

# Identity parade: managing contextual personal information for archival data

Toby Burrows

**Dr Toby Burrows** manages the Scholars' Centre in the University of Western Australia Library, where his responsibilities include rare books, personal archives and manuscripts. He also directs the Digital Services Programme of the Australian Research Council's Network for Early European Research (NEER), which is developing web services for the Australian research community in medieval and early modern studies. He is a medieval historian and librarian by training and has published widely in these and related fields.

*For researchers in the humanities, biographical data are central to most of their investigations and the sheer size and scope of biographical research has produced an enormous range of printed and digital information. One area of major research interest covers the connections between people, and their place in social, intellectual and cultural networks, particularly in the past. This approach is increasing in importance, spurred on by a much wider and more general interest in the phenomenon of networks and in their potential as an explanatory framework for human behaviour. This kind of research into social, intellectual and cultural networks appears to pose a major challenge to existing digital resources and to the metadata structures which underpin them. This paper investigates the extent to which existing metadata frameworks – particularly in the archival and library sectors – are able to present this type of contextual information and looks at the potential value of new and emerging approaches. It also suggests some ways in which the latest semantic web and Web 2.0 developments can be applied to build the next generation of biographical services for humanities researchers.*

## Introduction

For researchers in the humanities, biographical data are central to most of their investigations. After all, humanities research focuses on human beings – their lives, their creations, their relationships, their achievements. It can range from writing a biography or assessing the contribution of an individual person, through to studying large numbers of people involved in major social, intellectual or cultural movements and events.

As a result of this focus, there is an extensive range of biographical information available to researchers today, in both digital and printed forms. Probably the most obvious sources are the national biographical dictionaries, such as the *Oxford Dictionary of National Biography* or the *Australian Dictionary of Biography*. More general encyclopedias – particularly *Wikipedia* – also contain a huge amount of biographical data. And then there are the innumerable subject-specific biographical dictionaries and listings, which cover almost every possible subject from the *Prosopography of the Later Roman Empire* to *Contemporary Authors*. Sources which focus on genealogy are even more extensive.

In this enormous field of biographical research, one area of growing interest deals with the connections between people, and their place in social, intellectual and cultural networks, particularly in the past. This approach is increasing in importance, spurred on by a much wider and more general interest in the phenomenon of networks and in their potential as an explanatory framework – exemplified by the claim by Albert-László Barabási that the ‘new science of networks’ represents ‘the next scientific revolution’.<sup>1</sup> No doubt there is also some influence from the popularity of associative web services like Facebook and MySpace.

This kind of research into social, intellectual and cultural networks appears to pose a major challenge to existing digital resources and to the metadata structures which underpin them. This paper investigates the extent to which existing frameworks are hospitable to this type of research, and looks at the potential value of new and emerging approaches. It also suggests some ways in which the latest semantic web and Web 2.0 developments can be applied to build the next generation of biographical services for humanities researchers.

## Tracking networks of people: researchers' requirements

The Australian Research Council's Network for Early European Research (NEER) is a virtual organisation which brings together more than 350 researchers across Australia who are interested in medieval and early modern European culture and history.<sup>2</sup> It uses a shared digital environment which includes a web-based collaborative workspace (Confluence), a digital repository for research outputs and data (PioNEER), and a knowledge space for research linked to objects in cultural heritage collections (Europa Inventa). NEER was formed in 2004–05 as part of a new government program aimed at developing and encouraging large groups of researchers to collaborate on a national scale, modelled on a similar program in the European Union.

One of NEER's major research themes is 'intellectual formations', particularly in science, technology and medicine. The emergence of a scientific culture in early modern Europe is a major element within this theme. Closely associated with it is research into the changing relationship between science and religion, and between the sciences and the arts. The major focus of this theme is research into the structures which fostered the development of the scientific culture, and particularly scientific networks, in early modern Europe. Early European perceptions of the new phenomenon of 'information overload' and 'information explosion' are also important areas of research. NEER has provided seed-funding to several smaller research clusters which focus on specific topics related to this general theme:

- knowledge networks and reading communities in late medieval England;
- literary, monastic and intellectual culture in twelfth- and thirteenth-century Europe;
- notebooks and note-taking in early modern Europe; and
- the letter in Europe, 1150–1850.

A common interest of all these researchers is the way in which knowledge networks developed and operated during these periods. How did individual people fit into these networks? How was knowledge shared and transmitted? To what extent were women involved in these networks? What characteristics drew people together

and held them together? What techniques did they use to communicate with each other?

One of these researchers is Ludmilla Jordanova (King's College London). Her work focuses on 'Networked images: portraits, knowledge and collecting'. She looks at the role of visual and material culture in Britain from the Restoration until the beginning of the nineteenth century, particularly in relation to items which were considered to be valuable, although not necessarily in a monetary sense. Portraits are the most characteristic item of this type, and Jordanova examines how they were made, acquired, exchanged, and valued. She is especially interested in networks, both personal and professional, and their relationship to gifts, patronage, and artistic practices.<sup>3</sup>

A focus on intellectual and cultural networking is also central to the research of John Schuster (University of New South Wales). His work on what he describes as 'the organisational dynamics of the experimental life in the early Royal Society' looks at the network of connections, influences, conflicts, and relationships between English scientists in the seventeenth century, and their roles in experimentation and publication.<sup>4</sup> Another Australian researcher, Luciano Boschiero (Campion College), is investigating networking and experimental science in Italy and England during the seventeenth century, concentrating mainly on the connections between members of the Royal Society and members of academies in Bologna and Florence.<sup>5</sup> Also working in this general field is Richard Yeo (Griffith University), whose research examines the ways in which scientific knowledge was communicated and shared across the networks of correspondents established by such groups as the circle of Samuel Hartlib, the Oxford Philosophical Society, and the Royal Society of London.<sup>6</sup>

The standard information available in biographical sources and services about individual people is likely to cover the following major areas:

- name: including collocation of variant names and disambiguation of similar names;
- life events: biography, chronology, roles and occupations, dates, places; and
- evidence: works by and about the person (including citations, documents, images and so on).

But this information is not enough for the kinds of research being carried out under the rubric of 'intellectual formations'. What these researchers need, above all, is the ability to identify, record and map *connections* between individuals. They also need to be able to identify and specify the types of relationships involved.

These requirements are not the same as the network analysis being carried out by sociologists like Duncan J Watts and his colleagues.<sup>7</sup> Their work involves statistically aggregating data about groups of people and their relationships, and using the results to develop models which can predict how social networks operate. The work of researchers like Jordanova and Yeo, on the other hand, requires the ability to identify specific historical individuals and to track their inter-connections and the nature of their relationships. Addressing this kind of requirement has significant implications for the design of future e-research systems in the humanities, and for the metadata on which they need to be based.

### **Biographical services on the web**

There are innumerable biographical services on the web. They vary dramatically in the extent to which they provide information about relationships between people. Traditional dictionaries of biography – even in electronic form – do not explicitly link people who are connected in some way. Both a commercial service like the *Oxford Dictionary of National Biography* and a free service like the *Australian Dictionary of Biography* remain very close to the printed biographical dictionaries on which they are based and do not attempt to build internal links between related entries. *Wikipedia* goes slightly further in that each biographical entry is likely to contain hyperlinks to entries for other named persons, though there is no formal semantic structure to these connections.

Some newer biographical services are beginning to formalise the way in which they record relationships. The *Dictionary of Australian Artists Online*, for example, records the 'associates' of an Australian artist, with an internal hyperlink if the associate also has an entry in the dictionary. These associates are not limited to other artists, but the nature of the relationship is not specified. They appear to be manually entered, and the coverage is far from systematic or thorough.

Other biographical services are experimenting with re-engineering metadata from existing schemas and databases. Particularly interesting is Online Computer Library Center's *WorldCat Identities* service, which builds on the large-scale name authority files assembled over many years by the library sector, and on recent efforts to harmonise these files internationally through projects like Virtual International Authority File (VIAF), Linked European Authority File (LEAF) and ONE Shared Authority Control (ONESAC). The traditional library approach is typified by the MACHine-Readable Cataloging (MARC) record, currently expressed in the MARC 21 format for authority data, as well as related standards such as the International Standard Bibliographic Description (ISBD), Metadata Object Description Schema (MODS), and Anglo-American Cataloguing Rules, Second Edition (AACR2), which is soon to be superseded by Resource Description and Access (RDA).

MARC-type approaches are aimed at establishing a normalised form of personal names to serve as entry-points to bibliographical records in library catalogues. *WorldCat Identities* builds on the fact that many bibliographic records contain more than one name, and uses this to construct a list of 'related identities' for each person. These can be co-authors, editors, subjects, translators and so on. *WorldCat Identities* is important because it separates the personal records from the bibliographic records, and fills out the personal entries with other automatically derived data, including a link to the *Wikipedia* entry for that person.<sup>8</sup> The result is a quasi-biographical service, though one which is largely limited to the person as a bibliographical entity. The nature of the links between related identities is left unspecified more often than not, though some entries include role statements derived from the MARC record, such as editor or composer. Curiously, the relationships are not necessarily reciprocal, that is, the links do not necessarily go both ways.

Biographical services from the archival sector have also been experimenting with recording and identifying relationships between people. The *Bright Sparcs* service, which covers Australian scientists and their archives, includes some information about related people, with lists of 'related entries' which are hyperlinked to other biographical records in this service.<sup>9</sup> *Bright Sparcs* uses ISAAR (CPF) as its metadata

standard, and these related entries are based on the 'relationships area' which is included as part of this standard. As currently implemented, the entries appear to be limited to 'colleague' and a variety of family relationships (husband/wife, parent/child, sibling, uncle/nephew, and so on). The *Australian Women's Register* is based on the same standard. Each biographical entry includes 'related entries' which link to other biographical records in the service. In these entries, family relationships are distinguished from other relationships, which are then explained in a note.<sup>10</sup>

An important new service which brings together elements of both archival and library approaches is *People Australia*, currently under development by the National Library of Australia.<sup>11</sup> Though its name records are being derived largely from the Australian name authority file, *People Australia* uses the Encoded Archival Context (EAC) standard as its metadata schema. EAC is based on ISAAR (CPF) and makes provision for specifying relationships between persons using a list of 'type' attributes. The pilot version available through the prototype of the National Library's Single Business Discovery Service provides lists of 'related people and organisations', derived from the entries in the *Australian Women's Register*, but does not show the specific type or nature of these relationships.<sup>12</sup>

As these services show, we are beginning to see the emergence of people-centred databases in the library and archival sectors. This is an important development, based on a growing recognition that cultural heritage institutions need to go beyond their traditional preoccupation with databases focused on collections and on documentary evidence. For the most part, however, these newer services are based on re-engineered elements from existing metadata schemas, especially the MARC record, and from existing name authority files.

These newer people-centred services are also beginning to move away from the earlier model of largely standalone biographical services, which had little if any interoperability with other services. We are now seeing linkages being built across biographical services, such as those between *WorldCat Identities* and *Wikipedia*, and from *People Australia* to the *Australian Dictionary of Biography*.

These developments are undoubtedly a significant improvement in the presentation of biographical and personal information on the web. They greatly improve access to contextual information about resources – publications, archives, images, objects – which relate to a specific person. But they only go part of the way towards meeting the needs of researchers like those in the ‘intellectual formations’ area of NEER, who are particularly interested in tracing the connections and relationships between people themselves.

### **Using metadata to trace associative relations between people**

Meeting the needs of these researchers will require a software environment where users can follow a network or graph of connected points, each representing a specific person. This network could cover fictional or legendary people as well as real people, to accommodate literary and cultural research as well as historical. In addition to linking to each other, the records for each person would point outwards to biographical information and related resources about that person in other services on the web – including archival and library holdings.

What is required to build a contextual system capable of meeting these requirements? A fundamental building block is an ontology or vocabulary capable of describing a wide range of different types of relationships. Closely linked to this will need to be a metadata schema which can specify how to associate the names of two or more persons using such relationship types.

Underpinning these elements are several more technical components, which are outside the scope of this paper. They include a formal language in which to express these ontologies and schemas, employing RDF (Resource Description Framework), OWL (Web Ontology Language) and SKOS (Simple Knowledge Organization System).<sup>13</sup> Other essential ingredients are software for managing, presenting and browsing semantic networks of this kind, and unique identifiers (such as URIs) representing individual persons or (strictly speaking) individual points on the graph of relationships.

When existing metadata schemas and vocabularies from the archival and library sectors are assessed against these requirements, they are neither



specific nor comprehensive enough. In the 2004 revision of ISAAR (CPF), the archival standard for recording personal and corporate names, there is a section dealing specifically with relationships. Within this section, the instructions offer two approaches for recording the 'category of relationship' - either a separate classification scheme and controlled vocabulary, or one of four general categories: hierarchical, temporal (predecessor/successor), family, and associative (for any other kind of relationship).<sup>14</sup> The category of relationship is supplemented by a 'description of relationship', which can be in a narrative form and does not prescribe a controlled vocabulary.

A similar approach is presented in *Describing Archives: a Content Standard* (DACS) - the North American standard for archival description, published in 2004, which aims to harmonise archival formats like EAD, ISAD (G) and ISAAR (CPF) as well as library standards like MARC and AACR2. DACS deals at some length with rules for describing creators of archival materials and recommends a 'separate system of archival authority records that are linked to the archival descriptions rather than being embedded in them'. Rules 11.14 to 11.16 provide instructions for recording persons, families, and corporate bodies which 'have a significant relationship with the entity named in the authority record'.<sup>15</sup> These include members of families, hierarchical relationships between parts of organisations, chronological relationships between organisations or their parts, and offices held by a person within an organisation. The description may be given as narrative text or as a word or phrase.

The EAC approach is based on that described in ISAAR (CPF) and DACS.<sup>16</sup> In EAC there are several defined values for the RELTYPE attribute, which can be applied to personal and corporate names encoded under the <eacrel/eacrels> elements for related entities:

- 'superior, subordinate': any hierarchical relation;
- 'earlier, later': any temporal relations, such as predecessor, successor;
- 'parent, child': a biological or adoptive relation;
- 'associative': any other relationship, equivalent to 'see also'; and
- 'identity': for linking different EAC instances describing the same entity (for linking to external systems or when it is not possible to remove the duplicate).

EAC also makes it possible to use the TYPE or TYPEAUTH attributes to define additional values describing relationships, using a thesaurus or a local list of some kind. While this is potentially very helpful, it still begs the question of a suitable source of such additional values.

From the library sector, there is a standard list of role designations which can be attached to personal names in a MARC record.<sup>17</sup> They only cover relationships of an authorial or bibliographical kind: editor, translator, and so on. Even within this limited sphere they are far from comprehensive, since some roles – notably author – are taken for granted and not explicitly stated. The draft of the new international cataloguing code (*RDA: Resource Description and Access*) also contains rules for describing relationships between persons, families and corporate bodies, and between these entities and bibliographic records. These include nearly 100 ‘relationship designators’ – terms describing a specific type of relationship.<sup>18</sup> But about 75 of these describe specific roles in the production of a cultural or intellectual work. Most of the remaining designators cover relationships between families and corporate bodies, or relationships between corporate bodies. Here, RDA draws heavily on work done by the International Federation of Library Associations (IFLA) to develop a conceptual model of authority data, now published as *Functional Requirements for Authority Data* (FRAD).<sup>19</sup>

In the world of art museums, the *Getty Union List of Artists' Names* (ULAN) provides a model for describing relationships between artists. It covers both ‘associative relationships’ and ‘parent relationships’, each of which has a list of types. For associative relationships, there is a set of type codes covering such associations as teacher, student, influence, patron, and spouse. For parent relationships, the types are simply ‘parent’ or ‘child’. This semantic structure in ULAN has been used by the Finnish museum project CultureSampo as the basis for a ‘relational semantic search’ system, which enables users to search for chains of relationships between two named people.<sup>20</sup> A similar approach is being tested by the MultimediaN project in the Netherlands, which is using ULAN as the basis for developing a ‘relation search’ across the collections of art galleries and museums.<sup>21</sup>

The CultureSampo project found, however, that the ULAN model for relationships was not sufficient on its own, and ended up drawing on other sources to develop its own 'actor ontology'.<sup>22</sup> One of these sources was FOAF (Friend of a Friend), which began as a web-based method for describing and linking information about individual people and groups in a structured way. The property *foaf:knows* is used to record links to other FOAF descriptions.<sup>23</sup> It is deliberately limited to this vague level of connection, but is hospitable to extensions through other vocabularies. In particular, the *Relationship* vocabulary devised by Ian Davis and Eric Vitiello is intended as an extension for FOAF.<sup>24</sup> This relationship vocabulary defines 33 specific types of relationship, including such properties as: *acquaintanceOf*, *antagonistOf*, *collaboratesWith*, *lostContactWith*, and *wouldLikeToKnow*. It is mainly directed at describing contemporary social relationships – like a more detailed version of Facebook's 'how do you know this person?' – and is unsuitable by itself for documenting relationships identified by researchers. Some of the ULAN relationship types are paralleled in the *Relationship* vocabulary, but many are not.

There are various other ontologies and vocabularies which deal specifically with familial and genealogical relationships. The CIDOC Conceptual Reference Model (CRM) is a high-level ontology for cultural heritage, aimed particularly at the museum sector.<sup>25</sup> It has only limited support for modelling social and familial relationships, though there has been some recent discussion as to how best to expand this.<sup>26</sup> There are also several XML schemas derived from models for encoding genealogical information like GEDCOM and GENTECH, though none of these is expressed as a formal ontology.

The various experiments and projects currently underway give some indication of the potential value of this kind of approach. But none of the metadata schemas and vocabularies discussed above, on its own, is suitable for meeting the needs of researchers like those in the 'intellectual formations' groups within NEER. Exhaustively modelling the types of relationships being traced by this research will require a combination of these sources – ULAN, the FOAF *Relationship* vocabulary, MARC role designators, and EAC RELTYPE attributes – as well as others like the genealogical models.

## Other approaches

These approaches rely, of course, on using pre-existing metadata schemas or ontologies which have been developed by an authoritative body of some kind. A different approach might be the one being taken by the DBPedia project, which is extracting information from *Wikipedia* and using it to derive semantically enhanced content. The extraction focuses on structured templates within *Wikipedia* pages, which are then used to populate ontologies automatically. The extracted content can be searched through a 'relationship finder', which takes two terms input by the user and finds the relationship between them, up to ten steps apart. This approach relies heavily on the quality and structure of the information in *Wikipedia*, and the DBPedia researchers note that the results would be significantly improved by implementing more explicit rules for populating templates and by 'reasonably small modifications' to the MediaWiki software.<sup>27</sup> Instead of using a pre-existing metadata schema or ontology and applying it in the creation of new data, the DBPedia approach is to derive relationships information based on terms used in the textual corpus. But DBPedia's total reliance on *Wikipedia* significantly limits its value in the eyes of scholars, and it is unclear at this stage whether this approach could ever be sufficiently reliable and authoritative for use by researchers. More experimentation with the application of this kind of semantic data mining to text corpora in the humanities – especially those with a biographical focus – is required before an answer can begin to be given.

An alternative is to ensure that text markup includes encoding for people and their relationships. The Text Encoding Initiative (TEI), which is widely used for encoding scholarly editions of humanities texts, has both a <person> element and a <relation> element. The latter could be used to store information about relationships and to categorise them into family, social and so on. The TEI itself does not have a vocabulary for different types of relationships, however, so at this stage it offers no more than a potential framework for extracted embedded information about personal relationships from a text – albeit one which is created by the human encoder of the text, rather than being derived by automated data mining techniques.

A solution to the apparent difficulty and complexity involved in developing an ontology of relationships might be offered by the web service known as *Freebase*.<sup>28</sup> Rather than trying to define a complete data model to encompass all possible aspects of the global knowledge it aims to collect, *Freebase* allows users to define their own types and properties as they go. Types can be either unpublished (private) or published, and properties can be simple (core) or complex. But, analogous to *Wikipedia*, there is no authoritative schema – just a continually evolving structure which reflects the combined work of the contributors to *Freebase*. It remains to be seen whether this can evolve into the kind of service which meets the needs of researchers like those in the ‘intellectual formations’ group. It is hard to imagine that such a bottom-up development would ever develop the rigour and consistency required by researchers, nor that these researchers themselves would be willing – let alone able – to invest the time required to learn and participate in the technicalities of this data-modelling process.

The most feasible approach would seem to be a combination of pre-existing ‘authoritative’ ontologies and a controlled *Freebase*-like method for allowing researchers to define additional types of relationships of interest to them. The potential scale of the work required will almost certainly demand new forms of collaboration. The newer services looked at in this paper are still being largely constructed within an institutional framework, where archivists, librarians, and similar curatorial experts assemble and supply data to end-users. Comparatively little use is being made of distributed ‘social web’ approaches which can directly harness the knowledge of individual scholars and researchers. While most researchers are likely to be sceptical of folksonomy-based approaches, they may be convinced of the value of contributing additional data over the web if this can occur within a suitably authoritative environment.

## Conclusion

Biographical data form a central element of most humanities research, but they have traditionally been published and communicated in ways which are largely self-contained and lacking in interoperable semantic

content. Replicating this kind of approach in the digital environment, while it undoubtedly enriches and assists research, will never result in the kind of systemic change which e-research approaches are producing in the sciences.

Developing a semantic framework for recording and tracing relationships between people for the purposes of humanities research is a major undertaking. It will require an international collaborative effort along the lines of the *Encyclopaedia of Life*, which aims to be a global reference source with entries for each of the 1.8 million species known and named to date.<sup>29</sup> As well as the issues relating to metadata schemas and ontologies discussed above, there are fundamental questions to be addressed about the most feasible and appropriate technical environment and system architecture – particularly the roles of decentralised content providers and centralised resolver services. Also of major importance is how best to harness existing knowledge, by employing a combination of automated data mining, machine conversion of existing datasets, and a framework for direct contributions by scholars and experts.

The archive and library sectors, with their extensive data about people, have a major role to play in these developments. But they will need to contribute to efforts to pool personal data across different sectors rather than simply continuing to build their own services. This means, at a minimum, exposing their data in formats which can be used effectively by other systems. The full implementation of standards like EAC and RDA will enable personal and name data to be managed separately from descriptive records. But more work is needed on clarifying and defining the use of relationship descriptors, and ensuring that they are encoded in a way suitable for machine processing and matching. The use of unique identifiers for persons also needs further work; while a single global identifier service is unrealistic (despite the work of ISNI),<sup>30</sup> each record for a person ought to have a globally unique identifier which can be processed by external resolver services. Data relating to persons must be capable of being exposed in RDF, or in a format which can be transformed readily into RDF triples, for use in semantic services. This includes the vocabularies underpinning these data (geographical, chronological, and conceptual).

All this is part of a much bigger question: the future handling of personal entities and their names by library and archival systems, in a world where new services are being built by extracting, reformatting, and combining data. Standalone systems describing the contents of collections – even across a range of different cultural institutions – are no longer sufficient. Archives and libraries need to be able to make their data available for reuse by others, if they are to contribute effectively to the next generation of e-research technologies and services for the humanities.

### Endnotes

<sup>1</sup> Albert-László Barabási, *Linked: the New Science of Networks*, Perseus Publishing, Cambridge, Mass., 2002, p. 8.

<sup>2</sup> Toby Burrows, 'Reinventing the Humanities in a Networked Environment: the Australian Network for Early European Research', *Humanities, Computers and Cultural Heritage: Proceedings of the XVI International Conference of the Association for History and Computing 14-17 September 2005*, Royal Netherlands Academy of Arts and Sciences, Amsterdam, 2005, pp. 95-99; Stephanie Trigg, "'Medieval Literature" or "Early Europe"? How to Win Grants and Change the Course of Scholarship', *Literature Compass*, vol. 3, no. 3, 2006, pp. 318-330.

<sup>3</sup> Ludmilla Jordanova, *Defining Features: Scientific and Medical Portraits 1660-2000*, Reaktion Books, London, 2000.

<sup>4</sup> John A Schuster and Alan BH Taylor, 'Blind Trust: The Gentlemanly Origins of Experimental Science', *Social Studies of Science*, vol. 27, no. 3, 1997, pp. 503-36.

<sup>5</sup> Luciano Boschiero, 'Natural Philosophizing Inside the Late Seventeenth-Century Tuscan Court', *British Journal for the History of Science*, vol. 35, 2002, pp. 383-410.

<sup>6</sup> Richard Yeo, 'Between Memory and Paperbooks: Baconianism and Natural History in Seventeenth-Century England', *History of Science*, vol. 45, 2007, pp. 1-46.

<sup>7</sup> Peter Sheridan Dodds, Duncan J Watts, and Charles F Sabel, 'Information Exchange and the Robustness of Organizational Networks', *Proceedings of the National Academy of Sciences*, vol. 100, no. 21, 2003, pp. 12516-21; Gueorgi Kossinets and Duncan J Watts, 'Empirical Analysis of an Evolving Social Network', *Science*, vol. 311, no. 1, 2006, pp. 88-90.

<sup>8</sup> Thomas B Hickey, 'WorldCat Identities: Another View of the Catalog', *NextSpace: the OCLC Newsletter*, no. 6, 2007, pp. 18-19.

<sup>9</sup> Gavan McCarthy, 'The Structuring of Context: New Possibilities in an XML Enabled World Wide Web', *JAHc: Journal of the Association for History and Computing*, vol. 3, no. 1, 2000, available at <<http://mcel.pacificu.edu/jahc/2000/issue1/articles/mccarthy/>>, accessed 20 November 2009.

<sup>10</sup> *Australian Women's Register*, available at <<http://www.womenaustralia.info/index.html>>, accessed 20 November 2009.

- <sup>11</sup> Basil Dewhurst, 'People Australia: A Topic Based Approach to Resource Discovery', VALA 2008, *Libraries: Changing Places – Virtual Spaces*, 14th Biennial Conference, Melbourne, February 2008, available at <[http://www.valaconf.org.au/vala2008/papers2008/116\\_Dewhurst\\_Final.pdf](http://www.valaconf.org.au/vala2008/papers2008/116_Dewhurst_Final.pdf)>, accessed 20 November 2009.
- <sup>12</sup> National Library of Australia, Single Business Discovery Service (prototype) was available at <<http://sbdsproto.nla.gov.au>>, the beta service named *Trove* is available at <<http://trove.nla.gov.au/>>, accessed 20 November 2009.
- <sup>13</sup> KK Breitman, MA Casanova, and W Truszkowski, *Semantic Web: Concepts, Technologies and Applications*, Springer, New York, 2007; Alistair Miles and José R Pérez-Agüera, 'SKOS: Simple Knowledge Organization for the Web', *Cataloging and Classification Quarterly*, vol. 43, no. 3–4, 2007, pp. 69–83.
- <sup>14</sup> International Council on Archives, *ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families*, second edition, 2004, available at <[http://www.icacds.org.uk/eng/ISAAR\(CPF\)2ed.pdf](http://www.icacds.org.uk/eng/ISAAR(CPF)2ed.pdf)>, accessed 20 November 2009.
- <sup>15</sup> *Describing Archives: a Content Standard*, Society of American Archivists, Chicago, 2004, pp. 86, 108–9.
- <sup>16</sup> Daniel V Pitti, 'Creator Description: Encoded Archival Context', *Cataloging and Classification Quarterly*, vol. 38, no. 3–4, 2004, pp. 201–26.
- <sup>17</sup> Library of Congress, *MARC 21 Format for Authority Data*, available at <<http://www.loc.gov/marc/authority/ecadhome.html>>, accessed 20 November 2009.
- <sup>18</sup> *RDA*, full draft 24 November 2008, available at <<http://www.rdaonline.org/constituencyreview/>>, accessed 20 November 2009.
- <sup>19</sup> Glenn E Patton (ed.), *Functional Requirements for Authority Data – a Conceptual Model*, Saur, München, 2009.
- <sup>20</sup> Jussi Kurki and Eero Hyvönen, 'Relational Semantic Search: Search Social Paths on the Semantic Web', poster, Proceedings of the International Semantic Web Conference, 2007, Busan, Korea, available at <<http://www.seco.hut.fi/publications/2007/kurki-hyvonen-relational-2007.pdf>>, accessed 20 November 2009.
- <sup>21</sup> Jacco van Ossenbruggen, et al, 'Searching and Annotating Virtual Heritage Collections with Semantic-Web Techniques', *Museums and the Web 2007*, 11–14 April 2007, available at <<http://www.archimuse.com/mw2007/papers/ossenbruggen/ossenbruggen.html>>, accessed 20 November 2009.
- <sup>22</sup> Jussi Kurki and Eero Hyvönen, 'Authority Control of People and Organizations on the Semantic Web', May 2009, available at <<http://www.seco.tkk.fi/publications/2009/kurki-hyvonen-onki-people-2009.pdf>>, accessed 20 November 2009.
- <sup>23</sup> Dan Brickley and Libby Miller, *FOAF Vocabulary Specification 0.91: Namespace Document*, available at <<http://xmlns.com/foaf/spec/>>, accessed 20 November 2009.
- <sup>24</sup> Ian Davis and Eric Vitiello, *RELATIONSHIP: a Vocabulary for Describing Relationships Between People*, available at <<http://purl.org/vocab/relationship/>>, accessed 20 November 2009.
- <sup>25</sup> M Doerr, 'The CIDOC Conceptual Reference Model: An Ontological Approach to Semantic Interoperability of Metadata', *AI Magazine*, vol. 24, no. 3, 2003, pp. 75–92.
- <sup>26</sup> Christian-Emil Ore, *Social and Family Relations and the CIDOC CRM – Some Thoughts*, available at <[http://cidoc.ics.forth.gr/docs/frbr\\_oo/frbr\\_docs/meeting\\_presentations/10th\\_meeting\\_presentations/Fam\\_relation\\_CRM.rtf](http://cidoc.ics.forth.gr/docs/frbr_oo/frbr_docs/meeting_presentations/10th_meeting_presentations/Fam_relation_CRM.rtf)>, accessed 20 November 2009.



<sup>27</sup> Sören Auer and Jens Lehmann, 'What Have Innsbruck and Leipzig in Common? Extracting Semantics From Wiki Content', Proceedings of European Semantic Web Conference 2007, Springer, New York, 2007, pp. 503–17.

<sup>28</sup> K Bollacker, 'Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge', Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, Vancouver, Canada, 9–12 June 2008, ACM, New York, 2008, pp. 1247–50.

<sup>29</sup> Emma Marris, 'First Chapter of Book of Life Goes Live', *Nature*, vol. 451, 2008, p. 1038.

<sup>30</sup> *International Standard Name Identifier*, available at <<http://www.isni.org/index.html>>, accessed 20 November 2009.