

Sharing humanities data for e-research: conceptual and technical issues

Toby Burrows

Introduction

The humanities, as defined by the Australian Academy of the Humanities, encompass the following disciplines: Archaeology; Asian Studies; Classical Studies; English; European Languages and Cultures; History; Linguistics; Philosophy, Religion and the History of Ideas; Cultural and Communication Studies; the Arts. Researchers in some of these fields employ quantitative and qualitative methodologies similar to those used in the sciences and social sciences, but most research in the humanities is perceived as distinctive and different from research in other fields, both in its methodologies and in its approach to data.

Archiving and sharing humanities data for reuse by other researchers is crucial in the development and application of e-research in the humanities. There has been considerable debate about the applicability of e-research in the humanities, particularly around the relevance of programs to digitise source materials on a large scale. Conceptualised and designed properly, however, a humanities data archive can provide the platform on which data-intensive e-research can be based, and to which e-research processes and tools can be applied.

This paper looks at the distinctive characteristics of humanities data, and examines how various models of the humanities research process help in understanding the meaning of ‘data’ in the humanities. It reviews existing services and approaches to building data archives and e-research services for the humanities, and the assumptions they make about the nature of data. It also analyses some conceptual and technical frameworks which could serve as the basis for future developments, focusing particularly on the place of Linked Open Data in building large-scale humanities e-research environments.

1. The data deluge and e-research

The ‘data deluge’ is widely recognised as a major problem for scientific research. The scale and the complexity of the data now being gathered have been threatening to

overwhelm researchers in most areas of the sciences and have been making it increasingly difficult to design effective research strategies (Bell, Hey and Szalay 2009). One major response has been the development and systematic application of e-research solutions (also called ‘e-science’ or ‘cyberinfrastructure’) on a national and international scale, with extensive government investment in appropriate digital infrastructure (Hey and Trefethen 2005, Jankowski 2009). E-research in this context refers specifically to an environment where digital data are gathered from sophisticated instruments like radio telescopes, electron microscopes and synchrotrons, collected into very large (and often dispersed) datasets stored on supercomputers, and processed using sophisticated software for description, analysis, modelling and visualisation. E-research also involves the automation of experimental and analytical processes, demonstrated through such services as MyExperiment (Goble *et al.* 2010). This usually includes the automation of data capture and management processes into an integrated workflow. In Australia, many projects of this kind are currently being funded by the Australian National Data Service (ANDS).

The e-research approach has recently been described as ‘a new, fourth paradigm for science based on data-intensive computing’ (Bell 2009: xiii) and is based on an essentially scientific model of the research process. In Australia, successful applications of large-scale e-research can be found in such ‘big science’ fields as radio astronomy, marine science, climate science and geoscience. The humanities (defined as the disciplines covered by the Australian Academy of the Humanities, listed earlier) have been largely absent from the process of funding and building large-scale e-research solutions. And yet the data deluge is real and evident in the humanities as well. Although the size of humanities data may not reach the petabyte scale of the sciences, the digital information landscape for the humanities is characterised by proliferating digital resources and software tools and by rapidly increasing complexity and heterogeneity (Borgman 2007: 212-224). This enormous proliferation of relevant digital resources in widely varying formats, which are very difficult to navigate and use effectively, is an issue of major significance for humanities researchers.

There are numerous humanities digital infrastructure services and projects in Australia, as well as internationally. They include large collections of digitised objects and texts; online dictionaries, directories and encyclopedias; catalogues, indexes and

lists; and linguistic corpora. There are also many thematic databases and Web sites, each with its own customised interface, often produced as the result of a specific research project. There is little if any interoperability between all these services, however, other than their availability for searching through Google. As a result, they have been adding to the complexity of the landscape and making the data deluge worse, rather than serving as components of a systematic and coherent e-research environment.

Approaches to tackling the data deluge in the humanities to date have largely relied on providing a search capability which is larger and broader in scope and covers a greater range of data sources. A common way of doing this has been to harvest metadata from various sources into a central store, often using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). This has usually occurred on an international or national scale and is typified by the *Europeana* digital library, which aggregates information about digitised cultural objects from many European countries into a standard metadata schema. This approach typically provides search results which consist of records within the central database.

Federated searching of multiple data sources has been widely applied to bibliographic data in the library world, often in the form of commercial software designed for cross-searching many indexing databases. This method typically provides search results which are direct links to the original datasets. It has been extended to full-text sources in more recent projects by creating central indexes to a body of distributed content of this kind. The British *Connected Histories* service is a particularly sophisticated example of this approach, joining eleven major sources relevant to early modern and nineteenth-century British history. Another group of five sources was added in September 2011. Australian humanities services like AustLit have also been experimenting with federated searching of relevant databases. The National Library of Australia's *Trove* service combines federated searching of some distributed sources (e.g. digitised Australasian newspapers) with a central metadata store (Holley 2010). Search results are a mixture of records in the central database and links to results in distributed sources.

The idea of discipline-based 'virtual research environments' for the humanities has also been widely promoted and investigated as a solution to the data deluge. The Joint Information Systems Committee (JISC) in the United Kingdom has sponsored a

range of projects in this area. This kind of environment brings together data sources, annotations and analytical tools on a researcher's personal desktop, but it is aimed at improving data management for individual researchers rather than at enabling data archiving, sharing and reuse on a national or global scale. The current SUDAMIH (Supporting Data Management Infrastructure for the Humanities) project at the University of Oxford is a good example of this approach (Wilson 2010).

Another approach involves extending existing digital libraries and curatorial databases to accommodate scholarly annotations and other data-oriented features. This can be seen in recent work by the Perseus Digital Library for classical languages and literature to add semantic tagging and automatic entity recognition to its text collections (Babeu *et al.* 2007). Although this is an important initiative, reuse of data outside Perseus is made difficult by its reliance on a centralised and integrated database which integrates data sources, metadata and research findings. Related work has been done by the AusStage service in the Australian context, although the central building block in this case is a metadata store rather than a collection of texts or digital objects (Bollen *et al.* 2009).

All of these approaches have their value, but none of them is equivalent to such large-scale, integrated but decentralised scientific e-research frameworks as IMOS, auScope, TERN, and the Virtual Observatory. At best, the humanities services consist of sophisticated ways of searching digital object collections and the descriptive metadata assembled by curatorial and institutional experts. They tend to ignore the research processes required to exploit these sources. The NeAT-funded Aus-e-Lit project (Gerber, Hyland and Hunter 2010 and Chapter 8 in this volume) is a significant exception, with its annotation and visualization tools, and its ability to save and share sets of connected resources. *Connected Histories* also enables users to tag, save and share lists of 'connections' between resources.

2. Data and research processes in the humanities

A workable and consistent definition of 'data' for the humanities is an essential first step in building a large-scale e-research environment. After all, e-research in the sciences rests ultimately on the manipulation of 'data', in the scientific understanding of that word. There are *quantitative* data in the humanities similar to those in the sciences,

e.g. statistical spreadsheets and databases. There are also *qualitative* data similar to those in the social sciences, e.g. interviews, surveys, and questionnaires. Both types of data are present in major European and Australian data archives. The U.K. Data Archive contains a range of quantitative historical datasets, for example, and the Australian Data Archive includes historical census statistics from the Australian colonial period which are mainly of interest to historians.

But humanities research also produces and makes use of other kinds of evidence which are more difficult to define and categorise, and which do not fit readily into these quantitative and qualitative frameworks of the sciences and social sciences. There is a tendency among commentators to assert that primary sources are the humanities researcher's data and therefore that primary sources (including documents, texts, and images) and 'data' are one and the same thing (Borgman 2007: 215-217). This is not particularly helpful, since it blurs the distinction between 'data' and 'sources of data' - or between evidence and sources of evidence. It also conflates the objects of research with the descriptive and representational data derived from them by researchers. It would be analogous to describing the stars and galaxies as an astronomer's 'data' when, in fact, the actual physical objects are clearly distinguishable from the observations relating to them—and these observations form the data which the researcher uses and analyses. The difficulty for the humanities is that they do not deal exclusively with physical phenomena. They are also concerned with more abstract entities like texts and works, which are conceptual entities as well as their physical manifestations.

An analysis of the current digital landscape suggests that there are several reasons why the humanities have tended to fall outside the scope of existing e-research frameworks:

- It is difficult to define 'data' in the humanities in a consistent (i.e., machine-processable) way;
- It is difficult to identify and model generic research processes, since research methods tend to be poorly documented and little discussed, or regarded as matters of common sense;
- There has been a strong tendency towards project-specific digital solutions: integrated sites with a mixture of digital source materials, analysis, commentary

and annotations, which cannot be aggregated into a more general e-research framework;

- It is difficult to separate analysis and research outcomes from the source materials—one researcher's publications quickly become another researcher's evidence or data;
- There is a gulf between the research processes of academic researchers and the curatorial processes of the cultural institutions which hold most of the source materials—these institutions have their own ways of organising and describing source materials which may be quite different from the information produced by the research process;
- The digitisation of source materials has tended to be promoted as a substitute for (or equivalent) of e-research, with source materials seen as equivalent to 'data'.

In the light of all these difficulties, how can the modelling which underpins scientific e-research environments be applied in the humanities? A useful starting-point is provided by some recent attempts to develop a model of the humanities research lifecycle. Unsworth (2000) identified seven basic activities which he called 'scholarly primitives': discovering, annotating, comparing, referring, sampling, illustrating, and representing. He also emphasised the activity of *linking*—'either in the classic form of annotation, or in the more abstract sense of creating operative associations between, among, and within digital objects.' A subsequent study by the University of Minnesota Libraries (2006) reduced these to four basic activities: discover, gather, create, and share. Project Bamboo, a major investigation funded by the Mellon Foundation, uses a generally similar approach, defining annotations as 'notes, tags, links, and/or citations', but also identifying a closely related 'community curation' function which covers 'the ability to categorize, annotate, review, rate along multiple spectra, and discuss' (Project Bamboo 2009).

These models suggest that there are two basic components to humanities data. The first consists of the various annotations, tags, links, associations, ratings, reviews and comments produced during the humanities research process. The second consists of the *entities* to which these annotations refer: concepts, persons, places and events, as well as creative works, artworks, publications, texts and other physical and digital

objects. An e-research framework for the humanities needs to be able to identify these entities as well as capture the annotations and other scholarly outputs which refer to them.

3. Linked Open Data

The concepts and technologies of the emerging Linked Open Data movement appear to offer a realistic basis for designing, testing and evaluating a systematic framework which can serve as the foundation for an e-research environment in the humanities. First articulated by Tim Berners-Lee (the inventor of the Web) (Bizer, Heath and Berners-Lee 2009), the idea of Linked Open Data¹ focuses on the identification and management of information about entities (people, objects, concepts, places, events, creative works and the like) and the relationships between them. It provides the standards, tools and technical structures for managing these identifiers in a systematic way. It is hospitable to multiple interconnected vocabularies, code lists, ontologies and other naming systems, without enforcing artificial and inappropriate uniformity. It makes use of unique machine-processable codes (Uniform Resource Identifiers or URIs) to identify each entity, and employs the Resource Description Framework (RDF) as the syntax for expressing relationships between entities. RDF and URIs are specifications of the World Wide Web Consortium.

A Linked Open Data system is designed for expressing, tracking and analysing relationships between entities. The most common technical architecture for managing the complex network or graph of entities and their relationships is an RDF ‘triplestore’. Software for managing these triplestores is now relatively mature, and considerable research has been done on ways of improving their performance at an ever-increasing scale, encompassing billions of RDF statements (Hertel, Broekstra, and Stuckenschmidt 2009). Interfaces for working with and reusing the data can then be built on top of the linked data, as can tools for capturing annotations and for constructing links between entities and different types of data sources. A ‘Research Objects’ model has recently been proposed as a standard framework for this process of aggregating linked data for use by Web services (Bechhofer *et al.* 2010).

¹ Strictly speaking, Linked Data is the name of the technical framework; the term ‘Linked Open Data’ adds the dimension of open, freely-available linked data as opposed to proprietary, restricted-access data.

Linked Open Data is rapidly maturing as an internationally applied and tested approach, and is already being deployed for some scientific and government research data (e.g., Baker and Keizer 2010). Linked Open Data formats are also being used and tested in several community-sourced public knowledge projects, such as DBpedia and Freebase (Bizer *et al.* 2009, Bollacker *et al.* 2008). The European Commission, under its 7th Framework Programme, recently awarded a €6.45m grant to the Linking Open Data project (known as LOD2). Building on work originally done for DBpedia, this new project involves academic, commercial and community partners, and will focus on the development and deployment of tools, standards and methodologies for ‘Creating Knowledge out of Interlinked Data’ on a large scale.

The applicability of Linked Open Data to the humanities—and particularly to curatorial institutions—is the subject of growing discussion and investigation. The inaugural International Linked Open Data in Libraries, Archives, and Museums Summit (LOD-LAM) was held in San Francisco in June 2011. It built on small-scale pilot projects already carried out by various national collecting institutions, notably the Library of Congress. A particularly interesting project is *Civil War Data 150*, led by the Archives of Michigan, which is using the Linked Open Data framework to share and connect Civil War related data across local, state and federal libraries, archives and museums².

A small-scale example might illustrate how Linked Open Data works in the humanities. There are numerous statements like this in historical and literary texts: ‘New Zealand was discovered by Captain James Cook (1728-1779) in the ship H.M.S. Endeavour’. There are several entities referred to in this statement, together with explicit assertions of relationships between them:

<i>entity</i>	<i>entity</i>	<i>relationship</i>
New Zealand	James Cook	discovered by / discoverer of
James Cook	1728	born in
James Cook	1779	died in
James Cook	Captain	held title / rank of
H.M.S. Endeavour	James Cook	was ship of / was captain of
H.M.S. Endeavour	ship	example (instance) of

² <http://www.civilwardata150.net/>

There are also various implicit assertions in this statement, particularly about the larger classes of concepts to which specific entities belong:

<i>entity</i>	<i>entity</i>	<i>relationship</i>
James Cook	person / man	example (instance) of
Captain	title / rank	example (instance) of
1728	year	example (instance) of
New Zealand	country	example (instance) of

The relatively simple statement also conceals a body of much larger statements and conceptual assumptions about, for example, geographic naming systems (New Zealand, rather than Aotearoa or Nova Zeelandia), calendar systems (Christian era, as opposed to Māori), and languages (English, rather than Māori or any other language of scholarship). The word ‘discovered’ is particularly problematic, and illustrates the way in which concepts change their meaning over time as the result of cumulative scholarship. This kind of pervasive complexity, ambiguity and variation is at the heart of humanities research. Any useful e-research framework must be able to represent and process complex networks of assertions of this kind.

It is very important, therefore, that Linked Open Data is neutral about vocabularies, names, languages and concepts. A Unique Resource Identifier (URI) can be created for each name or concept in a different vocabulary or language, and different types of links can be created between them. URIs denoting types of links like ‘is equivalent to’ or ‘is the same entity in a different language’ can be used to connect these resources. In the case above, there can be a URI for ‘New Zealand’, a different URI for ‘Aotearoa’, and another URI for ‘Nova Zeelandia’³. Links can express which language is used for each name, and the chronological period in which that name was current.

The resulting graph of relationships could also express the fact that New Zealand is now an instance of the concept ‘country’, with a variety of names in different languages, but was not a ‘country’ in this sense at the time of Cook’s discovery. This would involve invoking an alternative ontology of geographical regions based on Māori terminology. Similarly, using the statement ‘was discovered by’ to express the

³ <http://sws.geonames.org/2186224/> is the URI for New Zealand in the Geonames dataset.
<http://sws.geonames.org/2186224/about.rdf> contains the full set of statements for that resource.

relationship between Captain Cook and New Zealand reflects European assumptions about the exploration of the world. It could be expressed quite differently from the Māori point of view (even using the English language)—using a more neutral statement like ‘was visited by’.

4. Road map for medieval manuscript research

Because of the inherent complexity of the conceptual and semantic structures of humanities research, an initial evaluation of the Linked Open Data approach might best be carried out using a representative corpus of humanities data which is sufficiently complex but also sufficiently manageable. Medieval manuscript research is an excellent example of the complex digital information landscape which has emerged in the humanities. It is a rich, fragmented, multilingual field of knowledge, which is difficult to navigate, analyse and exploit. There are hundreds of Web services, some commercial and many in the public domain. At present, these services have to be consulted separately and individually to gain the full value of their knowledge. Search engines like Google cover some—but not all—of them, and provide relatively unsophisticated access to their contents.

These existing Web 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. There are a range of national databases as well as a small number of international databases. Some of these services provide digital images of manuscripts as well as descriptive information about them. There are many Web sites which list, transcribe, or provide digital images of manuscripts relating to a specific text or a specific author. Ancillary Web services include sites devoted to manuscript terminology and vocabularies, *incipits*⁴, subjects, authors, and people more generally. Other services provide indexes to journal articles, scholarly books and other secondary literature about specific manuscripts.

⁴ *Incipits* are the opening words of medieval texts, often used to identify anonymous works.

Manuscripts are central to research in medieval studies, as the major surviving source of evidence, together with buildings and art. They are inherently interdisciplinary, covering the full range of humanities disciplines, including music, literature, philosophy, and art (e.g., in illuminated manuscripts). They pose significant technical challenges associated with linguistic variation and concept-shifting over time. An effective e-research environment would make it possible for the first time to study the entire corpus of medieval manuscripts as a whole, and to ask research questions across this whole corpus.

This is an area where considerable international planning and scoping have already been undertaken. A European Science Foundation Exploratory Workshop on ‘Applying Semantic Web Technologies to Medieval Manuscript Research’ was held at the University of Birmingham in March 2009. Organized by Wendy Scase, Orietta Da Rold and Toby Burrows, this workshop brought together specialists from the fields of manuscript studies, information and computer science, and library science, as well as from public and commercial organizations and institutions. The Exploratory Workshop resulted in the development of a Road Map for the research and development required to implement the proposed new Linked Open Data technologies (Burrows 2010, Scase 2009). Efforts are now underway to identify the funding required to turn the Road Map into a work program.

The Exploratory Workshop was initiated by the Medieval Manuscript Research Group of the Co-operative for the Advancement of Research through a Medieval European Network (CARMEN). CARMEN members include the many European centres for medieval research, as well as professional associations, cultural institutions and publishing companies with expertise in this field. The former ARC Network for Early European Research played a key role in establishing CARMEN. The Road Map draws on the expertise and needs of the entire European research community in this field through its connection with CARMEN. The work envisaged in the Road Map will also be able to draw on existing databases and vocabularies, such as those from the Council of European Research Libraries (CERL), the Manuscriptorium service (Czech National Library), and the specialist publishing company Brepols NV (Belgium), as well as the Europa Inventa database developed by the ARC Network for Early European Research (Burrows 2008).

The Road Map is not intended to replace these resources, but is aimed instead at identifying how new ways of exposing and inter-linking their data can be designed and implemented. The main technical components required to build the kind of Linked Open Data environment envisaged by the Road Map are as follows:

- Unique identifiers for individual manuscripts and their components;
- Unique identifiers for names, places, works and other entities related to manuscripts;
- Terminology mapping between the many different vocabularies and ontologies used to describe manuscripts;
- Schema mapping between the various different descriptive structures currently employed.

While these components can be derived from existing collection databases, it will also be important to test and implement methods for identifying and extracting entity information from research publications and text-based manuscript descriptions, using techniques like text mining.

The goal is to develop a Web-based environment which can provide unified browsing and searching across multiple sites and datasets using these identifiers, terminology services and mapping services. An important aim will be to link the manuscripts to the scholarly outputs (articles, books, published catalogues, editions and so on) which are derived from them, via the entities which reference (and are referenced by) both types of material.

5. Outcomes and benefits

The result envisaged by the Road Map will be the first systematic attempt to embed Linked Open Data into a model of the humanities research process and to identify the key elements for designing large-scale digital infrastructure for the humanities in the future. It will be built around a workable definition of ‘data’ in the humanities from an e-research perspective, linked to an analysis of humanities research processes. It will develop a working digital environment which can serve as a proof-of-concept for humanities e-research. This environment will enable researchers to browse and search across the international corpus of medieval manuscripts, using the entities referenced in

and by those manuscripts. Researchers will also be able to apply a range of software services to the Linked Open Data service, including tools for annotation, visualisation, mapping, network analysis and collaboration.

By demonstrating and testing a highly innovative new model for the design of humanities e-research infrastructure, the Road Map has the potential to establish an entirely new blueprint for e-research in the humanities. This will enable larger-scale research questions to be pursued more effectively and will also address the increasingly serious effects of the ‘data deluge’. The value of the Linked Open Data approach for modelling and managing complex bodies of knowledge in the humanities will be able to be assessed, together with an evaluation of the ways in which this approach can be embedded into humanities research processes.

An e-research approach of this kind can address other important issues affecting the humanities, as well as the ‘data deluge’. There is a significant gap between curatorial databases and research processes, since most analysis and publication takes place in an entirely different context (even when both are online). The use of Linked Open Data services should make it possible to achieve a closer connection between academic research processes and curatorial activities, by linking research analysis and outputs to the same entities used in curatorial databases. If e-research environments are designed with this in mind, it will be a major step towards increasing the effectiveness of researchers’ use of cultural collections, and towards ensuring that research results are fed back into the management of these collections. Bridging the gap between academic research and curatorial activities is a vital method of ensuring the future value of public cultural collections.

More generally, Linked Open Data infrastructure will be valuable for exploring more immediate and effective ways of sharing data in the humanities. Data sharing is now a major issue for scientific research, and is a high priority for initiatives like the Australian National Data Service (ANDS). But the practical and theoretical applicability of data sharing for the humanities in a digital world is yet to be examined and investigated in a systematic way. Using a Linked Open Data approach will at least make it possible to envisage collaborative ways of building a shared body of semantically rich information. In turn, this will contribute to ways of addressing the theoretical agendas emerging from the broadly conceived (‘big tent’) arena of the

digital humanities, and will be a major step towards ‘getting the Web to think like a humanist’⁵.

References

- Babeu, Alison, David Bamman, Gregory Crane, Robert Kummer and Gabriel Weaver. 2007. Named entity identification and cyberinfrastructure. *Proceedings of the 11th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2007), Budapest, Hungary*. Berlin: Springer-Verlag. 259-270.
- Baker, Thomas and Johannes Keizer. 2010. Linked data for fighting global hunger: Experiences in setting standards for Agricultural Information Management. In David Wood (ed.), *Linking Enterprise Data*. Washington, DC: Springer. 177-201.
- Bechhofer, S., Ainsworth, J., Bhagat, J., Buchan, I., Couch, P., Cruickshank, D., Delderfield, M., Dunlop, I., Gamble, M., Goble, C., Michaelides, D., Missier, P., Owen, S., Newman, D., De Roure, D. and Sufi, S. 2010. Why linked data is not enough for scientists. *IEEE Sixth International Conference on e-Science: e-Science 2010: Proceedings*. Los Alamitos: IEEE Computer Society, 300-307.
- Bell, Gordon. 2009. Foreword. In Tony Hey, Stewart Tansley and Kristin Tolle (eds.), *The Fourth Paradigm*. Redmond: Microsoft Research. xiii-xvii.
- Bell, Gordon, Tony Hey and Alex Szalay. 2009. Beyond the data deluge. *Science* 323: 1297-1298.
- Bizer, Christian, Tom Heath and Tim Berners-Lee. 2009. Linked data—The story so far. *International Journal on Semantic Web and Information Systems* 5.3: 1–22.
- Bizer, Christian, Jens Lehman, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak and Sebastian Hellman. 2009. DBpedia - a crystallization point for the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web* 7.3: 154-165.
- Bollacker, Kurt, Colin Evans, Praveen Paritosh, Tim Sturge and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data (SIGMOD '08)*. New York: ACM, 1247-1250.
- Bollen, Jonathan, Neal Harvey, Julie Holledge and Glen McGillivray. 2009. AusStage: e-Research in the Performing Arts. *Australasian Drama Studies* 54: 178-194.
- Borgman, Christine L. 2007. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, Mass.: MIT Press.

⁵ My thanks to Philip Mead for this phrase.

- Burrows, Toby. 2008. Discovering Early Europe in Australia: The Europa Inventa Resource Discovery Service. In George Buchanan, Masood Masoodian and Sally Jo Cunningham (eds.), *Lecture Notes in Computer Science: Digital Libraries: Universal and Ubiquitous Access to Information (ICADL 2008)*. Berlin: Springer-Verlag. 394-395.
- . 2010. Applying Semantic Web technologies to medieval manuscript research. In Franz Fischer, Christiane Fritze and Georg Vogeler (eds.), *Kodikologie und Paläographie im Digitalen Zeitalter 2—Codicology and Palaeography in the Digital Age 2*. Norderstedt: Books on Demand. 117-131.
- Gerber, Anna, Andrew Hyland and Jane Hunter. 2010. A collaborative scholarly annotation system for dynamic web documents—A literary case study. In Gobinda Chowdhury, Chris Koo and Jane Hunter (eds.), *Lecture Notes in Computer Science: The Role of Digital Libraries in a Time of Global Change (ICADL 2010)*. Berlin: Springer-Verlag. 29-39.
- Goble, Carole A., Jiten Bhagat, Sergejs Aleksejevs, Don Cruickshank, Danius Michaelides, David Newman, Mark Borkum, Sean Bechhofer, Marco Roos, Peter Li and David De Roure. 2010. myExperiment: a repository and social network for the of bioinformatics workflows. *Nucleic Acids Research* 38 (suppl 2), W677-W682.
- Hertel, Alice, Jeen Broekstra and Heiner Stuckenschmidt. 2009. RDF storage and retrieval systems. *Handbook on Ontologies*. Berlin: Springer. 489-508.
- Hey, Tony and Anne E. Trefethen. 2005. Cyberinfrastructure for e-Science. *Science* 308: 817-821.
- Holley, Rose. 2010. Trove: Innovation in access to information in Australia. *Ariadne*, 64. [<http://www.ariadne.ac.uk/issue64/holley/>]. Accessed 14 August 2011.
- Jankowski, Nicholas W. (ed.) 2009. *e-Research: Transformation in Scholarly Practice*. New York: Routledge.
- Project Bamboo. 2009. *Commons Entity and Artifact Description*. Available from: [<https://wiki.projectbamboo.org/display/BPUB/Commons+Entity+and+Artifact+Description>]. Accessed 14 August 2011.
- Scase, Wendy. 2009. *Applying Semantic Web Technologies to Medieval Manuscript Research*, European Science Foundation Exploratory Workshop Report. Strasbourg: European Science Foundation. [<http://www.esf.org/activities/exploratory-workshops/workshops-list.html?year=2009&domain>]. Accessed 14 August 2011.
- University of Minnesota Libraries. 2006. *A Multi-Dimensional Framework for Academic Support: a Final Report*. [http://www2.lib.umn.edu/about/mellon/UMN_Multi-dimensional_Framework_Final_Report.pdf]. Accessed 14 August 2011.

Unsworth, John. 2000. *Scholarly Primitives: What Methods Do Humanities Researchers Have in Common and How Might Our Tools Reflect This?* [<http://jefferson.village.virginia.edu/~jmu2m/Kings.5-00/primitives.html>]. Accessed 14 August 2011.

Wilson, James A. J. 2010. *Supporting Data Management Infrastructure for the Humanities (SUDAMIH): Project Plan*. [http://sudamih.oucs.ox.ac.uk/docs/SudamihPP_2.1_nobudget.pdf]. Accessed 14 August 2011.