

## BETH A. PLALE

Michael A and Laurie Burns McRobbie Bicentennial Professor of Computer Engineering  
Executive Director, Pervasive Technology Institute  
Director, Data To Insight Center  
Indiana University Bloomington

### EXECUTIVE SUMMARY

Beth A. Plale is the Michael A and Laurie Burns McRobbie Bicentennial Professor of Computer Engineering in the Luddy School of Informatics, Computing, and Engineering at Indiana University Bloomington. Plale is the Executive Director of the campus-wide Pervasive Technology Institute (PTI). Plale recently returned to IU from a 3-year term in Washington DC at the U.S. National Science Foundation (NSF).

**RESEARCH POLICY:** Dr. Plale served at the U.S. National Science Foundation over the 3-year period (2017 - 2021) in a cross-agency policy role in open science. Plale formulated a research agenda around open science for the agency and spearheaded the NSF Public Access Repository (PAR) 2.0 effort that secured funding and support inside agency to expand the NSF PAR to include metadata records about datasets and software in addition to records about publications. Plale regularly represented the agency in OSTP-led intra-agency work, with academic organizations (AAU/APLU), and with the science community.

**BUILDING SCIENCE COMMUNITY:** Dr. Plale is one of a dozen international leaders in research data sharing that founded the [Research Data Alliance](#) (RDA). The volunteer organization, now 10,000 members strong, was launched in 2013 to foster consensus building around the social and technical infrastructure to enable open share and re-use of data.

**INCLUSION ADVOCACY:** Plale lead the development of the 2021 Luddy School [Strategic Plan for Inclusive Excellence](#). The plan identifies a vision and objectives (with metrics) for attaining inclusive excellence. The plan is structured to give departments and units the flexibility to implement the objectives to fit their needs. Plale is one of the founders of the Indiana University Center of Excellence for Women in Technology (CEWIT), a highly successful campus effort that touches the lives of thousands of women faculty, staff, and students across campus. Earlier in her career, Dr. Plale co-founded the grassroots school-wide Women in Computing (WIC) for graduate and undergraduate women (2002 – 2007).

**ENTREPRENEURIAL ACTIVITY:** Plale is a founder of the HathiTrust Research Center (HTRC), an entrepreneurial activity of the Data To Insight Center (D2I) funded in its startup phase by the Lilly Endowment. HTRC is a successful organization today providing computational analytical access to the nearly 17.5 million volumes of digitized texts of which 60% are in copyright.

**ACTIVE RESEARCHER:** Dr. Plale has served in a principal investigator (PI/PD) or co-principal investigator (co-PI) role on 63 external grant funded awards totaling \$70,000,000. The Data To Insight Center has received external grant funding from public and private sources including the National Science Foundation (NSF), Department of Energy (DOE), Institute for Museum and Library Services (IMLS), National Aeronautics and Space Administration (NASA), the Alfred P. Sloan Foundation, the Andrew W. Mellon Foundation, and Microsoft.

Dr. Plale has authored over 150 scientific and scholarly publications in highly selective peer reviewed journals and conference proceedings. She has delivered nearly 100 keynote or invited talks and served as general chair and program chair for numerous leading peer-reviewed conferences. In the position of advisor, Dr. Plale has mentored, advised and graduated 15 PhD candidates and 7 Postdoctoral Fellows to date, and advised 24 master's theses. Dr. Plale was awarded the prestigious DOE Early Career award by the US Department of Energy.

**CAMPUS LEADER:** Dr. Plale served as the Associate Dean for Research (ADR) of the School of Informatics (2007 – 2009) where she initiated policies that improve research incentives. She chaired a campus-wide internal grant funding program (\$1M+/year) in 2010, and chaired the search committee for the Associate Vice President for Research at IU. She has led an Indiana University system-wide task force on digitization, and has served on numerous campus level committees on research compliance and integrity including the IU Restricted and Classified Task Force, the University Conflict of Interest Committee, and the Intellectual Property Council.

## EDUCATION AND APPOINTMENTS

---

### EDUCATION

Georgia Institute of Technology	College of Computing	Postdoctoral Fellow	1999-2001
State Univ of New York (SUNY) Binghamton	Computer Science	PhD	1998
Dissertation: <i>Software Approach to Hazard Detection using On-line Analysis of Safety Constraints</i>			
Chairs: Sudhir Aggarwal, SUNY Binghamton and Karsten Schwan, Georgia Tech			
Temple University	Computer & Info Science	M.S.	1992
Thesis: <i>Object Oriented System for Image Rendering</i>			
Advisor: Frank Friedman			
University of LaVerne	Business and Public Management	M.B.A.	1986
University of Southern Mississippi	Computer Science, Math Minor	BSc	1984

### APPOINTMENTS

2021 -	<i>Executive Director, Pervasive Technology Institute, IU</i>
--------	---

2020 -	<i>Michael A and Laurie Burns McRobbie Bicentennial Professor of Computer Engineering, IUB</i>
2018 -	<i>Professor, Dept. of Intelligent Systems Engineering, IUB</i>
2017 - 2020	<i>Science Advisor for Public Access, National Science Foundation (NSF)</i>
2015 -	<i>Affiliated Faculty, Ostrom Workshop on Political Theory and Policy Analysis</i>
2011 - 2018	<i>Professor, Dept. of Informatics and Computing, IU</i>
2014 - 2017	<i>Science Director, Pervasive Technology Institute (PTI), IU</i>
2019 -	<i>Director, Data to Insight Center (D2I), IU</i>
2008 - 2017	
2011 - 2017	<i>Director, HathiTrust Research Center (HTRC)</i>
2011 - 2014	<i>Managing Director, Pervasive Technology Institute, IU</i>
2007 - 2009	<i>Associate Dean of Research (ADR), School of Informatics and Computing, IUB</i>
2006 - 2009	<i>Director, Center for Data and Search Informatics, Indiana University</i>
2006 - 2011	<i>Associate Professor, Dept. of Informatics and Computing, IUB</i>
2002 - 2007	<i>Founder and mentor, Women in Computing (WIC), School of Informatics, IUB</i>
2001 - 2006	<i>Assistant Professor, Dept. of Computer Science, IUB</i>
1998 - 2001	<i>Postdoctoral Fellow, Georgia Institute of Technology, with: Karsten Schwan</i>
1996 - 1997	<i>Graduate Research Assistant, College of Computing, Georgia Institute of Technology</i>
1994 - 1996	<i>Adjunct Instructor, Georgia Perimeter College, Atlanta, Georgia</i>
1991 - 1994	<i>Graduate Research Assistant, Dept. Computer Science, SUNY Binghamton</i>
1993 - 1994	<i>Adjunct Instructor, Continuing Ed, SUNY Binghamton</i>
1989 - 1991	<i>Teaching Assistant, Dept. Computer and Information Science, Temple University</i>
1988 - 1989	<i>Lead Software Engineer, GTE Federal Systems, Westlake Village, California</i>
1986 - 1987	<i>Software Engineer, GTE Federal Systems, Westlake Village, California</i>
1984 - 1986	<i>Software Developer, Vitro Corporation, Oxnard, California</i>

## PROFESSIONAL MEMBERSHIIPS

Member, Institute of Electrical and Electronics Engineers (IEEE)  
 Senior Member, Association of Computing Machinery (ACM)  
 Member, American Association for the Advancement of Science

## BROAD IMPACT

---

Site Visit Chair: Science Foundation Ireland (SFI) \$100M INSIGHT project. 2021

NASEM Panel Participant: Accelerating Scientific Discovery through Intelligent and Automated Research Workflows, Board on Research Data and Information, The National Academies of Sciences, Engineering, and Medicine, March 2020

Interview: Open data for better science, Weijie Zhao, *National Science Review*, Volume 5, Issue 4, July 2018, pp 593-597, <https://doi.org/10.1093/nsr/nwy059>, Published:07 June 2018

Interview: Data science perspectives, State University of New York Binghamton, April 2019, <https://www.binghamton.edu/news/story/1780/points-of-view>

Interview: On the work at Hathi Trust Research Center, Big data, Aggregations and Text Mining, Open University, UK, Dec 2013, <http://kmi.open.ac.uk/news/article/18573>

## Boards / Leadership

Advisory Committee	NASA Earth Science Advisory Committee. A FACA committee of NASA	2021 -
Board of Directors	DONA Foundation, a non-profit organization in Geneva, Switzerland on internet governance	2021 -
Working Group	NOAA Data Archive and Access Requirements Working Group (subcommittee of NOAA Science Advisory Board)	2014 - 2018
External Advisory Board member	HathiTrust Research Center (HTRC)	2017 - 2021
External Advisory Board member	NSF funded Whole Tale project, University of Illinois	2016—2017
External Advisory Board member	Stanford University Center for Expanded Data Annotation and Retrieval (CEDAR)	2015 - 2017
Inaugural Chair	Research Data Alliance (RDA) Technical Advisory Board	2013 - 2015
Executive Steering Committee member	Pacific Rim Applications and Grid Middleware Assembly (PRAGMA)	2020 - 2015 - 2017
Steering Committee	Open Grid Forum	2005 - 2007

## GRANTS, CONTRACTS, and GIFTS (63)

AWARD	FUNDER
AI Institute: ICICLE: Intelligent Cyberinfrastructure with Computational Learning in the Environment, National Science Foundation (NSF), \$20,000,000, Plale co-PI and IU PI (IU portion \$1,999,000), 11/2021 - 10/2026	NSF
Data Capsule Appliance for Research Analysis of Restricted and Sensitive Data in Academic Libraries, Institute of Museum and Library Services (IMLS), \$320,546, 06/2017 - 05/19	IMLS
SEADTrain: Cloud-hosted Hands On Training Environment for Data Science where Data Frequently Restricted in Some Form, Microsoft Research, valued at \$60,000, 04/2017 - 03/2018	Microsoft
CC* Storage: Robust Persistent Identification of Data (RPID), Plale (PI) with Tufts and CNRI, National Science Foundation, \$199,047, 04/2017 - 03/2018. NSF Award 1659310.	NSF #1659310
Hardware Acquisition HathiTrust Research Center, Plale PI, HathiTrust Board of Governors, \$102,000, 07-2016 - 06/2017	HathiTrust
Bringing Visibility to Food Security Data Results: Harvests of PRAGMA and RDA, Plale PI with National Institute of Advanced Industrial Science and Technology (AIST) Japan, MacArthur Foundation through Rensselaer Polytechnic University, \$40,000, 01/2016 - 11/2016	MacArthur Foundation
Workshop on Data Quality in an Era of Big Data, Plale PI, National Science Foundation through CCC award to Midwest Big Data Hub, \$30,000, 06/2016 - 10/2016	NSF through CCC
Worksets and Data Capsules: Laying foundations for secure computation with copyrighted data in HathiTrust, Plale co-PI and IU PI. Andrew W. Mellon Foundation, \$1,170,000, 01/2016 - 12/2017	Andrew W. Mellon Foundation
HathiTrust Research Center, Plale PI with UIUC, HathiTrust, \$1,000,000, 07/2014 - 06/2018	HathiTrust

BD Hubs: Midwest: SEEDCorn: Sustainable Enabling Environment for Data Collaboration, Ed Seidel PI (UIUC), Plale co-PI. National Science Foundation, \$1,250,000, 10/2015 – 09/2018	NSF #1550320
Hazards SEES: Understanding cross-scale interactions of trade and policy to improve resilience to drought risk in Zambia, Kelly Caylor (Princeton) PI, IU portion Plale co-PI with Tom Evans (PI), National Science Foundation, \$1,862,385, 09/2014 – 08/2019.	NSF #1534544
WSC-Category 2 Collaborative: Impacts of Agricultural Decision Making and Adaptive Management on Food Security, Kelly Caylor (Princeton) PI. IU portion Plale co-PI with Tom Evans (PI) and Shahzeen Attari, \$1,862,385, National Science Foundation, 09/2014 – 08/2019	NSF #1360463
Research Data Alliance (RDA) Data Share, Plale PI/PD, with Rensselaer Polytechnic Institute, Alfred P. Sloan Foundation Grant G-2014-13746, \$748,000, 01/2015 - 12/2017	Alfred P. Sloan Foundation
Socio-Eco-Informatics: Enhancing Predictive Capability of Social Ecological Systems Research, Plale with David Leake and Xiaozhong Liu, Faculty Research Support Program, Indiana University, \$73,317, 03/2014 - 02/2016	Indiana University
Building the Research Data Alliance Community Through U.S. and International Engagement (RDA2), Plale co-PI, with Fran Berman (PI) Rensselaer Polytechnic Institute, \$6,022,637, National Science Foundation, 10/2013 - 09/2018	NSF #1349002
DataNet: Sustainable Environments Actionable Data (SEAD), National Science Foundation, Plale co-PI and IU PI, with UMich (PI), UIUC, National Science Foundation, total \$8,000,000, 10/2011 – 09/2016 (IU portion \$2.4M)	NSF #0940824
SAVI: PRAGMA - Enabling Scientific Expeditions and Infrastructure Experimentation for Pacific Rim Institutions and Researchers, IU Lead, with UC San Diego (PI), UWisconsin, UFlorida, \$5,693,064, National Science Foundation, 10/2012 - 09/2017	NSF #1234983
Workset Creation for Scholarly Analysis: Prototyping Project, Plale co-PI, with J. Stephen Downie (PI) UIUC, \$436,525, Andrew W. Mellon Foundation, 07/2013 - 09/2015	Andrew W. Mellon Foundation
Software Sustainability: an SI2 PI Workshop, Plale PI, with Doug Thain (Notre Dame) and Matt Jones (UC Santa Barbara), \$72,478, National Science Foundation, 01/2014 – 12/2014	NSF #1419131
CLIR/DLF Data Curation Fellowship, Council on Library and Information Resources, \$77,400, 07/2012 – 06/2014	CLIR
Building Trident Community, Beth Plale and Stacy Kowalczyk, \$50,000, Microsoft, 07/12 – 06/13	Microsoft
A Data Consortium: Coming Together Around Data, \$98,204, National Science Foundation, 08/12 – 07/13	NSF #1238168
SI2 SSE: Pipeline Framework for Ensemble Runs on Clouds, Plale PI with U. Miami, National Science Foundation, \$492,588, IU: \$292,863, 04/2012 – 03/2014	NSF #1148359 (IU)
Coming Together Around Data, A PI Project Meeting for DataNet and INTEROP, National Science Foundation, Plale PI, Scott Jensen co-PI, \$86,571, 08/2011 – 08/2012	NSF #1152946
Data Capsule for Non-Consumptive Research, Plale PI, with Atul Prakash, U Michigan, Alfred P. Sloan Foundation, \$606,161, 07/2011 – 12/2014	Alfred P. Sloan Foundation
HathiTrust Research Center: Computational Research on the HathiTrust Repository, Plale PI with Scott Poole, UIUC and others. HathiTrust Consortium, \$0*, 07/01/2011 – 06/30/2014 [*Funding of \$5M was conditional on outcome of	HathiTrust

class action lawsuit by Author's Guild and publishers against Google; suit was eventually thrown out of court.]

Microsoft Exploratory Research in Workflow and Provenance, Microsoft Research, gift through IU Foundation, \$125,000, 12/2010	Microsoft
EAGER: In-situ Archiving of Digital Scientific Data, Plale (PI) with Elinor Ostrom and Tom Evans, National Science Foundation, \$204,991, 10/2010 – 09/2012	NSF #1058452
Instant Karma: Applying a Proven Provenance Tool to NASA AMSR-E Data Production Stream, Plale (co-PI, IU lead) with U of Alabama Huntsville, NASA ACCESS 2009, (\$260,066 to IU), 10/2009 – 09/2011	NASA
netKarma: GENI Provenance Registry, Plale (PI) with Christopher Small of Global Research Network Operations Center (GRNOC), Global Environments for Networked Innovations (GENI) program, NSF through BBN Corp., approx. \$484,486, 09/2009 – 08/2012 [Plale .60 portion]	BBN
NextGen Weather Observation Networks, Plale (co-PI) with Oklahoma University, Federal Aviation Administration (FAA) through subcontract from Oklahoma University, \$247,500 total (\$90,000 to IU), 05/2009 – 03/2010	FAA
Towards Broad and Deep Community Deployment of LEAD, a Conference Grant, Plale (PI) with University of Alabama and University of Illinois Urbana Champaign, National Science Foundation, \$172,664 (\$94,664 to IU), 02/2009 – 09/2009	NSF
Pervasive Technology Institute, with Brad Wheeler, Craig Stewart, Geoffrey Fox and Fred Cate; Lilly Endowment, \$15,000,000, 01/2009 – 12/2014 [Data To Insight Center portion \$4.2M.]	Lilly Endowment
Cyberinfrastructure Software: Availability, Persistence, and Support Workshop, Plale (co-PI) with Brad Wheeler, Geoffrey Brown, Stacy Kowalczyk, and C. Stewart, National Science Foundation. \$131,691, 06/2008 – 05/2011 [Workshop]	NSF
III:Small: Assisted Emulation for Digital Preservation, PI: Geoffrey Brown, National Science Foundation, \$181,969 [minor role at project end]	NSF #1016967
LEAD-in-a-Box, Plale (PI) with Oklahoma University and University of North Carolina, Microsoft Corporation gift, \$100,000 (\$50,000 to IU), 01/2010 – 05/2011	Microsoft
ETF Grid Infrastructure Group: Providing System Management and Integration for the Teragrid, Plale (co-PI) with Craig Stewart, University of Chicago/National Science Foundation, \$2,519,000, 08/2005 – 07/2011 [Co-PI 2009-2011, minor role]	NSF
Future Grid: Experimental High-Performance Grid Test-bed, Plale (Sr. Investigator) with Geoffrey Fox (PI), Craig Stewart, et al., National Science Foundation, \$10,000,000, 10/09 - 09/13 [Minor role.]	NSF
Knowledge Discovery through Provenance Collection, Representation, and Use in the Lilly Science Grid (LSG), Plale (PI) with Dennis Groth and University of Manchester, UK, Eli Lilly Corp., \$220,000 (\$120,000 to IU), 03/2008 – 08/2009 [Plale .75 portion]	Eli Lilly Corp.
SDCI Data: New Toolkit for Provenance Collection, Publishing, and Use, Plale (PI) with David Leake and Yogesh Simmhan. National Science Foundation, \$437,954, 09/2007 -08/2009 [Plale .66 portion]	NSF #0721674
SDCI NMI: Improvement: Open Grid Computing Environments Software for Science Gateways Plale (Sr. Investigator) with Geoffrey Fox, Dennis Gannon, Marlon Pierce, and Nancy Wilkins-Diehr. National Science Foundation, NSF 07-503, \$1,698,347, 07/2007 - 06/2010 [Minor role]	NSF

CSR-CSI: An Adaptive Programming Framework for Data and Event Driven Computation, Plale (PI) with Dennis Gannon, National Science Foundation, \$300,000, 08/2007 - 07/2009 [Plale .50 portion]	NSF #0720580
Remote Digital Signatures and Signature Logging: A Proposal to Develop a Tool to Increase the Productivity of Physicians, Plale (PI), Midwest Proton Radiotherapy Institute, \$8,953, 03/2007-08/2008	Midwest Proton Radiotherapy Institute
Center for Research on Multicore Computing (CRMC), with Geoffrey Fox and Dennis Gannon, Microsoft, \$749,996, 07/2006 - 06/2009 [Minor role]	
Visual Search Tools using Existing Toolkits in LEAD, Plale (PI) with Sangmi Pallickara, National Science Foundation Research Experiences for Undergraduates (REU) associated with NSF ITR LEAD, \$14,000, 06/2006 - 05/2007	NSF
Chemical Informatics Cyberinfrastructure, Plale with Gary Wiggins, Geoffrey Fox (PI), Mookie Baik, Dennis Gannon, Randy Bramley, John Huffman, and Marlon Pierce, National Institutes of Health, \$731,750, 09/2005 - 07/2007 [minor role]	NIH
Science of Search: Data, Analytics, and Architectures Center (DSAAC), Plale (PI) with Dennis Gannon, National Science Foundation, \$10,000, 08/2006 - 07/2007	NSF #0630322
2005-2006 Distinguished Lecture Series, Plale with Kay Connelly and Butler University, DePauw University, Rose-Hulman Institute of Technology, Computing Research Association CRA-W, 2005 - 2006, \$10,000. [Seminar series]	Computing Research Association
MRI: Acquisition of a High-Speed, High Capacity Storage System to Support Scientific Computing: the Data Capacitor, Plale (co-PI) with Craig Stewart, Randy Bramley, Tom Hacker, and Cathy Pilachowski, National Science Foundation, 09/2005 - 08/2008, \$1,999,981. [Equipment grant]	NSF #0521433
DIALOGUE - Data Integration Applications: Linking Organisations to Gain Understanding and Experience", Plale (co-PI) with Malcolm Atkinson, University of Edinburgh, Joel Statz, Ohio State, Peter Brezeny, University of Vienna. Engineering and Physical Sciences Research Council (EPSRC), United Kingdom, 01/2005 - 01/2007. [Travel grant, workshop series]	EPSRC, United Kingdom
Expanding Science by Unified Access to Radar Data and Algorithms, Plale (PI), Microsoft Research Equipment Grant, \$35,000 equip + \$53,000 gift, 09/2004 - 01/2007	Microsoft
DOE Early Career: Time-based Data Streams: Fundamental Concepts for Data Resource for Streams, Plale (PI), Department of Energy, 09/2004 - 08/2007, \$298,343	DOE
An Itanium Environment for Grid-Based Data Mining, Plale (co-PI) with Dennis Groth, Hewlett Packard Philanthropy and Education Equipment Grant, 09/2004 - 09/2005, \$60,000. [Equipment grant]	Hewlett Packard
ITR Collaborative Research: Linked Environments for Atmospheric Discovery (LEAD), with Dennis Gannon, Oklahoma University, University of Alabama Huntsville, UCAR, University of Illinois Urbana Champaign, Howard University, Colorado State University, and Millersville University. National Science Foundation, 10/2003 - 9/2009, total approx. \$11,000,000 (\$2,234,101 IU). [Plale portion .50]	NSF #0331480 (IU)
Large-scale Streaming Data in Scientific Portals, Plale (PI), National Science Foundation through NCSA/University of Illinois, Scientific Portals Expedition, 10/2003 - 3/2005, \$200,000	NSF
Collaborative Proposal: Middleware for Grid Portal Development, Plale with Dennis Gannon, Marlon Pierce (PI), and Geoffrey Fox, University of Texas Austin,	NSF

Argonne National Labs, National Science Foundation, 09/2003 - 08/2006, \$868,803 to IU. [IU funds roughly evenly distributed]	
A Research Infrastructure for Collaborative, High-Performance Grid Applications, Plale (co-PI) with David Wise (PI), Andrew Lumsdaine, Geoffrey Fox, Randy Bramley, NSF, 09/2002 - 08/2007, \$1,311,875 plus IU match of \$803,709. [Equipment grant]	NSF
Crane-IU-Purdue Knowledge Projection for Fleet Maintenance" Plale with Randy Bramley, Geoffrey Fox, David Leake, and Rik McMullen (PI), National Science Foundation, 6/15/02 - 9/27/05, \$500,000 [funds roughly evenly distributed]	NSF
Middleware Technology to Support Science Portals: A Gateway to the Grid, Plale (co-PI) with Dennis Gannon, Department of Energy, 01/2001 - 07/2007, \$1,045,141. [Plale .33 portion]	DOE
ITR/SY Collaborative Research: A Unified Relational Approach to Grid Information Services, Plale (co-PI) with Northwestern University, National Science Foundation, 09/2001 - 08/2005, IU: \$234,702	NSF #0128390
Dynamic Querying of Large-scale Streaming Data, Plale (PI), National Science Foundation through NCSA/University of Illinois, \$62,000, 10/2001 - 09/2002	NSF
POWRE: Applying Database Techniques to Management of Large Data Flows in Interactive Scientific Simulations, Plale (PI), National Science Foundation, \$75,005, 1999 - 2002	NSF #9973834
Profiling the Performance of SMP Servers, Plale with Karsten Schwan, Intel Corporation, \$32,000, 1998. [equipment grant]	Intel

## PEER REVIEWED JOURNAL or CONFERENCE PROCEEDINGS (152) \_\_\_\_\_

2021	Beth Plale, Tanu Malik, Line Pouchard (2021). Reproducibility Practice in High Performance Computing: Community Survey Results, in <i>Computing in Science and Engineering</i> , 23(05), pp. 55-60. DOI: <a href="https://doi.org/10.1109/MCSE.2021.3096678">10.1109/MCSE.2021.3096678</a>
2021	Sachith Withana, Kshitij Mehta, Matthew Wolf, Beth Plale (2021). Towards System for Knowledge Representation of Campaign Experimentation, <i>2021 IEEE 17th Int'l Conf on eScience (eScience)</i> , 2021, pp. 257-258, doi: <a href="https://doi.org/10.1109/eScience51609.2021.00049">10.1109/eScience51609.2021.00049</a>
2021	Sachith Withana, Kshitij Mehta, Matthew Wolf, Beth Plale (2021). Campaign Knowledge Network: Building Knowledge for Campaign Efficiency. arXiv:2112.03435
2021	Kouper, Inna, Scheidt, Lois A. and Plale, Beth A. (2021). Fostering Interdisciplinary Data Cultures through Early Career Development: The RDA/US Data Share Fellowship. <i>Data Science Journal</i> , 20(1), DOI: <a href="https://doi.org/10.5334/dsj-2021-002">10.5334/dsj-2021-002</a>
2020	Zong Peng and Beth Plale (2020). Reliable Access to Massive Restricted Texts: Experience-based Evaluation, <i>Concurrency and Computation: Practice and Experience</i> , Wiley, 32:e5225. DOI: <a href="https://doi.org/10.1002/cpe.5255">10.1002/cpe.5255</a>
2019	Venice Juanillas, Alexis Dereeper, Nicolas Beaume, Gaetan Droc, Joshua Dizon, John Robert Mendoza, Jon Peter Perdon, Locedie Mansueto, Lindsay Triplett, Jillian Lang, Gabriel Zhou, Kunalan Ratharanjan, Beth Plale, Jason Haga, Jan E Leach, Manuel Ruiz, Michael Thomson, Nikolai Alexandrov, Pierre Larmande, Tobias Kretzschmar, Ramil P Mauleon (2019), Rice Galaxy: an open resource for plant science, <i>GigaScience</i> , 8(5), DOI: <a href="https://doi.org/10.1093/gigascience/giz156">10.1093/gigascience/giz156</a>

- Beth Plale, Eleanor Dickson, Inna Kouper, Samitha Liyanage, Yu Ma, Robert McDonald, John Walsh and Sachith Withana (2019), Safe Open Science for Restricted Data, *Data and Information Management*, De Gruyter, Vol. 3(1), 50-60. DOI: [10.2478/dim-2019-0005](https://doi.org/10.2478/dim-2019-0005)
- Yolanda Gil, Suzanne A. Pierce, Hassan Babaie, Arindam Banerjee, Kirk Borne, Gary Bust, Michelle Cheatham, Imme Ebert-Uphoff, Carla Gomes, Mary Hill, John Horel, Leslie Hsu, Jim Kinter, Craig Knoblock, David Krum, Vipin Kumar, Pierre Lermusiaux, Yan Liu, Chris North, Victor Pankratius, Shanan Peters, Beth Plale, Allen Pope, Sai Ravela, Juan Restrepo, Aaron Ridley, Hanan Samet, Shashi Shekhar (2019). Intelligent Systems for Geosciences: An Essential Research Agenda, *Communications of the ACM*, 62(1), pp. 76-84. DOI: [10.1145/3192335](https://doi.org/10.1145/3192335)
- 2018 Isuru Suriarachchi, Sachith Withana, and Beth Plale (2018), Big Provenance Stream Processing for Data Intensive Computations, *13th IEEE Int'l Conf on e-Science*, IEEE Computer Society, Amsterdam, Netherlands.
- Sachith Withana, Inna Kouper, Beth Plale (2018). Data Capsule for Restricted Data in Libraries, Extended Abstract. *Workshop on Cyberinfrastructure and Machine Learning for Digital Libraries and Archives, in conjunction with Joint Conference on Digital Libraries 2018*, Fort Worth, TX
- 2017 Mostafa Elag, Praveen Kumar, Luigi Marini, James D. Myers, Margaret Hedstrom, and Beth Plale (2017). Identification and characterization of information-networks in long-tail data collections, *Environmental Modelling & Software*, Elsevier, vol. 94, pp. 100-111. DOI:10.1016/j.envsoft.2017.03.032
- Beth Plale and Inna Kouper (2017). The Centrality of Data: Data Lifecycle and Data Pipelines, In *Data Analytics in Intelligent Transportation Systems*, M.A. Chowdhury, A. Apon, and K. Dey (Eds.) Elsevier Inc. Cambridge, MA. DOI:10.1016/B978-0-12-809715-1.00004-3
- Guangchen Ruan, Paul Hanson, Hilary A. Dugan, Beth Plale (2017). Mining Lake time series using symbolic representation, *Ecological Informatics*, Elsevier, vol. 39, pp. 10-22. DOI:10.1016/j.econif.2017.03.001
- Jaimie Murdock, Jacob Jett, Tim Cole, Yu Ma, J. Stephen Downie, Beth Plale (2017). Towards Publishing Secure Capsule-based Analysis (short paper), *17th ACM-IEEE-CS Joint Conference on Digital Libraries*, Toronto, CA Jun 2017
- 2016 J. Stephen Downie, Mike Furlough, Robert McDonald, Beth Namachchivaya, Beth A. Plale, John Unsworth (2016). The HathiTrust Research Center: Exploring the Full-Text Frontier, *Educause Review*, 51(3).
- Jiaan Zeng and Beth Plale (2016). Argus: A Multi-tenancy NoSQL store with workload-award resource reservation, *Parallel Computing*, Elsevier, vol. 58, pp. 76-89. DOI:10.1016/j.parco.2016.06.003
- S. Kowalczyk, Y. Sun, Z. Peng, B. Plale, C. Willis, J. Zeng, M. Pathirage, S. Liyanage, A. Todd, G. Ruan (2016). Big Data at Scale for Digital Humanities: An Architecture for the HathiTrust Research Center. In *Big Data: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*, IGI Global. DOI: 10.4018/978-1-4666-4699-5.ch011
- Beth Plale, Inna Kouper, Allison Goodwell, and Isuru Suriarachchi (2016). Trust Threads: Minimal Provenance for Data Publishing and Reuse, *Big Data is Not a Monolith: Policies, Practices and Problems*, Cassidy R. Sugimoto, Hamid Ekbia, and Michael Mattioli, Eds., MIT Press
- Peng Chen, Tom Evans, Michael Frisby, Eduardo Izquierdo, and Beth Plale (2016), A Hybrid Approach to Population Construction for Agricultural Agent-Based Simulation, *11th IEEE Int'l Conf. on e-Science*, IEEE Computer Society, Baltimore, MD, Oct 2016
- Isuru Suriarachchi and Beth Plale (2016), Crossing Analytics Systems: A Case for Integrated Provenance in Data Lakes, *IEEE 12th Int'l Conference on e-Science*, Baltimore, Oct 2016

Guangchen Ruan and Beth Plale (2016). Horne: Random Access Big Data Analytics, *IEEE Conf on Cluster Computing*, IEEE Computer Society, Taipei, Taiwan, Sep 2016

Isuru Suriarachchi and Beth Plale (2016). Provenance as Essential Infrastructure for Data Lakes, Short Paper. *6<sup>th</sup> Int'l Provenance and Annotation Workshop (IPAW)*, McLean Virginia, Jun 2016

Peng Chen, Tom Evans, and Beth Plale (2016), Analysis of Memory Constrained Live Provenance, *6<sup>th</sup> Int'l Provenance and Annotation Workshop (IPAW)*, McLean, Virginia, Jun 2016

Milinda Pathirage, Julian Hyde, Yi Pan, and Beth Plale (2016). SamzaSQL: Scalable Fast Data Management with Streaming SQL, *IEEE Int'l Workshop on High-Performance Big Data Computing, as part of IEEE Int'l Parallel and Distributed Processing Symp Workshops (IPDPSW 2016)*, Chicago, IL, May 2016.

Jiaan Zeng and Beth Plale (2016). KVLIGHT: A Lightweight Key-Value Store for Distributed Access in Cloud, *16th IEEE/ACM Int'l Symp on Cluster, Cloud and Grid Computing (CCGrid)*, Cartagena, Columbia, May 2016 (20% acceptance)

2015 Provenance and Annotation of Data and Processes: 5th International Provenance and Annotation Workshop, IPAW 2014 (2015). Cologne, Germany, June 9-13, 2014. Revised selected papers, Bertram Ludaescher and Beth Plale, Eds, Lecture Notes in Computer Science 8628, Springer, 2015

Isuru Suriarachchi, Quan Zhou and Beth Plale (2015). Komadu: A Capture and Visualization System for Scientific Data Provenance. *Journal of Open Research Software* 3(1):e4, DOI:<http://dx.doi.org/10.5334/jors.bq>

Jiaan Zeng and Beth Plale (2015). Workload-aware Resource Reservation for Multi-Tenant NoSQL, *IEEE Cluster 2015*, pp. 32-41, DOI:10.1109/CLUSTER.2015.14. Best paper candidate.

Jiaan Zeng, Beth Plale (2015). Short paper: Towards Building a Lightweight Key-Value Store on Parallel File System, *17th IEEE Int'l Conference on Cluster (CLUSTER)*, 2015

Peng Chen and Beth Plale (2015). ProvErr: System Level Statistical Fault Diagnosis using Dependency Model, *15th IEEE/ACM Int'l Symp on Cluster, Cloud and Grid Computing*, IEEE, pp. 525-534 10.1109/CCGrid.2015.86 (25.7% acceptance)

2014 You-Wei Cheah and Beth Plale (2014). Provenance Quality Assessment Methodology and Framework, *ACM Journal of Data and Information Quality, Special issue on Provenance, Data and Information Quality*, Vol 5(3).

Beth Plale (2014). Synthesis of Working Group and Interest Group Activity One Year into the Research Data Alliance, *D-Lib Magazine* (2014) DOI 10.1045/january2014-plale

Yuan Luo, Beth Plale, Zhenhua Guo, Wilfred W. Li, Judy Qiu, Yiming Sun (2014). Hierarchical MapReduce: Towards Simplified Cross-Domain Processing, *Concurrency and Computation: Practice and Experience*, on-line publication 24 Sep 2012 10.1002/cpe.2929. Issue publication Vol 26(4), pp. 878-893, 25 Mar 2014.

Abhirup Chakraborty, Milinda Pathirage, Isuru Suriarachchi, Kavitha Chandrasekar, Craig Mattocks, Beth Plale (2014). Executing Storm Surge Ensembles on PAAS Cloud, *Cloud Computing for Data-Intensive Applications*, X. Li and J. Qiu, Eds. Springer, pp. 257-276.

Guangchen Ruan, Hui Zhang, Beth Plale (2014). Parallel and Quantitative Sequential Pattern Mining for Large-Scale Interval-based Temporal Data, *5th SC Workshop on Big Data Analytics: Challenges and Opportunities*, New Orleans, LA, Nov 2014

Quan (Gabriel) Zhou, Devarshi Ghoshal, Beth Plale (2014). Study of Usefulness of Middleware-Only Provenance, *IEEE 10<sup>th</sup> Int'l Conf on e-Science*, IEEE, pp. 215-222, [10.1109/eScience.2014.49](https://doi.org/10.1109/eScience.2014.49)

Zong Peng, Miao Chen, Stacy Kowalczyk, Beth Plale (2014). Short Paper: Author Gender Metadata Augmentation of HathiTrust Digital Library, Connecting Collections, Cultures, and

Communities, *77<sup>th</sup> Annual Meeting of American Society for Information Science and Technology (ASIS&T)*, Seattle WA, 2014.

Guangchen Ruan, Hui Zhang, Beth Plale (2014). Parallel and Quantitative Sequential Pattern Mining for Large-scale Interval-based Temporal Data, *Workshop in Advances in Software and Hardware for Big Data to Knowledge Discovery, IEEE Int'l Conf on Big Data*, IEEE, pp. 32-39, Oct 2014

Jiaan Zeng and Beth Plale (2014). Multi-tenant Fair Share in No-SQL Data Stores, *IEEE Cluster 2014*, Madrid, Spain Sep 2014. (23.5% acceptance)

Jiaan Zeng, Guangchen Ruan, Alexander Crowell, Atul Prakash, Beth Plale (2014). Cloud Computing Data Capsules for Non-consumptive Use of Texts, *5<sup>th</sup> Workshop on Scientific Cloud Computing, co-located with ACM High Performance Distributed Computing (HPDC)*, Vancouver, CA, Jun 2014. Best paper runner up.

Guangchen Ruan, H. Zhang, Eric Wernert, Beth Plale (2014). TextRWeb: Large-Scale Text Analytics with R on the Web. *Proceedings of XSEDE '14*. Atlanta, GA, July 2014.

Devarshi Ghoshal, Arun Chauhan, Beth Plale (2014). Regenerating and Quantifying Quality of Benchmarking Data using Static and Dynamic Provenance, *Int'l Provenance and Annotation Workshop (IPAW 2014), Lecture Notes in Computer Science, Vol. 8628*, pp 56-67

Mostafa Elag, Praveen Kumar, Margaret Hedstrom, James Myers, Beth Plale, Luigi Marini, and Robert McDonald (2014). Characterization of Emergent Data Networks Among Long-Tail Data, *European Geosciences Union General Assembly 2014*, Vienna, Austria, Apr 2014.

2013 S. Jensen, B. Plale, M. Aktas, Y. Luo, P. Chen, and H. Conover (2013). Provenance Capture and Use in a Satellite Data Processing Pipeline, *IEEE Transactions on Geoscience and Remote Sensing, special issue on Data Provenance*, (51)11, pp. 5090-5097. DOI 10.1109/TGRS.2013.2266929

Mehmet Aktas, Beth Plale, David Leake, Nirmal K. Mukhi (2013). Unmanaged Workflows: Their Provenance and Use, Chapter 3, *Data Provenance and Data Management in eScience, Studies in Computational Intelligence series*, Q. Bai, Q. Liu eds., Springer, Vol 426, pp. 59-81.

Peng Chen, Beth Plale, Mehmet Aktas (2013). Temporal Representation for Mining Scientific Data Provenance, *Future Generation of Computer Systems*, Elsevier, Vol 36, pp. 363-378

Mehmet Aktas, Beth Plale, David Leake, Nirmal K. Mukhi (2013). Unmanaged Workflows: Their Provenance and Use, Q. Bai, Q. Liu eds. *Data Provenance and Data Management in eScience, Studies in Computational Intelligence series*, Springer, Vol 426, pp. 59-81

Beth Plale, Robert H. McDonald, Kavitha Chandrasekar, Inna Kouper, Stacy Konkiel, Margaret Hedstrom, James Myers, Praveen Kumar (2013). SEAD Virtual Archive: Building a Federation of Institutional Repositories for Long-Term Data Preservation in Sustainability Science, *Int'l Journal of Digital Curation*, 8(2), pp 172-180. DOI 10.2218/ijdc.v8i2.281

Peng Chen, Beth Plale, and Tom Evans (2013). Dependency Provenance in Agent Based Modeling, *9<sup>th</sup> IEEE Int'l Conf on e-Science*, IEEE Computer Society, Beijing, China, Oct 2013.

Jiaan Zeng and Beth Plale (2013). Data pipeline in MapReduce, *9<sup>th</sup> IEEE Int'l Conf on e-Science*, IEEE Computer Society, Beijing, China, Oct 2013.

Maryam Rahnemoonfar and Beth Plale (2013). Automatic Performance Evaluation of Dewarping Methods in Large Scale Digitization of Historical Documents, *Joint Conf on Digital Libraries*, pp 331-334, ACM, New York, NY. DOI: 10.1145/2467696.2467744

Miao Chen, Uma Pavalanathan, Scott Jensen, and Beth Plale (2013). Modeling Heterogeneous Data Resources for Social-Ecological Research: A Data-Centric Perspective, *Joint Conf on Digital Libraries (JCDL 2013)*, Indianapolis, Jul 2013.

- Abhirup Chakraborty, Milinda Pathirage, Isuru Suriarachchi, Kavitha Chandrasekar, Craig Mattocks and Beth Plale (2013). Storm Surge Simulation and Load Balancing in Azure Cloud, *High Performance Computing Symp*, Society for Computer Simulation International San Diego, CA, USA
- Guangchen Ruan, Hui Zhang, and Beth Plale (2013). Exploiting MapReduce and Data Compression for Data-intensive Applications, *XSEDE 2013*, San Diego, Jul 2013.
- You-Wei Cheah, Richard Canon, Beth Plale, and Lavanya Ramakrishnan (2013). Milieu: Lightweight and Configurable Big Data Provenance for Science, Research Session 3 – Big Data Mining, *2013 IEEE Second Int'l Congress on Big Data (BigData 2013)*, Santa Clara, CA, Jun 2013.
- Devarshi Ghoshal and Beth Plale (2013). Provenance from Log Files: a BigData Problem, *1<sup>st</sup> Int'l Workshop on Managing and Querying Provenance Data at Scale Held in conjunction with EDBT/ICDT 2013*, Mar 2013
- B. Plale, P. Kumar, J. Myers, M. Hedstrom, R. McDonald, S. Konkiel, and K. Chandrasekar (2013). SEAD Virtual Archive: Building a Federation of Institutional Repositories for Long Term Data Preservation, *8<sup>th</sup> Int'l Digital Curation Conf*, Amsterdam, Netherlands. <http://hdl.handle.net/2022/15247>
- 2012 Eran Chinthaka Withana and Beth Plale (2012). Sigiri: Uniform Resource Abstraction for Grids and Clouds, *Concurrency and Computation, Practice and Experience*, Vol 24, Issue 18, pp. 2362-2380, 2012.
- Peng Chen, Beth Plale, You-Wei Cheah, Devarshi Ghoshal, Scott Jensen, and Yuan Luo (2012). Visualization of Network Provenance Data, *Workshop on Massive Data Analytics on Scalable Systems (DataMASS), co-located with High Performance Computing Conf*, Pune India, 2012.
- You-Wei Cheah and Beth Plale (2012). Provenance Analysis: Towards Quality Provenance, *8<sup>th</sup> IEEE Int'l Conf on e-Science*, IEEE Computer Society 2012, Chicago, IL.
- Beth Plale (2012). Managing the long tail of science: data and communities. *1<sup>st</sup> Conf of Extreme Science and Engineering Discovery Environment (XSEDE'12)*, ACM, article 68. [10.1145/2335755.2335866](https://doi.org/10.1145/2335755.2335866)
- Peng Chen and Beth Plale (2012). Poster Abstract: Visualizing Large-scale Data Provenance, *IEEE/ACM Supercomputing*, Salt Lake City, pp. 1385-1386, Salt Lake City, UT 2012. [10.1109/SC.Companion.2012.205](https://doi.org/10.1109/SC.Companion.2012.205)
- Scott Jensen, Beth Plale, Xiaozhong Liu, Miao Chen, David Leake, and Julie England (2012).
- Peng Chen, Beth Plale, and Mehmet S. Aktas (2012). Temporal Representation for Scientific Data Provenance, *8<sup>th</sup> IEEE Int'l Conf on e-Science*, IEEE Computer Society, [10.1109/eScience.2012.6404477](https://doi.org/10.1109/eScience.2012.6404477)
- Beth Plale, Eran Chinthaka Withana, Chathura Herath, Kavitha Chandrasekar, and Yuan Luo (2012). Effectiveness of Hybrid Workflow Systems for Computational Science, *Int'l Conf on Computational Science (ICCS), Procedia Computer Science, Elsevier*, Vol 9, pp. 508-517
- Kavitha Chandrasekar, Milinda Pathirage, Samindra Wijeratne, Craig Mattocks, and Beth Plale (2012). Middleware Alternatives for Storm Surge Predictions in Windows Azure, *3<sup>rd</sup> Workshop on Scientific Cloud Computing*, pp 3-12, ACM, NY, NY [10.1145/2287036.2287040](https://doi.org/10.1145/2287036.2287040)
- Scott Jensen, Miao Chen, Xiaozhong Lin, Beth Plale, David Leake (2012). Short Paper: Mining Classifications from Social-Ecological Databases, *75<sup>th</sup> Annual Meeting of American Society for Information Science and Technology (ASIS&T 2012)*, Baltimore, MD, 2012
- Miao Chen and Beth Plale (2012). Short Paper: From Metadata to Ontology Representation: A Case of Converting Severe Weather Forecast Metadata to an Ontology, *75<sup>th</sup> Annual Meeting*

- of American Society for Information Science and Technology (ASIS&T 2012)*, Baltimore, MD, 2012
- 2011 Beth Plale, Bin Cao, Chathura Herath, and Yiming Sun (2011). Data Provenance for Preservation of Digital Geoscience Data, *Societal Challenges and Geoinformatics*, A. Krishna Sinha, David Arctur, Ian Jackson, and Linda Gundersen, Eds., in *Geological Society of America (GSA)*, Special Paper 482, 01 Nov 2011, ISBN 9780813724829
- Yogesh Simmhan and Beth Plale (2011). Using Provenance for Personalized Quality Ranking of Scientific Datasets, Artem Chebotko, Yogesh Simmhan and Paolo Missier, eds. *Int'l Journal of Computers and Their Applications: Special Issue on Scientific Workflows, Provenance and Their Applications*, 18(3), pp. 180-196
- Luc Moreau, Ben Clifford, Juliana Freire, Joe Futrelle, Yolanda Gil, Paul Groth, Natalia Kwasnikowska, Simon Miles, Paolo Missier, James Myers, Beth Plale, Yogesh Simmhan, Eric Stephan, Jan Van den Bussche, The Open Provenance Model Core Specification (V1.1), *Future Gener. Comput. Syst.*, Elsevier Science Publishers B. V., Amsterdam Netherlands, Vol. 27, No. 6, June, 2011, pp. 743-756, DOI:10.1016/j.future.2010.07.005
- P. Crippa, G. El Afandi, B.A. Plale, and S.C. Pryor (2011). Short Paper: Understanding the effects of boundary layer and synoptic meteorology on new particle formation based on WRF simulations and measurements in Southern Indiana, *American Geophysical Union Fall 2011 meeting*, 2011
- Chathura Herath and Beth Plale (2011). Programming abstraction for resource aware stream processing for scientific workflows, D3Science Workshop, in workshop proceedings of 9<sup>th</sup> *IEEE Int'l Conf on e-Science*, Dec 2011
- S. Jensen, M. Cox, D. Bender, M. Chen, J. England, B. Plale, and D. Leake (2011). Spatial Data in an Ontology for Research on Forest Resources, *COSIT11 Workshop Ontology of Spatial Thinking and Reasoning: Multidisciplinary Reconciliation*, Belfast, Maine, Sep 12, 2011
- You-Wei Cheah, Beth Plale, Joey Kendall-Morwick, David Leake, and Lavanya Ramakrishnan (2011). A Noisy 10GB Provenance Database, *2nd Int'l Workshop on Traceability and Compliance of Semi-Structured Processes (TC4SP2011)*, co-located with *Business Process Management (BPM 2011)*, Clermont-Ferrand, France, Aug 2011
- Yuan Luo, Zenhua Guo, Yiming Sun, Beth Plale, Judy Qiu, and Wilfred W. Li (2011). Hierarchical Framework for Cross-domain MapReduce Execution, *Emerging Computational Methods for Life Sciences*, co-located with *ACM High Performance Distributed Computing*, Chicago, IL, Jun 2011
- Beth Plale (2011). Challenges and Opportunities of Workflow Systems in Environmental Research, invited, *Water Information Research and Development Alliance (WIRADA) Science Symposium*, Melbourne, AU, Aug 2011
- Dean Guo, Beth Plale, Leon Welicki, and Eran Chinthaka (2011). Scientific Workflow Challenges, *Water Information Research and Development Alliance (WIRADA) Science Symposium*, Melbourne, AU, Aug 2011
- Eran Chinthaka Withana, Beth Plale, and Craig Mattocks (2011). Towards Enabling Mid-Scale Geoscience Experiments Through Microsoft Trident and Windows Azure, *Microsoft Cloud Futures Workshop*, Jun 2011
- 2010 Daniel S. Katz, Scott Callaghan, Robert Harkness, Shantenu Jha, Krzysztof Kurowski, Steven Manos, Sudhakar Pamidighantam, Marlon Pierce, Beth Plale, Carol Song, and John Towns (2010). Science on the Teragrid, *Computational Methods in Science and Technology, Special Issue*, Polish Academy of Sciences, pp. 81-97.
- Scott Jensen and Beth Plale (2010). Trading Consistency for Scalability in Scientific Metadata, *6<sup>th</sup> Int'l IEEE Conf on e-Science*, Brisbane, AU, 7-10 Dec 2010.

- Eran Chinthaka and Beth Plale (2010). Usage Patterns to Provision for Time Critical Scientific Experimentation in Clouds, *2<sup>nd</sup> IEEE Int'l Conf on Cloud Computing Technology and Science (CloudCom 2010)*, Indianapolis, IN Nov 2010.
- Chathura Herath and Beth Plale (2010). Streamflow - Programming Model for Data Streaming in Scientific Workflows, *Proceedings of the 10th IEEE/ACM Int'l Symp on Cluster, Cloud, and Grid Computing (CCGrid 2010)*, Melbourne Australia, May 2010.
- Lavanya Ramakrishnan, Dennis Gannon, and Beth Plale (2010). WORKEM: Representing and Emulating Distributed Scientific Workflow Execution State, *Proceedings of the 10th IEEE/ACM Int'l Symp on Cluster, Cloud and Grid Computing (CCGrid 2010)*, Melbourne Australia, May 2010. 10.1109/CCGRID.2010.89
- Lavanya Ramakrishnan and Beth Plale (2010). Multidimensional Classification Model for Scientific Workflow Characteristics, *1st Int'l Workshop on Workflow Approaches for New Data-Centric Science, co-located with ACM SIGMOD International Conf on Management of Data*, Jun 2010.
- Eran Chinthaka Withana, Beth Plale, Roger Barga, and Nelson Araujo (2010). Versioning for Workflow Evolution, *Data Intensive Distributed Computing Workshop (DIDC), co-located with High Performance Distributed Computing*, Chicago, Jun 2010
- 2009 Bin Cao, Beth Plale, Girish Subramanian, Paolo Missier, Carole Goble, and Yogesh Simmhan (2009). Semantically Annotated Provenance in the Life Science Grid, *1<sup>st</sup> Int'l Workshop on the role of Semantic Web in Provenance Management*, co-located with 8th Int'l Semantic Web Conference, Oct 2009, Washington D.C., USA.
- Bin Cao, Beth Plale, Girish Subramanian, Ed Robertson, and Yogesh Simmhan (2009). Provenance Information Model of Karma Version 3, *IEEE 2009 3<sup>rd</sup> Int'l Workshop on Scientific Workflows (SWF'09), Proceedings of 2009 Congress on Services*, IEEE Computer Society, pp. 348-351, Jul 2009, <http://doi.ieeecomputersociety.org/10.1109/SERVICES-I.2009.54>
- Eran Chinthaka Withana, Jalyia Ekanayke, David Leake, and Beth Plale (2009). CBR Based Workflow Composition Assistant, *IEEE 2009 3<sup>rd</sup> Int'l Workshop on Scientific Workflows (SWF'09), Proceedings of 2009 Congress on Services*, IEEE Computer Society, pp. 348-351, Jul 2009. <http://doi.ieeecomputersociety.org/10.1109/SERVICES-I.2009.51>
- Srinath Perera, Suresh Marru, Thilina Gunarathne, Dennis Gannon, and Beth Plale (2009). Application of Management Frameworks: A Case Study on Managing Workflow-related Systems, *IEEE Int'l Conf on Web Services (ICWS)*, Los Angeles, Jul 2009. DOI 10.1109/ICWS.2009.52
- Kathleen M. Baker, Beth Plale, Ilya Zaslavsky, and Suresh Marru (2009). Towards Cyberinfrastructure for Multi-scale Crop Disease Early Warning Systems, *World Congress on Computers in Agriculture*, Reno, NV Jun 2009
- Beth Plale, Dennis Groth, Bin Cao, Girish Subramanian, Carole Goble, and Paolo Missier (2009). Short Paper: Knowledge Discovery through Provenance Collection, Representation, and Use in the Life Science Grid, *Sixth Int'l Conference on Data Integration in the Life Sciences (DILS)*, Manchester, UK, 2009
- 2008 Nithya Vijayakumar and Beth Plale (2008). Missing Event Prediction in Sensor Data Streams Using Kalman Filters. *Book chapter in Knowledge Discovery from Sensor Data*, Eds. A. R. Ganguly, J. Gama, O. A. Omitaomu, M. Gaber and R. R. Vatsavai, Taylor and Francis/CRC Press pp. 149-170
- Yogesh Simmhan, Beth Plale, and Dennis Gannon (2008). Query Capabilities of the Karma Provenance Framework, *Concurrency and Computation: Practice and Experience*, John Wiley and Sons, Vol 20(5), pp. 441-451

- Xiang Li, Beth Plale, Nithya Vijayakumar, Rahul Ramachandran, Sara Graves, and Helen Conover (2008). Real-time Storm Detection and Weather Forecast Activation through Data Mining and Events Processing, *Earth Science Informatics*, H.A. Babaie, Ed., Springer Berlin/Heidelberg, Vol 1(2), pp. 49-57. DOI 10.1007/s12145-008-0010-7
- Yogesh Simmhan, Beth Plale, and Dennis Gannon (2008). Karma2: Provenance Management for Data Driven Workflows, Extended and invited from ICWS 2006. *Int'l Journal of Web Services Research*, IGI Publishing, Vol 5 (2), pp. 1-22
- Scott Jensen and Beth Plale (2008). Schema-Independent and Schema-Friendly Scientific Metadata Management, *Proceedings 4th Int'l IEEE Conf on e-Science*, IEEE Computer Society, pp. 428-429, <http://doi.ieeecomputersociety.org/10.1109/eScience.2008.114>
- The Open Provenance Model (v1.01). Moreau, L. (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. Van den Bussche, T. Ellkvist, J. Frieire, P. Growth (2008). *Technical Report, Electronics and Computer Science, University of Southampton*, 2008. <http://eprints.ecs.soton.ac.uk/16148>
- Yiming Sun, Suresh Marru, and Beth Plale (2008). Experience with Bursty Workflow-driven Workloads in LEAD Science Gateway *3rd Annual Teragrid 2008*, Las Vegas, NV, Jun 2008.
- Scott Jensen and Beth Plale (2008). Using Characteristics of Computational Science Schemas for Workflow Metadata Management, *Proceedings of the 2008 IEEE Congress on Services*, Computer Society Press, Washington DC, <http://doi.ieeecomputersociety.org/10.1109/SERVICES-1.2008.42>
- 2007** Yogesh Simmhan, Sangmi Lee Pallickara, Nithya A. Vijayakumar, and Beth Plale (2007). Data Management in Dynamic Environment-driven Computational Science. Book chapter in *Grid-Based Problem Solving Environments: Implications for Development and Deployment of Numerical Software*, *IFIP Int'l Federation for Information Processing Vol. 239*, P.W. Gaffney and J.C.T. Pool, Eds, Springer Boston, pp. 317-333, DOI: 10.1007/978-0-387-73659-4\_17
- Dennis Gannon, Beth Plale, Marcus Christie, Yi Huang, Scott Jensen, Ning Liu, Suresh Marru, Sangmi Lee Pallickara, Srinath Perera, Saotshi Shirasuna, Yiming Simmhan, Alex Slominski, Dennis Gannon, Beth Plale, and Dan Reed (2007). Service Architectures for e-Science Grid Gateways: Opportunities and Challenges, in *On the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS*, Lecture Notes in Computer Science Vol 4804, pp. 1179-1185, DOI 10.1007/978-3-540-76843-2
- Dennis Gannon, Beth Plale, Marcus Christie, Yi Huang, Scott Jensen, Nigh Liu, Suresh Marru, Sangmi Lee Pallickara, Srinath Perera, Satoshi Shirisuna, Yogesh Simmhan, Alexander Slominski, Yiming Sun, Nithya Vijayakumar (2007). Building Grid Portals for e-Science: A Service Oriented Architecture, *Volume 16: High Performance Computing and Grids in Action*, IOS Press - Amsterdam, Lucio Grandinetti editor
- Lavanya Ramakrishnan, Yogesh Simmhan, and Beth Plale (2007). Realization of Dynamically Adaptive Weather Analysis and Forecasting in LEAD: Four Years Down the Road, *Dynamic Data-Driven Application Systems Workshop, at International Conference on Computational Science (ICCS)*, Beijing, China, May 27-30, 2007
- Dennis Gannon, Beth Plale, Suresh Marru, Gopi Kandaswamy, Yogesh Simmhan, and Satoshi Shirasuna (2007). Dynamic, Adaptive Workflows for Mesoscale Meteorology. Book chapter in *Workflows for e-Science: Scientific Workflows for Grids*, Taylor, I.J.; Deelman, E.; Gannon, D.B.; Shields, M. (Eds.) Springer, pp. 129-145
- Nithya Vijayakumar and Beth Plale (2007). Tracking Stream Provenance in Complex Event Processing Systems for Workflow-driven Computing. *2nd Int'l Workshop on Event-driven Architecture, Processing, and Systems (EDA-PS'07)*, in conjunction with VLDB'07. Vienna, Austria, Sep 2007

- Nithya Vijayakumar and Beth Plale (2007). Prediction of Missing Events in Sensor Data Streams Using Kalman Filters, *1st Int'l Workshop on Knowledge Discovery from Sensor Data, in conjunction with ACM 13th Int'l Conf on Knowledge Discovery and Data Mining (KDD)*, San Jose, California, Aug 2007.
- 2006 Beth Plale, Dennis Gannon, Jerry Brotzge, Kelvin Droegemeier, Jim Kurose, Doug McLaughlin, Robert Wilhelmson, Sara Graves, Mohan Ramamurthy, Richard Clark, Sepi Yalda, Dan Reed, Edward Joseph, and V. Chandrasekar (2006). CASA and LEAD: Adaptive Cyberinfrastructure for Real-Time Multiscale Weather Forecasting, *Computer special issue on System-Level Science*, IEEE Computer Science Press, Vol. 39(11), pp. 56-63, <http://doi.ieeecomputersociety.org/10.1109/MC.2006.375>
- Yiming Sun, Scott Jensen, Sangmi Lee Pallickara, and Beth Plale (2006). Personal Workspace for Large-scale Data-driven Computational Experimentation, *7th IEEE/ACM Int'l Conf on Grid Computing (Grid'06)*, IEEE Computer Society, pp. 112-119. DOI 10.1.1.142.4780
- Liu, Y., N. N. Vijayakumar, and B. Plale (2006). Stream Processing in Data-driven Computational Science *7th IEEE/ACM Int'l Conf on Grid Computing (Grid'06)*, Barcelona Spain, September 2006. DOI 10.1109/ICGRID.2006.311011
- Vijayakumar, Nithya, Beth Plale, Rahul Ramachandran, and Xiang Li (2006). Dynamic Filtering and Mining Triggers in Mesoscale Meteorology Forecasting, *IEEE Int'l Geoscience and Remote Sensing Symp (IGARSS'06)*, Denver, CO, Aug 2006.
- Yogesh Simmhan, Beth Plale and Dennis Gannon (2006). Towards a Quality Model for Effective Data Selection in Collaboratories, *IEEE Workshop on Workflow and Data Flow for Scientific Applications (SciFlow06)*, held in conjunction with ICDE, Atlanta, GA, April 2006.
- Yogesh Simmhan, Beth Plale, and Dennis Gannon (2006). A Framework for Collecting Provenance in Data-Centric Scientific Workflows, *IEEE Int'l Conf on Web Services (ICWS'06)*, IEEE Computer Society Press, pp. 427-436, DOI 10.1109/ICWS.2006.5
- Yogesh Simmhan, Beth Plale, and Dennis Gannon (2006). Performance Evaluation of the Karma Provenance Framework for Scientific Workflows, *Int'l Provenance and Annotation Workshop (IPAW)*, in *Provenance and Annotation of Data, Lecture Notes in Computer Science*, Springer Berlin/Heidelberg, Vol 4145, 2006. DOI 10.1007/11890850
- Ying Liu and Beth Plale (2006). Multi-model Based Optimization for Stream Query Processing, *KSI 18th Int'l Conf on Software Engineering and Knowledge Engineering (SEKE'06)*, San Francisco, Jul 2006. DOI 10.1.1.85.4855
- Ying Liu and Beth Plale (2006). Query Optimization for Distributed Data Streams, *ISCA 15th Int'l Conf on Software Engineering and Data Engineering (SEDE'06)*, Los Angeles, Jul 2006. DOI 10.1.1.133.7271
- Sangmi Lee Pallickara, Beth Plale, Liang Fang, and Dennis Gannon (2006). Trust Cell: Towards the End-to-End Trust in Data-Oriented Scientific Computing, *IEEE Cluster Computing and Grid (CCGrid)*, May 2006.
- Sangmi Lee Pallickara and Beth Plale (2006). Enabling End-to-End Trustworthiness in Data-Oriented Scientific Computing, *Int'l Conf Workshops on Parallel Processing (ICPPW)*, Columbus, Ohio, Aug 2006. <http://doi.ieeecomputersociety.org/10.1109/ICPPW.2006.76>
- Scott Jensen, Beth Plale, Sangmi Lee Pallickara, and Yiming Sun (2006). A Hybrid XML-Relational Grid Metadata Catalog, *Int'l Conf Workshops on Parallel Processing (ICPPW)*, IEEE Computer Society, pp. 15-24, Aug 2006. <http://doi.ieeecomputersociety.org/10.1109/ICPPW.2006.10>
- Nithya Vijayakumar and Beth Plale (2006). Towards Low Overhead Provenance Tracking in Near Real-Time Stream Filtering, *Int'l Provenance and Annotation Workshop (IPAW'06)*, in *Provenance and Annotation of Data, Lecture Notes in Computer Science*, Springer Berlin/Heidelberg, Vol 4145, 2006. DOI 10.1007/11890850

- Katie A. Siek, Kay Connelly, Amanda Stephano, Suzanne Menzel, Jacki Bauer, Beth Plale (2006). Breaking the Geek Myth: Addressing Young Women's Misperceptions about Technology Careers, *Learning & Leading with Technology*, v. 33 no. 7 p19—22 Apr 2006
- 2005 K. Droegemeier, K. Brewster, M. Xue, D. Weber, D. Gannon, B. Plale, D. Reed, L. Ramakrishnan, J. Alameda, R. Wilhelmson, T. Baltzer, B. Domenico, D. Murray, A. Wilson, R. Clark, S. Yalda, S. Graves, R. Ramachandran, J. Rushing, E. Joseph (2005). Service-Oriented Environments for Dynamically Interacting with Mesoscale Weather", *Computing in Science and Engineering*, IEEE Computer Society Press and American Institute of Physics, Vol. 7(6), pp. 12-29
- Beth Plale, Dennis Gannon, Yi Huang, Gopi Kandaswamy, Sangmi Pallickara, and Aleksander Slominski (2005). Cooperating Services for Managing Data Driven Computational Experimentation, *Computing in Science and Engineering*, IEEE Computer Society Press and American Institute of Physics, Vol. 7(5), pp. 34 – 43.  
<http://doi.ieeecomputersociety.org/10.1109/MCSE.2005.91>
- Yogesh Simmhan, Beth Plale, and Dennis Gannon (2005). A Survey of Data Provenance in e-Science, *ACM SIGMOD Record*, ACM Press, Vol. 34(3), pp. 31-36
- Gannon, D., B. Plale, M. Christie, L. Fang, Y. Huang, S. Jensen, G. Kandaswamy, S. Marru, S. Lee Pallickara, S. Shirasuna, Y. Simmhan, A. Slominski, and Y. Sun (2005). Service Oriented Architectures for Science Gateways on Grid Systems, *International Conference on Service Oriented Computing 2005*, B. Benatallah, F. Casati, P. Traverso (Eds.), *Lecture Notes in Computer Science 3826*, Springer-Verlag Berlin Heidelberg pp. 21-32
- Gannon, D., J. Alameda, O. Chipara, M. Christie, V. Dukle, L. Fang, M. Farellee, G. Fox, S. Hampton, G. Kandaswamy, D. Kodeboyina, C. Moad, M. Pierce, B. Plale, A. Rossi, Y. Simmhan, A. Sarangi, A. Slominski, S. Shirasauna, T. Thomas (2005). Building Grid Portal Applications from a Web-Service Component Architecture, *Proceedings of the IEEE*, IEEE Press, Vol. 93, No. 3, pp. 551-563
- Beth Plale, Dennis Gannon, Jay Alameda, Bob Wilhelmson, Shawn Hampton, Alex Rossi, and Kelvin Droegemeier (2005). Active Management of Scientific Data, *IEEE Internet Computing special issue on Internet Access to Scientific Data*, IEEE Computer Science Press, Vol. 9(1), pp. 27-34
- Pallickara, S. L., B. Plale, S. Jensen, and Y. Sun (2005). Structure, Sharing, and Preservation of Scientific Experiment Data, *IEEE 3<sup>rd</sup> Int'l Workshop on Challenges of Large Applications in Distributed Environments (CLADE'05)*, Research Triangle Park, North Carolina, July 2005.
- Plale, B., Gannon, D., Reed, D., Graves, S., Droegemeier, K., Wilhelmson, B., and Ramamurthy, M. (2005). Towards Dynamically Adaptive Weather Analysis and Forecasting in LEAD, *Proceedings of Computational Science - ICCS Workshop on Dynamic Data Driven Applications, Lecture Notes in Computer Science (LNCS)*, No. 3515, Part II, Springer-Verlag GmbH, pp. 624 – 631
- Pallickara, S. L., B. Plale, S. Jensen, Y. Sun (2005). Monitoring Access to Stateful Resources in Grid Environments, *IEEE Int'l Conference on Services Computing (SCC'05)*, IEEE Computer Society Press, Orlando, Florida July 2005.
- Plale, B. and N. Vijayakumar (2005). Evaluation of Rate-based Adaptivity in Asynchronous Data Streams, *Proceedings of ACM/IEEE 19th Int'l Parallel and Distributed Processing Symposium (IPDPS)*, IEEE Computer Society Press, p. 69b,  
<http://dx.doi.org/10.1109/IPDPS.2005.205>, April 2005.
- 2004 Beth Plale (2004). Framework for Bringing Data Streams to the Grid, *Scientific Programming*, IOS Press, Amsterdam, Vol. 12(4), pp. 213-223
- Plale, B., C. Jacobs, S. Jensen, Y. Liu, C. Moad, R. Parab, and P. Vaidya (2004). Understanding Grid Resource Information Management through a Synthetic Database

- Benchmark/Workload, *Proceedings of 4<sup>th</sup> IEEE/ACM Int'l Symp on Cluster Computing and the Grid (CCGrid2004)*, Chicago, Illinois, April 2004.
- Plale, B. (2004). Using Global Snapshots to Access Data Streams on the Grid (2004). *Proceedings of 2<sup>nd</sup> European Across Grids Conf (AxGrids) published as Lecture Notes in Computer Science*, Springer-Verlag GmbH, Vol. 3165, pp. 191-201.
- 2003 Beth Plale and Karsten Schwan (2003). Dynamic Querying of Streaming Data with the dQUOB System, *IEEE Transactions of Parallel and Distributed Systems*, IEEE Computer Science Press, Vol. 14(3), pp. 422-432
- Kodeboyina, D. and B. Plale (2003). Experiences with OGSA-DAI: Portlet Access and Benchmark, *Global Grid Forum Workshop on Designing and Building Grid Services*, Chicago, Illinois, [http://www-unix.mcs.anl.gov/~keahey/DBGS/DBGS\\_files/dbgs\\_papers/kodeboyina.pdf](http://www-unix.mcs.anl.gov/~keahey/DBGS/DBGS_files/dbgs_papers/kodeboyina.pdf) September 2003.
- 2002 Plale, B., G. Turner, and A. Sharma (2002). Real Time Response to Streaming Data on Linux Clusters, *Proceedings of 3<sup>rd</sup> Int'l Conf on Linux Clusters: the HPC Revolution*, Linux Clusters Institute, <http://www.linuxclustersinstitute.org/Linux-HPC-Revolution/Archive/2002techpapers.html>, October 2002.
- Plale, B. (2002). Leveraging Runtime Knowledge about Event Rates to Improve Memory Utilization in Wide Area Data Stream Filtering 2002. *Proceedings of 11<sup>th</sup> IEEE Int'l Symp on High Performance Distributed Computing (HPDC)*, IEEE Computer Society, Washington, DC, p. 171, <http://dx.doi.org/10.1109/IPDPS.2001.925038> 2002.
- Plale, B., P. Dinda, and G. von Laszewski (2002). Key Concepts and Services of a Grid Information Service, *Proceedings of 15<sup>th</sup> ISCA Int'l Parallel and Distributed Computing Systems (PDCS'02)*, International Society for Computers and their Applications, Cary, North Carolina, pp. 437-442, 2002.
- 2001 Plale, B., P. Widener, and K. Schwan (2001). Taking the Step from Metadata to Communication Middleware in Computational Data Streams, *Proceedings of 10<sup>th</sup> Heterogeneous Computing Workshop*, April 2001, IEEE Computer Society Press, Washington, DC, p. 20085b <http://dx.doi.org/10.1109/IPDPS.2001.925038>, 2001.
- Plale, B. and Schwan, K. (2001). Optimizations Enabled by Relational Data Model View to Querying Data Streams, *Proceedings of the 15<sup>th</sup> Int'l Parallel and Distributed Processing Symp (IPDPS)*, IEEE Computer Society Press, Washington, DC, p. 10022a <http://dx.doi.org/10.1109/IPDPS.2001.924953>, 2001.
- 2000 Plale, B. and K. Schwan (2000). dQUOB: Managing Large Data Flows by Dynamic Embedded Queries, *Proceedings of IEEE High Performance Distributed Computing (HPDC'00)*, IEEE Computer Society Press, Washington DC, p. 263 <http://dx.doi.org/10.1109/HPDC.2000.868658>, 2000. Extended version available as Georgia Institute of Technology Technical Report GIT-TR-00-07.
- Plale, B., G. Eisenhauer, L. K. Daley, P. Widener, and K. Schwan (2000). Fast Heterogeneous Binary Data Interchange for Event-based Monitoring, *ISCA Int'l Conference on Parallel and Distributed Computing Systems (PDCS)*, August 2000.
- Oleson, V., K. Schwan, G. Eisenhauer, B. Plale, C. Pu and D. Amin (2000). Operational Information Systems - An Example from the Airline Industry, *Proceedings of 1<sup>st</sup> Workshop on Industrial Experiences with Systems Software (WIESS)*, USENIX Advanced Computing Systems Association, October 2000.
- 1999 Beth Plale, Volker Elling, Greg Eisenhauer, Karsten Schwan, D. King, and Vernard Martin (1999). Realizing Distributed Computational Laboratories, *International Journal of Parallel and Distributed Systems and Networks*, International Association of Science and Technology for Development (IASTED) Press, Vol. 2(3)

---

Plale, B. and K. Schwan, K. (1999). Run-time Detection in Parallel and Distributed Systems: Application to Safety-Critical Systems, *19<sup>th</sup> IEEE Int'l Conference on Distributed Computing Systems (ICDCS)*, IEEE Computer Society Press, p. 0163, DOI:10.1109/ICDCS.1999.776517, 1999.

1998 Beth Plale, Greg Eisenhauer, Karsten Schwan, Jeremy Heiner, Victor Martin, and Jeff Vetter (1998). From Interactive Applications to Distributed Laboratories, *IEEE Concurrency*, IEEE Computer Society Press, Vol. 6(2), pp. 78-90

Greg Eisenhauer, Beth Plale, Karsten Schwan (1998). DataExchange: High Performance Communications in Distributed Laboratories, *Journal of Parallel Computing*, Elsevier, Vol. 24(12-13), pp. 1713-1733

1997 Schroeder (Plale), B., S. Aggarwal, and K. Schwan, K. (1997). Software Approach to Hazard Detection Using On-line Analysis of Safety Constraints, *16<sup>th</sup> IEEE Symp on Reliable Distributed Systems (SRDS)*, IEEE Computer Society Press, October, pp. 80, <http://dx.doi.org/10.1109/RELDIS.1997.632801>, 1997.

Eisenhauer, G., B. Plale-Schroeder, K. Schwan, V. Martin, and J. Vetter (1997). DataExchange: High Performance Communications in Distributed Laboratories, *IASTED Int'l Conf on Parallel and Distributed Computing and Systems (PDCS)*, October 1997.

1996 Eisenhauer, G., B. Plale-Schroeder and K. Schwan (1996). From Interactive High Performance Programs to Distributed Laboratories: A Research Agenda, *IEEE SPDP'96 Workshop on Program Visualization and Instrumentation*, Oct 1996

1995 Beth Schroeder (Plale) (1995). On-line Monitoring: A Tutorial, *Computer*, IEEE Computer Science Press, Vol. 28(6), pp. 72-78

---

## INVITED KEYNOTES(\*) and TALKS (96)

		EVENT HOST OR LOCATION
2021	<i>Pervasive Technology Institute: Shaping the Future of IT innovation for Indiana University</i> , CENTRA 5, Sept 2021	Porto, Portugal
	<i>Trustworthy Artificial Intelligence and Open Science</i> , Beilstein Open Science Symposium, Beilstein Institute, Oct 2021	Frankfurt Germany
	<i>Open Science: Trusted Products</i> , Center for Informatics Research in Science and Scholarship, University of Illinois Urbana Champaign, May 2021	UIUC
2020	<i>Communicating through Research and Researcher</i> , SC20 Early Career Panel on Communications, International Conference for High Performance Computing, Networking, Storage, and Analysis, Nov 2020	Atlanta, Georgia
	<i>*Open Science: The Challenge and the Promise</i> , Michigan Institute for Data Science, University of Michigan, Sep 2020	Univ of Michigan
	<i>Artificial Intelligence Research and NSF: a Conversation</i> , Luddy AI Symposium, Jun 2020	Indiana Univ
	<i>*Open Science in Biodiversity, the Challenge and the Promise</i> , 4 <sup>th</sup> Digital Data Biodiversity Conference, Jun 2020	Bloomington, IN
	<i>Knowledge Graphs in AI: Trust and FAIR Reuse</i> , AI Technology, Application, and Innovation for Digital Cities, 2020 Smart City Summit and Expo, Taipei, Taiwan, Jun 2020	Taipei, Taiwan
	<i>Generalist repositories: NSF policy and perspective</i> , Establishing a FAIR Biomedical Data Ecosystem: The Role of Generalist and Institutional Repositories to Enhance Data Discoverability and Reuse, National Institutes of Health, Bethesda, MD, Feb 2020	National Institutes of Health (NIH)

	<i>*Open Science Research Ecosystem: a Multi-Legged Stool</i> , PIDapalooza: Open Festival of Persistent Identifiers, Lisbon, Portugal, Jan 2020	Lisbon, Portugal
2019	<i>Towards Fully Realized Open Science and Technology Trustworthiness</i> , Halicioğlu Data Science Institute, University of California San Diego, Dec 2019	UC San Diego
	<i>Transparency by Design in eScience Research (invited, visionary session talk)</i> , 15 <sup>th</sup> Int'l Conference on eScience, Sept 2019	San Diego California
	<i>*Open Science is (Good) Data Science</i> , Indiana University Peebles Memorial Lecture, May 2019	Indiana Univ
	<i>Open Science, FAIR data, and Cyberinfrastructure</i> , 36 <sup>th</sup> Pacific Rim Applications and Grid Middleware (PRAGMA), Jeju South Korea, Mar 2019	Jeju, South Korea
2018	<i>Open Science and Data Sharing: Policy and Infrastructure</i> , Institute for Systems and Computer Engineering, Technology and Science (INESTEC), Mar 2018	INESTEC Porto, Portugal
	<i>Making Open Science work for science and society</i> , IEEE Big Data Governance and Metadata Management workshop, Mar 2018	Berlin, Germany
	<i>Capsule Model for Open Science with Restricted Data</i> , Data Preservation Alliance for the Social Sciences (Data-PASS), Sep 2018.	Boston, Massachusetts
	<i>*Open Science and Data Sharing: Policy and Infrastructure</i> , Future Visions Symposium, Apr 2018	Colorado State Univ
	<i>*Open Science and Data Sharing: Policy and Infrastructure</i> , 2 <sup>nd</sup> Annual Texas A&M Research Computing Symposium, Jun 2018	Texas A&M
	<i>Capsule Framework for Open Science with Restricted Data</i> , Technical Solutions to Advance Evaluation and Replication in the Social Sciences: What's New, What's Next workshop. Co-located with American Political Science Association (APSA) annual meeting, Aug 2018	Boston, Massachusetts
	<i>Capsule Computing: Safe Open Science</i> , Computer Science Seminar, Binghamton University, Dec 2018	SUNY Binghamton
	<i>*Open Science as Roadmap to Better Data Science Research</i> , Data Science Initiative Seminar, Binghamton University, Dec 2018	SUNY Binghamton
2017	<i>*FAIR Open Science with PID Kernel Information: the RPID Testbed</i> , Basarim 2017: 5 <sup>th</sup> High Performance Computing Conference, Sep 2017	Istanbul, Turkey
	<i>Analyzing the HathiTrust Digital Library: 5.5 billion pages of knowledge</i> , Earth Science Information Partners (ESIP) Summer Meeting, Jul 2017	Bloomington, Indiana
	<i>Foundations of a Data Ecosystem: Global Persistent IDs and Data Provenance</i> , Int'l Symposium on Open Data and Innovation, Jul 2017	Beijing, China
	<i>PID Kernel Information: data handles and provenance</i> , Collaborations to Enable Transnational Cyberinfrastructure Applications CENTRA2 All Hands Meeting, Apr 2017	Gainesville, Florida
	<i>Foundations of a Data Ecosystem: Global Persistent IDs and Data Provenance</i> , Computer Science and Computer Engineering Colloquium, University of Arkansas, Mar 2017	Univ of Arkansas
	<i>*Data Mining Meets the Research Library</i> , Computing and Humanity Speaker Series, Valparaiso University, Feb 2017	Valparaiso Univ
2016	<i>Power of PID Kernel Information</i> , Southeast Asia International Joint-Research and Training Program, Tainan City, Taiwan, Dec 2016	Tainan City, Taiwan
	<i>Opening a Massive Cultural Record</i> , Dean's Advisory Council, School of Informatics and Computing, Indiana University Oct 2016	Indiana Univ
	<i>Data Science of Big Data</i> , School of Public and Environmental Affairs, Indiana University, Oct 2016	Indiana Univ

2015	<i>HathiTrust Research Center: Secure Commons</i> , University of Toronto, Jun 2015	Univ of Toronto
	<i>*Trust Threads: Minimal Provenance for Data Publication and Reuse</i> , National Conference on Data Integrity, Department of Information and Library Science, Colorado State University, May 2015	Colorado State Univ
	<i>Threads of Trust: Provenance of Data Reuse in Long Tail Science</i> , Department of Information and Library Science, Indiana University, Apr 2015	Indiana Univ
	<i>*HathiTrust Research Center: Unlocking the Secrets of 4.6 Billion Pages</i> , University of Missouri Cyberinfrastructure Days, Mar 2015	Univ of Missouri
2014	<i>Big data, publishing, and data sharing</i> , University of Tennessee i-School seminar, Dec 2014	Univ of Tennessee
	<i>Research Data Alliance and Big Data</i> , Oak Ridge National Labs, Dec 2014	Knoxville, Tennessee
	<i>HathiTrust Data Capsules</i> , CLIR/CNI Workshop on Expanding Access to Research Collections, Dec 2014	Washington, DC
	<i>Integrating Domain Repositories into the National Data Infrastructure</i> , Workshop on National Data Infrastructures, ICPSR, Nov 2014	University of Michigan
	<i>Semantics, Data Provenance, Agent based Models</i> , Ostrom Workshop on Political Theory and Policy Analysis, Oct 2014	Bloomington, Indiana
	<i>HathiTrust Research Center Data Capsule v1.0: An overview of functionality</i> , with Robert McDonald, Miao Chen, IU Scholarly Data Commons, Sep 2014 <a href="http://hdl.handle.net/2022/18936">http://hdl.handle.net/2022/18936</a>	Bloomington, Indiana
	<i>Case Study in Big Data: the Socio-Technical Issues of HathiTrust Digital Texts</i> , Women's Institute for Summer Enrichment (TRUST WISE), Jun 2014	Ithaca, New York
	<i>*Bridging Digital Humanities Research and Large Repositories of Digital Text</i> , 2nd Encuentro de Humanistas Digitales, Biblioteca Vasconceles, May 2014 <a href="http://www.slideshare.net/BethPlale/keynote-2nd-encuentro-de-humanistas-digitales">http://www.slideshare.net/BethPlale/keynote-2nd-encuentro-de-humanistas-digitales</a>	Mexico City, Mexico
	<i>HathiTrust and HathiTrust Research Center: the Changing Digital Library</i> , El Colegio de Mexico, May 2014	Mexico City, Mexico
	<i>HathiTrust Research Center: Challenges and Opportunities in Big Text Data</i> , Miao Chen and Beth Plale, IU Digital Libraries Brown Bag, Mar 2014 <a href="http://hdl.handle.net/2022/17276">http://hdl.handle.net/2022/17276</a>	Bloomington, Indiana
	<i>Archiving a social-ecological database: challenges, solutions, and lessons learned</i> , Beth Plale and Inna Kouper, IU Digital Libraries Brown Bag, Feb 2014 <a href="http://hdl.handle.net/2022/17301">http://hdl.handle.net/2022/17301</a>	Bloomington, Indiana
	<i>Research Data Alliance</i> , Meeting of NSF EarthCube Community, Jan 2014	Washington, DC
2013	<i>Big Data Opportunities and Challenges for IR, Text Mining, and NLP</i> , The British Library, Dec 2013	London, UK
	<i>Big Data Opportunities and Challenges for IR, Text Mining, and NLP</i> , Knowledge Media Institute (KMi), The Open University, Dec 2013	Milton Keynes, UK
	<i>Data Sets, Ensemble Cloud Computing, and the University Library</i> , American Geophysical Union (AGU) Meeting, Dec 2013 (presented by co-author, Jim Myers)	San Francisco, California
	<i>Opportunities and Challenges of Text Mining HathiTrust Digital Library</i> , The National Library of the Netherlands (Koninklijke Bibliotheek), Nov 2013	Den Haag, The Netherlands
	<i>*Big Data Opportunities and Challenges for IR, Text Mining and NLP</i> , Int'l Workshop on Mining Unstructured Big Data Using Natural Language Processing (MNL 2013), co-located with ACM Int'l Conf. on Information	San Francisco, California

	and Knowledge Management (CIKM), Oct 2013	
	<i>*Big Data and Open Access: On Track for Collision of Cosmic Proportions?</i> , 2 <sup>nd</sup> Int'l LSDMA-Symposium - The Challenge of Big Data in Science - with a focus on Big Data Analytics, Sep 2013	Karlsruhe, Germany
	<i>HathiTrust Research Center (HTRC): Exploration of the World's First Massive Digital Library</i> , with Miao Chen and Robert McDonald, Catapult Center for Digital Humanities and Computational Analysis of Texts Digital History and Philosophy of Science (HPS) Workshop, Sep 2013	Bloomington, Indiana
	<i>Research Data Alliance: Researchers Sharing and Using Research Data Without Barriers</i> , NIEHS-European Union Workshop on Identifying Opportunities for Global Integration of Toxicogenomics Databases, Jun 2013.	Research Triangle Park, North Carolina
	<i>Digital Humanities Text Mining at Scale: HathiTrust Research Center</i> , Notre Dame Digital Humanities talk series, May 2013	South Bend, Indiana
	<i>Studies in Social-Ecological Systems Data Management</i> , Interuniversity Consortium for Political and Social Research (ICPSR), Apr 2013	Ann Arbor, Michigan
	<i>International Data Sharing, Open Access and the Research Data Alliance</i> , Advanced Regional Networks Envision Workshop on Big Data, Apr 2013	Boston, Massachusetts
2012	<i>Research Data Alliance (RDA), EarthCube, and SEAD DataNet</i> , Bridging Big Data Infrastructures Workshop, Dec 2012	Taichung, Taiwan
	<i>Explicit and Hidden Workflows in Environmental Science: Opportunities Enabled By</i> , Microsoft Open Data for Open Science 2012 workshop, Apr 2012	Redmond, Washington
	<i>Metadata and Provenance: Fins in the Data Sea</i> , Purdue University, Mar 2012	West Lafayette, Indiana
	<i>Digital Humanities at Scale: HathiTrust Research Center</i> , University of Maryland College Park, Feb 2012	College Park, Maryland
2011	<i>Creating Functionality Around Non-consumptive Research</i> , American University School of Law, Jul 2011	Washington, DC
	<i>Metadata and Provenance Capture: Antecedent to Scientific Data Preservation</i> , IU School of Library and Information Science course Data Curation Feb 2011.	Bloomington, Indiana
	<i>Provenance Collection of Unmanaged Workflows</i> , Data To Insight Center Seminar, Indiana University, Jan 2011	Bloomington, Indiana
	<i>Metadata and Provenance Collection and Representation: Antecedent to Scientific Data Preservation</i> , EECS Seminar, Oregon State University, Jan 2011	Corvallis, Oregon
	<i>Metadata and Provenance Collection and Representation: Antecedent to Scientific Data Preservation</i> , EECS Seminar, Moratuwa University, May 2011	Columbo, Sri Lanka
2010	<i>Keynote talk: LEAD II Hybrid Workflows for Timely Weather Products</i> , 19 <sup>th</sup> Pacific Rim Applications and Grid Middleware Assembly (PRAGMA), Sep 2010	Changchun, China
	<i>Metadata and Provenance Collection and Representation: Antecedent to Scientific Data Preservation</i> , Open Data Seminar, University of Michigan, Nov 2010.	Ann Arbor, Michigan
	<i>Provenance and Workflows</i> , Computer Network Information Center of Chinese Academy of Sciences, Oct 2010.	Beijing, China

	<i>LEAD II/Trident workflows for timely weather products: challenge of Vortex2, American-Chinese Cyberinfrastructure and E-science Workshop (ACCESS) on Data Intensive Sciences and Computing (DISC), Aug 2010.</i>	Urbana, Illinois
	<i>LEADII: hybrid workflows in atmospheric science, DemoFest, Microsoft Faculty Research Summit, Jul 2010.</i>	Seattle, Washington
	<i>LEAD II / Trident Workflows for Timely Weather Products: the Challenge of Vortex2, Microsoft External Research Symposium, Apr 2010</i>	Seattle, Washington
	<i>Earth Systems Data in Real Time Applications: Low Latency, Metadata, and Preservation, Data-Intensive Research: how should we improve our ability to use data, e-Science Institute, University of Edinburgh, Mar 2010</i>	Edinburgh, Scotland
2009	<i>Metadata and Preservation in Geosciences: Issues at Scale, IU Digital Libraries Program Brown Bag, Sep 2009</i>	Bloomington, Indiana
	<i>Keynote talk: Discovery, Process and Preservation: the ABC's of Data in a Collaborative World, Indiana Universities Pervasive Technologies Institute Research Technologies Data Services Day, Sep 2009</i>	Indianapolis, Indiana
	<i>It's All in the Data, Workflow Systems and the Weather, American Geophysical Union (AGU) Meeting, Geoinformatics Session, May 2009</i>	Toronto, Canada
	<i>Integrating Weather/Climate Research Models and Data into End-Use Knowledge: Data to Insight, IBM Lecture Series, Dept. of Computer Science and Engineering, University of Notre Dame, Feb 2009</i>	South Bend, Indiana
	<i>Integrating Weather/Climate Research Models and Data into End-Use Knowledge: Data to Insight, Computer Science Dept. Honors Lecture Series, Indiana University, Feb 2009</i>	Bloomington, Indiana
2008	<i>Beyond LEAD: Impact, Education, and Future Plans, Supercomputing '08, Nov 2008</i>	Austin, Texas
	<i>*Research Instrumentation for Cyberinfrastructure, Data-intensive Computing and Weather Forecasting, NSF Workshop on Instrumentation Needs of Computer and Information Science Engineering, Snowbird, Jul 2008</i>	Salt Lake City, Utah
	<i>Provenance and Metadata in Data-Intensive Computing, Seminar Series, Engineering, Architecture, and Computer Sciences, Howard University, Oct 2008</i>	Washington, DC
	<i>Provenance of Digital Scientific Data, IEEE/ACM Supercomputing '08, Nov 2008</i>	Austin, Texas
2007	<i>Improving Data Capture in Science Discovery Cyberinfrastructure to Enable Educational Outcomes, Learning Sciences Professional Seminar, School of Education, Indiana University, Nov 2007</i>	Bloomington, Indiana
	<i>Provenance of workflow generated data: bridge to future and aid to workflow interoperability, Science and Scholarly Workflows Workshop, Oct 2007</i>	Baltimore, Maryland
	<i>Metadata, Provenance, and Search in e-Science, Complex Networks Seminar, Indiana University, Sep 2007</i>	Bloomington, Indiana
	<i>Data Management, Metadata, and Search in Workflow-driven Computational e-Science, invited talk Renmin University, May 2007</i>	Beijing, China
	<i>Data Integration, Search, and Analysis in Workflow-driven Computational Science, Workshop on Grid Portals and Data Management techniques for Earth Science Applications, Jun 2007</i>	Sardinia, Italy
	<i>SOA and Events Processing for Next Generation Weather Forecasting: the LEAD Project, Object Management Group, Mar 2007</i>	San Diego, California

	<i>Data Integration, Search, and Analysis in Workflow-driven Computational Science</i> , Purdue University, June 2007	West Lafayette, Indiana
2006	<i>*Distinguished Talk: Opening the Gates to Data Driven Computational Science through Cyberinfrastructure</i> , Indiana University Office of Women's Affairs Distinguished Lecture Series, May 2006	Bloomington, Indiana
	<i>Metadata Catalogs and Stream Processing: Key Cyberinfrastructure for Data Driven Computational Science</i> , University of Houston Computer Science Department Colloquia Series, Apr 2006	Houston, Texas
	<i>Transforming the Sensing and Prediction of Intense Local Weather Through Dynamic Adaptation</i> , NSF Dynamic Data Driven Application Systems (DDDAS) Workshop, Jan 2006.	Washington, DC
2005	<i>Calder, OGSA-DAI Access to Data in Streams</i> , Argonne National Labs, Aug 2005	Chicago, Illinois
	<i>LEAD Data Subsystem: Overview, Current Approach to Integration, and Challenges</i> , DIALOGUE Workshop, Aug 2005	Columbus, Ohio
	<i>Temporally Changing Geospatial Data in LEAD and DopplerSource</i> , Microsoft e-Science Workshop, Aug 2005, <a href="https://www.microsoft.com/en-us/research/video/dopplersource-net-framework-for-accesing-doppler-radar-data/">https://www.microsoft.com/en-us/research/video/dopplersource-net-framework-for-accesing-doppler-radar-data/</a>	Seattle, Washington

## TEACHING and RESEARCH SUPERVISION

---

### RESEARCH SUPERVISION

#### Current PhD Advisor (3):

Yu Luo (enter 2016) – Information science and cloud computing  
 Sachith Withana (enter 2018) – High performance computing and data

#### Graduated PhD Students (Advisor and Chair) (15):

Milinda Pathirage	KPMG	2018
Zong Peng	Microsoft Research AI (MSR AI)	2018
Isuru Suriarachchi	Amazon	2018
Guangchen Ruan	Research staff, Indiana University	2016
Peng Chen	Facebook	2016
Jiaan Zeng	Electronic Arts	2015
Yuan Luo	Facebook	2015
Devarshi Ghoshal	Lawrence Berkeley National Labs	2014
You-Wei Cheah	Lawrence Berkeley National Labs	2013
Chathura Herath	Knight Capital Group	2011
Eran Chinthaka Withana	Comprehend Systems	2011
Scott Jensen	Asst. Professor, San Jose State Univ	2010
Ying Liu	2007. Cisco	2007
Yogesh Simmhan	Assoc. Professor, India Institute of Science Bangalore	2007

Nithya Vijayakumar	Apple	2007
--------------------	-------	------

*Postdoctoral Advisor (7):*

Inna Kouper	Research Faculty, School of Informatics and Computing, Indiana University	2012 – 2016
Miao Chen	Alibaba, Hangzhou, China	2011 – 2015
Abhirup Chakraborty	Google	2012 – 2014
Stacy Kowalczyk	Assoc. Professor, Dominican University, Illinois	2010 – 2012
Mehmet Aktas	Professor, Yildiz Teknik University, Budapest, Turkey	2010 – 2011
Bin Cao	Teradata Corp	2008 – 2009
Sangmi Lee Pallickara	Professor, Colorado State University	2005 – 2007

*Advisor and mentor (2):*

Umashanthi Pavalanathan (2012-2013)	Junior visiting research scholar	now PhD candidate at Georgia Institute of Technology
Kavitha Chandrasekar (2011-2014)	Data To Insight Center developer	now PhD candidate at Georgia Institute of Technology

*Master's Thesis Advisor (24):*

Quan (Gabriel) Zhou (2018) eBay, San Francisco, CA	Aparna Venkatraman (2007) Cummins, Columbus IN
Liang Ran (2017), Amazon, Seattle, WA	Vinay Pandey (2007) Dell Technologies, Austin, TX
Charitha Madurangi Dandeniya Arachchi (2017), HathiTrust Research Center, Bloomington, IN	Srilatha Marru (2007) RightRez, Bloomington, IN
Pradeep Ravilla (2017), Global Network Operations Center, Bloomington, IN	Humin (Lily) Fang (2007)
Aravindh Varadharaju,	Ning Liu (2006) Audible, NYC
Yiming Sun (2013), Amazon, Seattle, WA	Ryan Baula (2005)
Bina Bhaskar (2012), Amazon, Orange County, CA	Craig Jacobs (2005)
Bimalee Salpitikorala, Indiana University, Knoxville, TN	Charlie Moad (2004) Costello, Indianapolis IN
Prajakta Purhoit (2011), Chef Software, Seattle, WA	Poornima Venkatakrishnan (2004) PayPal, San Francisco, CA
Shobana Krishnan (2011) Cummins, IN	Nithya Sivaraman (2004) Microsoft, Seattle, WA
Ashish Bhangale (2010) Microsoft, Seattle, WA	Deepti Kodeboyina (2004) ChargePoint, San Francisco, CA
Felix Terkhorn (2009) Cake Solutions Limited, NYC	
Ai Zhang (2008), Cisco, San Francisco, CA	

**PhD External Reviewer / External Committee member:**

- University of Melbourne, Melbourne, Australia
- Rensselaer Polytechnic University

## TEACHING

INSTITUTION	COURSES TAUGHT (* NEW OR SUBSTANTIAL COURSE DEVELOPMENT)	DATES/ENROLLMENTS
IUB	*ENGR E222 <i>Intelligent Systems Engineering II</i>	S 2021 (35)
IUB	*ENGR E599, INFO I590, CSCI 659 <i>Artificial Intelligence (AI) Trust and Safety</i>	F 2020 (12)
IUB	* I535, B669, I435: <i>Management Access and Use of Big and Complex Data</i> . Hybrid course. Data lifecycle, data storage, cloud computing	F 2016 (290), Fall 2015 (60), Fall 2014 (40)
IUB	* CSCI P434: <i>Distributed Systems</i> . Foundations of distributed computing	Spr 2012 (20)
IUB	* CSCI B669: <i>Scientific Data Management and Preservation</i> . Data lifecycle, scientific workflows.	Spr 2011 (35), Spr 2013 (35)
IUB	* I590/B669: <i>Topics in Data and Search Informatics</i> . Data provenance, data indexing, case based reasoning, metadata models, and visualization. Indiana University, co-taught with Dennis Groth.	Fall 2008 (35)
IUB	* CSCI B534 <i>Distributed Systems, Indiana University</i> . Foundations of distributed computing including models, consistency, global time, architectures. Projects: 2010 - Parallel ray tracing application using Amazon Elastic Map Reduce and Web Services. 2007 - performance benchmark of virtual machines on identical Dell and HP hardware platforms. 2005: sponsored several teams to participate in a Kelley Business School and HP sponsored competition on course-wear for handheld devices.	2003 (as B649), 2004, 2005, 2007, 2009, 2010
IUB	CSCI P436 <i>Introduction to Operating Systems</i>	2001, 2002, 2007
IUB	CSCI B649: <i>Service Architecture and Science</i> . Co-taught with Dennis Gannon	2006
IUB	CSCI B438 <i>Fundamentals of Networking</i>	2005, 2006
IUB	* CSCI B649 <i>Topics in Systems: Systems Support for Wide Area Applications</i>	2002
Georgia Tech	CS4210 <i>Advanced Operating Systems</i>	Sum 2000, Sum 2001
Georgia Perimeter College	<i>Fortran for Scientists and Engineers</i>	Fall 1995, Spr 1996
Georgia Perimeter College	<i>Introduction to Pascal</i>	Fall 1995, Spr 1996

North Springs High School	* <i>Introduction to Visual Basic</i> , Advanced Placement course taught at North Springs High School, <i>Georgia Perimeter College</i>	Fall 1995
Georgia Perimeter College	* <i>Visual Basic for Professionals</i>	Spr 1996
SUNY Binghamton	C Programming and the UNIX operating system. Off campus continuing education	Spr 1994
SUNY Binghamton	C Programming and the UNIX operating system, graduate course, <i>SUNY Binghamton</i>	Su 1994

## SERVICE TO RESEARCH COMMUNITY

---

### Research Community Leadership

General Chair or  
Program Chair for  
peer reviewed  
conferences:

Chair, SC20 Transparency and Reproducibility Initiative, part of IEEE/ACM Int'l Conference on High Performance Computing, Networking, Storage, and Analysis (SC20), 2020

Program Chair, 2019 IEEE Int'l Congress on Big Data, Jul 2019

General Chair, ACM High Performance Distributed Computing (HPDC) 2014

General Chair, 5<sup>th</sup> Int'l Provenance and Annotation Workshop (IPAW) 2014

Program Chair, IEEE 20<sup>th</sup> Int'l Conference on Web Services (ICWS) 2013

Program Vice Chair, 2006 IEEE/ACM Int'l Conference on Grid Computing (Grid06)

Community Leader:

Organized Workshop on Data Quality in Era of Big Data, Bloomington, IN Sep 2016

Organized 1<sup>st</sup> PRAGMA Int'l Clouds for Data Science, Indonesia, Oct 2015. NSF Funded.

Organized PRAGMA 27: Networking in a Big Data World, Bloomington, IN 2014

Organized 2014 NSF Software Infrastructure for Sustained Innovation (SI2) PI meeting. Report at <https://hdl.handle.net/2022/19760>

Organized Coming Together Around Data Workshop for NSF DataNet and INTEROP Projects, a Principal Investigator meeting, NSF funded, Indianapolis, IN, Jan 2012.

Organized 1st Int'l Workshop on Traceability and Compliance of Semi-Structured Processes (TC4SP2010), Sep 2010

Organized GeoCloud: Cloud Computing, Collaborative Technologies and the Geosciences, NSF funded. Sep 2009.

Planning Committee: National Forum for Geosciences Information Technology (FGIT), 2005

Organized Riding the Geoscience Cyberinfrastructure Wave of Data: Real Time Data Use in Education, Dec 2008

Organized Indiana Women in Computing (inWIC) 2006

Organized Central Indiana Celebration of Women in Computing (CICWIC) 2004

Co-lead, NCSA Alliance Scientific Portal Expedition, 2003-2005

Chair, Global Grid Forum NOMCOM, 2004

Co-Chair Global Grid Forum (GGF) Relational Grid Information Systems Group, 2001-2002

**Technical Program Committee FOR HIGHLY SELECTIVE peer reviewed CONFERENCES AND WORKSHOPS**

	<b>YEAR SERVED</b>
Int'l Provenance and Annotation Workshop (IPAW)	2014, 2015, 2016, 2018
IEEE Int'l Conf. on Parallel and Distributed Systems (ICPADS)	2015
Int'l Conf. on Scientific and Statistical Database Management (SSDBM)	2014, 2015, 2019
Workshop on Scientific Cloud Computing (ScienceCloud)	2014, 2016, 2018
IEEE Int'l Conf. on eScience (e-Science)	2014, 2015, 2017
ACM Int'l Conf on Information and Knowledge Management (CIKM)	2016
IEEE/ACM Supercomputing (SC) tutorials committee,	2011, 2012
IEEE Int'l Conf. on High Performance Computing (HiPC)	2012
IEEE/ACM Int'l Parallel and Distributed Processing Symposium (IPDPS)	2011
Int'l ACM Symp. on High Performance Distributed Computing (HPDC)	2011, 2012
USENIX Conf. on File and Storage Technologies (FAST)	2010
Int'l Provenance and Annotation Workshop (IPAW)	2010, 2012
1st Int'l Workshop on Workflow Approaches for New Data-Centric Science (WANDS)	2010
3 <sup>rd</sup> Int'l Workshop on Data Intensive Distributed Computing (DIDC)	2010
Challenges of Large Applications in Distributed Environments (CLADE)	2010
IEEE Int'l Conf. on e-Science (e-Science)	2010, 2011, 2012
Int'l Conf. on Distributed Event Based Systems (DEBS)	2008, 2009
IEEE Int'l Conf. on e-Science and Grid Computing	2007, 2009
IEEE Int'l Conf. on Information Reuse and Integration	2008
Grace Hopper Conference	2006, 2009
1 <sup>st</sup> Int'l Workshop on the role of Semantic Web in Provenance Management	2009

Int'l Conf. for High Performance Computing and Communications (SC)	2006, 2007
Semantic Scientific Knowledge Integration, AAAI/SSS Workshop	2007
Second Int'l Workshop on Event-driven Architecture, Processing and Systems (EDA-PS'07)	2007
7 <sup>th</sup> IEEE Int'l Symposium on Cluster Computing and the Grid (CCGrid)	2007
IEEE/ACM Int'l Parallel and Distributed Processing Symp. (IPDPS)	2006, 2005
IEEE High Performance Distributed Computing (HPDC)	2002
IEEE Heterogeneous Computing Workshop (HCW)	2002
<b>SESSION CHAIR for peer reviewed CONFERENCES AND WORKSHOPS</b>	<b>YEAR SERVED</b>
Int'l Conf. on Distributed Event Based Systems (DEBS)	2009
Int'l Conf. on Data Engineering (ICDE), Atlanta, GA	2006
IEEE/ACM Int'l Conf. on Grid Computing (Grid06)	2006
IEEE High Performance Distributed Computing (HPDC)	2002
ICSA Parallel and Distributed Computing Systems (PDCS)	2002
ICSA Parallel and Distributed Computing Systems (PDCS)	2000
<b>FUNDER MERIT REVIEW panel MEMBER</b>	<b>YEAR SERVED</b>
National Science Foundation, technical reviewer for OAC, CISE, GEO, SBE, OISE, MPS, USA	yearly until 2017
Academy of Finland, Finland	2015
Department of Energy, USA	2006, 2007, 2009, 2010, 2013, 2014
Department of Homeland Security, USA	2013
Skoltech, Russia	2013
Engineering and Physical Sciences Research Council (EPSRC), United Kingdom	2006

## SERVICE TO INDIANA UNIVERSITY

---

ROLE	UNIVERSITY LEVEL ACTIVITY	YEAR SERVED
Member	IUB Promotion Committee	2020 –
Member	IU Center of Excellence for Women and Technology (CEWIT) Advisory Board	2015 –
Chair	IU Associate Vice President for Research (AVPR) search committee	2014
Chair	IU Digitization Master Plan Task Force	2014
Member	IUB Provost Strategic Planning Faculty Team	2013
Chair	IUB Faculty Research Support Funding Program (FRSP) (\$1M annual internal grant program)	2010
Panel member	IUB Faculty Research Support Funding Program (FRSP)	2005, 2006, 2007, 2008, 2009
Member	IUB Provost search committee	2012
Member	IU Vice President of Research search committee	2009, 2010
Member	IUB Dean of Libraries search committee	2008
Member	University Intellectual Property Council	2008 – 2011

Member	Bloomington Faculty Council (BFC) Faculty Mediation Committee	2009 – 2011
Member	University Conflict of Interest Committee	2008 – 2009
Member	University Information Technology Committee	2008
Member	Dean of School of Informatics, Computing, and Engineering search committee	2007
Faculty advisor	IUB Office of Women's Affairs Women in Science Program	2006 – 2007
Faculty advisor	Midwest Crossroads Alliance for Graduate Education and the Professoriate (AGEP)	2006 – 2009
Member	IU Restricted and Classified Task Force	2005 – 2006
Member	University IT Director of Systems search committee	2005
Member	University IT Director of Applications search committee	2005
<b>ROLE SCHOOL OR DEPT LEVEL ACTIVITY YEAR SERVED</b>		
Chair	Luddy School Diversity and Inclusion Task Force	2020 –
Chair	Data Science Program Curriculum Committee	2015 – 2017
Member	Data Science Program Admissions Committee	2016 – 2017
Member	SICE Promotion & Tenure Committee (School level)	2014 – 2015, 2015 – 2016
Member	Informatics Faculty hiring committee	2009 – 2010, 2010 – 2011, 2013 – 2014
ADR	Associate Dean of Research	2007 – 2009
Member	SICE School Visioning Task Force	2009
Member	SICE Structure Task Force	2008
Co-chair	Dept of Computer Science Diversity committee	2005 – 2006
Member	SICE Policy Committee	2005 – 2006
Member	SICE School Structure Committee	2005 – 2006
Organizer	Data and Search Informatics Seminar	2007 – 2009
Founder, Advisor	Women In Computing @ IU (WIC@IU)	2002 – 2007
Member	Computer Science Hiring Committee	2008 – 2011
Chair or member	Computer Science Ph.D. Qualifiers committee (chair or me)	2002, 2003, 2004, 2005, 2006, 2007, 2008, 2010
Member	Computer Science Faculty Hiring committee	2004 – 2005, 2008 – 2009
Member	Dean's Faculty Advisory committee	2002 – 2003
Organizer	Systems Seminar	2001 – 2006
Member	Computer Science Dept Graduate Education committee (GEC)	2006 – 2008
Member	Undergraduate Education Committee (UEC)	2004 – 2005, 2012 – 2013

## INSTRUCTIONAL MATERIAL, DATASETS, PRE-PRINTS, WORKSHOP REPORTS (51)

---

CITATION	TYPE
C.A. Stewart, V. Welch, T.G. Doak, T. Miller, B. Plale, et al. The Pervasive Technology Institute at 20: Two decades of success and counting. Scholarworks. <a href="http://hdl.handle.net/2022/22607">http://hdl.handle.net/2022/22607</a>	Project Report
Yu Luo and Beth Plale, Pilot evaluation of Collection API with PID Kernel Information (2019, Jul 3). ArXiv. arXiv:1905.02896v2 [cs.DL]	Preprint
Tobias Weigel, Beth Plale, Mark Parsons, Gabriel Zhou, Yu Luo, Ulrich Schwardmann, and Robert Quick. (2018, August 9). RDA Recommendation on PID Kernel Information (draft). Zenodo. <a href="http://doi.org/10.5281/zenodo.1462543">http://doi.org/10.5281/zenodo.1462543</a>	Preprint
Venice Margarette J Juanillas, Alexis Dereeper, Nicolas Beaume, Gaetan Droc, Joshua Dizon, John Robert Mendoza, Jon Peter Perdon, Locedie Mansueto, Lindsay Triplett, Jillian Lang, Gabriel Zhou, Kunalan Ratharanjan, Beth Plale, Jason Haga, Jan E Leach, Manuel Ruiz, Michael Thomson, Nickolai Alexandrov, Pierre Larmande, Tobias Kretzschmar, Ramil P Mauleon (2019), Rice Galaxy: an open resource for plant science, bioRxiv 358754; doi: <a href="https://doi.org/10.1101/358754">https://doi.org/10.1101/358754</a>	Preprint
Yu Luo, Kunalan Ratharanjan, Quan Zhou and Beth Plale (2018). Poster Abstract: Persistent IDs: Application to Workflow and Sensor Applications, <i>34th Pacific Rim Applications and Grid Middleware Assembly</i> , <a href="http://hdl.handle.net/2022/22313">http://hdl.handle.net/2022/22313</a>	Poster
Inna Kouper, Yu Luo, Isuru Suriarachchi, Beth Plale (2017). Poster Abstract: Provenance Enriched PID Kernel Information as OAI-ORE Map Replacement for SEAD Research Objects, <i>17<sup>th</sup> ACM-IEEE-CS Joint Conf. on Digital Libraries</i> , Toronto, CA Jun 2017	Poster
Beth Plale and Inna Kouper (2017). Tutorial: SEADTrain Data Analysis, <i>ESIP Summer 2017 meeting</i> , Bloomington IN, <a href="http://hdl.handle.net/2022/22312">http://hdl.handle.net/2022/22312</a>	Tutorial
Beth Plale (2016). Tutorial: Persistent Identifiers: their role in Data Quality. <i>Data Quality in Era of Big Data Workshop</i> , Bloomington, IN Sep 2016	Tutorial
Beth Plale, Matt Jones, Douglas Thain (2015), Workshop report: Software in Science: a Report of Outcomes of the 2014 National Science Foundation (NSF) Software Infrastructure for Sustained Innovation (SI2) Meeting, Final workshop report, <a href="http://hdl.handle.net/2022/19760">http://hdl.handle.net/2022/19760</a> , Mar 2015	Workshop report
Peng Chen and Beth Plale (2015). Big Data Provenance Analysis and Visualization, the <i>Doctoral Symposium of 15th IEEE/ACM Int'l Symp. on Cluster, Cloud and Grid Computing (CCGrid)</i> , 2015	Doctoral symposium
Beth Plale, Atul Prakash, Robert McDonald (2015). Final Project Report: The Data Capsule for Non-Consumptive Research, URI: <a href="http://hdl.handle.net/2022/19277">http://hdl.handle.net/2022/19277</a>	Project final report
Guangchen Ruan and Beth Plale (2014), Evaluation of Data Storage in HathiTrust Research Center Using Cassandra. <a href="http://hdl.handle.net/2022/18472">http://hdl.handle.net/2022/18472</a>	Preprint
K. Caylor, T. Evans, L. Estes, J. Sheffield, B. Plale, and S. Attari (2014). Impacts of agricultural decision making and adaptive management on food security in Africa. <i>AGU Fall Meeting Abstracts</i> , 1:06	Preprint
Quan Zhou, Beth Plale, Keith Danielson, Robert J. Ping, Janae Cummings, and Alan Mauro (2013). Demo: Daily Regional Weather Forecasts in Support of Vortex2, <i>IEEE Cluster 2013</i> , Indianapolis, IN, Sep 2013	Demo
Luo, Yuan, Plale, Beth, Jensen, Scott, Cheah, You-Wei, Conover, Helen (2012). <i>Provenance of AMSR-E Data from the National Snow and Ice Data Center</i>	Dataset

(NSIDC). OPM XML Ver. 1.1., September 2 - October 4, 2011. Bloomington, Indiana: Data To Insight Center. <a href="http://dx.doi.org/10.5967/M0F47M2D">http://dx.doi.org/10.5967/M0F47M2D</a> (Dataset 13MB)	
Quan Zhou, Beth Plale (2012). Vortex2 Visualization, DOI: 10.5976/MOST7MST	Dataset
Scott Jensen, Devarshi Ghoshal, and Beth Plale (2011), Evaluation of Two XML Storage Approaches for Scientific Metadata , <i>Indiana University Computer Science Technical Report TR698</i> .	Preprint
Scott Jensen, Beth Plale, John Cobb, Rebecca Koskella (2011). Big Data Means Your Metadata Must Work, <i>IEEE/ACM Supercomputing 2011</i> , half day tutorial, Seattle, WA Nov 2011. 61 attendees.	Tutorial
Chathura Herath, Kavitha Chandrasekar, Beth Plale (2011). Integration of CEP into Scientific Workflows, 2 hour tutorial at <i>ACM Distributed Event Based Systems (DEBS)</i> , Hawthorne, New York, 2011.	Tutorial
Jeff Cox and Beth Plale (2011), Improving Automatic Weather Observations with the Public Twitter Stream, <i>Indiana University Computer Science Technical Report TR691</i> .	Preprint
Scott Jensen and Beth Plale (2011), Schema-Independent and Schema-Friendly Scientific Metadata Management, <i>Indiana University Computer Science Technical Report TR689</i> . A (shorter) extended abstract of this paper was published in Proceedings of 4th IEEE Int'l Conf. on eScience.	Preprint
Eran Chinthaka, Suresh Marru, and Beth Plale (2009) Sigiri: Towards A Light-Weight Job Management System for Large Scale Systems, <i>Indiana University Computer Science Technical Report TR681</i> .	Preprint
Yiming Sun, Beth Plale, Chathura Herath, and Scott Jensen (2009). Tutorial: Event Processing in Weather Responsive Scientific Workflow Cyberinfrastructure: tutorial and abstract, <i>3<sup>rd</sup> Int'l Conference on Distributed Event Based Systems (DEBS)</i> , 2009.	Tutorial
Droegemeier, Kelvin, Beth Plale, Mohan Ramamurthy, and Craig Mattocks (2009). A New Approach for Using Web Services, Grids, and Virtual Organizations in Mesoscale Meteorological Research, <i>American Meteorological Society (AMS) 25<sup>th</sup> Conf. on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology</i> , Jan 2009	Preprint
Bin Cao, Girish Subramanian, Sribabhu Doddapaneni, and Beth Plale (2008). Poster: Provenance Collection in an Industry Biochemical Discovery Cyberinfrastructure, <i>4th Int'l IEEE Conf. on e-Science (e-Science)</i> , Indianapolis, IN pp. 424-425, IEEE Computer Society, DOI 10.1109/eScience.2008.104 Dec 2008	Poster
Beth Plale, You-Wei Cheah, and Yiming Sun (2008). <a href="#">Towards Quantifying Limits in Automated Curation of e-Science Data</a> , <i>Indiana University Computer Science Technical Report TR672</i> , Nov 2008.	Preprint
You-Wei Cheah and Beth Plale (2008). Representing LEAD Experiments in a FEDORA digital repository, <i>Indiana University Computer Science Technical Report TR666</i> , June 2008.	Preprint
Aparna Venkatraman, Vinay Pandey, Beth Plale, Shing-Shong Shei (2007). <a href="#">Benchmarking Effort of Virtual Machines on Multicore Machines</a> , <i>Indiana University Computer Science Technical Report TR654</i> , Master's Research Project, Dec 2007.	Preprint (class project)
Xiang, X. and B. Plale (2007). Performance Evaluation of MySQL 5.0 and Berkeley DB XML as a Grid Resource Information Manager (GRIM) with a Benchmark/Workload, <i>Indiana University Computer Science Technical Report TR-645</i> .	Preprint
Plale, Beth, Rahul Ramachandran, and Steve Tanner (2006). Data Management Support for Adaptive Analysis and Prediction of the Atmosphere in LEAD, <i>American</i>	Preprint

<i>Meteorological Society (AMS) 22<sup>nd</sup> Conf. on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology</i> , 2006.	
Droegemeier, K. K., V. Chandrasekar, R. Clark, D. Gannon, S. Graves, E. Joseph, M. Ramamurthy, R. Wilhelmson, K. Brewster, B. Domenico, T. Leyton, V. Morris, D. Murray, B. Plale, R. Ramachandran, D. Reed, J. Rushing, D. Weber, A. Wilson, M. Xue, S. Yalda (2005). Linked Environments for Atmospheric Discovery (LEAD): Architecture, Technology Roadmap and Deployment Strategy, <i>American Meteorological Society (AMS) 21<sup>st</sup> Conf. on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology</i> , Jan 2005.	Preprint
Clark, Richard D., Sepideh Yalda, Dennis Gannon, Beth Plale, and Tom Baltzer (2005). Integrating LEAD Research in Undergraduate Education, <i>American Meteorological Society (AMS) 22<sup>nd</sup> Conf. on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology</i> , Jan 2006.	Preprint (undergrad education)
Yogesh L. Simmhan, Beth Plale and Dennis Gannon (2005). <a href="#">A Survey of Data Provenance Techniques</a> , <i>Indiana University Computer Science Technical Report TR618</i> . Full length version of ACM SIGMOD Record, Sep 2005.	Preprint
Ying Liu, Beth Plale, and Nithya Vijayakumar (2005). Poster Abstract: Distributed Query Planner in the Calder System, <i>14<sup>th</sup> IEEE Int'l Symposium on High Performance Distributed Computing (HPDC)</i> , Research Triangle, NC, Jul 2005.	Poster
Nithya Vijayakumar, Ying Liu, and Beth Plale (2005). Poster Abstract: Calder: Enabling Grid Access to Data Streams, <i>14<sup>th</sup> IEEE Int'l Symposium on High Performance Distributed Computing (HPDC)</i> , Research Triangle, NC, Jul 2005.	Poster
Nithya Vijayakumar and Beth Plale (2005). dQUOBEC Event Channel Communication System, <i>Indiana University Computer Science Technical Report TR-614</i> .	Preprint
Liu, Y. B. Plale, and N. Vijayakumar (2005). Realization of GGF DAIS Data Service Interface for Grid Access to Data Streams, <i>Indiana University Computer Science Technical Report TR-613</i> .	Preprint
Droegemeier, K., V. Chandrasekar, R. Clark, D. Gannon, S. Graves, E. Joseph, M. Ramamurthy, R. Wilhelmson, K. Brewster, B. Domenico, T. Leyton, V. Morris, D. Murray, B. Plale, R. Ramachandran, D. Reed, J. Rushing, D. Weber, A. Wilson, M. Xue, S. Yalda (2004). Linked Environments for Atmospheric Discovery (LEAD): A Cyberinfrastructure for Mesoscale Meteorology Research and Education, <i>American Meteorological Society (AMS) 20<sup>th</sup> Conf. on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology (IIPS)</i> , Jan 2004.	Preprint
Vijayakumar, N. and B. Plale (2004). RS- Algo: an Algorithm for Improved Memory Utilization in Continuous Query System under Asynchronous Data Streams, <i>Indiana University Computer Science Technical Report TR-601</i> .	Preprint
Moad, C. and B. Plale (2004). Portal Access to Parallel Visualization of Scientific Data on the Grid, <i>Indiana University Computer Science Technical Report TR-593</i> .	Preprint
Ying Liu and Beth Plale (2003). Survey of Publish-Subscribe Event Systems, <i>Indiana University Computer Science Technical Report TR-574</i> .	Preprint
Gannon, D., G. Fox, M. Pierce, B. Plale, G. von Laszewski, C. Severance, J. Hardin, J. Alameda, M. Thomas, J. Boisseau 2003. Grid Portals: A Scientist's Access Point for Grid Services, <i>GGF Community Practice document, working draft 1</i> .	Preprint
Plale, B., C. Jacobs, Y. Liu, C. Moad, R. Parab, and P. Vaidya 2003. Benchmark Details of Synthetic Database Benchmark/Workload for Grid Resource Information, <i>Indiana University Computer Science Technical Report TR-583</i> .	Preprint

Peter Dinda, and Beth Plale (2003). Poster Abstract: A Unified Relational Approach to Grid Information Services, <i>23rd IEEE Int'l Conference on Distributed Computing Systems (ICDCS)</i> , May 2003	Poster
Plale, B., C. Jacobs, Y. Liu, C. Moad, R. Parab, P. Vaidya, and N. Vijaykumar (2003). Poster Abstract: Understanding Grid Resource Information Management through a Synthetic Database Benchmark/Workload, <i>Int'l Conference on High Performance Computing (HiPC)</i> , Hyderabad, India, Dec 2003.	Poster
Vaidya, P. and B. Plale (2003). Benchmark Evaluation of Xindice as a Grid Information Server, <i>Indiana University Computer Science Technical Report TR-585</i> .	Preprint
Plale, B. (2001). Performance Impact of Streaming Doppler Radar Data on Geospatial Visualization System, <i>Georgia Institute of Technology Technical Report GIT-CC-01-07</i> .	Preprint
Plale, B. and K. Schwan 2000. dQUOB: Managing Large Data Flows Using Dynamic Embedded Queries, <i>Georgia Institute of Technology Technical Report GIT-CC-00-07</i> .	Preprint
Isert, C., King, Schwan, K., Plale, B., and Eisenhauer, G. (1999). Poster Abstract: Steering Data Streams in Distributed Computational Laboratories, <i>19th IEEE Int'l Symposium on High Performance Distributed Computing (HPDC)</i> , IEEE Computer Science Press, Aug 1999.	Poster
Plale, B. and Schwan, K. (1998). Poster Abstract: Multi-level Steering in Distributed Laboratories, <i>Proceedings of SIGMETRICS Symposium on Parallel and Distributed Tools</i> , ACM Press, p. 162, Aug 1998.	Poster
Schroeder (Plale), B. and K. Schwan (1997). Language Issues in Hazard Detection Using Queries, <i>Georgia Institute of Technology Technical Report GIT-CC-97-36</i> .	Preprint
Schroeder (Plale), B., S. Aggarwal, and K. Schwan (1997). Software Approach to Hazard Detection Using On-line Analysis of Safety Constraints, <i>Georgia Institute of Technology Technical Report GIT-CC-97-01</i> .	Preprint