denkbar. Die vorgeschlagene Infrastruktur konzentriert sich auf die Verknüpfung von Daten, und sie wird die Grundlage für neuartige Angebote bilden, um mit diesen Daten zu arbeiten. Sie zielt nicht darauf, einen bestimmten Standard für Metadaten durchzusetzen, sondern baut auf den bestehenden Datenbanken und Vokabularien auf. Der Beitrag beschreibt die Infrastruktur im Hinblick auf Systemaufbau, Services und Daten und erörtert Strategien zu ihrer Realisierung.

1. Introduction

Among the greatest treasures in the cultural heritage of Europe are its medieval manuscripts. Many hundreds of thousands still survive today, in collections around the world, where they are intensively studied by researchers and admired by visitors to libraries, museums and art galleries. There are numerous Web sites and projects devoted to medieval manuscripts, originating both from research groups working on these manuscripts and from the cultural heritage institutions in which they are held. In many ways, medievalists have been at the forefront of the application of digital technologies to research in the humanities.

Despite their undoubted value, these Web sites and services suffer from several major limitations. There is a lack of integration and interoperability between the many different sites, and it is difficult to find out systematically what research and digitization are being undertaken in collections around the world. Most of these services have to be consulted separately and individually, though search engines like Google cover some of them. The use of terminology and of descriptive standards is inconsistent and unsatisfactory, especially across different European languages and cultures. There is often no easy way of connecting the descriptions of the manuscripts with the reports of the research which has been based on them. As a result, researchers around the world still face major difficulties in finding, using and sharing knowledge about medieval manuscript collections.

This paper presents a proposal for a new international collaborative infrastructure for organizing and interlinking knowledge and research about medieval European manuscripts, based on technologies associated with the Semantic Web and the

Linked Open Data movement. The proposal draws on a Road Map developed for a European Science Foundation Exploratory Workshop held at the University of Birmingham (United Kingdom) in March 2009. Convened by Wendy Scase (University of Birmingham) and Orietta Da Rold (University of Leicester), this Workshop brought together specialists from the fields of manuscript studies, librarianship, and information and computer science, as well as from a range of public and commercial institutions and organizations.

The Exploratory Workshop was organized by the Medieval Manuscript Research Group of CARMEN, the Co-operative for the Advancement of Research through a Medieval European Network. CARMEN includes representatives of the many European centres for medieval research, as well as professional associations, cultural institutions and commercial publishing companies with expertise in this field. Participants from outside Europe include the Australian Research Council's Network for Early European Research (NEER) and several North American associations and institutions. A large proportion of the global research community in this field (estimated at more than 14,000 researchers) is represented in CARMEN.

2. Medieval Manuscripts: Research Questions

Medieval manuscripts are used in addressing a wide range of research questions. Most obviously, these include **research into the characteristics of manuscripts themselves as physical objects**. These characteristics include: the place of origin, the date or period of origin, the materials used, the decoration and illumination, the handwriting, the scribe, the binding, arrangement of the physical volume, and the language. Research into the subsequent history of a manuscript looks at its owners and at changes to its appearance over time, as well as at its modern location, and its place in modern collections.

Relationships between manuscripts are a common topic, including research which reunites dispersed leaves of what was originally a single manuscript. Identifying connections between specific medieval manuscripts and other materials which survive from the medieval period (especially art works, buildings, and other material objects) is another closely related area of research.

Defining these different physical characteristics also forms the starting-point of many research projects. These include such investigations as defining specific styles of handwriting, establishing different categories of decoration, and identifying different ways of constructing and creating physical volumes in the medieval period.

The second major area of research involves **the use of manuscripts as evidence for all aspects of life in the medieval period**. This requires knowledge and understanding of the *content* of a manuscript—the text, the images, the music, and so on. This kind of

research is heavily dependent on the descriptors used to identify the content, including authors' names, titles of works, incipits, subject and concept terms, and so on.

Both of these two major areas of research also draw on an extensive body of secondary literature relating to specific manuscripts: catalogues and descriptions (both medieval and modern), bibliographies, and secondary works. These are likely to reflect changes over time—as concepts shift, and descriptions and attributions are revised. All aspects of the body of knowledge in this field are multilingual; descriptions and descriptors may be in a variety of languages, mainly European.

Fundamental to both types of research projects is the availability of detailed descriptions of manuscripts as physical objects. In the world of printed catalogues, there are numerous different formats for such descriptions. In the digital world, a similar multiplicity of formats was already evident as long ago as the late 1980s and early 1990s. The major change since then has been the emergence of the Manuscript Description element (<msDesc>) and chapter (10. Manuscript Description) within the encoding guidelines of the Text Encoding Initiative (TEI), which is being increasingly regarded as a de facto standard. The North American Association of College and Research Libraries has also published rules for the *Descriptive Cataloging of Ancient, Medieval, Renaissance, and Early Modern Manuscripts (Pass).*

3. Web Resources for Medieval Manuscript Research

Medieval manuscript research is a complex, fragmented, multilingual field of knowledge, which is difficult to navigate, analyse and exploit. Though printed sources are still of great importance and value to researchers, there are now many services on the Web, some commercial and many in the public domain. The number of such services is hard to estimate, but listings of the major authoritative sites (such as those in the Labyrinth and the Virtual Library) contain at least fifty entries. These services employ a range of different descriptive standards and vocabularies, and use a variety of different technologies to make their information available on the Web.

Numerous collecting institutions provide information about the manuscripts they hold, either as part of more general databases or as specific manuscript databases. Examples of the latter approach include the British Library's Manuscripts Catalogue and the Codices Electronici Sangallenses (CESG) of the Abbey Library of St Gall. There are also a range of national databases, including Medieval Manuscripts in Dutch Collections (hosted by the Koninklijke Bibliotheek), MEDIUM (hosted by the Institut de recherche et d'histoire des textes), and the Digital Scriptorium, hosted by Columbia University and the University of California Berkeley. The Europa Inventa project has developed a national database for Australia.

The small number of international databases are particularly important in the context of this paper. The CERL (Council of European Research Libraries) Portal provides a union catalogue of manuscript descriptions, harvested from sites distributed across Europe, North America and Australia. The Manuscriptorium service, hosted by the Národní knihovna České republiky, contains descriptive records contributed by more than ninety institutions across Europe.

Some of these services provide digital images of some manuscripts as well as descriptive information about them. Others focus specifically on providing digital versions. The European digital library service Europeana, for instance, only includes digitized materials, though its scope is much broader than medieval manuscripts. Its content is being enhanced by a new project, Europeana Regia, which aims to digitize a corpus of illuminated royal manuscripts from France, Belgium, Germany and Spain. Large-scale digitization projects are underway in several major libraries, including the Bayerische Staatsbibliothek and the Biblioteca Apostolica Vaticana.

There are many Web sites which list, transcribe, or provide digital images of manuscripts of a specific text or relating to a specific medieval author. Two examples are the site devoted to Dante's *Divina Commedia*, maintained by the Società Dantesca Italiana, and the interrelated group of sites focusing on *Le Chevalier de la Charrette* by Chrétien de Troyes, hosted by Princeton University, Baylor University and the Université de Poitiers.

Ancillary Web services include sites devoted to manuscript terminology and vocabularies, incipits, subjects, authors, and people more generally. Some of these sites are commercial, notably the various databases managed by Brepols Publishers NV in Belgium. These include In Principio (incipits), the International Medieval Bibliography (IMB) with its subject thesaurus and lists of authors, and Europa Sacra (medieval prelates). Many other Web resources of this kind are hosted and maintained by cultural institutions. The CERL Thesaurus is an extensive collection of place names and personal names. The Medieval Manuscripts in Dutch Collections service provides its own list of authors. Personennamen des Mittelalters is hosted by the Deutsche Nationalbibliothek as part of its larger database of personal names (Personennamendatei). The Fasti Ecclesiae Anglicanae 1066–1300 focuses on English church prelates. Also important is the Web version of Denis Muzerelle's Vocabulaire codicologique, hosted by the IRHT. In Italy, the SISMEL research group publishes two major databases of medieval authors' names through its Mirabile service: Bibliotheca Scriptorum Latinorum Medii Recentiorisque Aevi (BISLAM) and Compendium Auctorum Latinorum Medii Aevi (CALMA).

Other services provide indexes to references to specific manuscripts in journal articles, scholarly books and other secondary literature. They include the indexes to the journal Scriptorium and the manuscripts index within the International Medieval Bibliography.

This extraordinary—and continually growing—collection of riches brings its own set of complications and difficulties, and leaves us with something resembling the Tower of Babel. There are simply too many sources of information, with manuscript descriptions in different formats and multiple languages, and highly variable uses of names, titles and concepts. It is very difficult to trace systematically the relationships between different manifestations of the same manuscript (e.g., digital images, transcriptions and editions). It can also be very difficult to trace relationships between manuscripts and what is written about them (e.g., articles, books and commentaries).

While the proliferation of Web resources is undoubtedly of great value to manuscript research, their sheer complexity and variety impose a significant barrier to our ability to ask large-scale research questions—both about manuscripts as physical objects and about their content. Trying to impose rigorous standards and continuing to pursue uniformity, in order to resolve these difficulties, is unlikely to be successful. Expecting Google—or some future global search engine—to provide a solution through keyword searching is equally unrealistic. A different kind of approach is required.

4. Semantic Web Technologies and the Linked Data Movement

If a serious effort is to be made to overcome the complexities of the current digital landscape, the most promising approach appears to lie in what are somewhat misleadingly called Semantic Web technologies. The underlying idea of the Semantic Web is to implement methods of making digital content available for processing by Web-based software in a more effective way, though its originator, Tim Berners-Lee, now prefers to use the term "Web of data", to emphasize the data-oriented nature of this framework. The primary purpose of these developments is to enable knowledge to be found, shared and interlinked more easily.

Semantic Web technologies are methods for adding semantic structures to Web data and documents, with the broad aim of making them more interoperable and automatically discoverable (Antoniou and van Harmelen). The main building-blocks for this process are as follows:

- **Object identifiers**: alpha-numeric addresses such as URIs (Uniform Resource Identifiers) which can be used to identify an object uniquely. It is possible to assign identifiers to abstract "objects" like concepts, subject terms, personal names and place names, as well as to physical objects like manuscripts.
- **Ontologies and ontological languages**: ways of structuring the relationships within the vocabulary and terminology of a body of knowledge, expressed in a formal language like OWL (Web Ontology Language) or SKOS (Simple Knowledge Organization System). The result is a machine-readable conceptual map of a domain of knowledge.

- **RDF databases**: collections of statements about objects, their properties, and their relationships ("triples"), expressed in the RDF (Resource Description Framework) syntax. These statements can be used to show how and where an object fits in the ontological structure of the body of knowledge.
- **Agent systems** and **Web services**: software environments which can be built to explore, analyse and exploit the knowledge embedded in ontologies and RDF databases.

Several major European research projects have already been applying Semantic Web technologies to the cultural heritage domain, including MultimediaN in the Netherlands (Schreiber et al.) and CultureSampo in Finland (Hyvönen et al.). MultimediaN, in particular, has been used to provide the prototype for the Europeana service. But these projects are still aimed at developing standalone Web services—albeit of a particularly sophisticated kind—and in some ways might be seen as adding to the complexity of the digital landscape rather than enabling researchers to make more effective use of it.

Semantic Web technologies are also being employed to develop a different kind of service, which aims to make data available on the Web in formats which can be processed automatically. A rapidly growing number of datasets are now available under the umbrella of the Linked Data movement (Bizer et al.). As of April 2010 they already included the Personennamendatei of the Deutsche Nationalbibliothek, as well as authoritative library vocabularies like the Library of Congress Subject Headings and RAMEAU (Bibliothèque nationale de France). These services simply provide the data; other services which enable the navigation and use of the whole body of linked data are developed separately and independently.

These technologies are intended to represent a complex body of knowledge in a connected and interoperable way, and to provide a platform for applying sophisticated discovery and analytical tools to data across the Web. They are likely to be of particular value for medieval manuscript research in the following ways:

- improving interoperability and interconnection between the many manuscript-related Web sites;
- interlinking the variety of different terminologies, vocabularies and data standards relating to manuscripts, particularly in the European multilingual and multicultural environment;
- enabling more direct connections between the manuscript descriptions and catalogues produced by cultural institutions and the continually developing apparatus of scholarly annotation, editing, study and commentary derived from and based on those manuscripts.

5. Linked Data for Medieval Manuscript Research

Semantic Web technologies and the Linked Data movement can be harnessed to build a new international collaborative infrastructure for organizing and interlinking knowledge and research about medieval European manuscripts. This collaborative infrastructure will be an open space on the Web where data about medieval manuscripts can be shared, stored, and exchanged for research purposes. It will focus on building links between data and will provide the basis for developing new kinds of services which exploit these data.

On the other hand, this infrastructure will not encroach on areas where cultural institutions exercise intellectual property rights over the medieval manuscripts in their custody, such as image reproduction. Nor should it aim to impose a single metadata standard on existing manuscript-related services; instead it should join up existing databases and vocabularies, such as those in services like the CERL Portal and Manuscriptorium. The proposed collaborative infrastructure is not intended to replace these existing databases and catalogues, but instead to connect them and act as a broker between them.

The proposed infrastructure will not, of itself, form a "virtual research environment". It will not include a transcription or digitization service, nor will it consist of a TextGrid for the analysis and manipulation of medieval texts. Nevertheless, such services will be able to link into the proposed infrastructure and to use its features—such as its manuscript identifiers—as points of reference. It will serve as the framework on which virtual research environments can be built. It is a necessary first stage before global services for asking research questions across different databases and different formats (digital, print, and manuscript) can be constructed.

Making this first stage a reality will require the development of a Linked Data environment for medieval manuscript research. This will involve transforming existing knowledge into Semantic Web formats and making it available on the Web. The main source of this knowledge is the extensive semantic network already embedded in existing manuscript descriptions, covering such semantic categories as names, identifiers, quantities, concepts, manifestations and dependents. Some existing services focus on just one of these categories, while others—particularly manuscript databases and library catalogues—cover a wider range.

Assigning unique identifiers to manuscripts is the key starting-point.[1] A URI for each manuscript will provide the crucial reference point which can be used as the basis for linking other kinds of information about that manuscript. A service for resolving these identifiers to gain access to relevant descriptions and related objects will be an essential component. Identifier services will also need to cover names, places, events and other

[1] I thank Antoine Isaac for his valuable contribution to the analysis of the requirements for this framework.

Semantic categories	Data types	Existing sources and services
IDENTIFIERS	shelf numbers, catalogue numbers, reference numbers	used in manuscript databases and library catalogues, printed catalogues, IMB, Scriptorium
NAMES	people (authors, owners, editors, bibliographers, commentators, artists), institutions, places	Europa Sacra, IMB, CERL Thesaurus, Personennamen des Mittelalters
CONTENTS	titles of works, incipits	In Principio, IMB, Scriptorium, manuscript databases and library catalogues, printed catalogues
QUANTITIES	measurements, sizes	recorded in manuscripts databases, library catalogues, and in printed catalogues
CONCEPTS	subject matter, materials, scripts, type of work, languages	IMB, Getty vocabularies, IconClass
EVENTS	dates and times	IMB, manuscripts databases and library catalogues, printed catalogues
MANIFESTATIONS	editions, transcriptions, images (printed and digital)	Web sites and books; also listed in IMB, library catalogues, some manuscript databases, printed catalogues
DEPENDENTS	scholarly writings, annotations, commentaries, bibliographical entries	listed in IMB, Scriptorium, some manuscript databases

Table 1. Semantic categories in medieval manuscript descriptions.

conceptual structures and entities. Some of these have already been developed by other disciplines, and can be reused or adapted.

Vocabularies will be another crucial part of the service, formed either by transforming existing vocabularies into an appropriate format like SKOS (Isaac and Summers) or by extracting terms from descriptive database records through techniques like text mining. Building alignments between these vocabularies is an important requirement; this will enable different vocabularies to be mapped and the relationships between terms to be indicated, without necessarily having to select one term as more authoritative than another. This will be of particular value in dealing with such areas as scripts and handwriting, where there is no generally accepted vocabulary and where the terminology used varies from country to country. Mapping between different languages will also be a significant issue here.

Closely related is the requirement to develop schemas for descriptive structures—in other words, identifying the different components of manuscript descriptions and mapping their variations. Building alignments and mapping between different languages will also be important in this area. These schemas—together with the vocabularies and their alignments—will need to be stored in repositories which are capable of providing terminology services. In practice, this will require a RDF "triple store" or similar technology (Hertel et al.).

Inherent in this framework will be a graph (in the sense of an abstract data structure intended to implement mathematical graph theory) showing the many different relationships within the data. These relationships will include those between different vocabularies, between different entities, between the component elements of a schema, between manuscripts and the entities which describe them, between manuscripts and their different manifestations, between manuscripts and their various dependents, and so on. Maintaining these interoperable knowledge bases for constructing manuscript descriptions will be an important challenge; both automated and manual methods will need to be tested and applied.

Once these data stores are in place, it will become possible to build Web services which exploit them and which add value to our existing knowledge about manuscripts. These should include services for querying and browsing the data, as well as visualizations which draw on the relationships inherent in the data structures, and map-based interfaces derived from the geographic information included in the data stores. Interfaces which enable annotation and comments from the scholarly community are another possible layer, together with services which allow the semantic network to be updated—either manually or in an automated way.

6. Organizational Requirements

The scale and complexity of the proposed infrastructure mean that it is unlikely to be developed quickly or as the result of a single, centralized project. A more feasible approach would involve making a start with a small number of datasets—and possibly several parallel projects—within an agreed overall technical framework. Existing services and institutions can be encouraged to make their data available for reformatting and inclusion.

Research groups and individual researchers need to be involved, in order to contribute to the analysis of use cases and user requirements, and to ensure that the resulting products and services are relevant to their needs. They will also be required for testing, correcting and updating, particularly once annotation and commenting become available. Their involvement can be harnessed through organizations like CARMEN, as well as directly through suitable interfaces to the data stores. The interest and

enthusiasm of the postgraduate and postdoctoral student community can also be drawn on, through training networks and other capacity-building initiatives.

Libraries and other collecting institutions are of crucial importance. As well as being the custodians of most of the manuscripts, they continue to make a major contribution by constructing and maintaining descriptive databases and managing digitization programmes and services. Their involvement can be harnessed through existing co-operative arrangements for cataloguing and digitization and through organizations like CERL, as well as individually. Commercial firms can also make a significant contribution; publishing houses specializing in medieval studies will have data and subject knowledge to contribute, while technology companies can contribute their technical expertise.

The proposed infrastructure provides an excellent opportunity for the public and private sectors to work together through the pooling of mutually relevant data and technological expertise. It will also work to strengthen links between academic researchers and the curatorial and professional staff in cultural institutions (libraries, museums, galleries and archives). Both groups have a vital and enduring interest in medieval manuscripts, but their priorities and perspectives are significantly different. Developing a Linked Data infrastructure of this kind will encourage them to share and pool their knowledge and expertise.

It should be possible to learn from other disciplines—especially in the sciences—which are already actively building global knowledge bases. This is not to say that the model based mainly on a single institution or on a loose coalition of institutions, which is probably more characteristic of the humanities, is completely inappropriate. The Perseus project, for example, has assembled a remarkably successful digital library service in classical studies from a base at Tufts University. But the scale involved in building a global infrastructure for manuscript data is more likely to require a globally coordinated approach.

An interesting model is provided by the Shared Names project, associated with the Science Commons and NeuroCommons collaborative initiatives. This project aims to assign URIs as names for biomedical information records (primarily genes), using a community-managed shared infrastructure for maintenance and development. Also of relevance from the biomedical sciences is the Encyclopedia of Life. Though not a Semantic Web project, it uses an innovative administrative structure which draws on the contributions of researchers and individuals around the world. These and other similar projects suggest that decentralized and distributed organizational arrangements which encourage individual contributions may be the most effective way of approaching a global initiative of this kind.

7. Conclusion

The surviving manuscripts are one of the key sources for research into medieval Europe and are used in addressing a wide range of research questions. Most obviously, these include research into the characteristics of manuscripts themselves as physical objects. The other major focus of research involves the use of manuscripts as evidence for all aspects of life in the medieval period.

Medieval manuscript research is a complex, fragmented, multilingual field of knowledge, which is difficult to navigate, analyse and exploit. Though printed sources are still of great importance and value, there is a large and rapidly growing body of material on the Web. Much of this Web material consists of descriptive information about manuscripts, though a considerable amount of digitization and transcription has also been carried out. At present, however, the digital landscape is difficult to navigate and overwhelming in its richness and complexity.

The infrastructure proposed in this paper focuses on the possibilities for applying new Semantic Web technologies to medieval manuscript research. These technologies have the potential to enhance research greatly, by enabling much more effective access to, and use of, relevant materials and knowledge. It will be possible to ask large-scale research questions across the global manuscript collection, in a quicker and more effective way than has ever been feasible in the past. The ultimate goal should be a Web service through which a researcher can readily find all manuscripts of relevance to the research question she is investigating, and be pointed to previous work about them and to digital representations of them.

The technologies which can make this happen are now coming to maturity, though it will ultimately require a global effort to harness their full potential. But we are in a position to envisage what the "Web of data" for medieval manuscript research might look like, and to start organizing ourselves to make it a reality.

Bibliography

Angjeli, Anila, et al. "Semantic Web and vocabulary interoperability: An experiment with illumination collections." *International Cataloguing and Bibliographic Control*, 38.2 (2009): 25–29.

Antoniou, Grigoris, and Frank van Harmelen. *A Semantic Web Primer*. 2nd ed. Cambridge, MA: MIT Press, 2008.

Berners-Lee, Tim. "The next Web of open, linked data." 2009. <http://www.youtube.com/watch?v=OM6XIICm_qo>

Bizer, Christian, Tom Heath and Tim Berners-Lee. "Linked Data – The Story So Far." *International Journal on Semantic Web and Information Systems* (2009). <http://linkeddata.org/docs/ijswis-special-issue>.

British Library's Manuscripts Catalogue. [Search entry to the main catalogues of the British Library's collection of Western manuscripts.] London: British Library, 2002–2005. <http://www.bl.uk/catalogues/manuscripts/INDEX.asp>.

CARMEN. Co-operative for the Advancement of Research through a Medieval European Network. Groningen: University of Groningen, Research School Medieval Studies 2003–2007. <http://www.carmen-medieval.eu/>.

Carusi, Annamaria and Torsten Reimer. "Text-Grid Virtual Research Environment in the e-Humanities." *Virtual research environment collaborative landscape study – A JISC funded project.* January 2010: 92–94. <http://www.jisc.ac.uk/media/documents/publications/vrelandscapereport.pdf>.

The CERL Portal: Manuscripts and Early Printed Material. Uppsala: Consortium of European Research Libraries, 2006–2007. <http://cerl.epc.ub.uu.se/portal>.

CERL Thesaurus. Göttingen: Consortium of European Research Libraries, 2008 <http://thesaurus.cerl.org/>.

[Chrétien de Troyes, *Le Chevalier de la Charrette.*] The Charrette Project 2. *Princeton University, Baylor University, and Université de Poitiers, 2005.*<http://lancelot.baylor.edu/>.

CESG: Codices Electronici Sangallenses. University of Fribourg and Abbey Library of St. Gall, 2005–2010. <http://www.cesg.unifr.ch/en>.

Crane, Gregory., Brent Seales and Melissa Terras. (2009). "Cyberinfrastructure for classical philology." *Digital Humanities Quarterly*, 3.1 (2009). <http://www.digitalhumanities.org/dhq/vol/3/1/000023.html>.

[Dante Alighieri, *La Divina Commedia.* Index of Manuscripts] Dante online. Indice dei Manoscritti. Società Dantesca Italiana, 2002. <http://www.danteonline.it/italiano/codici_indice.htm>.

Digital Scriptorium. New York: Columbia University Libraries, 2007. <http://www.scriptorium.columbia.edu/>.

Encyclopedia of Life. Washington D.C.: Smithsonian Institution et al., 2010. <http://www.eol.org/>.

Europa Inventa. Perth: ARC Network for Early European Research, 2010. <http://europa.arts.uwa.edu.au/>.

Europa Sacra. Turnhout: Brepols Publishers NV, 2004–2010. <http://www.brepolis.net/>.

Europeana. Den Haag: Koninklijke Bibliotheek, 2009–2010. <http://europeana.eu/portal>.

Fasti Ecclesiae Anglicanae 1066–1300. London: University of London, Institute of Historical Research, 2003–2010. <http://www.british-history.ac.uk/subject.aspx?subject=2&gid=39>.

Good, Benjamin M. and Mark D. Wilkinson. "The Life Sciences Semantic Web is full of creeps!" *Briefings in Bioinformatics* 7.3 (2006): 275–286.

Hertel, Alice, Jeen Broekstra and Heiner Stuckenschmidt. "RDF Storage and Retrieval Systems." *Handbook on Ontologies*, ed. S. Staab and R. Studer. Berlin: Springer-Verlag, 2009 (International Handbooks on Information Systems): 489–508.

Hitzler, Pascal, Markus Krötzsch and Sebastian Rudolph. *Foundations of Semantic Web Technologies.* Hoboken: Chapman & Hall/CRC, 2009.

Hyvönen, Eero. et al. "CultureSampo – A national publication system of cultural heritage on the Semantic Web 2.0." *The Semantic Web: Research and Applications: ESWC 2009 Heraklion,*

Crete, Greece, May 31–June 4, 2009 Proceedings. Ed. Lora Aroyo et al. Berlin: Springer-Verlag, 2009 (Lecture Notes in Computer Science 5554): 851–856. <http://www.seco.tkk.fi/publications/2009/hyvonen-et-al-culsa-demo-eswc-2009.pdf>.

In Principio: Incipit Index of Latin Texts. Turnhout: Brepols Publishers NV, 2004–2010. <http://www.brepolis.net/>.

International Medieval Bibliography. Turnhout: Brepols Publishers NV, 2001–2010. <http://www.brepolis.net/>.

Isaac, Antoine and Thierry Bouchet. "Rameau et SKOS." *Arabesques*, 54 (Apr.–June 2009), 13–14.

Isaac, Antoine and Ed Summers, eds. *SKOS Simple Knowledge Organization System Primer*. W3C Group Note, August 18, 2009. <http://www.w3.org/TR/skos-primer>.

Isaac, Antoine et al. "Evaluating thesaurus alignments for semantic interoperability in the library domain." *IEEE Intelligent Systems*, 24.2 (2009), 76–86.

Labyrinth: Resources for Medieval Studies – Manuscripts. Washington D.C: Georgetown University 1994–2007. <http://labyrinth.georgetown.edu/display.cfm?Action=View&Category=Manuscripts>.

Library of Congress Subject Headings. Washington D.C.: Library of Congress 2009. <http://id.loc.gov/authorities>.

Manuscriptorium. Praha: Národní knihovna České republiky, 2006–2010. <http://beta.manuscriptorium.com/>.

Medieval Manuscripts in Dutch Collections. Den Haag: Koninklijke Bibliotheek, 2007. <http://www.mmdc.nl/static/site/index.html>.

MEDIUM. Paris: Institut de recherche et d'histoire des textes, 2008. <http://www.irht.cnrs.fr/ressources/medium_frame.htm>.

Mirabile: Archivio digitale della cultura latina medievale. Firenze: SISMEL, 2010. <http://www.mirabileweb.it/>.

Muzerelle, Denis. *Vocabulaire codicologique*. Paris: Institut de recherche et d'histoire des textes, 2002–2003. <http://vocabulaire.irht.cnrs.fr/vocab.htm>.

NeuroCommons. Cambridge, MA: Science Commons, 2009. <http://neurocommons.org/page/Main_Page>.

Pass, Gregory A. *Descriptive Cataloging of Ancient, Medieval, Renaissance, and Early Modern Manuscripts*. Chicago: Association of College and Research Libraries, 2002.

Perseus Digital Library. Version 4.0. Boston, MA: Tufts University, 2005–2010. <http://www.perseus.tufts.edu/>.

Personennamendatei [including *Personennamen des Mittelalters*]. Frankfurt: Deutsche Nationalbibliothek, 2005–2010. <http://www.d-nb.de/eng/standardisierung/normdateien/pnd.htm>.

RAMEAU: Répertoire d'autorité-matière encyclopédique et alphabétique unifié. Paris: Bibliothèque nationale de France 2010. <http://rameau.bnf.fr/>.

Schreiber, Guus et al. "Semantic annotation and search of cultural-heritage collections: The MultimediaN E-Culture demonstrator." *Journal of Web Semantics*, 6.4 (2008), 243–249.

Science Commons. Cambridge, MA: Creative Commons, 2005–2010. <http://sciencecommons.org/>.

Scriptorium. [with catalogues 1946–2008]. Bruxelles: Centre International de Codicologie 2002–2010. <http://www.scriptorium.be/>.

Shared Names. Boston: Science Commons, 2009–2010. <http://sharedname.org/>.

Stevens, Wesley M., ed. *Bibliographic Access to Medieval and Renaissance Manuscripts: a Survey of Computerized Databases and Information Services*. Binghamton, NY: Haworth Press, 1992.

Text Encoding Initiative. P5: Guidelines for Electronic Text Encoding and Interchange. 10 Manuscript Description. TEI Consortium, 2007. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>.

TextGrid. Göttingen et al.: TextGrid Project, 2009–2010. <http://www.textgrid.de/>.

Virtual Library – Historical Auxiliary Sciences – Codicology. München: Ludwig-Maximilians-Universität, 2000–2009. <http://www.vl-ghw.lmu.de/kodikologie_en.html>.