Ontologies in a Data-driven World

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

Textbook

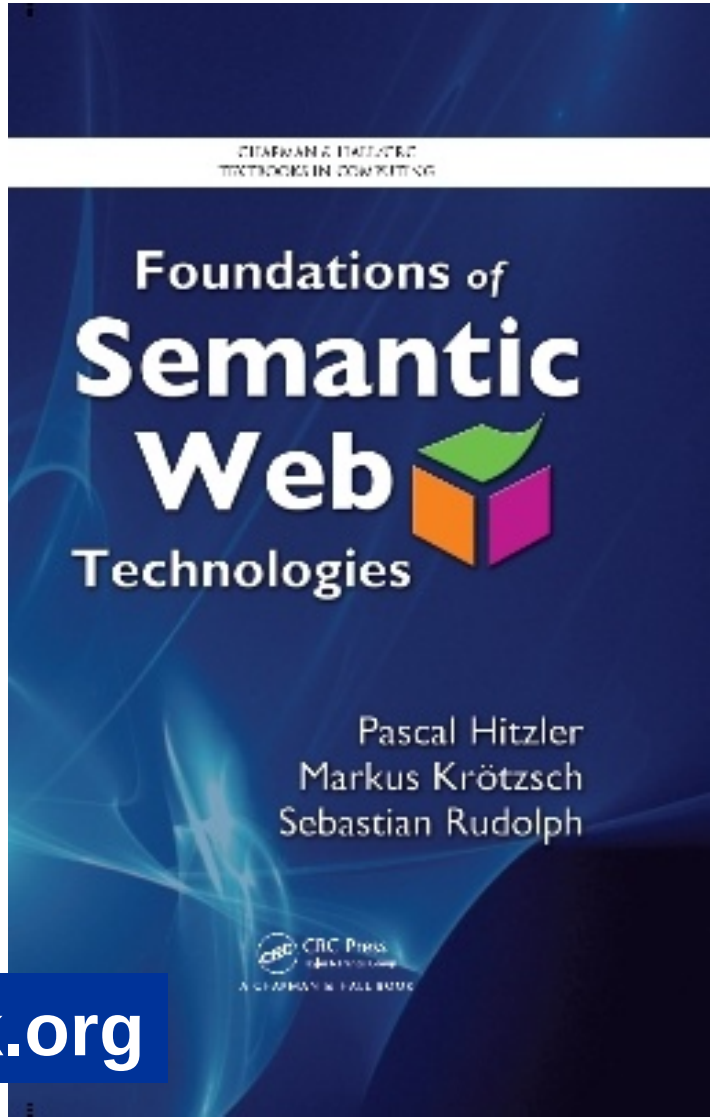
DaSe Lab

Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph

Foundations of Semantic Web Technologies

Chapman & Hall/CRC, 2010

Choice Magazine Outstanding Academic Title 2010 (one out of seven in Information & Computer Science)



http://www.semantic-web-book.org

Semantic Web journal

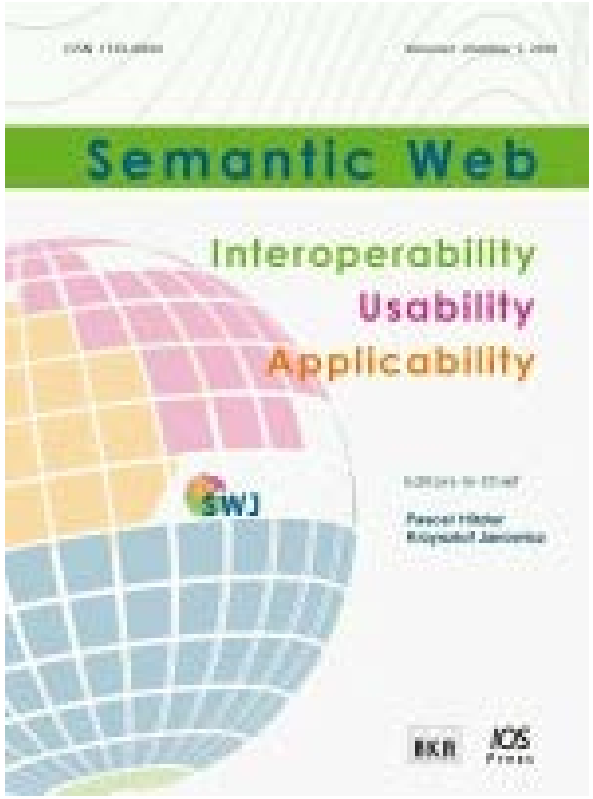
DaSe Lab

- **EiCs:** **Pascal Hitzler**
  Krzysztof Janowicz

- **New journal with significant initial uptake.**

- **We very much welcome contributions at the "rim" of traditional Semantic Web research – e.g., work which is strongly inspired by a different field.**

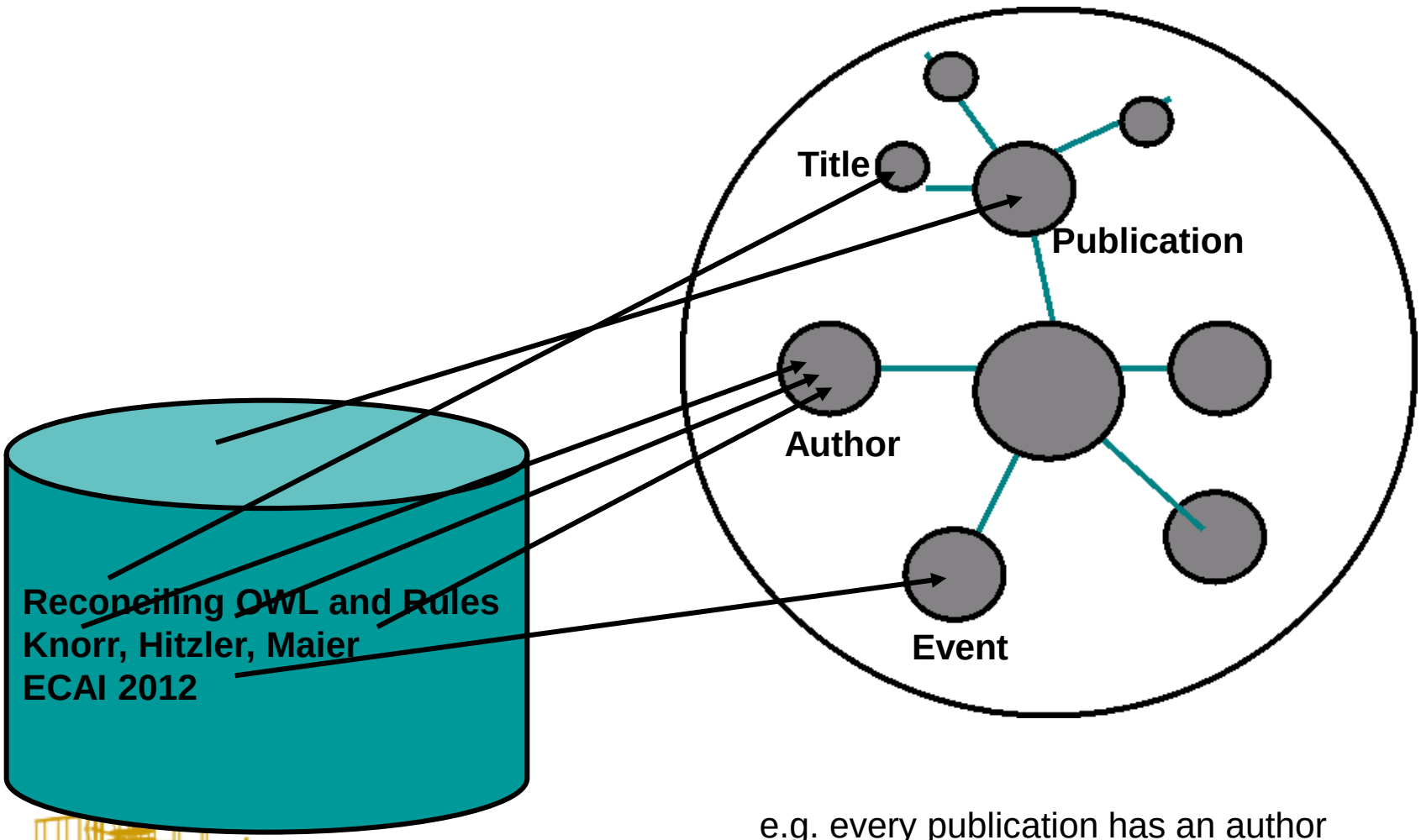
- **Non-standard (open & transparent) review process.**



- **http://www.semantic-web-journal.net/**

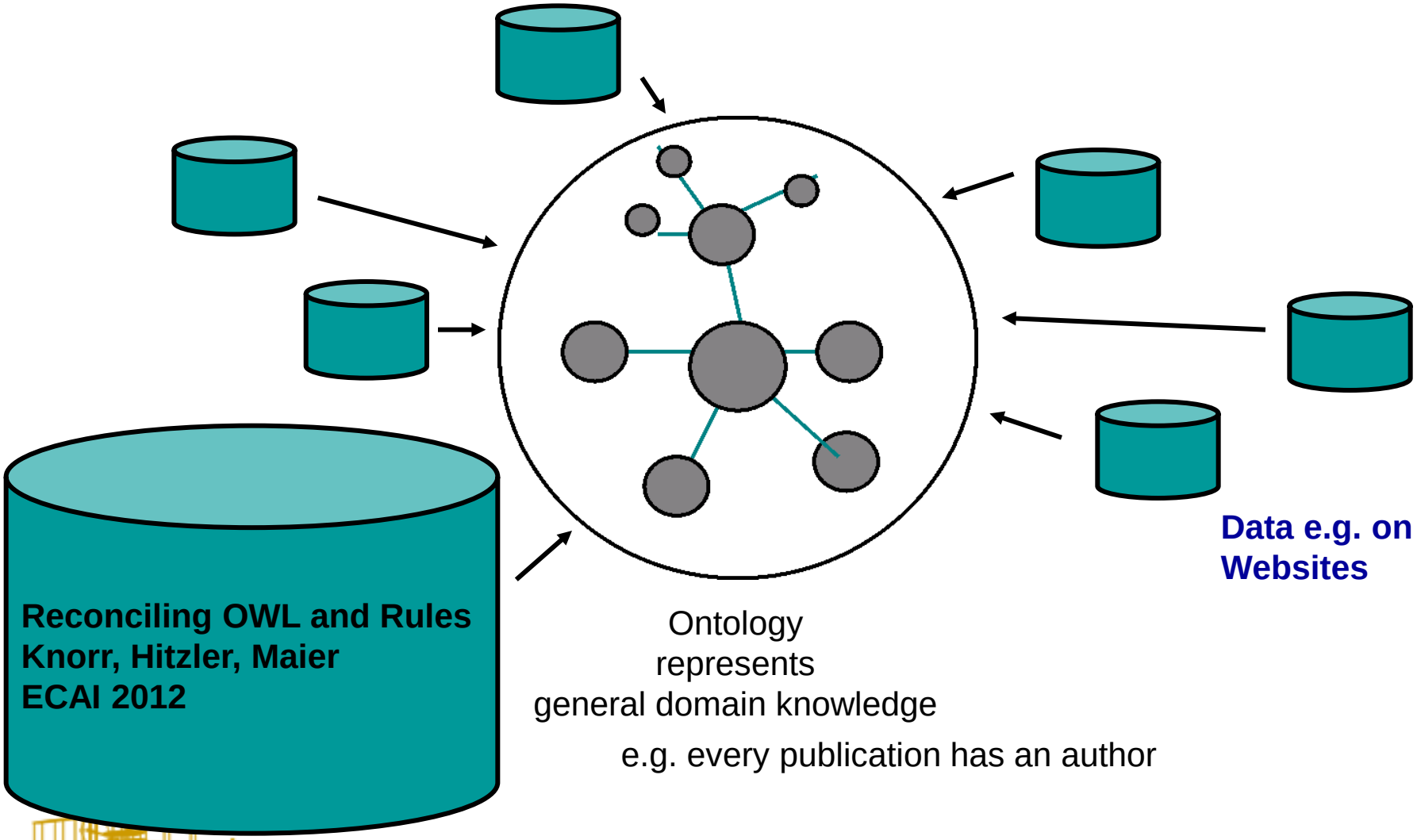
Ontologies?

A Basic Idea of the Semantic Web



e.g. every publication has an author

A Basic Idea of the Semantic Web



The ontology hype

- Large, well-thought-out ontologies (foundational/domain/etc).
- Networked, interlinked ontologies

- "You just have to get your formal definitions right, and a lot of the rest will just fall into place."

  – But this does not even work for
    - scientists
    - wanting to share and reuse scientific data
    - through well-kept data repositories

  – So how is this supposed to work for the web at large?

Multiple perspectives

- Try to find a universal definition for
  - Forest
  - Mountain
  - City
  - River

  - Etc.

- The stronger our ontological commitments, the more we loose reusability.

- We need to accept that conceptualizations are often very local, resulting in “micro-ontologies”.

Multiple perspectives

DaSe Lab

$$\text{a:flowsInto} \sqsubseteq \text{a:IsConnected} \quad (1)$$
$$\text{a:IrrigationCanal} \sqsubseteq \text{a:Canal} \quad (2)$$
$$\exists \text{a:flowsInto.a:AgriculturalField} \sqsubseteq \text{a:IrrigationCanal} \quad (3)$$
$$\text{a:Waterbody} \sqcap \text{a:Land} \sqsubseteq \bot \quad (4)$$
$$\text{a:AgriculturalField} \sqsubseteq \text{a:Land} \quad (5)$$

$$\text{b:flowsInto} \sqsubseteq \text{b:IsConnected} \quad (6)$$
$$\text{b:Canal} \sqsubseteq (\geq 2\ \text{b:IsConnected.b:Waterbody}) \quad (7)$$
$$\text{b:IrrigationCanal} \equiv (=1\ \text{b:isConnected.b:Waterbody}) \sqcap (=1\ \text{b:flowsInto.b:AgriculturalField}) \quad (8)$$

Two ontologies.

Left: transportation domain

Right: agriculture domain

We cannot simply equate a:Canal and b:Canal !

The well-done ontologies

- **Brittle**
- **Expensive**
- **Sometimes unintuitive**
- **Unwieldy**
- **Single-perspective**
- **Difficult to reuse**

- **Work in some contexts.**
- **Work if a lot of central control is imposed.**
- **Take a lot of manpower.**

Pre-LOD Semantic Web

- **Foundational ontologies**
- **Networked ontologies**
- **Sophisticated ontology languages**

Scientific Hypothesis:

These will solve your data and information management problems

Remember that scientific progress is fundamentally about falsification, not verification ☺

Linked Data?

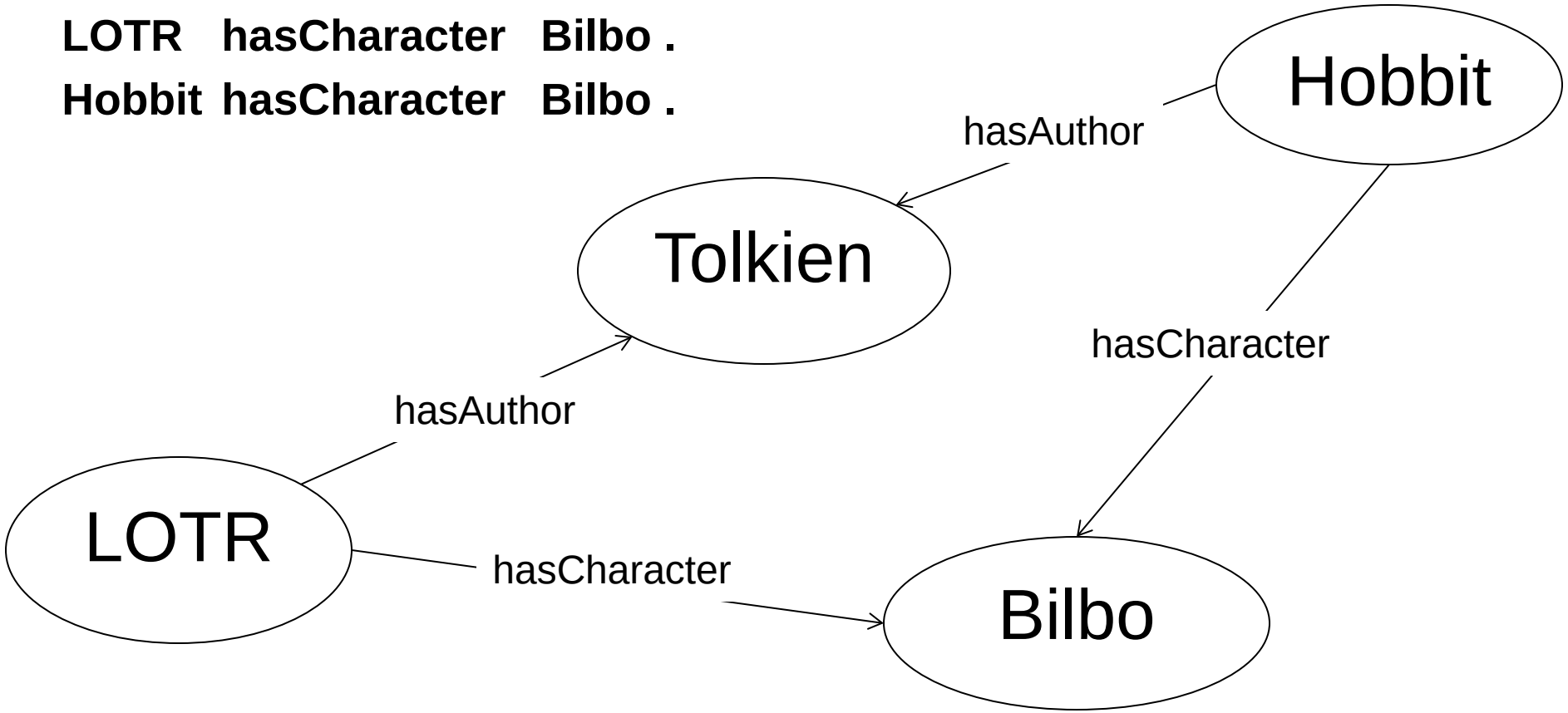
The linked data counter-hype

- **“Ontologies don’t work, let’s just link data”**
- **“Okay, with a little bit of ontologies on top.”**
- **“The Linked Data Web is the true Semantic Web.”**

Information as RDF graph

DaSe Lab

LOTR hasAuthor Tolkien .
Hobbit hasAuthor Tolkien .
LOTR hasCharacter Bilbo .
Hobbit hasCharacter Bilbo .



DBpedia: LOTR page

dbpedia-owl:thumbnail	▪ http://upload.wikimedia.org/wikipedia/commons/thumb/6/62/Jrrt_lotr_cover_design.jpg/200px-Jrrt_lotr_cover_design.jpg
dbpedia-owl:wikiPageExternalLink	▪ http://lotr.wikia.com ▪ http://www.glyphweb.com/arda/ ▪ http://www.tolkienlibrary.com/ ▪ http://www.tolkien.co.uk/ ▪ http://www.houghtonmifflinbooks.com/features/lordoftheringstrilogy/
dbpprop:author	▪ dbpedia:J._R._R._Tolkien
dbpprop:books	▪ dbpedia:The_Two_Towers ▪ dbpedia:The_Return_of_the_King ▪ dbpedia:The_Fellowship_of_the_Ring ▪ "Volumes:"
dbpprop:country	▪ England
dbpprop:expiry	▪ 20 (xsd:integer)
dbpprop:genre	▪ dbpedia:Adventure_novel ▪ dbpedia:High_fantasy
dbpprop:hasPhotoCollection	▪ http://www4.wiwiss.fu-berlin.de/flickrwrappr/photos/The_Lord_of_the_Rings
dbpprop:imageCaption	▪ Tolkien's own cover designs for the three volumes
dbpprop:language	▪ English
dbpprop:mediaType	▪ Print
dbpprop:name	▪ The Lord of the Rings
dbpprop:pages	▪ 1216 (xsd:integer)
dbpprop:precededBy	▪ dbpedia:The_Hobbit
dbpprop:pubDate	▪ 21 (xsd:integer)
dbpprop:publisher	▪ dbpedia:Allen_&_Unwin
dbpprop:small	▪ yes
dbpprop:wikiPageUsesTemplate	▪ dbpedia:Template:Infobox_book_series ▪ dbpedia:Template:Pp-vandalism
dcterms:subject	▪ category:Monomyths ▪ category:High_fantasy_novels ▪ category:Middle-earth_books ▪ category:British_fantasy_novels ▪ category:Fantasy_books_by_series ▪ category:1950s_fantasy_novels ▪ category:Sequel_novels ▪ category:The_Lord_of_the_Rings ▪ category:English_novels

Linked Data: Volume

DaSe Lab

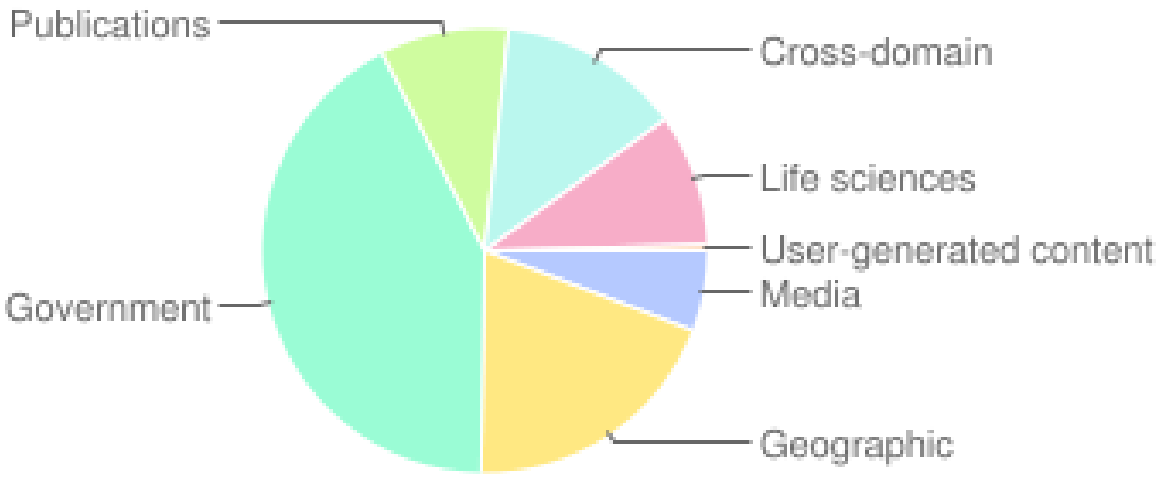
Number of Datasets

2011-09-19	**295**
2010-09-22	**203**
2009-07-14	**95**
2008-09-18	**45**
2007-10-08	**25**
2007-05-01	**12**

Number of triples (Sept 2011)

31,634,213,770

with 503,998,829 out-links



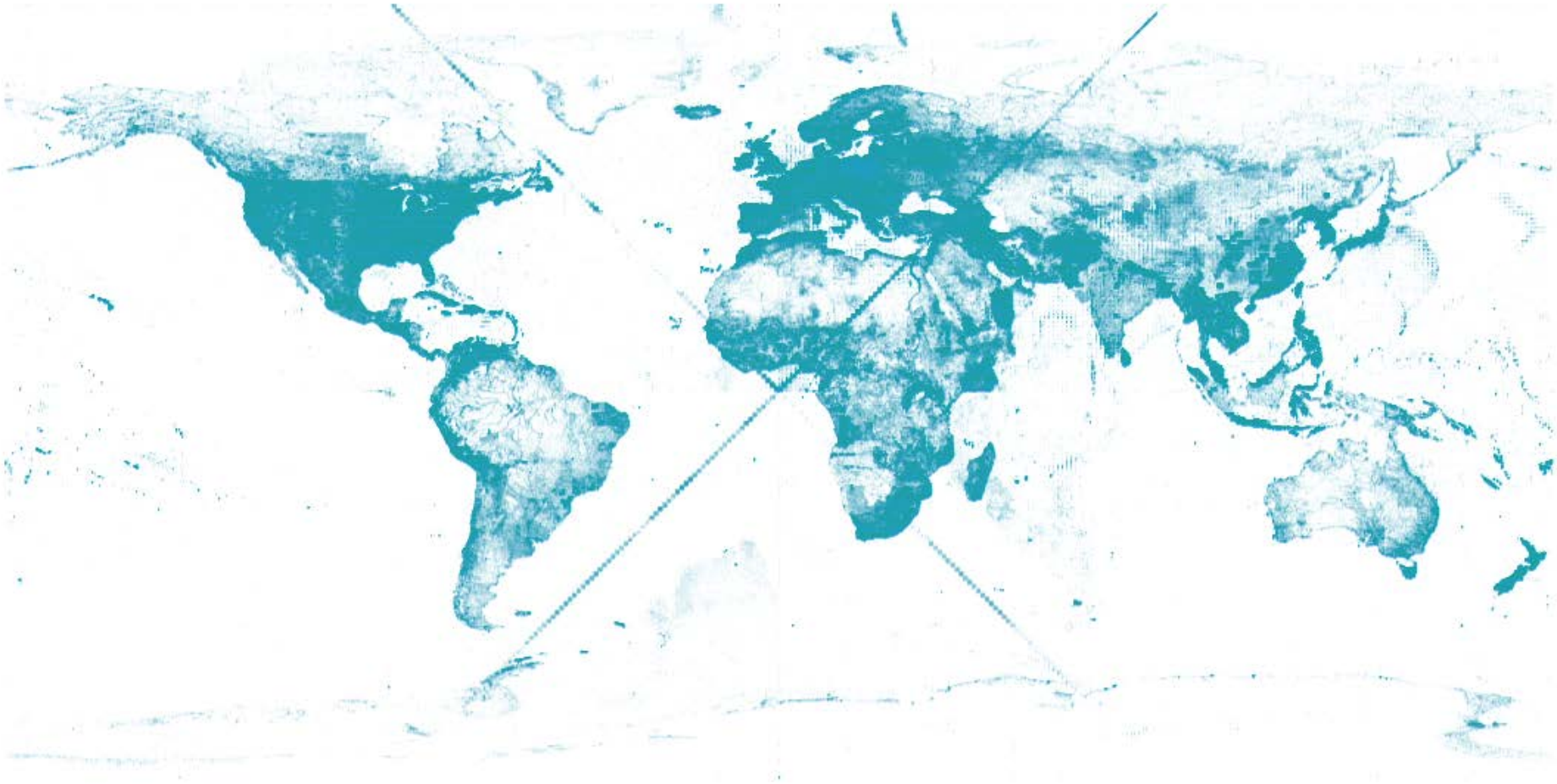
From http://www4.wiwiss.fu-berlin.de/lodcloud/state/

Linked Data: Volume

DaSe Lab

Geoindexed Linked Data – courtesy of Krzysztof Janowicz

http://stko.geog.ucsb.edu/location_linked_data



Linked Data: Volume

DaSe Lab

October 2013:

Ca. 25,000,000,000 schema.org references on the web.

15% of all pages now have schema.org markup.

That's just schema.org references …

Example querying LoD

DaSe Lab

"Identify congress members, who have voted "No" on pro environmental legislation in the past four years, with high-pollution industry in their congressional districts."

In principle, all the knowledge is there:

- **GovTrack**
- **GeoNames**
- **DBPedia**
- **US Census**

But even with LoD we cannot answer this query.

Example querying LoD

DaSe Lab

"Identify congress members, who have voted "No" on pro environmental legislation in the past four years, with high-pollution industry in their congressional districts."

Some missing puzzle pieces:

- Where is the data?
  - GovTrack
    GeoNames
    US Census
    requires intimate knowledge of the LoD data sets

Example querying LoD

DaSe Lab

"Identify congress members, who have voted "No" on pro environmental legislation in the past four years, with high-pollution industry in their congressional districts."

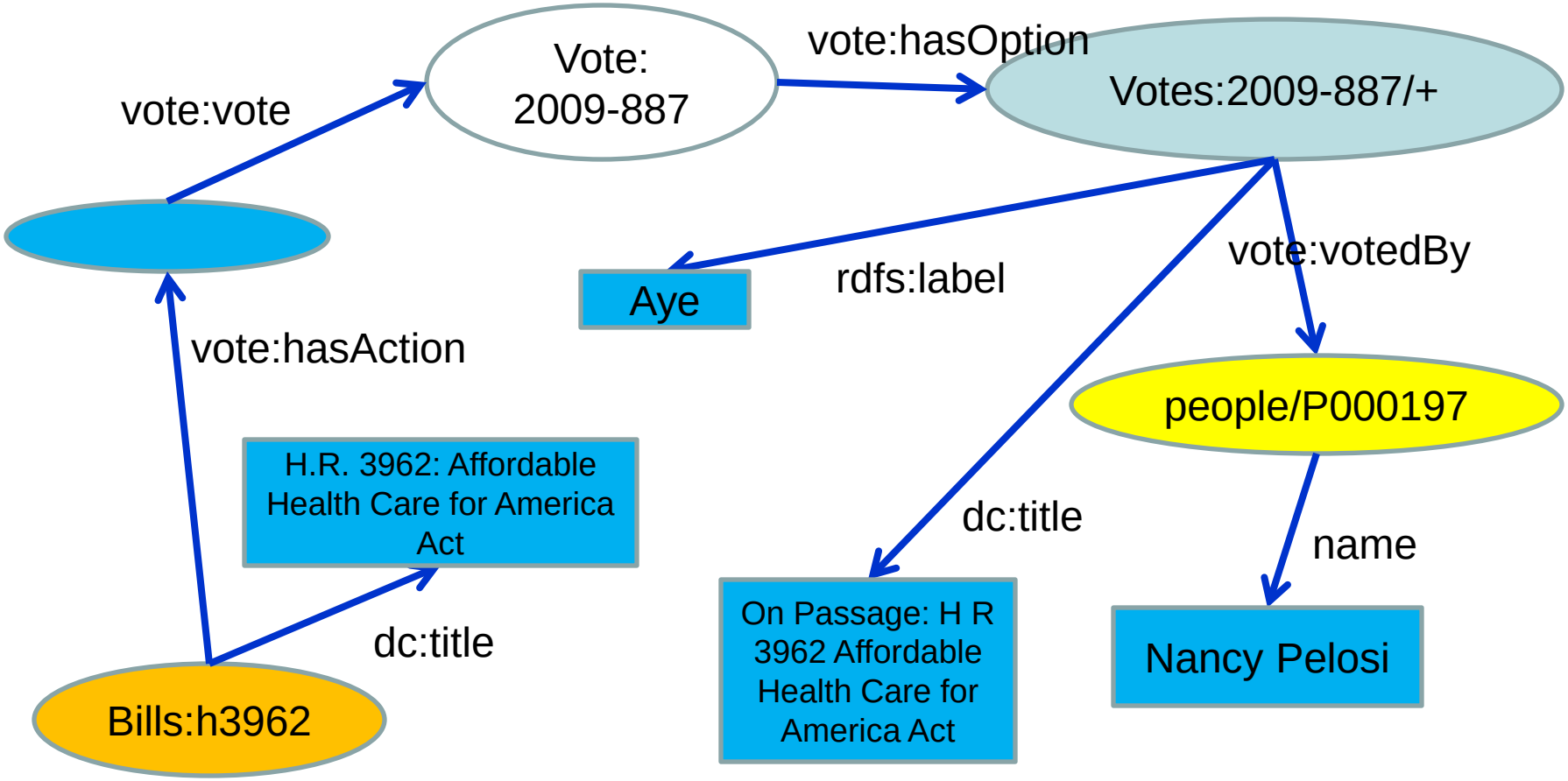
Some missing puzzle pieces:

- Where is the data?
  (smart federation needed)
- Missing background (schema) knowledge.
  (enhancements of the LoD cloud)
- Crucial info still hidden in texts.
  (ontology learning from texts)
- Added reasoning capabilities (e.g., spatial).
  (new ontology language features)

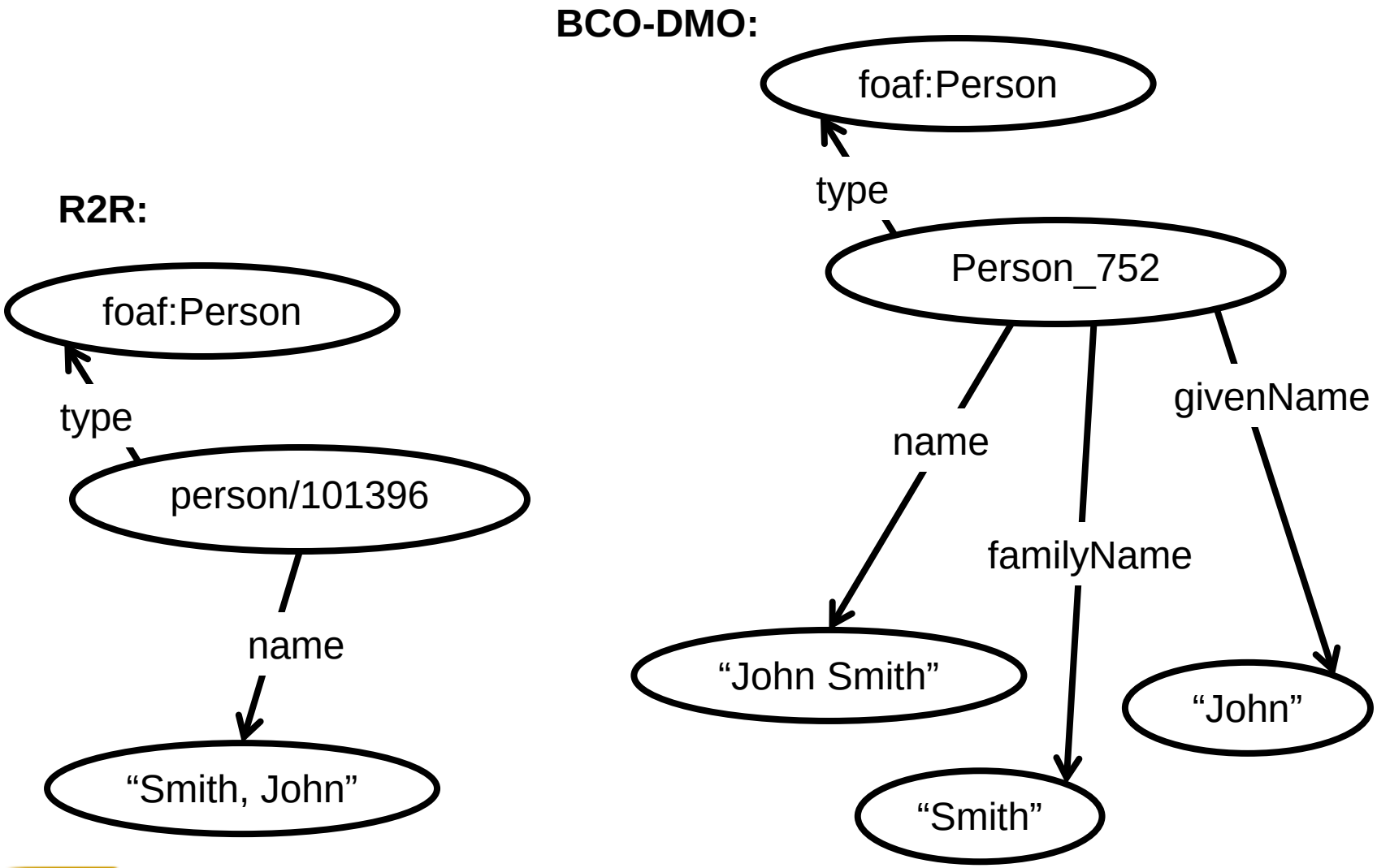
Linked Data: Variety

DaSe Lab

“Nancy Pelosi voted in favor of the Health Care Bill.”



Data Variety



Data Variety

$$a{:}hasWife \sqsubseteq a{:}hasSpouse$$
$$symmetric(a{:}hasSpouse)$$
$$\exists a{:}hasSpouse.a{:}Female \sqsubseteq a{:}Male$$
$$\exists a{:}hasSpouse.a{:}Male \sqsubseteq a{:}Female$$
$$a{:}hasWife(a{:}john, a{:}mary)$$
$$b{:}Male(a{:}john)$$
$$b{:}Female(a{:}mary)$$
$$a{:}Male \sqcap a{:}Female \sqsubseteq \bot$$

$$symmetric(b{:}hasSpouse)$$
$$b{:}hasSpouse(b{:}mike, b{:}david)$$
$$b{:}Male(b{:}david)$$
$$b{:}Male(b{:}mike)$$
$$b{:}Female(b{:}anna)$$

Data Variety

DaSe Lab

Copernicus lunar crater located on earth – courtesy of Krzysztof Janowicz http://stko.geog.ucsb.edu/location_linked_data (missing reference coordinate system)

Information Workbench

Print Help Login

Copernicus (lunar crater)

You do not have permission to edit this page. View Revisions

Copernicus is a lunar impact crater named after the astronomer Nicolaus Copernicus, located in eastern Oceanus Procellarum. It is estimated to be about 800 million years old, and typifies craters that formed during the Copernican period in that it has a prominent ray system.

Contents

Characteristics

Copernicus is visible using binoculars, and is located slightly northwest of the center of the Moon's Earth-facing hemisphere. South of the crater is the Mare Insularum, and to the south-south west is the crater Reinhold. North of Copernicus are the Montes Carpatus, which lie at the south edge of Mare Imbrium. West of Copernicus is a group of dispersed lunar hills. Due to its relative youth, the crater has remained in a relatively pristine shape since it formed.

The circular rim has a discernible hexagonal form, with a terraced inner wall and a 30 km wide, sloping rampart that descends nearly a kilometer to the surrounding mare. There are three distinct terraces visible, and arc-shaped landslides due to slumping of the inner wall as the crater debris subsided.

Most likely due to its recent formation, the crater floor has not been flooded

Location of Copernicus.

Location of Copernicus.

Image

Google Map

The linked data counter-hype

- **“Ontologies don’t work, let’s just link data”**
- **“Okay, with a little bit of ontologies on top.”**
- **But then we don’t even know how to effectively query over multiple linked datasets (without using a lot of manpower to manually integrate them).**
- **It seems rather obvious that we need to get ontologies into the picture, but how to do it while avoiding the drawbacks of strong ontological commitments?**

So What Now?

Ways forward?

How to establish a flexible conceptual architecture using data and ontological modeling?

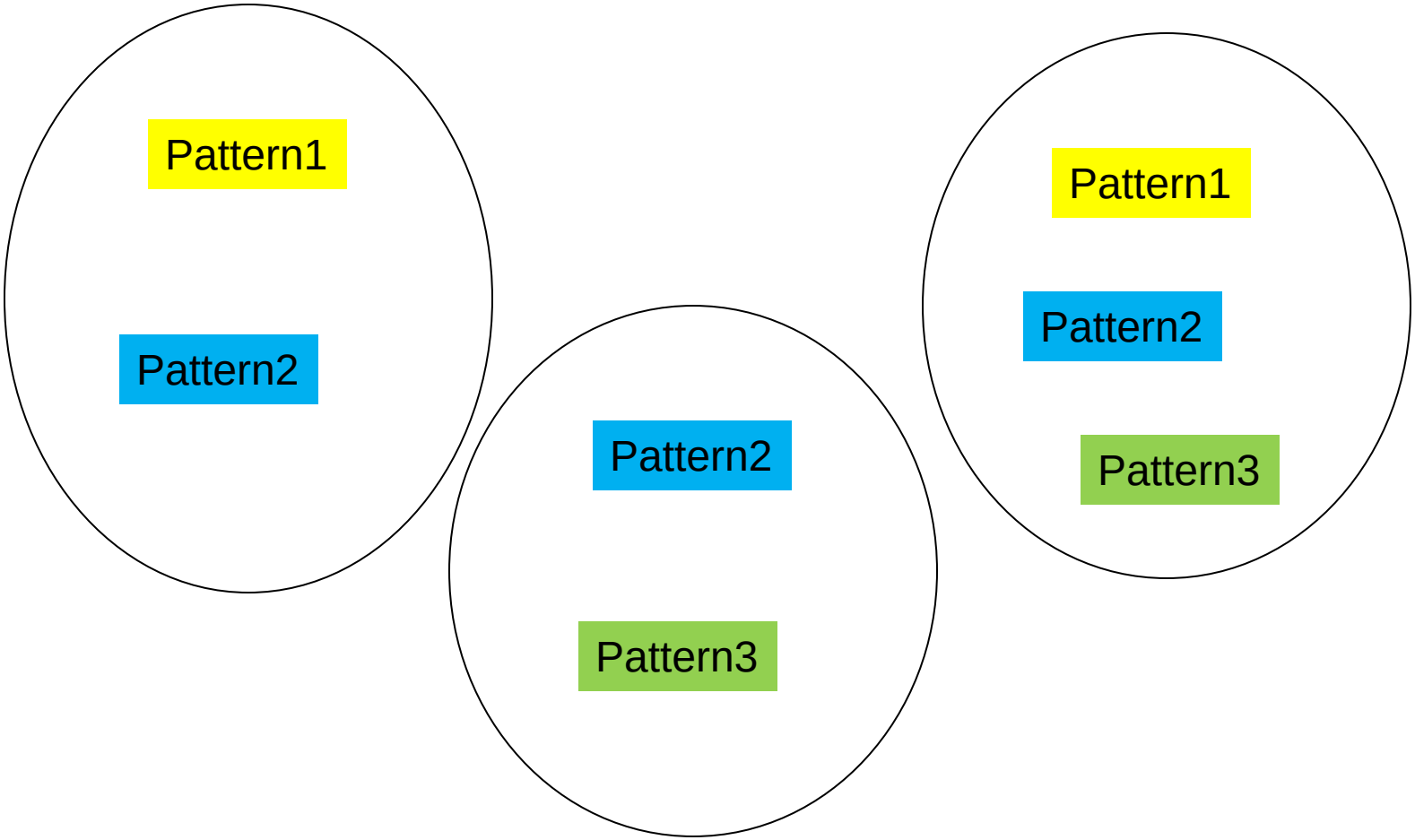
Ontology Design Patterns

DaSe Lab

- **Bottom-up homogenization of data representation.**
- **Avoidance of strong ontological commitments.**
- **Avoidance of standardization of specific modeling details.**
- **Well thought-out patterns can be very strong and versatile, thus serve many needs.**

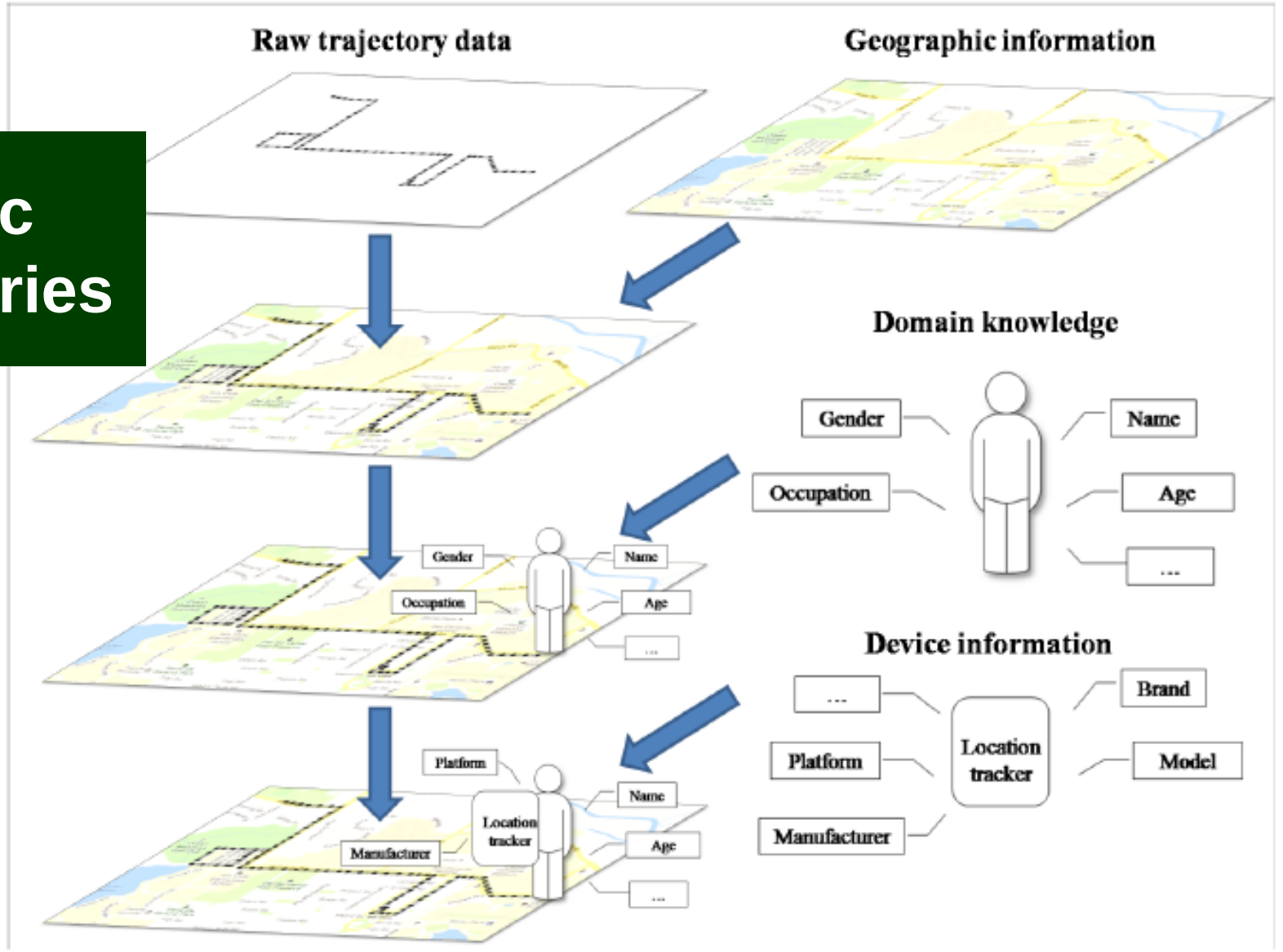
We are currently establishing many geo-patterns in a series of hands-on workshops, the GeoVoCamps, see http://vocamp.org/

Ontology Design Patterns



“Horizontal” alignment via patterns

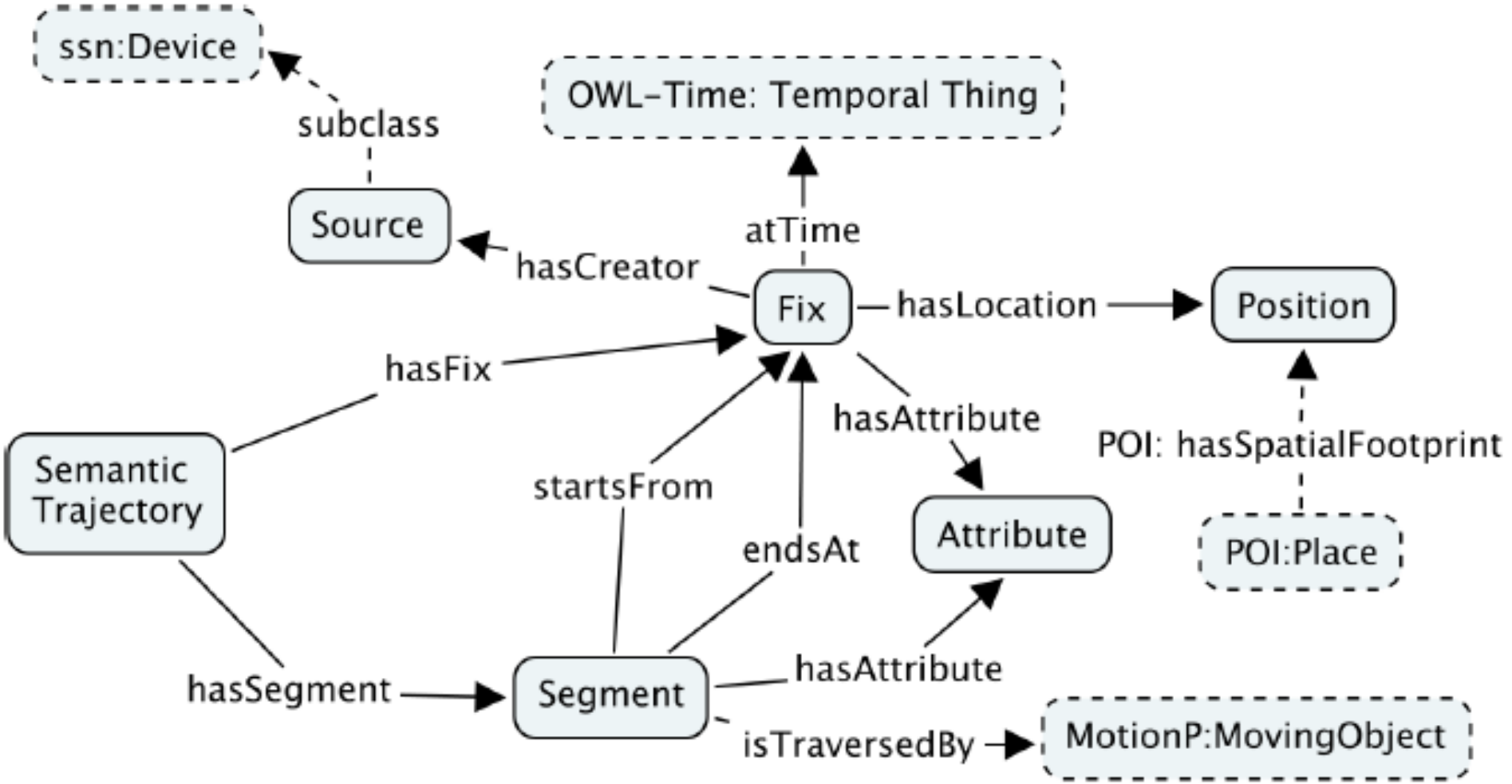
Semantic Trajectories



[Hu, Janowicz, Carral, Scheider, Kuhn, Berg-Cross, Hitzler, Dean, COSIT2013]

Semantic Trajectories

DaSe Lab



Semantics in OWL

$$Fix \sqsubseteq \exists atTime.OWL\text{-}Time{:}Temporal\ \ Thing \sqcap \exists hasLocation.Position \sqcap \exists hasFix^{-}.SemanticTrajectory \quad (1)$$

$$Segment \sqsubseteq \exists startsFrom.Fix \sqcap \exists endsAt.Fix \quad (2)$$

$$\top \sqsubseteq \leq 1 startsFrom.\top \quad (3)$$

$$\top \sqsubseteq \leq 1 endsAt.\top \quad (4)$$

$$Segment \sqsubseteq \exists hasSegment^{-}.SemanticTrajectory \quad (5)$$

$$startsFrom^{-} \circ endsAt \sqsubseteq hasNext \quad (6)$$

$$hasNext \sqsubseteq hasSuccessor \quad (7)$$

$$hasSuccessor \circ hasSuccessor \sqsubseteq hasSuccessor \quad (8)$$

$$hasNext^{-} \sqsubseteq hasPrevious \quad (9)$$

$$hasSuccessor^{-} \sqsubseteq hasPredecesor \quad (10)$$

Semantics in OWL

$$Fix \sqcap \neg \exists endsAt.Segment \sqsubseteq StartingFix \quad (11)$$

$$Fix \sqcap \neg \exists startsFrom.Segment \sqsubseteq EndingFix \quad (12)$$

$$Segment \sqcap \exists startsFrom.StartingFix \sqsubseteq StartingSegment \quad (13)$$

$$Segment \sqcap \exists endsAt.EndingFix \sqsubseteq EndingSegment \quad (14)$$

$$SemanticTrajectory \sqsubseteq \exists hasSegment.Segment \quad (15)$$

$$hasSegment \circ startsFrom \sqsubseteq hasFix \quad (16)$$

$$hasSegment \circ endsAt \sqsubseteq hasFix \quad (17)$$

$$\exists hasSegment.Segment \sqsubseteq SemanticTrajectory \quad (18)$$

$$\exists hasSegment^{-}.SemanticTrajectory \sqsubseteq Segment \quad (19)$$

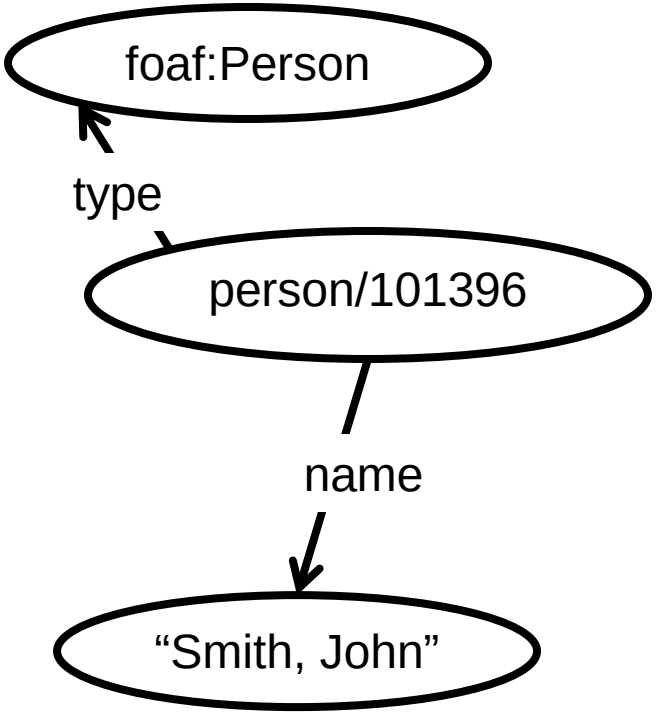
$$\exists hasFix.Segment \sqsubseteq SemanticTrajectory \quad (20)$$

$$\exists hasFix^{-}.SemanticTrajectory \sqsubseteq Fix \quad (21)$$

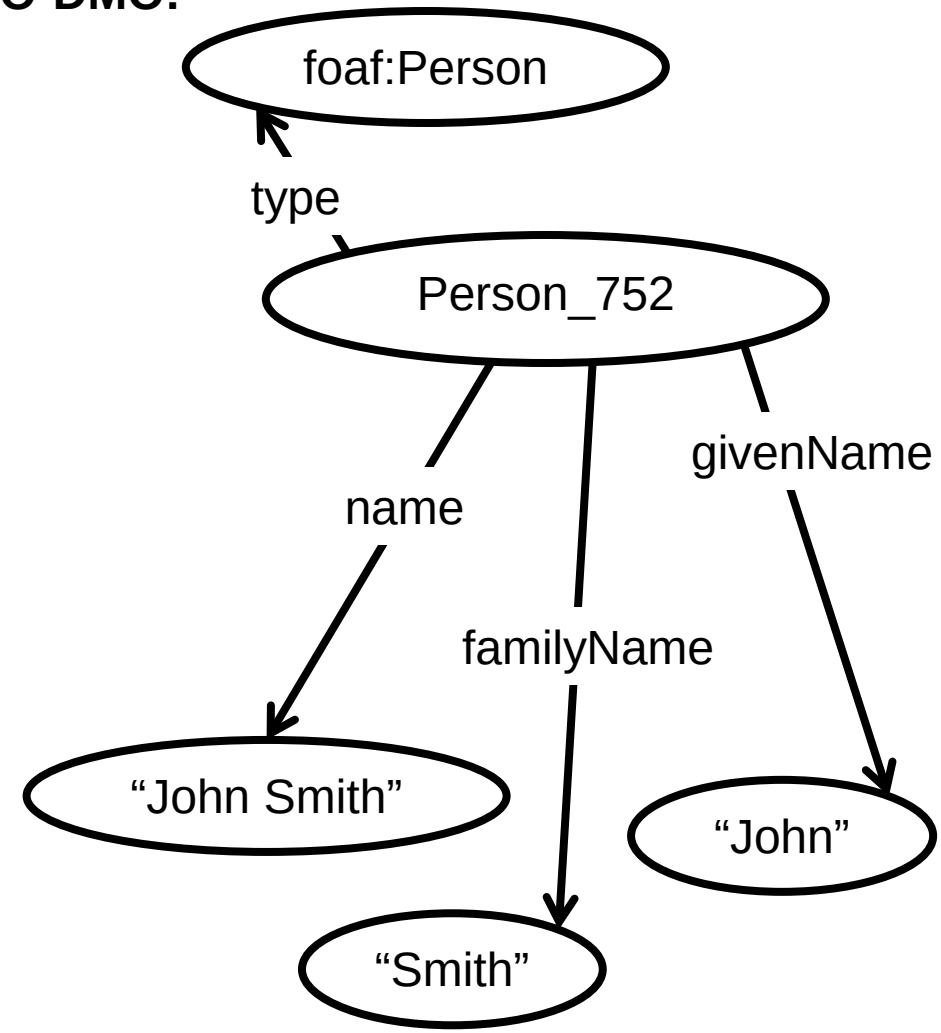
Helpfulness of patterns

Even minimalistic reuse is helpful:

R2R:



BCO-DMO:



Patterns

- Help to focus when modeling (one key notion at a time).
- Good ontology modeling implicitly employs the patterns idea anyway. It's just that you expose the patterns.
- An ontology composed of patterns exposes its internal conceptual structure (as a composition of formal vocabulary pieces).
- Well-designed patterns are widely reusable and adaptable.
- You don't have to buy a whole ontology when you adopt a few patterns from it.
- You can easily modify a pattern without giving up on a lot of similarity to the original pattern (which can be leveraged for data integration).
- You can separate the patterns from specific (application-driven) modifications.
- You can separate the patterns from specific axiomatically defined "views".

Patterns Example

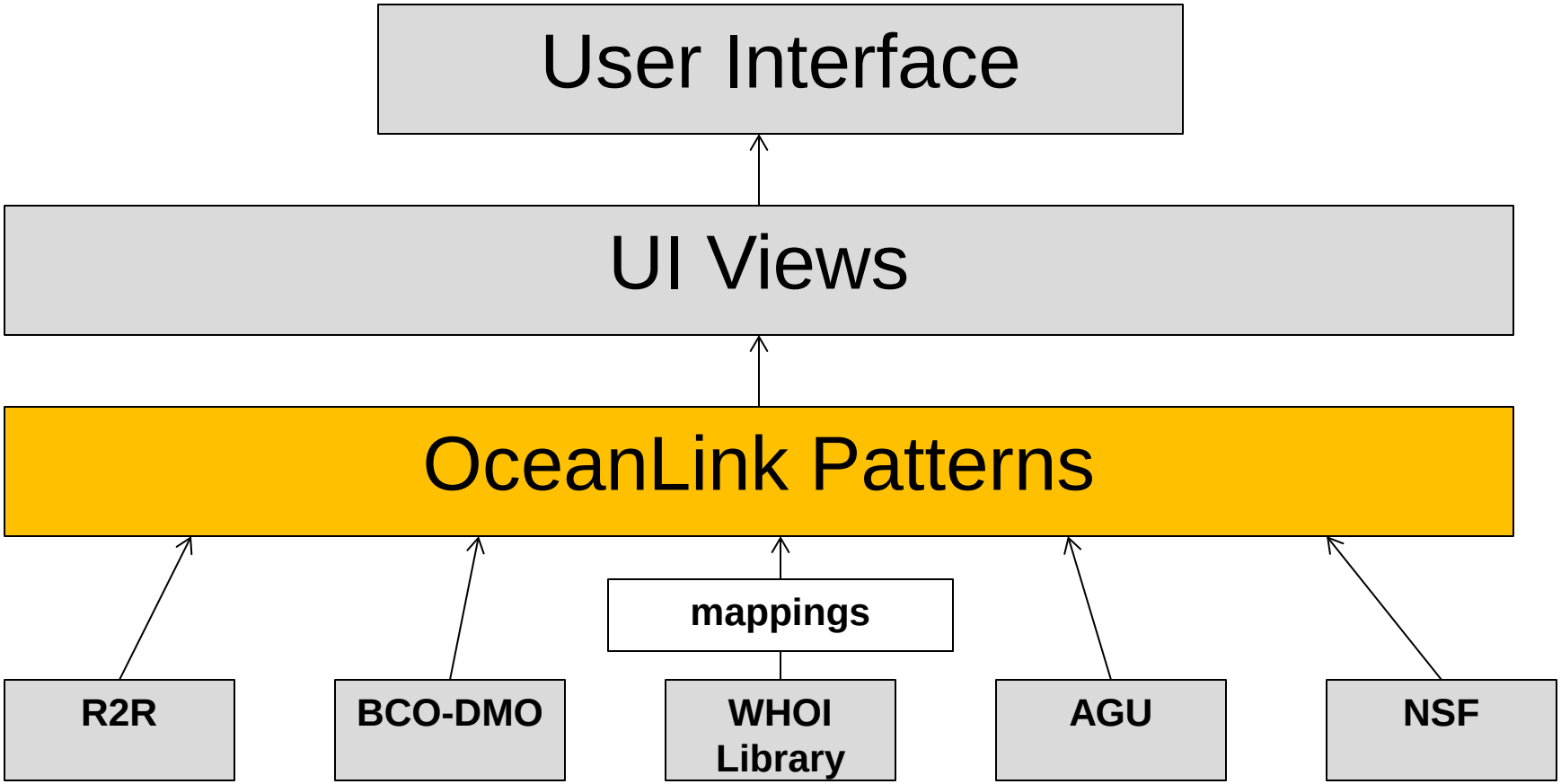
DaSe Lab

NSF EarthCube project “OceanLink”:

- **Integration of existing ocean science data repositories.**
- **For faceted browsing and semantic search.**
- **To be done in a flexible, extendable, modular way.**
- **With minimal effort for additional data providers to integrate their content.**

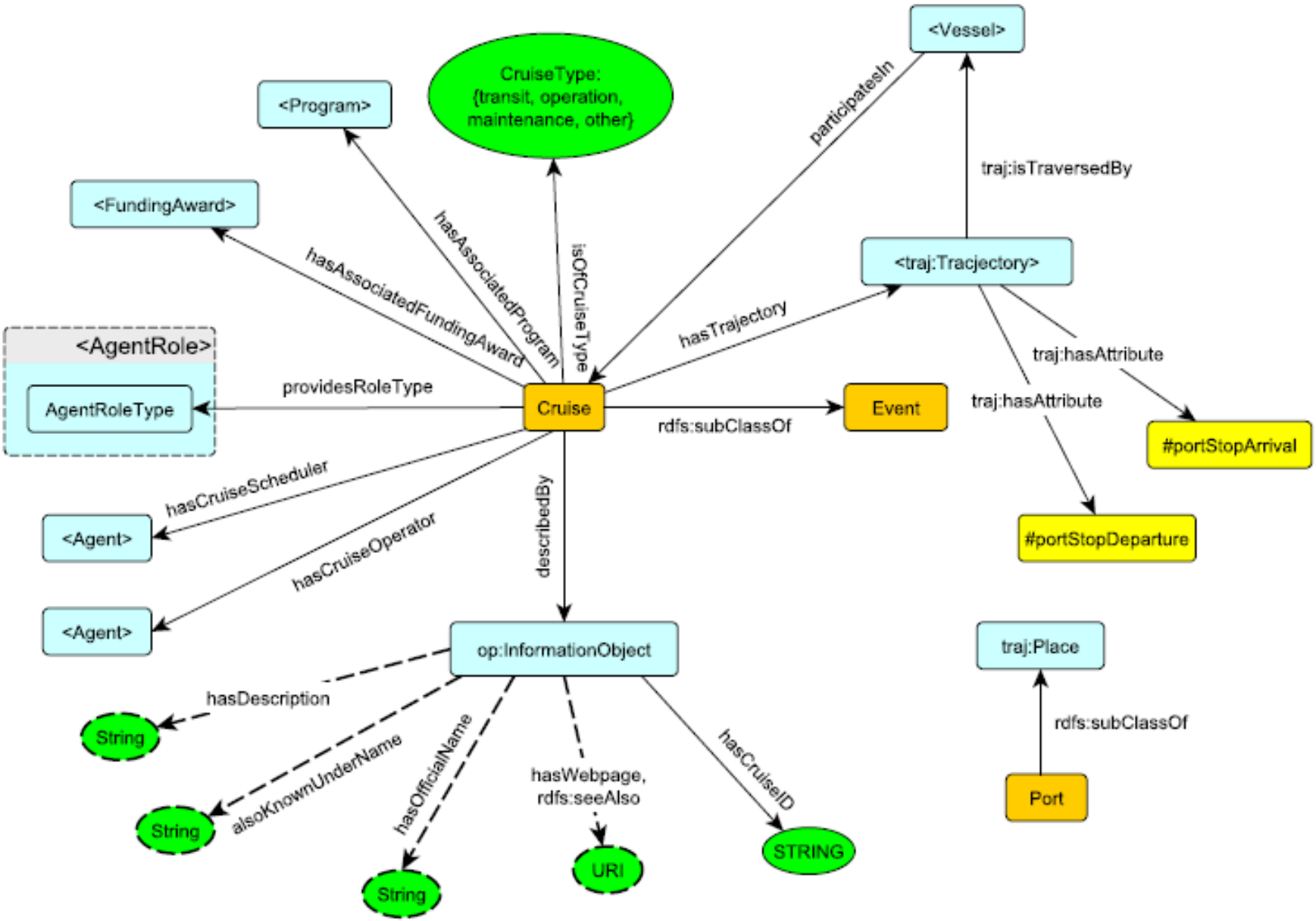
National Science Foundation award 1354778 "EAGER: Collaborative Research: EarthCube Building Blocks, Leveraging Semantics and Linked Data for Geoscience Data Sharing and Discovery."

OceanLink setup



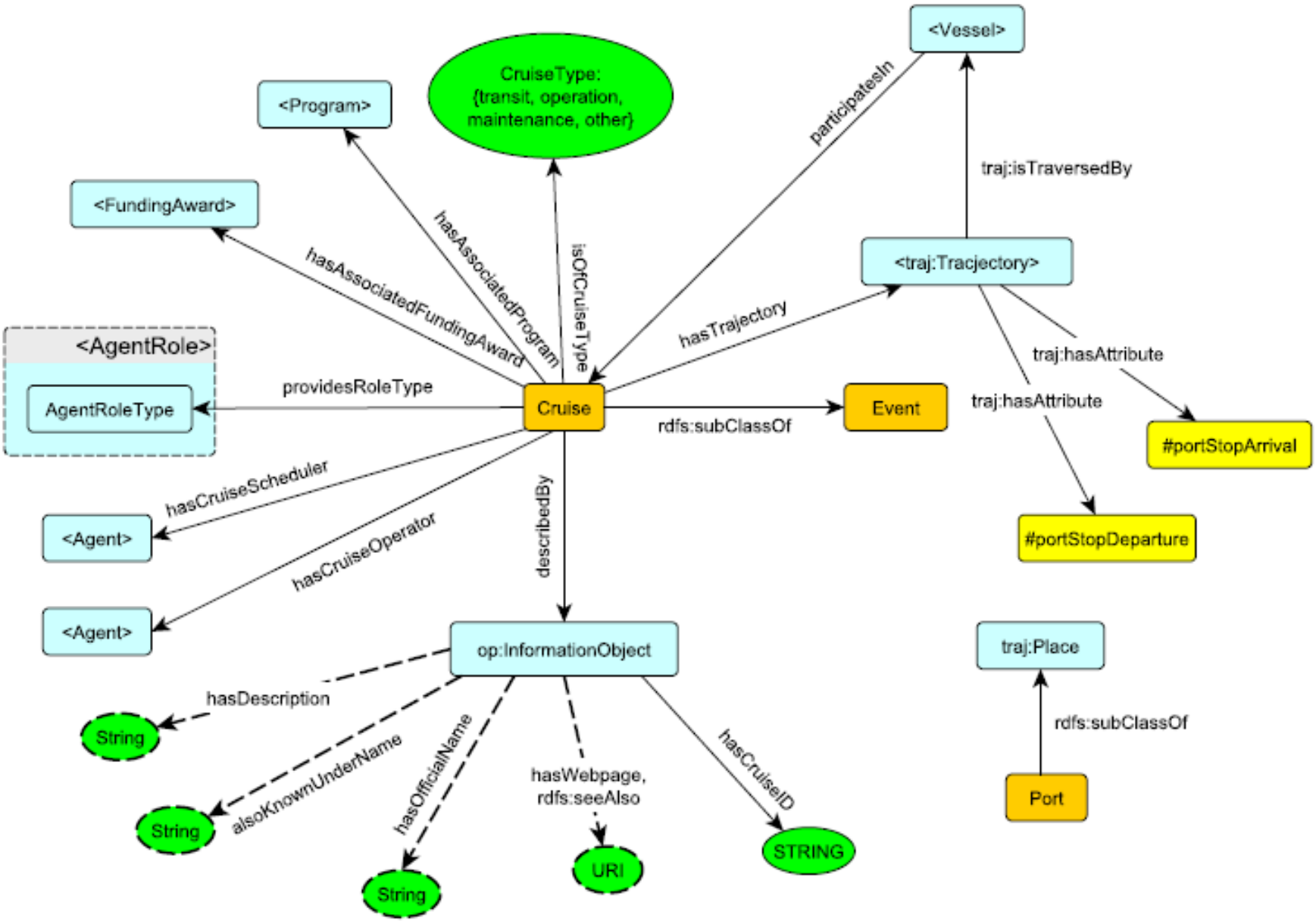
Ocean Science Cruise (draft)

DaSe Lab



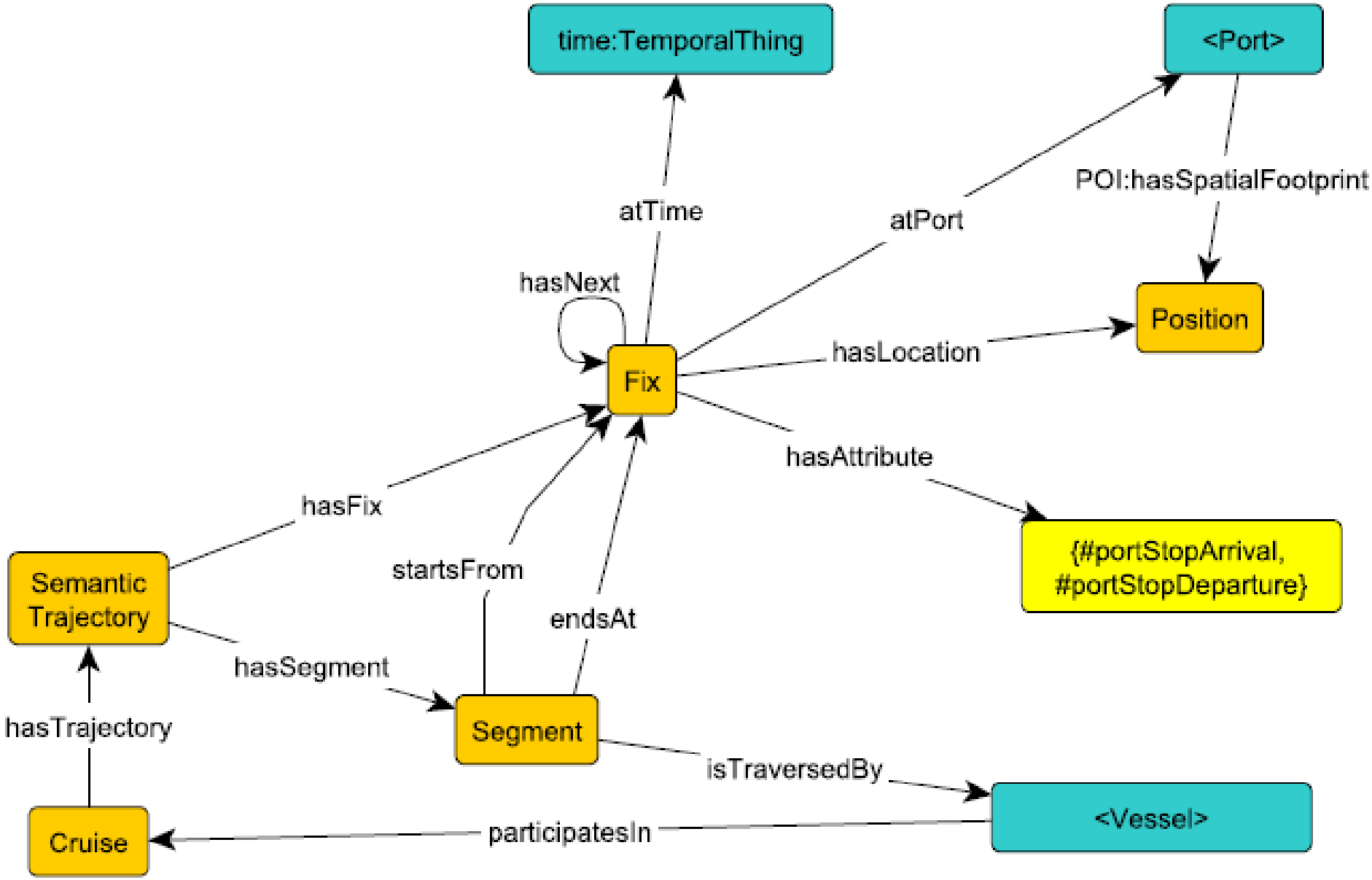
Ocean Science Cruise (draft)

DaSe Lab



Cruise trajectory (draft)

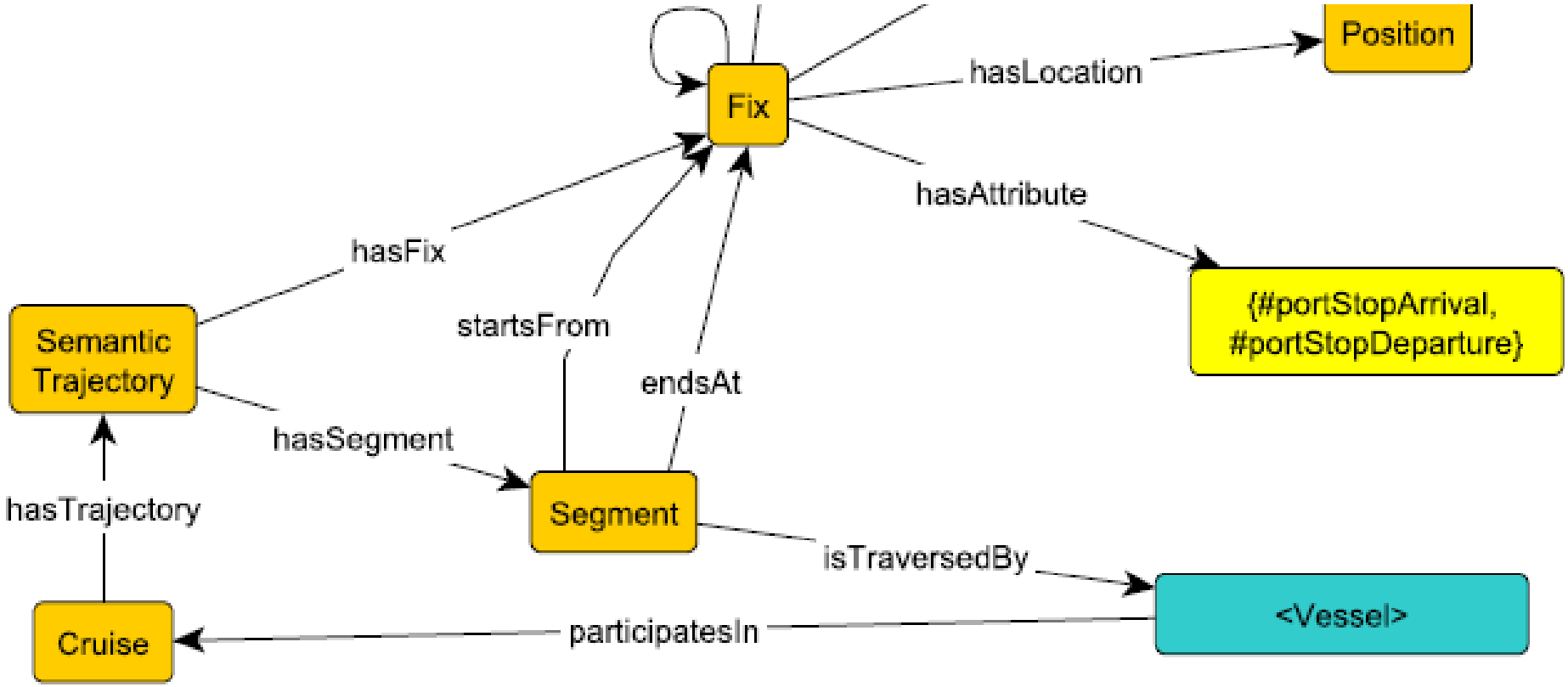
DaSe Lab



Cruise trajectory (draft)

DaSe Lab

$$\text{Cruise}(x) \wedge \text{hasTrajectory}(x, y)$$
$$\wedge\, \text{hasSegment}(y, z) \wedge \text{isTraversedBy}(z, v)$$
$$\rightarrow \text{participatesIn}(v, z)$$



Cruise trajectory (draft)

$$\begin{aligned}\mathrm{Cruise}(x) \wedge \mathrm{hasTrajectory}(x, y) \\ \wedge\, \mathrm{hasSegment}(y, z) \wedge \mathrm{isTraversedBy}(z, v) \\ \rightarrow \mathrm{participatesIn}(v, z)\end{aligned}$$

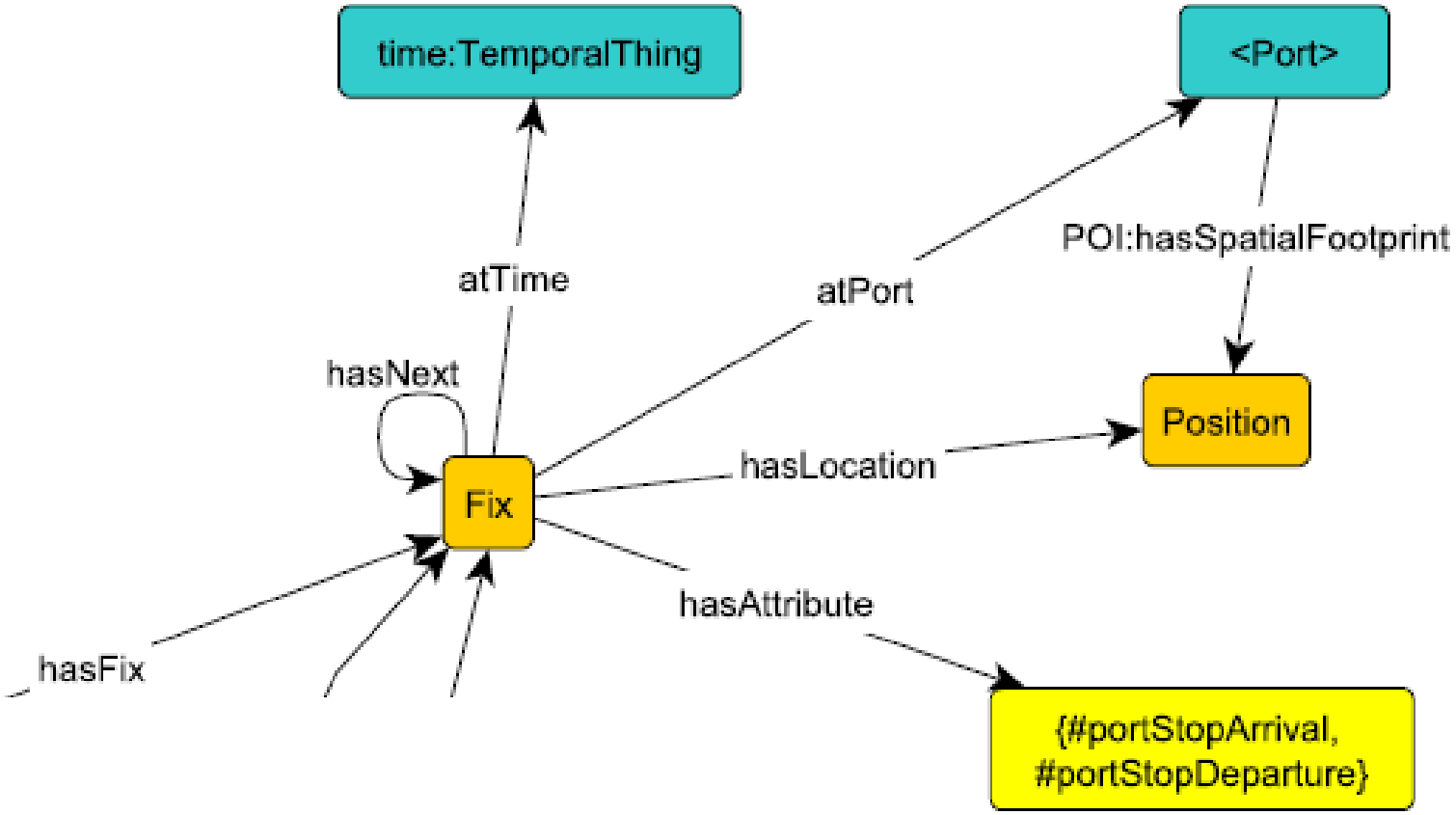
$$\mathrm{Cruise} \equiv \exists \mathrm{cruise}.\mathrm{Self}$$

$$\mathrm{cruise} \circ \mathrm{hasTrajectory} \circ \mathrm{hasSegment} \circ \mathrm{isTraversedBy} \sqsubseteq \mathrm{hasParticipant}$$

$$\mathrm{hasParticipant} \equiv \mathrm{participatesIn}^{-}$$

Cruise trajectory (draft)

DaSe Lab



$$\mathrm{Fix}(x) \wedge \mathrm{hasAttribute}(x, \#\mathrm{portStopArrival})$$
$$\wedge\, \mathrm{atPort}(x, y) \wedge \mathrm{hasSpatialFootprint}(y, z)$$
$$\wedge\, \mathrm{hasLocation}(x, w) \rightarrow \mathrm{locatedIn}(w, z)$$

Cruise trajectory (draft)

$$\mathrm{Fix}(x) \wedge \mathrm{hasAttribute}(x, \#\mathrm{portStopArrival}) \wedge \mathrm{atPort}(x, y) \wedge \mathrm{hasSpatialFootprint}(y, z) \wedge \mathrm{hasLocation}(x, w) \rightarrow \mathrm{locatedIn}(w, z)$$

$$\mathrm{Fix} \wedge \exists \mathrm{hasTrajectory}.\{\#\mathrm{portStopArrival}\} \equiv \exists \mathrm{fixps}.\mathrm{Self} \sqcap \mathrm{hasLocation}^{-} \circ \mathrm{fixps} \circ \mathrm{atPort} \circ \mathrm{hasSpatialFootprint} \sqsubseteq \mathrm{locatedIn}$$

OceanLink patterns

DaSe Lab

Some central patterns:

- **Cruise**
- **Trajectory**
- **Person**
- **Organization**
- **Roles of Agents**
- **Repository Object**
- **Data Set**
- **Document**

We’re not starting from zero of course.

Ways forward

- **Establish a flexible conceptual architecture using data and ontological modeling.**
- **A principled use of patterns, including**
  - **the development of a theory of patterns and**
  - **the provision of a critical amount of central patterns**

  may provide a primary path forward.

Other things I'm doing

DaSe Lab

- **Automated ontology alignment**
- **Overcoming limitations of ontology languages**
- **Reasoning algorithms for ontologies**
- **Scientometrics (for our journal, but branching out with support from the publisher, IOS Press)**

Thanks!

References

- 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.

References

- 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.

References

- 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.

References

- 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: Partonomy 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.

References

- Markus Krötzsch, Frederick Maier, Adila Alfa Krisnadhi, Pascal Hitzler, A Better Uncle For OWL - Nominal Schemas for Integrating Rules and Ontologies. In: S. Sadagopan, Krithi Ramamritham, Arun Kumar, M.P. Ravindra, Elisa Bertino, Ravi Kumar (eds.), WWW '11 20th International World Wide Web Conference, Hyderabad, India, March/April 2011. ACM, New York, 2011, pp. 645-654.
- 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.