The Foundations for Provenance on the Web

The Foundations for Provenance on the Web

Luc Moreau

University of Southampton
UK

L.Moreau@ecs.soton.ac.uk

now

the essence of knowledge

Boston – Delft

Foundations and Trends® in Web Science

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is L. Moreau, The Foundations for Provenance on the Web, Foundation and Trends® in Web Science, vol 2, nos 2–3, pp 99–241, 2010

ISBN: 978-1-60198-386-2

Foundations and Trends® in Web Science
Volume 2 Issues 2–3, 2010

Editorial Board

Editorial Scope

Foundations and Trends® in Web Science will publish survey and tutorial articles in the following topics:

- Agents and the Semantic Web
- Content Management
- Databases on the Web
- Data Mining
- Dependability
- Emergent behaviour
- Human-Computer Interaction
- Hypertext/hypermedia
- Languages on the Web
- Mobile/pervasive
- Multimedia
- Network infrastructures
- Performance
- Scalability
- Security
- Semantic Web — metadata, ontologies
- Standards
- Trust and Provenance (Policy)
- Universal Usability
- User Interfaces
- Web Searching/Information Retrieval
- Web Services

Information for Librarians
Foundations and Trends® in Web Science, 2010, Volume 2, 4 issues. ISSN paper version 1555-077X. ISSN online version 1555-0788. Also available as a combined paper and online subscription.

Foundations and Trends® in
Web Science
Vol. 2, Nos. 2–3 (2010) 99–241

DOI: 10.1561/1800000010

now
the essence of knowledge

The Foundations for Provenance on the Web

Luc Moreau

University of Southampton, UK, L.Moreau@ecs.soton.ac.uk

Abstract

Provenance, i.e., the origin or source of something, is becoming an important concern, since it offers the means to verify data products, to infer their quality, to analyse the processes that led to them, and to decide whether they can be trusted. For instance, provenance enables the reproducibility of scientific results; provenance is necessary to track attribution and credit in curated databases; and, it is essential for reasoners to make trust judgements about the information they use over the Semantic Web.

As the Web allows information sharing, discovery, aggregation, filtering and flow in an unprecedented manner, it also becomes very difficult to identify, reliably, the original source that produced an information item on the Web. Since the emerging use of provenance in niche applications is undoubtedly demonstrating the benefits of provenance, this monograph contends that provenance can and should reliably be tracked and exploited on the Web, and investigates the necessary foundations to achieve such a vision.

Multiple data sources have been used to compile the largest bibliographical database on provenance so far. This large corpus permits the analysis of emerging trends in the research community. Specifically, the CiteSpace tool identifies clusters of papers that constitute

research fronts, from which characteristics are extracted to structure a foundational framework for provenance on the Web. Such an endeavour requires a multi-disciplinary approach, since it requires contributions from many computer science sub-disciplines, but also other non-technical fields given the human challenge that is anticipated.

To develop such a vision, it is necessary to provide a definition of provenance that applies to the Web context. A conceptual definition of provenance is expressed in terms of processes, and is shown to generalise various definitions of provenance commonly encountered. Furthermore, by bringing realistic distributed systems assumptions, this definition is refined as a query over assertions made by applications.

Given that the majority of work on provenance has been undertaken by the database, workflow and e-science communities, some of their work is reviewed, contrasting approaches, and focusing on important topics believed to be crucial for bringing provenance to the Web, such as abstraction, collections, storage, queries, workflow evolution, semantics and activities involving human interactions.

However, provenance approaches developed in the context of databases and workflows essentially deal with closed systems. By that, it is meant that workflow or database management systems are in full control of the data they manage, and track their provenance within their own scope, but not beyond. In the context of the Web, a broader approach is required by which chunks of provenance representation can be brought together to describe the provenance of information flowing across multiple systems. For this purpose, this monograph puts forward the Open Provenance Vision, which is an approach that consists of controlled vocabulary, serialisation formats and interfaces to allow the provenance of individual systems to be expressed, connected in a coherent fashion, and queried seamlessly. In this context, the Open Provenance Model is an emerging community-driven representation of provenance, which has been actively used by some 20 teams to exchange provenance information, in line with the Open Provenance Vision.

After identifying an open approach and a model for provenance, techniques to expose provenance over the Web are investigated. In particular, Semantic Web technologies are discussed since they have been successfully exploited to express, query and reason over provenance.

Symmetrically, Semantic Web technologies such as RDF, underpinning the Linked Data effort, are analysed since they offer their own difficulties with respect to provenance.

A powerful argument for provenance is that it can help make systems transparent, so that it becomes possible to determine whether a particular use of information is appropriate under a set of rules. Such capability helps make systems and information accountable. To offer accountability, provenance itself must be authentic, and rely on security approaches, which are described in the monograph. This is then followed by systems where provenance is the basis of an auditing mechanism to check past processes against rules or regulations. In practice, not all users want to check and audit provenance, instead, they may rely on measures of quality or trust; hence, emerging provenance-based approaches to compute trust and quality of data are reviewed.

Contents

1 Introduction **1**

1.1 Drivers for Provenance 2
1.2 Provenance for Web Science 4
1.3 A Web Science View of Provenance 5

2 Analysis of the Provenance Literature **7**

2.1 The Provenance Bibliography 7
2.2 New Research Fronts 9
2.3 Analysis of Research Trends 12
2.4 Summary 18

3 Definition of Provenance **21**

3.1 Dictionary Definition 22
3.2 Definition of Provenance in Computer Systems 23
3.3 Mashup Exemplar Application 24
3.4 Alternative Definitions of Provenance 26
3.5 Assumptions 29
3.6 Provenance: A Query over Process Assertions 32
3.7 Summary 33

4 Provenance in Workflows and Databases **35**

4.1 Views and Abstraction 38
4.2 Data Collections and Streams 40
4.3 Efficient Storage of Provenance 43
4.4 Querying Provenance 45
4.5 Workflow Evolution 48
4.6 Provenance Semantics 48
4.7 Human-Driven Workflows 50
4.8 Summary 51

5 The Open Provenance Vision **53**

5.1 Broadening the Scope of Provenance 54
5.2 Provenance-Aware Monolithic Application 56
5.3 Provenance Inter-operability Across Components 57
5.4 The Provenance Challenge Series 60
5.5 The Open Provenance Model 62
5.6 Provenance in Open Systems 64
5.7 Summary 67

6 Provenance, the Web and the Semantic Web **69**

6.1 Publishing Provenance on the Web 70
6.2 Semantic Web Techniques for Provenance 71
6.3 Provenance for RDF 73
6.4 Knowledge and Web Provenance 75
6.5 Summary 76

7 Accountability **79**

7.1 Provenance and Security 81
7.2 Accountability 86
7.3 Data Quality and Trust 89
7.4 Alternative Approaches 92
7.5 Summary 92

8 Conclusion **95**

8.1 The Benefits of Provenance on the Web 97
8.2 Future Research 98

Acknowledgments **101**

Provenance Bibliography **103**

References **119**

1

Introduction

Provenance, i.e., the origin or source of something, is becoming an important concern for several research communities in computer science, since it offers the means to verify data products, to infer their quality, to analyse the processes that led to them, and to decide whether they can be trusted. In fact, provenance is an intrinsic property of data, which gives data value, when accurately captured. To motivate the need for provenance, its potential benefits are reviewed in several contexts: e-science, curated databases and Semantic Web. Furthermore, the provenance philosophy is showed not to be restricted to data in computer systems, but also to apply to real-life artefacts, such as ingredients in the food industry, parts in manufacturing and works of art (Section 1.1). Building on theoretical and practical results related to provenance, a new multi-disciplinary perspective of provenance is proposed, so that it can be developed on the Web (Section 1.2). A bibliography-based methodology, capable of identifying trends in the provenance research community, is outlined; the results of this analysis are used to structure a vision of provenance on the Web (Section 1.3).

1.1 Drivers for Provenance

As the e-science vision becomes reality [155, 156], researchers in the scientific community are increasingly perceived as providers of online data, which take the form of raw data sets from sensors and instruments, data products produced by workflow-based intensive computations [122], or databases resulting from sophisticated curation [39]. While science is becoming computation and data intensive, the fundamental tenet of the scientific method remains unchanged: experimental results need to be reproducible. In contrast to a workflow, which can be viewed as a recipe that can be applied in the future, *provenance* is regarded as the equivalent of a logbook, capturing all the steps that were involved in the actual derivation of a result, and which could be used to replay the execution that led to that result so as to validate it.

Curated databases typically represent the efforts of a dedicated group of people to produce a definitive description of some subject area [39]. They cover a vast range of application domains from Swiss-Prot,[1] which is a protein knowledgebase that is manually annotated and reviewed, to Wikipedia,[2] a crowd-sourced encyclopaedia, with increasingly sophisticated editorial processes. Such databases are generally published on the Web; and they are heavily cross-referenced with, and include data from, other databases. Curated databases therefore bring out the problem of attribution (who initially created a description), and raise questions about the source, or *provenance*, of such descriptions (where were descriptions initially published).

Meanwhile, the Web has evolved into a network of blogs, information portals and social bookmarking services which provide automated feeds between subscribers. As the Web allows information sharing, discovery, aggregation, filtering and flow in an unprecedented manner, it also becomes very difficult to identify, reliably, the original source that produced an information item on the Web. Without knowing the provenance of information, information services may not be able to undertake the necessary due diligence about their content, they may be the subject of fraud or spam, and overall they may be judged as unreliable.

[1] www.uniprot.org.
[2] www.wikipedia.org.

Provenance is in fact identified as one of the many salient factors that affect how users determine trust in content provided by Web information sources [121]. This view is echoed by Lynch [178], who argues that among the consequences of this shift to new highly distributed dissemination systems, will be a new emphasis on the provenance of data and metadata, and the need for information retrieval systems to permit users to factor in trust preferences about this information.

Where the Web originally consisted largely of documents intended to be read by humans, the Semantic Web [21] envisions a Web of information and knowledge processable by computer systems which undertake automated reasoning. Central to this effort are RDF [169] and OWL [90], the frameworks in which to express metadata, vocabularies and perform associated reasoning. This vision is being deployed by means of Linked Data [25, 243], an information space in which data is being enriched by typed links expressed in RDF [169], cross-referencing data sets, in a machine-processable fashion. Given the possibility for anybody (or system) to publish sets of Linked Data that refer to others, reasoners will need explicit representations of provenance information in order to make trust judgements about the information they use [274].

The issue of provenance is in no way limited to data, information or knowledge. It also applies to physical artefacts, for example in the food industry. From wine to meat, from dairy products to whisky, and from coffee to vegetables, the food industry is very keen to be able to demonstrate the origin of the ingredients consumers purchase and eat. Understanding the provenance of food, i.e., its origin, how it is produced, transported, and delivered to consumers, is turned into a competitive advantage by the food industry, since it allows it to demonstrate quality (in taste, in carbon footprint, or in ethics). Furthermore, across the world, governments and associated regulatory authorities are interested in food safety, and typically require the traceability of food. Likewise, manufacturers focus on compliance and traceability initiatives for a variety of reasons. Understanding past processes is critical to discover bottlenecks, inefficiencies, wastage, and learn how to improve them. Exact traceability is essential to manage product recalls efficiently and minimise their economic impact. Similar to the food

industry, provenance of products is used to build customer trust. And of course, in the context of art, the provenance of art objects is so important that available evidence is typically produced before auctions in order to maximise the price obtained for these objects.

1.2 Provenance for Web Science

Web science is the emerging interdisciplinary field that aims to understand the Web, engineer its future and ensure its social benefit [20]. In the context of Web science, trust is recognised as one of the important concerns associated with the Web [128]: there is a broad consensus that trust in content could be derived if the transformations and derivations that resulted in such content are known. Hence, given that the Web currently provides little support for provenance, the topic of provenance is becoming recognised as an important subject of investigation [243] in this context.

Like Web science, there is a multidisciplinary facet to provenance. First, within computer science, multiple sub-disciplines are involved including database, systems, eScience, grid, Semantic Web, and security. Second, provenance can be exploited to provide new services to the scientific community, businesses, and all Web users. It has the potential to make systems more transparent, and therefore auditable. As a result, it is a strong contender technology to underpin information accountability [281]. While it can be used to perform compliance checks (such as conformance to process or checking that terms of data licensing are met), it also raises issues related to privacy. Thus, societal, legal, and business perspectives on provenance could potentially have a wide impact on its use on the Web.

The aim of this monograph is to survey the technical aspects of provenance that are relevant to Web science, but also to draw attention to the potential multidisciplinary opportunities that they bring. Provenance, as a technical subject of study, is by no means a green field. The oldest publications discussed in this survey dates back to the late 1980s. Importantly, the interest of provenance has been growing dramatically, as illustrated by the number of publications on the topic (see Figure 2.1, page 9, to be discussed in the next chapter). Over

400 publications on provenance have been identified, 200 of which have been published over the last two years.

Several surveys already exist, but, to some extent, work has so far been broadly surveyed on a per discipline basis, as illustrated by Simmhan et al.'s review of provenance in e-Science [248], Bose and Frew's survey of provenance for scientific processing [30], Cheney et al.'s survey of database provenance [67], and Glavic and Dittrich's classification of approaches [125]. This monograph aims to break such silos, and tries to investigate cross-cutting concerns that are relevant to providing provenance of information on the Web.

In fact, society is now at a turning point since it is presented with a unique opportunity, which will require social and technical changes: it is the author's belief that *society can and should reliably track and exploit the provenance of information on the Web*. To achieve this vision, the research output from all disciplines investigating provenance should be integrated into a coherent approach, for which a foundational framework is proposed here. For instance, the work undertaken by the workflow community on provenance is very relevant to the flow of information that is becoming commonplace on the Web, since it can help track provenance as information flows through distributed services. Given that much of the data available on the Web is actually stored in databases, provenance research in the context of databases is essential, since it tracks provenance as data changes within databases. The work focusing on making provenance secure and non-forgeable is also relevant to the goal of reliably tracking information on the Web.

1.3 A Web Science View of Provenance

Having compiled the most extensive bibliography on provenance so far, developments in Web science provide tools and techniques to analyze this research topic. Simple metrics such as citation count can help identify the most popular papers. However, citation analysis can help gain a deeper insight into the different subfields of this subject of study. Using clustering techniques, emerging research fronts dealing with different concerns can be identified; using tag clouds [13], these concerns can be summarised in a visual manner.

This monograph is structured as follows. Section 2 undertakes an analysis of the provenance literature, discusses key topics of interest and identifies landmark papers. Section 3 discusses a broad definition of provenance that would apply to the Web and compares it with various alternative definitions that are recast in a Web context. Section 4 contrasts the work undertaken by the workflows and databases, which have traditionally studied provenance. By some measures, these approaches can be regarded as closed; hence, Section 5 introduces a vision of provenance for open systems such as the Web. Section 6 then discusses issues of provenance related to Web technologies. Finally, the problem of accountability is tackled in Section 7, before some concluding remarks in Section 8.

References

[1] P. Agrawal, O. Benjelloun, A. D. Sarma, C. Hayworth, S. Nabar, T. Sugihara, and J. Widom, "Trio: A system for data, uncertainty, and lineage," in *Proceedings of the 32nd International Conference on Very Large Data Bases*, pp. 1151–1154, Seoul, Korea, September 2006. (url: http://www.vldb.org/conf/2006/p1151-agrawal.pdf).

[2] R. Aldeco-Perez and L. Moreau, "Provenance-based auditing of private data use," in *BCS International Academic Research Conference, Visions of Computer Science*, September 2008. (url: http://eprints.ecs.soton.ac.uk/16580/).

[3] I. Altintas, O. Barney, and E. Jaeger-Frank, "Provenance collection support in the kepler scientific workflow system," in *Proceedings of the International Provenance and Annotation Workshop 2006 (IPAW'2006)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 118–132, Springer, 2006. (doi: http://dx.doi.org/10.1007/11890850_14).

[4] M. K. Anand, S. Bowers, T. M. McPhillips, and B. Ludaescher, "Efficient provenance storage over nested data collections," in *Proceedings of the 12th International Conference on Extending Database Technology (EDBT'09)*, (M. L. Kersten, B. Novikov, J. Teubner, V. Polutin, and S. Manegold, eds.), pp. 958–969, 2009. (doi: http://doi.acm.org/10.1145/1516360.1516470).

[5] M. K. Anand, S. Bowers, T. M. McPhillips, and B. Ludaescher, "Exploring scientific workflow provenance using hybrid queries over nested data and lineage graphs," in *Proceedings of the 21st International Conference on Scientific and Statistical Database Management (SSDBM'09)*, pp. 237–254, New Orleans, LA, USA, 2009. (doi: http://dx.doi.org/10.1007/978-3-642-02279-1_18).

[6] D. W. Archer, L. M. L. Delcambre, and D. Maier, "A framework for fine-grained data integration and curation, with provenance, in a dataspace," in

TAPP'09: First Workshop on Theory and Practice of Provenance, (J. Cheney, ed.), San Francisco, CA, February 2009. USENIX Association. (url: http://www.usenix.org/event/tapp09/tech/ full_papers/archer/archer.pdf).

[7] B. Balis, M. Bubak, M. Pelczar, and J. Wach, "Provenance querying for end-users: A drug resistance case study," in *ICCS '08: Proceedings of the 8th International Conference on Computational Science, Part III*, pp. 80–89, Berlin, Heidelberg, 2008. Springer-Verlag. (doi: http://dx.doi.org/10.1007/978-3-540-69389-5_11).

[8] B. Balis, M. Bubak, M. Pelczar, and J. Wach, "Provenance tracking and querying in the virolab virtual laboratory," in *CCGRID '08: Proceedings of the 2008 Eighth IEEE International Symposium on Cluster Computing and the Grid*, pp. 675–680, Washington, DC, USA, 2008. IEEE Computer Society. (doi: http://dx.doi.org/10.1109/CCGRID.2008.83).

[9] B. Balis, M. Bubak, and J. Wach, "Provenance tracking in the virolab virtual laboratory," in *Parallel Processing and Applied Mathematics, 7th International Conference, PPAM 2007, Gdansk, Poland, September 9–12, 2007, Revised Selected Papers*, pp. 381–390, 2007. (doi: http://dx.doi.org/10.1007/978-3-540-68111-3_40).

[10] B. Balis, M. Bubak, and J. Wach, "User-oriented querying over repositories of data and provenance," in *E-SCIENCE '07: Proceedings of the Third IEEE International Conference on e-Science and Grid Computing*, pp. 187–194, Washington, DC, USA, 2007. IEEE Computer Society. (doi: http://dx.doi.org/10.1109/E-SCIENCE.2007.81).

[11] Z. Bao, S. Cohen-Boulakia, S. B. Davidson, A. Eyal, and S. Khanna, "Differencing provenance in scientific workflows," in *IEEE 25th International Conference on Data Engineering (ICDE'09)*, pp. 808–819, IEEE Computer Society, 2009. (doi: http://doi.ieeecomputersociety.org/10.1109/ICDE.2009.103).

[12] R. S. Barga and L. A. Digiampietri, "Automatic capture and efficient storage of escience experiment provenance," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, 2008. (doi: http://dx.doi.org/10.1002/cpe.1235).

[13] S. Bateman, C. Gutwin, and M. Nacenta, "Seeing things in the clouds: the effect of visual features on tag cloud selections," in *HT '08: Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia*, pp. 193–202, New York, NY, USA, 2008. ACM. (doi: http://doi.acm.org/10.1145/1379092.1379130).

[14] L. Bavoil, S. P. Callahan, P. J. Crossno, J. Freire, C. E. Scheidegger, C. T. Silva, and H. T. Vo, "VisTrails: Enabling interactive multiple-view visualizations," in *Proceedings of IEEE Visualization*, p. 18, Los Alamitos, CA, USA, 2005. IEEE Computer Society. (doi: http://doi.ieeecomputersociety.org/10.1109/VIS.2005.113).

[15] R. A. Becker and J. M. Chambers, "Auditing of data analyses," in *SSDBM'86: Proceedings of the 3rd International Workshop on Statistical and Scientific Database Management*, (R. E. Cubitt, B. Cooper, and G. Ozsoyoglu, eds.), pp. 78–80, Berkeley, CA, US, 1986. Lawrence Berkeley Laboratory.

[16] R. A. Becker and J. M. J. M. Chambers, "Auditing of data analyses," *SIAM Journal of Scientific and Statistical Computing*, vol. 9, no. 4, pp. 747–760, 1988. (doi: http://dx.doi.org/10.1137/0909049).

[17] O. Benjelloun, A. D. Sarma, A. Halevy, M. Theobald, and J. Widom, "Databases with uncertainty and lineage," *The VLDB Journal*, vol. 17, no. 2, pp. 243–264, 2008. (doi: http://dx.doi.org/10.1007/s00778-007-0080-z).

[18] O. Benjelloun, A. D. Sarma, A. Halevy, and J. Widom, "ULDBs: Databases with uncertainty and lineage," in *VLDB '06: Proceedings of the 32nd International Conference on Very Large Data Bases*, pp. 953–964, VLDB Endowment, 2006. (url: http://ilpubs.stanford.edu:8090/703/).

[19] O. Benjelloun, A. D. Sarma, C. Hayworth, and J. Widom, "An introduction to ULDBs and the trio system," *IEEE Data Engineering Bulletin*, March 2006. (url: http://ilpubs.stanford.edu:8090/793/).

[20] T. Berners-Lee, W. Hall, J. A. Hendler, K. O'Hara, N. Shadbolt, and D. J. Weitzner, "A framework for web science," *Foundations and Trends in Web Science*, vol. 1, no. 1, pp. 1–130, 2006. (doi: http://dx.doi.org/10.1561/1800000001).

[21] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, vol. 284, no. 5, pp. 34–43, 2001. (url: http://www.scientificamerican.com/article.cfm?id=the-semantic-web).

[22] D. Bhagwat, L. Chiticariu, W.-C. Tan, and G. Vijayvargiya, "An annotation management system for relational databases," in *VLDB '04: Proceedings of the Thirtieth International Conference on Very Large Data Bases*, pp. 900–911, VLDB Endowment, 2004. (url: http://www.vldb.org/conf/2004/RS23P1.PDF).

[23] O. Biton, S. C. Boulakia, and S. B. Davidson, "Zoom*userviews: Querying relevant provenance in workflow systems," in *VLDB '07: Proceedings of the 33rd International Conference on Very Large Data Bases*, (C. Koch, J. Gehrke, M. N. Garofalakis, D. Srivastava, K. Aberer, A. Deshpande, D. Florescu, C. Y. Chan, V. Ganti, C.-C. Kanne, W. Klas, and E. J. Neuhold, eds.), pp. 1366–1369, ACM, 2007. (url: http://www.vldb.org/conf/2007/papers/demo/p1366-biton.pdf).

[24] O. Biton, S. Cohen-Boulakia, S. B. Davidson, and C. S. Hara, "Querying and managing provenance through user views in scientific workflows," in *International Conference Data Engineering (ICDE'08)*, pp. 1072–1081, Los Alamitos, CA, USA, 2008. IEEE Computer Society. (doi: http://doi.ieeecomputersociety.org/10.1109/ICDE.2008.4497516).

[25] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data — the story so far," *International Journal on Semantic Web and Information Systems (IJSWIS)*, 2009. (url: http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf).

[26] B. T. Blaustein, L. Seligman, M. Morse, M. D. Allen, and A. Rosenthal, "Plus: Synthesizing privacy, lineage, uncertainty and security," in *ICDE Workshops*, pp. 242–245, 2008. (doi: http://dx.doi.org/10.1109/ICDEW.2008.4498325).

[27] S. Boag, D. Chamberlin, M. F. Fernandez, D. Florescu, J. Robie, and J. Simeon, "Xquery 1.0: An xml query language," W3c recommendation,

World Wide Web Consortium, January 2007, (url: http://www.w3.org/TR/xuqery/).

[28] R. Bose, "A conceptual framework for composing and managing scientific data lineage," in *Proceedings of the 14th International Conference on Scientific and Statistical Database Management (SSDBM'02)*, pp. 15–19, Washington, DC, USA, July 2002. IEEE Computer Society. (doi: http://dx.doi.org/10.1109/SSDM.2002.1029701).

[29] R. Bose, I. Foster, and L. Moreau, "Report on the international provenance and annotation workshop (ipaw'06)," *Sigmod Records*, vol. 35, no. 3, pp. 51–53, September 2006. (doi: http://doi.acm.org/10.1145/1168092.1168102).

[30] R. Bose and J. Frew, "Lineage retrieval for scientific data processing: A survey," *ACM Computing Surveys*, vol. 37, no. 1, pp. 1–28, March 2005. (doi: http://doi.acm.org/10.1145/1057977.1057978).

[31] S. Bowers, T. McPhillips, B. Ludaescher, S. Cohen, and S. B. Davidson, "A model for user-oriented data provenance in pipelined scientific workflows," in *Proceedings of the International Provenance and Annotation Workshop 2006 (IPAW'2006)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 133–147, Springer, 2006. (doi: http://dx.doi.org/10.1007/11890850_15).

[32] S. Bowers, T. McPhillips, M. Wu, and B. Ludaescher, "Project histories: Managing data provenance across collection-oriented scientific workflow runs," in *Proceedings of the International Workshop on Data Integration in the Life Sciences (DILS)*, vol. 4544 of *Lecture Notes in Computer Science*, pp. 122–138, 2007. (doi: http://dx.doi.org/10.1007/978-3-540-73255-6_12).

[33] S. Bowers, T. M. McPhillips, and B. Ludaescher, "Provenance in collection-oriented scientific workflows," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, 2008. (doi: http://dx.doi.org/10.1002/cpe.1226).

[34] U. Braun, S. Garfinkel, D. A. Holland, K.-K. Muniswamy-Reddy, and M. I. Seltzer, "Issues in automatic provenance collection," in *Proceedings of the International Provenance and Annotation Workshop 2006 (IPAW'2006)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 171–183, Springer, 2006. (doi: http://dx.doi.org/10.1007/11890850_18).

[35] U. Braun, D. A. Holland, K.-K. Muniswamy-Reddy, and M. I. Seltzer, "Coping with cycles in provenance," Technical report, Harvard University, 2006, (url: http://www.eecs.harvard.edu/∼syrah/pubs/cycles.pdf).

[36] U. Braun, A. Shinnar, and M. Seltzer, "Securing provenance," in *HOTSEC'08: Proceedings of the 3rd Conference on Hot Topics in Security*, pp. 1–5, Berkeley, CA, USA, 2008. USENIX Association. (url: http://www.usenix.org/event/hotsec08/tech/full_papers/braun/braun.pdf).

[37] P. Buneman, A. Chapman, and J. Cheney, "Provenance management in curated databases," in *SIGMOD '06: Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data*, pp. 539–550, New York, NY, USA, 2006. ACM Press. (doi: http://doi.acm.org/10.1145/1142473.1142534).

[38] P. Buneman, A. Chapman, J. Cheney, and S. Vansummeren, "A provenance model for manually curated data," in *Proceedings of the International Provenance and Annotation Workshop 2006 (IPAW'2006)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 162–170, Springer, 2006. (doi: http://dx.doi.org/10.1007/11890850_17).

[39] P. Buneman, J. Cheney, W.-C. Tan, and S. Vansummeren, "Curated databases," in *PODS '08: Proceedings of the Twenty-Seventh ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pp. 1–12, New York, NY, USA, 2008. ACM. (doi: http://doi.acm.org/10.1145/1376916.1376918).

[40] P. Buneman, J. Cheney, and S. Vansummeren, "On the expressiveness of implicit provenance in query and update languages," in *11th International Conference on Database Theory (ICDT 2007)*, vol. 4353 of *Lecture Notes in Computer Science*, pp. 209–223, 2007. (doi: http://dx.doi.org/10.1007/11965893_15).

[41] P. Buneman, J. Cheney, and S. Vansummeren, "On the expressiveness of implicit provenance in query and update languages," *ACM Transactions on Database Systems*, vol. 33, no. 4, pp. 1–47, 2008. (doi: http://doi.acm.org/10.1145/1412331.1412340).

[42] P. Buneman, S. Khanna, K. Tajima, and W.-C. Tan, "Archiving scientific data," in *Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data*, pp. 1–12, ACM Press, 2002. (doi: http://doi.acm.org/10.1145/564691.564693).

[43] P. Buneman, S. Khanna, K. Tajima, and W.-C. Tan, "Archiving scientific data," *ACM Transactions on Database Systems*, vol. 29, pp. 2–42, 2004. (doi: http://doi.acm.org/10.1145/974750.974752).

[44] P. Buneman, S. Khanna, and W.-C. Tan, "Why and Where: A Characterization of Data Provenance," in *Proceedings of 8th International Conference on Database Theory (ICDT'01)*, vol. 1973 of *Lecture Notes in Computer Science*, pp. 316–330, London, UK, 2001. Springer. (doi: http://dx.doi.org/10.1007/3-540-44503-X_20).

[45] P. Buneman, S. Khanna, and W.-C. Tan, "Data provenance: Some basic issues," in *FST TCS 2000: Foundations of Software Technology and Theoretical Computer Science*, vol. 1974 of *Lecture Notes in Computer Science*, pp. 87–93, 2000. (doi: http://dx.doi.org/10.1007/3-540-44450-5).

[46] P. Buneman, D. Maier, and J. Widom, "Where was your data yesterday, and where will it go tomorrow?," in *Position Paper for NSF Workshop on Information and Data Management (IDM '00)*, Chicago IL, 2000. (url: http:// hermes.dpi.inpe.br:1910/col/dpi.inpe.br/banon/2004/04.21.11.45/doc/BunemanWhereTomorrow.pdf).

[47] P. Buneman and W.-C. Tan, "Provenance in databases," in *SIGMOD '07: Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, pp. 1171–1173, New York, NY, USA, 2007. ACM. (doi: http://doi.acm.org/10.1145/1247480.1247646).

[48] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging it platforms: Vision, hype, and reality for

delivering computing as the 5th utility," *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599–616, 2009. (doi: http://dx.doi.org/10.1016/j.future.2008.12.001).

[49] S. P. Callahan, J. Freire, C. E. Scheidegger, C. T. Silva, and H. T. Vo, "Towards provenance-enabling paraview," in *Second International Provenance and Annotation Workshop, IPAW'2008*, vol. 5272 of *Lecture Notes in Computer Science*, (J. Freire, D. Koop, and L. Moreau, eds.), pp. 120–127, Springer, June 2008. (doi: http://dx.doi.org/10.1007/978-3-540-89965-5_13).

[50] J. J. Carroll, C. Bizer, P. Hayes, and P. Stickler, "Named graphs, provenance and trust," in *WWW '05: Proceedings of the 14th International Conference on World Wide Web*, pp. 613–622, New York, NY, USA, 2005. ACM Press. (doi: http://doi.acm.org/10.1145/1060745.1060835).

[51] M. C. Cavalcanti, M. L. Campos, and M. Mattoso, "Managing scientific models in structural genomic projects updated, At [101]," October 2002, (url: http://people.cs.uchicago.edu/∼yongzh/papers/CCM.ps).

[52] A. Chapman, "Incorporating provenance in database systems," PhD thesis, University of Michigan, 2008, (url: http://hdl.handle.net/2027.42/61645).

[53] A. Chapman and H. V. Jagadish, "Issues in building practical provenance systems," *IEEE Data Engineering Bulletin*, vol. 30, no. 4, pp. 38–43, 2007. (url: http://sites.computer.org/debull/A07dec/chapman.pdf).

[54] A. P. Chapman, H. V. Jagadish, and P. Ramanan, "Efficient provenance storage," in *SIGMOD '08: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, pp. 993–1006, New York, NY, USA, 2008. ACM. (doi: http://doi.acm.org/10.1145/1376616.1376715).

[55] A. Chebotko, S. Chang, S. Lu, F. Fotouhi, and P. Yang, "Scientific workflow provenance querying with security views," in *Proceedings of the Ninth International Conference on Web-Age Information Management*, pp. 349–356, 2008. IEEE Computer Society. (doi: http://dx.doi.org/10.1109/WAIM.2008.41).

[56] A. Chebotko, X. Fei, C. Lin, S. Lu, and F. Fotouhi, "Storing and querying scientific workflow provenance metadata using an RDBMS," in *e-Science and Grid Computing, International Conference on*, pp. 611–618, Los Alamitos, CA, USA, 2007. IEEE Computer Society. (doi: http://doi.ieeecomputersociety.org/10.1109/E-SCIENCE.2007.70).

[57] C. Chen, "Citespace ii: Detecting and visualizing emerging trends and transient patterns in scientific literature," *Journal of American Society for Information Science & Technology*, vol. 57, no. 3, pp. 359–377, 2006. (doi: http://dx.doi.org/10.1002/asi.v57:3).

[58] C. Chen, Y. Chen, M. Horowitz, H. Hou, Z. Liu, and D. Pellegrino, "Towards an explanatory and computational theory of scientific discovery," *Journal of Informetrics*, 2009. (doi: http://dx.doi.org/10.1016/j.joi.2009.03.004).

[59] L. Chen and Z. Jiao, "Supporting provenance in service-oriented computing using the semantic web technologies," *IEEE Intelligent Informatics Bulletin*, vol. 7, no. 1, pp. 4–11, 2006. (url: http://www.comp.hkbu.edu.hk/∼cib/2006/Dec/iib_vol7no1_article1.pdf).

[60] L. Chen, Z. Jiao, and S. J. Cox, *On the Use of Semantic Annotations for Supporting Provenance in Grids*. Springer, 2006. (doi: http://dx.doi.org/10.1007/11823285).

[61] L. Chen, V. Tan, F. Xu, A. Biller, P. Groth, S. Miles, J. Ibbotson, M. Luck, and L. Moreau, "A proof of concept: Provenance in a service oriented architecture," in *Proceedings of the Fourth All Hands Meeting (AHM)*, September 2005. (url: http://www.allhands.org.uk/2005/proceedings/papers/503.pdf).

[62] L. Chen, X. Yang, and F. Tao, "A semantic web service based approach for augmented provenance," in *WI '06: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 594–600, Washington, DC, USA, 2006. IEEE Computer Society. (doi: http://dx.doi.org/10.1109/WI.2006.25).

[63] J. Cheney, "Program slicing and data provenance," *IEEE Data Engineering Bulletin*, vol. 30, no. 4, pp. 22–28, 2007. (url: http://sites.computer.org/debull/A07dec/cheney.pdf).

[64] J. Cheney, "Provenance, XML, and the scientific Web," in *Programming Language Techniques for XML (Plan-X'09)*, 2009. (url: http://db.ucsd.edu/planx2009/camera-ready/unpaginated/invited.pdf).

[65] J. Cheney, U. A. Acar, and A. Ahmed, "Provenance traces (extended report)," Technical Report http://arxiv.org/ abs/0812.0564v1, University of Edinburgh, December 2008, (url: http://homepages.inf.ed.ac.uk/jcheney/publications/drafts/provenance-traces-tr.pdf).

[66] J. Cheney, A. Ahmed, and U. A. Acar, "Provenance as dependency analysis," in *Proceedings of the 11th International Symposium on Database Programming Languages (DBPL 2007)*, number 4797 in *Lecture Notes in Computer Science*, (M. Arenas and M. I. Schwartzbach, eds.), pp. 139–153, 2007. (doi: 10.1007/978-3-540-75987-4_10).

[67] J. Cheney, L. Chiticariu, and W.-C. Tan, "Provenance in databases: Why, how, and where," *Foundations and Trends in Databases*, vol. 1, no. 4, pp. 379–474, 2009. (doi: http://dx.doi.org/10.1561/1500000006).

[68] K. Cheung and J. Hunter, "Provenance explorer — customized provenance views using semantic inferencing," in *5th International Semantic Web Conference (ISWC2006)*, vol. 4273 of *Lecture Notes in Computer Science*, Springer-Verlag, 2006. (doi: http://dx.doi.org/10.1007/11926078_16).

[69] L. Chiticariu and W.-C. Tan, "Debugging schema mappings with routes," in *VLDB '06: Proceedings of the 32nd International Conference on Very Large Data bases*, pp. 79–90, VLDB Endowment, 2006. (url: http://www.vldb.org/conf/2006/p79-chiticariu.pdf).

[70] L. Chiticariu, W.-C. Tan, and G. Vijayvargiya, "Dbnotes: A post-it system for relational databases based on provenance," in *SIGMOD '05: Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, pp. 942–944, New York, NY, USA, 2005. ACM. (doi: http://doi.acm.org/10.1145/1066157.1066296).

[71] S. Chong, "Towards semantics for provenance security," in *TAPP'09: First Workshop on Theory and Practice of Provenance*, (J. Cheney, ed.), San Francisco, CA, February 2009. USENIX Association. (url: http://www.usenix.org/event/tapp09/tech/full_papers/chong/chong.pdf).

[72] A. Chorley, P. Edwards, A. Preece, and J. Farrington, "Tools for tracing evidence in social science," in *Third International Conference on e-Social Science*,

October 2007. (url: http://users.cs.cf.ac.uk/A.D.Preece/publications/download/ess2007b.pdf).

[73] A. Cirillo, R. Jagadeesan, C. Pitcher, and J. Riely, "Tapido: Trust and authorization via provenance and integrity in distributed objects (extended abstract)," in *7th European Symposium on Programming (ESOP'08)*, vol. 4960 of *Lecture Notes in Computer Science*, (S. Drossopoulou, ed.), pp. 208–223, Springer, 2008. (doi: http://dx.doi.org/10.1007/978-3-540-78739-6_17).

[74] J. Clark and S. DeRose, "XML path language (xpath) version 1.0," W3c recommendation, World Wide Web Consortium, November 1999, (url: http://www.w3.org/TR/xpath/).

[75] B. Clifford, I. Foster, M. Hategan, T. Stef-Praun, M. Wilde, and Y. Zhao, "Tracking provenance in a virtual data grid," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 565–575, 2008. (doi: http://dx.doi.org/10.1002/cpe.1256).

[76] S. Cohen, S. Cohen-Bolakia, and S. B. Davidson, "Towards a model of provenance and user views in scientific workflows," in *Third International Workshop on Data Integration in the Life Sciences (DIL'06)*, vol. 4076 of *Lecture Notes in Computer Science*, pp. 264–279, Hinxton, UK, July 2006. Springer. (doi: http://dx.doi.org/10.1007/11799511).

[77] S. Cohen-Boulakia, O. Biton, S. Cohen, and S. Davidson, "Addressing the provenance challenge using zoom," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 497–506, 2008. (doi: http://dx.doi.org/10.1002/cpe.1232).

[78] Y. Cui, "Lineage tracing in data warehouses," PhD thesis, Stanford University, December 2001, (url: http://ilpubs.stanford.edu:8090/522/).

[79] Y. Cui and J. Widom, "Lineage tracing in a data warehousing system," in *Proceedings of the 16th International Conference on Data Engineering*, pp. 683–684, San Diego, California, 2000. Demonstration Description, (doi: http://dx.doi.org/10.1109/ICDE.2000.839493).

[80] Y. Cui and J. Widom, "Practical lineage tracing in data warehouses," in *Proceedings of the 16th International Conference on Data Engineering (ICDE'00)*, pp. 367–378, San Diego, California, February 2000. (doi: http://dx.doi.org/10.1109/ICDE.2000.839437).

[81] Y. Cui and J. Widom, "Storing auxiliary data for efficient maintenance and lineage tracing of complex views," in *Proceedings of the International Workshop on Design and Management of DataWarehouses (DMDW'00)*, Stockholm, Sweden, 2000. (url: http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-28/paper11.pdf).

[82] Y. Cui and J. Widom, "Lineage tracing for general data warehouse transformations," *The VLDB Journal*, vol. 12, no. 1, pp. 41–58, 2003. (doi: http://dx.doi.org/10.1007/s00778-002-0083-8).

[83] Y. Cui, J. Widom, and J. L. Wiener, "Tracing the lineage of view data in a warehousing environment," *ACM Transactions on Database System*, vol. 25, no. 2, pp. 179–227, 2000. (doi: http://doi.acm.org/10.1145/357775.357777).

[84] F. Curbera, Y. N. Doganata, A. Martens, N. Mukhi, and A. Slominski, "Business provenance — a technology to increase traceability of end-to-end

operations," in *On the Move to Meaningful Internet Systems: OTM'2008*, pp. 100–119, 2008. (doi: http://dx.doi.org/10.1007/978-3-540-88871-0_10).

[85] S. M. S. da Cruz, M. L. M. Campos, and M. Mattoso, "Towards a taxonomy of provenance in scientific workflow management systems," pp. 259–266, Los Alamitos, CA, USA, 2009. IEEE Computer Society. (doi: http://doi.ieeecomputersociety.org/10.1109/SERVICES-I.2009.18).

[86] P. P. da Silva, D. L. McGuinness, and R. McCool, "Knowledge provenance infrastructure," *IEEE Data Engineering Bulletin*, vol. 26, no. 4, pp. 26–32, December 2003. (url: http://www-ksl.stanford.edu/people/pp/papers/PinheirodaSilva_DEBULL_2003.pdf).

[87] C. Dai, D. Lin, E. Bertino, and M. Kantarcioglu, "An approach to evaluate data trustworthiness based on data provenance," in *SDM '08: Proceedings of the 5th VLDB Workshop on Secure Data Management*, pp. 82–98, Berlin, Heidelberg, 2008. Springer-Verlag. (doi: http://dx.doi.org/10.1007/978-3-540-85259-9_6).

[88] S. B. Davidson, S. C. Boulakia, A. Eyal, B. Ludaescher, T. M. McPhillips, S. Bowers, M. K. Anand, and J. Freire, "Provenance in scientific workflow systems," *IEEE Data Engineering Bulletin*, vol. 30, no. 4, pp. 44–50, 2007. (url: http://sites.computer.org/debull/A07dec/susan.pdf).

[89] S. B. Davidson and J. Freire, "Provenance and scientific workflows: Challenges and opportunities," in *SIGMOD Conference*, pp. 1345–1350, 2008. (doi: http://doi.acm.org/10.1145/1376616.1376772).

[90] M. Dean (ed), G. S. S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein, "OWL web ontology language reference," W3c recommendation, World Wide Web Consortium, February 2004, (url: http://www.w3.org/TR/owl-ref/).

[91] L. Ding, T. Finin, Y. Peng, P. P. da Silva, and D. L. McGuinness, "Tracking RDF graph provenance using RDF molecules," Technical report, UMBC, April 2005, (url: ftp://ksl.stanford.edu/pub/KSL_Reports/KSL-05-06.pdf).

[92] L. Ding, P. Kolari, T. Finin, A. Joshi, Y. Peng, and Y. Yesha, "On homeland security and the semantic web: A provenance and trust aware inference framework," in *Proceedings of the AAAI Spring Symposium on AI Technologies for Homeland Security*, AAAI Press, 2005. (url: http://ebiquity.umbc.edu/paper/html/id/209/On-Homeland-Security-and-the-Semantic-Web-A-Provenance-and-Trust-Aware-Inference-Framework).

[93] R. Dividino, S. Sizov, S. Staab, and B. Schueler, "Querying for provenance, trust, uncertainty and other meta knowledge in rdf," in *Web Semantics: Science, Services and Agents on the World Wide Web*, In Press, Corrected Proof:–, 2009. (doi: http://dx.doi.org/10.1016/j.websem.2009.07.004).

[94] R. Q. Dividino, S. Schenk, S. Sizov, and S. Staab, "Provenance, trust, explanations — and all that other meta knowledge," *KI*, vol. 23, no. 2, pp. 24–30, 2009. (url: http://www.uni-koblenz.de/~staab/Research/Publications/2009/provenance-ki-2009.pdf).

[95] L. M. Dusseault, Editor, "Http extensions for web distributed authoring and versioning (webdav)," Technical report, IETF, June 2007, (url: http://www.webdav.org/specs/rfc4918.html).

[96] M. Factor, E. Henis, D. Naor, S. Rabinovici-Cohen, P. Reshef, S. Ronen, G. Michetti, and M. Guercio, "Authenticity and provenance in long term digital preservation: Modeling and implementation in preservation aware storage," in *TAPP'09: First Workshop on Theory and Practice of Provenance*, (J. Cheney, ed.), San Francisco, CA, February 2009. USENIX Association. (url: http://www.usenix.org/event/tapp09/tech/full_papers/factor/factor.pdf).

[97] H. Fan, "Tracing data lineage using automed schema transformation pathways," in *Advances in Databases*, pp. 44–55, Springer-Verlag, 2002. (doi: http://dx.doi.org/10.1007/3-540-45495-0_6).

[98] H. Fan and A. Poulovassilis, "Tracing data lineage using schema transformation pathways," in *Knowledge Transformation for the Semantic Web. Frontiers in Artificial Intelligence and Applications*, (B. Omelayenko and M. Klein, eds.), pp. 64–79, IOS Press, 2003. (url: http://www.doc.ic.ac.uk/automed/publications/FP03a.ps).

[99] H. Fan and A. Poulovassilis, "Using schema transformation pathways for data lineage tracing," in *Database: Enterprise, Skills and Innovation*, vol. 3567, pp. 133–144, June 2005. (doi: http://dx.doi.org/10.1007/11511854_11).

[100] Y. Feng and W. Cai, "Provenance provisioning in mobile agent-based distributed job workflow execution," in *ICCS '07: Proceedings of the 7th International Conference on Computational Science, Part I*, vol. 4487 of *Lecture Notes in Computer Science*, pp. 398–405, Berlin, Heidelberg, 2007. Springer-Verlag. (doi: http://dx.doi.org/10.1007/978-3-540-72584-8_51).

[101] I. Foster and P. Buneman, "Workshop on data provenance and derivation," October 2002. (url: http://people.cs.uchicago.edu/∼yongzh/position_papers.html).

[102] I. Foster and C. Kesselman, eds., *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufman Publishers, 1998.

[103] I. Foster, C. Kesselman, and S. Tuecke, "The anatomy of the grid. Enabling scalable virtual organizations," *International Journal of Supercomputer Applications*, vol. 15, no. 3, pp. 200–222, 2001. (doi: http://dx.doi.org/10.1177/109434200101500302).

[104] I. Foster, J. Vockler, M. Wilde, and Y. Zhao, "The virtual data grid: A new model and architecture for data-intensive collaboration, At [101]," October 2002, (url: http://people.cs.uchicago.edu/∼yongzh/papers/CIDR.VDG.submitted.pdf).

[105] I. Foster, J.-S. Voeckler, M. Wilde, and Y. Zhao, "Chimera: A virtual data system for representing, querying and automating data derivation," in *Proceedings of the 14th Conference on Scientific and Statistical Database Management (SSDBM'02)*, pp. 37–46, Edinburgh, Scotland, July 2002. (doi: http://doi.ieeecomputersociety.org/10.1109/SSDM.2002.1029704).

[106] J. N. Foster and G. Karvounarakis, "Provenance and data synchronization," *IEEE Data Engineering Bulletin*, vol. 30, no. 4, pp. 13–21, 2007. (url: http://sites.computer.org/debull/A07dec/foster.pdf).

[107] M. S. Fox and J. Huang, "Knowledge provenance in enterprise information," *International Journal of Production Research*, vol. 43, no. 20, pp. 4471–4492, October 2005. (doi: http://dx.doi.org/10.1080/00207540500142415).

[108] J. Freire, C. T. Silva, S. P. Callahan, E. Santos, C. E. Scheidegger, and H. T. Vo, "Managing rapidly-evolving scientific workflows," in *Proceedings of the International Provenance and Annotation Workshop 2006 (IPAW'2006)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 10–18, Springer, 2006. (doi: http://dx.doi.org/10.1007/11890850_2).

[109] J. Frew and R. Bose, "Earth system science workbench: A data management infrastructure for earth science products," in *Proceedings of the 13th International Conference on Scientific and Statistical Database Management (SSDBM'01)*, pp. 180–189, Fairfax, VA, July 2001. (doi: http://dx.doi.org/10.1109/SSDM.2001.938550).

[110] J. Frew and R. Bose, "Lineage issues for scientific data and information, At [101]," October 2002, (url: http://people.cs.uchicago.edu/∼yongzh/papers/position-paper.html).

[111] J. Frew, D. Metzger, and P. Slaughter, "Automatic capture and reconstruction of computational provenance," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 485–496, 2008. (doi: http://dx.doi.org/10.1002/cpe.1247).

[112] J. Frew and P. Slaughter, "Es3: A demonstration of transparent provenance for scientific computation," in *Second International Provenance and Annotation Workshop, IPAW'2008*, vol. 5272 of *Lecture Notes in Computer Science*, (J. Freire, D. Koop, and L. Moreau, eds.), pp. 200–207, Springer, June 2008. (doi: http://dx.doi.org/10.1007/978-3-540-89965-5_21).

[113] J. Frey, D. De Roure, K. Taylor, J. Essex, H. Mills, and E. Zaluska, "Combechem: A case study in provenance and annotation using the semantic web," in *Proceedings of the International Provenance and Annotation Workshop 2006 (IPAW'2006)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 270–277, Springer, 2006. (doi: http://dx.doi.org/10.1007/11890850_27).

[114] J. Futrelle, "Harvesting RDF triples," in *Proceedings of the International Provenance and Annotation Workshop 2006 (IPAW'2006)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 64–72, Springer, 2006. (doi: http://dx.doi.org/10.1007/11890850_8).

[115] L. M. R. Gadelha and M. Mattoso, "Kairos: An architecture for securing authorship and temporal information of provenance data in grid enabled workflow management systems," in *e-Science and Grid Computing, International Conference on*, vol. 0, pp. 597–602, Los Alamitos, CA, USA, 2008. IEEE Computer Society. (doi: http://doi.ieeecomputersociety.org/10.1109/eScience.2008.161).

[116] A. Gehani, M. Kim, and J. Zhang, "Steps toward managing lineage metadata in grid clusters," in *TAPP'09: First Workshop on Theory and Practice of Provenance*, (J. Cheney, ed.), San Francisco, CA, February 2009. USENIX Association. (url: http://www.usenix.org/event/tapp09/tech/full_papers/gehani/gehani.pdf).

[117] A. Gehani and U. Lindqvist, "Bonsai: Balanced lineage authentication," in *Computer Security Applications Conference, 2007. ACSAC 2007. Twenty-Third Annual*, pp. 363–373, December 2007. (doi: http://dx.doi.org/10.1109/ACSAC.2007.45).

[118] A. Gehani and U. Lindqvist, "Veil: A system for certifying video provenance," in *ISM '07: Proceedings of the Ninth IEEE International Symposium on Multimedia*, pp. 263–272, Washington, DC, USA, 2007. IEEE Computer Society. (doi: http://dx.doi.org/10.1109/ISM.2007.10).

[119] A. Geraci, "IEEE Standard Computer Dictionary: Compilation of IEEE Standard Computer Glossaries," IEEE Press, Piscataway, NJ, USA, 1991.

[120] T. Gibson, K. Schuchardt, and E. Stephan, "Application of named graphs towards custom provenance views," in *TAPP'09: First Workshop on Theory and Practice of Provenance*, (J. Cheney, ed.), San Francisco, CA, February 2009. USENIX Association. (url: http://www.usenix.org/event/tapp09/tech/full_papers/gibson/gibson.pdf).

[121] Y. Gil and D. Artz, "Towards content trust of web resources," *Web Semantics*, vol. 5, no. 4, pp. 227–239, 2007. (doi: http://dx.doi.org/10.1016/j.websem.2007.09.005).

[122] Y. Gil, E. Deelman, M. Ellisman, T. Fahringer, G. Fox, D. Gannon, C. Goble, M. Livny, L. Moreau, and J. Myers, "Examining the challenges of scientific workflows," *IEEE Computer*, vol. 40, no. 12, pp. 26–34, December 2007. (doi: http://doi.ieeecomputersociety.org/10.1109/MC.2007.421).

[123] B. Glavic and G. Alonso, "Perm: Processing provenance and data on the same data model through query rewriting," in *IEEE 25th International Conference on Data Engineering (ICDE'09)*, pp. 174–185, IEEE Computer Society, 2009. (doi: http://doi.ieeecomputersociety.org/10.1109/ICDE.2009.15).

[124] B. Glavic and G. Alonso, "Provenance for nested subqueries," in *EDBT '09: Proceedings of the 12th International Conference on Extending Database Technology*, pp. 982–993, New York, NY, USA, 2009. ACM. (doi: http://doi.acm.org/10.1145/1516360.1516472).

[125] B. Glavic and K. R. Dittrich, "Data provenance: A categorization of existing approaches," in *Datenbanksysteme in Business, Technologie und Web (BTW'07)*, pp. 227–241, 2007. (url: http://www.ifi.uzh.ch/dbtg/fileadmin/storage/Glavic/publications/07_BTW_2007_long_version.pdf).

[126] C. Goble, "Position statement: Musings on provenance, workflow and (semantic web) annotations for bioinformatics, At [101]," October 2002, (url: http://people.cs.uchicago.edu/∼yongzh/papers/provenance_workshop_3.doc).

[127] J. Golbeck, "Combining provenance with trust in social networks for semantic web content filtering," in *Proceedings of the International Provenance and Annotation Workshop 2006 (IPAW'2006)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 101–108, Springer, 2006. (doi: http://dx.doi.org/10.1007/11890850_12).

[128] J. Golbeck, "Trust on the world wide web: A survey," *Foundations and Trends in Web Science*, vol. 1, no. 2, pp. 131–197, 2008. (doi: http://dx.doi.org/10.1561/1800000006).

[129] J. Golbeck and J. Hendler, "A semantic web approach to the provenance challenge," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 431–439, 2008. (doi: http://dx.doi.org/10.1002/cpe.1238).

[130] J. Golbeck and A. Mannes, "Using trust and provenance for content filtering on the semantic web," in *Proceedings of the WWW'06 Workshop on Models*

of Trust for the Web (MTW'06), 2006. (url: http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-190/paper02.pdf).

[131] J. M. Gomez-Perez and O. Corcho, "Problem-solving methods for understanding process executions," *Computing in Science and Engineering*, vol. 10, no. 3, pp. 47–52, 2008. (doi: http://doi.ieeecomputersociety.org/10.1109/MCSE.2008.78).

[132] D. Gotz and M. X. Zhou, "Characterizing users' visual analytic activity for insight provenance," in *Visual Analytics Science and Technology, 2008. VAST '08. IEEE Symposium on*, pp. 123–130, October 2008. (doi: http://dx.doi.org/10.1109/VAST.2008.4677365).

[133] T. J. Green, G. Karvounarakis, Z. G. Ives, and V. Tannen, "Update exchange with mappings and provenance," in *VLDB '07: Proceedings of the 33rd International Conference on Very Large Data Bases*, pp. 675–686, 2007. (url: http://repository.upenn.edu/cis_reports/763).

[134] T. J. Green, G. Karvounarakis, and V. Tannen, "Provenance semirings," in *PODS '07: Proceedings of the Twenty-Sixth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pp. 31–40, New York, NY, USA, 2007. ACM. (doi: http://doi.acm.org/10.1145/1265530.1265535).

[135] M. Greenwood, C. Goble, R. Stevens, J. Zhao, M. Addis, D. Marvin, L. Moreau, and T. Oinn, "Provenance of e-science experiments — experience from bioinformatics," in *Proceedings of the UK OST e-Science second All Hands Meeting 2003 (AHM'03)*, pp. 223–226, Nottingham, UK, September 2003. (url: http://www.ecs.soton.ac.uk/~lavm/papers/prov-ahm03.pdf).

[136] D. P. Groth and K. Streefkerk, "Provenance and annotation for visual exploration systems," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 6, pp. 1500–1510, 2006. (doi: http://doi.ieeecomputersociety.org/10.1109/TVCG.2006.101).

[137] P. Groth, E. Deelman, G. Juve, G. Mehta, and B. Berriman, "A pipeline-centric provenance model," in *The 4th Workshop on Workflows in Support of Large-Scale Science (WORKS'09)*, Portland, Oregon, 2009. (doi: http://doi.acm.org/10.1145/1645164.1645168).

[138] P. Groth, S. Jiang, S. Miles, S. Munroe, V. Tan, S. Tsasakou, and L. Moreau, "D3.1.1: An architecture for provenance systems," Technical Report http://eprints.ecs.soton.ac.uk/13216/, University of Southampton, November 2006, (url: http://eprints.ecs.soton.ac.uk/13216/).

[139] P. Groth, M. Luck, and L. Moreau, "A protocol for recording provenance in service-oriented grids," in *Proceedings of the 8th International Conference on Principles of Distributed Systems (OPODIS'04)*, vol. 3544 of *Lecture Notes in Computer Science*, pp. 124–139, Grenoble, France, December 2004. Springer-Verlag. (doi: http://dx.doi.org/10.1007/11516798_9).

[140] P. Groth, S. Miles, W. Fang, S. C. Wong, K.-P. Zauner, and L. Moreau, "Recording and using provenance in a protein compressibility experiment," in *Proceedings of the 14th IEEE International Symposium on High Performance Distributed Computing (HPDC'05)*, pp. 201–208, July 2005. (doi: http://dx.doi.org/10.1109/HPDC.2005.1520960).

[141] P. Groth, S. Miles, S. Modgil, N. Oren, M. Luck, and Y. Gil, "Determining the trustworthiness of new electronic contracts," in *Proceedings of the Tenth Annual Workshop on Engineering Societies in the Agents' World, (ESAW'09)*, Utrecht, The Netherlands, November 2009. (doi: http://dx.doi.org/10.1007/978-3-642-10203-5_12).

[142] P. Groth, S. Miles, and L. Moreau, "Preserv: Provenance recording for services," in *Proceedings of the UK OST e-Science Second All Hands Meeting 2005 (AHM'05)*, Nottingham, UK, September 2005. (url: http://www.ecs.soton.ac.uk/∼lavm/papers/Groth-AHM05.pdf).

[143] P. Groth, S. Miles, and L. Moreau, "A model of process documentation to determine provenance in mash-ups," *Transactions on Internet Technology (TOIT)*, vol. 9, no. 1, pp. 1–31, 2009. (doi: http://doi.acm.org/10.1145/1462159.1462162).

[144] P. Groth, S. Miles, and S. Munroe, "Principles of high quality documentation for provenance: A philosophical discussion," in *International Provenance and Annotation Workshop (IPAW'06)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), Springer, May 2006. (doi: http://dx.doi.org/10.1007/11890850_28).

[145] P. Groth and L. Moreau, "Recording process documentation for provenance," in *IEEE Transactions on Parallel and Distributed Systems*, publication, September 2009. (doi: http://doi.ieeecomputersociety.org/10.1109/TPDS.2008.215).

[146] P. T. Groth, "The origin of data: Enabling the determination of provenance in multi-institutional scientific systems through the documentation of processes," PhD thesis, Electronics and Computer Science, University of Southampton, 2007, (url: http://eprints.ecs.soton.ac.uk/14649/).

[147] C. Halaschek-Wiener, J. Golbeck, A. Schain, M. Grove, B. Parsia, and J. Hendler, "Annotation and provenance tracking in semantic web photo libraries," in *Proceedings of the International Provenance and Annotation Workshop 2006 (IPAW'2006)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 82–89, Springer, 2006. (doi: http://dx.doi.org/10.1007/11890850_10).

[148] A. Hameed, A. Preece, and D. Sleeman, "Ontology reconciliation," in *Handbook of Ontologies, International Handbooks on Information Systems,* Chapter 12, pp. 231–250, Springer Verlag, 2004. (url: eprints.aktors.org/332/01/p139.pdf).

[149] A. Harth, A. Polleres, and S. Decker, "Towards a social provenance model for the web," in *Workshop on Principles of Provenance (PrOPr)*, Edinburgh, Scotland, 2007. (url: http://sw.deri.org/2007/02/swsepaper/harth-propr.pdf).

[150] O. Hartig, "Provenance information in the web of data," in *Proceedings of the Linked Data on the Web Workshop (LDOW'09)*, Madrid, Spain, April 2009. (url: http://events.linkeddata.org/ldow2009/papers/ldow2009_paper18.pdf).

[151] O. Hartig and J. Zhao, "Using web data provenance for quality assessment," in *Proceedings of the 1st International Workshop on the Role of Semantic Web in Provenance Management (SWPM'09) at ISWC*, 2009. (url:

http://www.dbis.informatik.hu-berlin.de/fileadmin/research/papers/conferences/2009_swpm_hartig.pdf).

[152] R. Hasan, R. Sion, and M. Winslett, "Introducing secure provenance: Problems and challenges," in *StorageSS '07: Proceedings of the 2007 ACM Workshop on Storage Security and Survivability*, pp. 13–18, New York, NY, USA, 2007. ACM. (doi: http://doi.acm.org/10.1145/1314313.1314318).

[153] R. Hasan, R. Sion, and M. Winslett, "The case of the fake picasso: Preventing history forgery with secure provenance," in *Proceedings of 7th USENIX Conference on File and Storage Technologies, FAST 2009*, (M. I. Seltzer and R. Wheeler, eds.), pp. 1–14, San Francisco, CA, February 2009. (url: http://www.usenix.org/events/fast09/tech/full_papers/hasan/hasan.pdf).

[154] T. Heinis and G. Alonso, "Efficient lineage tracking for scientific workflows," in *SIGMOD '08: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, pp. 1007–1018, New York, NY, USA, 2008. ACM. (doi: http://doi.acm.org/10.1145/1376616.1376716).

[155] J. Hendler, "COMMUNICATION: Enhanced: Science and the semantic Web," *Science*, vol. 299, no. 5606, pp. 520–521, 2003. (doi: http://dx.doi.org/10.1126/science.1078874).

[156] A. J. G. Hey and A. E. Trefethen, "The data deluge: An e-science perspective," 2003, (url: http://eprints.ecs.soton.ac.uk/7648/1/The_Data_Deluge.pdf).

[157] D. A. Holland, U. Braun, D. Maclean, K.-K. Muniswamy-Reddy, and M. Seltzer, "Choosing a data model and query language for provenance," Technical report, Harvard University, 2008, (url: http://www.eecs.harvard.edu/∼kiran/pubs/ipaw08.pdf).

[158] D. A. Holland, M. Seltzer, U. Braun, and K.-K. Muniswamy-Reddy, "Passing the provenance challenge," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, 2008. (doi: http://dx.doi.org/10.1002/cpe.1227).

[159] J. Huang, T. Chen, A. Doan, and J. F. Naughton, "On the provenance of non-answers to queries over extracted data," *Proceedings of the VLDB Endowment*, vol. 1, no. 1, pp. 736–747, 2008. (doi: http://doi.acm.org/10.1145/1453856.1453936).

[160] J. Huang and M. S. Fox, "Trust judgment in knowledge provenance," in *Database and Expert Systems Applications, 2005. Proceedings of the Sixteenth International Workshop on*, pp. 524–528, August 2005. (doi: http://dx.doi.org/10.1109/DEXA.2005.193).

[161] J. Hunter and K. Cheung, "Provenance explorer — a graphical interface for constructing scientific publication packages from provenance trails," *International Journal of Digital Library*, vol. 7, no. 1, pp. 99–107, 2007. (doi: http://dx.doi.org/10.1007/s00799-007-0018-5).

[162] I. Jacobs and N. Walch, "Architecture of the world wide web, volume one," Technical report, World Wide Web Consortium, 2004, (url: http://www.w3.org/TR/webarch/).

[163] T. J. Jankun-Kelly, "Using visualization process graphs to improve visualization exploration," in *Second International Provenance and Annotation Workshop, IPAW'2008*, vol. 5272 of *Lecture Notes in Computer Science*, (J. Freire, D. Koop, and L. Moreau, eds.), pp. 78–91, Springer, June 2008. (doi: http://dx.doi.org/10.1007/978-3-540-89965-5_10).

[164] A. Kementsietsidis and M. Wang, "On the efficiency of provenance queries," in *Data Engineering, International Conference on*, vol. 0, pp. 1223–1226, Los Alamitos, CA, USA, 2009. IEEE Computer Society. (doi: http://doi. ieeecomputersociety.org/10.1109/ICDE.2009.206).

[165] T. Kifor, L. Varga, S. Alvarez, J. Vazquez-Salceda, and S. Willmott, "Privacy issues of provenance in electronic healthcare record systems," in *First International Workshop on Privacy and Security in Agent-Based Collaborative Environments (PSACE2006), AAMAS 2006*, 2006. (url: http://www.gridprovenance.org/publications/EHCR-Prov-Privacy.pdf).

[166] T. Kifor, L. Z. Varga, J. Vázquez-Salceda, S. Álvarez, S. Willmott, S. Miles, and L. Moreau, "Provenance in agent-mediated healthcare systems," *IEEE Intelligent Systems*, vol. 21, no. 6, pp. 38–46, Nov/Dec 2006. (doi: http://doi. ieeecomputersociety.org/10.1109/MIS.2006.119).

[167] J. Kim, E. Deelman, Y. Gil, G. Mehta, and V. Ratnakar, "Provenance trails in the wings/pegasus system," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 587–597, 2008. (doi: http://dx.doi.org/10.1002/cpe.1228).

[168] G. klyne. Contexts for RDF information modelling, (url: http://www.ninebynine.org/RDFNotes/RDFContexts.html).

[169] G. Klyne and J. J. Carroll, "Resource description framework (RDF): Concepts and abstract syntax," W3c recommendation, World Wide Web Consortium, February 2004, (url: http://www.w3.org/TR/rdf-concepts/).

[170] D. P. Lanter, "Design of a lineage-based meta-data base for GIS," *Cartography and Geographic Information Systems*, vol. 18, no. 4, pp. 255–261, 1991.

[171] A. E. Lawabni, C. Hong, D. H. C. Du, and A. H. Tewfik, "A novel update propagation module for the data provenance problem: A contemplating vision on realizing data provenance from models to storage," in *MSST '05: Proceedings of the 22nd IEEE/13th NASA Goddard Conference on Mass Storage Systems and Technologies*, pp. 61–69, Washington, DC, USA, 2005. IEEE Computer Society. (doi: http://dx.doi.org/10.1109/MSST.2005.2).

[172] J. Ledlie, C. Ng, D. A. Holland, K.-K. Muniswamy-Reddy, U. Braun, and M. Seltzer, "Provenance-aware sensor data storage," in *Data Engineering Workshops, 2005. 21st International Conference on*, April 2005. (doi: http://dx.doi.org/10.1109/ICDE.2005.270).

[173] B. N. Levine and M. Liberatore, "Dex: Digital evidence provenance supporting reproducibility and comparison," in *Proceedings of the Digital Forensic Research Workshop (DFRWS'09)*, 2009. (doi: http://dx.doi.org/10.1016/j.diin.2009.06.011).

[174] L. Lins, D. Koop, E. W. Anderson, S. P. Callahan, E. Santos, C. E. Scheidegger, J. Freire, and C. T. Silva, "Examining statistics of workflow evolution provenance: A first study," in *SSDBM '08: Proceedings of the 20th International Conference on Scientific and Statistical Database Management*, pp. 573–579, Berlin, Heidelberg, 2008. Springer-Verlag. (doi: http://dx.doi.org/10.1007/978-3-540-69497-7_40).

[175] P. Lord, P. Alper, C. Wroe, R. Stevens, C. Goble, J. Zhao, D. Hull, and M. Greenwood, "The semantic web: Service discovery and provenance

in my grid," 2004, (url: http://lists.w3.org/Archives/Public/public-swls-ws/2004Sep/att-0016/semantic_web_for_life_sciences_position.pdf).

[176] W. Lu and G. Miklau, "Auditing a database under retention restrictions," in *ICDE '09: Proceedings of the 2009 IEEE International Conference on Data Engineering*, pp. 42–53, Washington, DC, USA, 2009. IEEE Computer Society. (doi: http://dx.doi.org/10.1109/ICDE.2009.125).

[177] B. Ludaescher, N. Podhorszki, I. Altintas, S. Bowers, and T. M. McPhillips, "From computation models to models of provenance: The RWS approach," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 519–529, 2008. (doi: http://dx.doi.org/10.1002/cpe.1234).

[178] C. A. Lynch, "When documents deceive: Trust and provenance as new factors for information retrieval in a tangled web," *Journal of the American Society for Information Science and Technology*, vol. 52, no. 1, pp. 12–17, 2001. (doi: /10.1002/1532-2890(2000)52:1<12::AID-ASI1062>3.3.CO;2-M).

[179] S. P. Macleod, C. L. Kiernan, and Rajarajan and Vij (Issaquah, WA), "Data lineage data type," United States Patent 6434558, United States Patent, 2002, (url: http://www.freepatentsonline.com/6434558.html).

[180] D. W. Margo and M. Seltzer, "The case for browser provenance," in *TAPP'09: First Workshop on Theory and Practice of Provenance*, (J. Cheney, ed.), San Francisco, CA, February 2009. USENIX Association. (url: http://www.usenix.org/event/tapp09/tech/full_papers/margo/margo.pdf).

[181] A. Marins, M. A. Casanova, K. Breitman, and A. Furtado, "Modeling provenance for semantic desktop applications," in *Anais do XXVII Congresso da SBC (SBC'07)*, pp. 2100–2112, Rio de Janeiro, Brazil, July 2007. (url: http://cidoc.ics.forth.gr/docs/Modeling Provenance for Semantic Desktop Applications.pdf).

[182] A. Martin, "The ten page introduction to trusted computing," Technical Report RR-08-11, OUCL, December 2008, (url: http://web.comlab.ox.ac.uk/files/1873/RR-08-11.PDF).

[183] D. L. McGuinness and P. P. da Silva, "Infrastructure for web explanations," in *International Semantic Web Conference*, pp. 113–129, 2003. (doi: http://dx.doi.org/10.1007/b14287).

[184] D. L. McGuinness and P. P. da Silva, "Explaining answers from the semantic Web: The inference web approach," *Journal of Web Semantics*, vol. 1, no. 4, pp. 397–413, 2004. (doi: http://dx.doi.org/10.1016/j.websem.2004.06.002).

[185] A. Michlmayr, F. Rosenberg, P. Leitner, and S. Dustdar, "Service provenance in QoS-aware web service runtimes," in *Proceedings of the 7th IEEE International Conference on Web Services (ISWC'09)*, Los Angeles, CA, July 2009. (url: http://www.infosys.tuwien.ac.at/Staff/rosenberg/papers/icws2009.pdf).

[186] G. Miklau, B. N. Levine, and P. Stahlberg, "Securing history: Privacy and accountability in database systems," in *Third Biennial Conference on Innovative Data Systems Research (CIDR'07),* Asilomar, CA, USA, January 7–10, 2007, Online Proceedings, pp. 387–396, 2007. (url: http://www.cidrdb.org/cidr2007/papers/cidr07p44.pdf).

[187] S. Miles, "Agent-oriented data curation in bioinformatics," in *Proceedings of Workshop on Multi-agent Systems in Medicine, Computational*

*Biology, and Bioinformatics (MAS*BioMed'05)*, July 2005. (url: http://eprints.ecs.soton.ac.uk/10853/).

[188] S. Miles, "Electronically querying for the provenance of entities," in *Proceedings of the International Provenance and Annotation Workshop 2006 (IPAW'2006)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 184–192, Springer, 2006. (doi: http://dx.doi.org/10.1007/11890850_19).

[189] S. Miles, E. Deelman, P. Groth, K. Vahi, G. Mehta, and L. Moreau, "Connecting scientific data to scientific experiments with provenance," in *Proceedings of the Third IEEE International Conference on e-Science and Grid Computing (e-Science'07)*, pp. 179–186, Bangalore, India, December 2007. (doi: http://doi.ieeecomputersociety.org/10.1109/E-SCIENCE.2007.22).

[190] S. Miles, P. Groth, M. Branco, and L. Moreau, "The requirements of using provenance in e-science experiments," *Journal of Grid Computing*, vol. 5, no. 1, pp. 1–25, 2007. (doi: http://dx.doi.org/10.1007/s10723-006-9055-3).

[191] S. Miles, P. Groth, E. Deelman, K. Vahi, G. Mehta, and L. Moreau, "Provenance: The bridge between experiments and data," *Computing in Science and Engineering*, vol. 10, no. 3, pp. 38–46, May/June 2008. (doi: http://doi.ieeecomputersociety.org/10.1109/MCSE.2008.82).

[192] S. Miles, P. Groth, and M. Luck, "Handling mitigating circumstances for electronic contracts," in *Proceedings of the AISB 2008 Symposium on Behaviour Regulation in Multi-agent Systems*, pp. 37–42, Aberdeen, UK, April 2008. The Society for the Study of Artificial Intelligence and Simulation of Behaviour. (url: http://calcium.dcs.kcl.ac.uk/1283/).

[193] S. Miles, P. Groth, S. Munroe, S. Jiang, T. Assandri, and L. Moreau, "Extracting causal graphs from an open provenance data model," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 577–586, 2008. (doi: http://dx.doi.org/10.1002/cpe.1236).

[194] S. Miles, P. Groth, S. Munroe, M. Luck, and L. Moreau, "Agentprime: Adapting mas designs to build confidence," in *Agent-Oriented Software Engineering (AOSE'07)*, vol. 4951 of *Lecture Notes in Computer Science*, Springer, 2007. (doi: http://dx.doi.org/10.1007/978-3-540-79488-2_3).

[195] S. Miles, P. Groth, S. Munroe, and L. Moreau, "Prime: A methodology for developing provenance-aware applications," *ACM Transactions on Software Engineering and Methodology*, 2009.

[196] S. Miles, S. Munroe, M. Luck, and L. Moreau, "Modelling the provenance of data in autonomous systems," in *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'07)*, pp. 1–8, New York, NY, USA, 2007. ACM. (doi: http://doi.acm.org/10.1145/1329125.1329185).

[197] S. Miles, S. C. Wong, W. Fang, P. Groth, K.-P. Zauner, and L. Moreau, "Provenance-based validation of e-science experiments," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 5, no. 1, pp. 28–38, 2007. (doi: http://dx.doi.org/10.1016/j.websem.2006.11.003).

[198] A. Misra, M. Blount, A. Kementsietsidis, D. M. Sow, and M. Wang, "Advances and challenges for scalable provenance in stream processing systems," in *Second International Provenance and Annotation Workshop,*

IPAW'2008, vol. 5272 of *Lecture Notes in Computer Science*, (J. Freire, D. Koop, and L. Moreau, eds.), pp. 253–265, Springer, June 2008. (doi: http://dx.doi.org/10.1007/978-3-540-89965-5_26).

[199] L. Moreau, "Provenance architecture principles according to rest guidelines," Technical report, University of Southampton, 2009, In Preparation.

[200] L. Moreau and I. Foster, eds., *Provenance and Annotation of Data — International Provenance and Annotation Workshop, IPAW 2006*, vol. 4145 of *Lecture Notes in Computer Science*. Springer, May 2006. (doi: http://dx.doi.org/10.1007/11890850).

[201] L. Moreau, J. Freire, J. Futrelle, R. E. McGrath, J. Myers, and P. Paulson, "The open provenance model (v1.00)," Technical report, University of Southampton, December 2007, (url: http://eprints.ecs.soton.ac.uk/14979/).

[202] L. Moreau, P. Groth, S. Miles, J. Vazquez, J. Ibbotson, S. Jiang, S. Munroe, O. Rana, A. Schreiber, V. Tan, and L. Varga, "The provenance of electronic data," *Communications of the ACM*, vol. 51, no. 4, pp. 52–58, April 2008. (doi: http://doi.acm.org/10.1145/1330311.1330323).

[203] L. Moreau, N. Kwasnikowska, and J. V. den Bussche, "A formal account of the open provenance model," Technical report, University of Southampton, April 2009, (url: http://eprints.ecs.soton.ac.uk/17282/).

[204] L. Moreau and B. Ludaescher, *Special Issue on the First Provenance Challenge*, vol. 20. Wiley, April 2008. (doi: http://dx.doi.org/10.1002/cpe.1233).

[205] L. Moreau, B. Ludaescher, I. Altintas, R. S. Barga, S. Bowers, S. Callahan, G. Chin Jr, B. Clifford, S. Cohen, S. Cohen-Boulakia, S. Davidson, E. Deelman, L. Digiampietri, I. Foster, J. Freire, J. Frew, J. Futrelle, T. Gibson, Y. Gil, C. Goble, J. Golbeck, P. Groth, D. A. Holland, S. Jiang, J. Kim, D. Koop, A. Krenek, T. McPhillips, G. Mehta, S. Miles, D. Metzger, S. Munroe, J. Myers, B. Plale, N. Podhorszki, V. Ratnakar, E. Santos, C. Scheidegger, K. Schuchardt, M. Seltzer, Y. L. Simmhan, C. Silva, P. Slaughter, E. Stephan, R. Stevens, D. Turi, H. Vo, M. Wilde, J. Zhao, and Y. Zhao, "The first provenance challenge," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 409–418, April 2008. (doi: http://dx.doi.org/10.1002/cpe.1233).

[206] L. Moreau (Editor), B. Plale, S. Miles, C. Goble, P. Missier, R. Barga, Y. Simmhan, J. Futrelle, R. McGrath, J. Myers, P. Paulson, S. Bowers, B. Ludaescher, N. Kwasnikowska, J. V. den Bussche, T. Ellkvist, J. Freire, and P. Groth, "The open provenance model (v1.01)," Technical report, University of Southampton, July 2008, (url: http://eprints.ecs.soton.ac.uk/16148/1/opm-v1.01.pdf).

[207] P. Mouallem, R. Barreto, S. Klasky, N. Podhorszki, and M. Vouk, "Tracking files in the kepler provenance framework," in *Proceedings of 21st International Conference on Scientific and Statistical Database Management (SSDBM'09)*, pp. 273–282, New Orleans, LA, USA, 2009. (doi: http://dx.doi.org/10.1007/978-3-642-02279-1_21).

[208] K. Muniswamy-Reddy, D. Holland, U. Braun, and M. Seltzer, "Provenance-aware storage systems," in *ATEC '06: Proceedings of the Annual Conference on USENIX '06 Annual Technical Conference*, pp. 43–56, Berkeley,

CA, USA, June 2006. USENIX Association. (url: http://www.usenix.org/events/usenix06/tech/muniswamy-reddy.html).

[209] S. Munroe, S. Miles, L. Moreau, and J. Vázquez-Salceda, "PrIMe: A software engineering methodology for developing provenance-aware applications," in *ACM Digital Proceedings of the Software Engineering and Middleware Workshop (SEM'06)*, pp. 39–46, New York, NY, USA, 2006. ACM. (doi: http://doi.acm.org/10.1145/1210525.1210535).

[210] M. Mutsuzaki, M. Theobald, A. de Keijzer, J. Widom, P. Agrawal, O. Benjelloun, A. D. Sarma, R. Murthy, and T. Sugihara, "Trio-one: Layering uncertainty and lineage on a conventional DBMS," in *Proceedings of CIDR Conference (system demonstration)*, 2007. (url: http://ilpubs.stanford.edu:8090/805/).

[211] J. Myers, "Design constraints for scientific annotation systems, At [101]," October 2002, (url: http://people.cs.uchicago.edu/∼yongzh/papers/SAM1.Provenance.position.doc).

[212] J. Myers, C. Pancerella, C. Lansing, K. Schuchardt, and B. Didier, "Multi-scale science, supporting emerging practice with semantically derived provenance," in *Proceedings of the ISWC 2003 Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data*, Sanibel Island, Florida, October 2003. (url: http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-83/prov_1.pdf).

[213] J. D. Myers, A. R. Chappell, M. Elder, A. Geist, and J. Schwidder, "Re-integrating the research record," *IEEE Computing in Science and Engineering*, vol. 5, no. 3, pp. 44–50, 2003. (doi: http://dx.doi.org/10.1109/MCISE.2003.1196306).

[214] J. D. Myers, J. Futrelle, J. Gaynor, J. Plutchak, P. Bajcsy, J. Kastner, K. Kotwani, J. S. Lee, L. Marini, R. Kooper, R. E. McGrath, T. McLaren, A. Rodrguez, and Y. Liu, "Embedding data within knowledge spaces," in *UK e-sicence 2008*, 2008. (url: http://arxiv.org/abs/0902.0744).

[215] M. Nagappan and M. A. Vouk, "A model for sharing of confidential provenance information in a query based system," in *Second International Provenance and Annotation Workshop, IPAW'2008*, vol. 5272 of *Lecture Notes in Computer Science*, (J. Freire, D. Koop, and L. Moreau, eds.), pp. 62–69, Springer, June 2008. (doi: http://dx.doi.org/10.1007/978-3-540-89965-5_8).

[216] J. Ockerbloom, "Copyright and provenance: Some practical problems," *IEEE Data Engineering Bulletin*, vol. 30, no. 4, pp. 51–58, 2007. (url: http://sites.computer.org/debull/A07dec/ockerbloom.pdf).

[217] C. Pancerella, J. Myers, and L. Rahn, "Data provenance in the CMCS, At [101]," October 2002, (url: http://people.cs.uchicago.edu/∼yongzh/papers/ProvenanceWorkshopCMCS.pdf).

[218] M. P. Papazoglou and W.-J. van den Heuvel, "Service oriented architectures: Approaches, technologies and research issues," *VLDB Journal*, vol. 16, no. 3, pp. 389–415, 2007. (url: http://dx.doi.org/10.1007/s00778-007-0044-3).

[219] K. E. Pavlou and R. T. Snodgrass, "Forensic analysis of database tampering," *ACM Transactions on Database Systems*, vol. 33, no. 4, pp. 1–47, 2008. (doi: http://doi.acm.org/10.1145/1412331.1412342).

[220] P. Pediaditis, G. Flouris, I. Fundulaki, and V. Christophides, "On explicit provenance management in RDF/s graphs," in *TAPP'09: First Workshop on Theory and Practice of Provenance*, (J. Cheney, ed.), San Francisco, CA, February 2009. USENIX Association. (url: http://www.usenix.org/event/tapp09/tech/full_papers/pediaditis/pediaditis.pdf).

[221] L. Philip, A. Chorley, J. Farrington, and P. Edwards, "Data provenance, evidence-based policy assessment, and e-social science," in *Third International Conference on e-Social Science*, October 2007. (url: http://www.scientificcommons.org/40739576).

[222] A. Powell, M. Nilsson, A. Naeve, P. Johnston, and T. Baker, "DCMI abstract model," DCMI recommendation, Dublin Core Metadata Initiative, June 2007, (url: http://dublincore.org/documents/abstract-model/).

[223] N. Prat and S. Madnick, "Measuring data believability: A provenance approach," in *Proceedings of the 41st Hawaii International Conference on System Sciences — 2008*, IEEE Computer Society, 2008. (doi: http://doi.ieeecomputersociety.org/10.1109/HICSS.2008.243).

[224] E. Prud'hommeaux and A. Seaborne, "Sparql query language for RDF," W3c recommendation, World Wide Web Consortium, 2008, (url: http://www.w3.org/TR/rdf-sparql-query/).

[225] S. Rajbhandari, A. Contes, O. F. Rana, V. Deora, and I. Wootten, "Establishing workflow trust using provenance information," in *1st IEEE International Workshop on Modelling Autonomic Communications Environments (MACE 2006)*, October 2006. (url: http://www.gridprovenance.org/publications/manweek-ranaetal.pdf).

[226] S. Rajbhandari, A. Contes, O. F. Rana, V. Deora, and I. Wootten, "Trust assessment using provenance in service oriented applications," in *EDOCW '06: Proceedings of the 10th IEEE on International Enterprise Distributed Object Computing Conference Workshops*, p. 65, Washington, DC, USA, 2006. IEEE Computer Society. (doi: http://dx.doi.org/10.1109/EDOCW.2006.70).

[227] S. Rajbhandari, O. F. Rana, and I. Wootten, "A fuzzy model for calculating workflow trust using provenance data," in *MG '08: Proceedings of the 15th ACM Mardi Gras Conference*, pp. 1–8, New York, NY, USA, 2008. ACM. (doi: http://doi.acm.org/10.1145/1341811.1341823).

[228] S. Rajbhandari, I. Wootten, A. S. Ali, and O. F. Rana, "Evaluating provenance-based trust for scientific workflows," in *CCGRID '06: Proceedings of the Sixth IEEE International Symposium on Cluster Computing and the Grid*, pp. 365–372, Washington, DC, USA, 2006. IEEE Computer Society. (doi: http://dx.doi.org/10.1109/CCGRID.2006.43).

[229] S. Ram, J. Liu, N. Merchant, T. Yuhas, and P. Jansma, "Toward developing a provenance ontology for biological images," in *Eighth Annual Bio-Ontologies Workshop*, 2005. (url: http://kartik.eller.arizona.edu/Abstract4_29.doc).

[230] S. D. Ramchurn, D. Huynh, and N. R. Jennings, "Trust in multi-agent systems," *Knowledge Engineering Review*, vol. 19, no. 1, pp. 1–25, 2004. (doi: http://dx.doi.org/10.1017/S0269888904000116).

[231] C. Ré and D. Suciu, "Approximate lineage for probabilistic databases," *Proceedings VLDB Endowment*, vol. 1, no. 1, pp. 797–808, 2008. (doi: http://doi.acm.org/10.1145/1453856.1453943).

[232] A. Reggiori, D.-W. van Gulik, and Z. Bjelogrlic, "Indexing and retrieving semantic web resources: The RDF store model," in *SWAD-Europe Workshop on Semantic Web Storage and Retrieval*, Amsterdam, Netherlands, November 2003. (url: http://www.w3.org/2001/sw/Europe/events/20031113-storage/positions/asemantics.html).

[233] N. D. Rio and P. P. da Silva, "Probe-it! visualization support for provenance," in *Advances in Visual Computing, Third International Symposium, ISVC 2007, Lake Tahoe, NV, USA, November 26–28, 2007, Proceedings, Part II*, vol. 4842 of *Lecture Notes in Computer Science*, pp. 732–741, Springer, 2007. (doi: http://dx.doi.org/10.1007/978-3-540-76856-2_72).

[234] N. D. Rio, P. P. da Silva, A. Q. Gates, and L. Salayandia, "Semantic annotation of maps through knowledge provenance," in *GeoSpatial Semantics, Second International Conference, GeoS 2007, Mexico City, Mexico, November 29–30, 2007, Proceedings*, vol. 4853 of *Lecture Notes in Computer Science*, pp. 20–35, 2007. (doi: http://dx.doi.org/10.1007/978-3-540-76876-0_2).

[235] A. Rosenthal, L. Seligman, A. Chapman, and B. Blaustein, "Scalable access controls for lineage," in *TAPP'09: First Workshop on Theory and Practice of Provenance*, (J. Cheney, ed.), San Francisco, CA, February 2009. USENIX Association. (url: http://www.usenix.org/event/tapp09/tech/full_papers/rosenthal/rosenthal.pdf).

[236] S. Russell, Quads, (url: http://robustai.net/sailor/grammar/Quads.html).

[237] P. Ruth, D. Xu, B. K. Bhargava, and F. Regnier, "E-notebook middleware for accountability and reputation based trust in distributed data sharing communities," in *Proceedings 2nd International Conference on Trust Management (iTrust'04)*, vol. 2995 of *Lecture Notes in Computer Science*, pp. 161–175, Springer, 2004. (doi: http://dx.doi.org/10.1007/b96545).

[238] S. S. Sahoo, R. S. Barga, J. Goldstein, and A. P. Sheth, "Provenance algebra and materialized view-based provenance management," Technical Report 76523/tr-2008-170, Microsoft Research, 2008, (url: http://research.microsoft.com/pubs/76523/tr-2008-170.pdf).

[239] S. S. Sahoo, A. Sheth, and C. Henson, "Semantic provenance for escience: Managing the deluge of scientific data," *Internet Computing, IEEE*, vol. 12, no. 4, pp. 46–54, July–August 2008. (doi: http://dx.doi.org/10.1109/MIC.2008.86).

[240] A. D. Sarma, M. Theobald, and J. Widom, "Exploiting lineage for confidence computation in uncertain and probabilistic databases," Technical Report 2007-15, Stanford InfoLab, March 2007, (url: http://ilpubs.stanford.edu:8090/800/).

[241] C. Scheidegger, D. Koop, E. Santos, H. Vo, S. Callahan, J. Freire, and C. Silva, "Tackling the provenance challenge one layer at a time," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 473–483, 2008. (doi: http://dx.doi.org/10.1002/cpe.1237).

[242] C. Scheidegger, D. Koop, H. Vo, J. Freire, and C. Silva, "Querying and creating visualizations by analogy," *IEEE Transactions on Visualization and Computer Graphics*, 2007. (doi: http://doi.ieeecomputersociety.org/10.1109/TVCG.2007.70584).

[243] N. Shadbolt, T. Berners-Lee, and W. Hall, "The semantic web revisited," *IEEE Intelligent Systems*, vol. 21, no. 3, pp. 96–101, 2006. (doi: http://doi.ieeecomputersociety.org/10.1109/MIS.2006.62).

[244] S. Shah, C. A. N. Soules, G. R. Ganger, and B. D. Noble, "Using provenance to aid in personal file search," in *ATC'07: 2007 USENIX Annual Technical Conference on Proceedings of the USENIX Annual Technical Conference*, pp. 1–14, Berkeley, CA, USA, 2007. USENIX Association. (url: http://www.pdl.cmu.edu/PDL-FTP/ABN/usenix07.pdf).

[245] C. A. Silles and A. R. Runnalls, "Provenance tracking in cxxr," in *The R User Conference 2009*, Agrocampus-Ouest, Rennes, France, July 2009. (url: http://www.agrocampus-ouest.fr/math/useR-2009/abstracts/pdf/Silles_Runnalls.pdf).

[246] C. Silva, J. Freire, and S. P. Callahan, "Provenance for visualizations: Reproducibility and beyond," *Computing in Science and Engineering*, vol. 9, no. 5, pp. 82–89, 2007. (doi: http://doi.ieeecomputersociety.org/10.1109/MCSE.2007.106).

[247] Y. Simmhan, "Provenance Framework in Support of Data Quality Estimation," PhD thesis, University of Indiana, 2007, (url: http://gradworks.umi.com/32/97/3297094.html).

[248] Y. L. Simmhan, B. Plale, and D. Gannon, "A survey of data provenance in e-science," *SIGMOD Record*, vol. 34, no. 3, pp. 31–36, 2005. (doi: http://doi.acm.org/10.1145/1084805.1084812).

[249] Y. L. Simmhan, B. Plale, and D. Gannon, "A framework for collecting provenance in data-centric scientific workflows," in *International Conference on Web Service (ICWS'06)*, pp. 427–436, Washington, DC, USA, 2006. IEEE Computer Society. (doi: http://dx.doi.org/10.1109/ICWS.2006.5).

[250] Y. L. Simmhan, B. Plale, and D. Gannon, "Querying capabilities of the Karma provenance framework," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 441–451, 2008. (doi: http://dx.doi.org/10.1002/cpe.1229).

[251] Y. L. Simmhan, B. Plale, D. Gannon, and S. Marru, "Performance evaluation of the karma provenance framework for scientific workflows," in *Proceedings of the International Provenance and Annotation Workshop 2006 (IPAW'2006)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 222–236, Springer, 2006. (doi: http://dx.doi.org/10.1007/11890850_23).

[252] I. Souilah, A. Francalanza, and V. Sassone, "A formal model of provenance in distributed systems," in *TAPP'09: First Workshop on Theory and Practice of Provenance*, (J. Cheney, ed.), San Francisco, CA, February 2009. USENIX Association. (url: http://www.usenix.org/event/tapp09/tech/full_papers/souilah/souilah.pdf).

[253] L. Spery, C. Claramunt, and T. Libourel, "A spatio-temporal model for the manipulation of lineage metadata," *Geoinformatica*, vol. 5, no. 1, pp. 51–70, 2001. (doi: http://dx.doi.org/10.1023/A:1011459921552).

[254] P. Stahlberg, G. Miklau, and B. N. Levine, "Threats to privacy in the forensic analysis of database systems," in *SIGMOD '07: Proceedings of the*

2007 ACM SIGMOD International Conference on Management of Data, pp. 91–102, New York, NY, USA, 2007. ACM. (doi: http://doi.acm.org/10.1145/1247480.1247492).

[255] R. Stevens, J. Zhao, and C. Goble, "Using provenance to manage knowledge of in silico experiments," *Briefing in Bioinformatics*, vol. 8, no. 3, pp. 183–194, 2007. (doi: http://dx.doi.org/10.1093/bib/bbm015).

[256] P. Sun, Z. Liu, S. B. Davidson, and Y. Chen, "Detecting and resolving unsound workflow views for correct provenance analysis," in *SIGMOD '09: Proceedings of the 35th SIGMOD International Conference on Management of Data*, pp. 549–562, New York, NY, USA, 2009. ACM. (doi: http://doi.acm.org/10.1145/1559845.1559903).

[257] A. Syalim, Y. Hori, and K. Sakurai, "Grouping provenance information to improve efficiency of access control," in *Third International Conference and Workshops on Advances in Information Security and Assurance (ISA'09)*, vol. 5576 of *Lecture Notes in Computer Science*, pp. 51–59, 2009. (doi: http://dx.doi.org/10.1007/978-3-642-02617-1_6).

[258] M. Szomszor and L. Moreau, "Recording and reasoning over data provenance in web and grid services," in *International Conference on Ontologies, Databases and Applications of SEmantics (ODBASE'03)*, vol. 2888 of *Lecture Notes in Computer Science*, pp. 603–620, Catania, Sicily, Italy, November 2003. (doi: http://dx.doi.org/10.1007/b94348).

[259] V. Tan, P. Groth, S. Miles, S. Jiang, S. Munroe, S. Tsasakou, and L. Moreau, "Security issues in a SOA-based provenance system," in *Proceedings of the International Provenance and Annotation Workshop (IPAW'06)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 203–211, Chicago, Illinois, May 2006. Springer-Verlag. (doi: http://dx.doi.org/10.1007/11890850_21).

[260] W. C. Tan, "Research problems in data provenance," *IEEE Data Engineering Bulletin*, vol. 27, no. 4, pp. 45–52, 2004. (url: http://db.cs.ucsc.edu/node/216).

[261] W.-C. Tan, "Provenance in databases: Past, current, and future," *Bulletin of the Technical Committee on Data Engineering*, vol. 30, no. 4, pp. 3–12, December 2007. (url: ftp://ftp.research.microsoft.com/pub/debull/A07dec/wang-chiew.pdf).

[262] W. T. L. Teacy, J. Patel, N. R. Jennings, and M. Luck, "Coping with inaccurate reputation sources: Experimental analysis of a probabilistic trust model," in *AAMAS '05: Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 997–1004, New York, NY, USA, 2005. ACM. (doi: http://doi.acm.org/10.1145/1082473.1082624).

[263] The R Foundation for Statistical Computing, "R: Regulatory compliance and validation issues a guidance document for the use of R in regulated clinical trial environments," Technical report, Wirtschaftsuniversiat Wien, 2008, (url: http://www.r-project.org/doc/R-FDA.pdf).

[264] P. Townend, P. Groth, N. Looker, and J. Xu, "Ft-grid: A fault-tolerance system for e-science," in *Proceedings of the UK OST e-Science Fourth All Hands Meeting (AHM05)*, September 2005. (url: http://www.allhands.org.uk/2005/proceedings/papers/392.pdf).

[265] P. Townend, P. Groth, and J. Xu, "A provenance-aware weighted fault tolerance scheme for service-based applications," in *Proceedings of the 8th IEEE International Symposium on Object-Oriented Real-Time Distributed Computing (ISORC 2005)*, pp. 258–266, Los Alamitos, CA, USA, May 2005. IEEE Computer Society. (doi: http://doi.ieeecomputersociety.org/10.1109/ISORC.2005.3).

[266] W. Tsai, X. Wei, Y. Chen, R. Paul, J.-Y. Chung, and D. Zhang, "Data provenance in SOA: Security, reliability, and integrity," *Service Oriented Computing and Applications*, vol. 1, no. 4, pp. 223–247, December 2007. (doi: http://dx.doi.org/10.1007/s11761-007-0018-8).

[267] W.-T. Tsai, X. Wei, D. Zhang, R. Paul, Y. Chen, and J.-Y. Chung, "A new SOA data-provenance framework," in *Eighth International Symposium on Autonomous Decentralized Systems. (ISADS'07)*, pp. 105–112, March 2007. (doi: http://doi.ieeecomputersociety.org/10.1109/ISADS.2007.5).

[268] S. Vansummeren and J. Cheney, "Recording provenance for SQL queries and updates," *IEEE Data Engineering Bulletin*, vol. 30, no. 4, pp. 29–37, 2007. (url: http://sites.computer.org/debull/A07dec/stijn.pdf).

[269] J. A. Vaughan, L. Jia, K. Mazurak, and S. Zdancewic, "Evidence-based audit," in *CSF '08: Proceedings of the 2008 21st IEEE Computer Security Foundations Symposium*, pp. 177–191, Washington, DC, USA, 2008. IEEE Computer Society. (doi: http://dx.doi.org/10.1109/CSF.2008.24).

[270] J. Vázquez-Salceda, S. Alvarez, T. Kifor, L. Z. Varga, S. Miles, L. Moreau, and S. Willmott, "Agent Technology and E-Health," in *chapter EU PROVENANCE Project: An Open Provenance Architecture for Distributed Applications*, (R. Annicchiarico, U. Cortés, and C. Urdiales, eds.), pp. 45–63, Whitestein Series in Software Agent Technologies and Autonomic Computing. Birkhauser Verlag AG, Switzerland, December 2007. (doi: http://dx.doi.org/10.1007/978-3-7643-8547-7_4).

[271] J. Vazquez-Salceda and S. Alvarez-Napagao, "Using SOA provenance to implement norm enforcement in e-institutions," in *Proceedings of the Workshop on Coordination, Organizations, Institutions and Norms (COIN@AAAI08)*, pp. 188–203, Berlin, Heidelberg, 2009. Springer-Verlag. (doi: http://dx.doi.org/10.1007/978-3-642-00443-8_13).

[272] N. Vijayakumar, "Data management in distributed stream processing systems," PhD thesis, University of Indiana, 2007, (url: http://proquest.umi.com/pqdlink?did=1407508891&Fmt=7&clientId=79356&RQT=309&VName=PQD).

[273] N. N. Vijayakumar and B. Plale, "Towards low overhead provenance tracking in near real-time stream filtering," in *Proceedings of the International Provenance and Annotation Workshop 2006 (IPAW'2006)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 46–54, Springer, 2006. (doi: http://dx.doi.org/10.1007/11890850_6).

[274] W3C Incubator Activity, "Provenance incubator group charter," September 2009, (url: http://www.w3.org/2005/Incubator/prov/charter).

[275] L. Wang, S. Lu, X. Fei, A. Chebotko, H. V. Bryant, and J. L. Ram, "Atomicity and provenance support for pipelined scientific workflows," *Future*

Generation Computer Systems, vol. 25, no. 5, pp. 568–576, 2009. (doi: http://dx.doi.org/10.1016/j.future.2008.06.007).

[276] L. Wang, S. Lu, X. Fei, and J. Ram, "A dataflow-oriented atomicity and provenance system for pipelined scientific workflows," in *Proceedings of the 2nd International Workshop on Workflow Systems in e-Science (WSES 07), in conjunction with International Conference on Computational Science (ICCS) 2007*, vol. 4489 of *Lecture Notes in Computer Science*, Springer, 2007. (doi: http://dx.doi.org/10.1007/978-3-540-72588-6_42).

[277] M. Wang, M. Blount, J. Davis, A. Misra, and D. Sow, "A time-and-value centric provenance model and architecture for medical event streams," in *HealthNet '07: Proceedings of the 1st ACM SIGMOBILE International Workshop on Systems and Networking Support for Healthcare and Assisted Living Environments*, pp. 95–100, New York, NY, USA, 2007. ACM. (doi: http://doi.acm.org/10.1145/1248054.1248082).

[278] S. Wang, A. Padmanabhan, J. D. Myers, W. Tang, and Y. Liu, "Towards provenance-aware geographic information systems," in *GIS '08: Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 1–4, New York, NY, USA, 2008. ACM. (doi: http://doi.acm.org/10.1145/1463434.1463515).

[279] Y. R. Wang and S. E. Madnick, "A polygen model for heterogeneous database systems: The source tagging perspective," in *VLDB '90: Proceedings of the 16th International Conference on Very Large Data Bases*, pp. 519–538, San Francisco, CA, USA, 1990. Morgan Kaufmann Publishers Inc. (url: http://web.mit.edu/tdqm/www/tdqmpub/polygenmodelAug90.pdf).

[280] R. E. Watkins and D. A. Nicole, "Named graphs as a mechanism for reasoning about provenance," in *Frontiers of WWW Research and Development — APWeb 2006: 8th Asia-Pacific Web Conference*, vol. 3841 of *Lecture Notes in Computer Science*, pp. 943–948, Harbin, China, 2006. (doi: http://dx.doi.org/10.1007/11610113).

[281] D. J. Weitzner, H. Abelson, T. Berners-Lee, J. Feigenbaum, J. Hendler, and G. J. Sussman, "Information accountability," *Communications of the ACM*, vol. 51, no. 6, pp. 81–87, June 2008. (doi: http://doi.acm.org/10.1145/1349026.1349043).

[282] J. Widom, "Trio: A system for integrated management of data, accuracy, and lineage," in *Second Biennial Conference on Innovative Data Systems Research (CIDR 2005), Asilomar, Calif.*, January 2005. (url: http://www.cidrdb.org/cidr2005/papers/P22.pdf).

[283] S. C. Wong, S. Miles, W. Fang, P. Groth, and L. Moreau, "Provenance-based validation of e-science experiments," in *Proceedings of 4th International Semantic Web Conference (ISWC'05)*, vol. 3729 of *Lecture Notes in Computer Science*, pp. 801–815, Galway, Ireland, November 2005. Springer-Verlag. (doi: http://dx.doi.org/10.1007/11574620_57).

[284] A. Woodruff and M. Stonebraker, "Supporting fine-grained data lineage in a database visualization environment," in *Proceedings of the 13th International Conference on Data Engineering*, pp. 91–102, Birmingham, England, April 1997. IEEE Computer Society. (doi: http://doi.ieeecomputersociety.org/10.110910.1109/ICDE.1997.581742).

[285] J. Zhang, A. Chapman, and K. LeFevre, "Fine-grained tamper-evident data pedigree," Technical Report CSE-TR-548-08, University of Michigan, 2008, (url: https://www.eecs.umich.edu/eecs/research/techreports/cse_tr/database/reports.cgi?08).

[286] J. Zhao, "A conceptual model for e-science provenance," PhD. thesis, University of Manchester, June 2007, (url: http://users.ox.ac.uk/∼zool0770/jun_thesis_final_2007.pdf).

[287] J. Zhao, C. Goble, M. Greenwood, C. Wroe, and R. Stevens, "Annotating, linking and browsing provenance logs for e-science," in *Proceedings of the ISWC 2003 Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data*, Sanibel Island, Florida, October 2003. (url: http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-83/prov_2.pdf).

[288] J. Zhao, C. Goble, R. Stephens, and S. Bechhofer, "Semantically linking and browsing provenance logs for e-science," in *Proceedings of the 1st International Conference on Semantics of a Networked World*, vol. 3226 of *Lecture Notes in Computer Science*, pp. 158–176, Paris, France, June 2004. Springer. (doi: http://dx.doi.org/10.1007/b102069).

[289] J. Zhao, C. Goble, and R. Stevens, "An identity crisis in the life sciences," in *Proceedings of the International Provenance and Annotation Workshop 2006 (IPAW'2006)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 254–269, Springer, 2006. (doi: http://dx.doi.org/10.1007/11890850_26).

[290] J. Zhao, C. Goble, R. Stevens, and D. Turi, "Mining taverna's semantic web of provenance," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 463–472, 2008. (doi: http://dx.doi.org/10.1002/cpe.1231).

[291] J. Zhao, A. Miles, G. Klyne, and D. Shotton, "Linked data and provenance in biological data webs," *Brief Bioinform*, p. bbn044+, December 2008. (doi: http://dx.doi.org/10.1093/bib/bbn044).

[292] J. Zhao, C. Wroe, C. Goble, R. Stevens, D. Quan, and M. Greenwood, "Using semantic web technologies for representing e-science provenance," in *Proceedings of Third International Semantic Web Conference (ISWC2004)*, vol. 3298 of *Lecture Notes in Computer Science*, pp. 92–106, Hiroshima, Japan, November 2004. Springer-Verlag. (doi: http://dx.doi.org/10.1007/b102467).

[293] Y. Zhao, "A virtual data language and system for scientific workflow management in data grid environments," PhD thesis, The University of Chicago, August 2007, (url: http://proquest.umi.com/pqdlink?did=1397920731&Fmt=7&clientI d=79356&RQT=309&VName=PQD).

[294] Y. Zhao, M. Wilde, and I. Foster, "Applying the virtual data provenance model," in *Proceedings of the International Provenance and Annotation Workshop 2006 (IPAW'2006)*, vol. 4145 of *Lecture Notes in Computer Science*, (L. Moreau and I. Foster, eds.), pp. 148–161, Springer, 2006. (doi: http://dx.doi.org/10.1007/11890850_16).