

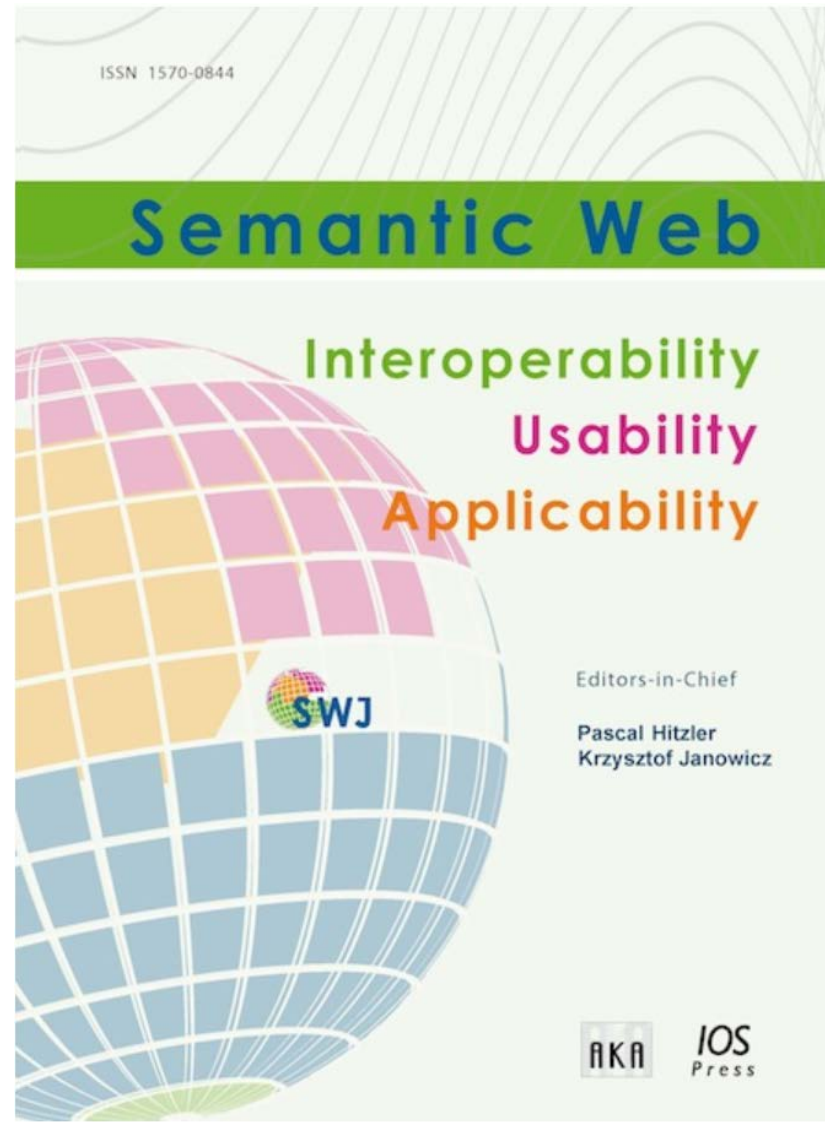
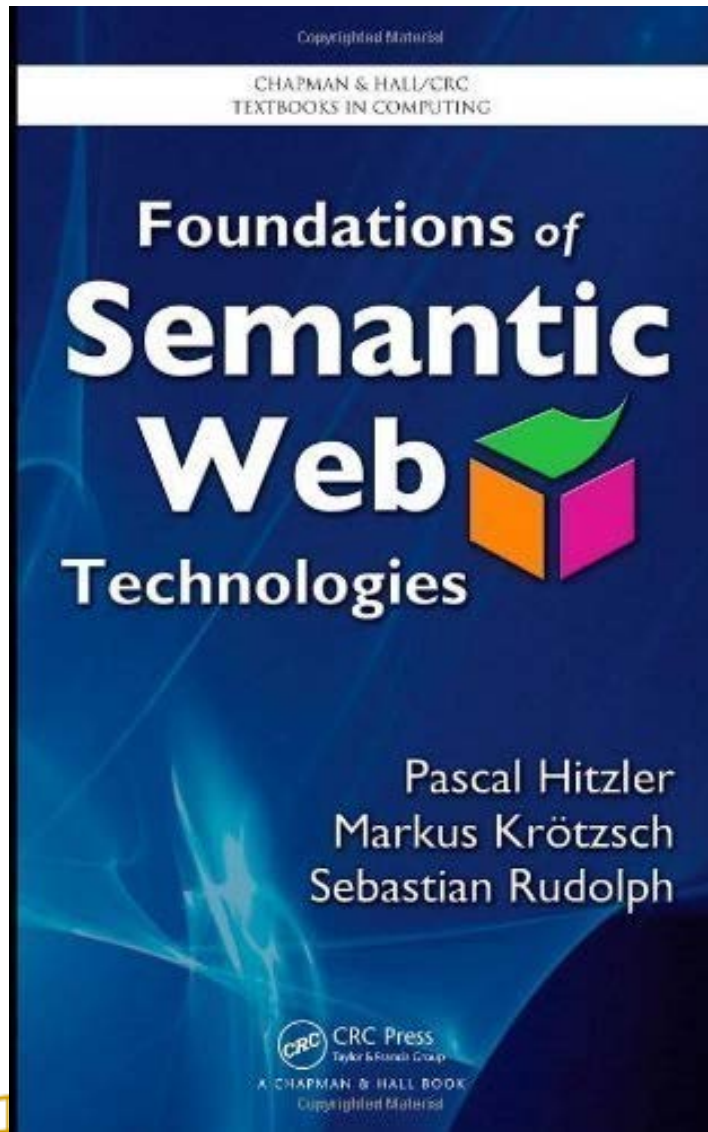


Ontologies, Vocabularies, Standards, and Linked Data

Pascal Hitzler

Data Semantics Laboratory
Data Science and Security Cluster
Wright State University
<http://daselab.org>







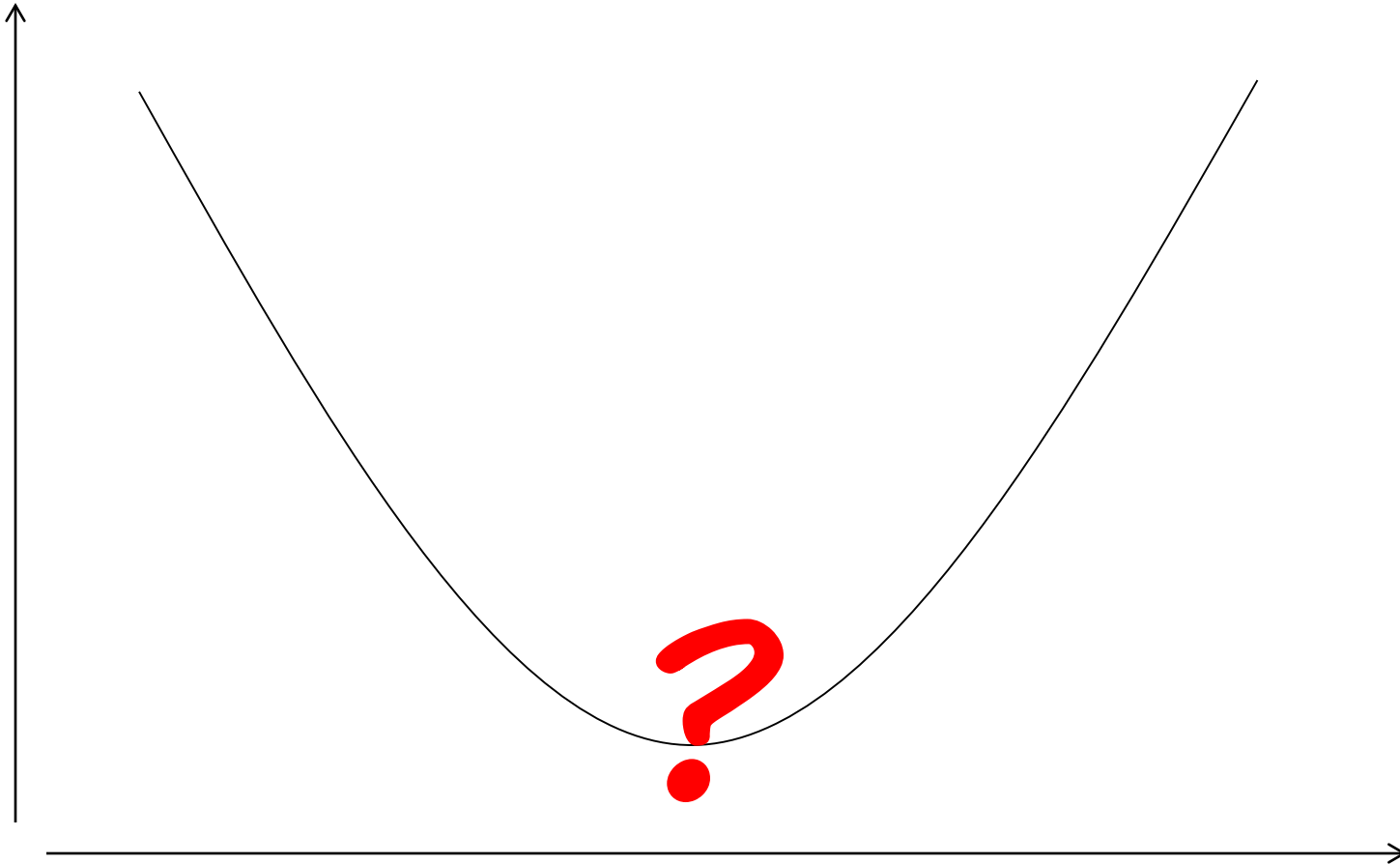
Community-generated best practices.

(Peter Fox, talk this morning)

Soft Spot Search



cost of data integration and reuse



What are the relevant dimensions?

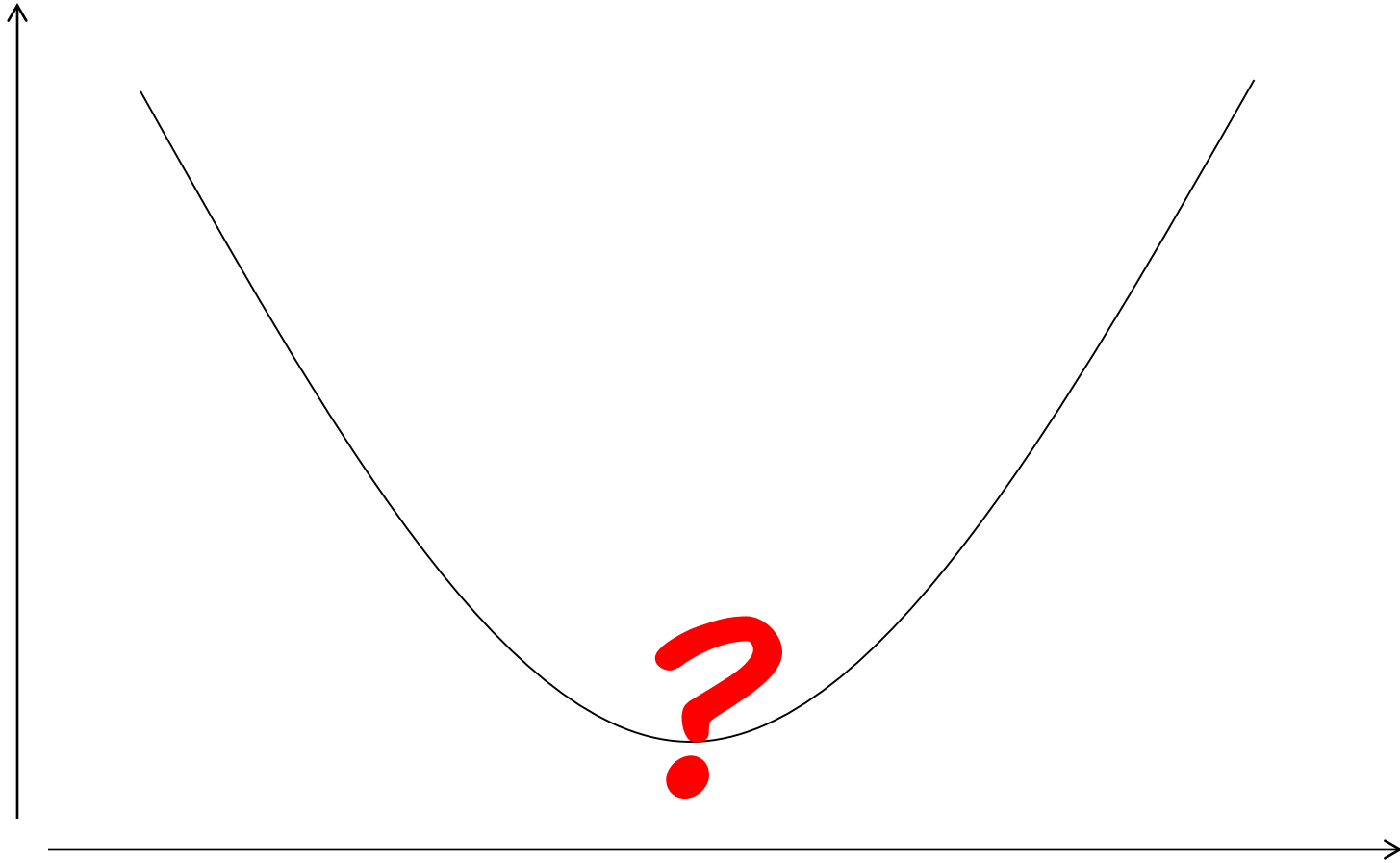
Making a world knowledge base?

Making a knowledge base of all earth science knowledge?

Soft Spot Search



cost of data integration and reuse



Too little centralization

Too much centralization

But ...



... we need to have, and be able to use,

Big Data originating from multiple, heterogeneous sources.

See Kerstin Lehnert's talk Wednesday.

So what about ...

... standardized vocabularies (or ontologies)?

**Can we really just agree on what a forest is?
or a mountain?**



... and consider this settled once and for all?

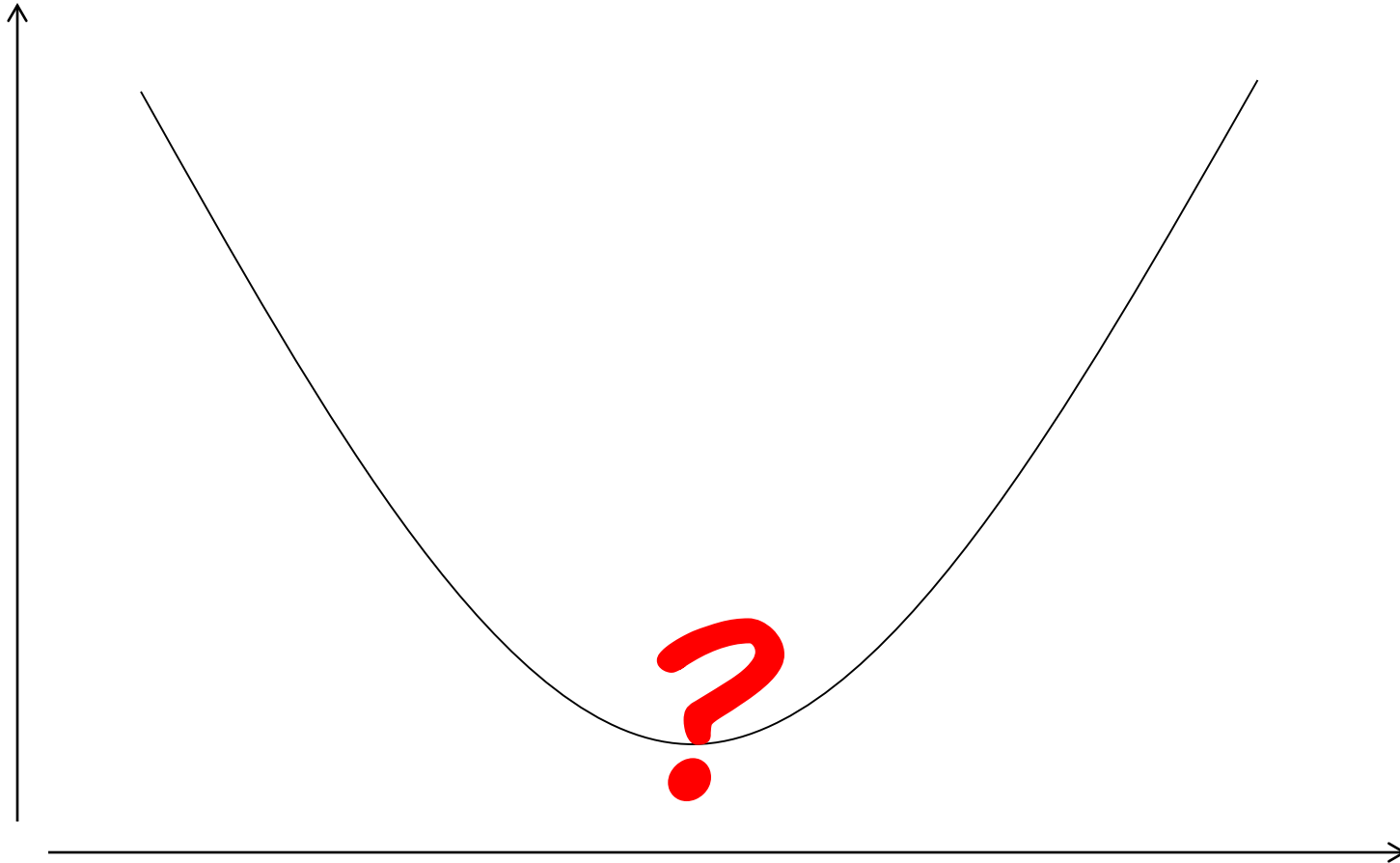
**Standards are useful in many cases.
But they are rigid.**

For every standard there is a use case which doesn't fit it.

Soft Spot Search



cost of data integration and reuse



Not enough use of standards

Too much use of standards



**Instead of standardizing vocabulary,
standardize a language for describing vocabularies.**

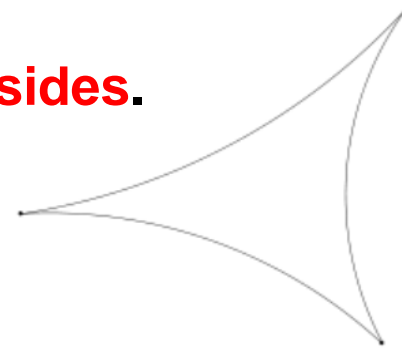
**Instead of aiming at global knowledge,
restrict to topics (specific content domains).**

This led to the rise of (large) domain ontologies.

In an ontology, you state relationships between *your* terms in a machine-readable form.



A **triangle** has exactly three **straight sides**.



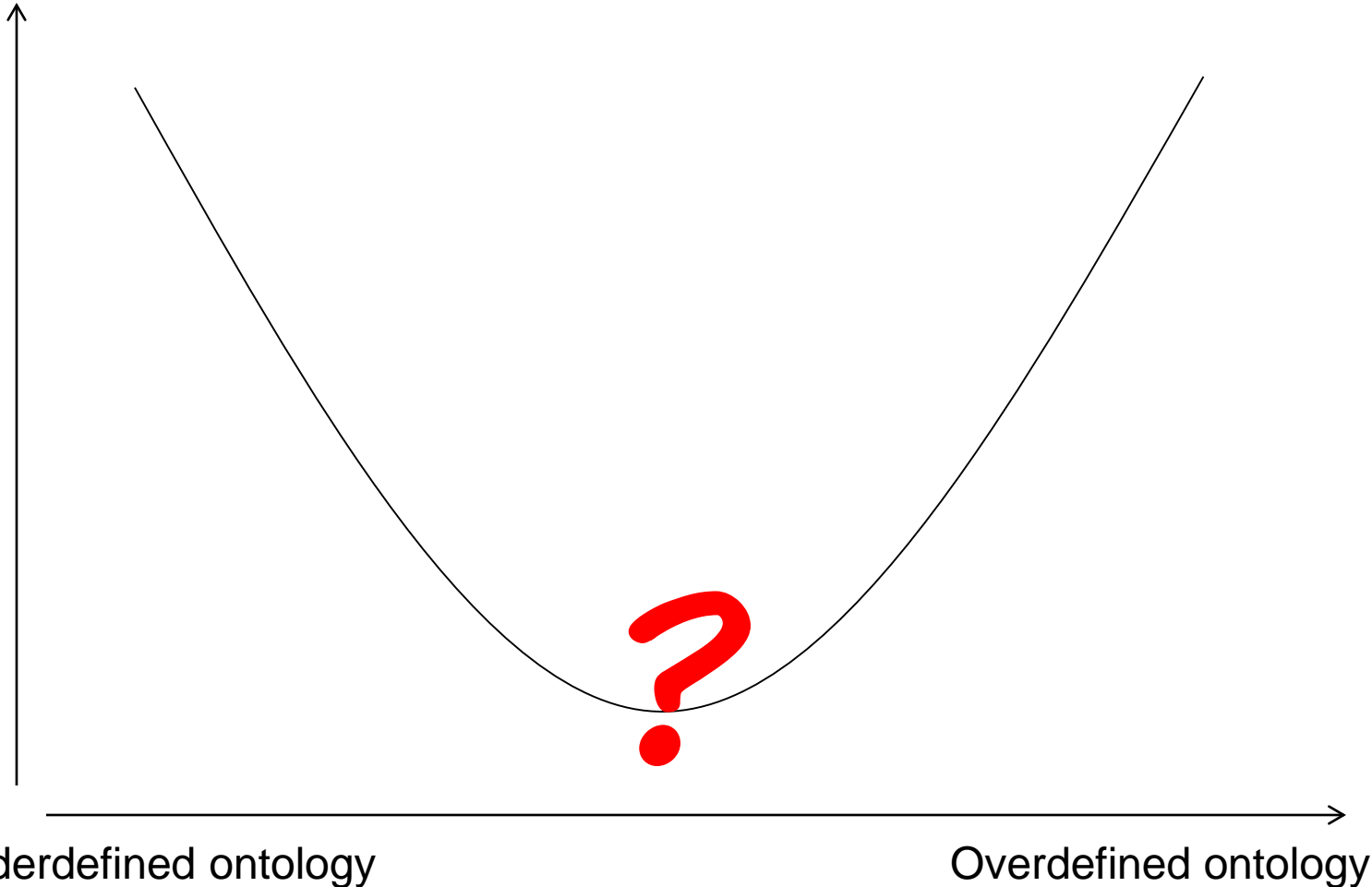
Ontologies **do not define** meaning.
They **constrain** meaning.

They do not attempt to address all ambiguities.
(Depicted a Euclidean projection of a hyperbolic triangle.)

Soft Spot Search

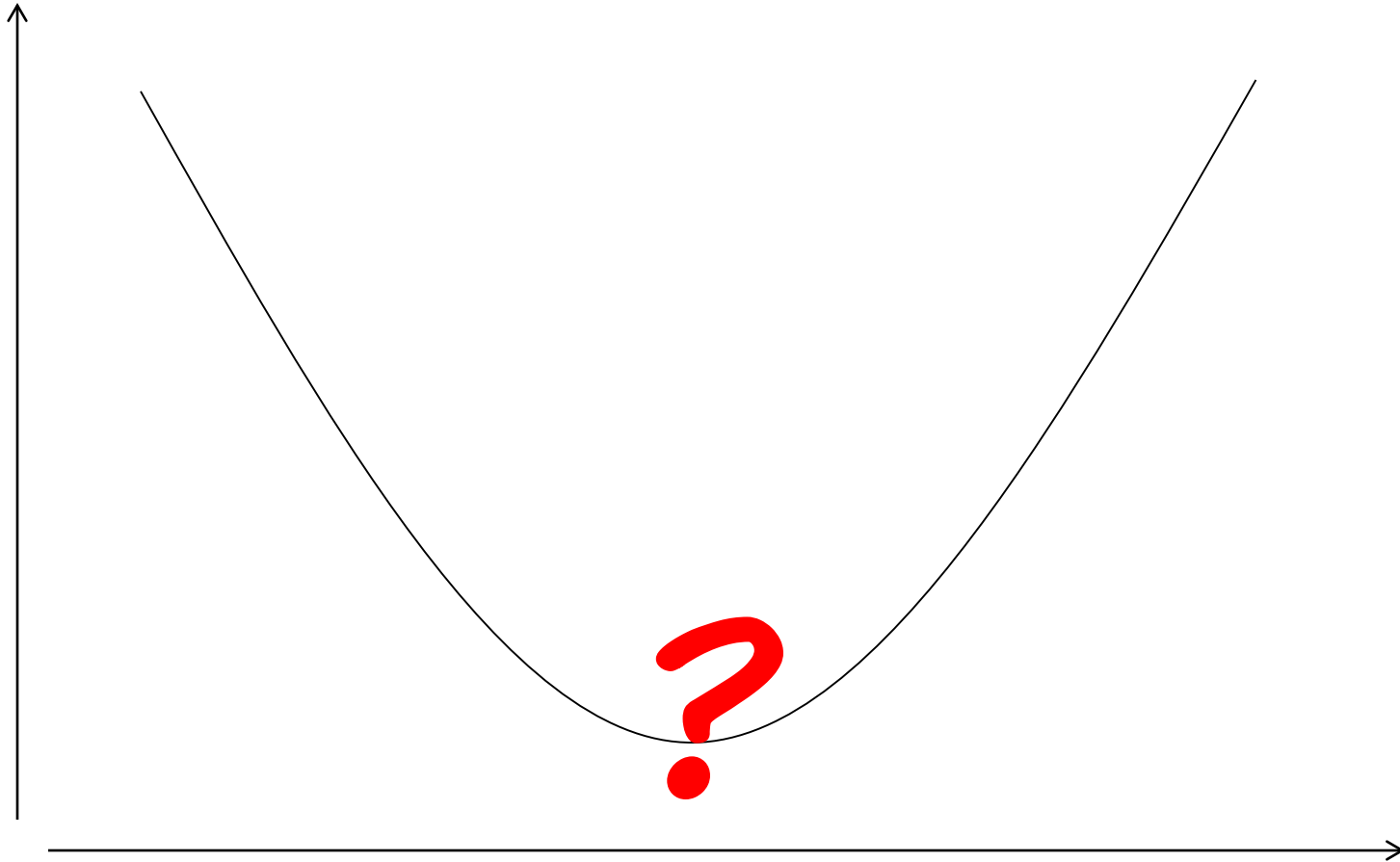


cost of data integration and reuse



Soft Spot Search

cost of data integration and reuse



Use of small, local ontologies

Use of large cross-domain ontologies

Ca. 2000 to 2004 hype:



“Ontologies are going to solve the world’s data integration and reuse problems.”

Of course they didn’t.

Falsification is a central part of scientific progress.

In Computer Science, the hype cycles are rather short.

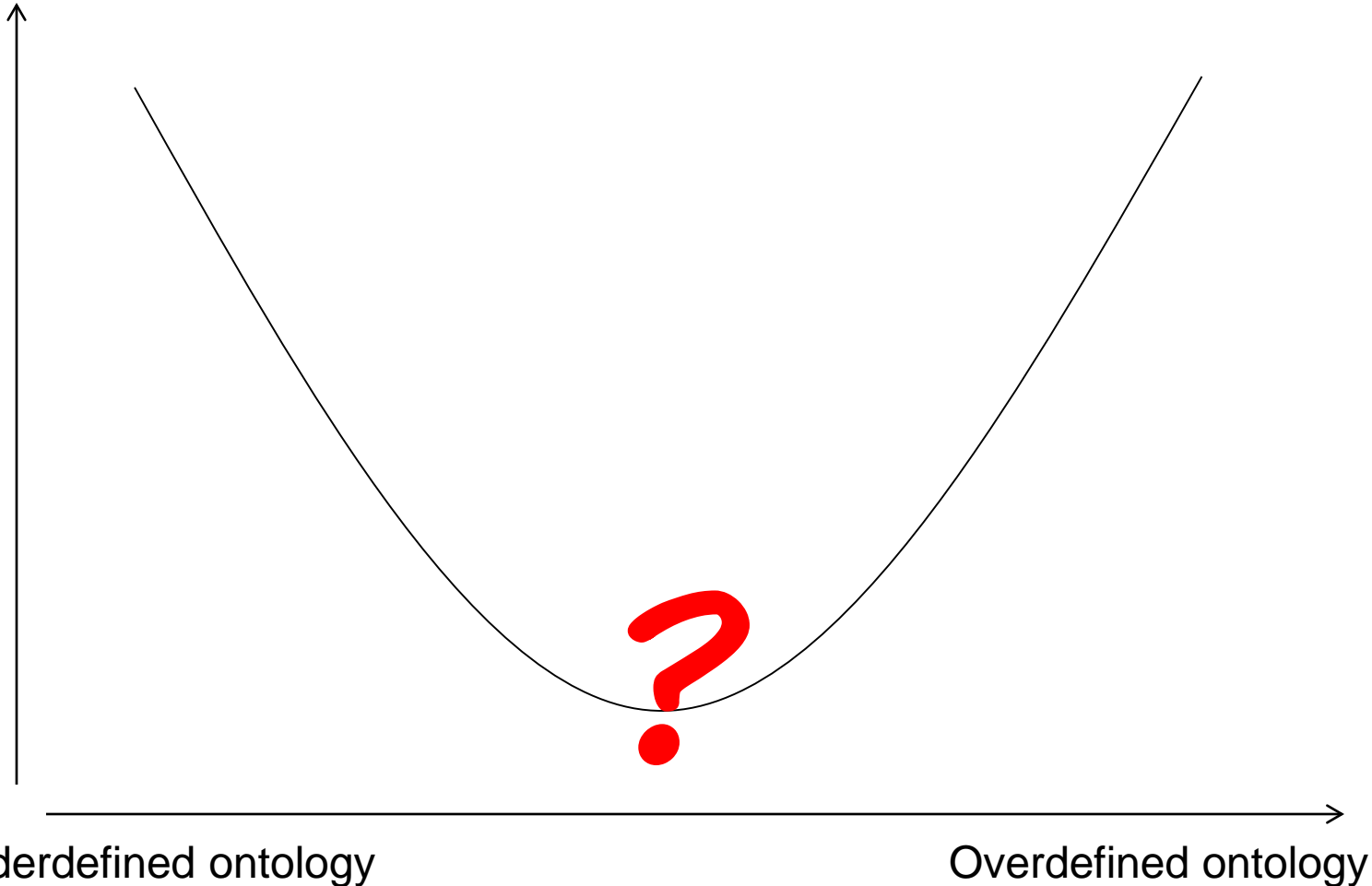
Ca. 2008 to 2013 hype:

“Linked Data is going to solve the world’s data integration and reuse problems.”

Soft Spot Search

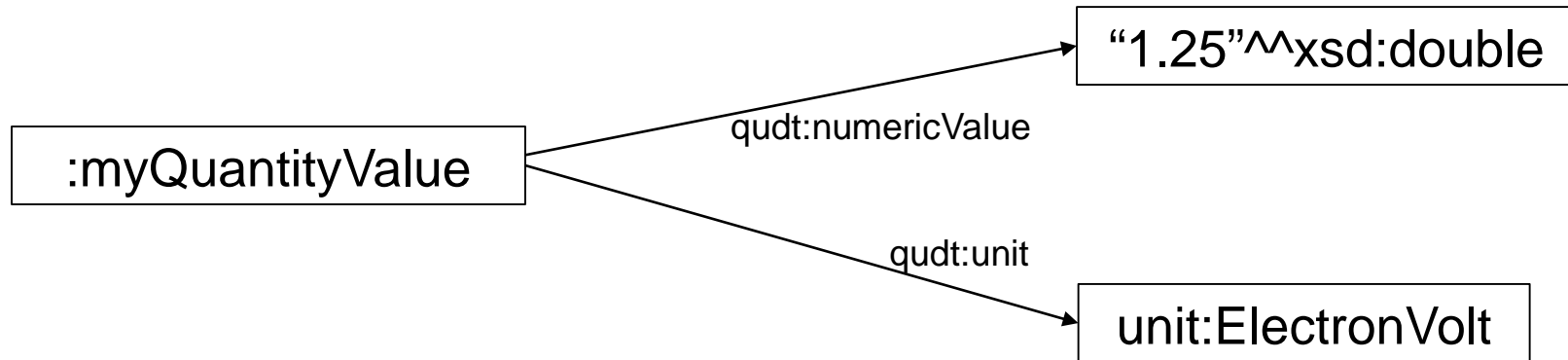


cost of data integration and reuse





Is data in graph structures, serialized as RDF (usually in XML).



(In RDF Turtle syntax:)

```
:myQuantityValue      qudt:numericValue      "1.25"^^xsd:double;
                        qudt:unit                          unit:ElectronVolt.
```



(In RDF Turtle syntax:)

```
:myQuantityValue      qudt:numericValue      "1.25"^^xsd:double;  
                        qudt:unit              unit:ElectronVolt.
```



Use of links / reuse of external vocabulary:

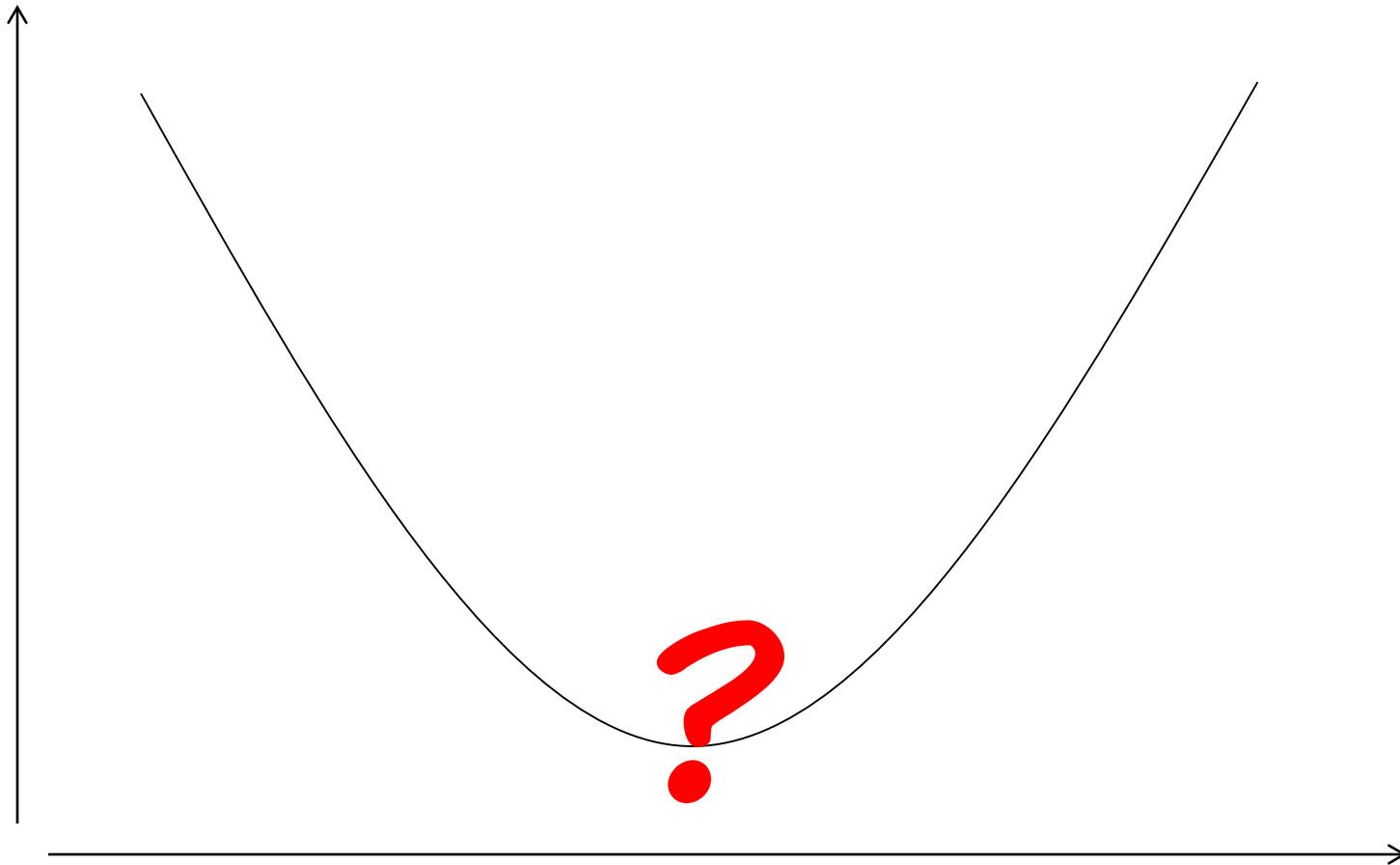
- Are you sure what their vocabulary really means?
- They may change their vocabulary.
- Ambiguity can lead to shifts in meaning and thus to inconsistencies.

- So need to be careful with reuse.

Soft Spot Search



cost of data integration and reuse



No vocabulary reuse

Over-reuse of vocabulary



Ontology defines the graph structure.
Constrains the meaning of data items.
Constrains ambiguities.

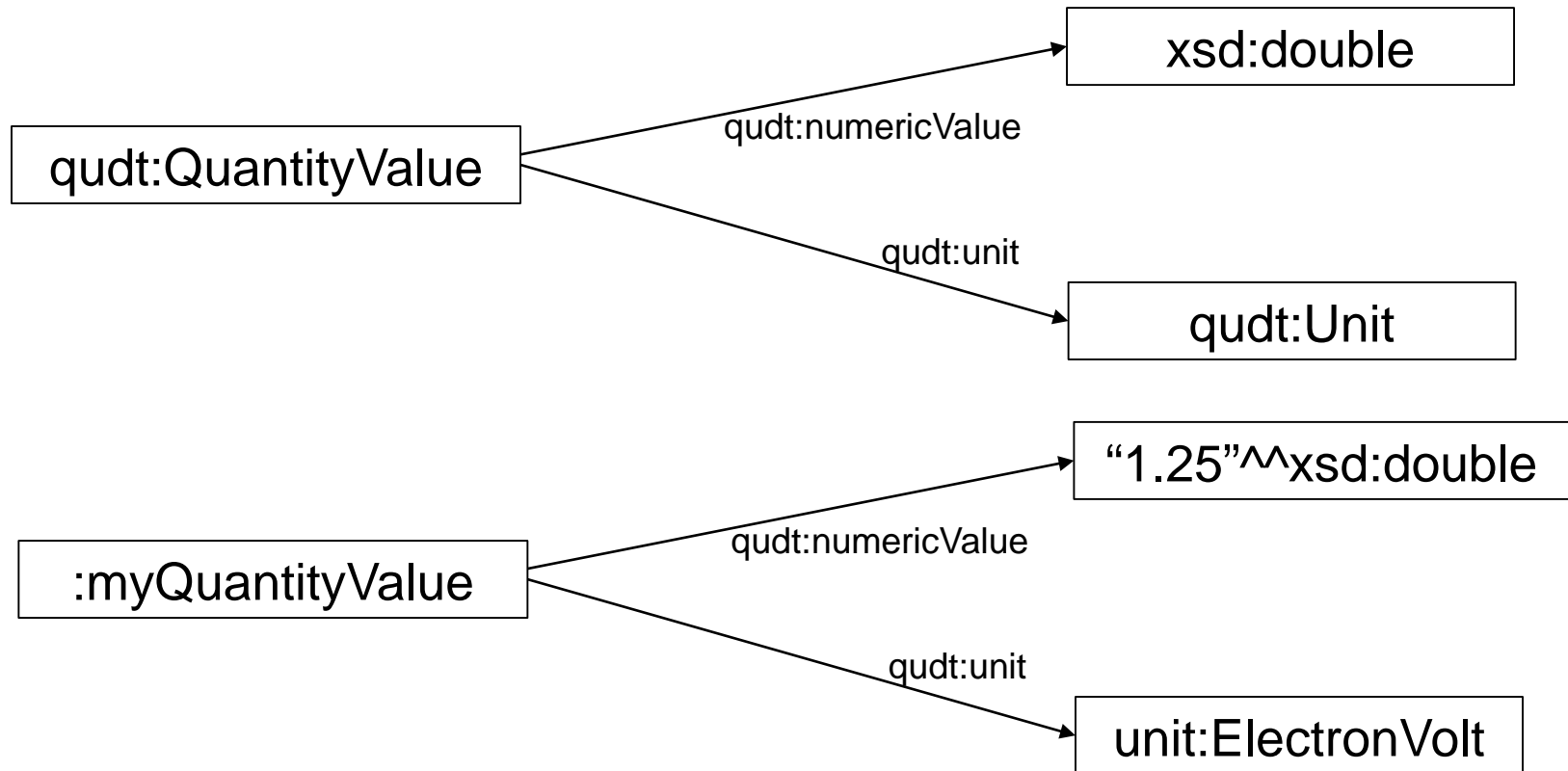
For some things (e.g., publications, persons), standardized vocabularies and identifiers are in addition extremely helpful (e.g., DOIs, ORCIDs).

For other things (e.g. forests, cities) the rigidity of standardized vocabularies is often counterproductive.



Ontology:

A **QuantityValue** has exactly one **NumericValue**
and exactly one **Unit**.



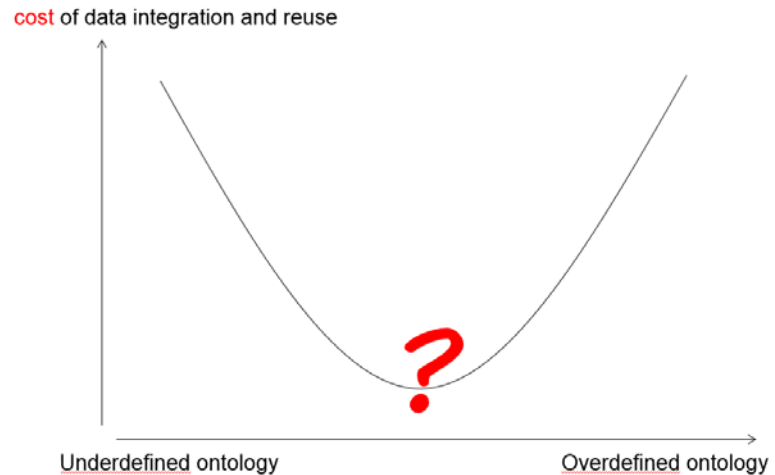


2000 to 2004: ontology hype
2008 to 2014: linked data hype

Ontologies are useful.
Linked Data is useful.

Dan Brickley at ISWC2015:

“Schema.org could be considered the biggest semantic web success story yet.”



Next: **The Linked Data Quality Frenzy?**



- Before publishing your linked data, be aware that you need a **good** ontology which informs your graph structure.
- The ontology should not (necessarily) closely model your data, but should faithfully describe the key notions in your data.
- Publish the ontology with your data.
- Be aware of the problems of over- and underspecification, of over- and underuse of other vocabularies and standards.

- Eva Blomqvist, Pascal Hitzler, Krzysztof Janowicz, Adila Krisnadhi, Thomas Narock, Monika Solanki, **Considerations regarding Ontology Design Patterns**. Semantic Web 7 (1), 2015, 1-7.
- 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.
- Pascal Hitzler, Krzysztof Janowicz, Adila Krisnadhi, **Ontology modeling with domain experts: The GeoVoCamp experience**. In: Claudia d'Amato, Freddy Lecue, Raghava Mutharaju, Tom Narock, Fabian Wirth (eds.), Diversity++ 2015, Proceedings of the 1st International Diversity++ Workshop co-located with the 14th International Semantic Web Conference (ISWC 2015), Bethlehem, Pennsylvania, USA, October 12, 2015. CEUR Workshop Proceedings Vol-1501, ceur-ws.org, pp. 31-36. Use. Semantic Web 5 (3), 2014, 173-176.





- Adila A. Krisnadhi, Yingjie Hu, Krzysztof Janowicz, Pascal Hitzler, Robert Arko, Suzanne Carbotte, Cynthia Chandler, Michelle Cheatham, Douglas Fils, Tim Finin, Peng Ji, Matthew Jones, Nazifa Karima, Audrey Mickle, Tom Narock, Margaret O'Brien, Lisa Raymond, Adam Shepherd, Mark Schildhauer, Peter Wiebe, **The GeoLink Modular Oceanography Ontology**. In: Marcelo Arenas, Óscar Corcho, Elena Simperl, Markus Strohmaier, Mathieu d'Aquin, Kavitha Srinivas, Paul T. Groth, Michel Dumontier, Jeff Heflin, Krishnaprasad Thirunarayan, Steffen Staab (eds.), The Semantic Web - ISWC 2015 - 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part II. Lecture Notes in Computer Science 9367, Springer, Heidelberg, 2015, 301-309.

- Víctor Rodríguez-Doncel, Adila A. Krisnadhi, Pascal Hitzler, Michelle Cheatham, Nazifa Karima, Reihaneh Amini, **Pattern-Based Linked Data Publication: The Linked Chess Dataset Case**. In: Olaf Hartig, Juan Sequeda, Aidan Hogan (eds.), Proceedings of the 6th International Workshop on Consuming Linked Data co-located with 14th International Semantic Web Conference (ISWC 2105), Bethlehem, Pennsylvania, US, October 12th, 2015. CEUR Workshop Proceedings 1426, CEUR-WS.org, 2015.