David Thau

CONTACT INFORMATION	956 York Street San Francisco, CA 94110	*E-mail:* davethau@gmail.com *Home:* (415) 282-3920

RESEARCH INTERESTS

Scientific data management, knowledge representation, data integration, biodiversity- and eco-informatics, and cognitive science.

EDUCATION

University of California, Davis

Ph.D. Computer Science, June, 2010
- Dissertation Topic: "Reasoning about Taxonomies"
- Advisor: Bertram Ludäscher

University of Michigan, Ann Arbor

M.Sc., Electrical Engineering and Computer Science, 1993
M.Sc., Psychology, 1991

University of California, Los Angeles

B.Sc., Cognitive Science, 1989

REFEREED PUBLICATIONS AND BOOK CHAPTERS

1. **Thau, D.**, 2010. Collaborative environmental in situ data collection: Experiences and opportunities for ambient data integration. *Proceedings of the 9th International Conference on Ontologies, Databases, and the Applications of Semantics (ODBASE)*, On the Move to Meaningful Internet Systems: OTM 2010. Lecture Notes in Computer Science.

2. **Thau, D.**, Bowers, S., 2010. Best effort data exchange of taxonomically organized data. 2nd International Workshop on New Trends in Information Integration (co-located with ICDE 2010).

3. **Thau, D.**, Bowers, S., Ludäscher, B., 2009. Merging sets of taxonomically organized data using concept mappings under uncertainty. *Proceedings of the 8th International Conference on Ontologies, Databases, and the Applications of Semantics (ODBASE)*, On the Move to Meaningful Internet Systems: OTM 2009, 1103-1120. Lecture Notes in Computer Science.

4. **Thau, D.**, Morris, R., White, S., 2009. Contemporary challenges in ambient data integration for biodiversity informatics. *Proceedings of the 2nd International Workshop on Ambient Data Integration (ADI)*, On the Move to Meaningful Internet Systems: OTM 2009 Workshops, 59-68. Lecture Notes in Computer Science

5. **Thau, D.**, Bowers, S., Ludäscher, B. 2009. On merging taxonomies under RCC-5 algebraic articulations. *Journal of Computer Science and Engineering*, 3(2), 109-126. KIISE.

6. **Thau, D.**, Bowers, S., Ludäscher, B., 2009. CleanTax: A framework for reasoning about taxonomies. *Proceedings of the AAAI Spring Symposium*, 49-50.

7. **Thau, D.**, Bowers, S., Ludäscher, B., 2008. Merging taxonomies under RCC-5 algebraic articulations. *Proceedings of the CIKM Workshop on Ontologies and Information Systems for the Semantic Web (ONISW)*, 47-54. ACM.

8. **Thau, D.**, 2008. Reasoning about taxonomies and articulations. *Proceedings of the EDBT Ph.D. Workshop*, 11–19. ACM.

9. **Thau, D.**, Ludäscher, B., 2007. Reasoning about taxonomies in first-order logic. *Ecological Informatics*, 2(3):195-209.

10. Bowers, S., **Thau, D.**, Williams, R., Ludäscher, B., 2004. Data procurement for enabling scientific workflows: On exploring inter-ant parasitism. *Proceedings of the VLDB International Workshop on the Semantic Web and Databases (SWDB)*, LNCS 3372, 5763.

11. **Thau, D.**, 1992. Primacy effects and selective attention in incremental clustering. Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society, 219 – 223. Lawrence Erlbaum Associates, Inc.

12. Medin, D.L., **Thau, D.**, 1991. Theories, constraints, and cognition. In *The Study of Cognition: Conceptual and Methodological Issues.* Herb Pick (Ed.). American Psychological Association.

BOOKS PUBLISHED   **Thau, D.**, 2006, 2000. The Book of JavaScript: A Practical Guide to Interactive Web Pages. No Starch Press. First edition: 404 pages. ISBN 978-1886411364. Second edition: 528 pages. ISBN 978-1593271060. Translated into Chinese, Czech, French, Italian, Japanese, Korean, Polish, Portuguese, and Russian.

INVITED PRESENTATIONS

Google at Yale: A Meeting of the Minds on Geotechnology. Organized and presented at a 2 day workshop on Geotechnology, Yale, February 2013.

Geo for Good. Organized and presented at a 3 day workshop on using Geotechnology for public benefit. Mountain Vew, 2012.

Terapixels for Everyone: Life in the Age of Massive Image Datasets. Invited keynote address for 2012 Annual Meeting of the American Society for Photogrammetry and Remote Sensing. Sacramento, 2012.

Mobile Data Collection. Invited presentation at the Annual Governor's Climate and Forest Taskforce meeting. Central Kalimantan, Indonesia, 2011.

Conference on Forest Monitoring and Biomass Estimation. Invited presentation at Woods Hole Research Center national policy workshop in Kigali, Rwanda, 2011.

Global validation in the Context of REDD+. Invited presentation at Global Land Cover Validation Workshop, Vienna, Austria, 2011.

Massively Parallel Environmental Analysis. Invited presentation at the Taxonomic Databases Working Group Annual Meeting, Woods Hole, Massachussettes, 2011.

Reasoning about Relationships Between Biological Taxonomies. An invited presentation to the University of California at Merced Cognitive Science Colloquium. University of California, Merced, 2009.

CleanTax: A Framework and System for Automated Reasoning with Taxonomies and Articulations. Talk presented at the Fourth Workshop on Data Integration in the Life Sciences, Philadelphia, Pennsylvania, 2007.

CleanTax: An Infrastructure for Reasoning about Biological Taxonomies. An invited presentation to the Stanford Research Institute. Stanford, California, 2007.

GUIDs and LSIDs. An invited presentation to the Taxonomic Data Working Group. Christchurch, New Zealand, 2004.

OWL and SDD. An invited presentation to the Taxonomic Data Working Group subgroup on the Structure of Descriptive Data. Federal Biological Research Centre for Agriculture and Forestry, Berlin, Germany, 2004.

Globally Unique Identifiers for Taxonomists. An invited presentation at the eScience GRID Environments conference. Napier University, Edinburgh, 2004.

Flexible Image Comparisons in AntWeb. An invited presentation at A Workshop on Establishing a Comprehensive Database for Plant Systematics. University of Florida, Gainesville, 2004.

Panelist in a Roundtable Discussion and Workshop on a "Biological Names Server." In the Trends and Developments in Biodiversity Informatics meeting sponsored by CRIA. Campinas, Brazil, 2003.

SERVICE	Editor for the Journal on Data Semantics	2010 - 2013
	Reviewer for the European Semantic Web Conferences	2009
	Reviewer for the Journal of Information Processing and Management	2009
	Reviewer for the International Conference on Data Engineering	2008
	Reviewer for the International Semantic Web Conference	2008
	Reviewer for the Journal of Web Semantics	2008
	Reviewer for the International Workshop on Data Integration in the Life Sciences	2008
	Reviewer for the International Conference on Web Reasoning and Rule Systems	2008
	Reviewer for the European Semantic Web Conferences	2007
	Reviewer for the Conference on Information and Knowledge Management	2007
	Reviewer for the Workshop on Service Oriented Technologies for Biological Databases	2007
	Reviewer for the International Conference on Very Large Data Bases	2006

Google, Inc., Mountain View, California

Developer Advocate **2010 − Present**
Teaching, project managing, and software engineering for Earth Engine, Google's geospatial analysis platform. Presented to thousands of people at academic, governmental, and commercial conferences and workshops. Developed and launched operational environmental monitoring systems.

California Academy of Sciences, San Francisco, California

Programming / Research **2003 − 2010**
Sole developer for AntWeb.org, a collaborative website dedicated to the collection and dissemination of scientific information about ants.

Natural History Museum, University of Kansas, Lawrence, Kansas
Computer Science Department, University of California, Davis, California

Programming / Research **2003 − 2008**
Acted as liaison between biologists and computer scientists in the SEEK project (The Science Environment for Ecological Knowledge), a five year NSF funded initiative to create cyberinfrastructure for ecological, environmental, and biodiversity research. Worked on ontologies for representing biological taxonomies, carried out research on, and developed a system for, merging biological taxonomies and integrating taxonomically organized data sets, and contributed to the code for the MetaCat XML metadata repository.

All Species Foundation, San Francisco, California

Director of Engineering **2002**
The All Species Foundation was a non-profit agency dedicated to accelerating the identification of new species. While there I created what was then the biggest database and search engine of species names, built a template-based system for the creation of species web pages, and undertook data integration work with the United States National Park Service.

biancaTroll Productions, San Francisco, California

Chief Executive Officer **2000 − 2001**
biancaTroll Productions, a company I co-founded in 1993, created some of the first 500 Web sites online, including the first web-based online community, a guide to Chicago, and websites for many political organizations. In 2000, the company was purchased and I officially became CEO, with all the responsibilities that entails (e.g., overall company direction, managing the staff, reporting to the Board of Directors). In 2001, the company was sold again, and I stepped down as CEO.

Wired Digital (HotWired), San Francisco, California

Director of Software Engineering (and other positions) **1995 − 1999**
As director of software engineering, I managed a team of up to 14 software developers who were responsible for the maintenance and creation of numerous software-related products at HotWired. These included the content management system, a number of database-driven websites, the advertising system, an early Java-based chat system, and a number of community-driven sites. I also oversaw the research and development of emerging products at Wired Digital. Among other projects, I oversaw the development of the first expert-system based advertising system and the integration of a number of artificial intelligence elements into the HotBot search engine.

Apollo Derivatives, Chicago, Illinois

Software Developer and Director of User Training **1995**
At Apollo, an options and derivatives trading organization, I wrote and supported several mission-critical programs ranging from an automated paging system to the company's trade resolution accounting system. I also developed the curriculum used by the company to train their traders how to use the company's network of NeXT computers.

Streams Online Media Development, Chicago, Illinois

Director of Software Development **1994**

As one of the founders of Streams, the first company in Chicago to develop websites for clients, I directed the creation of the software used in most of Streams' initial projects, wrote proposals, and acquired clients.

TEACHING

- **Instructor.** Ontologies in Ecology and Biodiversity Informatics, Pan-American Advanced Studies Institute, Costa Rica, June 2008.
- **Guest Lecturer.** Introduction to Programming, ECS-10, UC Davis, 2006.
- **Instructor.** Database Management Systems, Mills College, 2003.
- **Instructor.** XML and Related Technologies, San Francisco State University, 2003.
- **Instructor.** Building Online Communities, San Francisco State University, 1999-2000.
- **Instructor.** Beginning and Advanced JavaScript, San Francisco State University, 1998-2003.
- **Instructor.** Introduction to Programming, San Jose State University, 1998-1999.
- **Co-Founder, Instructor.** Humboldt Institute for Technological Studies (a K-12 technology literacy program), 1996-2002.
- **Teaching Assistant.** Computational Decision Making Algorithms, Northwestern University, 1994.
- **Visiting Lecturer.** Experimental Psychology, University of Michigan, 1994.
- **Teaching Assistant.** Laboratory in Experimental Psychology, University of Michigan, 1993.
- **Teaching Assistant.** Introduction to Cognitive Science, University of Michigan, 1991.
- **Teaching Assistant.** The Psychology of Thinking, University of Michigan, 1991.
- **Teaching Assistant.** Social Psychology, University of California, Los Angeles, 1989.

SKILLS

Programming Languages Expert in Java, Python, Perl, JavaScript, C, Lisp.
Good working experience in PHP, Ruby, Prolog, Objective-C, CLIPS, OPS5, Haskell, Snobol.

Operating Systems Expert in Linux, Mac OSX.
Good working experience with Windows.

Databases Expert in MySQL, Postgres.
Good working experience with Oracle, SQL Server, Sybase.

Management Managed many software development teams. Managed the creation, development and maintenance of high-traffic web sites.

<small>REFERENCES</small>

Bertram Ludäscher, Ph.D.
Professor, Department of Computer Sciences, University of California, Davis
Phone: (530) 754-8576
Email: `ludaesch@ucdavis.edu`
Role: Primary Graduate Advisor

Shawn Bowers, Ph.D.
Assistant Professor, Department of Computer Science, Gonzaga University
Phone: (509) 313-5712
Email: `bowers@gonzaga.edu`
Role: Colleague, Co-author

Robert A. Morris, Ph.D.
Professor of Computer Science (retired), University of Massachusetts, Boston
Phone: (857) 222-7992
Email: `morris.bob@gmail.com`
Role: Colleague, Co-author

Lydia Mann, Ph.D.
Salesforce
Phone: (510) 205-5879
Email: `lisanti.mann@sbcglobal.net`
Role: Professor of Education and Former student

Barbara Li Santi, Ph.D.
Professor of Mathematics and Computer Science, Mills College
Phone: (510) 430-2247
Email:`barbara@mills.edu`
Role: Department Chair when I taught at Mills

Brian Fisher, Ph.D.
Chairman, Department of Entomology, California Academy of Sciences
Phone: (415) 379 5314
Email: `bfisher@calacademy.org`
Role: Colleague, Employer

Stan Blum, Ph.D.
Research Information Manager, California Academy of Sciences
Phone: (415) 379-5189
Email:`sblum@calacademy.org`
Role: Colleague

Joel Truher
VP Product Development, Sustainable Spaces
Phone: (415) 269-2183
Email: `joel@truher.org`
Role: Former manager