Ontologies in a data-driven world: Finding the middle ground

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

Krzysztof Janowicz
STKO Lab
UC Santa Barbara
http://stko.geog.ucsb.edu/

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.”**

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

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

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

Two ontologies.

Left: transportation domain

Right: agriculture domain

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

Multiple perspectives

- 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”.

The well-done ontologies

- Where are they used on the web?
- Brittle
- Expensive
- Sometimes unintuitive
- Unwieldy
- Difficult to reuse

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

The linked data counter-hype

- "Ontologies don't work, let's just link data"

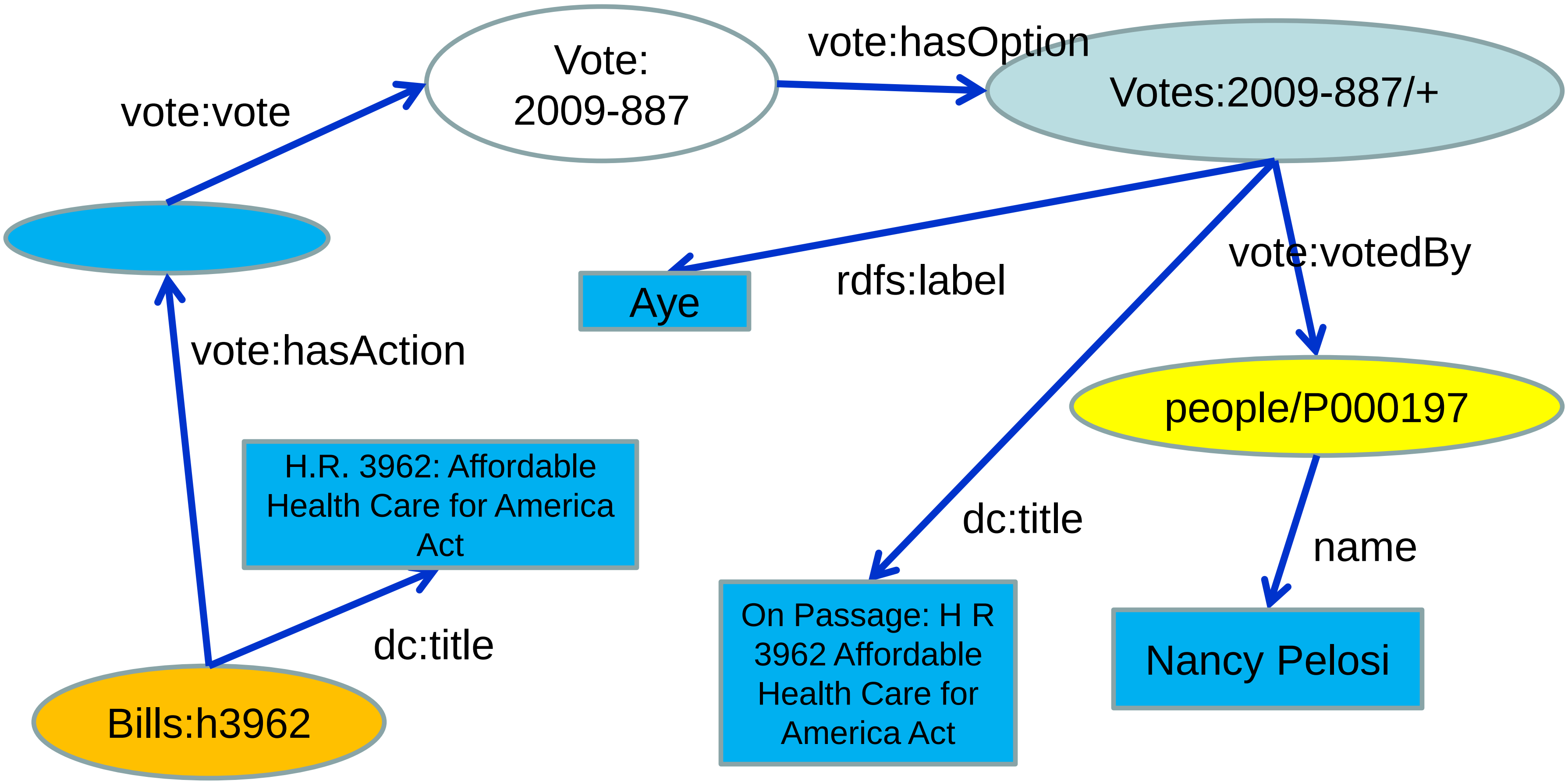
- "Okay, with a little bit of ontologies on top."

- 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?

Linked Data: Variety and Value

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



Ways forward

- Identify, which logical or conceptual depth of modeling is suitable for which purpose.

But even more importantly.

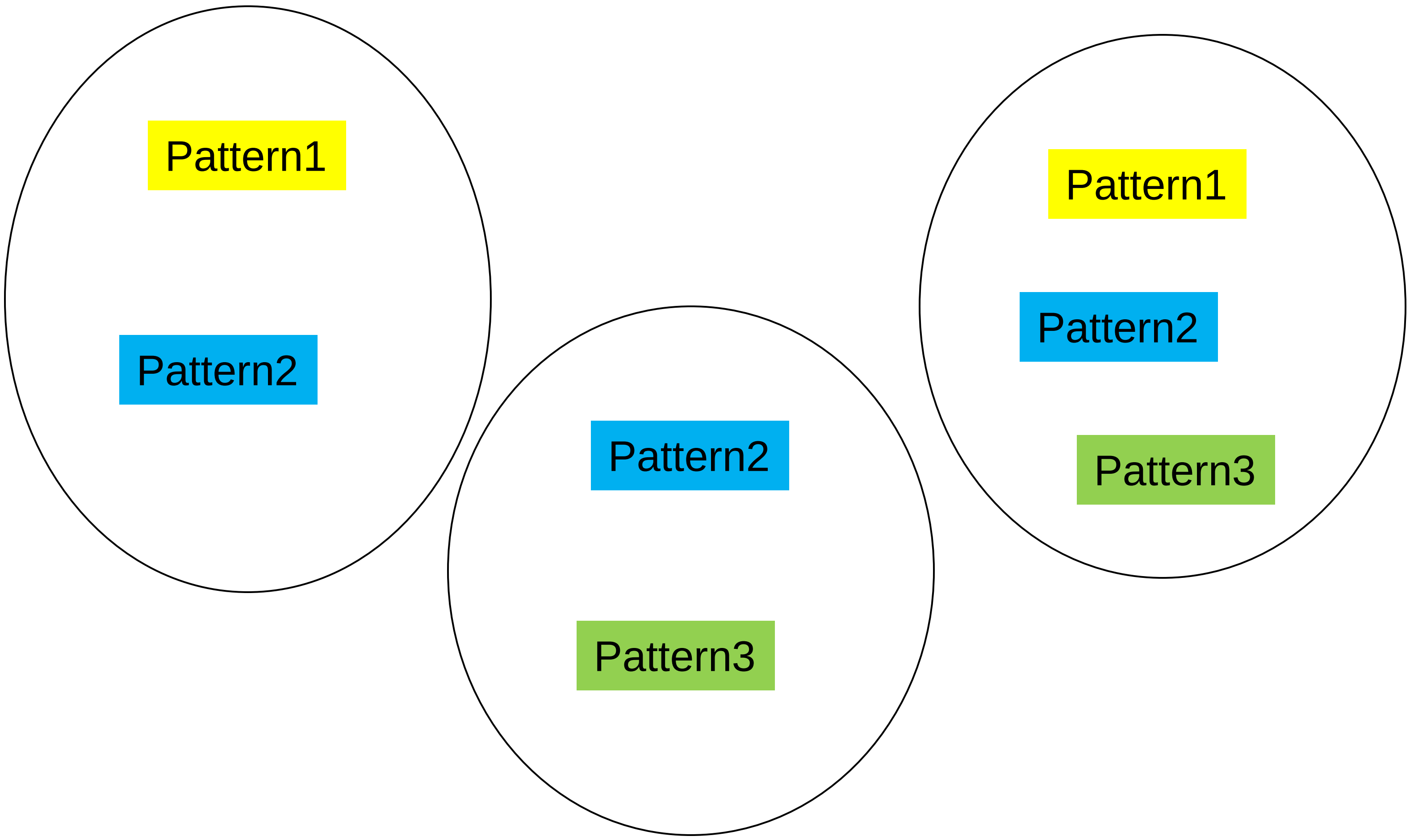
- Establish a flexible conceptual architecture using data and ontological modeling.

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.**

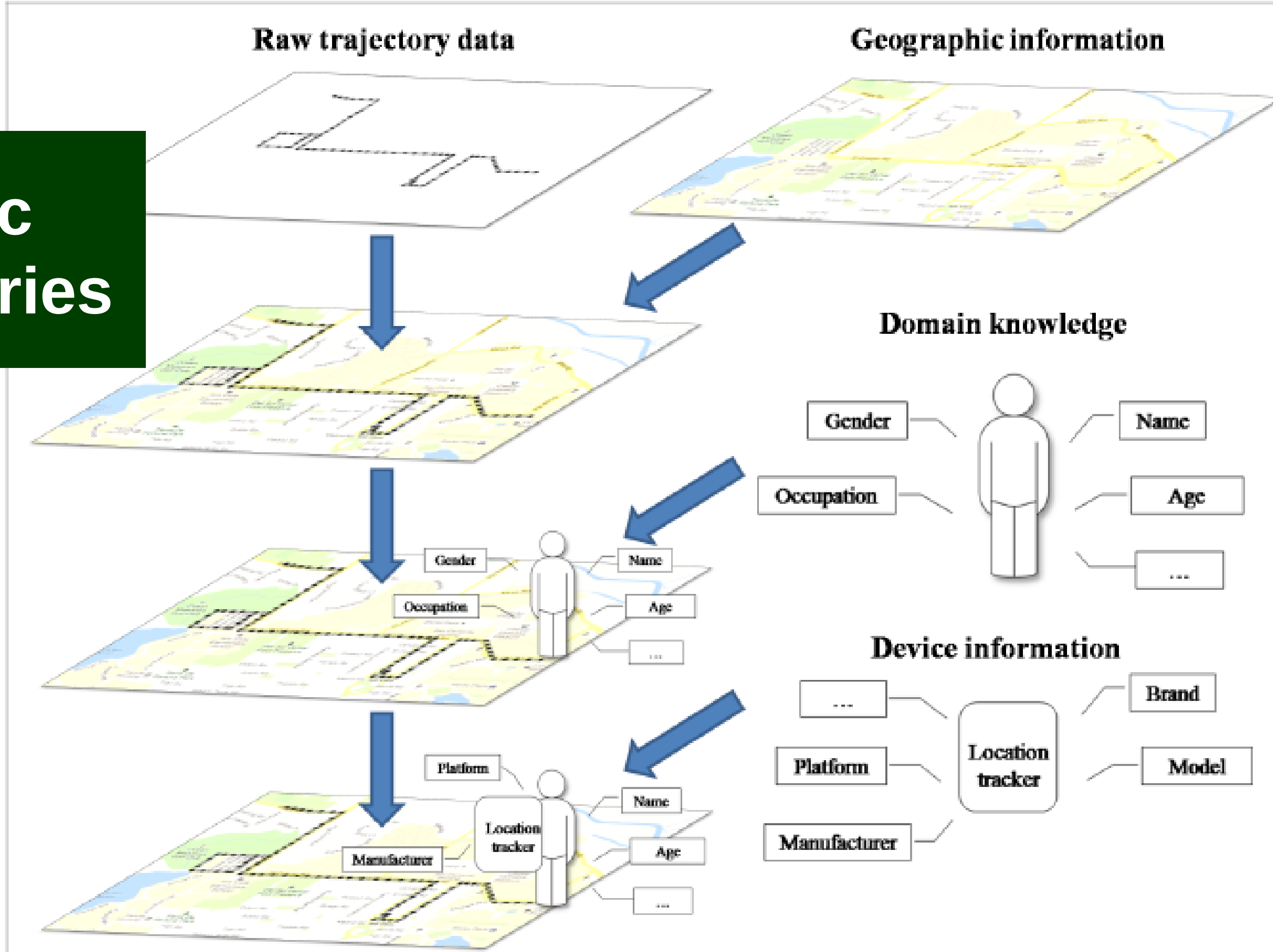
We are currently establishing many geo-patterns in a series of hands-on workshops, the GeoVoCamps, see http://vocamp.org/

Ontology Design Patterns



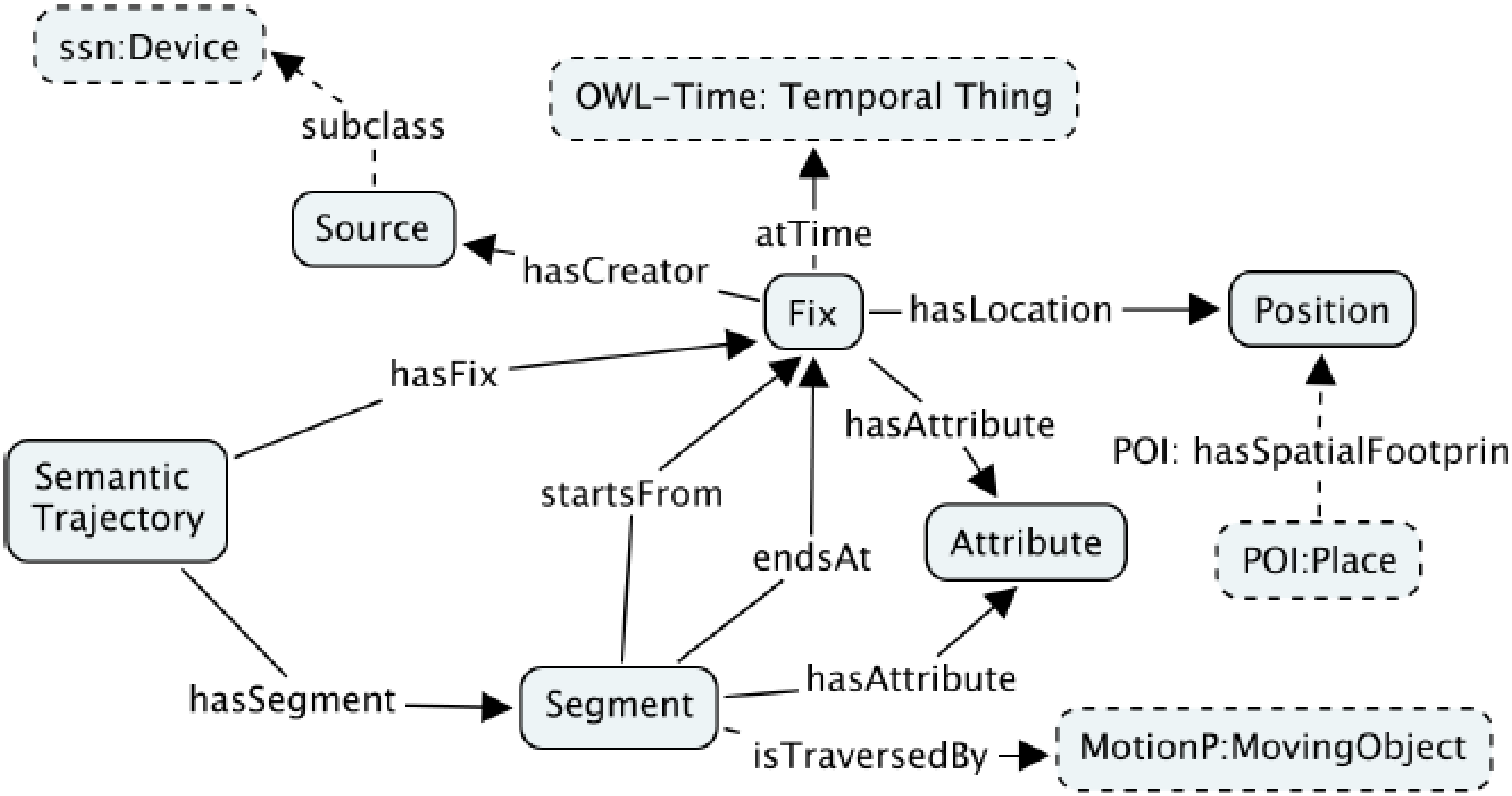
"Horizontal" alignment via patterns

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

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

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

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

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

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

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

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

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

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

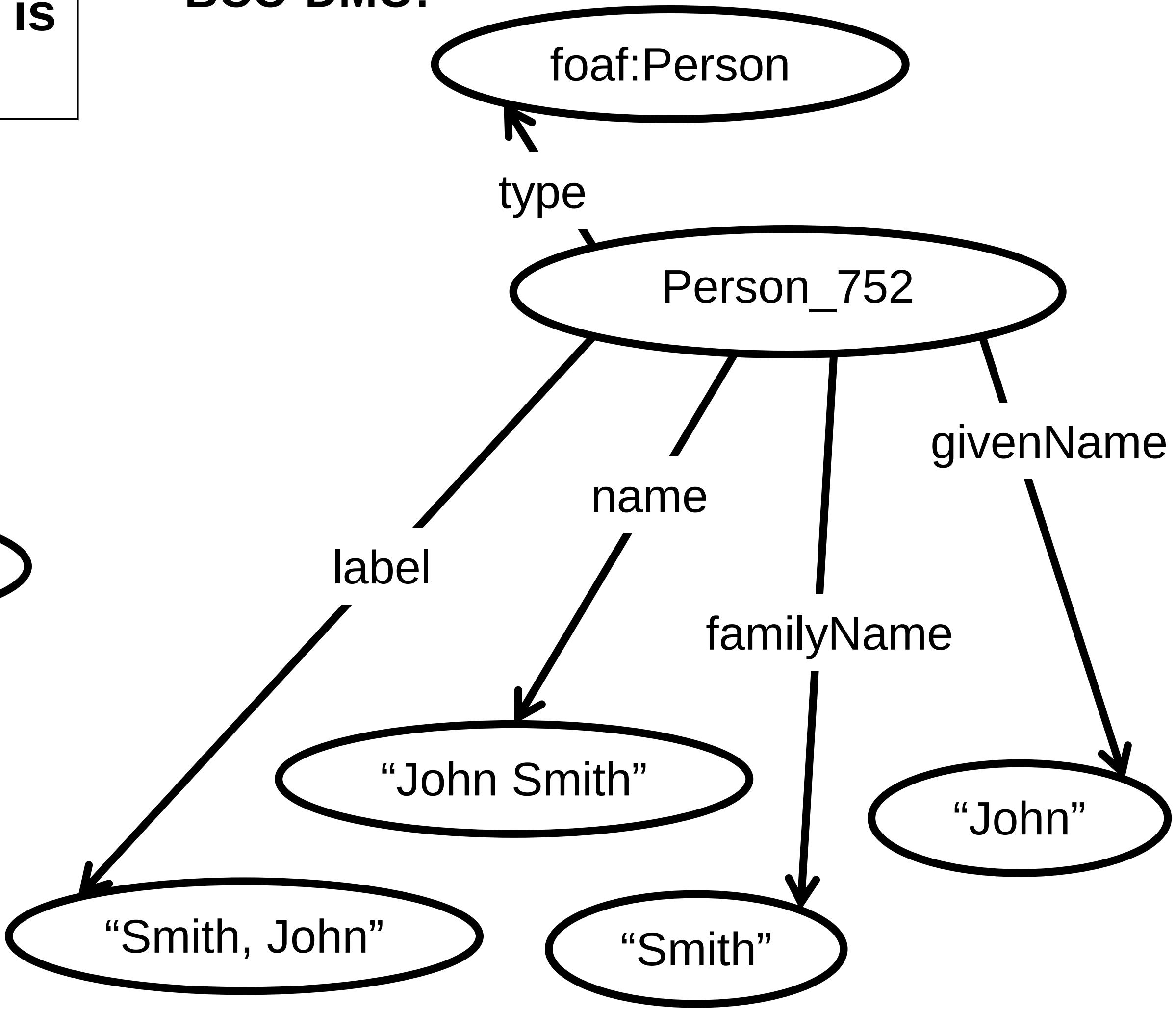
$$\exists hasFix.Segment \sqsubseteq SemanticTrajectory \quad (20)$$

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

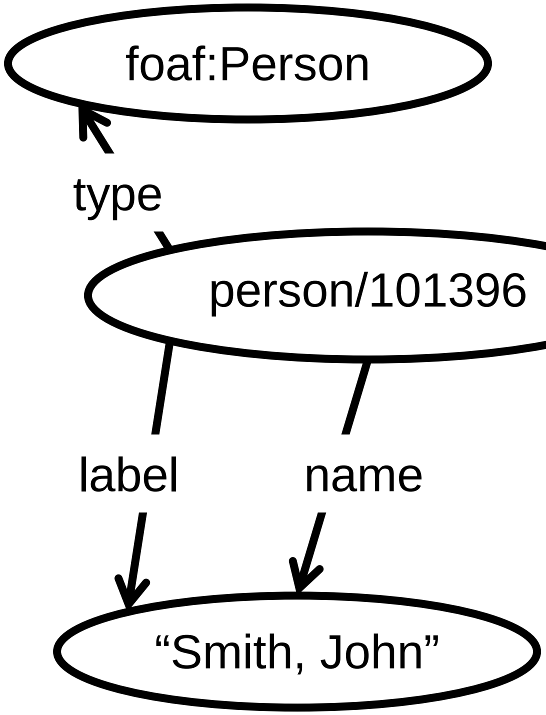
Helpfulness of patterns

Even minimalistic reuse is helpful:

BCO-DMO:



R2R:



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".

Ways forward

- Identify, which logical or conceptual depth of modeling is suitable for which purpose.

But even more importantly.

- 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.

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

- 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.
- 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

- David Carral Martinez, Krzysztof Janowicz, Pascal Hitzler, A Logical Geo-Ontology Design Pattern for Quantifying over Types. In: Isabel F. Cruz, Craig Knoblock, Peer Kröger, Egemen Tanin, Peter Widmayer (Eds.): SIGSPATIAL 2012 International Conference on Advances in Geographic Information Systems (formerly known as GIS), SIGSPATIAL'12, Redondo Beach, CA, USA, November 7-9, 2012. ACM 2012, pp. 239-248.
- David Carral, Simon Scheider, Krzysztof Janowicz, Charles Vardeman, Adila A. Krisnadhi, Pascal Hitzler, An Ontology Design Pattern for Cartographic Map Scaling. In: Philipp Cimiano, Oscar Corcho, Valentina Presutti, Laura Hollink, Sebastian Rudolph (Eds.), The Semantic Web: Semantics and Big Data. 10th International Conference, ESWC 2013, Montpellier, France, May 26-30, 2013. Proceedings. Lecture Notes in Computer Science Vol. 7882, Springer, Heidelberg, 2013, pp. 76-93.
- 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.