Ontology modeling with domain experts

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

Contents

Ontology modeling in practice

“We were developing an ontology on topic X to drive our application.”

Me: “That’s a tough topic to model, even if you have several domain experts in the room.”

“Well, we didn’t have any domain experts at hand, so we simply used common sense to make the ontology.”

[I’ve had this conversation in several variants.]

Ontology modeling in practice

“Semantic Web just doesn’t work.”

Me: “Why do you say that?”

“Well, several years ago, there was a big push in our community to develop Well-Known-Ontology X, but if you actually try to use X it’s just not really working.”

Me: “So how was X developed?”

“I understand they simply took a couple of undergrads and let them model away …”

What are you modeling for?

I used to hear (and say):

> **“Before you model an ontology, know what the ontology is going to be used for.”**

The problem with this, of course, is that then you’re going to end up with an ontology which can probably only be used for the purpose it was developed for.

Wasn’t it one of the key ideas of Semantic Web that we should be able to *reuse*?

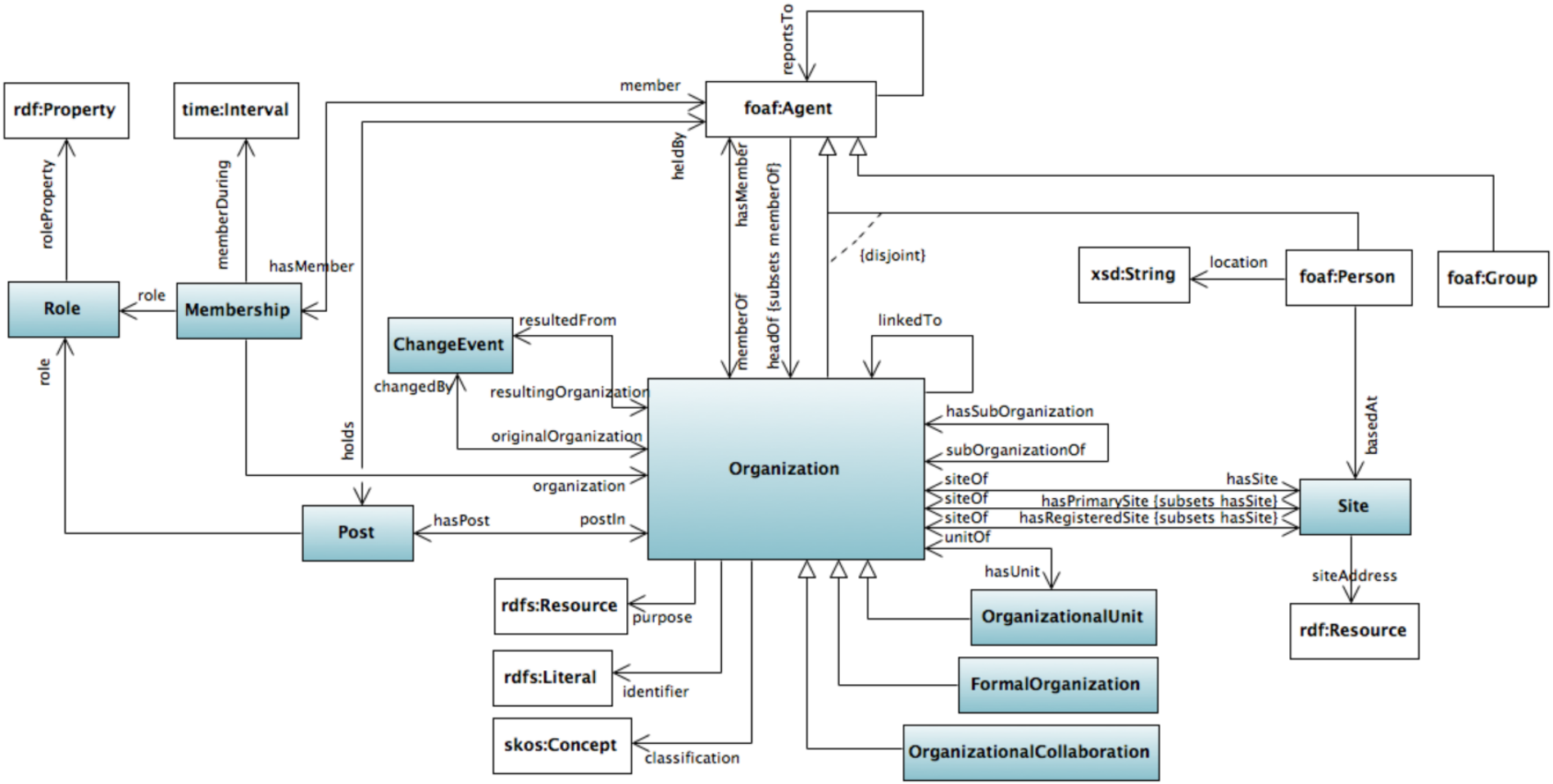
Application-driven modeling?

Just some example findings.

```
:AggregateModel
        owl:equivalentClass
                [ rdf:type owl:Restriction ;
                  owl:allValuesFrom :Aggregate ;
                  owl:onProperty :models
                ] .
```

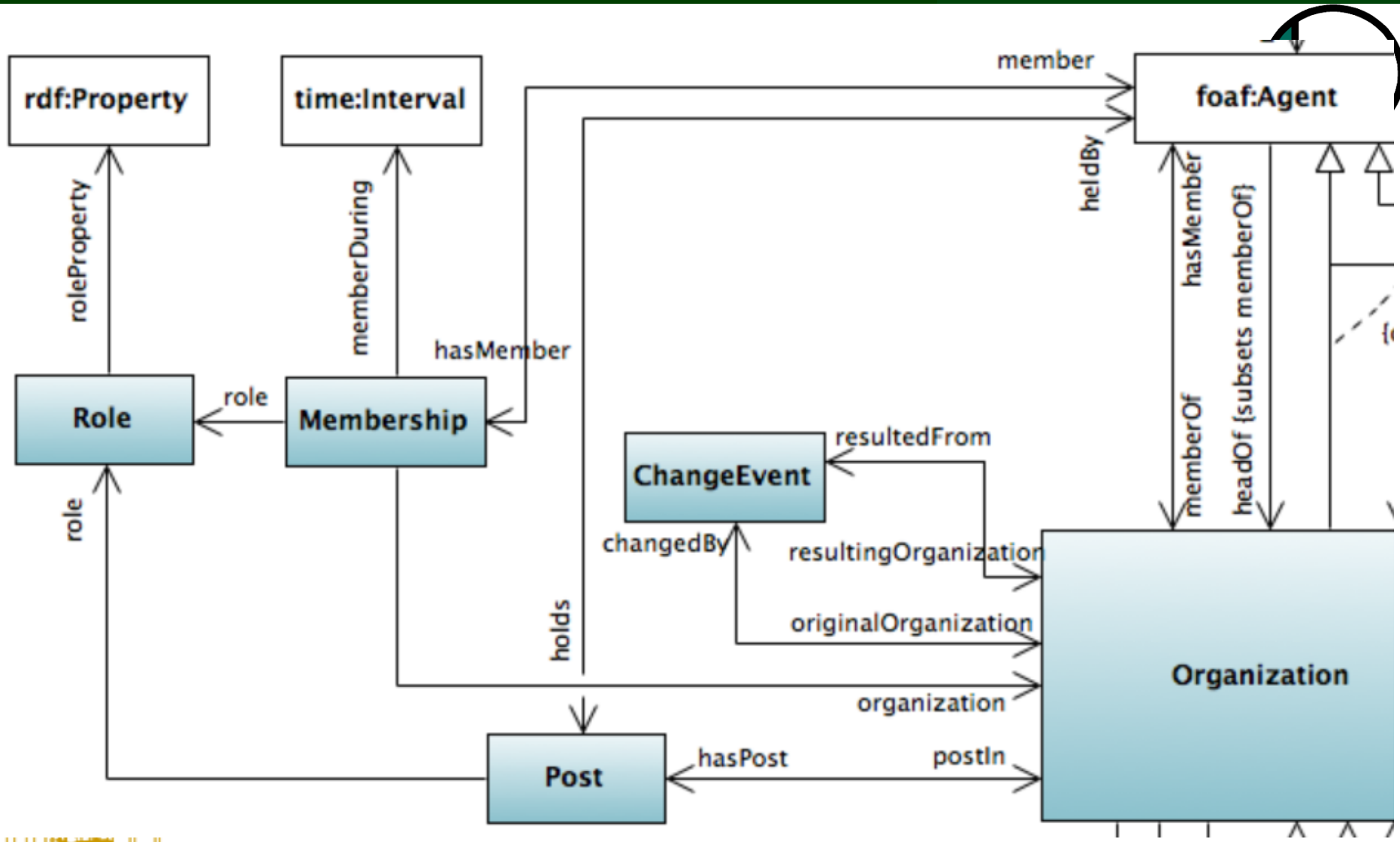
```
:ChildLabor     rdfs:subClassOf  :HumanRights
```

W3C Organization ontology

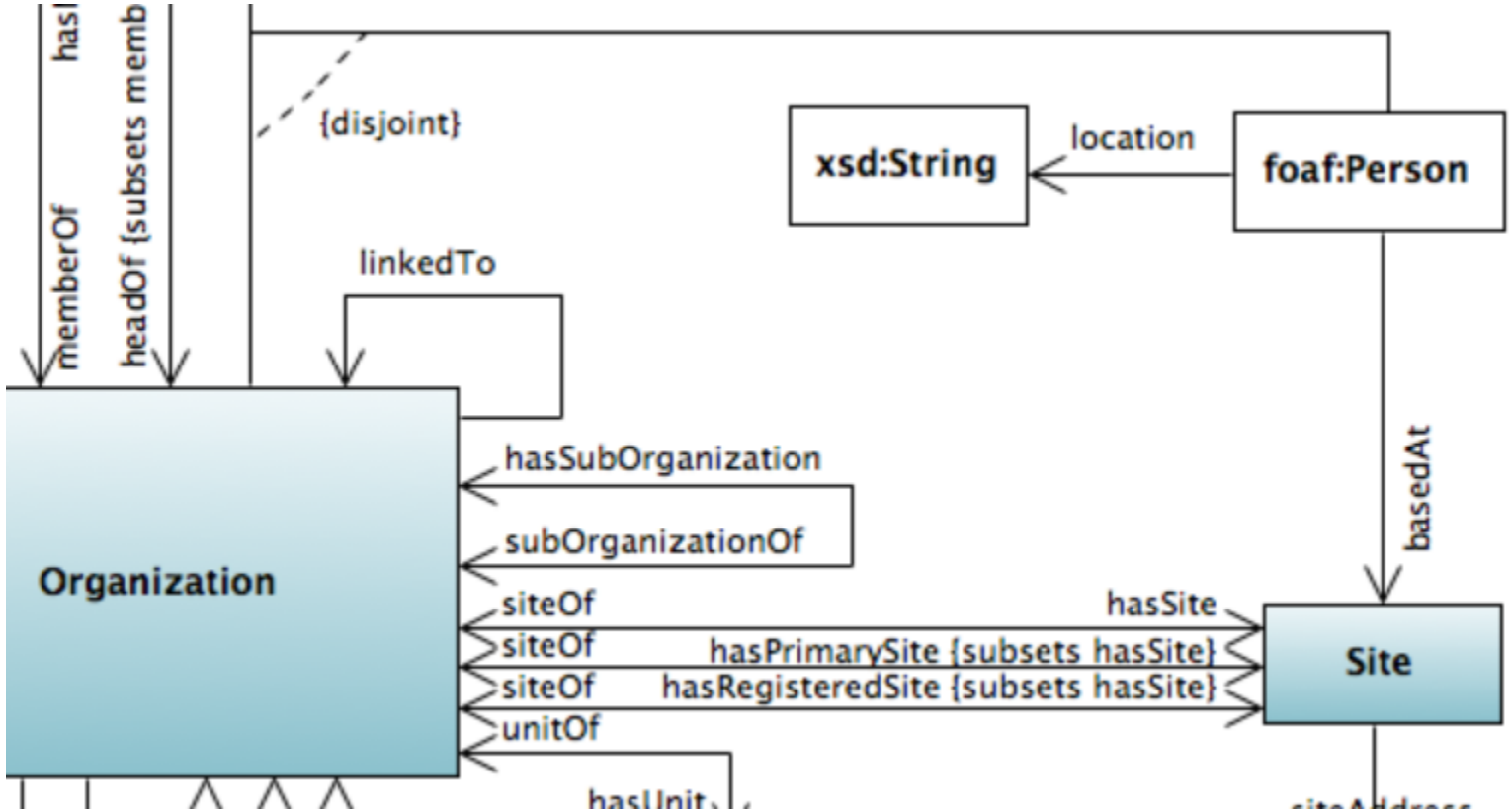


W3C Organization ontology

DaSe Lab



W3C Organization Ontology



Ontology Reuse

If …

(a) Ontologies are modeled with specific application purposes in mind and

(b) Ontologies commonly come with quite a few conceptual and logical issues

… is it then a surprise that ontologies are not being reused?

But: It's not necessarily a deficiency of the ontology-based approach *as such* that this happens.

Rather,

- **some of it is unavoidable using a traditional ontology modeling approach, and**
- **some of it is simply bad ontology engineering.**

Contents

1. Ontologies
2. Linked Data
3. Modeling

Linked Data Will Save Us?

Around 2006/2007 a countermovement started:

If ontology reuse doesn't work, then let's simply publish data in RDF.

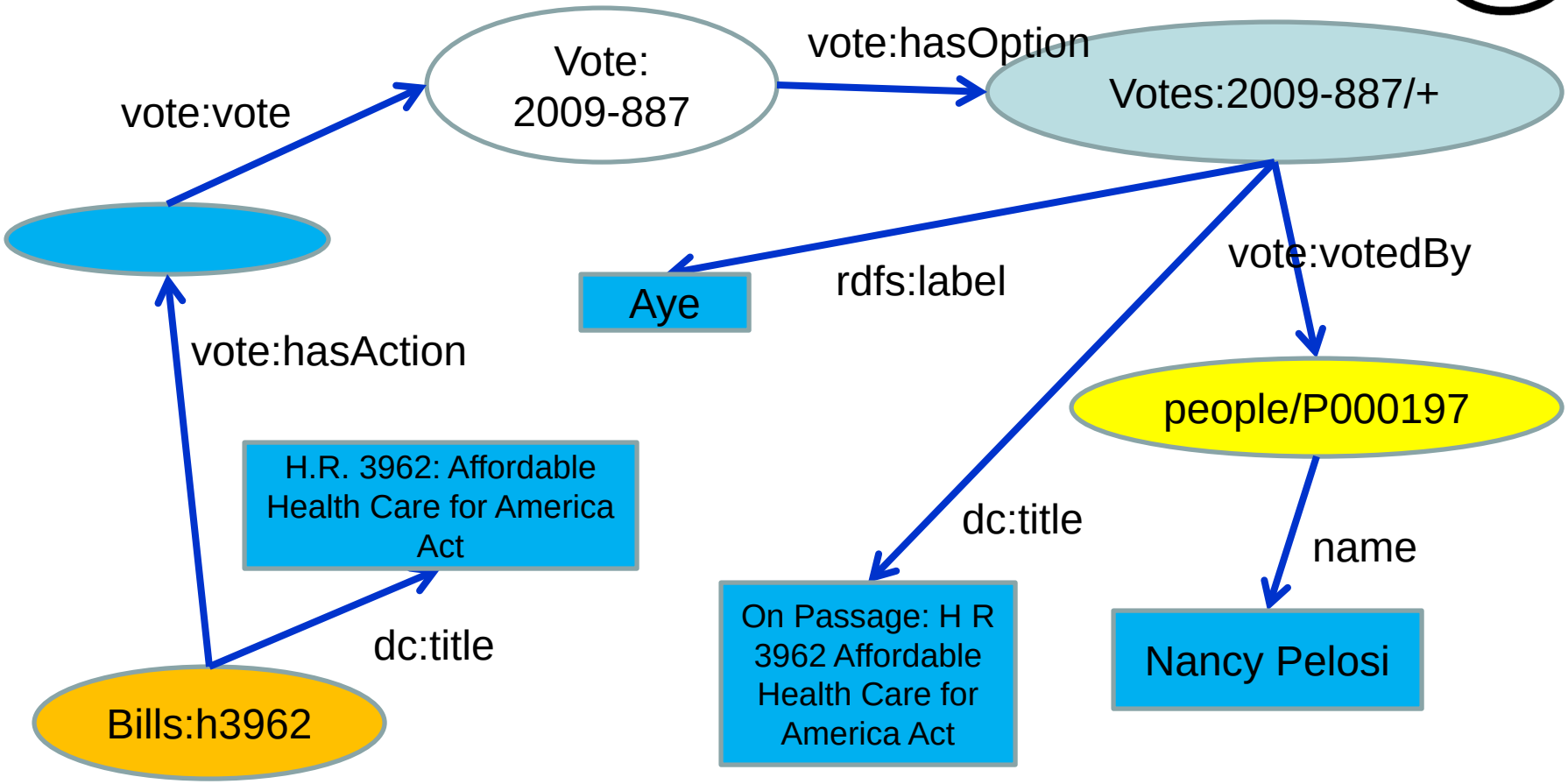
This gave rise to Linked Data.

So: We gave up on ontology reuse, and are instead banking on data reuse (without ontologies)?

Let's see …

Reusing Linked Data is tricky

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



Absence of schema?

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

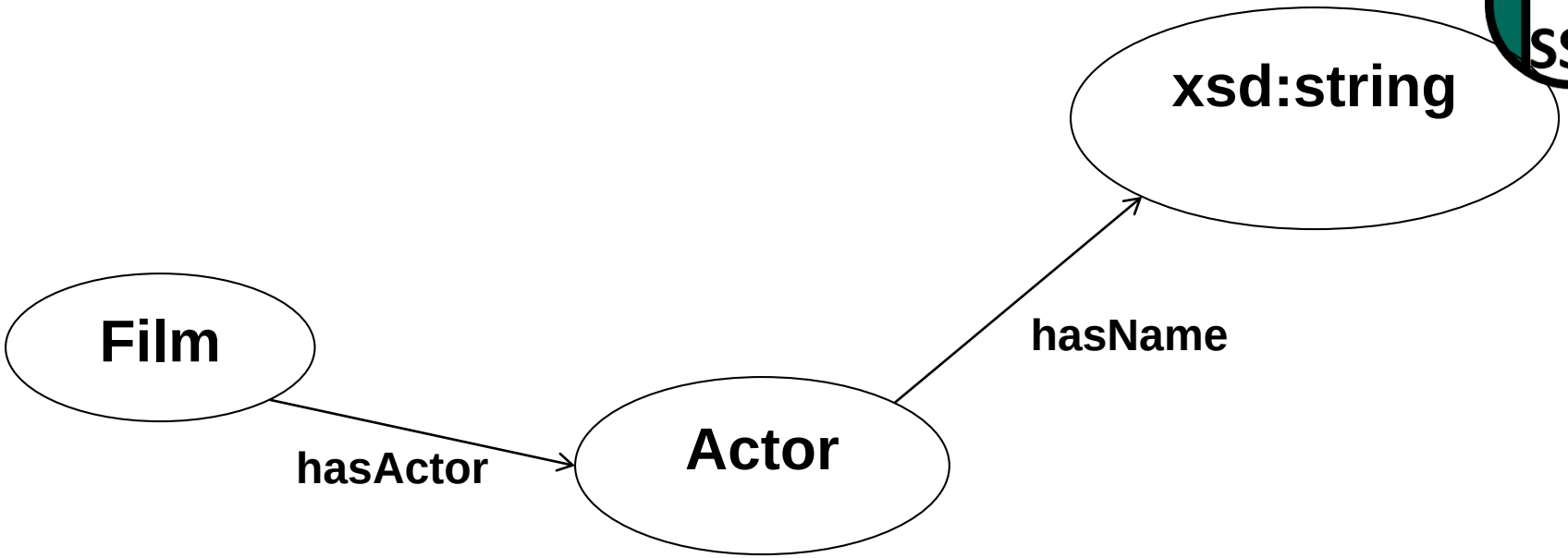
Location of Copernicus.

Location of Copernicus.

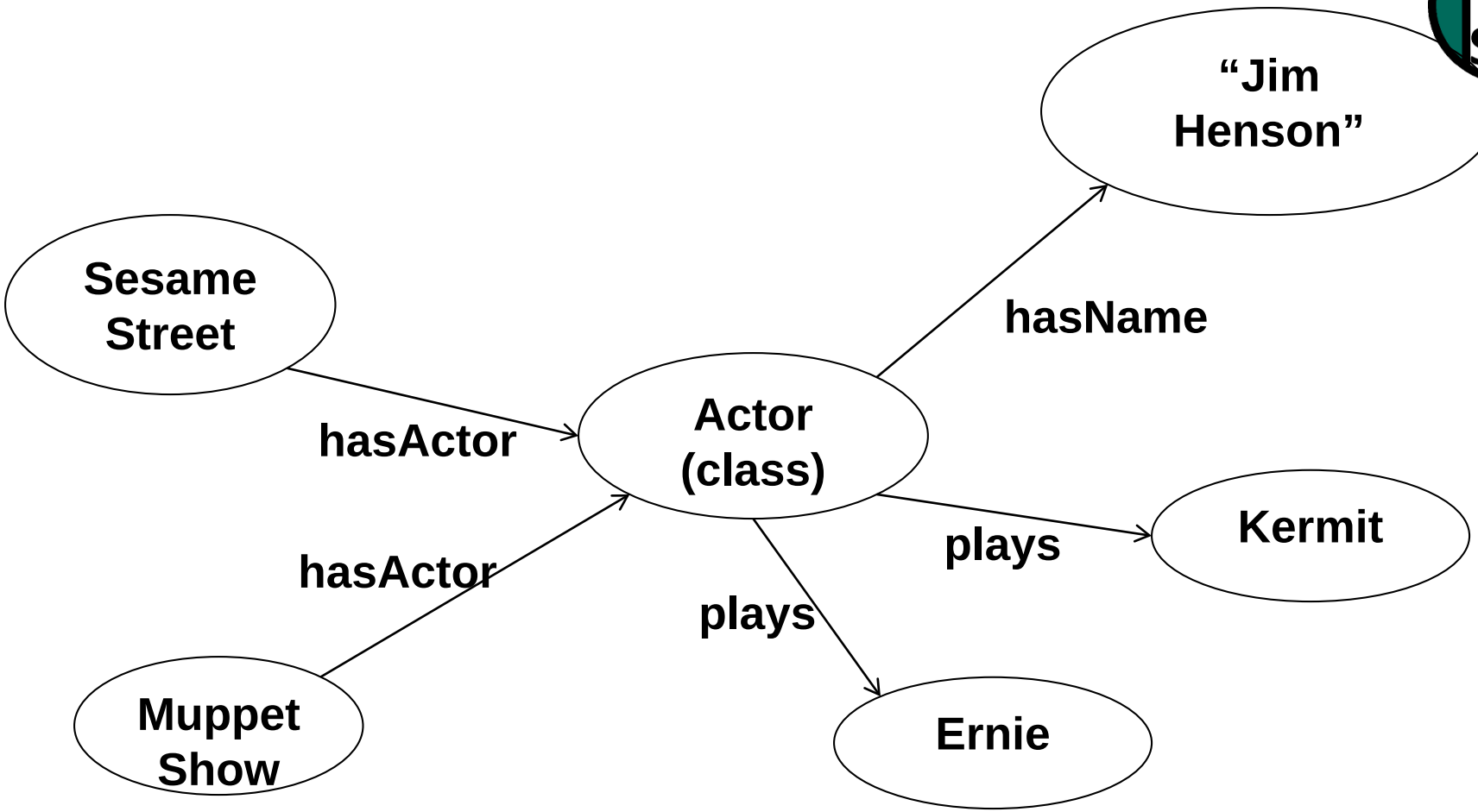
Image

Google Map

Example from Linked MDB



Problem!


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

Implicit Ontologies

When publishing Linked Data, there is always an underlying graph schema, which somebody has "designed."

In other words, there is always an underlying ontology

even if the provider hasn't bothered to write it up properly or share it.

The W3C RDF Shapes Working Group is probably out to make this explicit.

(I'm not sure why they don't call them "ontologies" though, but I can also live with "RDF Shapes" of course.)

No LOD without ontologies

You can't avoid the schema when dealing with Linked Data.

Which means you also can't avoid the ontology/schema modeling issues.

If your schema is not well-designed and well-documented, then it will not be easily reusable.

[Looking forward to seeing more about the RDF Shapes work.]

From ontologies to linked data

DaSe Lab

SSC

We recently realized we have a lot of chess players in the lab.

And that there's no linked dataset for chess games.

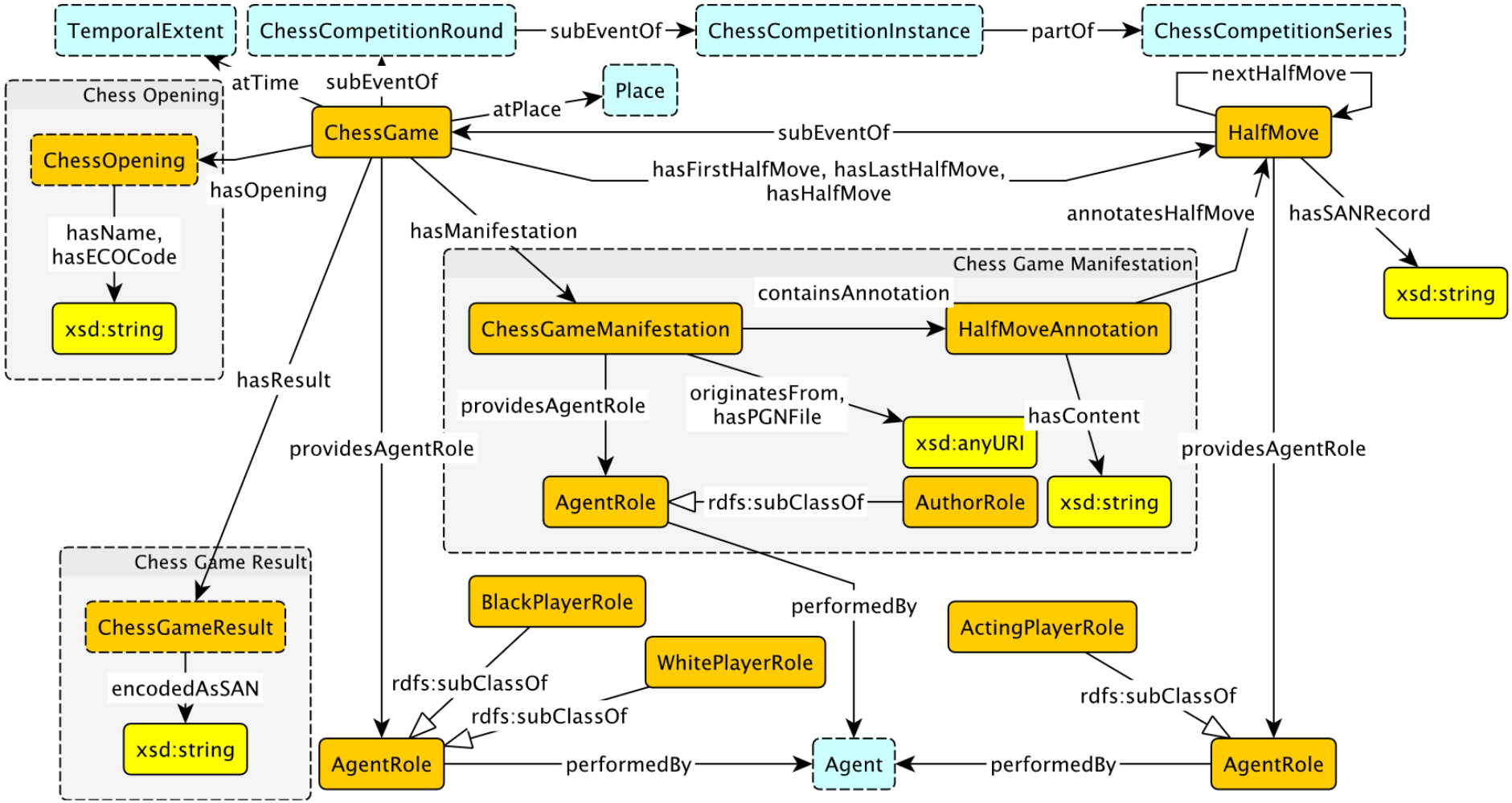
So we decided to change that.

[It's actually a COLD paper, presented this afternoon.]

There is already an established standard, the Portable Games Notation PGN (text-based, with some basic metadata), and lots of data available on the Web.

Following our own recommendations, we first made an ontology …

Chess ontology / ODP



Linked Chess Data

See http://salonica.dia.fi.upm.es:8080/rdfchess/

(should be available later on http://chessdata.org)

Contents

1. Ontologies
2. Linked Data
3. Modeling

How to model with experts

DaSe Lab

SSC

Good modeling practices are important

- **To make it easier for ontology engineers to understand the ontology, and thus to make it easier to reuse it or to do maintenance.**
- **To enable sustainability of the model (future changes or developments).**
- **To enable widening coverage, or finer granularity, down the road, without breaking things.**

- **To more easily discuss with experts during the modeling process.**
- **To more easily have the ontology assessed by experts.**
- **To make sure that the ontology captures not only a single expert perspective.**

GeoVoCamps modeling approach

- Collaborative modeling, group ideally has
  - More than one domain experts.
  - People familiar with the base data.
  - People understanding possible target use cases.
  - An ontology engineer familiar with the modeling approach.
  - Somebody who understands formal semantics of OWL.

- Domain experts are queried as to the main notions for the application domain.
  - E.g. for chess, these would include
    - Chess game; move; opening; tournament; players; commentary

GeoVoCamps modeling approach

- From available data and from application use cases, devise competency questions, i.e. questions which should be convertible into queries, which in turn should be answerable using the data.

  Retrieve all games where Fischer lost in the poisoned pawn variation of the Sicilian.

  Retrieve all games where Fischer opened 1. Nf3.

- Then prioritize which notions to model first. In the chess case, e.g.

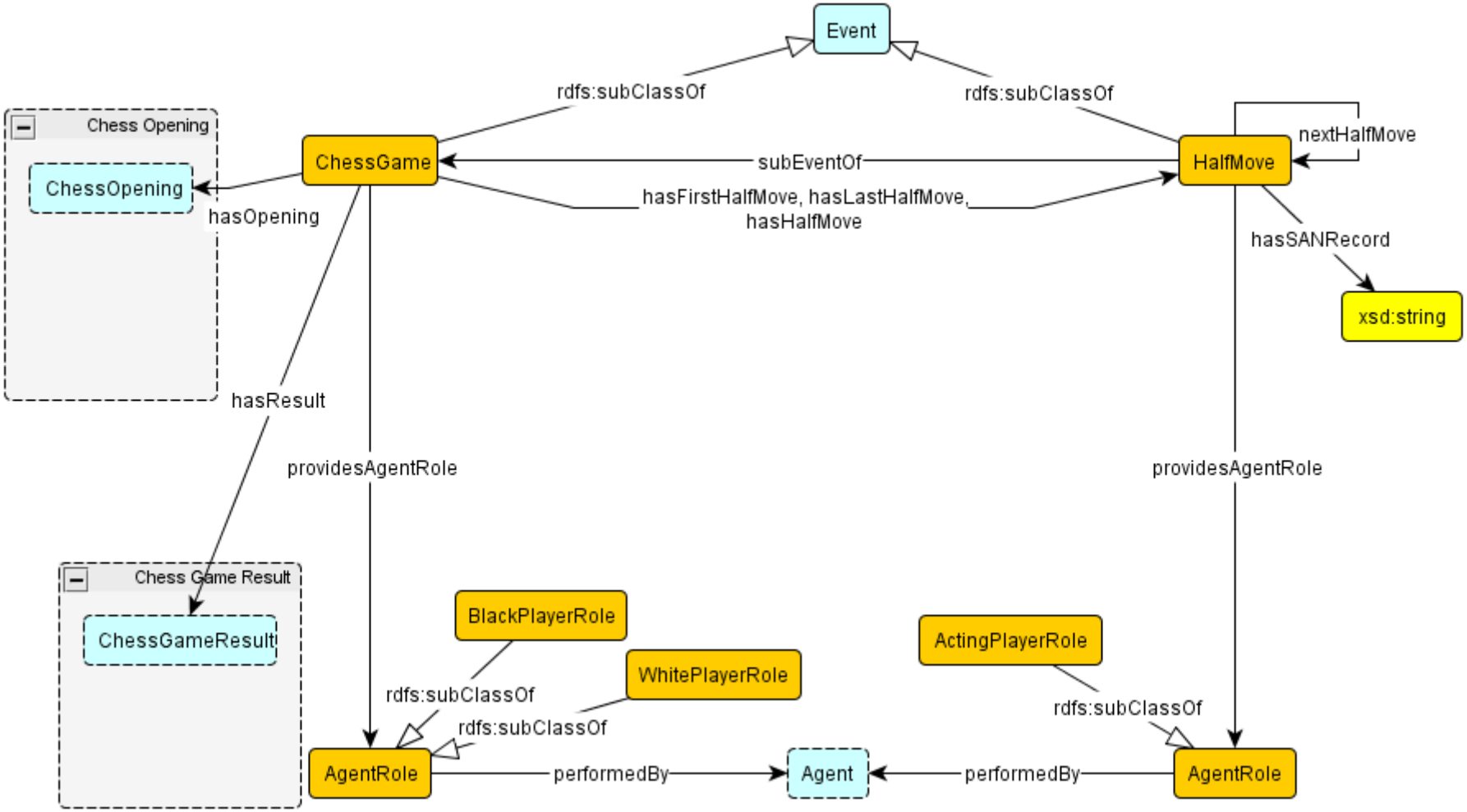
  chess game
  move/half-move
  players
  opening
  commentary
  tournaments

GeoVoCamps modeling approach

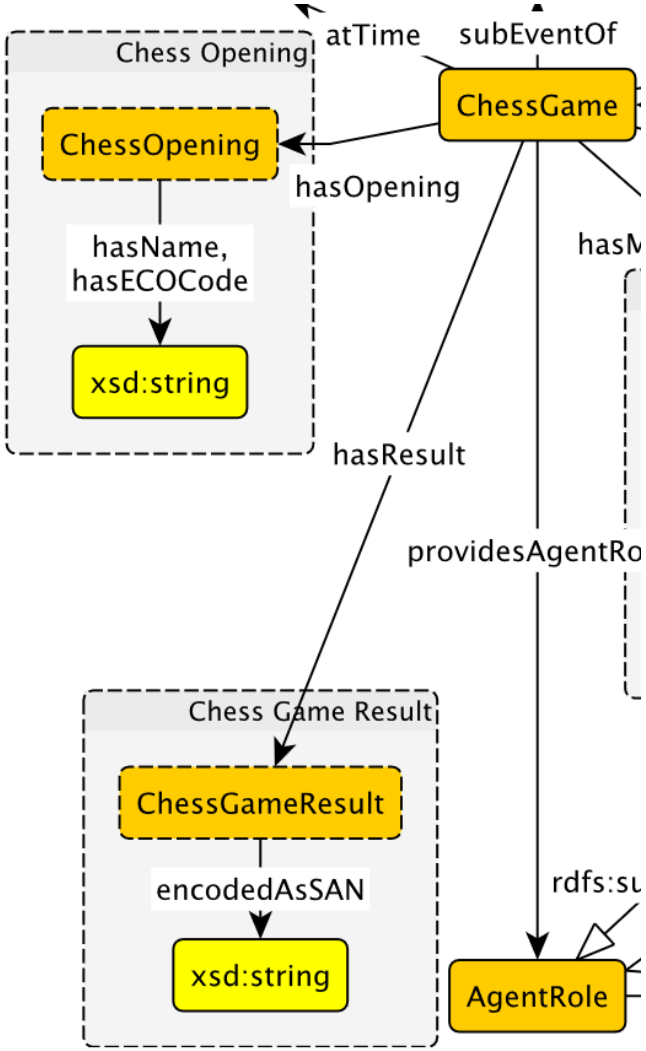
- **Understand the nature of the things you are modeling.**

Chess game	**...**	**An Event**
Half-move	**…**	**A Subevent of a chess game**
Player	**…**	**The Role of an Agent**
Opening	**…**	**this is probably complex**
commentary	**…**	**this is again more complex**
tournaments	**…**	**Events**

Chess game / player



Opening and games result



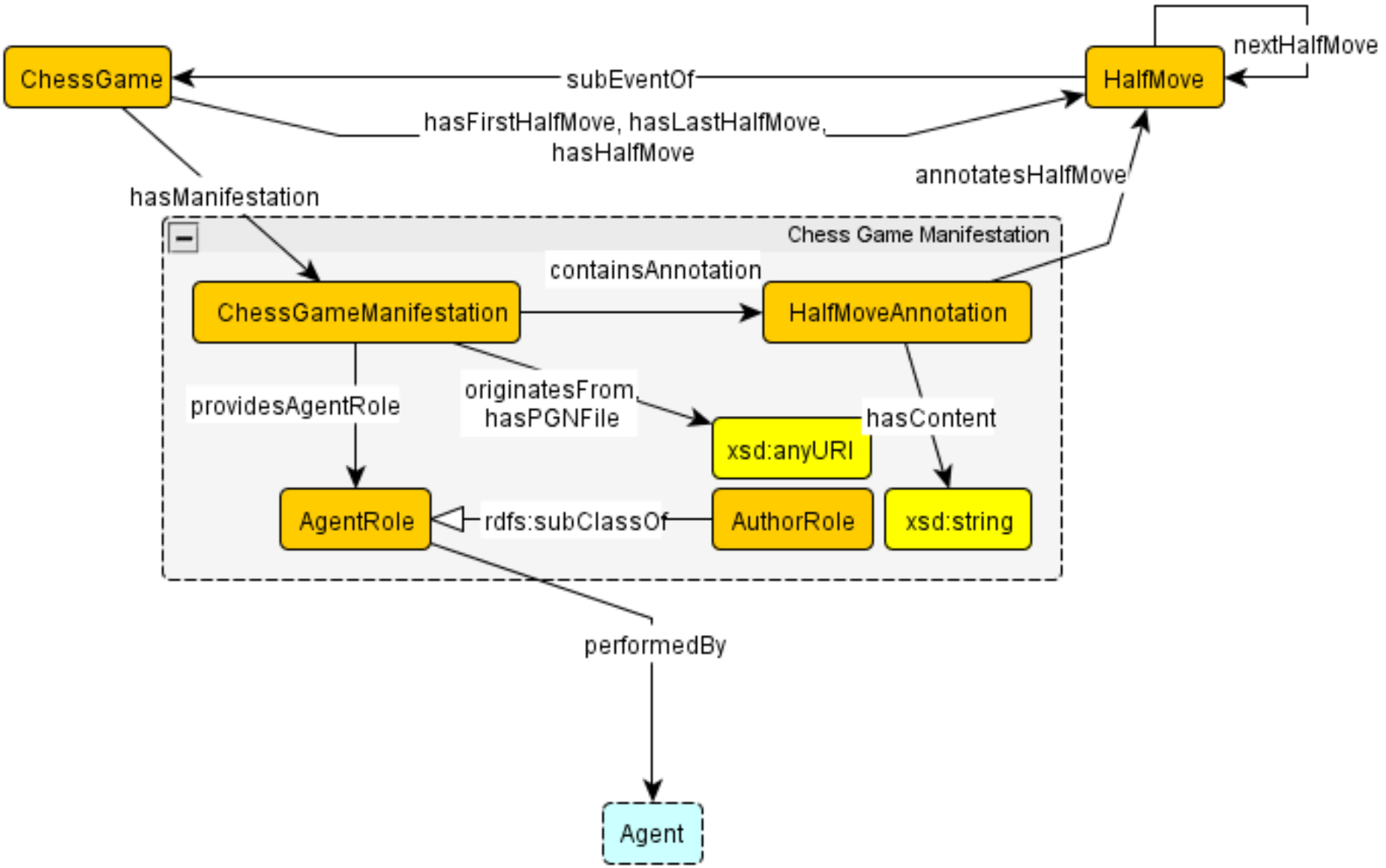
We call these “stubs”.

I.e. we’re aware that more fine-grained modeling will be needed for some use cases.

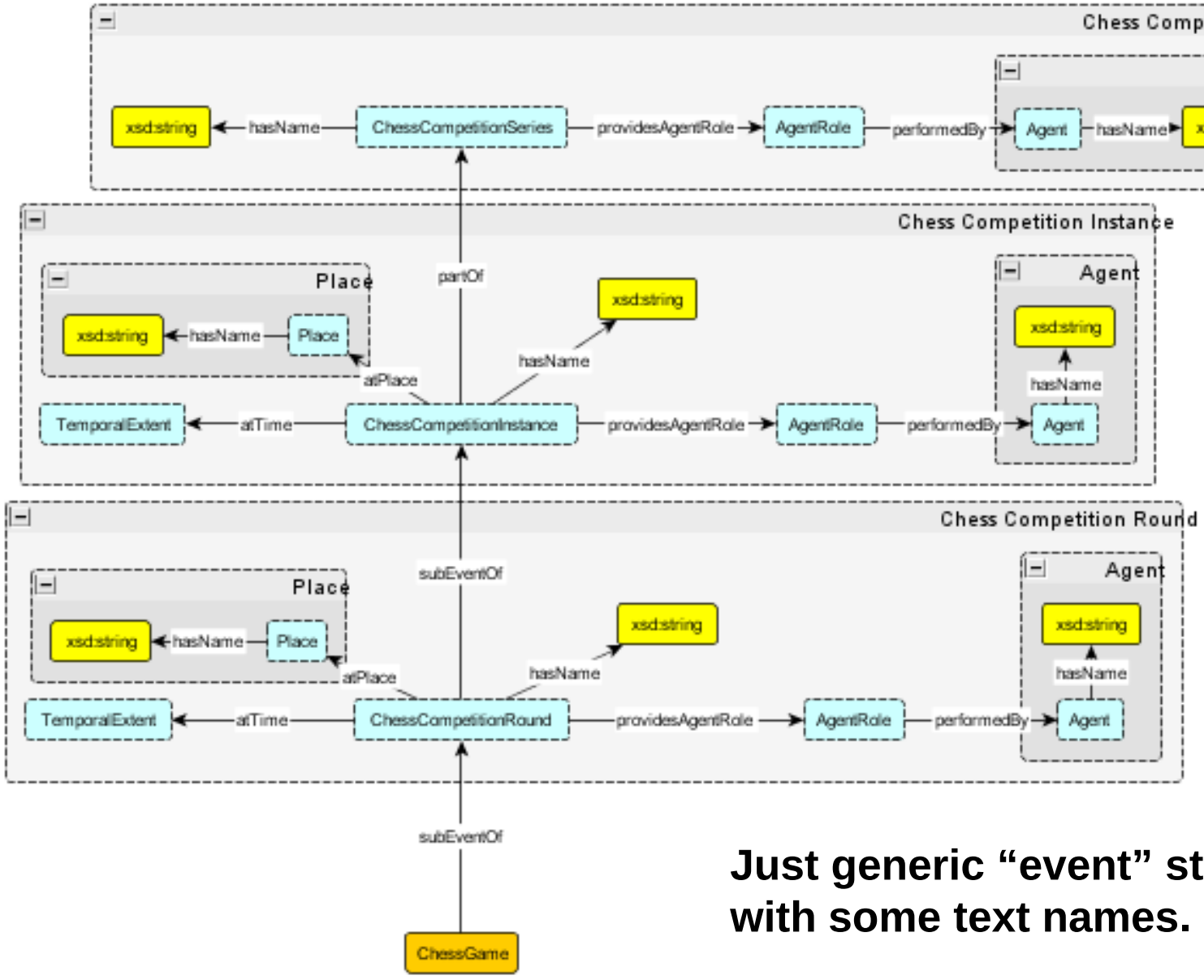
But currently there’s no reason to do it (not in use case, no data), so we only provide “hooks” for future development of the ontology.

Commentary and PGN file

DaSe Lab



Tournament

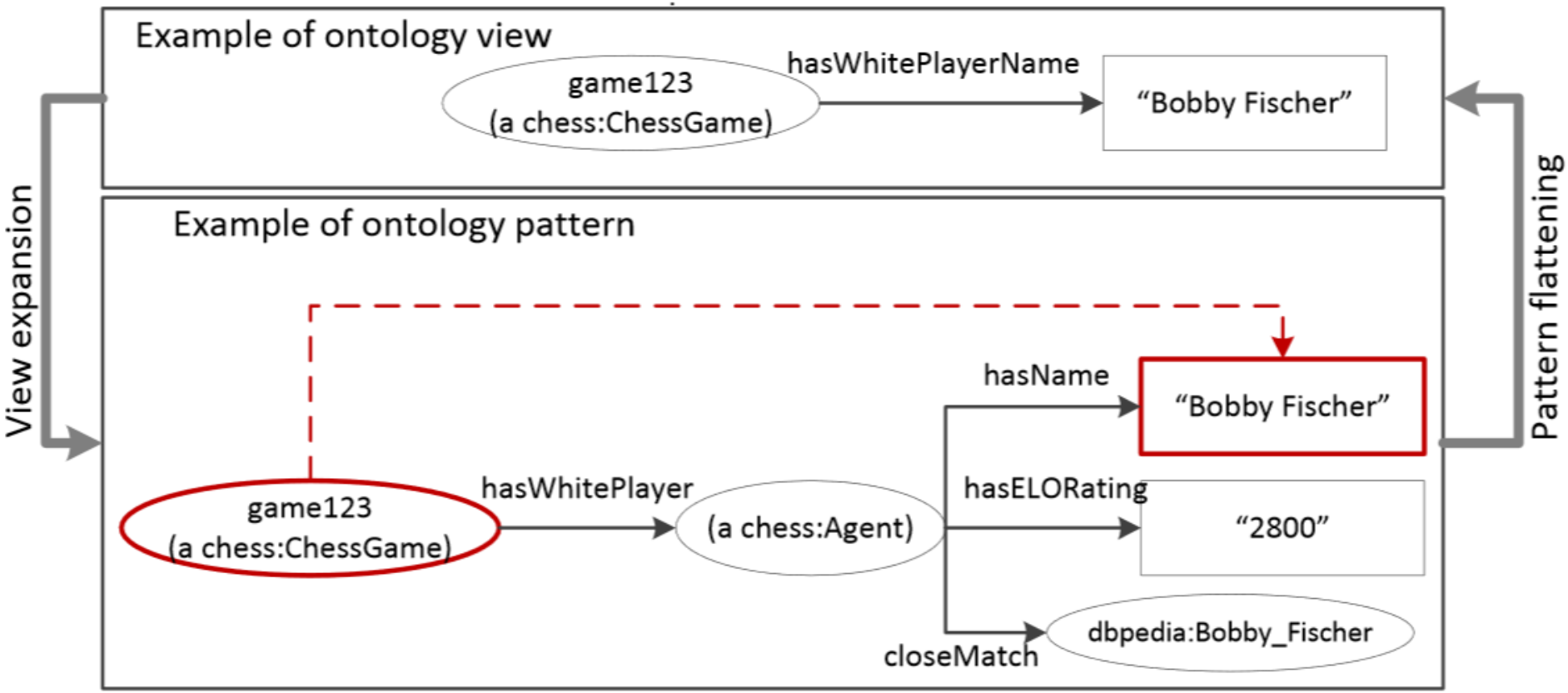


Just generic "event" stubs with some text names.

Adequacy check

- **Triplify sample data using the ontology.
  Does it work?**
- **Check if competency questions can be answered.**
- **Add axioms as appropriate (the graph is only for intuition, the OWL axioms are the actual ontology).**
- **(there are more post-hoc details to be taken care of, but let's leave it at that)**

Shortcuts (views)



Modular modeling

Note the modular modeling.
We find that it helps tremendously to

- **Focus on a single notion at a time.**
- **Discuss with domain experts on their grounds without the need to get into technical details.**
- **Relate to existing ontology design patterns, which helps with reuse and with quality modeling.**

Take Home Message

- **Semantic Web looses much of its added value unless we make it much easier to reuse data and metadata.**
- **Ontologies cannot be avoided: There is always a conceptual model, even if it's not explicated.**
- **Modular and thorough modeling makes reuse of data and ontologies easier.**

References

- Pascal Hitzler, Krzysztof Janowicz, Adila Krisnadhi, Ontology modeling with domain experts: The GeoVoCamp experience. In: Proceedings Diversity++ workshop at ISWC 2015. To appear.
- Adila A. Krisnadhi, Víctor Rodríguez Doncel, Pascal Hitzler, Michelle Cheatham, Nazifa Karima, Reihaneh Amini, Ashley Coleman, An Ontology Design Pattern for Chess Games. In: Proceedings WOP 2015. To appear.
- 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: Proceedings COLD 2015. To appear.

References

- Krzysztof Janowicz, Frank van Harmelen, James A. Hendler, Pascal Hitzler, Why the Data Train Needs Semantic Rails. AI Magazine 26 (1), 2015, 5-14.
- 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.
- Adila A. Krisnadhi, Pascal Hitzler, Krzysztof Janowicz, On capabilities and limitations of OWL regarding typecasting and ontology design pattern views. In: Proceedings OWLED 2015. To appear.

References

- Adila A. Krisnadhi, Robert Arko, Suzanne Carbotte, Cynthia Chandler, Michelle Cheatham, Timothy Finin, Pascal Hitzler, Krzysztof Janowicz, Thomas Narock, Lisa Raymond, Adam Shepherd, Peter Wiebe, Ontology Pattern Modeling for Cross-Repository Data Integration in the Ocean Sciences: The Oceanographic Cruise Example. In: Tom Narock and Peter Fox (eds.), The Semantic Web in Earth and Space Science: Current Status and Future Directions. Studies on the Semantic Web, IOS Press, Amsterdam, 2015, pp. 256-284.
- 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, Kerstin Lehnert, Audrey Mickle, Tom Narock, Margaret O'Brien, Lisa Raymond, Adam Shepherd, Mark Schildhauer, Peter Wiebe, The GeoLink Framework for Pattern-based Linked Data Integration. In: ISWC2015 poster. To appear.

References

- 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: Proceedings ISWC2015. To appear.
- Adila A. Krisnadhi, Robert Arko, Cynthia Chandler, Michelle Cheatham, Pascal Hitzler, Yingjie Hu, Krzysztof Janowicz, Peng Ji, Nazifa Karima, Adam Shepherd, Peter Wiebe, R2R+BCO-DMO - Linked Oceanographic Datasets. In: Proceedings Diversity++ workshop at ISWC 2015.