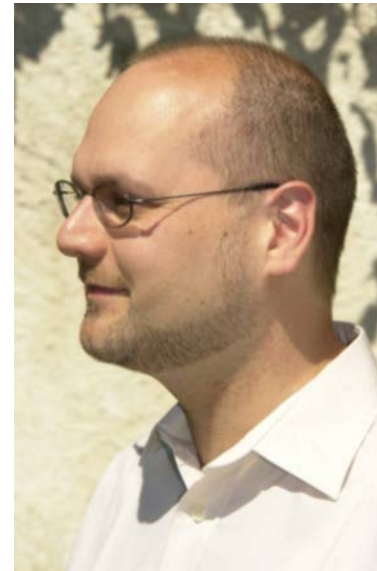


# Metadata Semantics: What Semantic Web technologies can contribute to scientific data and information sharing and discovery



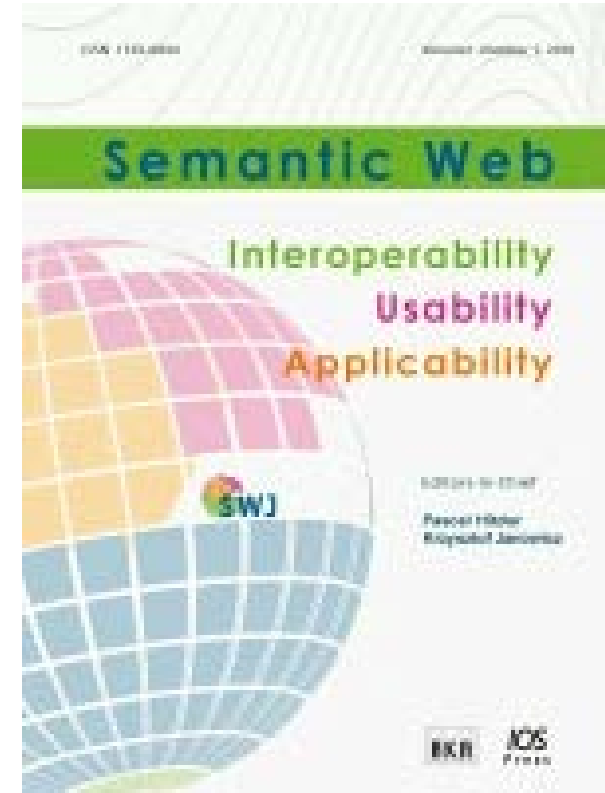
**Pascal Hitzler**

DaSe Lab for Data Semantics  
Wright State University  
<http://www.pascal-hitzler.de>

- **Since the rise of the World Wide Web, the role of publishing houses and scientific libraries is changing.**
- **Scientific publishing houses are redefining their roles and are investigating new revenue models.**
- **What exactly is the role of libraries?**
- **What will the role of libraries be in, say, 20 years?**

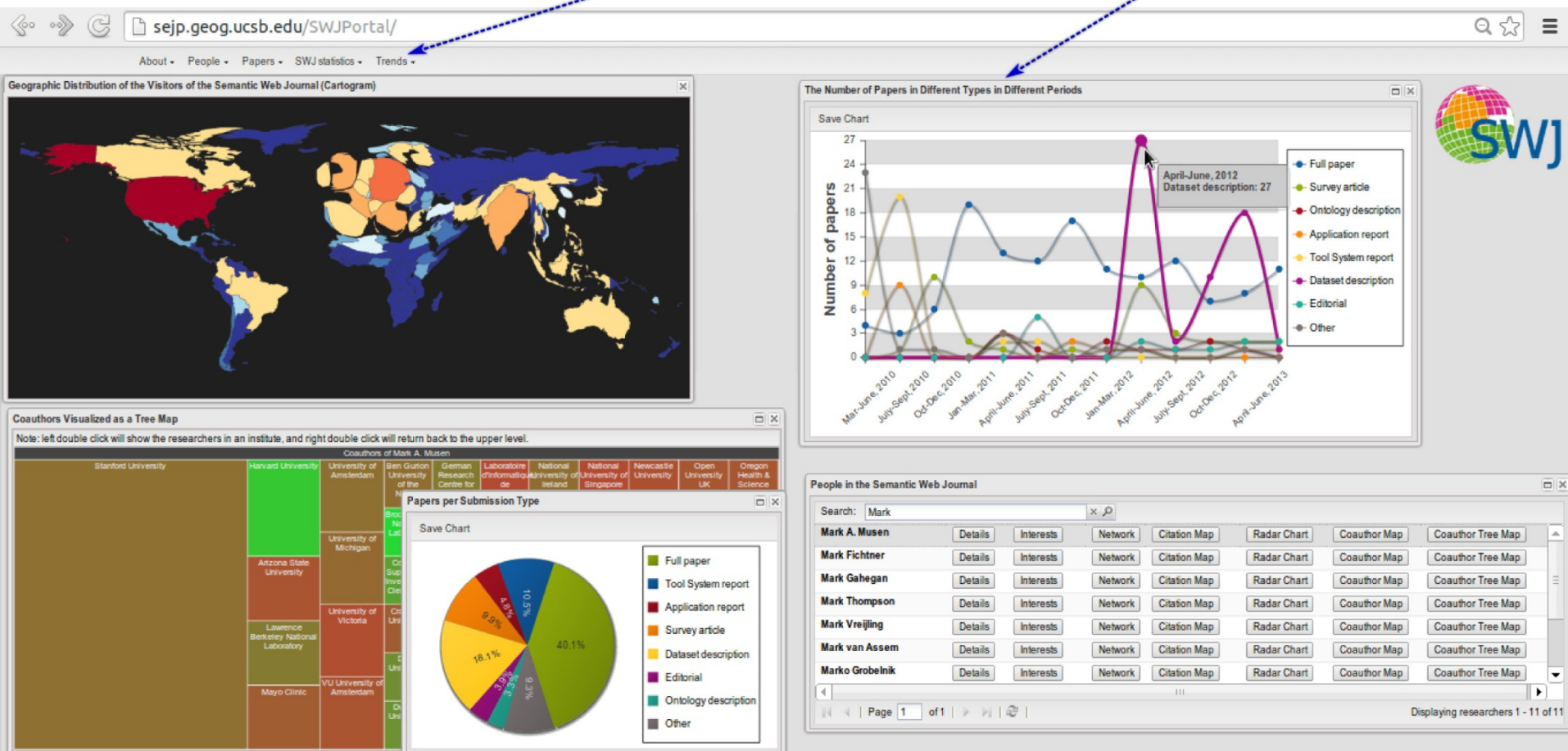
- **“I’m looking for an easy but introductory text on discrete mathematics suitable for computer scientists, with high quality in the mathematical formalization and notation, and including (besides the usual stuff) at least brief treatments of Russel’s paradox and of countable versus uncountable sets, e.g. uncountability of the real numbers.”**
- **“I’m looking for a textbook for a second-year introductory class on logic for computer scientists. Formal treatment of mathematics, tableaux algorithms for propositional and predicate logic, and preferably some coverage of datalog.”**

- **EiCs:** Pascal Hitzler  
Krzysztof Janowicz
- **Established 2010. Going strong.**
- **We very much welcome contributions at the “rim” of traditional Semantic Web research – e.g., work which is strongly inspired by a different field.**
- **Non-standard (open & transparent) review process.**
- **<http://www.semantic-web-journal.net/>**



Menu of modules

Individual rearrangeable modules



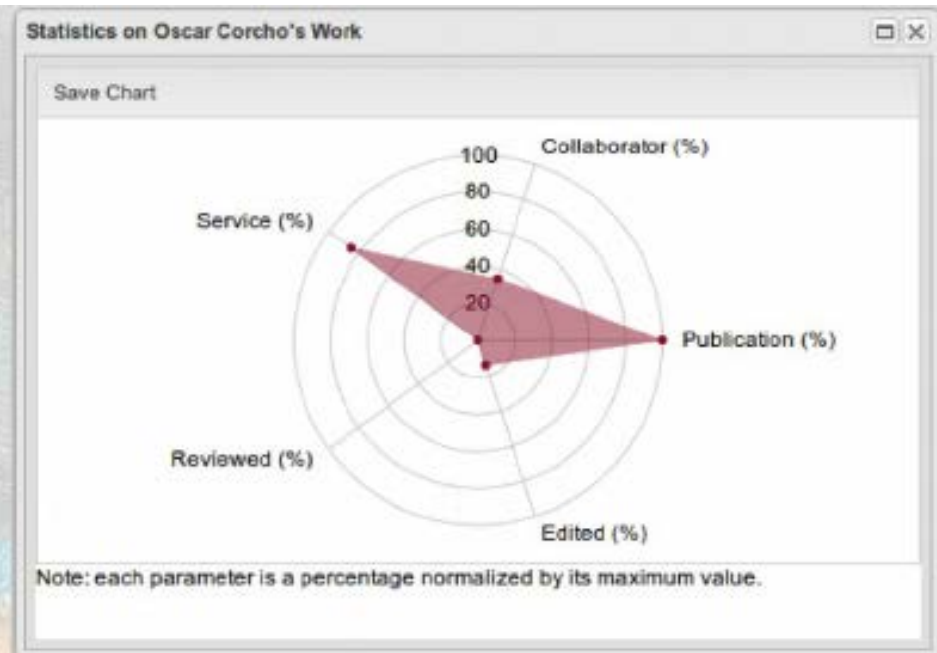
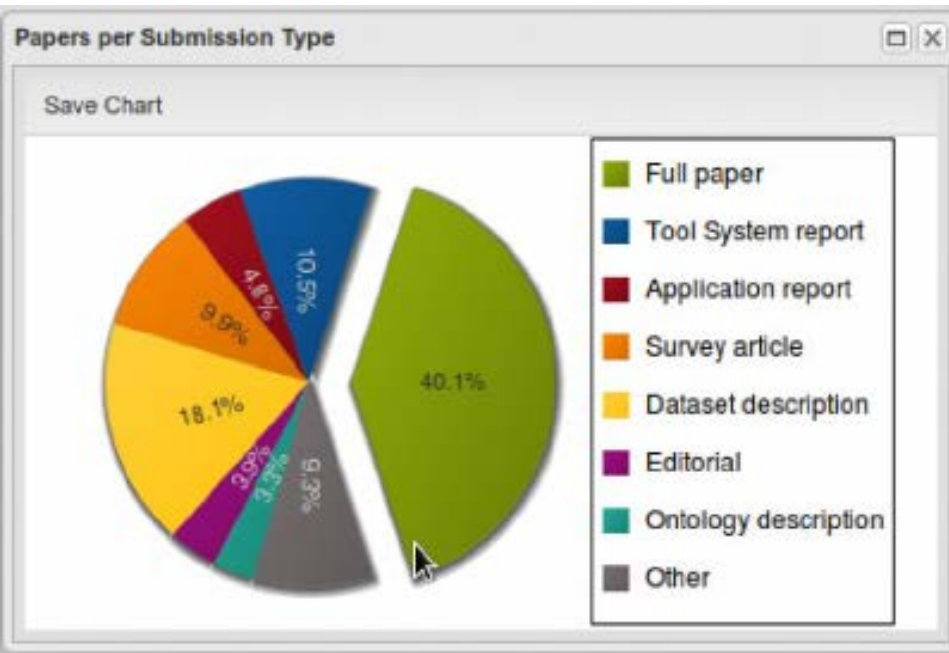
The screenshot displays the Semantic Web Journal portal at [sejp.geog.ucsb.edu/SWJPortal/](http://sejp.geog.ucsb.edu/SWJPortal/). The interface includes a navigation menu (About, People, Papers, SWJ statistics, Trends) and several interactive modules:

- Geographic Distribution of the Visitors of the Semantic Web Journal (Cartogram):** A map showing visitor density across different regions.
- The Number of Papers in Different Types in Different Periods:** A line chart showing the number of papers over time for various categories. A tooltip for April-June 2012 shows a dataset description count of 27.
- Coauthors Visualized as a Tree Map:** A treemap showing the distribution of coauthors by institution. A note indicates that double-clicking returns to the upper level.
- Papers per Submission Type:** A pie chart showing the distribution of papers across submission types.
- People in the Semantic Web Journal:** A list of researchers with search and filter options (Details, Interests, Network, Citation Map, Radar Chart, Coauthor Map, Coauthor Tree Map).

Submission Type	Percentage
Full paper	40.1%
Tool System report	10.5%
Application report	8.8%
Survey article	18.1%
Dataset description	2.8%
Editorial	3.9%
Ontology description	3.9%
Other	6.6%

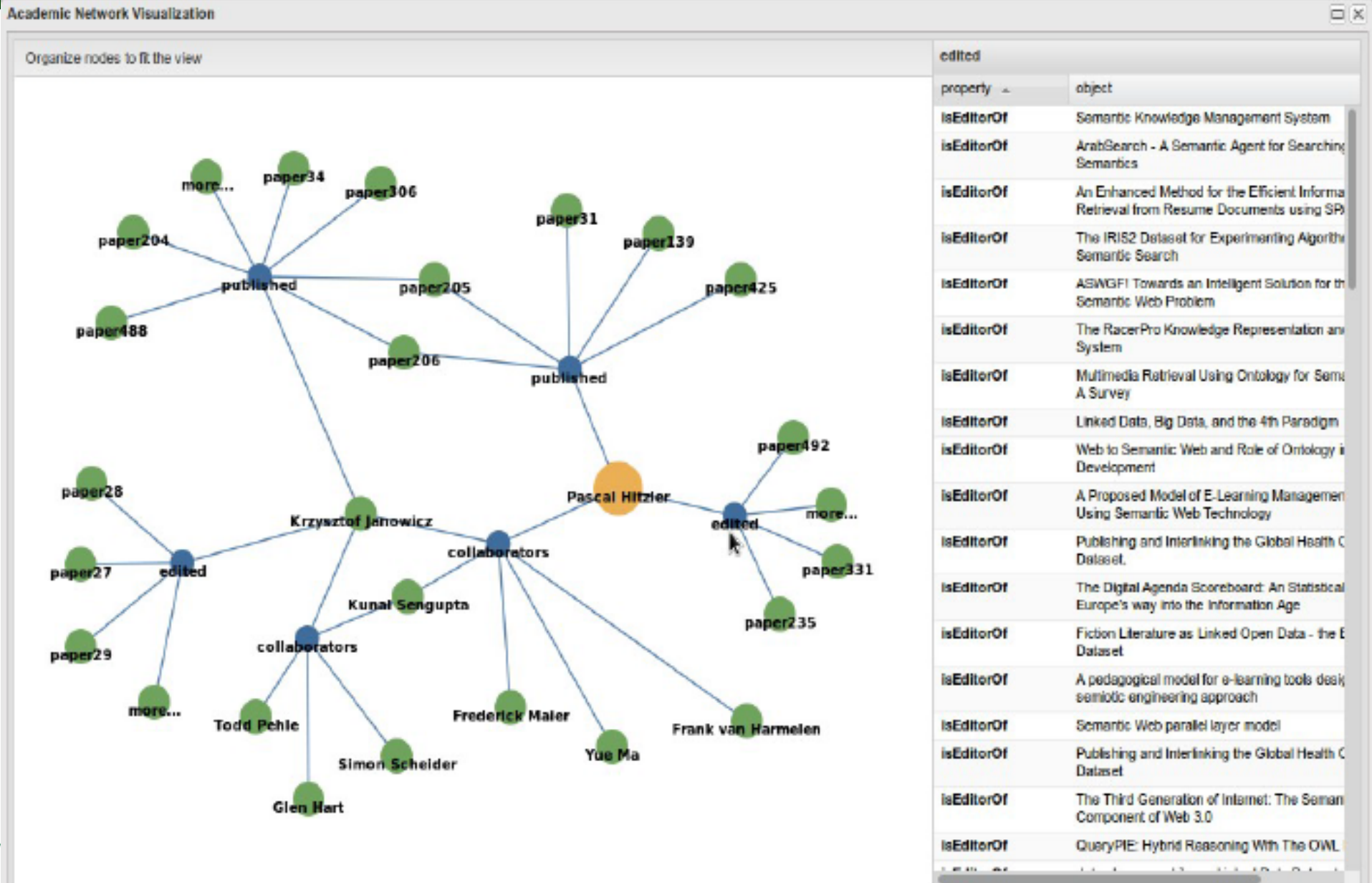
Researcher	Details	Interests	Network	Citation Map	Radar Chart	Coauthor Map	Coauthor Tree Map
Mark A. Musen	Details	Interests	Network	Citation Map	Radar Chart	Coauthor Map	Coauthor Tree Map
Mark Fichtner	Details	Interests	Network	Citation Map	Radar Chart	Coauthor Map	Coauthor Tree Map
Mark Gahegan	Details	Interests	Network	Citation Map	Radar Chart	Coauthor Map	Coauthor Tree Map
Mark Thompson	Details	Interests	Network	Citation Map	Radar Chart	Coauthor Map	Coauthor Tree Map
Mark Vreijing	Details	Interests	Network	Citation Map	Radar Chart	Coauthor Map	Coauthor Tree Map
Mark van Assem	Details	Interests	Network	Citation Map	Radar Chart	Coauthor Map	Coauthor Tree Map
Marko Grobelnik	Details	Interests	Network	Citation Map	Radar Chart	Coauthor Map	Coauthor Tree Map

# Summary Statistics





# Collaboration Networks







- **Provide analysis of citations, topic trends, research networks, etc., which can be obtained from (suitable!) metadata.**
- **Establish the social, economical and computational infrastructure to provide such data: open access, legal reusability of text and data, rich metadata (citations and beyond)**

# Data Discovery

Search

Return to classic map Start over

**Categories**

Select a category:

- People
- Programs
- Projects
- Deployments
- Platforms
- Instruments by type

**Category : People**

Wiebe, Peter Reset

- Werne, Josef
- Wertheimer, Alex
- Wethey, David
- Wheeler, Patricia
- White, Angelicque
- Whitledge, Terry

**You are looking at**

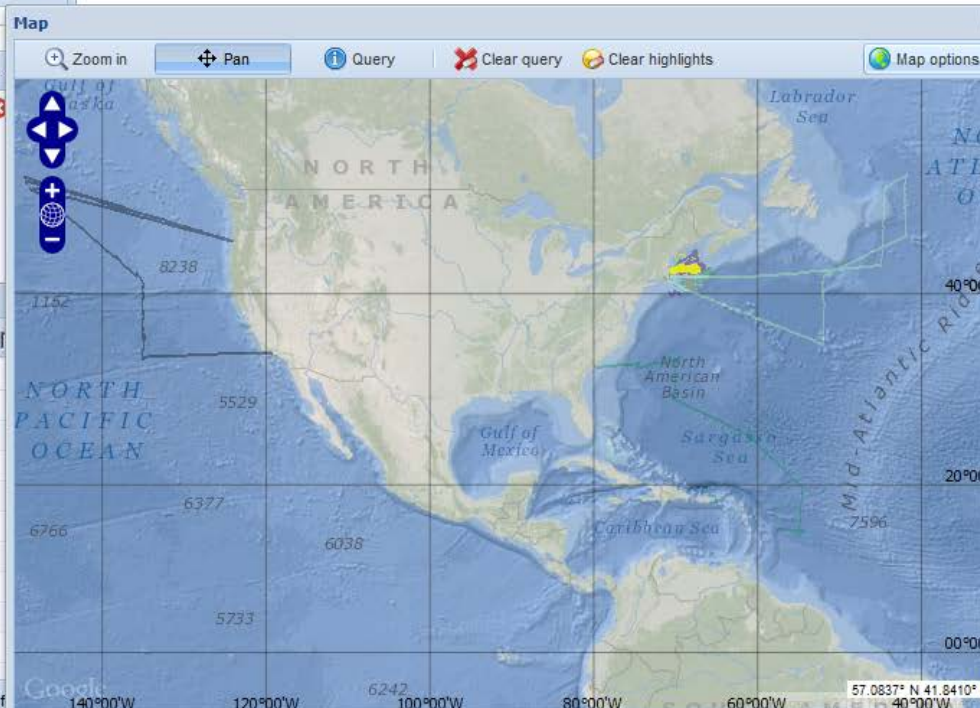
View or modify your current search.

**People**

- Wiebe, Peter

**Map**

Zoom in Pan Query Clear query Clear highlights Map options



The map displays the North American continent with a grid overlay. A yellow marker is placed in the North Atlantic Ocean, near the Mid-Atlantic Ridge. The map includes labels for the Gulf of Alaska, North Pacific Ocean, Gulf of Mexico, Caribbean Sea, North American Basin, Sargasso Sea, Labrador Sea, and Mid-Atlantic Ridge. Coordinates are shown at the bottom of the map.

**Visible deployments**

Highlight selected deployments on map?

Select deployment(s) to view dataset(s). Right-click a row for more options.

Deployment ID	Color
AL9205	Red
AL9404	Orange
AL9508	Pink
AL9607	Purple
AL9801	Brown
AL9906	Magenta
CT2010	Dark Purple
EN261	Red-Orange
EN307	Yellow
EN330	Light Yellow

Page 1 of 1 1 - 24 of

Determine if a **GMRT grid** contains high-resolution data from a ship's multibeam sonar in the proximity of a specified physiographic feature.

Return the **list of ship expeditions** that contributed high-resolution data to those grid cells.

For these expeditions, determine which, if any, are found in the **R2R** catalog and contain quality-controlled geophysical (gravity/magnetics) profiles along the same ship track.

Further determine which investigators are linked to those expeditions; which expeditions and investigators are linked to **journal publications and/or meeting abstracts** that contain thematic keywords pertaining to the physiographic feature; and what other data are available from the same expeditions in other repositories such as **BCO-DMO**.

# “Inside” and beyond the publications **aSe Lab**

- **Make paper *contents* available through rich metadata.**
- **Combine papers with data and datasets, and with information from “outside” the publishing process proper, such as funding awards, geographical information, affiliations, etc.**
- **More importantly, help in providing a social, economical and technological infrastructure where such information is provided to scientists and students.**

## **EarthCube:**

### **Developing a Community-Driven Data and Knowledge Environment for the Geosciences**

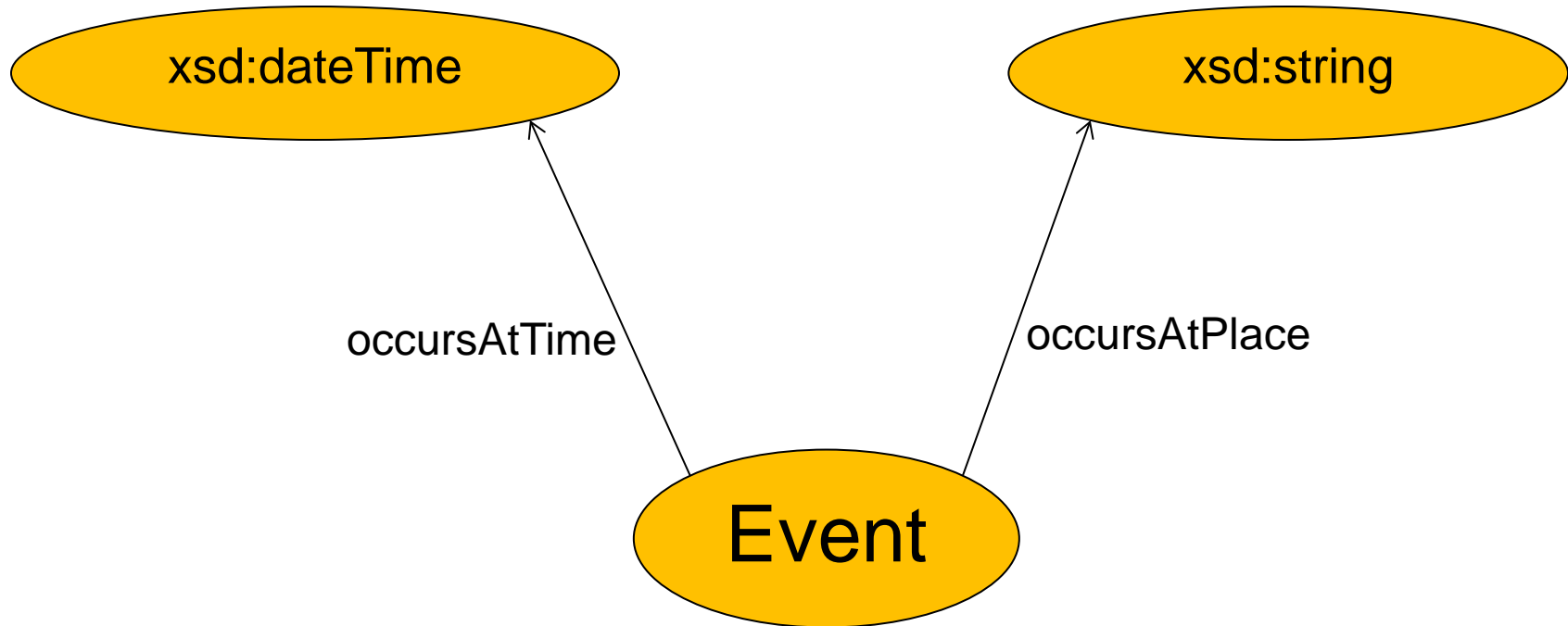
**“concepts and approaches to create integrated data management infrastructures across the Geosciences.”**

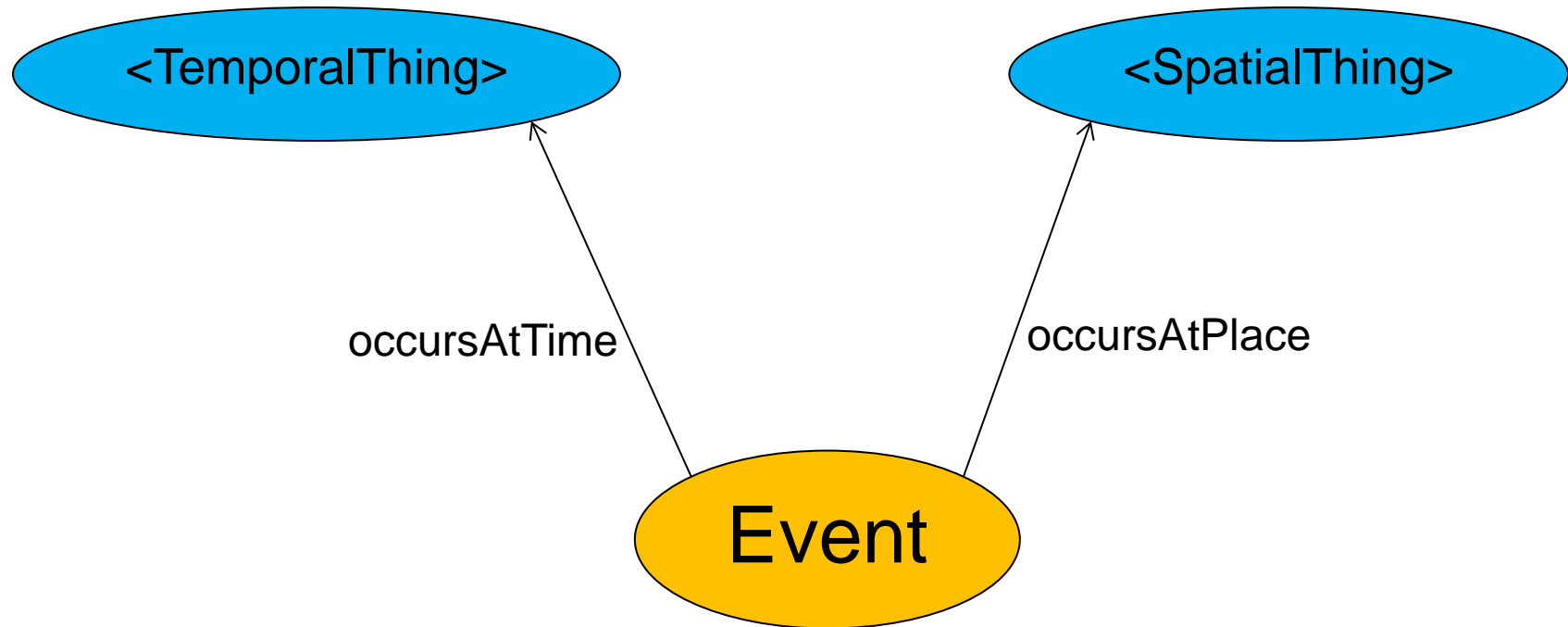
**“EarthCube aims to create a well-connected and facile environment to share data and knowledge in an open, transparent, and inclusive manner, thus accelerating our ability to understand and predict the Earth system.”**

- **An EarthCube Building Block**
- **Integrating ocean science repositories BCO-DMO and R2R, as well as datasets from the WHOI Library, AGU abstracts, NSF projects.**
- **Demonstrable added value (faceted integrated search).**
- **Key: extensible architecture that has the potential to grow to EarthCube size**

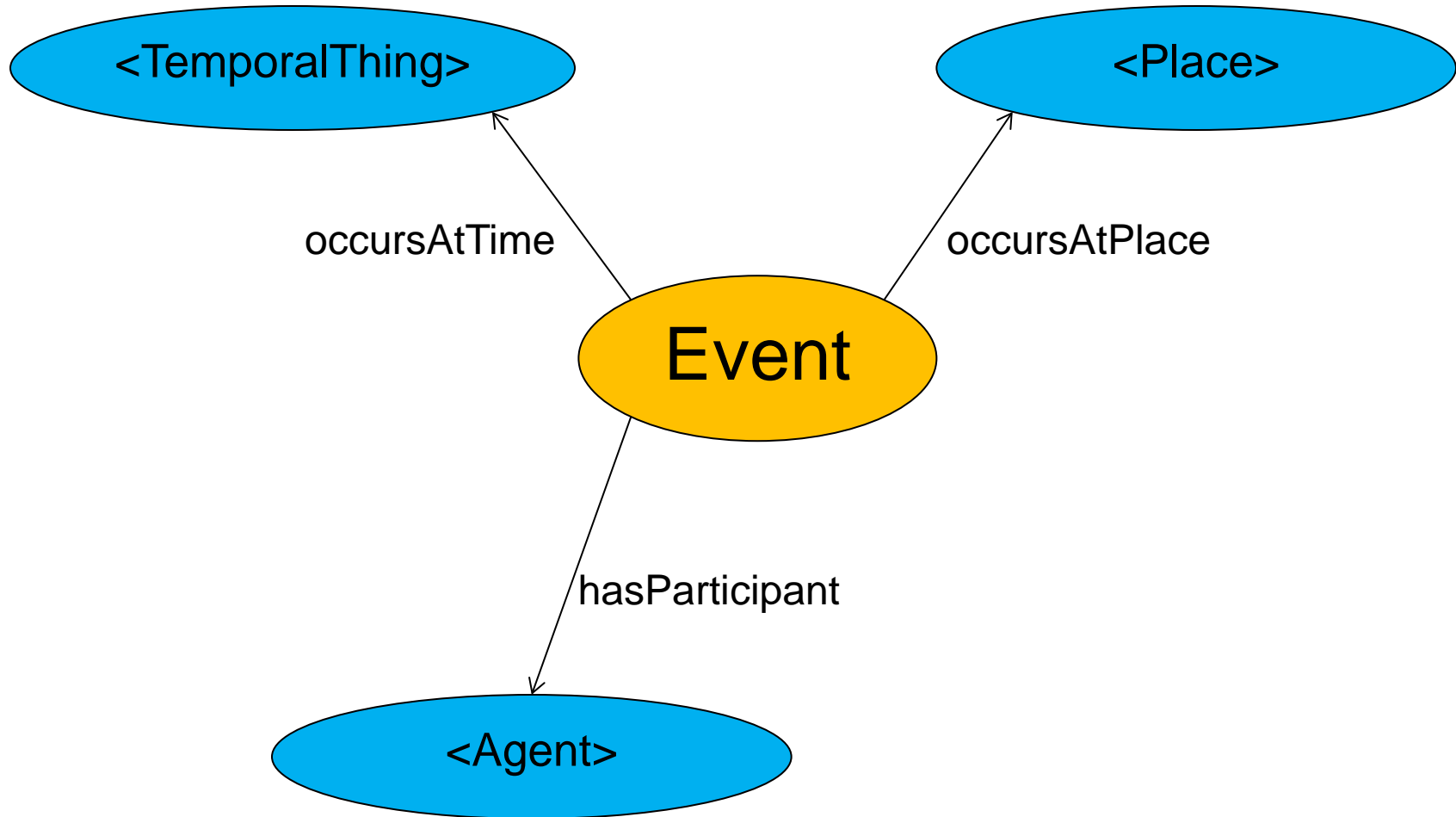
- **Well-established:**
  - using controlled vocabularies
  - which are standardized through a social process
  
- **How many vocabularies do you need to**
  - answer circumstantial queries?
  - cover all scientific paper contents?
  - even just to cover the earth sciences?
  
- **What do you do if scientific notions or perspectives change?**





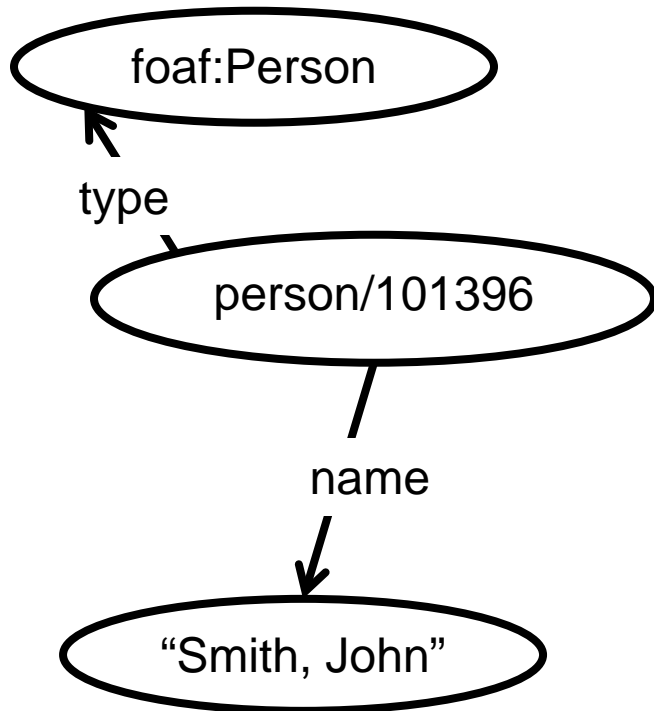


But what about events taking place in Second Life?

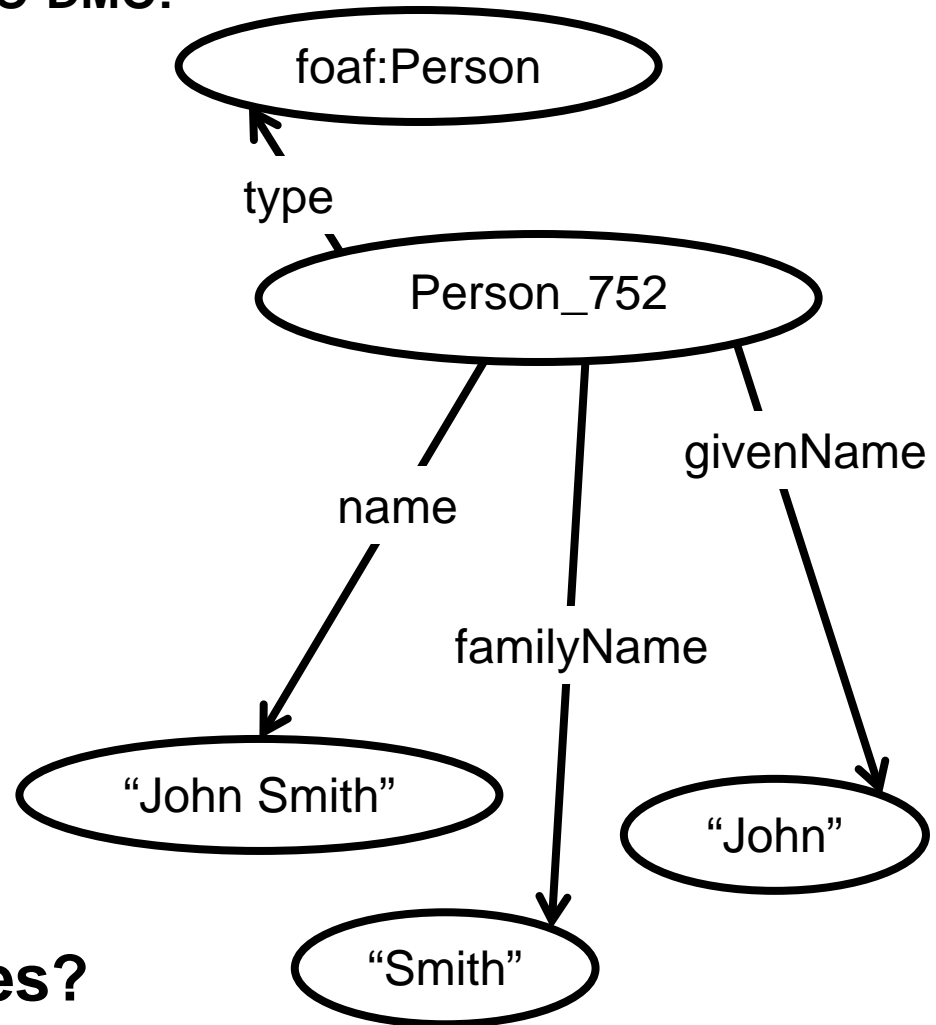


# Different representations

R2R:



BCO-DMO:



**What about other countries?**

- Research field in computer science.
- Took off in ca. the year 2000.
- Significant funding, initially DARPA, then large-scale in the EU.
- In the meantime, large international effort, with significant investment by funding agencies and companies.
- The Semantic Web *vision* is about seamless integration of data, knowledge, and services. It is not restricted to the WWW.
- The Semantic Web *approach* has (whatever type of) formal knowledge representation as a key ingredient.

- **Vocabularies on steroids.**
  - **Complex relationships between notions are part of the formal and machine-processable vocabulary definitions, e.g. “Every cruise must have a chief scientist who is PI on one of the research awards which pays for the expenses of the cruise.”**
- **Standardization of languages for defining vocabularies. E.g., the Web Ontology Language OWL.**
  - **Rather than standardizing vocabularies themselves.**
  - **Requires establishing best practices for defining and sharing vocabularies.**

- **Libraries could again be at the forefront of being providers for scientific information.**
- **Trends go towards integrated information spaces with a plethora of differing and heterogeneous information sources.**
- **How to organize this information space conceptually, technologically, and socially, is a key quest in the Big Data age.**

**Thanks!**



**Robert Arko, Columbia University**

**Suzanne Carbotte, Columbia University**

**Cynthia Chandler, Woods Hole Oceanographic Institution**

**Michelle Cheatham, Wright State University**

**Timothy Finin, University of Maryland, Baltimore County**

**Pascal Hitzler, Wright State University**

**Krzysztof Janowicz, University of California, Santa Barbara**

**Adila Krisnadhi, Wright State University**

**Thomas Narock, Marymount University**

**Lisa Raymond, Woods Hole Oceanographic Institution**

**Adam Shepherd, Woods Hole Oceanographic Institution**

**Peter Wiebe, Woods Hole Oceanographic Institution**

**Some of the presented work is part of the NSF *OceanLink* project:  
EarthCube Building Blocks, Leveraging Semantics and Linked Data  
for Geoscience Data Sharing and Discovery**

- Pascal Hitzler, Frank van Harmelen, A reasonable Semantic Web. *Semantic Web 1 (1-2)*, 39-44, 2010.
- Prateek Jain, Pascal Hitzler, Peter Z. Yeh, Kunal Verma, Amit P. Sheth, Linked Data is Merely More Data. In: Dan Brickley, Vinay K. Chaudhri, Harry Halpin, Deborah McGuinness: *Linked Data Meets Artificial Intelligence*. Technical Report SS-10-07, AAAI Press, Menlo Park, California, 2010, pp. 82-86. ISBN 978-1-57735-461-1. Proceedings of LinkedAI at the AAAI Spring Symposium, March 2010.
- Pascal Hitzler, Krzysztof Janowicz, *What's Wrong with Linked Data?* <http://blog.semantic-web.at/2012/08/09/whats-wrong-with-linked-data/> , August 2012.
- Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, *Foundations of Semantic Web Technologies*. Chapman and Hall/CRC Press, 2009.

- **Pascal Hitzler, Krzysztof Janowicz, Linked Data, Big Data, and the 4th Paradigm. Semantic Web 4 (3), 2013, 233-235.**
- **Krzysztof Janowicz, Pascal Hitzler, The Digital Earth as Knowledge Engine. Semantic Web 3 (3), 213-221, 2012.**
- **Gary Berg-Cross, Isabel Cruz, Mike Dean, Tim Finin, Mark Gahegan, Pascal Hitzler, Hook Hua, Krzysztof Janowicz, Naicong Li, Philip Murphy, Bryce Nordgren, Leo Obrst, Mark Schildhauer, Amit Sheth, Krishna Sinha, Anne Thessen, Nancy Wiegand, Ilya Zaslavsky, Semantics and Ontologies for EarthCube. In: K. Janowicz, C. Kessler, T. Kauppinen, D. Kolas, S. Scheider (eds.), Workshop on GIScience in the Big Data Age, In conjunction with the seventh International Conference on Geographic Information Science 2012 (GIScience 2012), Columbus, Ohio, USA. September 18th, 2012. Proceedings.**
- **Krzysztof Janowicz, Pascal Hitzler, Thoughts on the Complex Relation Between Linked Data, Semantic Annotations, and Ontologies. In: Paul N. Bennett, Evgeniy Gabrilovich, Jaap Kamps, Jussi Karlgren (eds.), Proceedings of the 6th International Workshop on Exploiting Semantic Annotation in Information Retrieval, ESAIR 2013, ACM, San Francisco, 2013, pp. 41-44.**

- Prateek Jain, Pascal Hitzler, Amit P. Sheth, Kunal Verma, Peter Z. Yeh, Ontology Alignment for Linked Open Data. In P. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Pan, I. Horrocks, B. Glimm (eds.), **The Semantic Web - ISWC 2010. 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010, Revised Selected Papers, Part I. Lecture Notes in Computer Science Vol. 6496. Springer, Berlin, 2010, pp. 402-417.**
- Amit Krishna Joshi, Prateek Jain, Pascal Hitzler, Peter Z. Yeh, Kunal Verma, Amit P. Sheth, Mariana Damova, Alignment-based Querying of Linked Open Data. In: Meersman, R.; Panetto, H.; Dillon, T.; Rinderle-Ma, S.; Dadam, P.; Zhou, X.; Pearson, S.; Ferscha, A.; Bergamaschi, S.; Cruz, I.F. (eds.), **On the Move to Meaningful Internet Systems: OTM 2012, Confederated International Conferences: CoopIS, DOA-SVI, and ODBASE 2012, Rome, Italy, September 10-14, 2012, Proceedings, Part II. Lecture Notes in Computer Science Vol. 7566, Springer, Heidelberg, 2012, pp. 807-824.**

- **Yingjie Hu, Krzysztof Janowicz, David Carral, Simon Scheider, Werner Kuhn, Gary Berg-Cross, Pascal Hitzler, Mike Dean, Dave Kolas, A Geo-Ontology Design Pattern for Semantic Trajectories. In: Thora Tenbrink, John G. Stell, Antony Galton, Zena Wood (Eds.): Spatial Information Theory - 11th International Conference, COSIT 2013, Scarborough, UK, September 2-6, 2013. Proceedings. Lecture Notes in Computer Science Vol. 8116, Springer, 2013, pp. 438-456.**
- **Yingjie Hu, Krzysztof Janowicz, Grant McKenzie, Kunal Sengupta, Pascal Hitzler, A Linked Data-driven Semantically-enabled Journal Portal for Scientometrics. In: H. Alani, L. Kagal, A. Fokoue, P. Groth, C. Biemann, J.X. Parreira, L. Aroyo, N. Noy, C. Welty, K. Janowicz (eds.), The Semantic Web - ISWC 2013. 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part II. Lecture Notes in Computer Science Vol. 8219, Springer, Heidelberg, 2013, pp. 114-129.**

- Prateek Jain, Peter Z. Yeh, Kunal Verma, Reymonrod G. Vasquez, Mariana Damova, Pascal Hitzler, Amit P. Sheth, Contextual Ontology Alignment of LOD with an Upper Ontology: A Case Study with Proton. In: Grigoris Antoniou, Marko Grobelnik, Elena Paslaru Bontas Simperl, Bijan Parsia, Dimitris Plexousakis, Pieter De Leenheer, Jeff Pan (Eds.): The Semantic Web: Research and Applications - 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29-June 2, 2011, Proceedings, Part I. Lecture Notes in Computer Science 6643, Springer, 2011, pp. 80-92.
- Prateek Jain, Pascal Hitzler, Kunal Verma, Peter Yeh, Amit Sheth, Moving beyond sameAs with PLATO: Paronymy detection for Linked Data. In: Ethan V. Munson, Markus Strohmaier (Eds.): 23rd ACM Conference on Hypertext and Social Media, HT '12, Milwaukee, WI, USA, June 25-28, 2012. ACM, 2012, pp. 33-42.

- **Sebastian Rudolph, Markus Krötzsch, Pascal Hitzler, Cheap Boolean Role Constructors for Description Logics. In: Steffen Hölldobler and Carsten Lutz and Heinrich Wansing (eds.), Proceedings of 11th European Conference on Logics in Artificial Intelligence (JELIA), volume 5293 of LNAI, pp. 362-374. Springer, September 2008.**
- **Adila Alfa Krisnadhi, Frederick Maier, Pascal Hitzler, OWL and Rules. In: A. Polleres, C. d'Amato, M. Arenas, S. Handschuh, P. Kroner, S. Ossowski, P.F. Patel-Schneider (eds.), Reasoning Web. Semantic Technologies for the Web of Data. 7th International Summer School 2011, Galway, Ireland, August 23-27, 2011, Tutorial Lectures. Lecture Notes in Computer Science Vol. 6848, Springer, Heidelberg, 2011, pp. 382-415.**
- **Adila Krisnadhi, Robert Arko, Suzanne Carbotte, Cynchia Chandler, Michelle Cheatham, Timothy Finin, Pascal Hitzler, Krzysztof Janowicz, Thomas Narock, Lisa Raymond, Adam Shepherd, Peter Wiebe, An Ontology Pattern for Oceanographic Cruises: Towards an Oceanographer's Dream of Integrated Knowledge Discovery. OceanLink Technical Report 2014.1.**