

Provenance: crossing boundaries

Jinfang Niu*

Jinfang Niu is an Assistant Professor at the School of Information, University of South Florida. She received her PhD from the University of Michigan, Ann Arbor. Prior to this, she worked as a librarian at the Tsinghua University Library for three years. Her current research focuses on electronic records and digital curation.

Through examining the concept of provenance and its use in several communities, including archives management, computer science, rare book cataloging and archaeology, this paper presents an expanded view of provenance. For end-users, provenance covers the whole life cycle of records, from creation and evolution to acquisition, processing, preservation and access. During each stage of this life cycle, both the sociopolitical context and technical details fall within the scope of provenance.

Keywords: provenance; archives management; computer science

Introduction

The archives management community has discussed the concept and principle of provenance many times. One of the topics in these discussions has been how the definition and use of provenance in neighbouring fields, such as rare book cataloging, museum studies and archaeology, contributes to the archival description of provenance. Further afield, there are multiple subdisciplines of computer science that are also involved in provenance research, such as database, e-science, workflow and the Semantic Web. There is, in fact, a large body of provenance research literature in computer science. According to Moreau,¹ from 1986 to November 2009, there have been 425 publications about provenance in the computer science community. Almost half of these papers were published in the two years prior to 2009. However, the archives community seems hesitant or perhaps intimidated when it comes to exploring the concept of provenance in computer science research literature. Similarly, while the concept and principle of provenance is so significant and fundamental in archives management, computer scientists seem unaware of this fact. They occasionally mention provenance research for assessing authenticity and justifying market values of artworks, but have rarely discussed how provenance is used and described in archives management.

It is important for archivists and computer scientists to learn about provenance research in each other's field. Archival records are increasingly produced and preserved in databases, workflows and other computer systems. Capturing provenance information in these systems, which is the concern of computer scientists, also means capturing provenance for electronic records and digital archives, which is a concern for archivists. Also, archival materials and description will be increasingly published and accessed on the Web, sometimes in the form of linked data, and might be automatically processed, aggregated or mashed-up with other information resources. Recording, managing and

*Email: Jinfang@usf.edu

accessing the provenance information of web resources, especially linked data, is an important topic in the computer science community and will be relevant for publishing archival descriptions on the Web.

In this paper, I will compare the meaning and use of the concept of provenance in several communities, mostly in archives management and computer science, and present an expanded view of provenance. It is anticipated that this analysis can bring provenance research in computer science and archives management closer and improve the dialogue between these two disciplines.

Provenance in traditional archives management

In the physical arrangement of records, provenance means the creator and/or the components of the creator, which can be individuals, corporate bodies or families. In arranging records from multiple origins, the principle of provenance essentially dictates that records from different creators are separated and records from the same creator are collocated, although preservation concerns sometimes require otherwise – for example, photographs are stored separately from paper records, even though they belong to the same series. In the internal arrangement of records from the same creator, especially large complex organisations, the principle of provenance further dictates that records from different components of the same creator are separated and records from the same component of the creator are collocated.

The ‘components’ of a creating organisation can be defined based on structure (organisational units) or function. In other words, records may be classified and arranged based on structural or functional provenance. In cases where records have already been well-organised, based on either structural or functional provenance, upon acquisition, compliance with the principle of provenance for the internal arrangement of records naturally entails respecting the original order. The archives management community has recognised the complex, many-to-many relationships between provenance and records.² When one archival collection is sourced from multiple provenances or one provenance contributes to multiple archival collections, the complex relationships cannot be represented entirely in physical arrangement of the records and, thus, need to be further described in archival descriptions. An arrangement based on provenance makes it possible for records to be retrieved from storage based on provenance. If records are also described and indexed based on provenance, then provenance can serve as an access point in an archival finding aid system. The creator, its functions and structure can all be used as access points. Several decades ago, Bearman and Lytle argued that provenance can be used as an access point and suggested allowing searches by function and documentary forms in archival information systems and even general information systems.³ Since then, there have been various functional thesauri created to provide controlled access points for functions, such as the *Australian Governments’ Interactive Functions Thesaurus* (AGIFT).⁴ Provenance is also the basis for macro-appraisal, which may directly appraise the creators themselves, as in the case of the Minnesota method,⁵ or the structure and function of the creating organisation, as in the functional and structural analysis method proposed by Cook.⁶

As discussed above, provenance means the creator and its functions and internal structures in archival appraisal, arrangement and access. In archival description, provenance has a richer meaning. Based on an examination of the term provenance in archaeology and museology, Millar suggested expanding the definition of archival provenance to encompass creator history, records history and custodial history. According to Millar, creator history enlarges existing archival provenance to accommodate organisational and functional changes over time. All of the agents who were involved

in the creation, accumulation and utilisation of the records over time and space should be described. Records history is the history of recordkeeping:

how records were created and used; who had them and when; where they were moved to and why; and whether any records were lost or destroyed, enhanced or altered, and why, up to and including the time they were transferred into archival custody.⁷

Finally, custodial history describes: ‘the transfer of ownership or custody of the records from the creator or custodian to the archival institution and the subsequent care of those records’.⁸

The three kinds of provenance information discussed by Millar can be seen in today’s archival description standards. In the General International Standard Archival Description (ISAD(G)), provenance-related elements include the name of creator, administrative/biographical history, archival history, immediate source of acquisition, appraisal, destruction and scheduling. This shows that not only the creator, but also the creator’s history, the custodial history of archival materials, as well as the appraisal and acquisition information, will be included in provenance description. Encoded Archival Description (EAD), as a standard based on ISAD(G), further enriches the description of provenance information. The *<origination>* element contains the creator, collectors, dealers and various other agents who are involved in the creation, accumulation and assembly of records. The *<processinfo>* element records both the provenance information before acquisition and activities occurring to the records after acquisition, including accessioning, arranging, describing, preserving, storing or otherwise preparing the archival materials for secondary use. While activities occurring to records after acquisition may not be provenance information for archivists upon acquisition, it is provenance information for end-users, because it helps them understand how the records came into their current state of existence. International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR(CPF)) and Encoded Archival Context (EAC) allow much more detailed descriptions of creators and other agents involved in creating and preserving archival records during the evolutionary history of records. They include not only the name and identity of the agent, its functions and internal structure, but also its history, mandates and relationships with other agents.

In addition to the kinds of provenance information discussed above, several scholars have suggested including the sociopolitical context in which the records were created and evolved in provenance description. For example, Nordland argued that political power structures affect how geographic features and ethnic groups are represented on historical maps and, thus, need to be included in archival description to help users interpret the map; Wurl wrote that the ethnicity of records creators should be included;⁸ and Beattie suggested including motivations for keeping records, intended audiences of records and even the change in the conventions involved in creating particular kinds of records, such as diaries.⁹

Provenance in rare book cataloguing and archaeology

While much can be discussed about the similarities and differences between provenance research in rare book cataloguing, archaeology and archives management, one finding derived from the provenance research literature in these fields is particularly relevant: when the provenance is uncertain or unknown, cataloguers record evidence that helps to determine or infer the provenance. The archival description standards mentioned above assume that the creator, history, function and structure are known; however, this

is often not the case with rare books and archaeological objects. Unlike archivists, who often write detailed narratives about the creator and its history, rare book cataloguers often record evidence of previous ownership, such as bookplates, signatures, inscription, stamps, marginal annotations (*marginalia*) and branded bindings, which are often present in rare books, as well as external evidence, such as auction records.¹⁰ This approach of recording evidence of provenance is even more evident in archaeology, where the creation provenance of discovered objects is usually unknown, due to the very long temporal distance between an object's creation and its discovery. The archaeology community created the term *provenience* – a derivation from the term *provenance* – to refer to discovery provenance. *Provenience* means the place where an object was found or recovered by archaeologists.¹¹ The description of the *provenience* of an object includes 'its juxtaposition to other objects in situ, its relationship to those objects, and the strata above and below the level at which the object is found in the excavation'.¹² This detailed description helps to infer the creation provenance. For example, based on the description of the provenance, an archaeologist may determine that a discovered stone was part of a temple built 3000 years ago.

Provenance research in computer science

Provenance in computer science is defined similarly to that used in archival description: the origin, creation, transformation and derivation of data and information resources. For example, Buneman et al. defined data provenance in database systems as the description of the origin of data and the process by which it arrived at the database.¹³ Lanter defined the provenance of Graphic Information System (GIS) data as information describing materials and transformations applied to derive the data.¹⁴ Greenwood et al. viewed provenance as metadata recording the process of experiment workflows, annotations and notes about experiments.¹⁵ Despite these similarities, there are many specific differences in the scope, understanding and use of provenance information used in computer science, when compared to those in archives management.

Compared with archivists, computer scientists are less concerned with the social and political context and more concerned with the technical details of the creation and transformation process. They do, however, recognise the social aspect of provenance. For example, Harth et al. proposed a 'social dimension to associate provenance with the originator (typically a person) of a given piece of information'.¹⁶ The *Provenance Vocabulary* includes human agents that are involved in the creation and access of linked data.¹⁷ However, many technical elements are also included. Non-human agents are defined in their provenance models and vocabularies. The creation, transformation and derivation processes described by those provenance models and vocabularies are often very technical. Data creation and transformation may be through workflows or algorithms, rather than human actions or historical events. For example, a data creation process can be the completion of a web form, and a data transformation process can be the addition of one to all numbers in a dataset. These kinds of technical provenance information are audit trails, which may capture every action that impacts upon the data. They are more detailed than archival history, which might span the whole life of an individual, family or organisation, in terms of years, decades or even centuries. These technical audit trails can also be proactively captured during the ongoing processes, rather than retrospectively traced, as has typically occurred in traditional archival description.

Much archival provenance information is created for, and consumed by, human users, often in narrative or loose-structured form, such as the administrative history or

biography and archival history. Computer scientists are very concerned with representing provenance information in machine-processing form, such as RDF/XML format. They also distinguish annotations (that is, provenance metadata generated either manually or automatically) from provenance information, deduced indirectly through inversion. In this latter method, the output data and derivation method are recorded as provenance information. Given this provenance information, the input data can be retrieved through technically inverting the derivations process.¹⁸ For example, given the query result and the query, the source data can be derived. This is a unique method of recording provenance information that is unknown in traditional archival provenance description. Archivists also derive provenance information from records. However, this is largely an intellectual process, which is based on human interpretation, rather than the technical process involved in computer science.

Similar to archivists, computer scientists recognise that provenance metadata applies to multiple levels of resources. They distinguish between coarse-grained and fine-grained provenance, which mean different things in different contexts. For Tan,¹⁹ coarse-grained and fine-grained provenance corresponds to workflow provenance and data provenance, respectively. Workflow provenance describes the entire history of the derivation of the final output of a workflow. It may involve the recording of software programs, the hardware and the instruments used in the workflow. Data provenance is about the derivation of single pieces of data. Ding et al. discuss provenance for linked data.²⁰ They consider provenance for Resource Description Framework (RDF) graphs, which contain many RDF triples,²¹ as coarse-grained, and provenance for single RDF triples as fine-grained. They have also introduced provenance for RDF molecules, which is an intermediary level between RDF graphs and RDF triples. Zhao et al. suggested that there can be a continuum between the two extremes (that is, provenance for an RDF graph or an RDF triple), and the appropriate level of granularity can be determined based on the needs and resources of a particular application.²² Compared with these, archival provenance can be far more coarse-grained and less fine-grained. Archivists usually describe the provenance of a whole archival collection. Occasionally, they describe the provenance of a single record, but do not go beyond the record level. These detailed provenance descriptions from computer science also apply to digital archives, where a single record can have lower-level components, and when archival records are produced in databases or workflows or published on the Web as linked data.

Computer scientists include information that helps assess the authenticity of data and determine the appropriate use of data in their scope of provenance information. Therefore, they consider technical measures for assessing authenticity, such as digital signatures and public keys,²³ as well as licensing information, use restriction and copyright and ownership as provenance information.²⁴ Accordingly, for computer scientists in the Dublin Core elements set, not only is there an element named *provenance*, which is defined as a 'statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation',²⁵ but there are also another 24 elements that are provenance-related, such as *dct:available* and *dct:valid*, *dct:contributor*, *dct:creator*, *dct:publisher*, *dct:rightsHolder*, *dct:isVersionOf*, *dct:isFormatOf*, *dct:replaces*, *dct:source*, *dct:references*, *dct:license* and *dct:Rights*.

The Web Science and Semantic Web communities include information about the access process as provenance information, because this information helps consumers assess the authenticity of retrieved web resources. The Provenance Vocabulary describes not only the data creation process, but also the data access process, such as how web servers retrieve the digital image and deliver it to the browser and how the browser ren-

ders the image.²⁶ When archival descriptions are published on the Web as linked open data, this kind of provenance information can be used to supplement archival provenance information to help users judge the authenticity and reliability of archival materials.

Due to its different scope and focus, provenance information in computer science has different functionalities than that of archives management. Archival provenance information traditionally helps archivists to arrange records and helps users understand and interpret records. Computer scientists do not seem much concerned with the role of provenance in data storage or as access points, probably because they often talk about provenance of data within particular software systems, unlike archival repositories, which aggregate data from multiple provenances. In Web Science, especially Web 2.0 and the Semantic Web, provenance information plays an important role in assessing the authenticity of data, due to the distributed nature of data creation. In an archival repository, although authenticity is of great concern, it is often presumed by users, because archival records are rigorously appraised, selected and preserved by a trustworthy custodian. Computer scientists also use provenance information to assess the currency and timeliness of data. This is very different from the archives management community, where archival information is, by default, historical. In e-science, detailed documentation of the data creation and transformation process helps to ensure the reproducibility of scientific data and verification of findings. Provenance information in workflow systems helps troubleshooting and optimisation of efficiency, demonstrating compliance with regulatory requirements and underpinning accountability.²⁷

The computer science community has created many models and vocabularies for recording provenance information in various contexts. The Open Provenance Model (OPM) is created for the interoperability of existing provenance models and vocabularies and exchange of provenance information across systems.²⁸ It consists of a common core found in many provenance vocabularies. OPM describes provenance in terms of processes, artefacts and agents, as well as the relationships among these entities. As a provenance model, it emphasises the causal and derivative relationships within entities: one artefact was derived from another artefact, and one process was triggered by another process. It also includes relationships between entities, describing how the artefacts are derived; how a process used and generated artefacts; and, how a process was controlled by an agent. It does not model the hierarchical relationships within entities: one agent is part of another agent; one process is part of another process; or, one artefact is part of another artefact.

The SPIRT recordkeeping metadata model has a structure similar to the OPM model, although it was not created solely for recording provenance information. The records, business and agent entities in SPIRT map onto the artefact, process and agent classes in OPM, respectively. The records entity in SPIRT is one kind of artefact in OPM. Similar to OPM, an agent in SPIRT can cover both human and non-human agents. In SPIRT, descriptions about businesses and agents can be considered provenance information for the records. Unlike the OPM model, SPIRT does not model causal and derivative relationships within entities. It emphasises the hierarchical structure of agents, businesses and records and supports the description of the component parts of a records collection, the internal structure of an organisational agent or family, as well as the breakdown of a function into events and transactions. ISAAR(CPF) and the International Standard for Describing Functions (ISDF)²⁹ expand the relationships between agents and functions to include sequential and temporal relationships – for example, an agent or function is succeeded by another agent or function. They also

describe various other associative relationships within agents and functions. However, causal and derivative relationships are still not the focus of archival description. Despite these differences, the similarities between the two models make it possible for archival provenance information to be converted into OPM and then processed and aggregated, together with provenance information from other sources.

Computer systems are where electronic records and digital archives are created. Provenance information captured in these computer systems can be used to enrich and expand provenance information that is traditionally included in archival description. Archivists and records managers may also need to appraise, select or summarise provenance information from these computer systems and transfer those with archival value to an archival repository.

Provenance in electronic records management and digital archiving

The approach taken to provenance in digital archiving and electronic records research has already demonstrated similarities with those of computer science. Similar to what has been revealed by the Provenance Vocabulary, archivists have also recognised the technical process of delivering and accessing digital records as a kind of provenance metadata. For example, Nordland discussed that computer terminals, software viewers and monitors may affect a user's perception of digital records and, thus, be part of the provenance of digital records.³⁰ Nesmith also argued that the capacity of information technologies to capture and preserve information at any given time is a kind of provenance information. The recordkeeping metadata standard ISO 23081 allows us to proactively capture very detailed audit trails in the records management process.

Electronic records researchers have paid close attention to the authenticity of electronic records, due to their fragility, the ease with which they can be tampered with and their dependence on technologies. The InterPARES project provides a list of benchmark requirements for preservers to appraise the authenticity of records upon acquisition. The requirements include: who created, handled or transmitted the records and at what time; whether there are access controls and protective procedures to prevent the loss and corruption of records; and, whether there is a guarantee of the integrity of records against media deterioration and technology obsolescence.³¹ The InterPARES project also provides a list of baseline requirements for preservers to attest to the authenticity of copies of archived electronic records. These requirements record the transfer of records to archival institutions, their preservation and the reproduction process. The benchmark and baseline requirements fall within the scope of provenance information defined by computer scientists. They also correspond to the two types of provenance information mentioned in the 2012 version of the OAIS model: provenance information provided by the producer and provenance information created by the archives from the point of ingest.³²

Table 4–1 in the 2012 version of OAIS presents an example of provenance information for space science data, digital library collections and software packages. A careful examination of these elements shows that they are consistent with the computer science view of provenance. The elements describe all three main classes in the OPM model and Provenance Vocabulary. There are elements for human agents, such as *principal investigator*, and for non-human agents, such as data-gathering sensors. There are also elements for processes, such as processing history, storage and handling history, digitisation process, preservation process, change history and revision. Some elements record the derivative relationships between artefacts, such as pointers to the originals,

master versions and earlier versions of digitised material. Copyright information and digital signatures are also included.

Broadened view of provenance

The various kinds of provenance information presented in the foregoing discussion demonstrate an expanded view of provenance. For end-users, provenance information may cover the whole life cycle of records, from creation and evolution to archival acquisition, processing, preservation and access. During each stage of the life cycle, both the sociopolitical context and technical details fall within the scope of provenance. However, not all kinds of provenance information need to exist in all kinds of contexts. Exactly what kinds of provenance information are needed depends on the purposes of use and the temporal, geographical and intellectual distance between the records and consumers. When the original provenance is uncertain or unknown, evidence of the provenance can be recorded to help users determine or infer the original provenance.

The most basic provenance description may only include the name or identity of the creator. This basic description may be sufficient in certain scenarios. For example, knowing that records are from the White House's website, most people would believe the information to be authentic and reliable. When record users lack background knowledge about the record creator, due to their temporal, geographic or intellectual distance from the provenance, more detailed information about the creator needs to be provided. For example, employees of an organisation do not need much provenance information to understand their organisational records, because they already have that knowledge in their minds. However, when the records are transferred to archives, tacit knowledge needs to be made explicit for secondary users. An expert in quantum physics only needs the author's name and affiliation as the provenance information for a paper in the same field. Whereas a user who is unfamiliar with quantum physics may need more detailed provenance information to be confident that what he or she is reading is authoritative and trustworthy. Temporal distance comes to all archival materials, which, by definition, are historical. Archivists need to keep this in mind and provide sufficient provenance information to help users cross the temporal distance, in order to understand records.

Moving forwards in the life cycle of records, provenance information itself may evolve, similar to the evolution of records. More provenance information may be created and accumulated along the way. In the meantime, unnecessary provenance information may be removed. For example, audit trails in workflow or database systems are very detailed. These audit trails may be pruned and summarised when records pass the archival threshold, in order to reduce storage cost and avoid confusing users. In the archival repository, provenance information for archival processing and preservation may be added, accumulated and then pruned again when needed. Eventually, archival provenance information will cover a longer time period, but become less detailed than typical audit trails in a records creation system.

Despite the broader view of provenance, this author does not suggest that archivists manually create more provenance description than what is traditionally undertaken. In fact, even the existing fields for provenance in archival description standards are woefully underused.³³ With the wider adoption of the More Product Less Processing (MPLP) concept and the reality of large backlogs, archivists are unlikely to spend more time on provenance description. However, several measures can be utilised to provide richer archival provenance description, without increased human labour. First, by

inheriting and re-using provenance information created in database systems, workflows, web servers and various other software applications. Second, by linking to provenance information already existing on the Web. For example, a detailed biography of the donor, description of the geographic or jurisdictional region where the donor comes from and the ethnic group that the donor belongs may already exist on Wikipedia. Archivists can link to these Wikipedia entries and save time in writing detailed biographies. Third, archivists can utilise the power of crowdsourcing, by allowing users and volunteers to create provenance information.

Conclusion

In traditional archives management, the meaning of provenance has been expanded from encompassing merely the creator and its functions and internal structures to encompassing creator history, records history and custodial history. By crossing boundaries and examining provenance research in other communities, especially the computer science community, archivists can see an even broader view of provenance. For the end-users, provenance information may cover the whole life cycle of records, from creation and evolution to archival acquisition, processing, preservation and access. During each stage of the life cycle, both the sociopolitical context and technical details fall within the scope of provenance. This broader view of provenance is especially important for managing and describing the provenance of electronic records that are created in computer systems and increasingly published on the Web as linked open data.

Endnotes

1. Luc Moreau, 'The foundations for provenance on the web', *Foundations and Trends in Web Science*, vol. 2, nos 2–3, November 2010, pp. 99–241.
2. Chris Hurley, 'Problems with Provenance', *Archives and Manuscripts*, vol. 23, no. 2, July 1995, pp. 234–59.
3. David A Bearman and Richard H Lytle, 'The Power of the Principle of Provenance', *Archivaria*, vol. 21, Winter 1985, pp. 14–27.
4. National Archives of Australia, 'A Standard Framework for Describing the Functions of Government (AGIFT)', available at <<http://www.naa.gov.au/records-management/publications/agift.aspx>>, accessed 9 March 2013.
5. Mark A Greene and Todd J Daniels-Howell, 'Documentation with an Attitude: A Pragmatist's Guide to the Selection and Acquisition of Modern Business Records', in James M O'Toole (ed.), *Records of American Business*, Society of American Archivists, Chicago, 1997, p. 168.
6. Terry Cook, 'Macroappraisal in Theory and Practice: Origins, Characteristics, and Implementation in Canada, 1950–2000', *Archival Science*, vol. 5, nos 2–4, December 2005, pp. 101–61.
7. Laura Millar, 'The Death of the Fonds and the Resurrection of Provenance: Archival Context in Space and Time', *Archivaria*, vol. 53, Spring 2002, p. 13.
8. *ibid.*, pp. 1–15.
9. Joel Wurl, 'Ethnicity as Provenance: In Search of Values and Principles for Documenting the Immigrant Experience', *Archival Issues*, vol. 29, no. 1, 2005, pp. 65–76.
10. Heather Beattie, 'Where Narratives Meet: Archival Description, Provenance and Women's Diaries', *Libraries and the Cultural Record*, vol. 44, no. 1, February 2009, pp. 82–100.
11. David Pearson, 'Exploring and Recording Provenance: Initiatives and Possibilities', *The Papers of the Bibliographical Society of America*, vol. 91, no. 12, December 1997, pp. 505–15.
12. Shelley Sweeney, 'The Ambiguous Origins of the Archival Principle of "Provenance"', *Libraries and the Cultural Record*, vol. 43, no. 2, May 2008, pp. 193–213.

13. Jessica L Darraby, *Art, Artifact and Architecture Law*, Clark Boardman Callaghan, Deerfield, Illinois, 1995, pp. 6–51.
14. Peter Buneman, Sanjeev Khanna and Tan Wang-Chiew, 'Why and Where: A Characterization of Data Provenance', in Jan Van den Bussche and Victor Vianu, *Database Theory*, ICDT 2001, Springer Berlin Heidelberg, 2001, pp. 316–330.
15. David P Lanter, 'Design of a Lineage-Based Meta-Data Base for GIS', *Cartography and Geographic Information Science*, vol. 18, no. 4, October 1991, pp. 255–61.
16. Mark Greenwood, CA Goble, Robert D Stevens, Jun Zhao, Matthew Addis, Darren Marvin, Luc Moreau and Tom Oinn, 'Provenance of E-Science Experiments-Experience from Bioinformatics', in Simon J Cox, *Proceedings of UK e-Science All Hands Meeting*, EPSRC, September 2003, pp. 223–226, available at <http://www.nesc.ac.uk/events/ahm2003/AHMCD/ahm_proceedings_2003.pdf>, accessed 17 June 2013.
17. Andreas Harth, Axel Polleres and Stefan Decker, 'Towards a Social Provenance Model for the Web', Workshop on Principles of Provenance (ProPr), 2007, available at: <<http://aran.library.nuigalway.ie/xmlui/bitstream/handle/10379/527/harth-et-al-2007.pdf?sequence=1>>, accessed 16 June 2013.
18. Olaf Hartig, 'Provenance Information in the Web of Data', in Christian Bizer, Tom Heath, Tim Berners-Lee, and Kingsley Idehen, *Proceedings of the 2nd Workshop on Linked Data on the Web*, LDOW 2009, Madrid, Spain, April 2009, available at <http://CEUR-WS.org/Vol-538/ldow2009_paper18.pdf>, accessed 17 June, 2013.
19. Yogesh L Simmhan, Beth Plale and Dennis Gannon, 'A Survey of Data Provenance in E-Science', *SIGMOD Record*, vol. 34, no. 3, September 2005, pp. 31–36.
20. Wang Chiew Tan, 'Provenance in Databases: Past, Current, and Future', *IEEE Data Engineering Bulletin*, vol. 30, no. 4, December 2007, pp. 3–12.
21. Li Ding, Tim Finin, Yun Peng, Paulo Pinheiro Da Silva and Deborah L McGuinness, 'Tracking RDF Graph Provenance using RDF Molecules', in Yolanda Gil, Enrico Motta, V. Richard Benjamins and Mark Musen, *Proceedings of the 4th International Semantic Web Conference (Poster)*, 2005, available at <ftp://www.ksl.stanford.edu/pub/KSL_Reports/KSL-05-06.pdf>, accessed 17 June 2013.
22. A RDF triple is a RDF statement that includes three parts: the subject, predicate and object. RDF triples are the foundations of linked data. On the Web of linked open data, data comes from all sorts of provenances. People need to know information about the provenance to assess the reliability and trustworthiness of the data. This is similar to the reason why provenance information is important on the document web.
23. Jun Zhao, Alistair Miles, Graham Klyne and David Shotton, 'Linked Data and Provenance in Biological Data Webs', *Briefings in Bioinformatics*, vol. 10, no. 2, March 2009, pp. 139–52.
24. Hartig, 'Provenance Information in the Web of Data'.
25. Yolanda Gil, James Cheney, Paul Groth, Olaf Hartig, Simon Miles, Luc Moreau and Paulo Pinheiro da Silva, 'Provenance XG Final Report', 8 December 2010, available at <<http://www.w3.org/2005/Incubator/prov/XGR-prov/>>, accessed 29 May 2013.
26. Simon Miles, Craig M. Trim and Michael Panzer, 'Dublin Core to PROV Mapping', December 2012, available at <<http://www.w3.org/TR/2012/WD-prov-dc-20121211/>>, accessed 17 June 2013.
27. Olaf Hartig and Jun Zhao, 'Publishing and Consuming Provenance Metadata on the Web of Linked Data', in Deborah L. McGuinness, Luc Moreau and James R. Michaelis, *Provenance and Annotation of Data and Processes*, Springer Berlin Heidelberg, 2011, pp. 78–90, available at <https://cs.uwaterloo.ca/~ohartig/files/HartigZhao_Provenance_IPAW2010_Preprint.pdf>, accessed 17 June 2013.
28. SB Davidson, SC Boulakia, A Eyal, B Ludäscher, TM McPhillips, S Bowers, MK Anand and J Freire, 'Provenance in Scientific Workflow Systems', *IEEE Data Engineering Bulletin*, vol. 30, no. 4, December 2007, pp. 44–50.
29. Luc Moreau, Ben Clifford, Juliana Freire, Joe Futrelle, Yolanda Gil, Paul Groth, Natalia Kwasnikowska, Simon Miles, Paolo Missiere, Jim Myersg, Yogesh Simmhanf, Eric Stephang and Jan Van den Bussche, 'The Open Provenance Model Core Specification (v1.1)', *Future Generation Computer Systems*, vol. 27, no. 6, June 2011, pp. 743–56.
30. International Council on Archives, 'International Standard for Describing Functions', available at <http://www.mcu.es/archivos/docs/CE/ISDF_ENG_definitiva.pdf>, accessed 12 March 2013.

31. Lori Podolsky Nordland, 'Studies in Documents: The Concept of "Secondary Provenance": Re-interpreting Ac ko mok ki's Map as Evolving Text', *Archivaria*, vol. 58, Fall 2004, pp. 147–59.
32. InterPARES Authenticity Task Force, 'Requirements for Assessing and Maintaining the Authenticity of Electronic Records', available at <http://www.interpares.org/book/inter-pares_book_k_app02.pdf>, accessed 12 March 2013.
33. CCSDS, 'Reference Model for an Open Archival Information System', available at <<http://public.ccsds.org/publications/archive/650x0m2.pdf>>, accessed 12 March 2013.
34. Millar, 'The Death of the Fonds and the Resurrection of Provenance', pp. 1–15.