



UvA-DARE (Digital Academic Repository)

Collaborative provenance for workflow-driven science and engineering

Altıntaş, İ.

Publication date
2011

[Link to publication](#)

Citation for published version (APA):

Altıntaş, İ. (2011). *Collaborative provenance for workflow-driven science and engineering*.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Bibliography

- Abraham, John, Pearl Brazier, Artem Chebotko, Jaime Navarro and Anthony Piazza (2010). Distributed storage and querying techniques for a semantic web of scientific workflow provenance. In *SCC '10: Proceedings of the 2010 IEEE International Conference on Services Computing*. IEEE Computer Society. Washington, DC, USA. pp. 178–185.
- Abramson, David, Blair Bethwaite, Colin Enticott, Slavisa Garic and Tom Peachey (2009). Parameter space exploration using scientific workflows. In *ICCS '09: Proceedings of the 9th International Conference on Computational Science*. Springer-Verlag. Berlin, Heidelberg. pp. 104–113.
- Abramson, David, Colin Enticott and Ilkay Altintas (2008). Nimrod/k: towards massively parallel dynamic grid workflows. In *SC '08: Proceedings of the 2008 ACM/IEEE conference on Supercomputing*. IEEE/ACM. Piscataway, NJ, USA. pp. 1–11.
- Abramson, David, Jagan Kommineni and Ilkay Altintas (2005). Flexible io services in the kepler grid workflow system. In *E-SCIENCE '05: Proceedings of the First International Conference on e-Science and Grid Computing*. IEEE Computer Society. Washington, DC, USA. pp. 255–262.
- Aggarwal, Charu C. and Yu, Philip S., Eds.) (2008). *Privacy-preserving data mining: Models and algorithms*. Vol. 34 of *Advances in Database Systems*. Springer.
- Altintas, Ilkay (2008). Lifecycle of scientific workflows and their provenance: A usage perspective. In *SERVICES '08: Proceedings of the 2008 IEEE Congress on Services - Part I*. IEEE Computer Society. Washington, DC, USA. pp. 474–475.
- Altintas, Ilkay, Abel W. Lin, Jing Chen, Chris Churas, Madhusudan Gujral, Shulei Sun, Weizhong Li, Ramil Manansala, Mayya Sedova, Jeffrey S. Grethe and Mark Ellisman (2010a). Camera 2.0: A data-centric metagenomics community infrastructure driven by scientific workflows. In *SWF 2010 in conjunction with 6th World Congress on Services (SERVICES 2010)*. IEEE Computer Society. pp. 352–359.
- Altintas, Ilkay, Adam Birnbaum, Kim K. Baldridge, Wibke Sudholt, Mark Miller, Celine Amoreira, Yohann Potier and Bertram Ludaescher (2004a). A framework for the design and reuse of grid workflows. In *Scientific Applications of Grid Computing: First International Workshop; Lecture Notes in Computer Science* (P. Herrero, M.S. Perez and

- V. Robles, Eds.). number 3 In *Lecture Notes in Computer Science*. Springer-Verlag GmbH. pp. 119–132.
- Altintas, Ilkay, Chad Berkley, Efrat Jaeger, Matthew B. Jones, Bertram Ludäscher and Steve Mock (2004b). Kepler: An extensible system for design and execution of scientific workflows. In *Proceedings of the 16th International Conference on Scientific and Statistical Database Management (SSDBM 2004)*. IEEE Computer Society. Washington, DC, USA. pp. 423–424.
- Altintas, Ilkay, Daniel Crawl, C.J. Crosby and Peter Cornillon (2010b). *Geoinformatics: Cyberinfrastructure for the Solid Earth Sciences (Keller, G. Randall and Baru, Chaitanya Eds.)*. Chap. Scientific Workflows for the Geosciences: An Emerging Approach to Building Integrated Data Analysis Systems. Cambridge University Press.
- Altintas, Ilkay, Efrat Jaeger, Kai Lin, Bertram Ludaescher and Ashraf Memon (2004c). A web service composition and deployment framework for scientific workflows. In *ICWS '04: Proceedings of the IEEE International Conference on Web Services*. IEEE Computer Society. Washington, DC, USA. p. 814.
- Altintas, Ilkay, Jing Jing Chen, Mayya Sedova, Amarnath Gupta, Shulei Sun, Abel W. Lin, Madhusudan Gujral, Manish K. Anand, Weizhong Li, Jeffrey S. Grethe and Mark Ellisman (2010c). Extending the data model for data-centric metagenomics analysis using scientific workflows in camera. In *Proceedings of HPC for Life Sciences Workshop at eScience 2010*.
- Altintas, Ilkay, Manish Anand, Daniel Crawl, Shawn Bowers, Adam Belloum, Paolo Missier, Bertram Ludäscher, Carole Goble and Peter Sloot (2010d). Understanding collaborative studies through interoperable workflow provenance. In *Provenance and Annotation of Data and Processes* (Deborah McGuinness, James Michaelis and Luc Moreau, Eds.). Vol. 6378. pp. 42–58. Springer Berlin / Heidelberg.
- Altintas, Ilkay, Manish Kumar Anand, Adam Belloum, Gokhan Ertaylan, Bartosz Balis, Marian Bubak and Peter M.A. Sloot (2010e). Collaborative provenance for workflow-driven science - a position paper. Submitted to FGCS.
- Altintas, Ilkay, Manish Kumar Anand, Trung Vuong, Shawn Bowers, Bertram Ludäscher and Peter M.A. Sloot (2010f). A data model for analyzing user collaborations in workflow-driven escience. Submitted to the International Journal of Computers and Their Applications (IJCA), Special Issue on Scientific Workflows, Provenance and Their Applications.
- Altintas, Ilkay, Oscar Barney and Efrat Jaeger-Frank (2006a). Provenance collection support in the kepler scientific workflow system. In *Provenance and Annotation of Data (IPAW 2006, Revised Selected Papers)* (Luc Moreau and Ian Foster, Eds.). Vol. 4145 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. pp. 118–132.
- Altintas, Ilkay, Oscar Barney, Zhengang Cheng, Terence Critchlow, Bertram Ludaescher, Steve Parker, Arie Shoshani and Mladen Vouk (2006b). Accelerating the scientific exploration process with scientific workflows. In *SciDAC 2006, SCIENTIFIC DISCOVERY*

- THROUGH ADVANCED COMPUTING* (Dr. William M Tang, Ed.). Vol. Vol. 46. pp. 468–478.
- Altintas, Ilkay, Sangeeta Bhagwanani, David Buttler, Sandeep Chandra, Zhengang Cheng, Matthew Coleman, Terence Critchlow, Amarnath Gupta, Wei Han, Ling Liu, Bertram Ludäscher, Calton Pu, Reagan Moore, Arie Shoshani and Mladen Vouk (2003). A modeling and execution environment for distributed scientific workflows. In *15th International Conference on Scientific and Statistical Database Management (SSDBM)*. Boston, Massachusetts. pp. 247–250.
- Anand, Manish Kumar, Shawn Bowers and Bertram Ludäscher (2009a). A navigation model for exploring scientific workflow provenance graphs. In *Proceedings of the 4th Workshop on Workflows in Support of Large-Scale Science. WORKS'09*. ACM. New York, NY, USA. pp. 2:1–2:10.
- Anand, Manish Kumar, Shawn Bowers and Bertram Ludäscher (2010a). Provenance browser: Displaying and querying scientific workflow provenance graphs (Demo). In *IEEE 26th International Conference on Data Engineering (ICDE), 2010*. pp. 1201–1204.
- Anand, Manish Kumar, Shawn Bowers and Bertram Ludäscher (2010b). Techniques for efficiently querying scientific workflow provenance graphs. In *Proceedings of the 13th International Conference on Extending Database Technology. EDBT'10*. ACM. New York, NY, USA. pp. 287–298.
- Anand, Manish Kumar, Shawn Bowers, Timothy McPhillips and Bertram Ludäscher (2009b). Efficient provenance storage over nested data collections. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology. EDBT'09*. ACM. New York, NY, USA. pp. 958–969.
- Anand, Manish Kumar, Shawn Bowers, Timothy McPhillips and Bertram Ludäscher (2009c). 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 2009*. Springer-Verlag. Berlin, Heidelberg. pp. 237–254.
- Anand, Manish, Shawn Bowers, Ilkay Altintas and Bertram Ludäscher (2010c). Approaches for exploring and querying scientific workflow provenance graphs. In *Provenance and Annotation of Data and Processes* (Deborah McGuinness, James Michaelis and Luc Moreau, Eds.). Vol. 6378 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. pp. 17–26.
- Angly, F., B. Rodriguez-Brito, D. Bangor, P. McNairnie, M. Breitbart, P. Salamon, B. Felts, J. Nulton, J. Mahaffy and F. Rohwer (2005). Phacces, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *BMC Bioinformatics*.
- Backstrom, Lars, Cynthia Dwork and Jon Kleinberg (2007). Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In *WWW '07*:

- Proceedings of the 16th international conference on World Wide Web*. ACM. New York, NY, USA. pp. 181–190.
- Baldrige, Kim K., Wibke Sudholt, Jerry P. Greenberg, Celine Amoreira, Yohann Potier, Ilkay Altintas, Adam Birnbaum, David Abramson, Colin Enticott and Slavisa Garic (2005). *Parallel Computing for Bioinformatics* (ed. A. Y. Zomaya). Chap. Cluster and Grid Infrastructure for Computational Chemistry and Biochemistry. John Wiley & Sons.
- Balis, Bartosz, Marian Bubak, M. Pelczar and J. Wach (2009). Provenance tracking and end-user oriented query construction. In *Handbook of Research on Computational Grid Technologies for Life Sciences, Biomedicine and Healthcare, Information Science Reference* (M. Cannataro, Ed.). pp. 60–75. IGI Global.
- Balis, Bartosz, Marian Bubak, Michal Pelczar and Jakub Wach (2008). Provenance tracking and querying in the virolab virtual laboratory. *Cluster Computing and the Grid, IEEE International Symposium on*. **0**, 675–680.
- Barga, Roger S. and Luciano A. Digiampietri (2008). Automatic capture and efficient storage of e-science experiment provenance. *Concurrency and Computation Practice and Experience*. **20**(5), 419–429.
- Barseghian, Derik, Ilkay Altintas, Matthew B. Jones, Daniel Crawl, Nathan Potter, James Gallagher, Peter Cornillon, Mark Schildhauer, Elizabeth T. Borer, Eric W. Seabloom and Parvizeh R. Hosseini (2010). Workflows and extensions to the kepler scientific workflow system to support environmental sensor data access and analysis. *Ecological Informatics*. **5**(1), 42–50.
- Berman, Francine (2008). Got data?: a guide to data preservation in the information age. *Communications of the ACM*. **51**(12), 50–56.
- Berthold, Michael R., Nicolas Cebron, Fabian Dill, Thomas R. Gabriel, Tobias Kötter, Thorsten Meinel, Peter Ohl, Kilian Thiel and Bernd Wiswedel (2009). Knime - the konstanz information miner: version 2.0 and beyond. *SIGKDD Explor. Newsl.* **11**(1), 26–31.
- Bose, Rajendra and James Frew (2005). Lineage retrieval for scientific data processing: a survey. *ACM Comput. Surv.* **37**(1), 1–28.
- Bowers, Shawn and Bertram Ludäscher (2004). An ontology-driven framework for data transformation in scientific workflows. In *Data Integration in the Life Sciences* (Rahm Erhard, Ed.). Vol. 2994 of *Lecture Notes in Computer Science*. pp. 1–16. Springer Berlin / Heidelberg.
- Bowers, Shawn and Bertram Ludäscher (2005). Actor-oriented design of scientific workflows. In *Proceedings of the International Conference on Conceptual Modeling (ER'2005)*. Vol. 3716 of *Lecture Notes in Computer Science*.
- Bowers, Shawn, Timothy M. McPhillips and Bertram Ludäscher (2008a). Provenance in collection-oriented scientific workflows. *Concurrency and Computation: Practice and Experience*. **20**(5), 519–529.
- Bowers, Shawn, Timothy McPhillips, Martin Wu Wu and Bertram Ludäscher (2007). *Data*

- Integration in the Life Sciences*. Chap. Project Histories: Managing Data Provenance Across Collection-Oriented Scientific Workflow Runs, pp. 122–138. Vol. 4544 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg.
- Bowers, Shawn, Timothy McPhillips, Sean Riddle, Manish Kumar Anand and Bertram Ludäscher (2008b). Kepler/ppod: Scientific workflow and provenance support for assembling the tree of life. In *Provenance and Annotation of Data and Processes: Second International Provenance and Annotation Workshop, IPAW 2008, Salt Lake City, UT, USA, June 17-18, 2008. Revised Selected Papers*. Springer-Verlag. Berlin, Heidelberg. pp. 70–77.
- Bubak, Marian T., T. Gubala, M. Kasztelnik, M. Malawski, P. Nowakowski and Peter M.A. Sloot (2007). Collaborative virtual laboratory for e-health. In *Expanding the Knowledge Economy: Issues, Applications, Case Studies, eChallenges e-2007 Conference Proceedings* (P. Cunningham and M. Cunningham, Eds.). IOS Press. Amsterdam. pp. 537–544.
- Buneman, Peter, Sanjeev Khanna and Wang Chiew Tan (2000). Data provenance: Some basic issues. In *Proceedings of the 20th Conference on Foundations of Software Technology and Theoretical Computer Science*. FST TCS 2000. Springer-Verlag. London, UK. pp. 87–93.
- Buneman, Peter, Sanjeev Khanna and Wang Chiew Tan (2001). Why and where: A characterization of data provenance. In *Proceedings of the 8th International Conference on Database Theory*. ICDT '01. Springer-Verlag. London, UK. pp. 316–330.
- Callahan, Steven P., Juliana Freire, Emanuele Santos, Carlos E. Scheidegger, Cláudio T. Silva and Huy T. Vo (2006). Vistrails: visualization meets data management. In *SIGMOD '06: Proceedings of the 2006 ACM SIGMOD international conference on Management of data*. ACM. New York, NY, USA. pp. 745–747.
- Cao, Bin, Beth Plale, G. Subramanian, E. Robertson and Yogesh Simmhan (2009). Provenance information model of karma version 3. In *World Conference on Services - I*. IEEE. pp. 348–351.
- Carey, Stephen S. (2003). *A Beginner's Guide to Scientific Method*. 3 ed.. Wadsworth Publishing.
- Carter, W., R. Shrestha, G. Tuell, D. Bloomquist and M. Sartori (2001). Airborne laser swath mapping shines new light on earth's topography. *Eos Trans. AGU*. **82**(46), 549.
- Cheney, James (2010). Causality and the semantics of provenance. In *DCM*. pp. 63–74.
- Cheney, James, Laura Chiticariu and Wang-Chiew Tan (2009). Provenance in databases: Why, how, and where. *Foundations and Trends in Databases*. pp. 379–474.
- Clifford, Ben, Ian Foster, Jens-S. Voeckler, Michael Wilde and Yong Zhao (2008). Tracking provenance in a virtual data grid. *Concurrency and Computation: Practice and Experience*. **20**(5), 565–575.
- Cohen, Shirley, Sarah Cohen-Boulakia and Susan Davidson (2006). Towards a model of provenance and user views in scientific workflows. In *Data Integration in the Life Sciences* (S. Istrail, P. Pevzner and M. Waterman, Eds.). Vol. 4075/2006 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. Berlin, Heidelberg. pp. 264–279.

- Committee on Metagenomics: Challenges and Functional Applications (2007). *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet*. The National Academies Press.
- Crawl, Daniel and Ilkay Altintas (2008). A provenance-based fault tolerance mechanism for scientific workflows. In *Provenance and Annotation of Data and Processes (IPAW 2008, Revised Selected Papers)* (Juliana Freire, David Koop and Luc Moreau, Eds.). Vol. 5272 of *Lecture Notes in Computer Science*. Springer. pp. 152–159.
- Cuadrado, D.L. (2008). *Automated Distribution Simulation in Ptolemy II*. PhD thesis. Aalborg University.
- Czajkowski, K., D.F. Ferguson, Ian Foster, J. Frey, S. Graham, I. Sedukhin, D. Snelling, S. Tuecke and W. Vambenepe (2004). The ws-resource framework. Technical report. Globus Project.
- Davidson, Susan B. and Juliana Freire (2008). Provenance and scientific workflows: challenges and opportunities. In *SIGMOD '08: Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. ACM. New York, NY, USA. pp. 1345–1350.
- Davidson, Susan B., Sanjeev Khanna, Sudeepa Roy and Sarah Cohen Boulakia (2010). Privacy issues in scientific workflow provenance. In *Wands '10: Proceedings of the 1st International Workshop on Workflow Approaches to New Data-centric Science*. ACM. New York, NY, USA. pp. 1–6.
- Davidson, Susan B., Sarah Cohen Boulakia, Anat Eyal, Bertram Ludäscher, Timothy M. McPhillips, Shawn Bowers, Manish Kumar Anand and Juliana Freire (2007). Provenance in scientific workflow systems. *IEEE Data Eng. Bull.* **30**(4), 44–50.
- Deelman, Ewa and Yolanda Gil (2006). Managing large-scale scientific workflows in distributed environments: Experiences and challenges. *e-Science and Grid Computing, International Conference on*. **0**(0-7695-2734-5), 144.
- Deelman, Ewa, Dennis Gannon, Matthew Shields and Ian Taylor (2009). Workflows and e-science: An overview of workflow system features and capabilities. *Future Generation Computer Systems*. **25**(5), 528–540.
- Deelman, Ewa, James Blythe, Yolanda Gil, Carl Kesselman, Gaurang Mehta, Sonal Patil, Mei-hui Su, Karan Vahi and Miron Livny (2004). Pegasus: Mapping scientific workflows onto the grid. In *Grid Computing*. Vol. 3165/2004 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. pp. 131–140.
- DeRemer, Frank and Hans Kron (1975). Programming-in-the large versus programming-in-the-small. *SIGPLAN Not., ACM*. **10**(6), 114–121.
- Djorgovski, S.G. (2005). Virtual astronomy, information technology, and the new scientific methodology. In *Proceedings of Seventh International Workshop on Computer Architecture for Machine Perception (CAMP 2005)*. IEEE.
- Ellqvist, T., D. Koop, J. Freire, C. Silva and L. Stromback (2009). Using mediation to achieve

- provenance interoperability. In *Proceedings of 2009 World Conference on Services - I*. IEEE. pp. 291–298.
- Elmroth, Erik, Francisco Hernández and Johan Tordsson (2010). Three fundamental dimensions of scientific workflow interoperability: Model of computation, language, and execution environment. *Future Generation Computer Systems*. **26**(2), 245–256.
- Fahringer, Thomas, Radu Prodan, Rubing Duan, Jüurgen Hofer, Farrukh Nadeem, Francesco Nerieri, Stefan Podlipnig, Jun Qin, Mumtaz Siddiqui, Hong-Linh Truong, Alex Villazon and Marek Wieczorek (2007). Askalon: A development and grid computing environment for scientific workflows. In *Workflows for e-Science* (Ian J. Taylor, Ewa Deelman, Dennis B. Gannon and Matthew Shields, Eds.). pp. 450–471. Springer London.
- Foster, Ian, Carl Kesselman, Jeffrey Nick and Steven Tuecke (2002). The physiology of the grid: An open grid services architecture for distributed systems integration. Technical report. Globus Project.
- Freeman, Linton C. (1977). A set of measures of centrality based on betweenness. *Sociometry*. **40**(1), 35–41.
- Freeman, Linton C. (1979). Centrality in social networks conceptual clarification. *Social Networks*. **1**(3), 215–239.
- Freire, Juliana, Cláudio T. Silva, Steven P. Callahan, Emanuele Santos, Carlos E. Scheidegger and Huy T. Vo (2006). Managing rapidly-evolving scientific workflows. In *Provenance and Annotation of Data* (Luc Moreau and Ian Foster, Eds.). Vol. 4145 of *Lecture Notes in Computer Science*. pp. 10–18. Springer Berlin / Heidelberg.
- Freire, Juliana, David Koop, Emanuele Santos and Claudio T. Silva (2008). Provenance for computational tasks: A survey. *Computing in Science and Engineering*. **10**(3), 11–21.
- Gadang, Siddharta S., Brajendra Panda and Joseph E. Hoag (2008). Provenance tracking with bit vectors. *International Symposium on Information Assurance and Security*. pp. 132–137.
- Gadelha Jr., Luiz M.R., Ben Clifford, Marta Mattoso, Michael Wilde and Ian Foster (2010). Provenance management in swift. *Future Generation Computer Systems*.
- Gao, Yong, June Kinoshita, Elizabeth Wu, Eric Miller, Ryan Lee, Andy Seaborne, Steve Cayzer and Tim Clark (2006). SWAN: A distributed knowledge infrastructure for alzheimer disease research. *Journal of Web Semantics*. **4**(3), 222–228.
- Gil, Yolanda, Ewa Deelman, Mark Ellisman, Thomas Fahringer, Geoffrey Fox, Dennis Gannon, Carole Goble, Miron Livny, Luc Moreau and Jim Myers (2007). Examining the challenges of scientific workflows. *IEEE Computer*. **40**(12), 24–32.
- Gil, Yolanda, Varun Ratnakar and Ewa Deelman (2006). Metadata catalogs with semantic representations. In *Provenance and Annotation of Data* (Luc Moreau and Ian Foster, Eds.). Vol. 4145 of *Lecture Notes in Computer Science*. pp. 90–100. Springer Berlin / Heidelberg.
- Goble, Carole A., Jiten Bhagat, Sergejs Aleksejevs, Don Cruickshank, Danius Michaelides, David Newman, Mark Borkum, Sean Bechhofer, Marco Roos, Peter Li and David

- De Roure (2010). myexperiment: a repository and social network for the sharing of bioinformatics workflows. *Nucleic Acids Research*.
- Goderis, Antoon, Christopher Brooks, Ilkay Altintas, Edward A. Lee and Carole A. Goble (2009). Heterogeneous composition of models of computation. *Future Generation Computer Systems*. **25**(5), 552–560.
- Golbeck, Jennifer and James Hendler (2007). A semantic web approach to tracking provenance in scientific workflows. *Concurrency and Computation Practice and Experience*. **20**(5), 431–439.
- Graves, Alvaro (2010). POMELo: A PML online editor. In *Provenance and Annotation of Data and Processes* (Deborah McGuinness, James Michaelis and Luc Moreau, Eds.). Vol. 6378 of *Lecture Notes in Computer Science*. pp. 91–97. Springer Berlin / Heidelberg.
- Groth, Paul, Michael Luck and Luc Moreau (2005). A protocol for recording provenance in service-oriented grids. In *Principles of Distributed Systems*. Vol. 3544 of *Lecture Notes in Computer Science*. pp. 124–139. Springer Berlin / Heidelberg.
- Groth, Paul, Simon Miles and Luc Moreau (2009). A model of process documentation to determine provenance in mash-ups. *ACM Trans. Internet Technol.* **9**(1), 1–31.
- Heinis, Thomas and Gustavo Alonso (2008). Efficient lineage tracking for scientific workflows. In *SIGMOD '08*. pp. 1007–1018.
- Holland, David A., Uri Braun, Diana Maclean, Kiran-kumar Muniswamy-reddy and Margo I. Seltzer (2008). Choosing a data model and query language for provenance. In *IPAW 2008*.
- Hull, Duncan, Katy Wolstencroft, Robert Stevens, Carole Goble, Mathew R. Pocock, Peter Li and Tom Oinn (2006). Taverna: a tool for building and running workflows of services. *Nucl. Acids Res.* **24**(2), 729–732.
- Jaeger-Frank, E., C.J. Crosby, A. Memon, V. Nandigam, J. Conner, J.R. Arrowsmith, I. Altintas and C. Baru (2006a). A three tier architecture applied to LiDAR processing and monitoring. *Scientific Programming*. **14**(3), 185–194.
- Jaeger-Frank, E., C.J. Crosby, A. Memon, V. Nandigam, J.R. Arrowsmith, J. Conner, I. Altintas and C. Baru (2006b). Three tier architecture for LiDAR interpolation and analysis. In *1st International Workshop on Workflow Systems in e-Science in Conjunction with ICCS 2006* (Vassil N. Alexandrov, G. Dick van Albada, Peter M. A. Sloot and Jack Dongarra, Eds.). Vol. 3 of *Lecture Notes in Computer Science*. Springer. Berlin, Heidelberg. pp. 920–927. 1st International Workshop on Workflow Systems in e-Science (WSES06).
- Kacsuk, Peter and Gergely Sipos (2006). Multi-grid, multi-user workflows in the p-grade grid portal. *Journal of Grid Computing*. pp. 1–18.
- Kahn, Gilles (1974). The semantics of a simple language for parallel programming. In *In Information Processing '74: Proceedings of the IFIP Congress*. pp. 471–475.
- Kifer, Michael, Georg Lausen and James Wu (1995). Logical foundations of object-oriented and frame-based languages. *J. ACM*. **42**(4), 741–843.

- Kim, Jihie, Ewa Deelman, Yolanda Gil, Gaurang Mehta and Varun Ratnakar (2008). Provenance trails in the wings-pegasus system. *Concurrency and Computation: Practice and Experience*. **20**(5), 587–597.
- Korkhov, Vladimir, Dmitry Vasyunin, Adianto Wibisono, Victor Guevara-Masis, Adam Beloum, Cees de Laat, Pieter Adriaans and L.O. Hertzberger (2007). Ws-vlam: towards a scalable workflow system on the grid. In *WORKS '07: Proceedings of the 2nd workshop on Workflows in support of large-scale science*. ACM. New York, NY, USA. pp. 63–68.
- Lee, Edward A. and Thomas M. Parks (1995). Dataflow process networks. *Proceedings of the IEEE*. **83**(5), 773–801.
- Leymann, Frank (2001). Web Services Flow Language (WSFL 1.0), IBM. Technical report. IBM.
- Lin, Cui, Shiyong Lu, Xubo Fei, Artem Chebotko, Darshan Pai, Zhaoqiang Lai, Farshad Fotouhi and Jing Hua (2009a). A reference architecture for scientific workflow management systems and the view soa solution. *IEEE Trans. Serv. Comput.*. **2**(1), 79–92.
- Lin, Cui, Shiyong Lu, Xubo Fei, Darshan Pai and Jing Hua (2009b). A task abstraction and mapping approach to the shimming problem in scientific workflows. In *Proceedings of the 2009 IEEE International Conference on Services Computing. SCC '09*. IEEE Computer Society. Washington, DC, USA. pp. 284–291.
- Ludäscher, Bertram and Goble, Carole, Eds.) (2005). *Special section on scientific workflows*. Vol. 34:3. ACM SIGMOD Record.
- Ludäscher, Bertram, Ilkay Altintas, Chad Berkley, Dan Higgins, Efrat Jaeger-Frank, Matthew Jones, Edward Lee, Jing Tao and Yang Zhao (2006). Scientific workflow management and the kepler system. *Concurrency and Computation: Practice and Experience, Special Issue on Scientific Workflows*. **18**(10), 1039–1065.
- Ludäscher, Bertram, Ilkay Altintas, Shawn Bowers, Julian Cummings, Terence Critchlow, Ewa Deelman, David De Roure, Juliana Freire, Carole Goble, Matthew Jones, Scott Klasky, Timothy McPhillips, Norbert Podhorszki, Cláudio T. Silva, Ian Taylor and Mladen A. Vouk (2009). *Scientific Data Management: Challenges, Technology, and Deployment (Shoshani, Arie and Rotem, Doron, Eds.)*. Chap. Chapter 13: Scientific Process Automation and Workflow Management. Computational Science Series. Chapman and Hall/CRC.
- Ludäscher, Bertram, Norbert Podhorszki, Ilkay Altintas, Shawn Bowers and Timothy M. McPhillips (2008). From computation models to models of provenance: the rws approach. *Concurrency and Computation: Practice and Experience*. **20**(5), 507–518.
- Malawski, Maciej, Tomasz Bartynski and Marian Bubak (2010). Invocation of operations from script-based grid applications. *Future Generation Computer Systems*. **26**(1), 138 – 146.
- Mandal, Nandita, Ewa Deelman, Gaurang Mehta, Mei-Hui Su and Karan Vahi (2007). Integrating existing scientific workflow systems: the kepler/pegasus example. In *WORKS*

- '07: *Proceedings of the 2nd workshop on Workflows in support of large-scale science, Monterey, California, USA*. ACM. New York, NY, USA. pp. 21–28.
- Margo, Daniel W. and Margo Seltzer (2009). The case for browser provenance. In *TAPP'09: First workshop on Theory and practice of provenance*. USENIX Association. Berkeley, CA, USA. pp. 1–5.
- McGuinness, Deborah, L. Ding, P. Pinheiro da Silva and C. Chang (2007). PML 2: A modular explanation interlingua.. In *ExaCt*. pp. 49–55.
- McPhillips, Timothy, Shawn Bowers and Bertram Ludäscher (2006). Collection-oriented scientific workflows for integrating and analyzing biological data. In *Data Integration in the Life Sciences* (Ulf Leser, Felix Naumann and Barbara Eckman, Eds.). Vol. 4075 of *Lecture Notes in Computer Science*. pp. 248–263. Springer Berlin / Heidelberg.
- Mesirov, Jill P. (2010). Accessible reproducible research. *Science*. **327**(5964), 415–416.
- Miles, Simon, Paul Groth, Miguel Branco and Luc Moreau (2007). The requirements of using provenance in e-science experiments. *Journal of Grid Computing*. **5**(1), 1–25.
- Miles, Simon, Paul Groth, Steve Munroe and Luc Moreau (2009). Prime: A methodology for developing provenance-aware applications. *ACM Transactions on Software Engineering and Methodology*.
- Missier, Paolo, Carole Goble, Saumen Dey, Anandarup Sarkar, Biva Shrestha, Bertram Ludäscher, Shawn Bowers, Ilkay Altintas and Manish Kumar Anand (2010a). Linking multiple workflow provenance traces for interoperable collaborative science. In *Proceedings of the 5th Workshop on Workflows in Support of Large-Scale Science*. WORKS'10. ACM. New York, NY, USA. pp. 1–8.
- Missier, Paolo, Satya Sahoo, Jun Zhao, Carole Goble and Amit Sheth (2010b). Janus: From workflows to semantic provenance and linked open data. In *Provenance and Annotation of Data and Processes* (Deborah McGuinness, James Michaelis and Luc Moreau, Eds.). Vol. 6378 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. pp. 129–141.
- Moreau, Luc, Ben Clifford, Juliana Freire, Yolanda Gil, Paul Groth, Joe Futrelle, Natalia Kwasnikowska, Simon Miles, Paolo Missier, Jim Myers, Yogesh Simmhan, Eric Stephan and Jan Van den Bussche (2010). The open provenance model core specification (v1.1). Submitted to Future Generation Computer Systems.
- Moreau, Luc, Bertram Ludäscher, Ilkay Altintas, Roger S. Barga, Shawn Bowers, Steven P. Callahan, George Chin Jr., Ben Clifford, Shirley Cohen, Sarah Cohen Boulakia, Susan B. Davidson, Ewa Deelman, Luciano A. Digiampietri, Ian T. Foster, Juliana Freire, James Frew, Joe Futrelle, Tara Gibson, Yolanda Gil, Carole A. Goble, Jennifer Golbeck, Paul T. Groth, David A. Holland, Sheng Jiang, Jihie Kim, David Koop, Ales Krenek, Timothy M. McPhillips, Gaurang Mehta, Simon Miles, Dominic Metzger, Steve Munroe, Jim Myers, Beth Plale, Norbert Podhorszki, Varun Ratnakar, Emanuele Santos, Carlos Eduardo Scheidegger, Karen Schuchardt, Margo I. Seltzer, Yogesh L. Simmhan, Cláudio T. Silva, Peter Slaughter, Eric G. Stephan, Robert Stevens, Daniele Turi, Huy T. Vo, Michael Wilde, Jun

- Zhao and Yong Zhao (2008). Special issue: The first provenance challenge. *Concurrency and Computation: Practice and Experience*. **20**(5), 409–418.
- Mouallem, Pierre, Daniel Crawl, Ilkay Altintas, Mladen A. Vouk and Ustun Yildiz (2010). A fault-tolerance architecture for kepler-based distributed scientific workflows. In *Proceedings of Scientific and Statistical Database Management, 22nd International Conference (SSDBM 2010)*. Vol. 6187 of *Lecture Notes in Computer Science*. Springer. Berlin, Heidelberg. pp. 452–460.
- Ngu, Anne H., Shawn Bowers, Nicholas Haasch, Timothy McPhillips and Terence Critchlow (2008). Flexible scientific workflow modeling using frames, templates, and dynamic embedding. In *SSDBM '08: Proceedings of the 20th international conference on Scientific and Statistical Database Management*. Springer-Verlag. Berlin, Heidelberg. pp. 566–572.
- Novotny, Jason, Michael Russell and Oliver Wehrens (2004). Gridsphere: An advanced portal framework. *EUROMICRO Conference*. pp. 412–419.
- Oinn, Tom, Mark Greenwood, Matthew Addis, M. Nedim Alpdemir, Justin Ferris, Kevin Glover, Carole Goble, Antoon Goderis, Duncan Hull, Darren Marvin, Peter Li, Phillip Lord, Matthew R. Pocock, Martin Senger, Robert Stevens, Anil Wipat and Chris Wroe (2006). Taverna: lessons in creating a workflow environment for the life sciences. *Concurrency and Computation: Practice and Experience, Special Issue on Scientific Workflows*. **18**(10), 1067–1100.
- Otte, Evelien and Ronald Rousseau (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*. **28**(6), 441–453.
- Pettifer, Steve, Katy Wolstencroft, Pinar Alper, Teresa Attwood, Alain Coletta, Carole Goble, Peter Li, Philip McDermott, Tom Marsh, James Oinn, James Sinnott and David Thorne (2007). mygrid and utopia: An integrated approach to enacting and visualising in silico experiments in the life sciences. In *Data Integration in the Life Sciences*. Vol. 4544/2007 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. pp. 59–70.
- Podhorszki, Norbert, Bertram Ludaescher and Scott A. Klasky (2007). Workflow automation for processing plasma fusion simulation data. In *WORKS '07: Proceedings of the 2nd workshop on Workflows in support of large-scale science*. ACM. New York, NY, USA. pp. 35–44.
- Podhorszki, Norbert, Scott Klasky, Qing Liu, Ciprian Docan, Manish Parashar, Hasan Ab-basi, Jay Lofstead, Karsten Schwan, Matthew Wolf, Fang Zheng and Julian Cummings (2009). Plasma fusion code coupling using scalable i/o services and scientific workflows. In *WORKS '09: Proceedings of the 4th Workshop on Workflows in Support of Large-Scale Science*. ACM. New York, NY, USA. pp. 1–9.
- Rex, D. E., J.Q. Ma and A.W. Toga (2003). The Ioni pipeline processing environment. *Neuroimage*. **19**(3), 1033–1048.
- Roure, David De, Carole Goble and Robert Stevens (2007). Designing the myexperiment virtual research environment for the social sharing of workflows. In *E-SCIENCE '07: Pro-*

- ceedings of the Third IEEE International Conference on e-Science and Grid Computing*. IEEE Computer Society. Washington, DC, USA. pp. 603–610.
- Rowe, Anthony, Dimitrios Kalaitzopoulos, Michelle Osmond, Moustafa Ghanem and Yike Guo (2003). The discovery net system for high throughput bioinformatics. *Bioinformatics*. **19**, 225–231.
- Sahoo, Satya S., Amit Sheth and Cory Henson (2008). Semantic provenance for escience: Managing the deluge of scientific data. *IEEE Internet Computing*. **12**(4), 46–54.
- Scott, John P (2000). *Social Network Analysis: A Handbook*. 2 ed.. Sage Publications Ltd.
- Serra da Cruz, Sérgio Manuel, Maria Luiza M. Campos and Marta Mattoso (2009). Towards a taxonomy of provenance in scientific workflow management systems. *IEEE Congress on Services*. pp. 259–266.
- Seshadri, R., S.A. Kravitz, L. Smarr, P. Gilna and M. Frazier (2007). Camera: A community resource for metagenomics. *PLoS Biology*. **5**(3), e75.
- Simmhan, Yogesh, Beth Plale, Dennis Gannon and Suresh Marru (2006a). Performance evaluation of the karma provenance framework for scientific workflows. In *Provenance and Annotation of Data* (Luc Moreau and Ian Foster, Eds.). Vol. 4145 of *Lecture Notes in Computer Science*. pp. 222–236. Springer Berlin / Heidelberg.
- Simmhan, Yogesh L., Beth Plale and Dennis Gannon (2005). A survey of data provenance in e-science. *SIGMOD Record*. **34**(3), 31–36.
- Simmhan, Yogesh L., Beth Plale and Dennis Gannon (2006b). A framework for collecting provenance in data-centric scientific workflows. In *ICWS '06: Proceedings of the IEEE International Conference on Web Services*. IEEE Computer Society. Washington, DC, USA. pp. 427–436.
- Simmhan, Yogesh L., Beth Plale and Dennis Gannon (2008). Query capabilities of the karma provenance framework. *Concurrency and Computation: Practice and Experience*. **20**(5), 441–451.
- Sloot, Peter M.A., Alfredo Tirado-Ramos, Ilkay Altintas, Marian T. Bubak and Charles A. Boucher (November 2006). From molecule to man: Decision support in individualized e-health. *IEEE Computer*. **39**(11), 40–46.
- Sloot, Peter M.A., PETER V. Coveney, Gokhan Ertaylan, V. Mueller, C.A. Boucher and Marian A Bubak, Marian (2009). Hiv decision support: from molecule to man. *Phil. Trans. R. Soc.* **367**(doi: 10.1098/rsta.2009.0043), 2691–2703.
- Smachat, Sucha, Maria Indrawan, Sea Ling, Colin Enticott and David Abramson (2009). Scheduling multiple parameter sweep workflow instances on the grid. In *E-SCIENCE '09: Proceedings of the 2009 Fifth IEEE International Conference on e-Science*. IEEE Computer Society. Washington, DC, USA. pp. 300–306.
- Stephenson, Karen and Marvin Zelen (1989). Rethinking centrality: Methods and examples. *Social Networks*. **11**(1), 1–37.

- Stevens, Robert, Jun Zhao and Carole Goble (2007). Using provenance to manage knowledge of in silico experiments. *Briefings in Bioinformatics*. **8**(3), 183–194.
- Sudholt, Wibke, Ilkay Altintas and Kim Baldridge (2006). Scientific workflow infrastructure for computational chemistry on the grid. In *Computational Science – ICCS 2006* (Vassil Alexandrov, Geert van Albada, Peter Sloot and Jack Dongarra, Eds.). Vol. 3993 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. pp. 69–76.
- Sun, Shulei, Jing Chen, Li Weizhong, Jeffrey Grethe, Ilkay Altintas, Abel W. Lin, Steve Peltier, Karen Stocks, Eric E. Allen, Mark Ellisman and John Wooley (2010). Community cyberinfrastructure for advanced microbial ecology research and analysis - the camera resource. *Nucleic Acids Research, Database Issue*.
- Tan, Wei, Paolo Missier, Ian Foster, Ravi Madduri, David De Roure and Carole Goble (2010). A comparison of using taverna and bpel in building scientific workflows: the case of cagrid. *Concurrency and Computation Practice and Experience*. **22**(9), 1098–1117.
- Taylor, Ian (2006). Triana generations. In *E-SCIENCE '06: Proceedings of the Second IEEE International Conference on e-Science and Grid Computing*. IEEE Computer Society. Washington, DC, USA. p. 143.
- Taylor, Ian J., Deelman, Ewa, Gannon, Dennis B. and Shields, Matthew, Eds.) (2007). *Workflows for e-Science*. Springer.
- Vouk, M., I. Altintas, R. Barreto, J. Blondin, Z. Cheng, T. Critchlow, A. Khan, S. Klasky, J. Ligon, B. Ludaescher, P. Moullem, S. Parker, N. Podhorszki, A. Shoshani and C. Silva (2007). Automation of network-based scientific workflows. In *Grid-Based Problem Solving Environments* (Patrick Gaffney and James Pool, Eds.). Vol. 239 of *IFIP International Federation for Information Processing*. Springer Boston. pp. 35–61.
- Wang, Jianwu, Daniel Crawl and Ilkay Altintas (2009a). Kepler + Hadoop: A general architecture facilitating data-intensive applications in scientific workflow systems. In *WORKS '09: Proceedings of the 4th Workshop on Workflows in Support of Large-Scale Science*. ACM New York, NY, USA. Portland, Oregon. pp. 1–8.
- Wang, Jianwu, Ilkay Altintas, Chad Berkley, Lucas Gilbert and Matthew B. Jones (2008). A high-level distributed execution framework for scientific workflows. *IEEE International Conference on eScience*. pp. 634–639.
- Wang, Jianwu, Ilkay Altintas, Parvizeh R. Hosseini, Derik Barseghian, Daniel Crawl, Chad Berkley and Matthew B. Jones (2009b). Accelerating parameter sweep workflows by utilizing ad-hoc network computing resources: An ecological example. In *Services, IEEE Congress on*. IEEE Computer Society. pp. 267–274.
- Wang, Jianwu, Prakashan Korambath, Seonah Kim, Scott Johnson, Kejian Jin, Daniel Crawl, Ilkay Altintas, Shava Smallen, Bill Labate and Kendall N. Houk (2010). Theoretical enzyme design using the kepler scientific workflows on the grid. In *Procedia Computer Science, ICCS 2010*. Vol. 1. pp. 1169 – 1178.
- Wilde, Michael, Ian Foster, Kamil Iskra, Pete Beckman, Zhao Zhang, Allan Espinosa, Mihael

- Hategan, Ben Clifford and Ioan Raicu (2009). Parallel scripting for applications at the petascale and beyond. *Computer*. **42**, 50–60.
- Wilson Jr., E. Bright (1991). *An Introduction to Scientific Research*. rev sub ed.. Dover Publications.
- Yang, Heng-Li and Jih-Hsin Tang (2003). A three-stage model of requirements elicitation for web-based information systems. *Industrial Management and Data Systems*. **103**(6), 398–409.
- Youn, Choonhan, Chaitan Baru, Karan Bhatia, Sandeep Chandra, Kai Lin, Ashraf Memon, Ghulam Memon and Dogan Seber (2007). Geongrid portal: design and implementations: Research articles. *Concurrency and Computation: Practice and Experience*. **19**(12), 1597–1607.
- Zhao, Jun (2007). *A conceptual model for e-science provenance*. PhD thesis. University of Manchester.
- Zhao, Jun, Carole Goble and Robert Stevens (2006a). An identity crisis in the life sciences. In *Provenance and Annotation of Data* (Luc Moreau and Ian Foster, Eds.). Vol. 4145 of *Lecture Notes in Computer Science*. pp. 254–269. Springer Berlin / Heidelberg.
- Zhao, Yong and Shiyong Lu (2008). A logic programming approach to scientific workflow provenance querying. In *Provenance and Annotation of Data and Processes: Second International Provenance and Annotation Workshop, IPAW 2008, Salt Lake City, UT, USA, June 17-18, 2008. Revised Selected Papers*. Springer-Verlag. Berlin, Heidelberg. pp. 31–44.
- Zhao, Yong, Michael Wilde and Ian Foster (2006b). Applying the virtual data provenance model. In *Provenance and Annotation of Data* (Luc Moreau and Ian Foster, Eds.). Vol. 4145 of *Lecture Notes in Computer Science*. pp. 148–161. Springer Berlin / Heidelberg.
- Zhao, Zhiming, Suresh Booms, Adam Belloum, Cees de Laat and Bob Hertzberger (2006c). VLE-WFBus: A scientific workflow bus for multi e-science domains. In *Proceedings of the Second IEEE International Conference on e-Science and Grid Computing*. E-SCIENCE '06. IEEE Computer Society. Washington, DC, USA. p. 11.
- Zhenning, Shangguan, Rui Huang, Paulo Pinheiro da Silva, Nicholas Del Rio and Deborah L. McGuinness (2009). Towards usable and interoperable workflow provenance: Empirical case studies using pml. In *Proceedings of the First International Workshop on the role of Semantic Web in Provenance Management (SWPM 2009), collocated with the 8th International Semantic Web Conference (ISWC-2009), Washington DC, USA, October 25, 2009*. Vol. 526 of *SWPM*.
- Zinn, Daniel, Shawn Bowers, Timothy McPhillips and Bertram Ludäscher (2009). Scientific workflow design with data assembly lines. In *Proceedings of the 4th Workshop on Workflows in Support of Large-Scale Science*. WORKS'09. ACM. New York, NY, USA. pp. 14:1–14:10.