Data Integration with Ontology Design Patterns

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

A hacker who studied ontology
Was famed for his sense of frivolity.
When his program inferred
That Clyde ISA Bird
He blamed, not his code, but zoology.

Henry Kautz
https://www.cs.rochester.edu/~kautz/misc/limericks.html

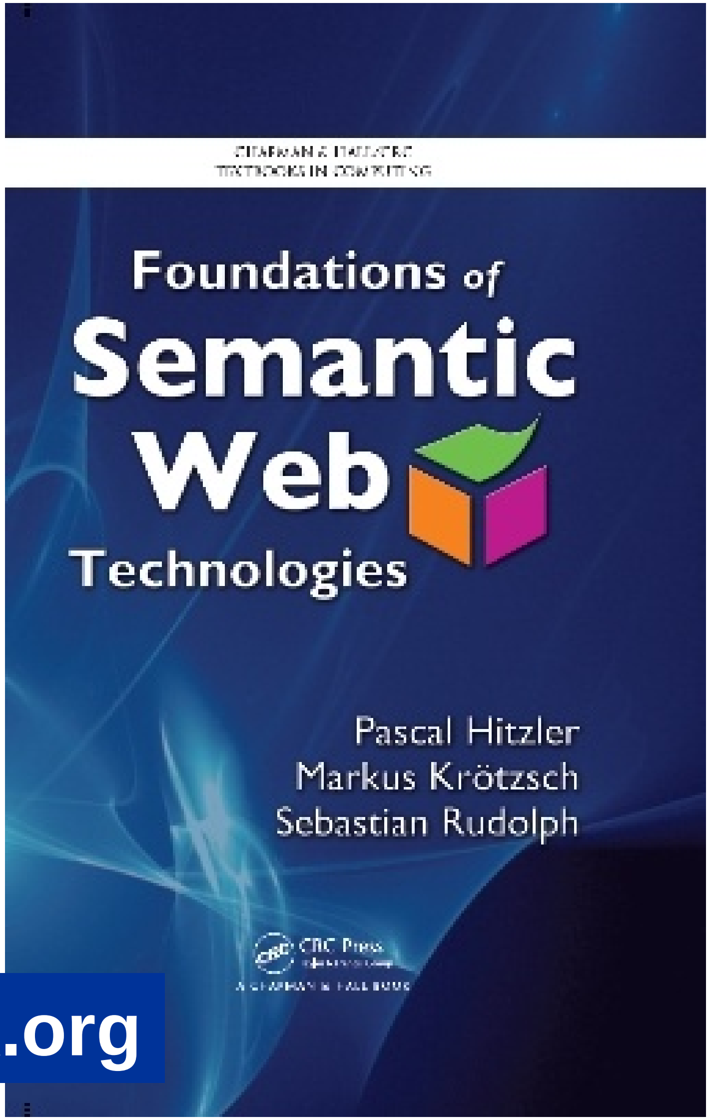
Textbook

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)

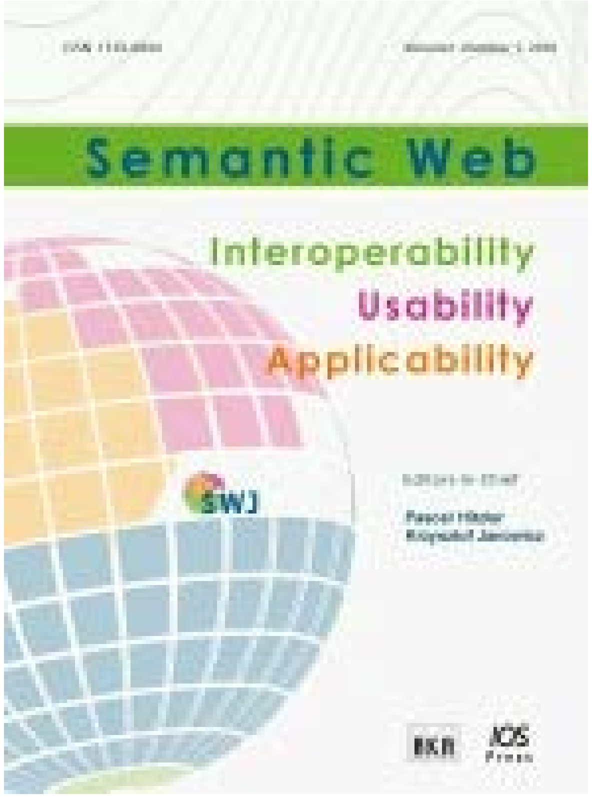


http://www.semantic-web-book.org

Semantic Web journal

DaSe Lab

- EiCs: Pascal Hitzler
  Krzysztof Janowicz
- Funded 2010
- SCImago ranks us 18th worldwide in Computer Science



- 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
- Pawel Grzebala
- 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
- data security
- applications in the sciences and elsewhere

EarthCube

EarthCube:

NSF Program, multiple projects, long run-time

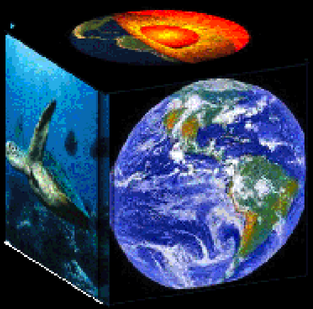
Goal: Developing a Community-Driven Data and Knowledge Environment for the Geosciences

“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 requires

- information integration
- interoperability
- conceptual modeling
- intelligent search
- data-model intercomparison
- data publishing support



Semantic Web studies

- information integration
- interoperability
- conceptual modeling
- intelligent search
- data-model intercomparison
- data publishing support

EarthCube Challenges

The EarthCube "Data Architecture" must be

- **modular**
- **extensible**
- **scalable**
- **sustainable**
- **sliceable (i.e. you can adopt part of it without adopting all)**
- **simple enough for easy adoption**
- **complex enough to solve real problems**
- **elastic, in that it allows partners to decide how much they want to share**
- **respectful of individual modeling choices**

EarthCube GeoLink project

Targeting data sharing and discovery in the Earth Sciences.

LDEO: Robert Arko, Suzanne Carbotte, Kerstin Lehnert, Peng Ji

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

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

Ocean Leadership: Douglas Fils

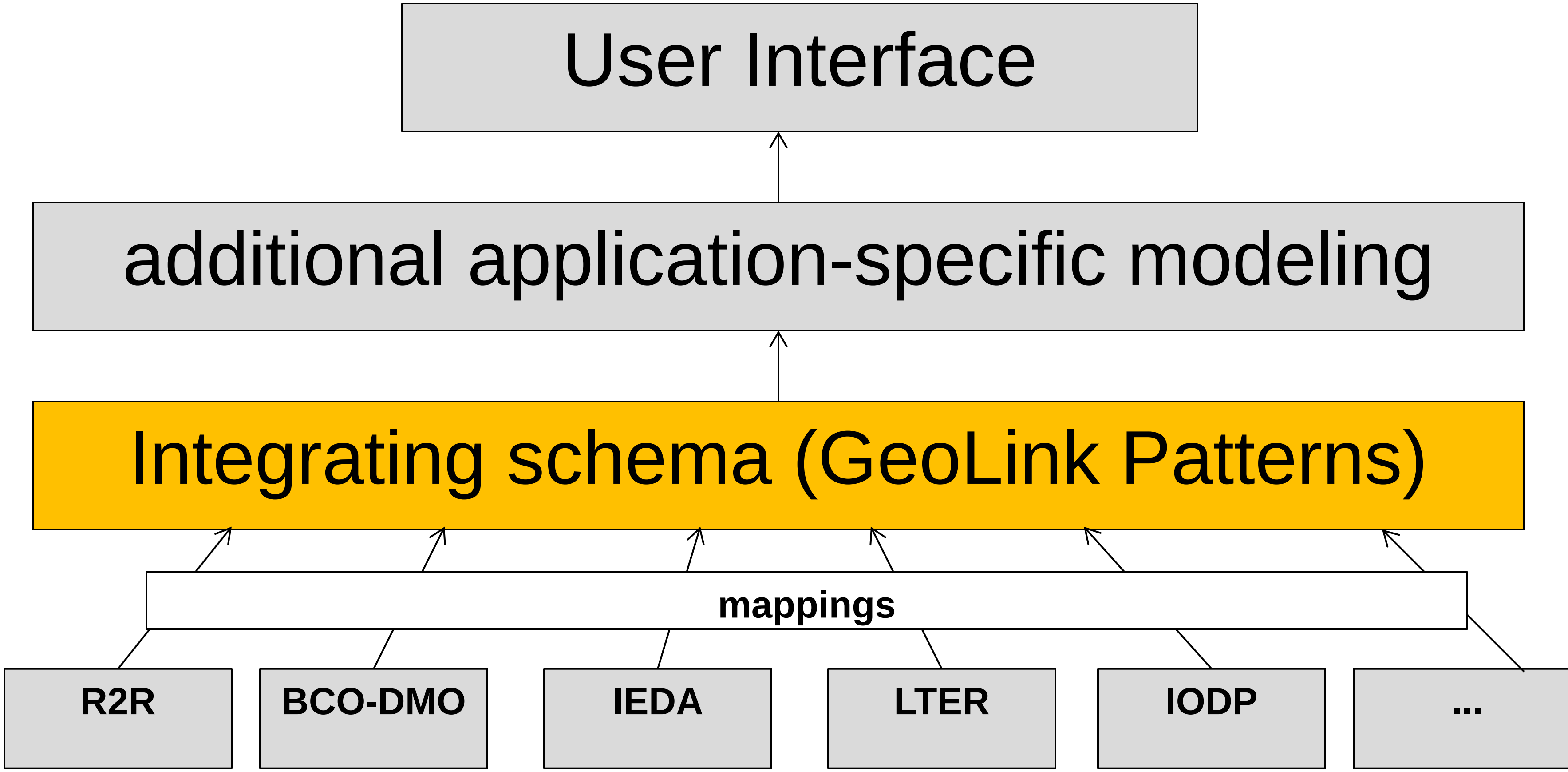
Marymount Univ: Thomas Narock

WSU: Pascal Hitzler, Michelle Cheatham, Adila Krisnadhi, Nazifa Karima, Brooke McCurdy

UMBC: Tim Finin

Featured in a January 2015 *Science* article.

GeoLink setup



Ontological Commitments

a.k.a.

modeling choices you may regret later

Ontological Commitments

http://data.kit.edu/person/pascalhitzler

owl:sameAs

http://data.wright.edu/person/pascalhitzler

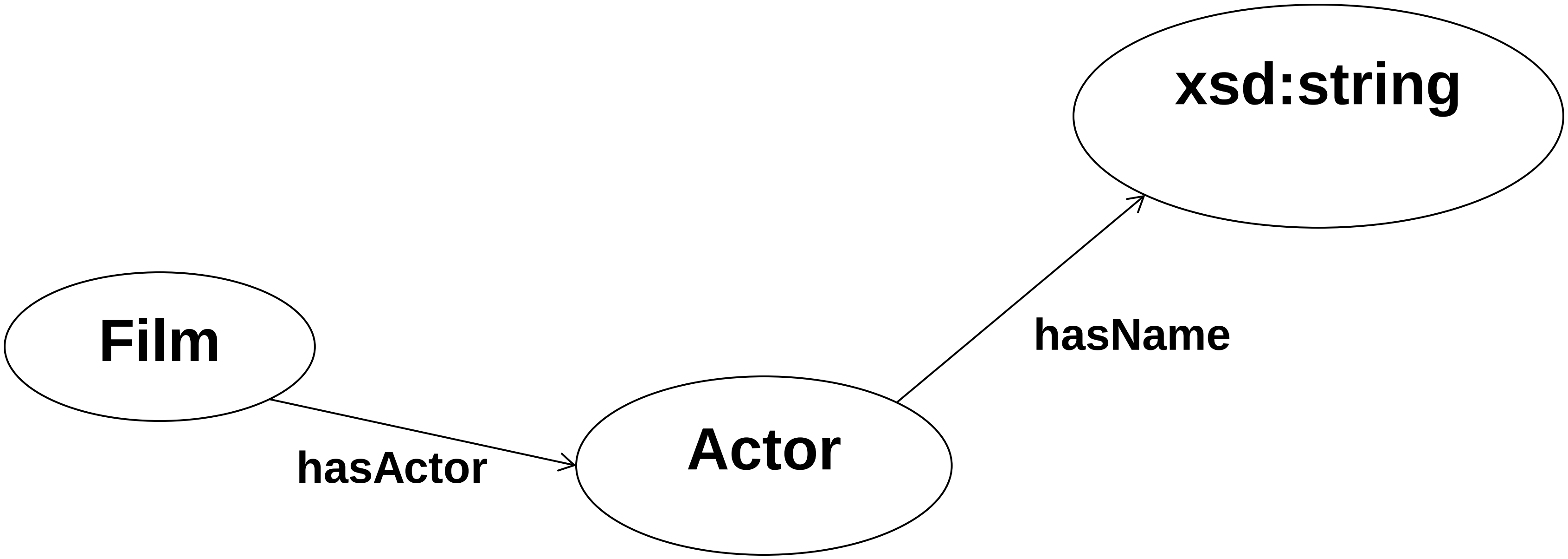
Ontological Commitments

http://data.kit.edu/person/pascalhitzler

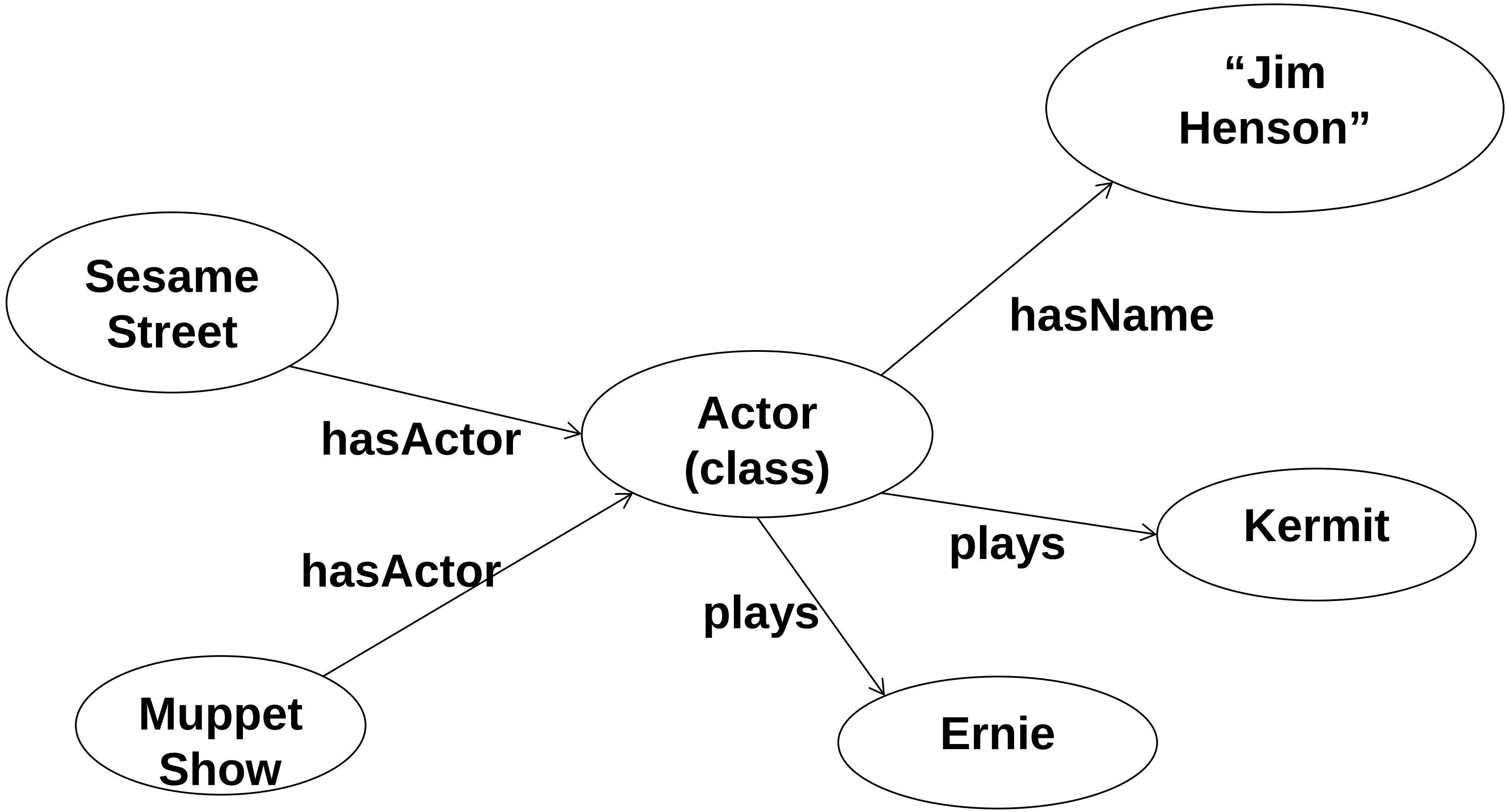
http://data.wright.edu/person/pascalhitzler

what does it mean that both may have the same ORCID ID?

From Linked MDB



Problem!



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

Multiple perspectives

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

$$
\begin{aligned}
&symmetric(b{:}hasSpouse) \\
&b{:}hasSpouse(b{:}mike, b{:}david) \\
&b{:}Male(b{:}david) \\
&b{:}Male(b{:}mike) \\
&b{:}Female(b{:}anna)
\end{aligned}
$$

Ontology Design Patterns

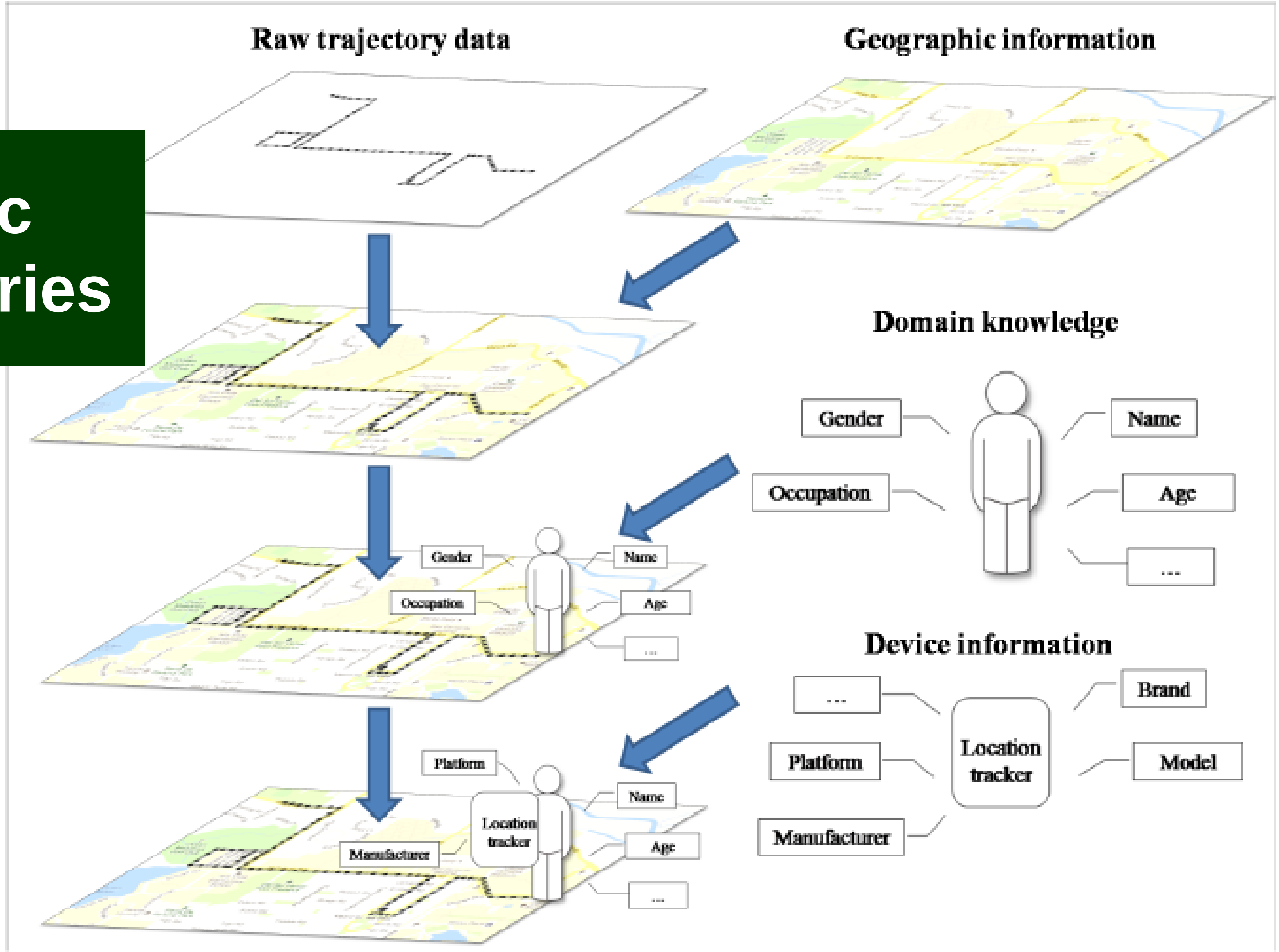
"An ontology design pattern is a reusable successful solution to a recurrent modeling problem."

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

Patterns provide modular, reusable, replaceable, pieces.

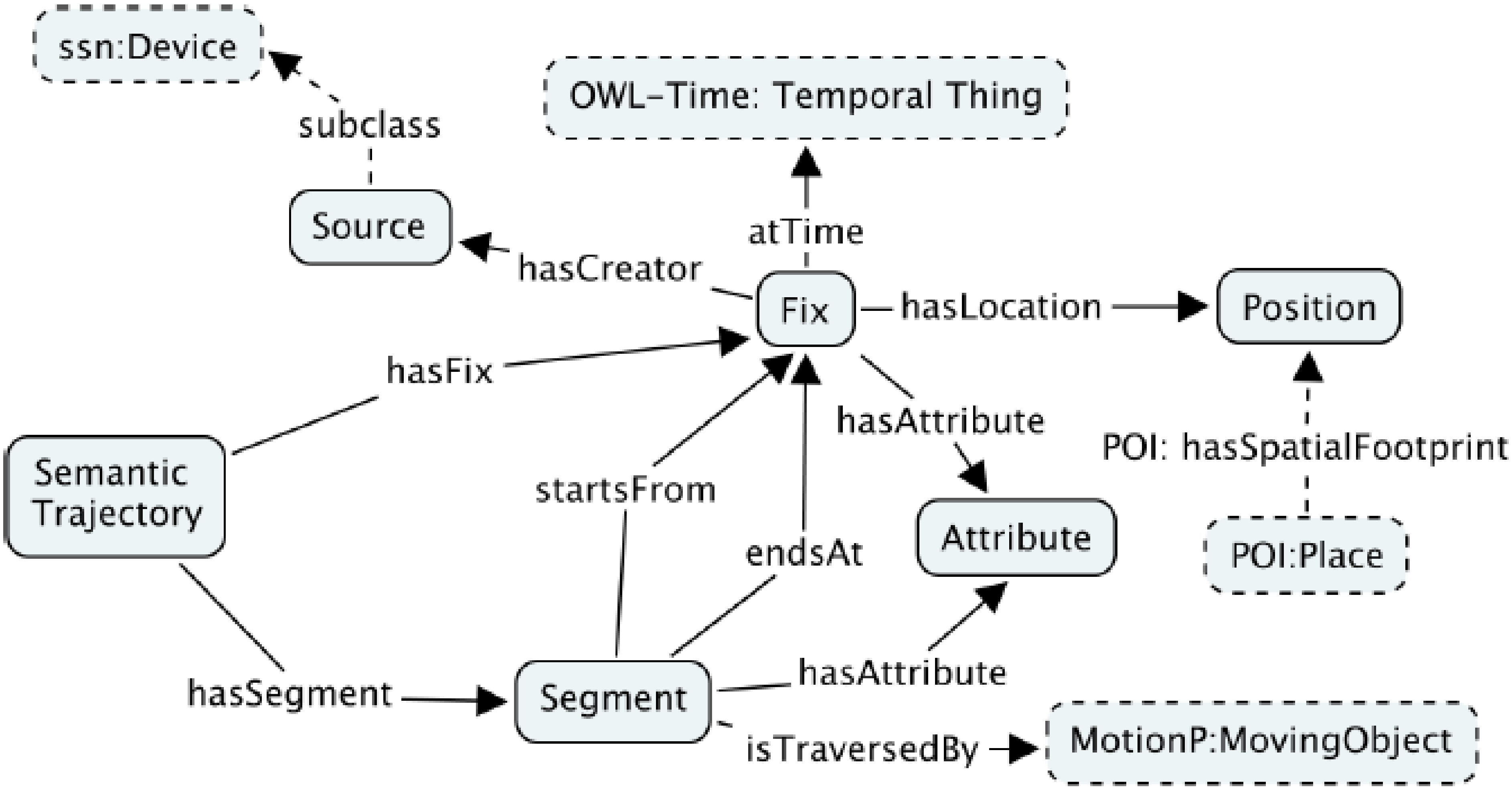
By agreeing on reuse of generic patterns (but leaving the relationships between the patterns to a specific assembly for a special purpose), we can have reuse while preserving heterogeneity.

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

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

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

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

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

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

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

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

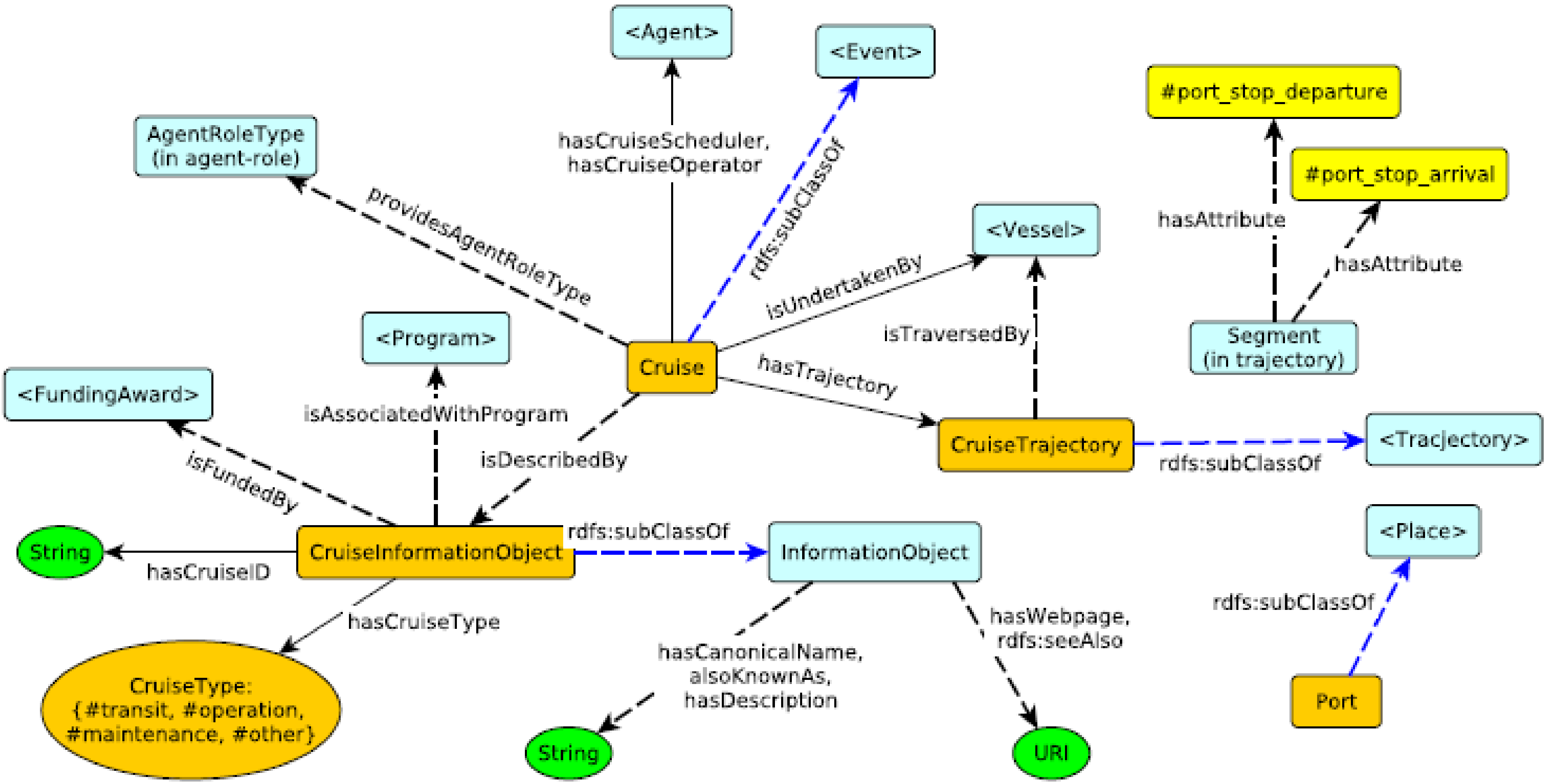
$$\exists hasSegment.Segment \sqsubseteq SemanticTrajectory \quad (18)$$

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

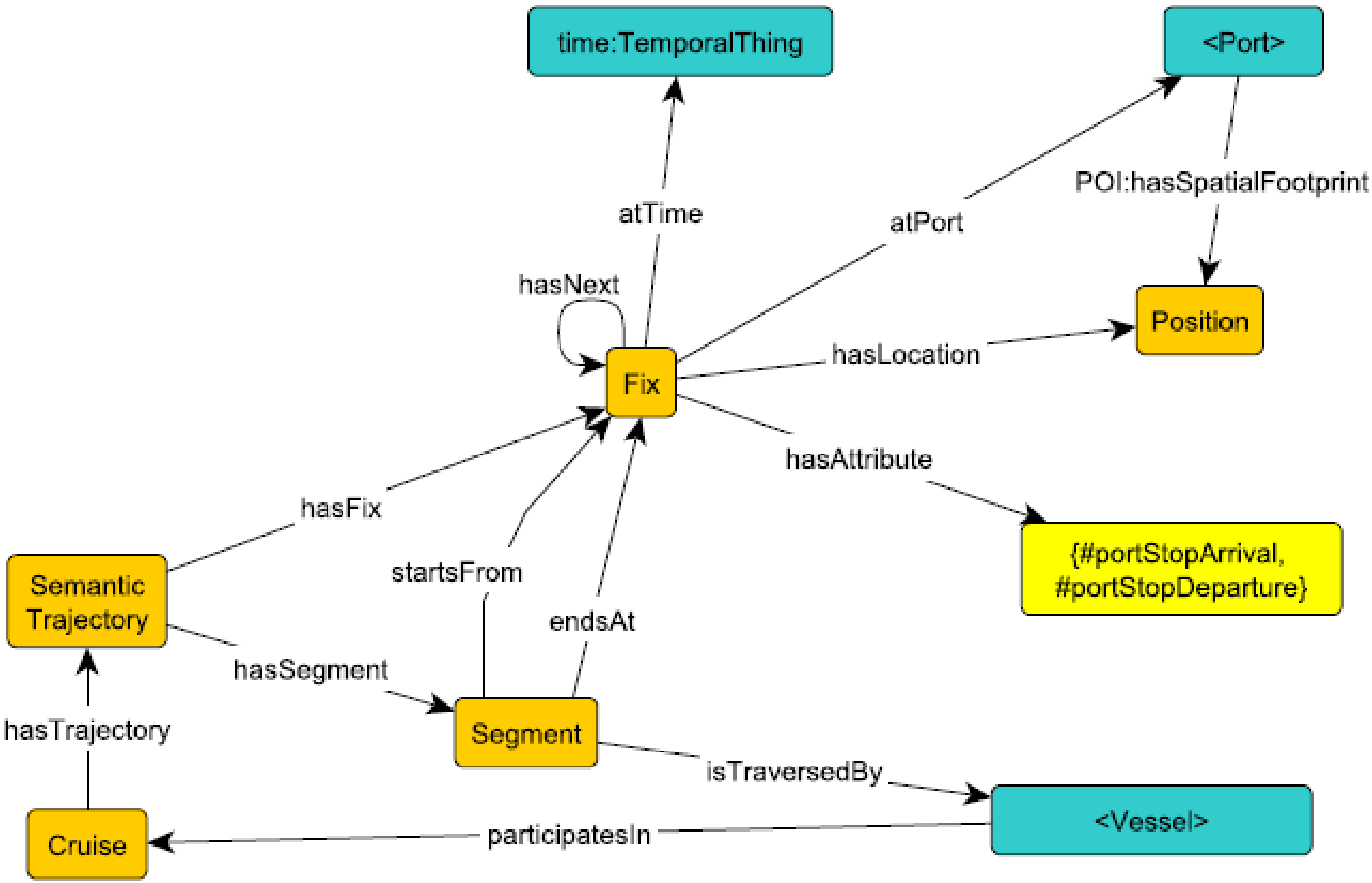
$$\exists hasFix.Segment \sqsubseteq SemanticTrajectory \quad (20)$$

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

Ocean Science Cruise (draft)

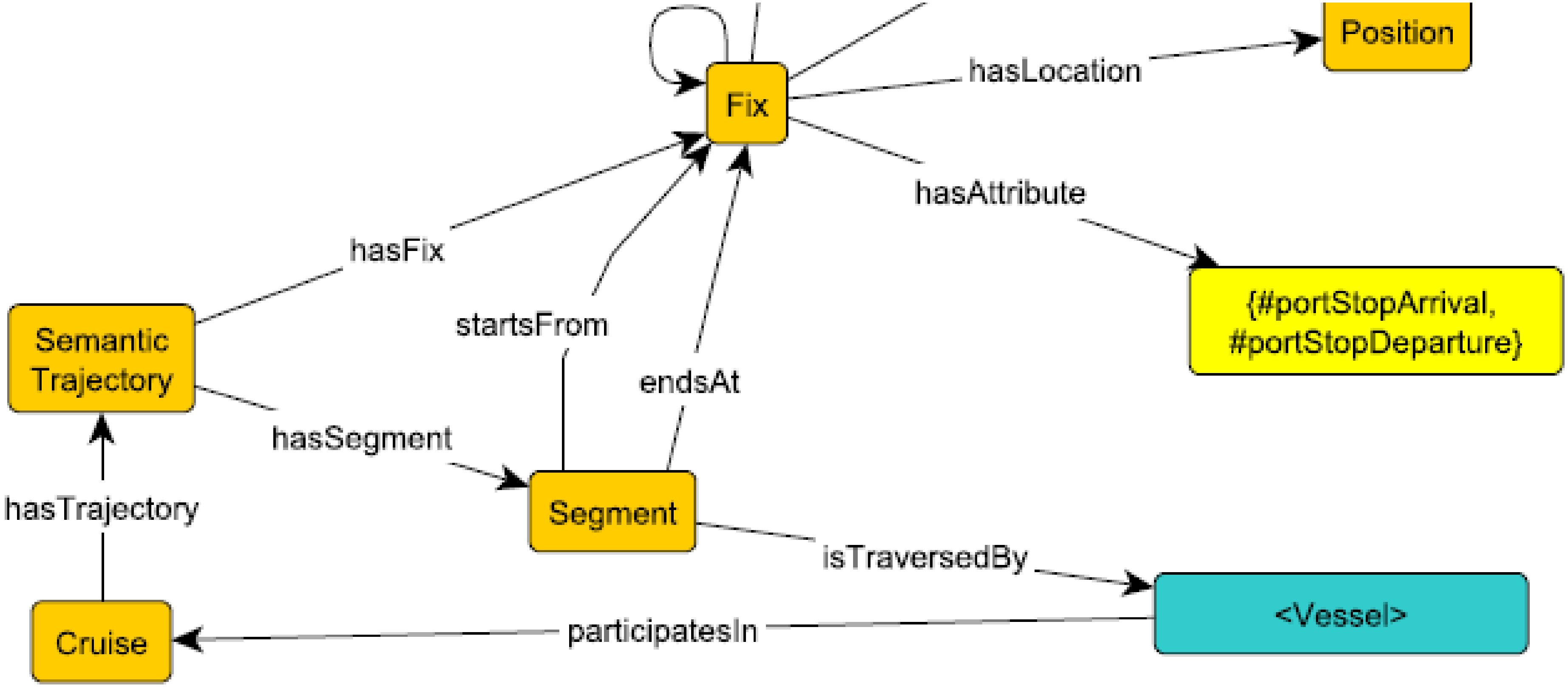


Cruise trajectory (draft)



Cruise trajectory

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



Cruise trajectory

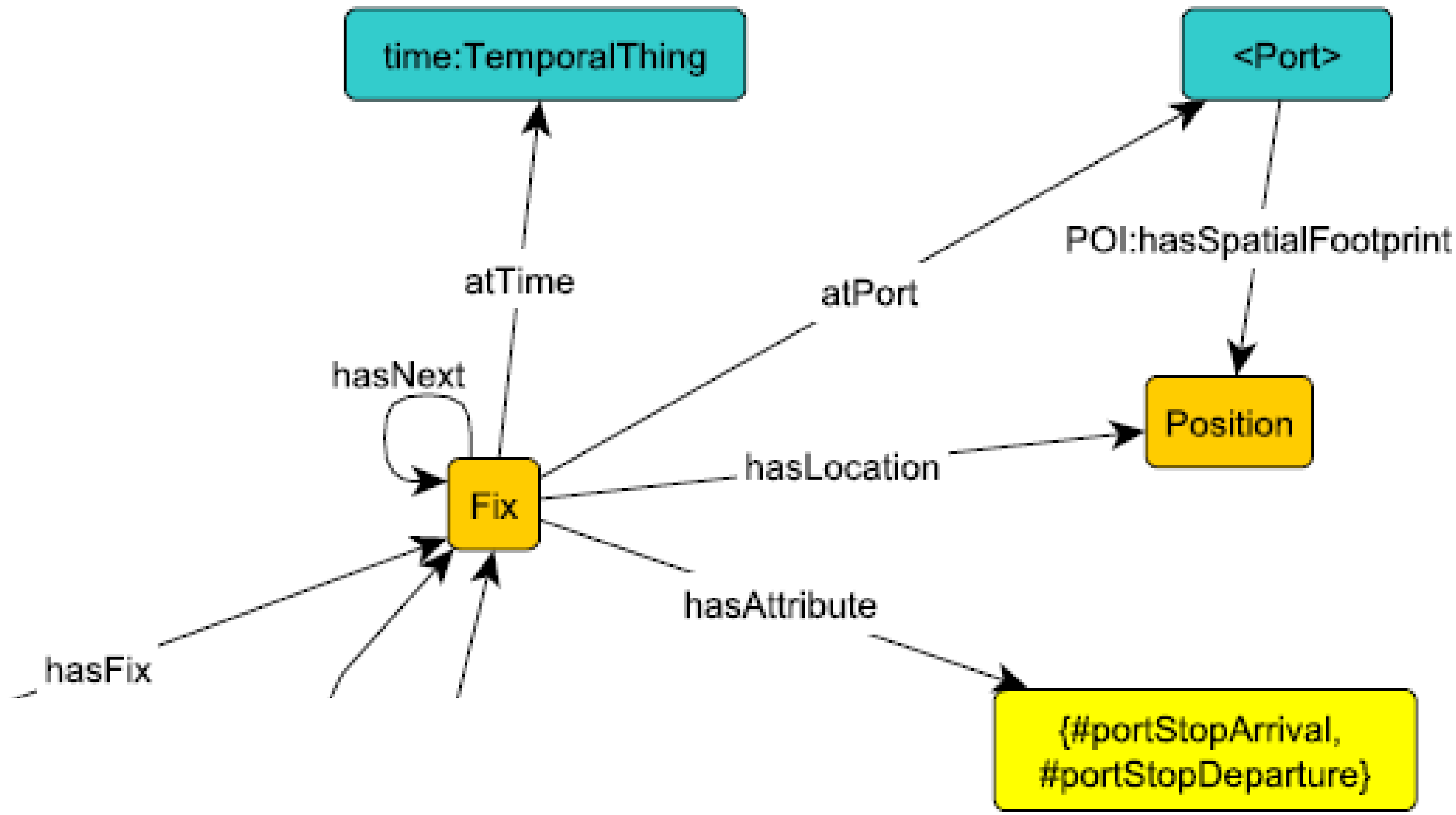
$$\text{Cruise}(x) \wedge \text{hasTrajectory}(x, y) \wedge \text{hasSegment}(y, z) \wedge \text{isTraversedBy}(z, v) \rightarrow \text{participatesIn}(v, z)$$

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

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

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

Cruise trajectory



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

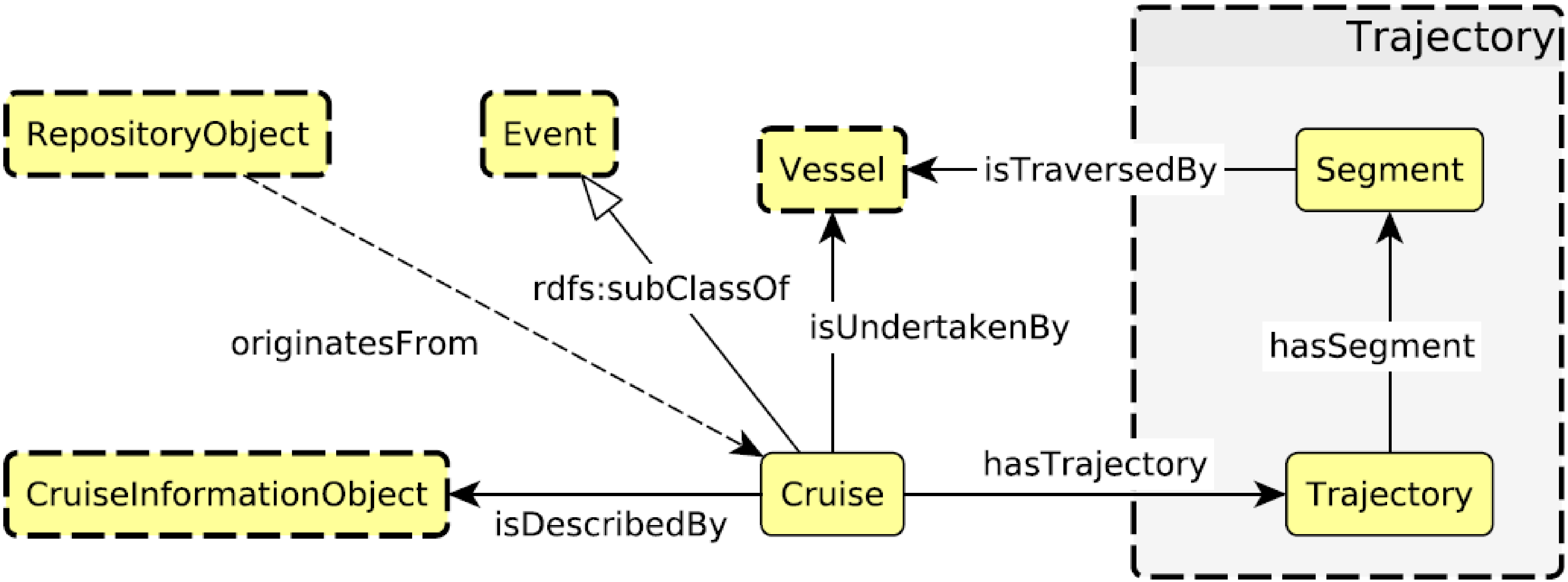
Cruise trajectory

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

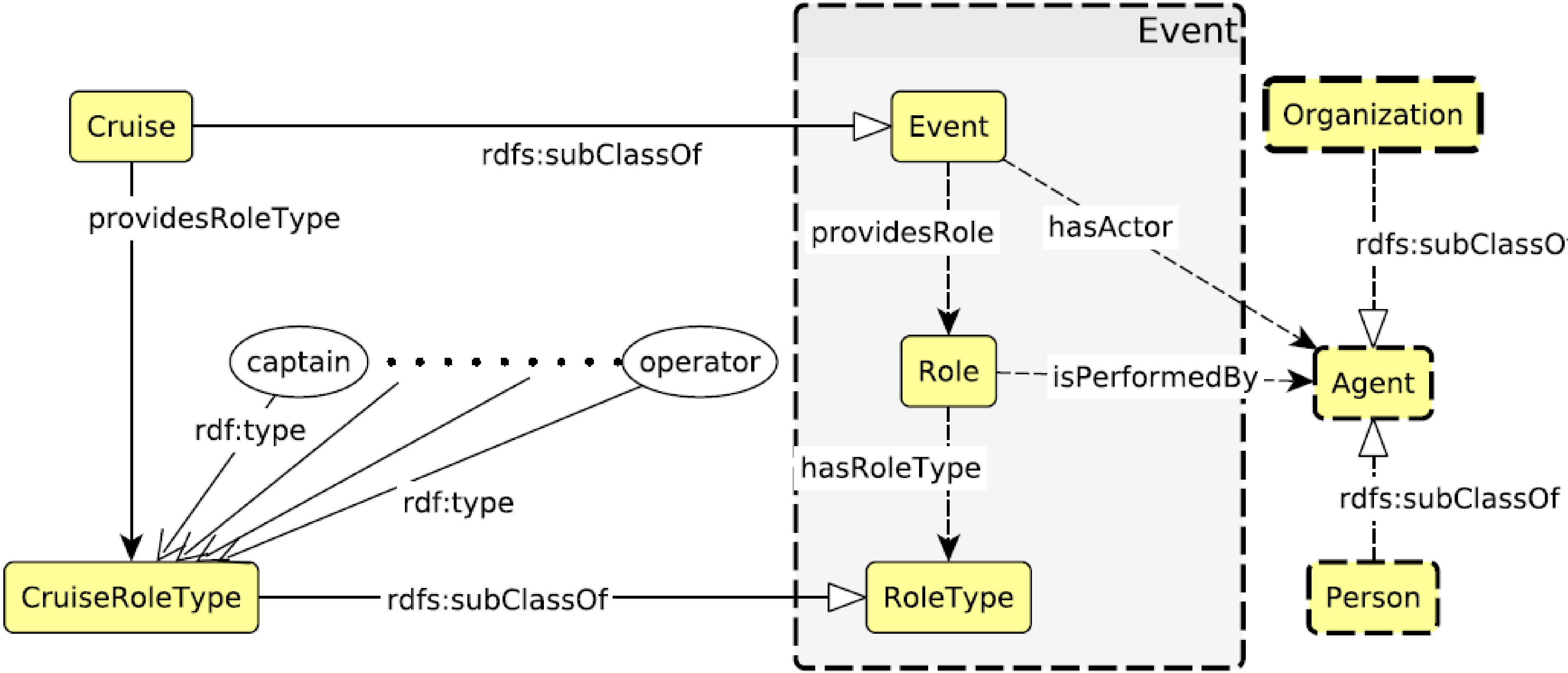
$$\mathrm{Fix} \wedge \exists \mathrm{hasTrajectory}.\{\#\mathrm{portStopArrival}\} \equiv \exists \mathrm{fixps}.\mathrm{Self}$$

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

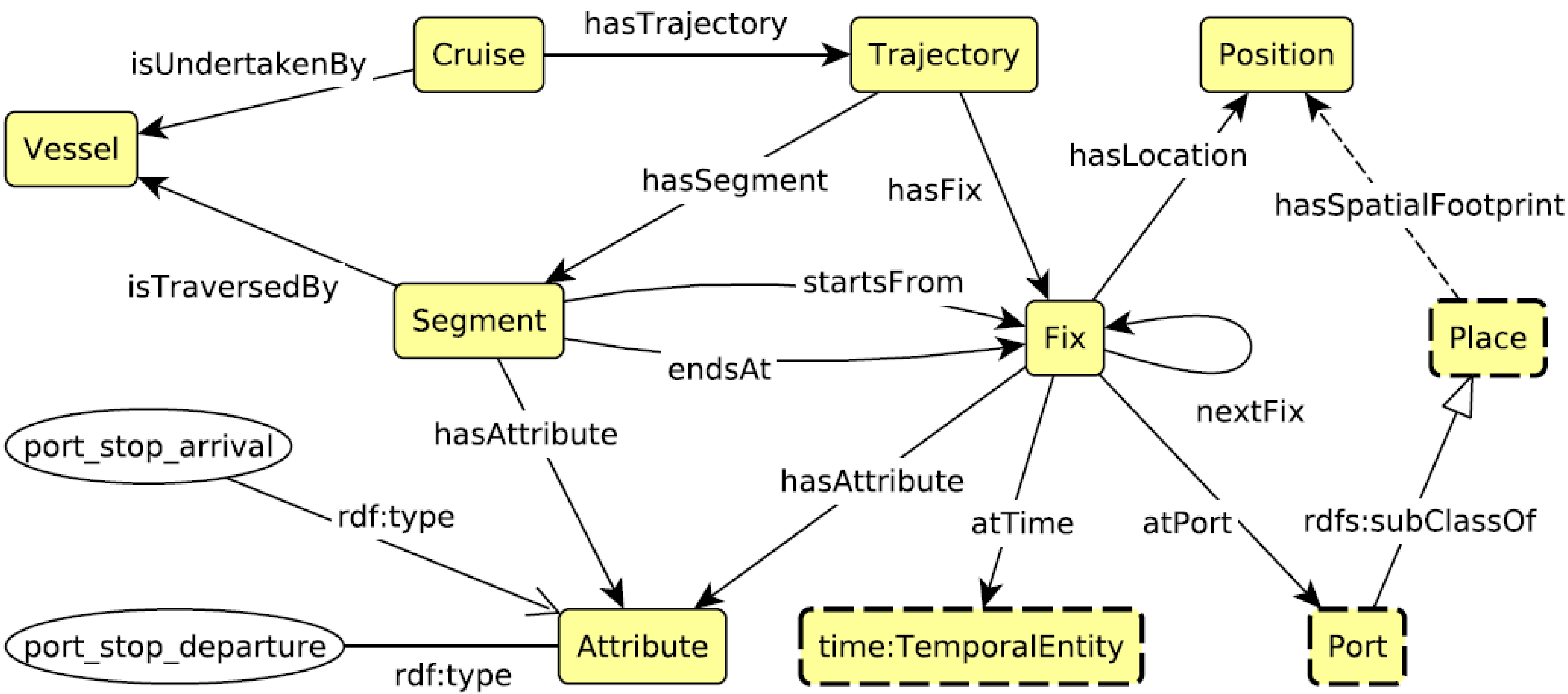
Oceanographic Cruise



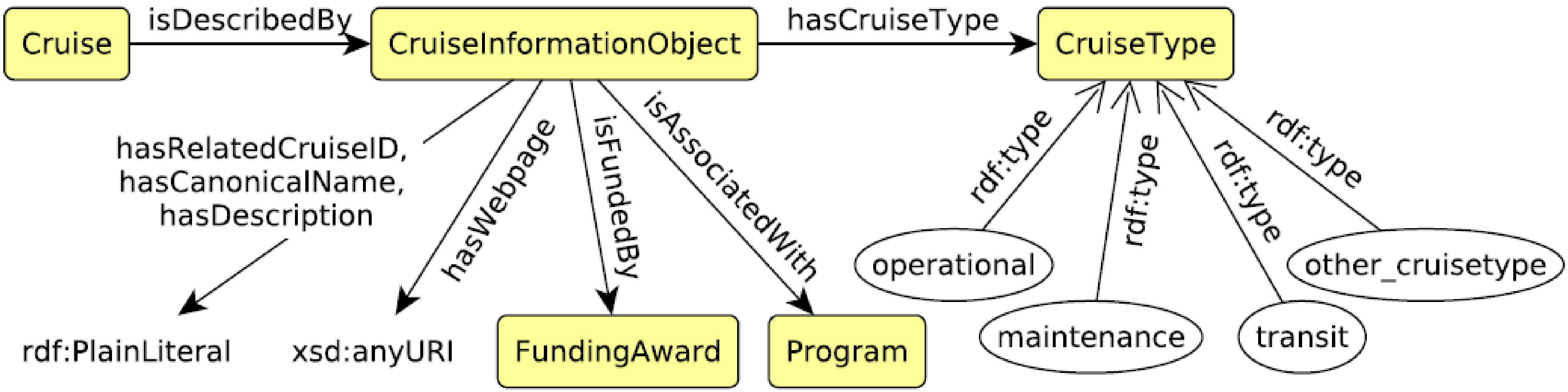
Roles (Cruise as Event)



Cruise Trajectories

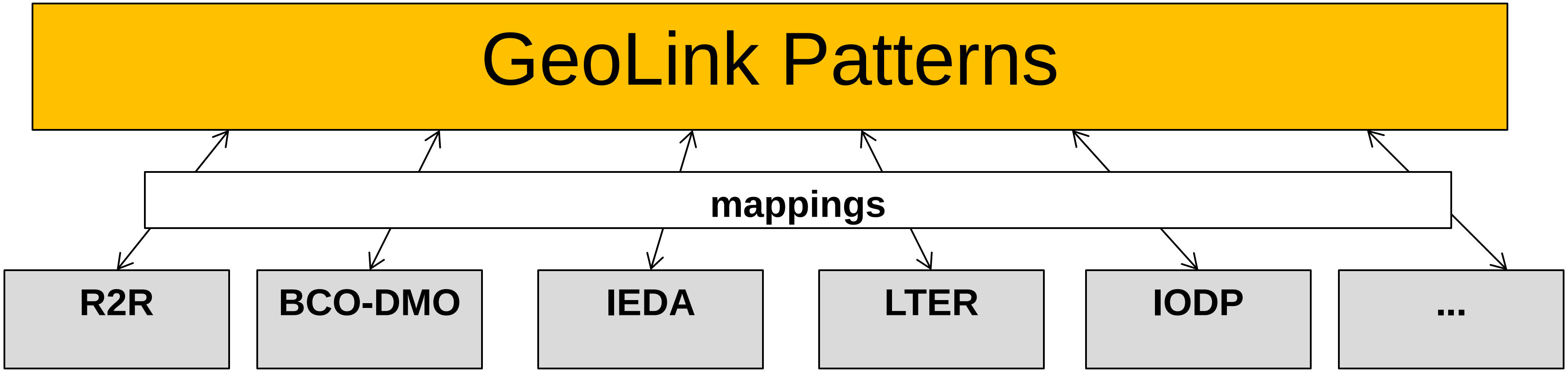


Information Objects



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.



Modeling process draft

1. 6-10 people with diverse backgrounds, including domain experts, data experts, ontology modeling experts
2. focus on one key notion, e.g. “cruise”
3. work out a set of competency questions to which the pattern shall contribute, e.g. “list all PIs of cruises going near ocean station Papa”
4. collaboratively, draft sketch of ODP, while making sure that the intended axiomatization is understood by the modeling experts
5. test sketch against competency questions and available data

Modeling process draft

Offline, modeling experts lead

1. **cleaning up sketch**
2. **detailed checking on related available patterns**
3. **rethinking class and property names**
4. **fixing axiomatization in OWL or other KR languages**
5. **possibly provision of mappings/alignment to other patterns, controlled vocabularies, etc.**

ODP paper standard layout

1. Motivating the general need for the pattern, including a discussion of what is considered in-scope and what is considered out-of-scope.

2. Intuitive presentation of graph sketch.

3. Formal presentation of OWL axiomatization, including discussion of KR aspects.

4. Presentation of use cases and pointers to available data for population.

5. Discussion of related modeling work.

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

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.

References

- Krzysztof Janowicz, Frank van Harmelen, James A. Hendler, Pascal Hitzler, Why the Data Train Needs Semantic Rails. AI Magazine. To appear.
- 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.

References

- **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.**