Ontology Design Patterns for Large-Scale Data Interchange and Discovery

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

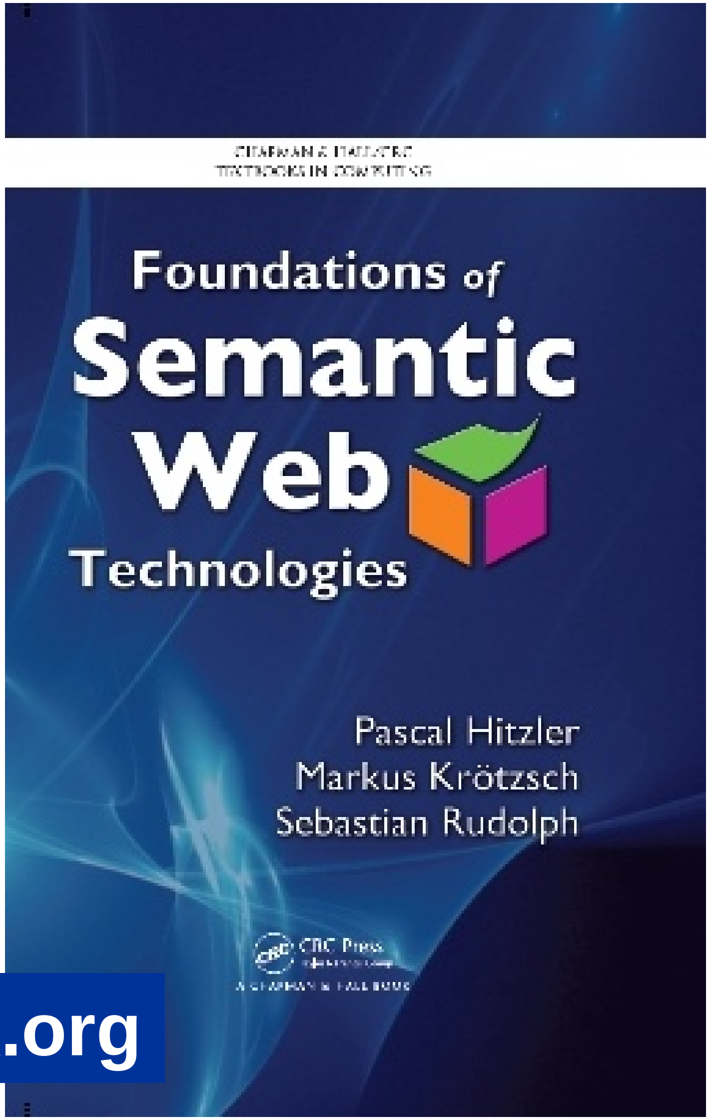
Textbook

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

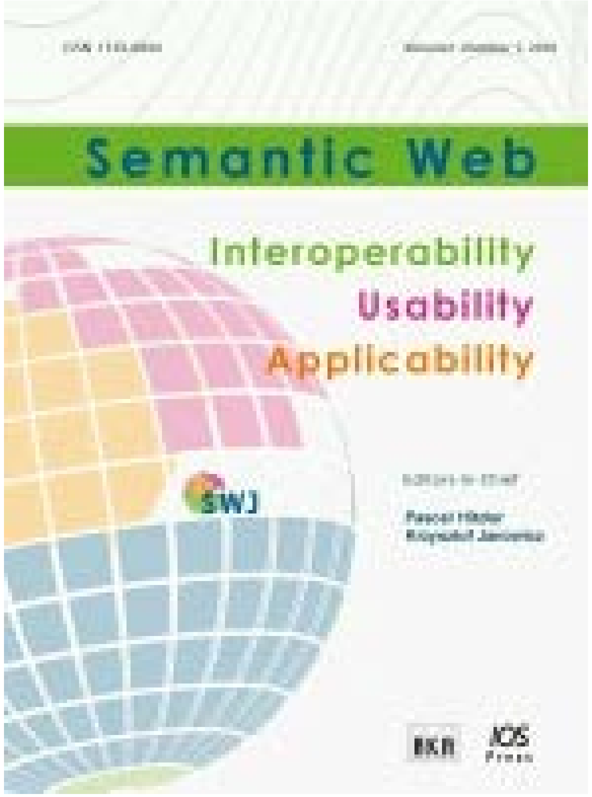
DaSe Lab

- **EiCs: Pascal Hitzler**
  Krzysztof Janowicz

- **Funded 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/**

Our Lab

Data Semantics (DaSe) Lab

Wright State University, Dayton, Ohio, USA

Directors: Michelle Cheatham & Pascal Hitzler

PhD students:
- Reihaneh Amini
- David Carral
- Amit Joshi
- Nazifa Karima
- Adila Krisnadhi
- Raghava Mutharaju
- Stella Sam
- Kunal Sengupta
- Cong Wang

Master students:
- Ashley Coleman
- Todd Huster
- Kylyn Magee
- Brooke McCurdy

Our Lab

DaSe Lab

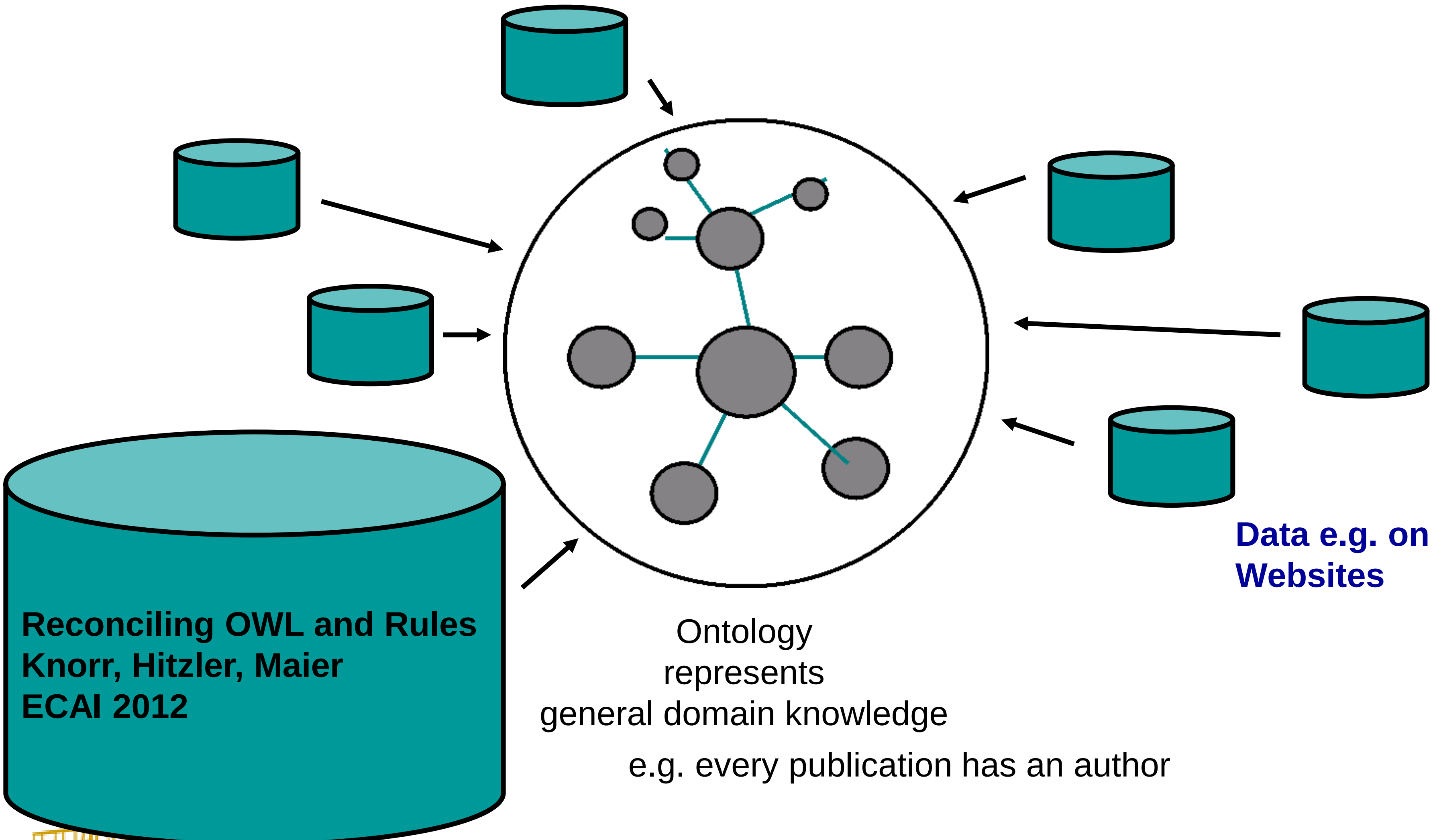
Current focus topics:

- ontology modeling
- ontology design patterns
- ontology and data alignment
- data and information integration
- use of formal semantics
- semantic web languages
- logical foundations
- efficient reasoning algorithms
- applications in the sciences and elsewhere

Ontologies?

A Basic Idea of the Semantic Web

DaSe Lab



The ontology hype

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

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

The ontology hype

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

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

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

Two ontologies.

Left: transportation domain

Right: agriculture domain

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

Multiple perspectives

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

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

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.
- Need a lot of manpower to create.

Pre-LOD Semantic Web

- **Large, monolithic ontologies**
- **Sophisticated ontology languages**

Scientific Hypothesis:

These will solve your data and information management problems

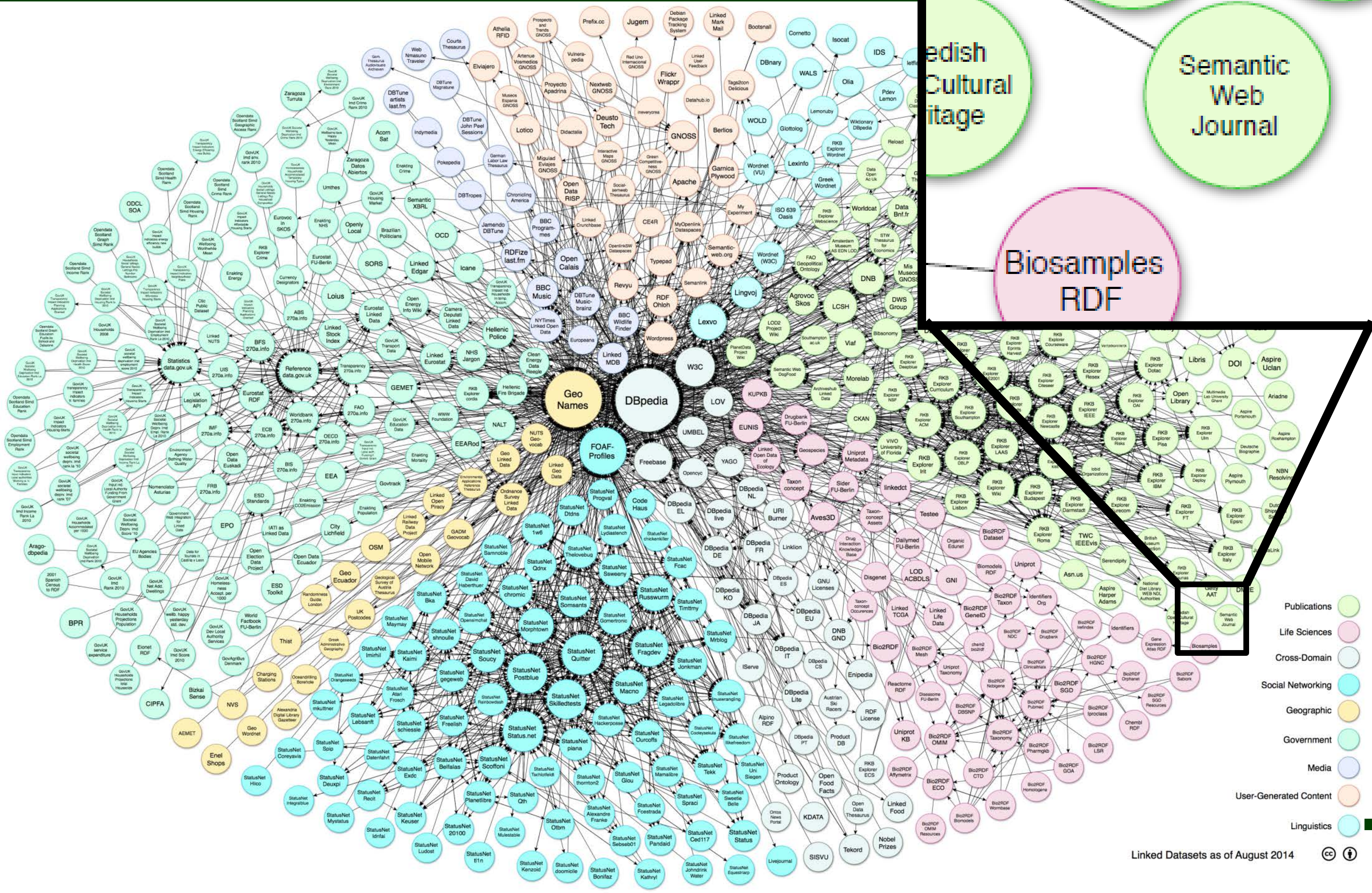
Remember that scientific progress is fundamentally about falsification, not verification ☺

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.”**

Linked Data Cloud 2014



Linked Data: Volume

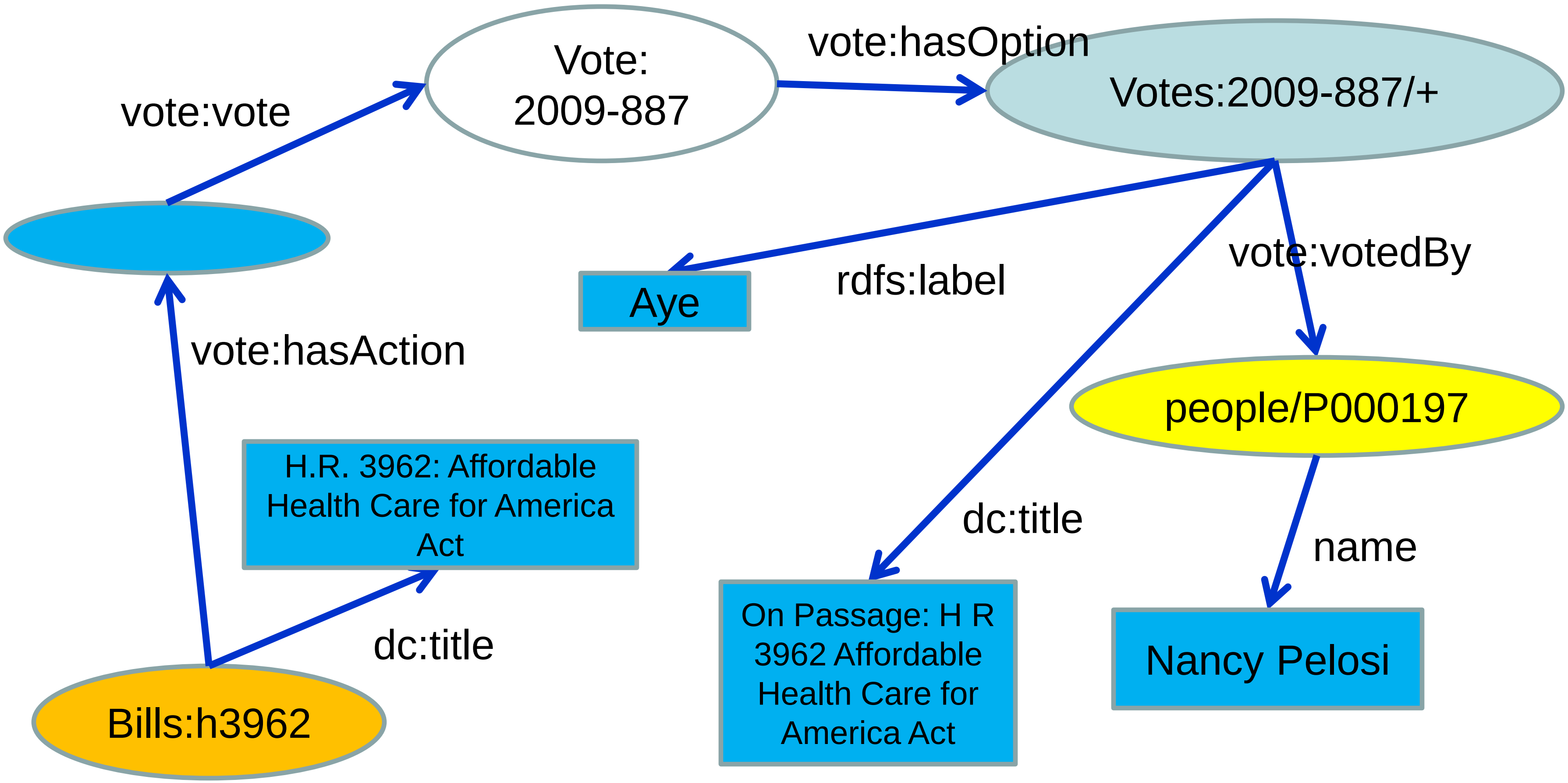
Geoindexed Linked Data – courtesy of Krzysztof Janowicz

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

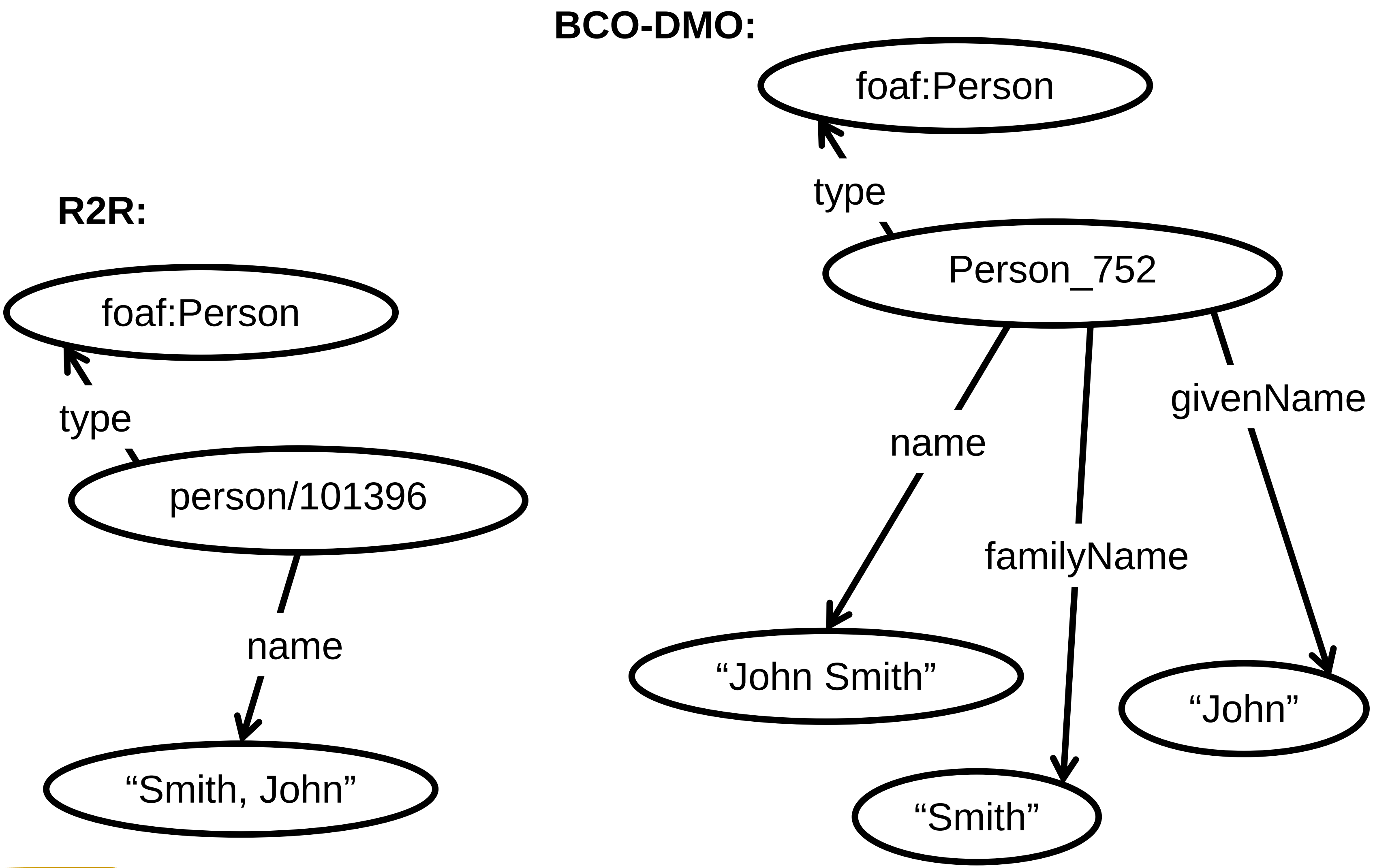
Using Linked Data is tricky

DaSe Lab

"Nancy Pelosi voted in favor of the Health Care Bill."



Alignment? Integration?



Absence of schema?

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
Names
Satellite craters
See also
References
External links

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

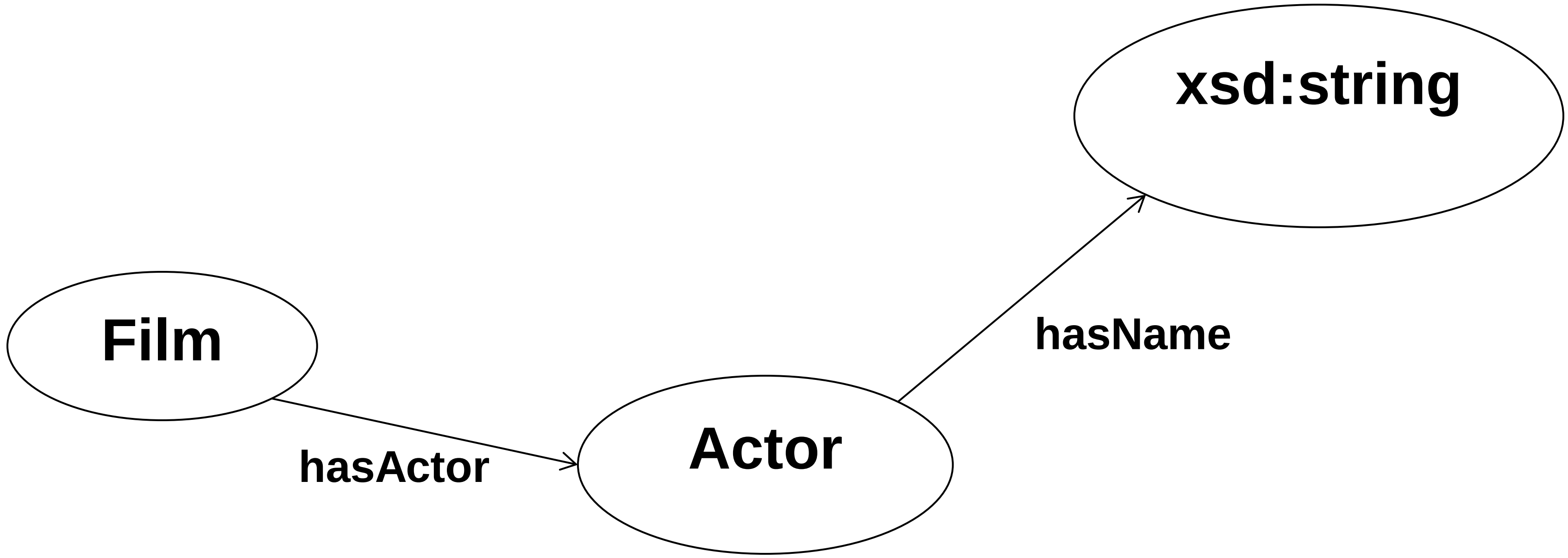
5 LOD Schema Stars

Krzysztof Janowicz, Pascal Hitzler, Benjamin Adams, Dave Kolas, Charles Vardeman II, Five Stars of Linked Data Vocabulary Use. Semantic Web 5 (3), 2014, 173-176.

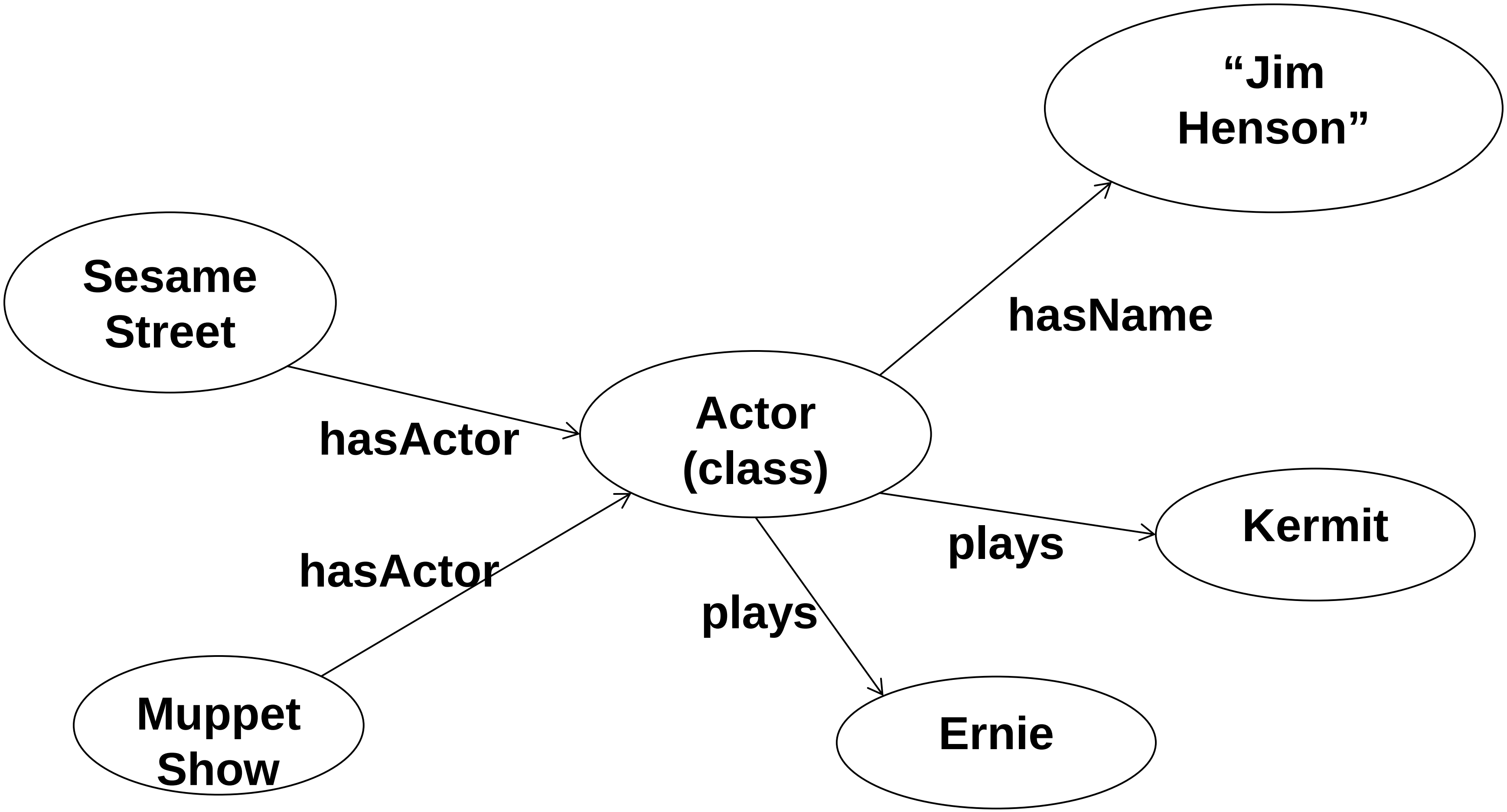
- Quality of schema and documentation.

- Level of reuseability.

Example from Linked MDB



Problem!



"Jim Henson"
Sesame Street
hasName
hasActor
Actor (class)
Kermit
hasActor
plays
plays
Muppet Show
Ernie

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

"An ontology design pattern is a reusable successful solution to a recurrent modeling problem."

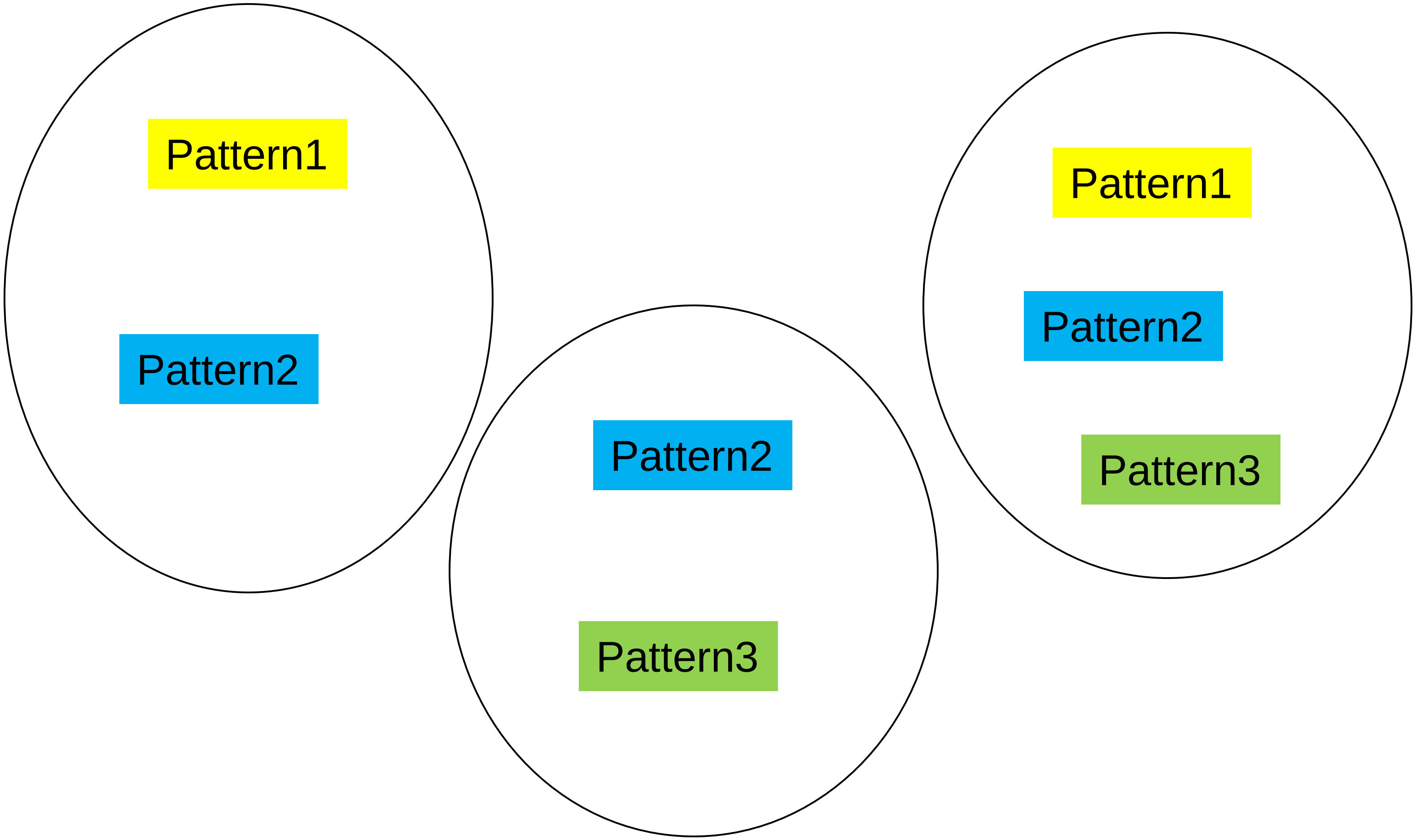
So-called *content patterns* usually encode specific abstract notions, such as process, event, agent, etc.

Ontology Design Patterns

- **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

Example: The NSF GeoLink Project

EarthCube

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.”

EarthCube GeoLink project

DaSe Lab

Targeting data sharing and discovery in the Earth Sciences.

LDEO: Robert Arko, Suzanne Carbotte, Kerstin Lehnert

WHOI: Cynthia Chandler, Peter Wiebe, Lisa Raymond, Adam Shepherd

UCSB: Mark Schildhauer, Krzysztof Janowicz, Matt Jones, Yingjie Hu

Ocean Leadership: Douglas Fils

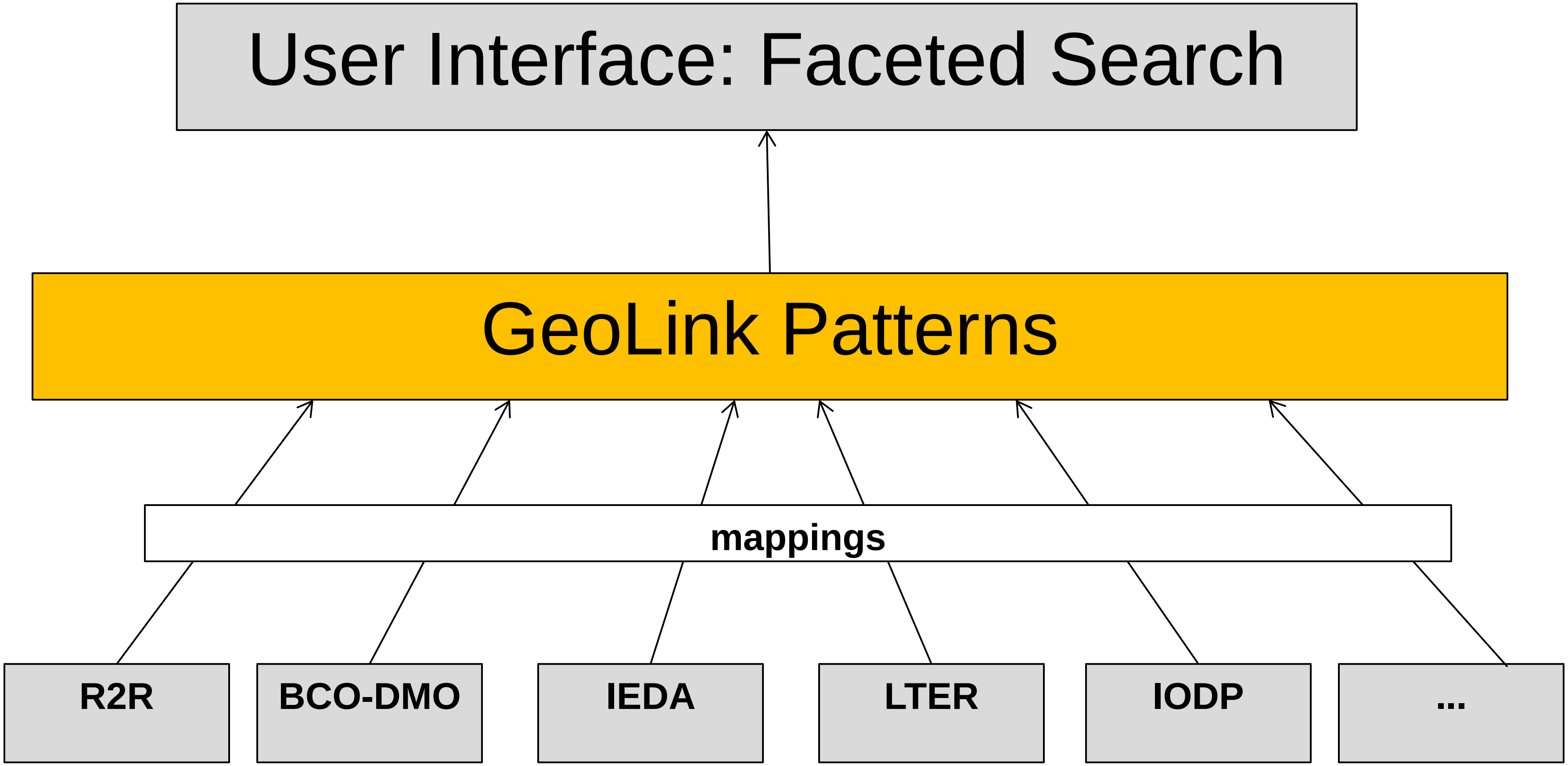
Marymount Univ: Thomas Narock

WSU: Pascal Hitzler, Michelle Cheatham, Adila Krisnadhi

UMBC: Tim Finin

~$1.9M, 2 years duration

GeoLink setup



Ontology Design Patterns

DaSe Lab

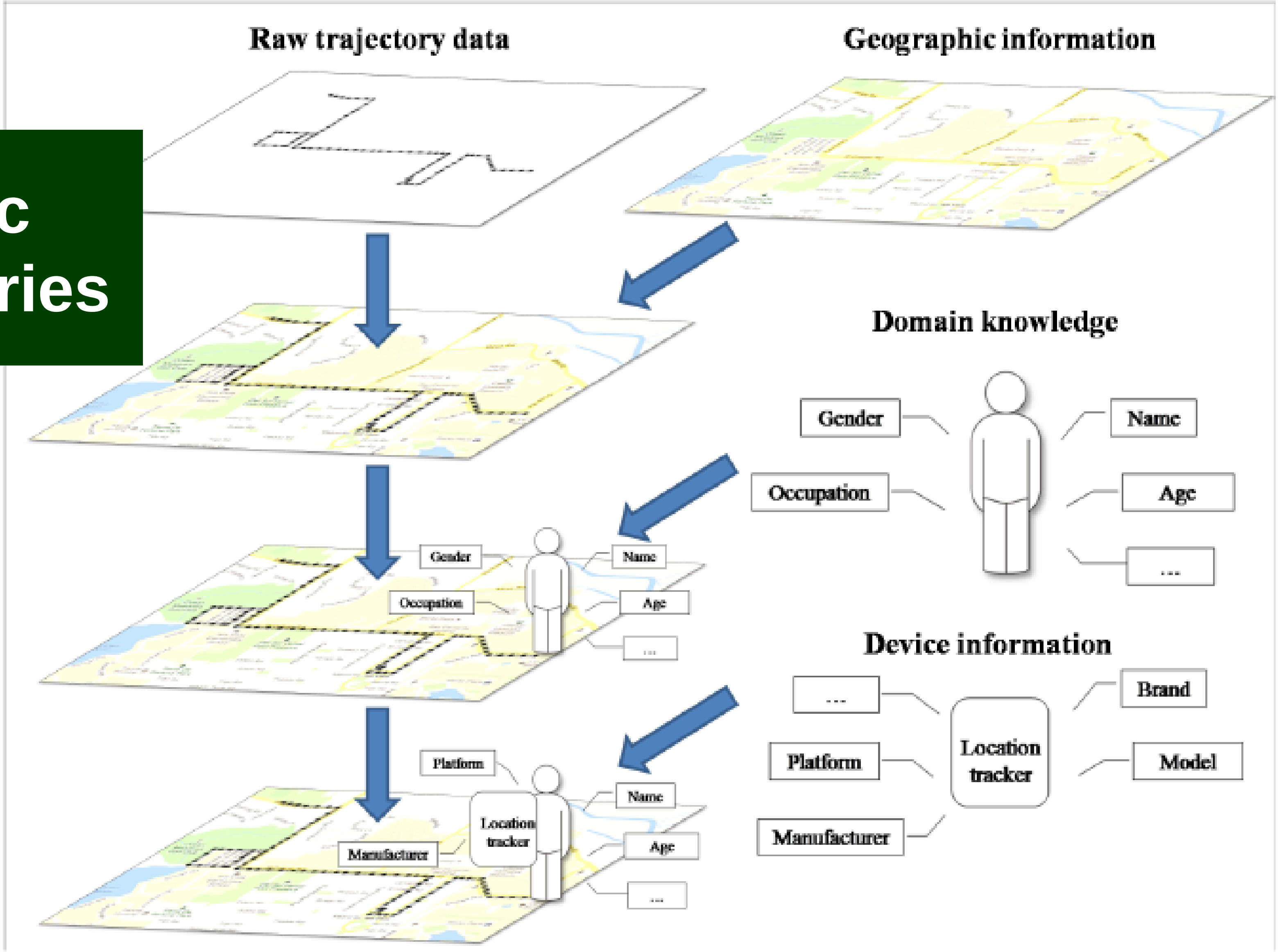
“An ontology design pattern is a reusable successful solution to a recurrent modeling problem.”

So-called *content patterns* usually encode specific abstract notions, such as process, event, agent, etc.

Patterns provide modular, reusable, replaceable, pieces.

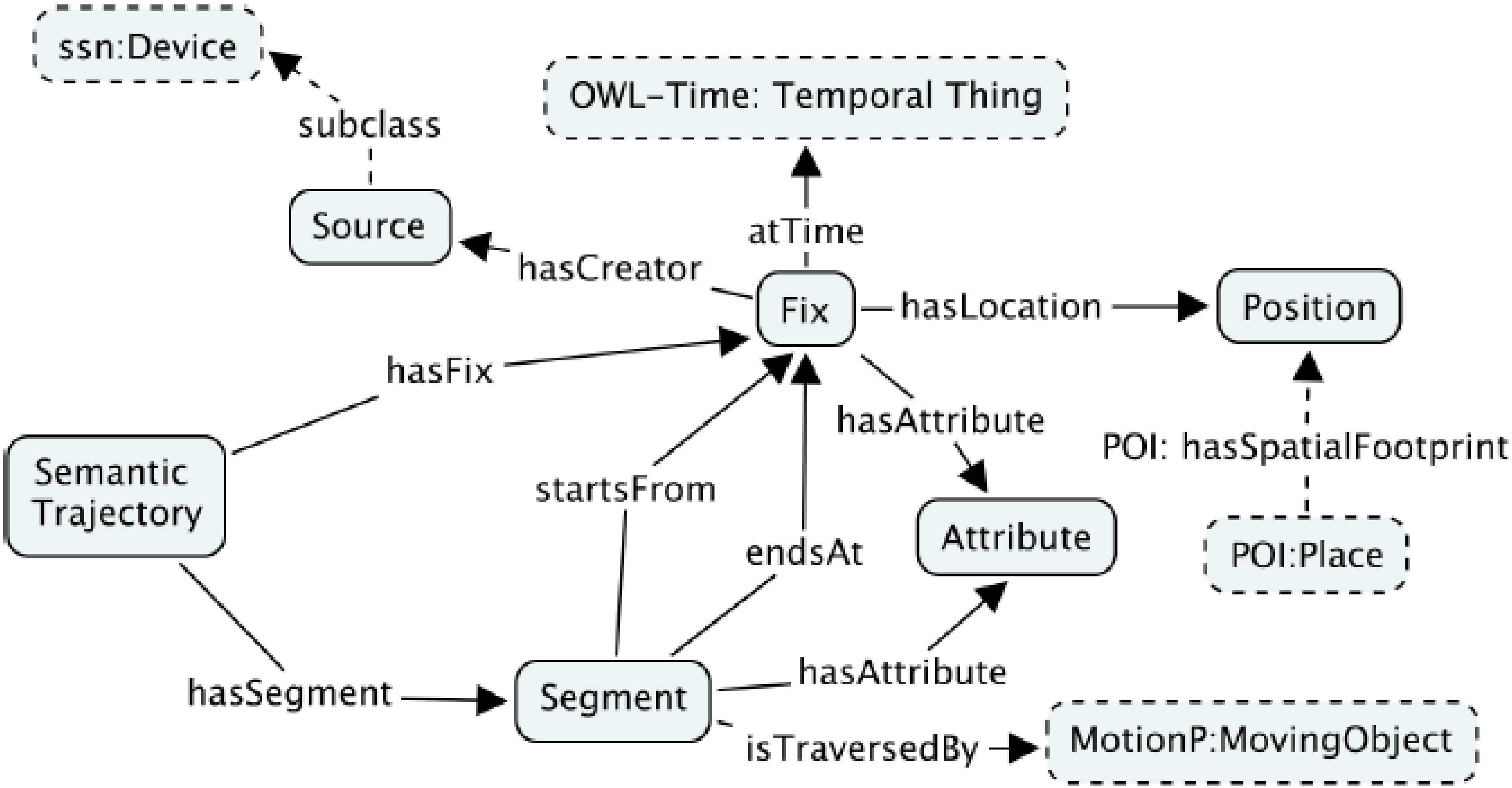
By agreeing on reuse of generic patterns (but leaving the relationships between the patterns to a specific assembly for a special purpose), we can have reuse while preserving heterogeneity.

Semantic Trajectories



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

Semantic Trajectories



Semantics in OWL

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

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

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

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

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

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

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

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

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

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

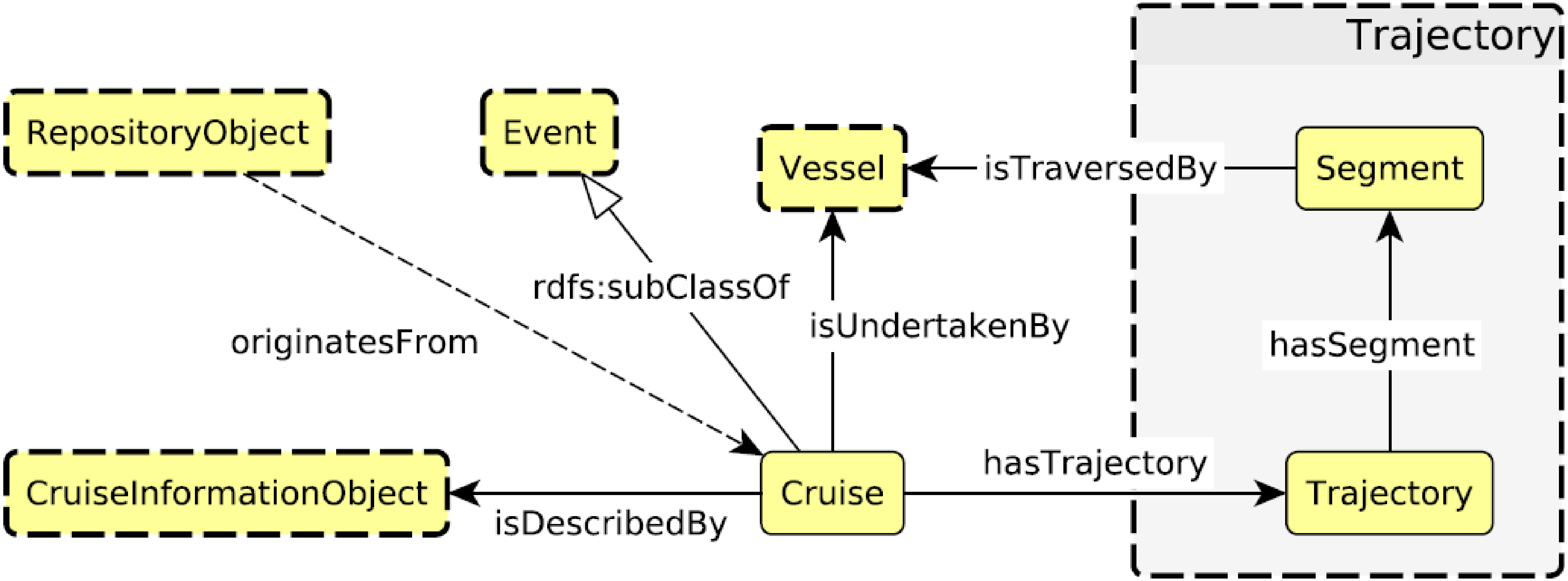
Semantics in OWL

$$Fix \sqcap \neg\exists endsAt.Segment \sqsubseteq StartingFix \quad (11)$$
$$Fix \sqcap \neg\exists startsFrom.Segment \sqsubseteq EndingFix \quad (12)$$
$$Segment \sqcap \exists startsFrom.StartingFix \sqsubseteq StartingSegment \quad (13)$$
$$Segment \sqcap \exists endsAt.EndingFix \sqsubseteq EndingSegment \quad (14)$$

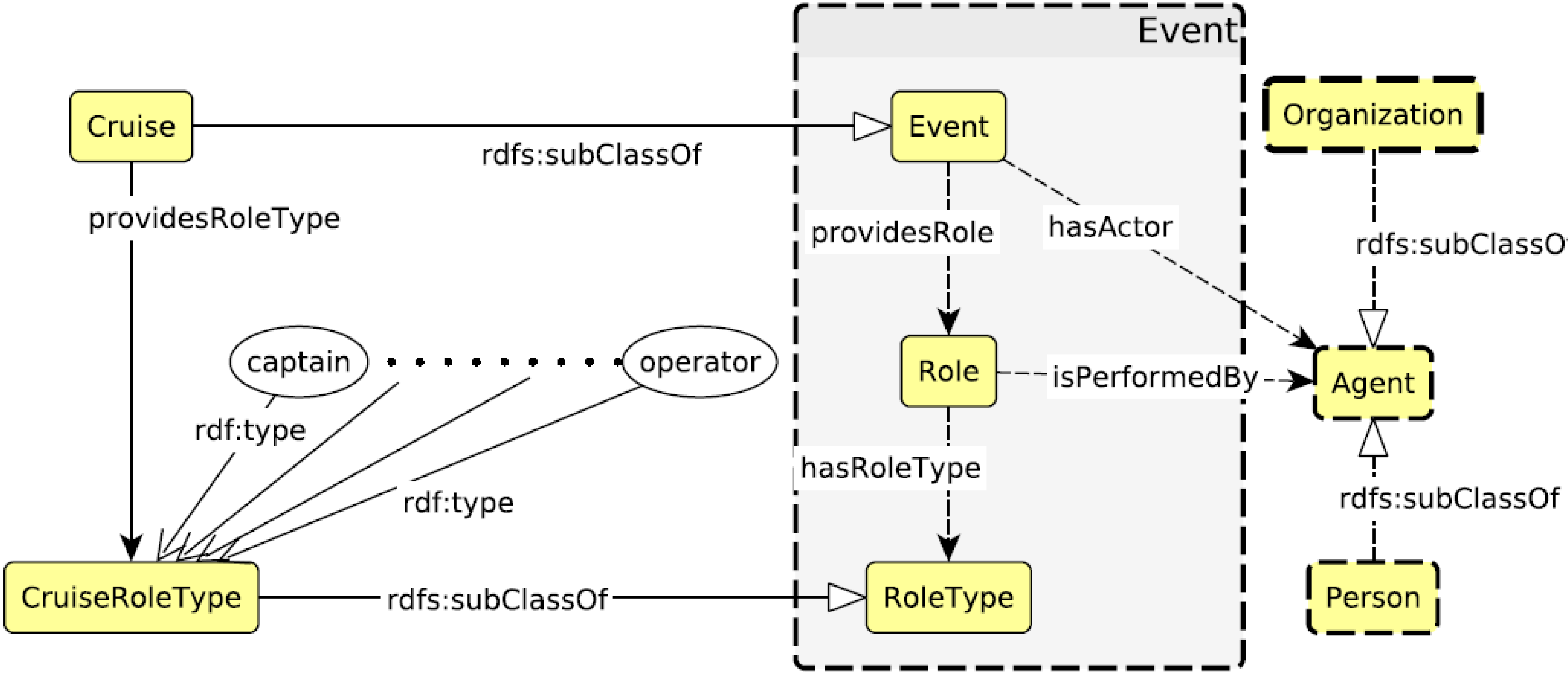
$$SemanticTrajectory \sqsubseteq \exists hasSegment.Segment \quad (15)$$
$$hasSegment \circ startsFrom \sqsubseteq hasFix \quad (16)$$
$$hasSegment \circ endsAt \sqsubseteq hasFix \quad (17)$$

$$\exists hasSegment.Segment \sqsubseteq SemanticTrajectory \quad (18)$$
$$\exists hasSegment^{-}.SemanticTrajectory \sqsubseteq Segment \quad (19)$$
$$\exists hasFix.Segment \sqsubseteq SemanticTrajectory \quad (20)$$
$$\exists hasFix^{-}.SemanticTrajectory \sqsubseteq Fix \quad (21)$$

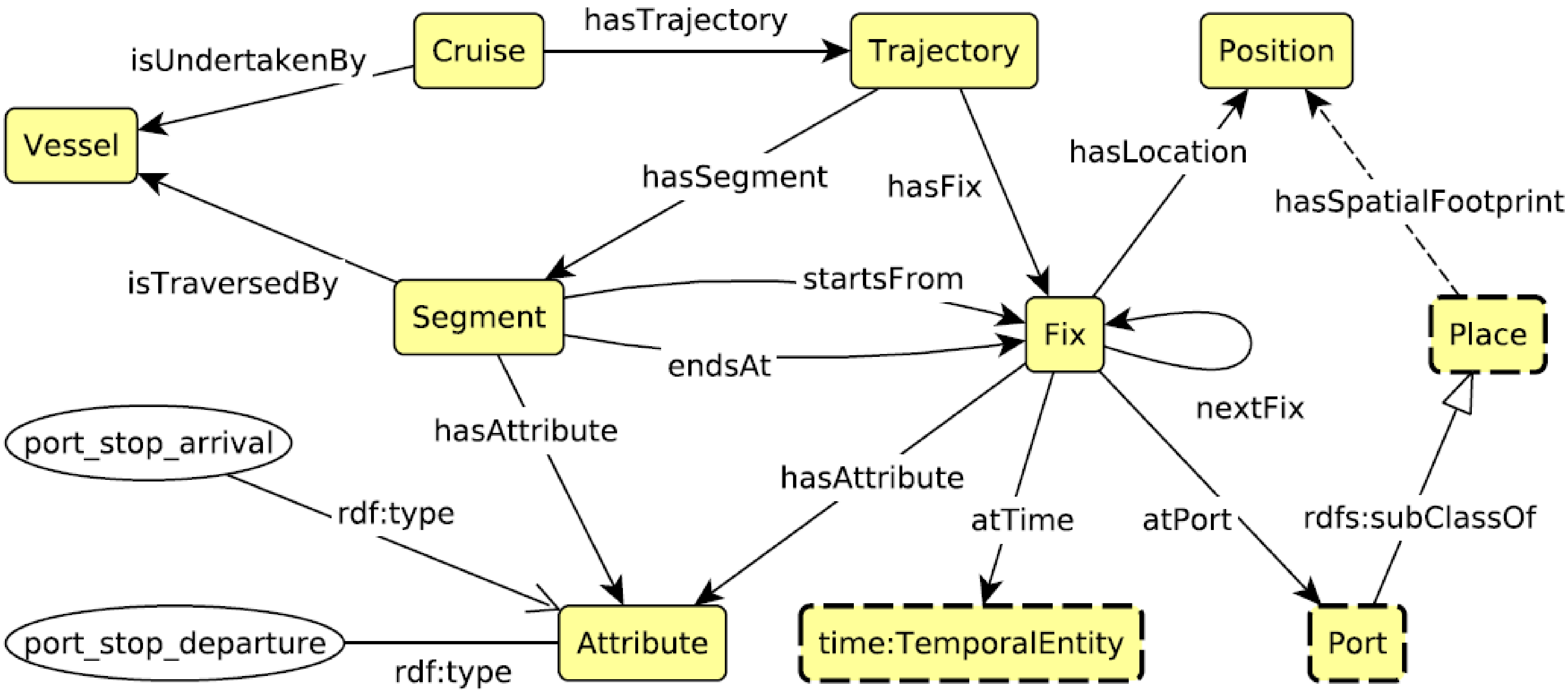
Oceanographic Cruise



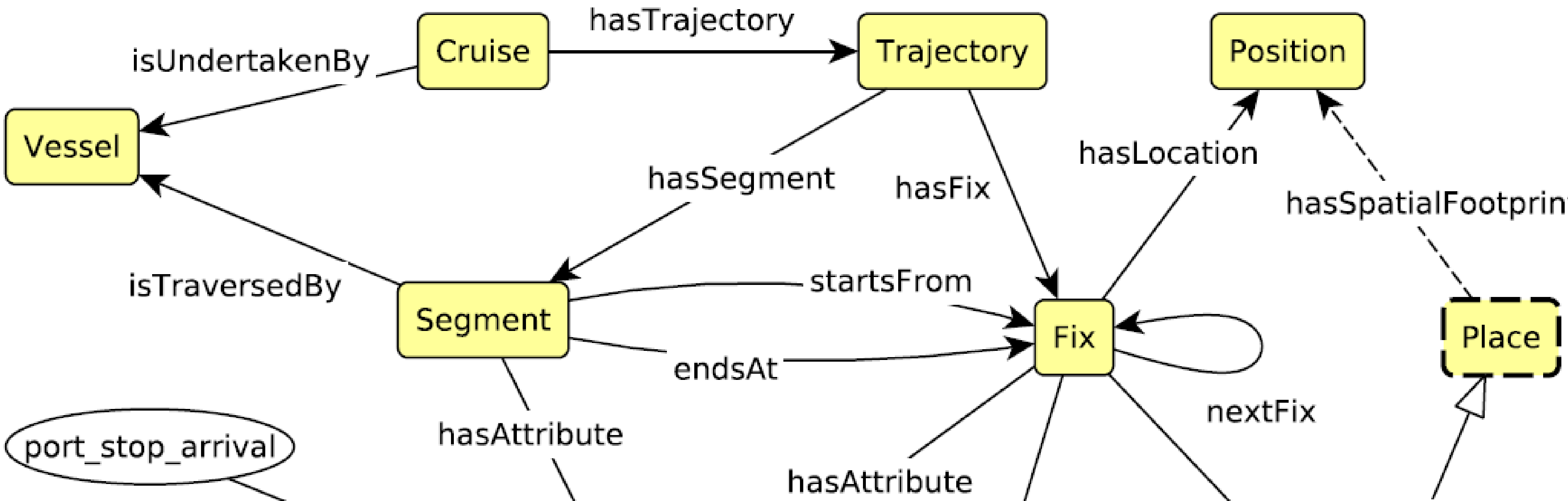
Roles (Cruise as Event)



Cruise Trajectories



Cruise Trajectories



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

Cruise trajectory

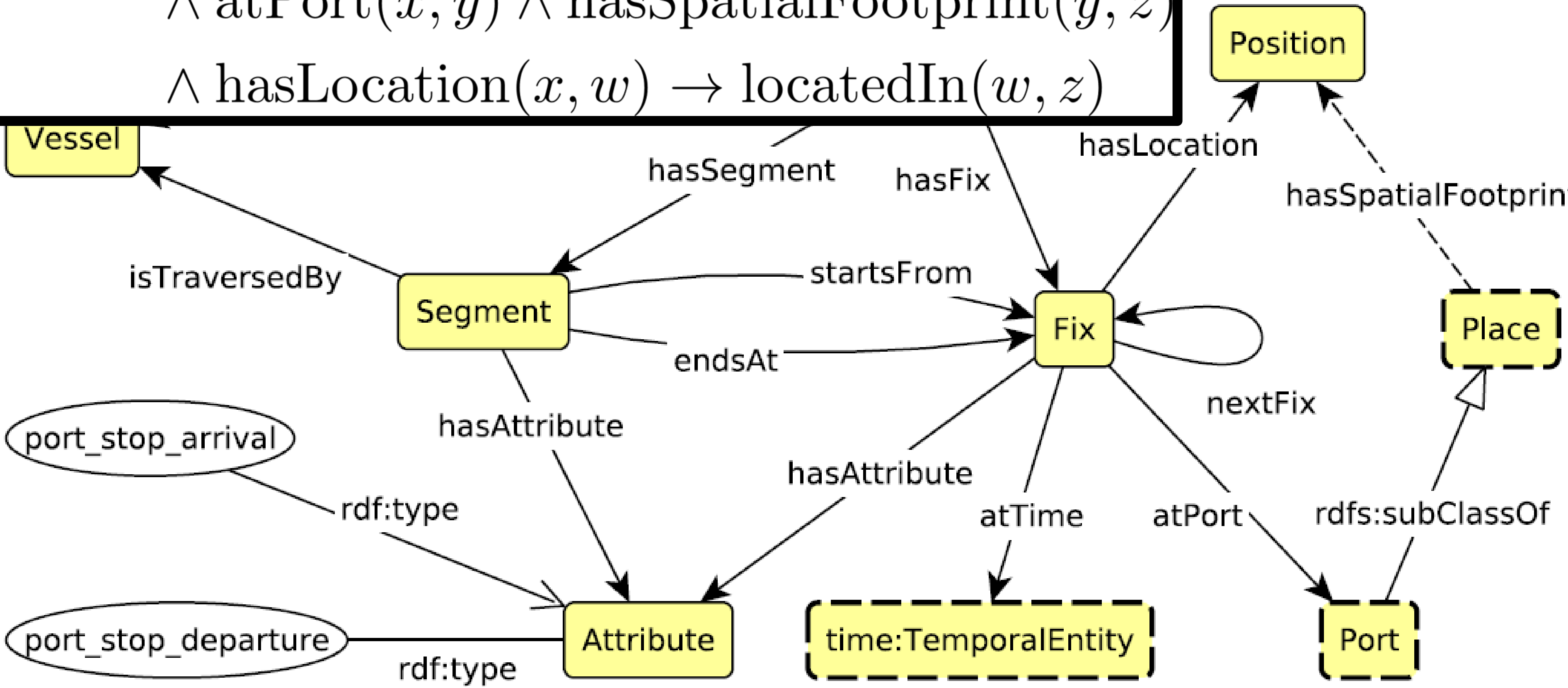
$$\text{Cruise}(x) \wedge \text{hasTrajectory}(x, y) \wedge \text{hasSegment}(y, z) \wedge \text{isTraversedBy}(z, v) \rightarrow \text{isUndertakenBy}(x, v)$$

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

$$\text{cruise} \circ \text{hasTrajectory} \circ \text{hasSegment} \circ \text{isTraversedBy} \sqsubseteq \text{isUndertakenBy}$$

Cruise Trajectories

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



Cruise trajectory

DaSe Lab

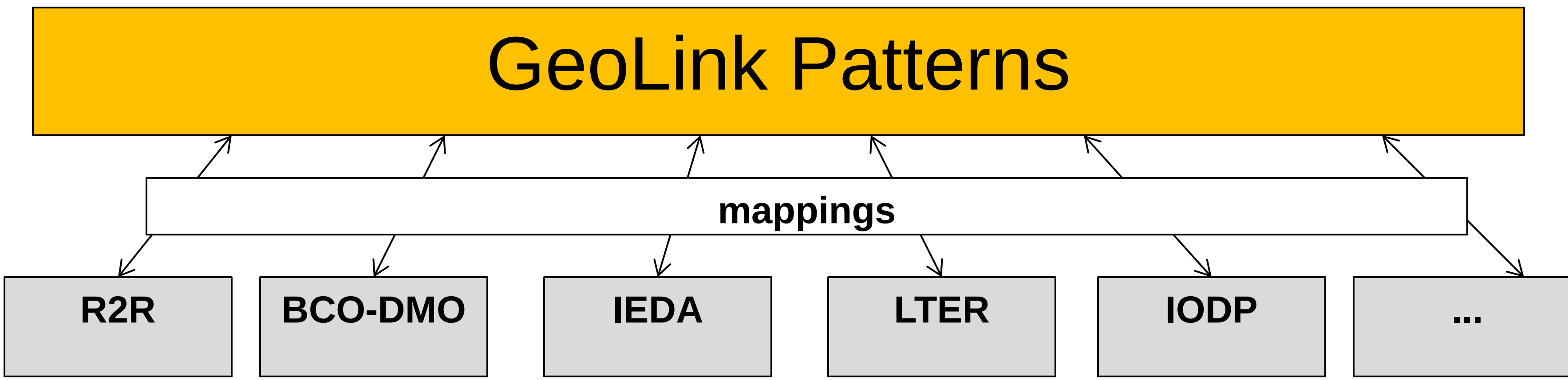
$$\begin{aligned}\mathrm{Fix}(x) &\wedge \mathrm{hasAttribute}(x, \mathrm{portStopArrival})\\ &\wedge \mathrm{atPort}(x, y) \wedge \mathrm{hasSpatialFootprint}(y, z)\\ &\wedge \mathrm{hasLocation}(x, w) \rightarrow \mathrm{locatedIn}(w, z)\end{aligned}$$

$$\mathrm{Fix} \sqcap \exists \mathrm{hasTrajectory}.\{\mathrm{portStopArrival}\} \equiv \exists \mathrm{fixps}.\mathrm{Self}$$

$$\mathrm{hasLocation}^{-} \circ \mathrm{fixps} \circ \mathrm{atPort} \circ \mathrm{hasSpatialFootprint} \sqsubseteq \mathrm{locatedIn}$$

Patterns as interchange format

- Aggregated data can be “pulled back” along the same mappings, if desired.
- Since the patterns are very generic, there is no loss of information by using them as interchange format.



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.

Some central questions

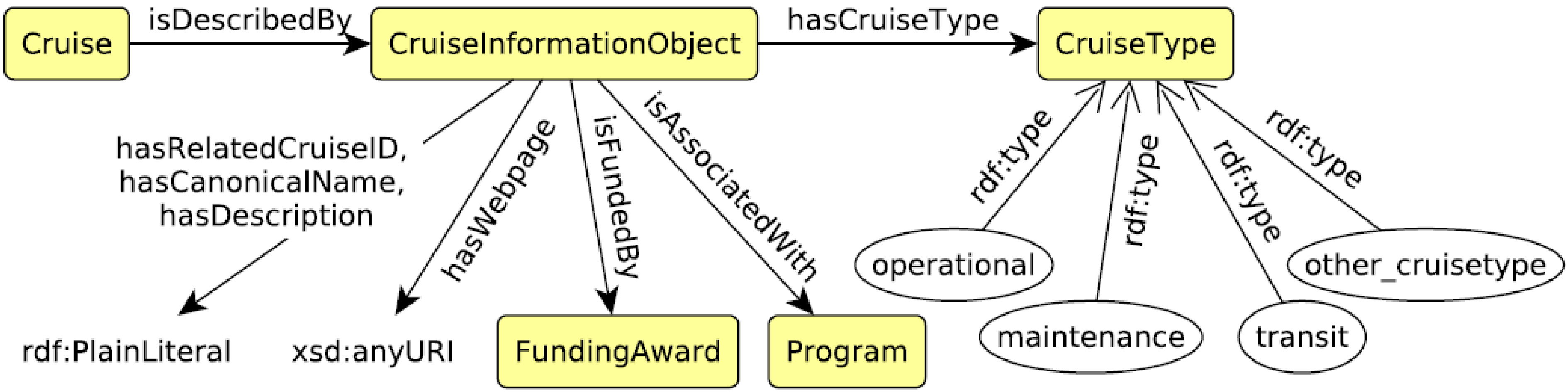
- ODPs as subject of study
- Understanding generic versus specific modeling in patterns.
- Developing pattern languages and tools
- Understanding and formalizing relationships between patterns, and making systematic use of it: ecosystems of patterns
- Evaluating the added value of patterns for ontology-based tasks or applications, e.g. ontology alignment, linked data visualization, information integration, ...

Thanks!

www.oceanlink.org

www.geo-link.org

Information Objects



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.**
- **Krzysztof Janowicz, Pascal Hitzler, Benjamin Adams, Dave Kolas, Charles Vardeman II, Five Stars of Linked Data Vocabulary Use. Semantic Web 5 (3), 2014, 173-176.**

References

- 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.
- Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, *Foundations of Semantic Web Technologies*. Chapman and Hall/CRC Press, 2010.
- 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.

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.