Towards Modular Ontology Modeling with Ontology Design Patterns

Pascal Hitzler

Data Semantics Laboratory (DaSe Lab)
Data Science and Security Cluster (DSSC)
Wright State University
http://www.pascal-hitzler.de

These two appear to be rather different:

1. **Ontologies used as knowledge bases for (deductive) expert systems.**
   Usually highly axiomatized.
   Constructed primarily with automated reasoning in mind.

2. **Ontologies constructed for data sharing, integration, and reuse.**
   Usually not constructed with focus on automated reasoning.

We will here focus on the latter, as it is core to Semantic Web.

After 17+ years, do we actually have scientific evidence that ontologies improve data sharing, integration and reuse?

Hardly.

There is plenty of anecdotal and conceptual evidence, and some hard evidence hidden behind confidentiality walls.

But there are hardly any hard scientific investigations into this.

(and yes, it is tricky)

sharing
integration
reuse

Was there ever hard evidence that a new programming paradigm is superior to a previous one, before it became mainstream?

Probably it was just that sufficient people (and companies) believed that it is, took it up, and it proved itself (or didn't).

So why didn't the same happen to ontologies?

Perhaps because the hype happened when we still didn't really know how to make or use them?

sharing
Integration
reuse

Ontologies for data sharing

We had/have the big linked data hype which started ca. 10 years ago.

But without much attention to schema/ontologies.

Now they are called knowledge graphs.

Data sharing seems easy. And almost everybody seems to do it.

But is it really easy? What are success measures?

sharing
integration
reuse

Ontologies for data reuse

Where are the reports on linked data reuse?

Reuse doesn't seem to happen large-scale.

If you've tried to do it, you know why. It is extremely hard to understand a linked data graph if you have nothing much but the graph to go on:

You simply don't understand the structure.

You have a hard time grasping which things belong together.

There is a lot of ambiguity in vocabulary names, even if documented.

Claim: A well-constructed and documented schema would help!

But what does that mean?

sharing
integration
reuse

Ontologies for data integration

Data integration often necessary before data reuse.

Almost impossible without a good schema.

And even having schemas means it's very tedious.

[Cheatham & Hitzler, ISWC 2013] shows that state-of-the-art ontology alignment is hardly better than string matching.

sharing
integration
reuse

Modularity

We need to advance ontology engineering to simplify data **sharing**, **integration** and **reuse**.

Claim: This can be achieved by switching to modular ontologies based on ontology design patterns.

In the following, I will:

1. Explain what we mean with this.

2. Argue why it helps with data **sharing**, **integration**, and **reuse**.

Worked Example: Chess

- **Establish a searchable repository for chess data.**

- **Starting point are PGN files.**
- **Should be extendable with other information from**
  - **Chess websites**
  - **Wikipedia**
  - **Geographic data**
  - **News**
  - **Etc.**

- **Use an ontology for information integration.**

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.**

1. Who played against Kasparov in the round 1994 Lineares tournament? Did (s)he play as a white or black player?
2. What is the first move taken by the black player in the Sicilian Defense opening?
3. Find all games in which Bobby Fischer, playing black, lost in the poisoned pawn variation of the Sicilian Defence opening.
4. Are there any recorded games using the Grünfeld Defence from before the 20th century?
5. What did Kasparov say about his opponent's first two moves in his commentary about his game against Topalov in the 1999 Tournament in Wijk ann Zee?
6. Who was the first non-Russian world champion after Fischer?
7. Did Bobby Fischer ever play against a grandmaster in Germany?
8. List all world championship games won by forfeit.

GeoVoCamps modeling approach

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

  chess game
  move/half-move
  players
  opening
  tournaments
  commentary

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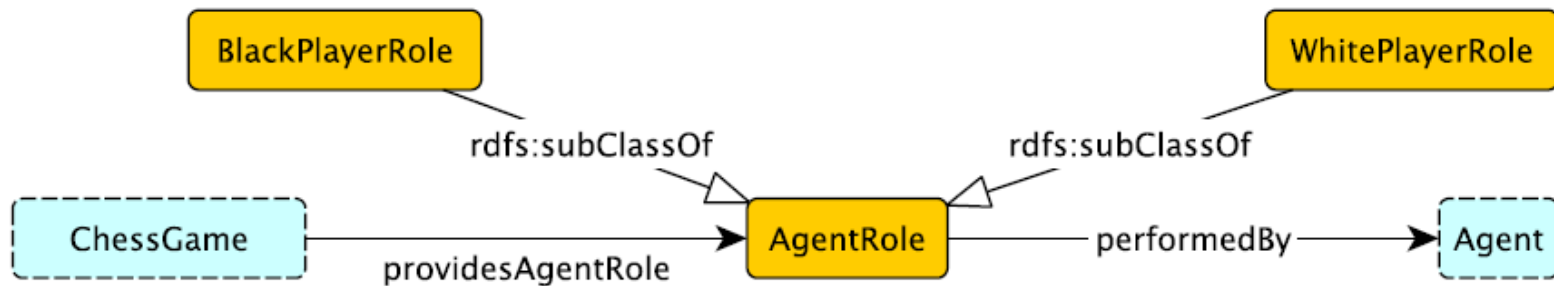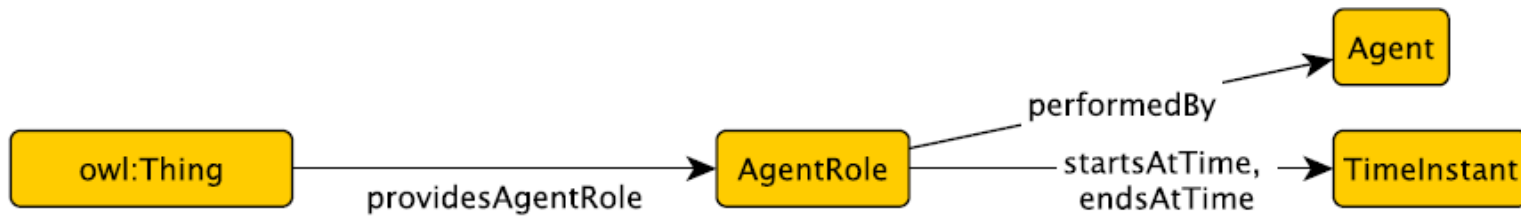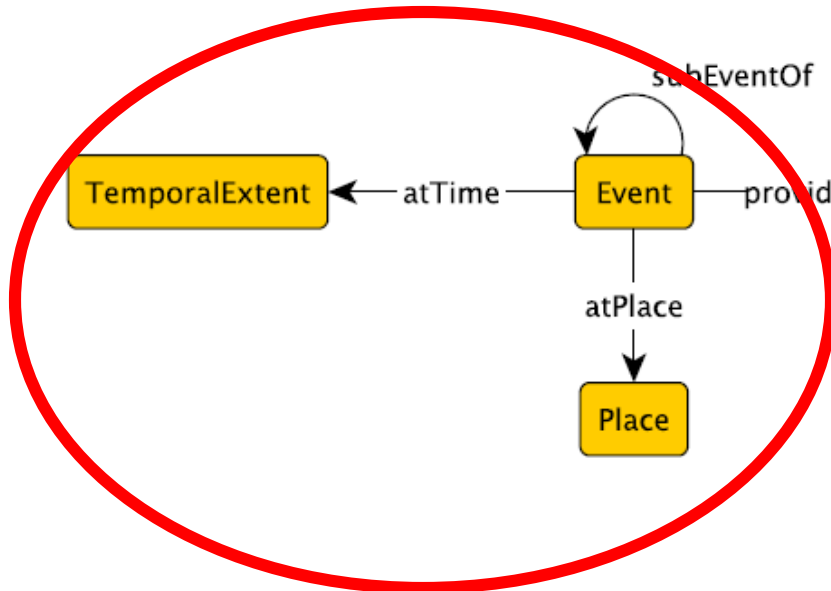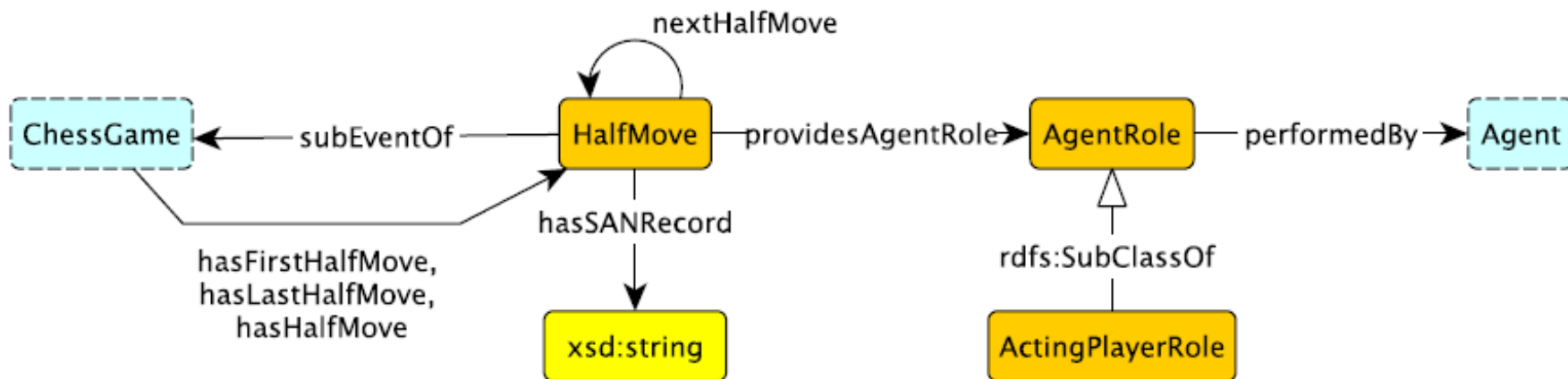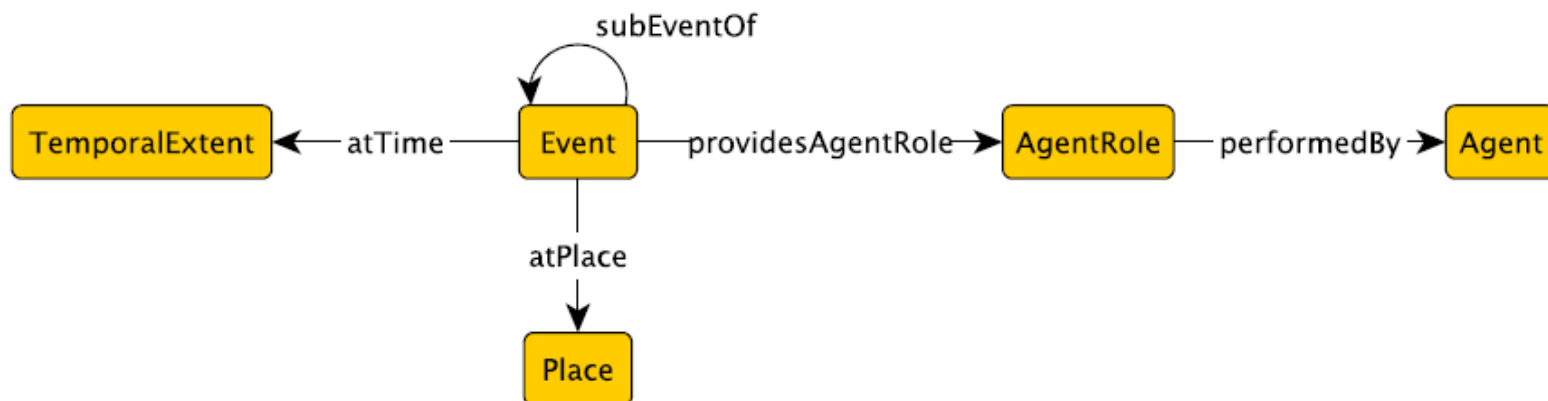
- **Understand the nature of the things you are modeling.**

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

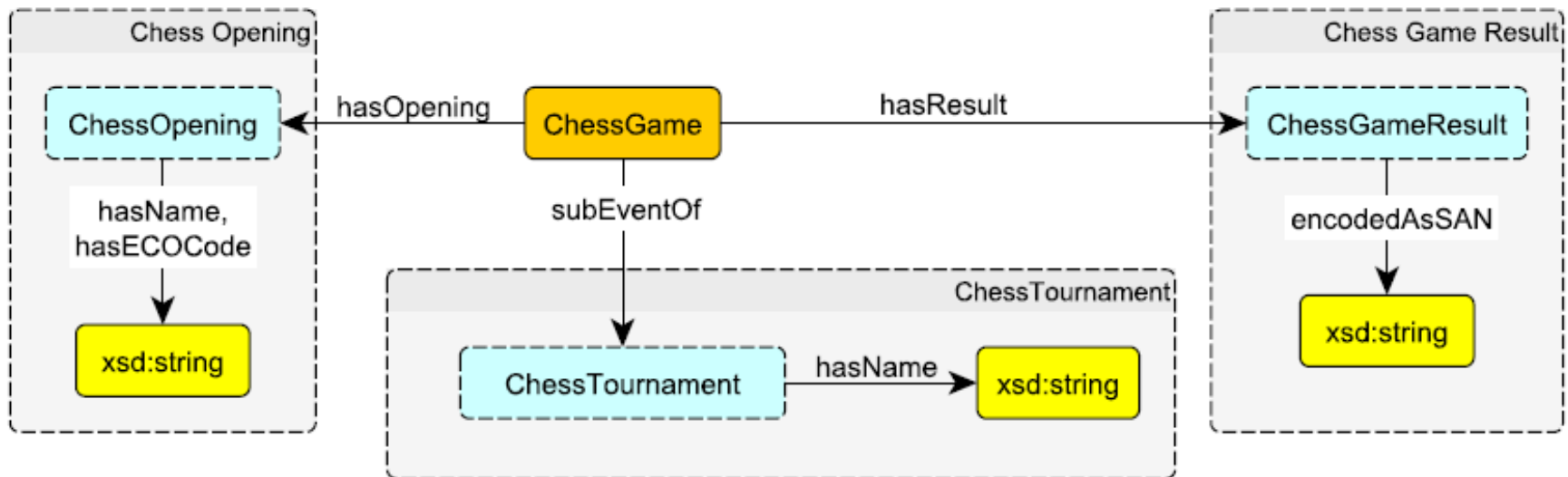# Half-moves
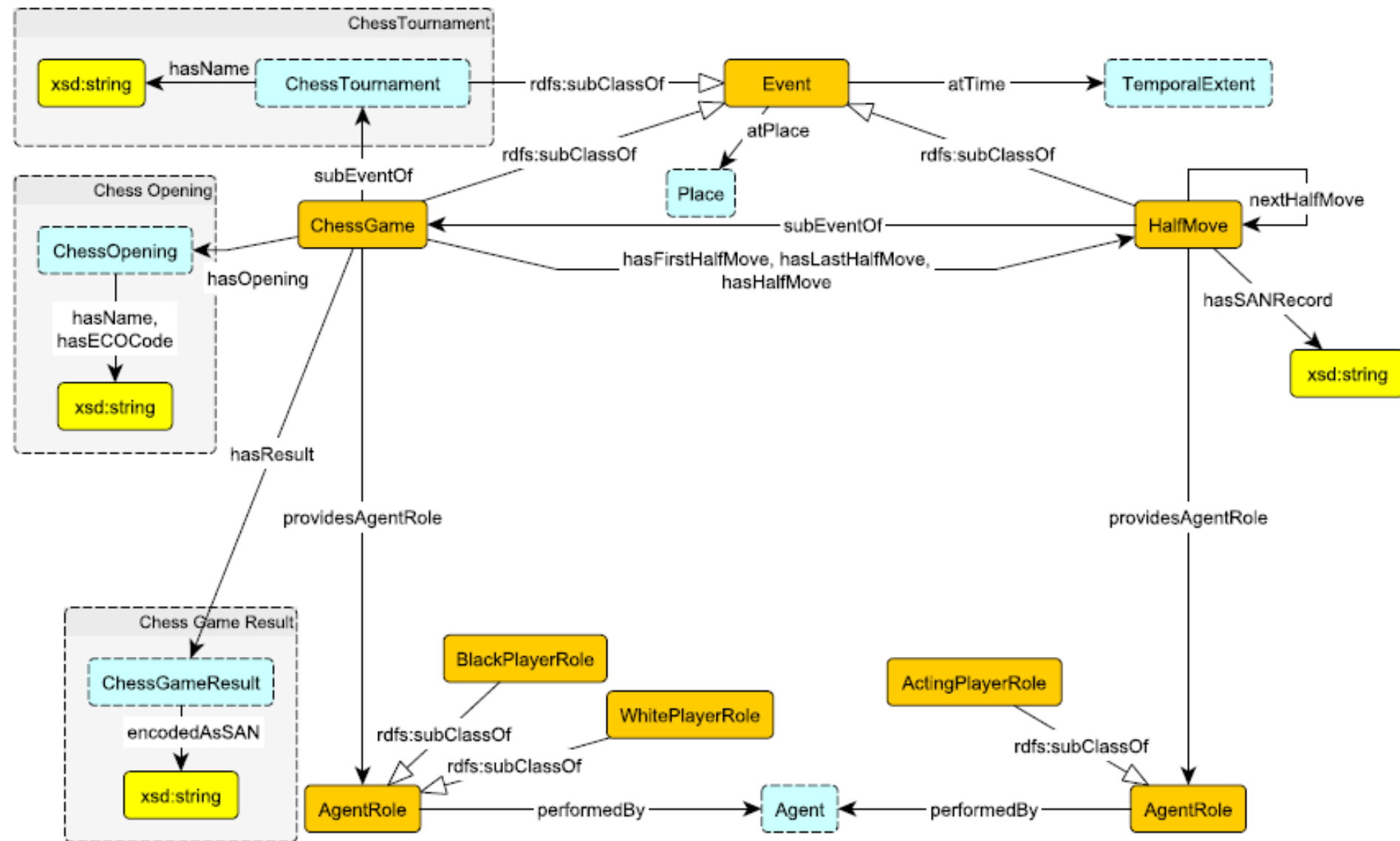
Opening, game result, etc.
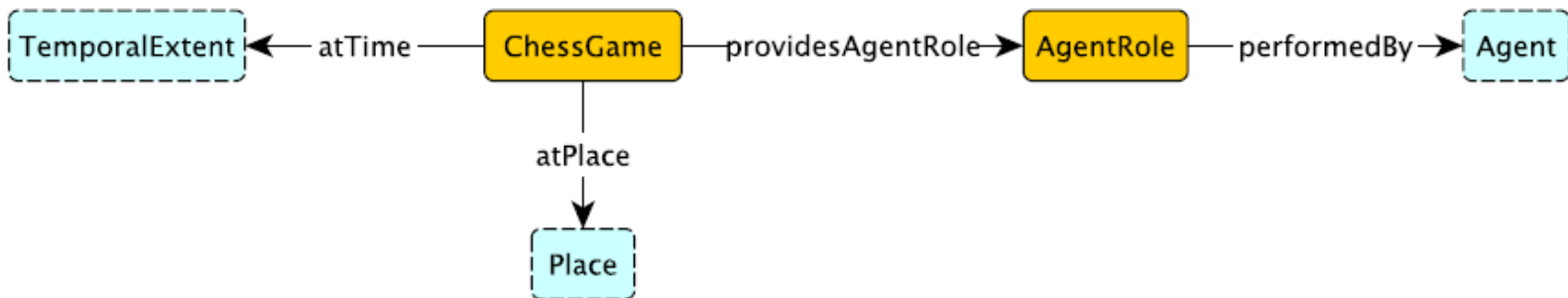
We call these "stubs".

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.

Adequacy check
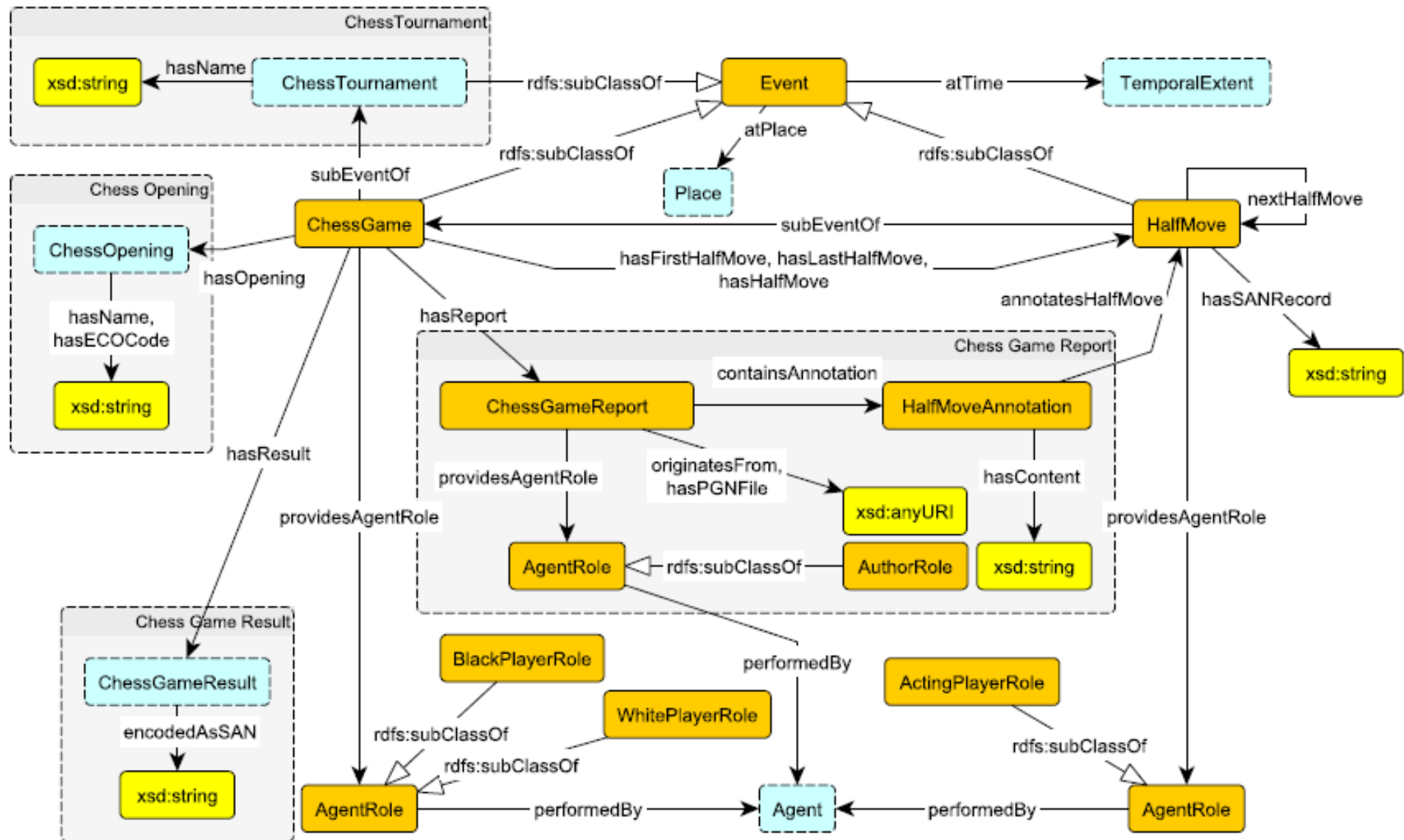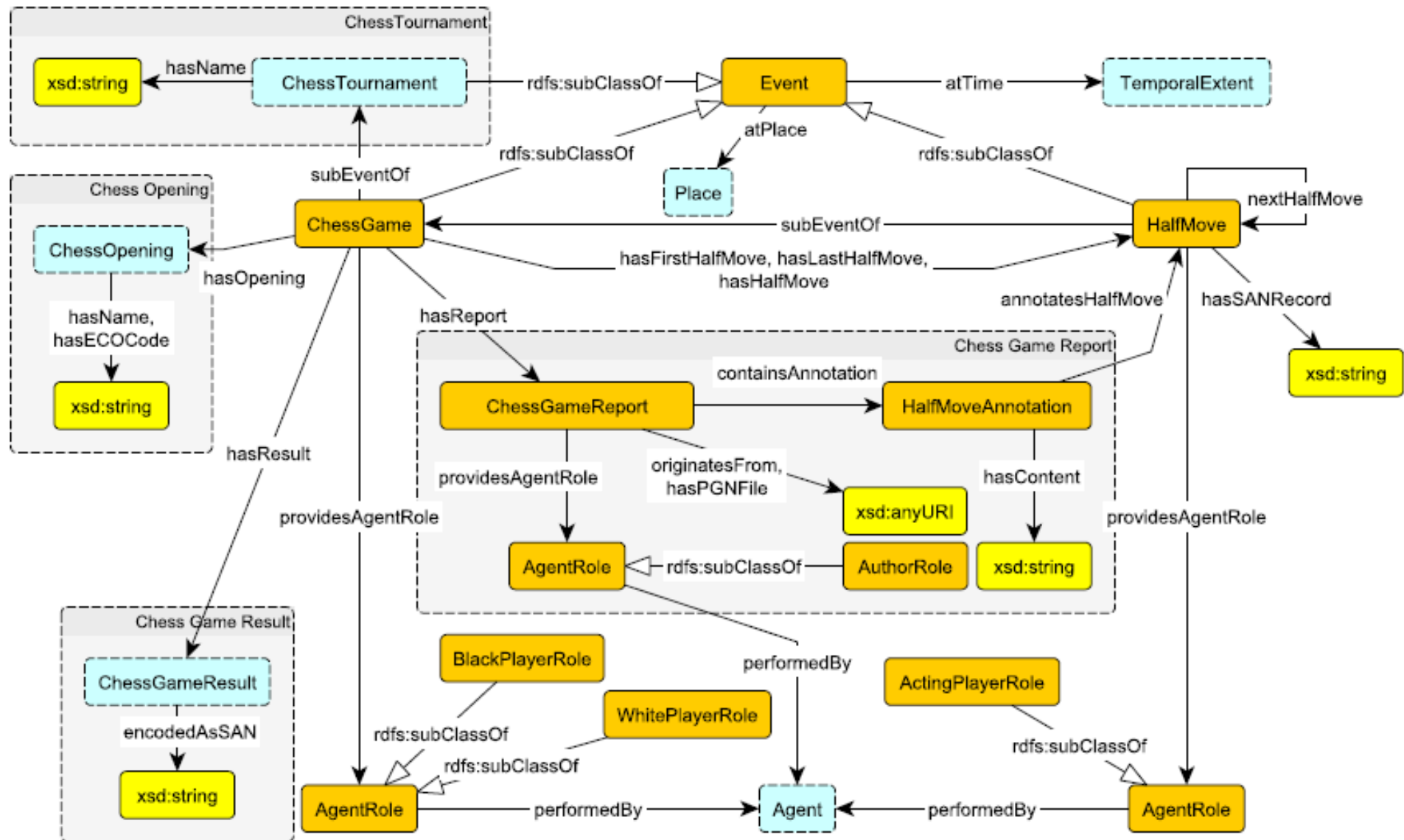
- **Triplify sample data using the ontology.
  Does it work?**

- **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)**

Axioms

Axioms in this case are mostly straightforward:

- **Inherited from Event or AgentRole**
- **Scoped domain/range restrictions, possibly with some cardinalities**
- **Basic existentials**
- **Non-cyclicity of half-move sequence**

- **Of course, some axioms may span several patterns/modules, and they would have to be added after assembly.**

What is a module?

A module is a part of an ontology which captures a conceptual entity (and its key properties) and which "makes sense" for a domain expert.

Of course, there can be modules within modules.

And modules can (and will!) overlap.

The axiomatization will usually not respect module boundaries.

Serializing modular ontologies

There needs to be a language for doing this. We have a preliminary version for this, called OPLa, using OWL annotation properties.

Pascal Hitzler, Aldo Gangemi, Krzystof Janowicz, Adila A. Krisnadhi, Valentina Presutti
Towards a simple but useful ontology design pattern representation language.
In: Proceedings WOP 2017.
To appear.

http://daselab.cs.wright.edu/pub2/wop2017-odpl.pdf

In particular, it can be used to identify the modules, and their provenance (which patterns they come from).

- **Focus on making a quality (modular) ontology first.**
- **Then create (and publish) the corresponding knowledge graph (aka ABox).**

- **The modular approach is designed to make it easier to arrive at good decisions for the schema.**

- **At the same time, the schema falls naturally into modules which correspond to experts' conceptualizations of key domain notions.**

sharing
integration
reuse

- It's much easier to understand what's in a knowledge graph if you have a schema which follows your (human) conceptualization of a domain. Modules provide exactly that.

- Modular modeling is extensible, and easier to modify. I.e. data management becomes more efficient (and cheaper) downstream.

We view ontologies as artefacts which mediate between data as structured for machine use and human conceptualizations.

**sharing
integration
reuse**

Data integration

If two different datasets each adhere to an ontology,

and both ontologies are built using modular principles,

then integration becomes easier.

- The two ontologies may share modules.
- The two ontologies may have modules which share patterns as their origins or starting points.

Key issue: Modelers often think of conceptual units anyway when modeling. It should be best practice to record this (in the ontology, in the documentation), as it's next to impossible to recreate later by others.

sharing
integration
reuse

What is missing?

- **We have a pretty good understanding of the modeling approach.**

- **We now even have a first language for representing modules.**

- **We are still missing a repository of quality patterns which can be used for assembly.**

- **We are still missing tools which support the complete workflow.**

- **We are still missing easy, step-by-step introductory material on how to use the process.**

- **How to deal with this notion of modularization from a logical perspective? How does it affect deduction?**

- **Schema is understood, informally, as a shape constraint. We may want to be able to interpret an axiomatization sometimes as open world, sometimes as closed world (shape contraints). This needs more work.**

- **We need role chains on right-hand-sides of subclass relationships, in order to meaningfully describe relationships between modules/patterns etc.**

- **We need ways to soften the regularity constraint on Rboxes.**

Thanks!

References

Hitzler, Krötzsch, Rudolph, Foundations of Semantic Web Technologies, CRC/Chapman & Hall, 2010

Adila Krisnadhi, Pascal Hitzler, Modeling With Ontology Design Patterns: Chess Games As a Worked Example. In: Pascal Hitzler, Aldo Gangemi, Krzysztof Janowicz, Adila Krisnadhi, Valentina Presutti (eds.), Ontology Engineering with Ontology Design Patterns: Foundations and Applications. Studies on the Semantic Web Vol. 25, IOS Press/AKA Verlag, 2016. Chapter 1, pp. 3-22.

Adila Krisnadhi, Nazifa Karima, Pascal Hitzler, Reihaneh Amini, Michelle Cheatham, Víctor Rodríguez-Doncel, Krzysztof Janowicz, Ontology Design Patterns for Linked Data Publishing. In: Pascal Hitzler, Aldo Gangemi, Krzysztof Janowicz, Adila Krisnadhi, Valentina Presutti (eds.), Ontology Engineering with Ontology Design Patterns: Foundations and Applications. Studies on the Semantic Web Vol. 25, IOS Press/AKA Verlag, 2016. Chapter 10, pp. 201-232.

References

Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider, Sebastian Rudolph, OWL 2 Web Ontology Language: Primer (Second Edition). W3C Recommendation, 11 December 2012.

Michelle Cheatham, Pascal Hitzler, String Similarity Metrics for Ontology Alignment. In: H. Alani et al. (eds.), The Semantic Web - ISWC 2013. 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part II. Lecture Notes in Computer Science Vol. 8219, Springer, Heidelberg, 2013, pp. 294-309.

A. Gangemi. Ontology design patterns for semantic web content. In Y. Gil et al. (eds), The Semantic Web - ISWC 2005 – 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005, Proceedings, volume 3729 of Lecture Notes in Computer Science, pages 262-276. Springer, 2005

Eva Blomqvist, Pascal Hitzler, Krzysztof Janowicz, Adila Krisnadhi, Thomas Narock, Monika Solanki, Considerations regarding Ontology Design Patterns. Semantic Web 7 (1) 1-7.

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.

References

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.

Adila Krisnadhi, Ontology Pattern-Based Data Integration. Dissertation, Department of Computer Science and Engineering, Wright State University, 2015.