Modular Schema Development for Wikibase

Pascal Hitzler
Data Semantics Laboratory (DaSe Lab)
Kansas State University

http://www.daselab.org

Collaborators:

Cogan Shimizu, Andrew Eells, Lu Zhou — DaSeLab, Kansas State University
Seila Gonzalez, Alicia Sheill, Catherine Foley, Dean Rehberger — MATRIX, Michigan State University

Knowledge Graphs vs. Linked (Open) Data

Knowledge Graphs are like linked data, with

- **linking de-emphasized**
- **openness de-emphasized**
- **renewed consideration of schema quality**
- **tighter and central control**
- **clear transition from academia to industry**

Goal is to produce a flexible, extendable, end-user friendly but in the end rather tightly controlled repository of integrated, re-useable data.

Authors from: Google, Microsoft, IBM, Facebook, eBay

PRACTICE

Industry-Scale Knowledge Graphs: Lessons and Challenges

By Natasha Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, Jamie Taylor

Communications of the ACM, August 2019, Vol. 62 No. 8, Pages 36-43
10.1145/3331166

Comments

Credit: Adempercem / Stutterstock

Knowledge graphs are critical to many enterprises today: They provide the structured data and factual knowledge that drive many products and make them more intelligent and "magical."

In general, a knowledge graph describes objects of interest and connections between them. For example, a knowledge graph may have nodes for a movie, the actors in this movie, the director, and so on. Each node may have properties such as an actor's name and age. There may be nodes for multiple movies involving a particular actor. The user can then traverse the knowledge graph to collect information on all the movies in which the actor appeared or, if applicable, directed.

Many practical implementations impose constraints on the links in knowledge graphs by defining a *schema* or *ontology*. For example, a link from a movie to its director must connect an object of type Movie to an object of type Person. In some cases the links themselves might have their own properties: a link connecting an actor and a movie might have the name of the specific role the actor

MORE NEWS & OPINIONS
MIT Robot Could Help People

Wikidata

Started 2012 at Wikimedia Deutschland, with funding by Google, Yandex, Allen Institute for AI.

Crowdsourced knowledge graph,

like Wikipedia is a crowdsourced encyclopedia.

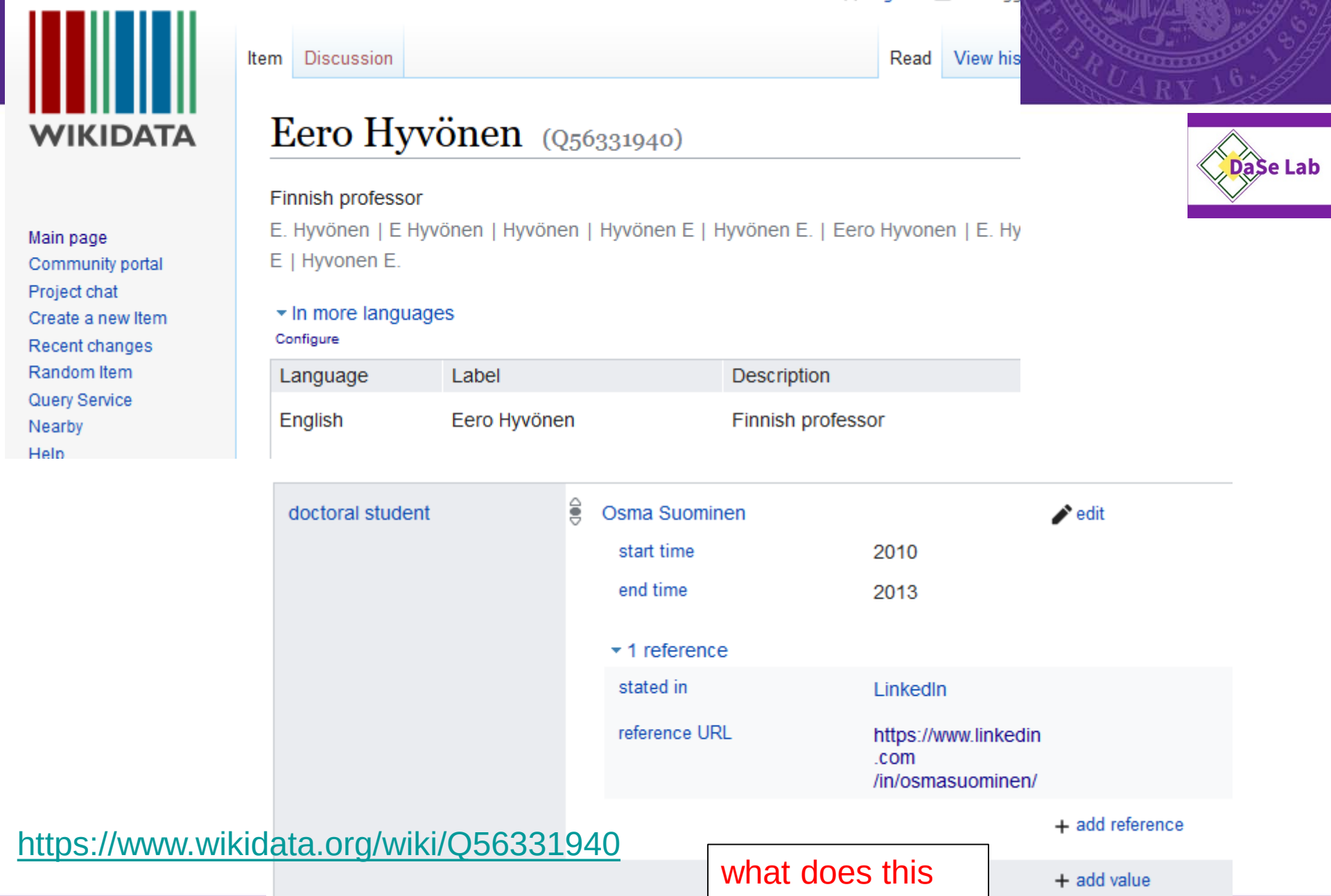
Wikibase Software, Mediawiki interface.

Constrains graph shapes wrt. context information, references, datatype use.

Otherwise crowdsourced.

Wikibase can be set up locally, used for your own knowledge graph.

- **powerful tool support for management, including interface**

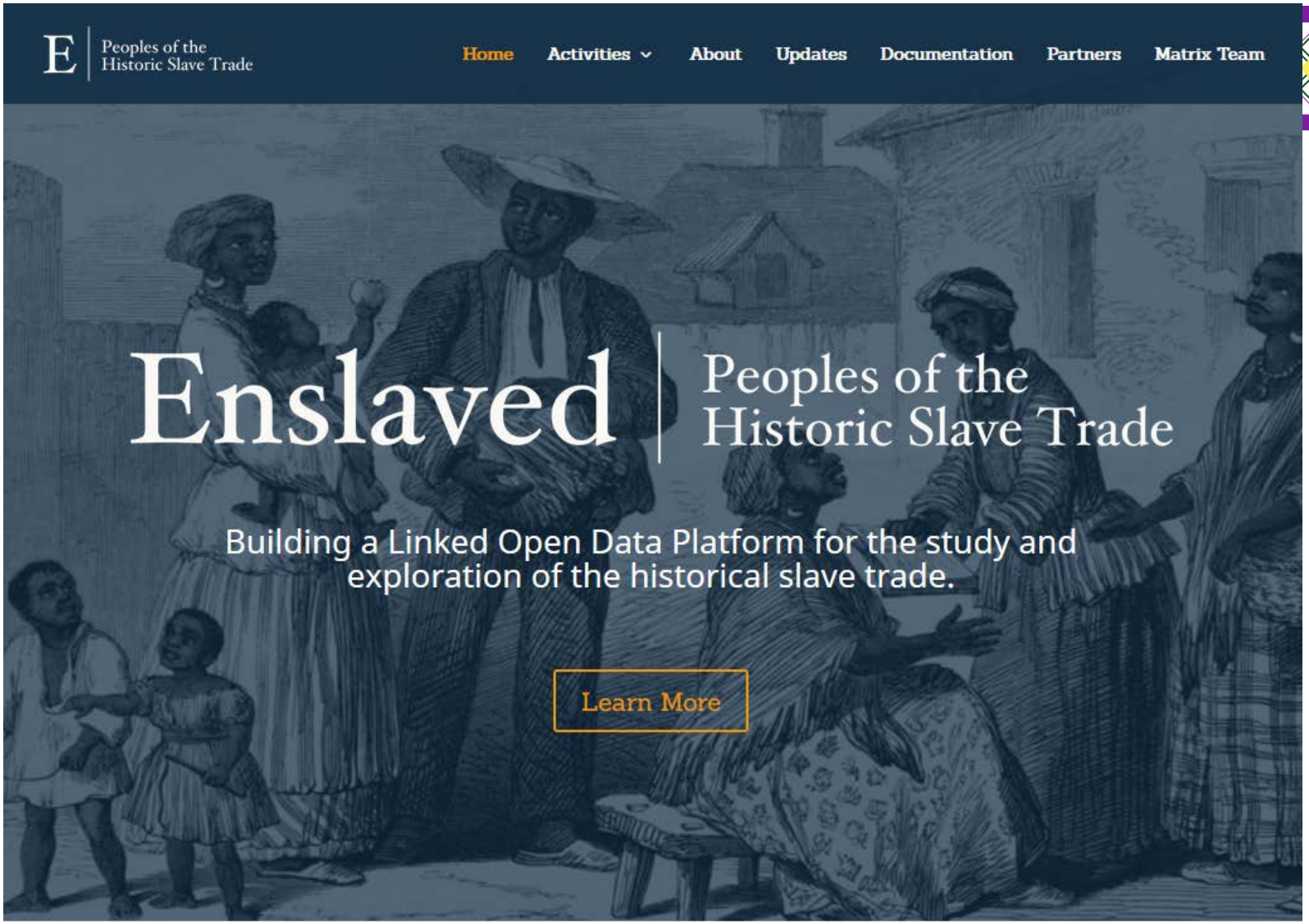


https://www.wikidata.org/wiki/Q56331940

what does this look like in RDF?

enslaved.org

https://enslaved.org/ https://lod.enslaved.org



DaSe Lab

enslaved.org process

1. **Quality Ontology Design.**
2. **Realization of ontology-based schema in Wikibase.**
3. **Knowledge graph construction and interaction through Wikibase as engine.**
4. **Additional front-end (simplified view)**

DaSe Lab

(4) https://enslaved.org/

(3) https://lod.enslaved.org/

>53M RDF triples from Wikibase export

Focus of this talk: Going from (1) to (2)

PEOPLE

Gender

Age Category

Ethnodescriptor

Role Types

Occupation

Status

EVENT

Event Type

Date

PLACE

Place Type

Showing 20 of 552009 Results

Sort By

20 Per Page

People

Ardealah

Sex Female

Person Status Enslaved Person

Place Multiple

Person's Connections

4 3 3

Chemorowee

Sex Female

Person Status Enslaved Person

Place Multiple

Person's Connections

4 3 3

Allarbah

Sex Female

Person Status Enslaved Person

Place Multiple

Person's Connections

4 3 3

Arnahyajumah

Sex Female

Person Status Enslaved Person

Place Multiple

Person's Connections

4 3 3

Sannom (Q358958)

LSD-PER-075163
LSD-PER-075163

▾ In more languages
Configure

Language	Label	Description
English	Sannom	LSD-PER-075163

Statements

instance of	Person ▾ 0 references
hasName	Sannom recordedAt: Sale of unnamed enslaved persons by Mathurin Guerin and sons (1817-6-7) ▸ 1 reference
hasSex	Male recordedAt: Sale of unnamed enslaved persons by Mathurin Guerin and sons (1817-6-7) ▸ 1 reference
hasPersonStatus	Enslaved Person hasStatusGeneratingEvent: Sale of unnamed

Sale of unnamed enslaved persons by Mathurin Guerin and sons (1817-6-7) (Q310252)

LSD-EVE-SAL-27571
LSD-EVE-SAL-27571

▾ In more languages
Configure

Language	Label	Description	Also known as
English	Sale of unnamed enslaved persons by Mathurin Guerin and sons (1817-6-7)	LSD-EVE-SAL-27571	LSD-EVE-SAL-27571

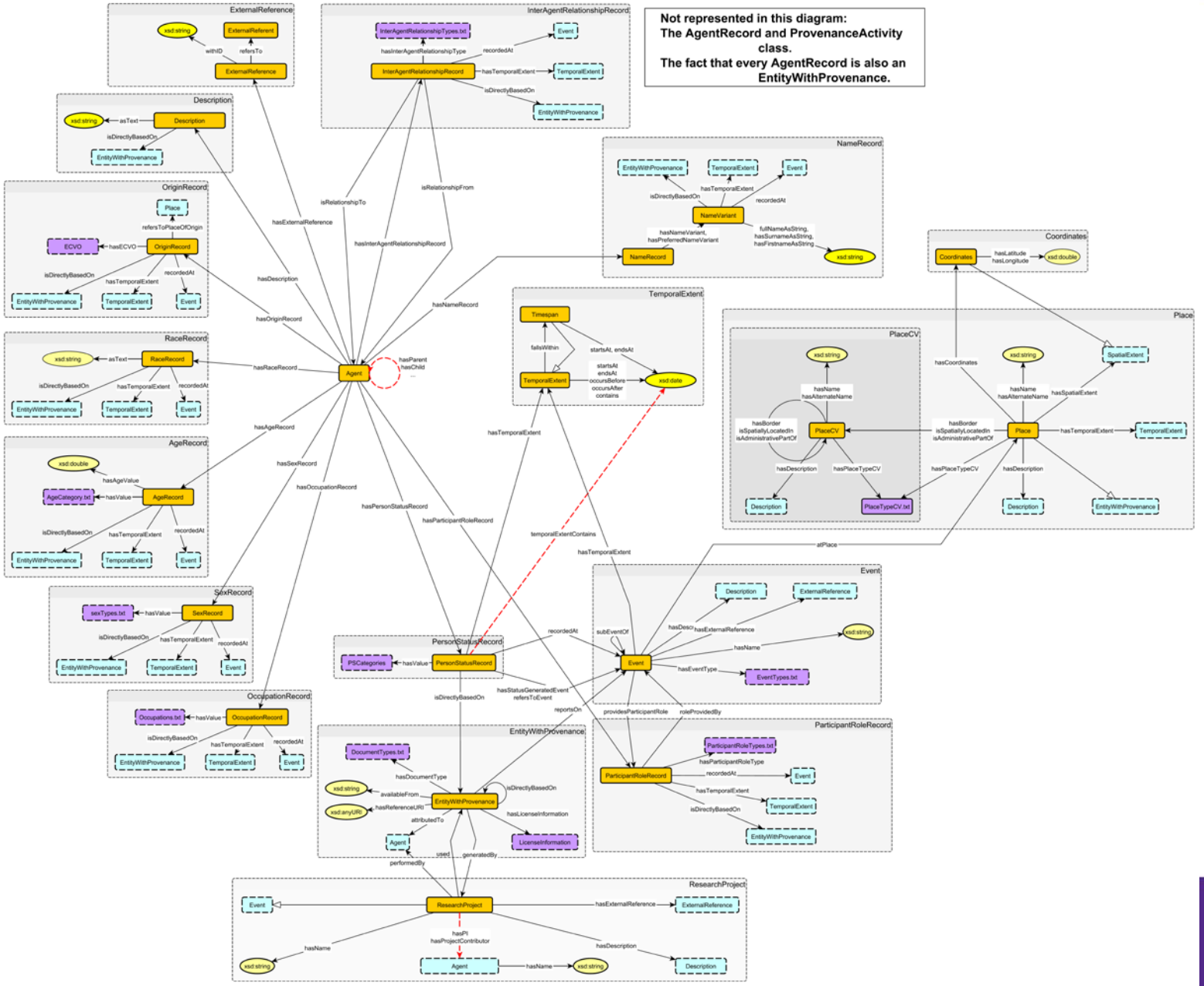
Statements

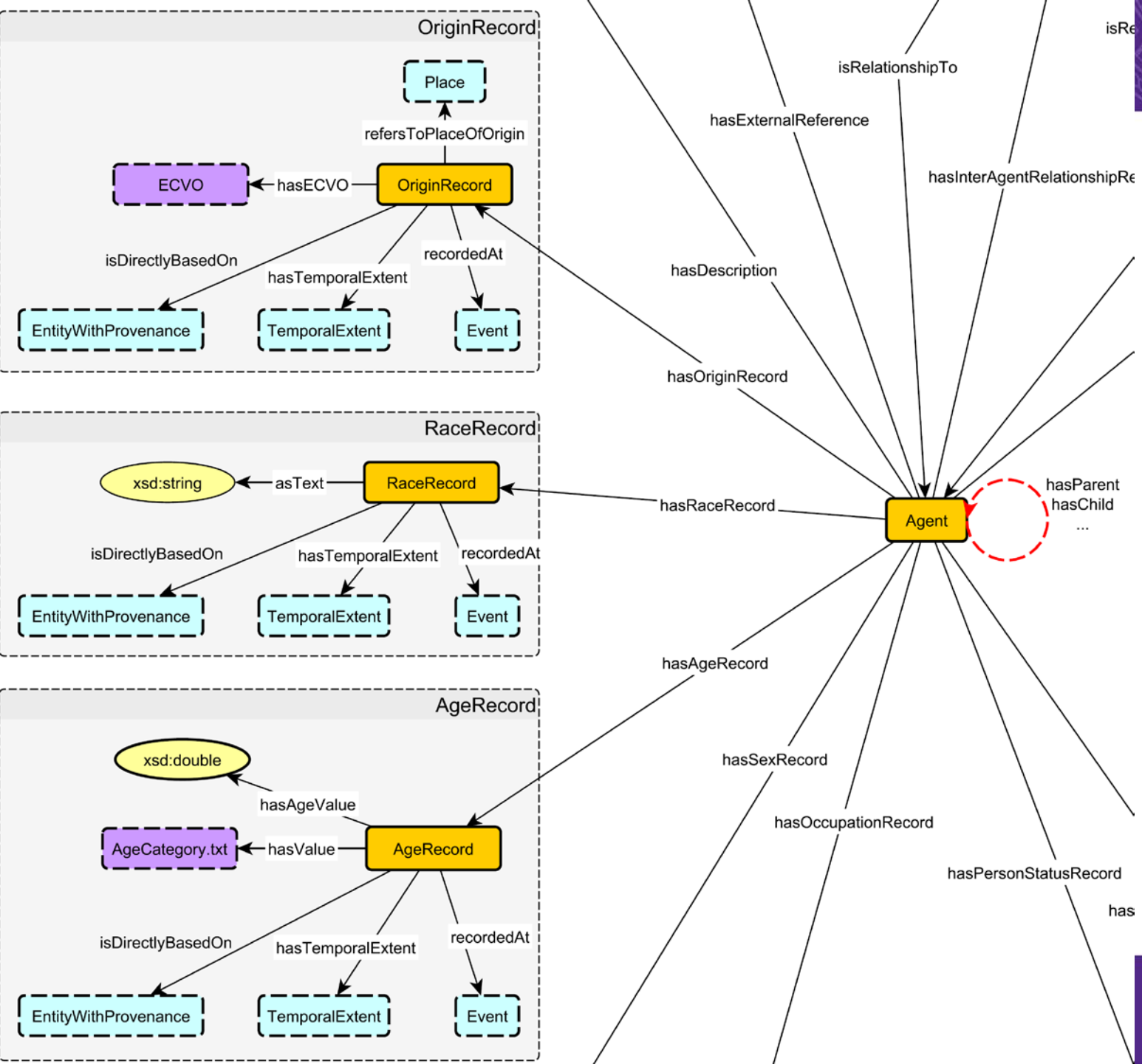
instance of	Event ▾ 0 references
hasName	Sale of unnamed enslaved persons by Mathurin Guerin and sons (1817-6-7) ▸ 1 reference
hasEventType	Sale ▸ 1 reference
date	7 June 1817 *Gregorian* ▸ 1 reference
atPlace	St. James ▸ 1 reference
providesParticipantRole	Sold Person hasParticipantRole: Isaac, James, John, Sarrah, Angelique, Congo Francoise, Alexandre

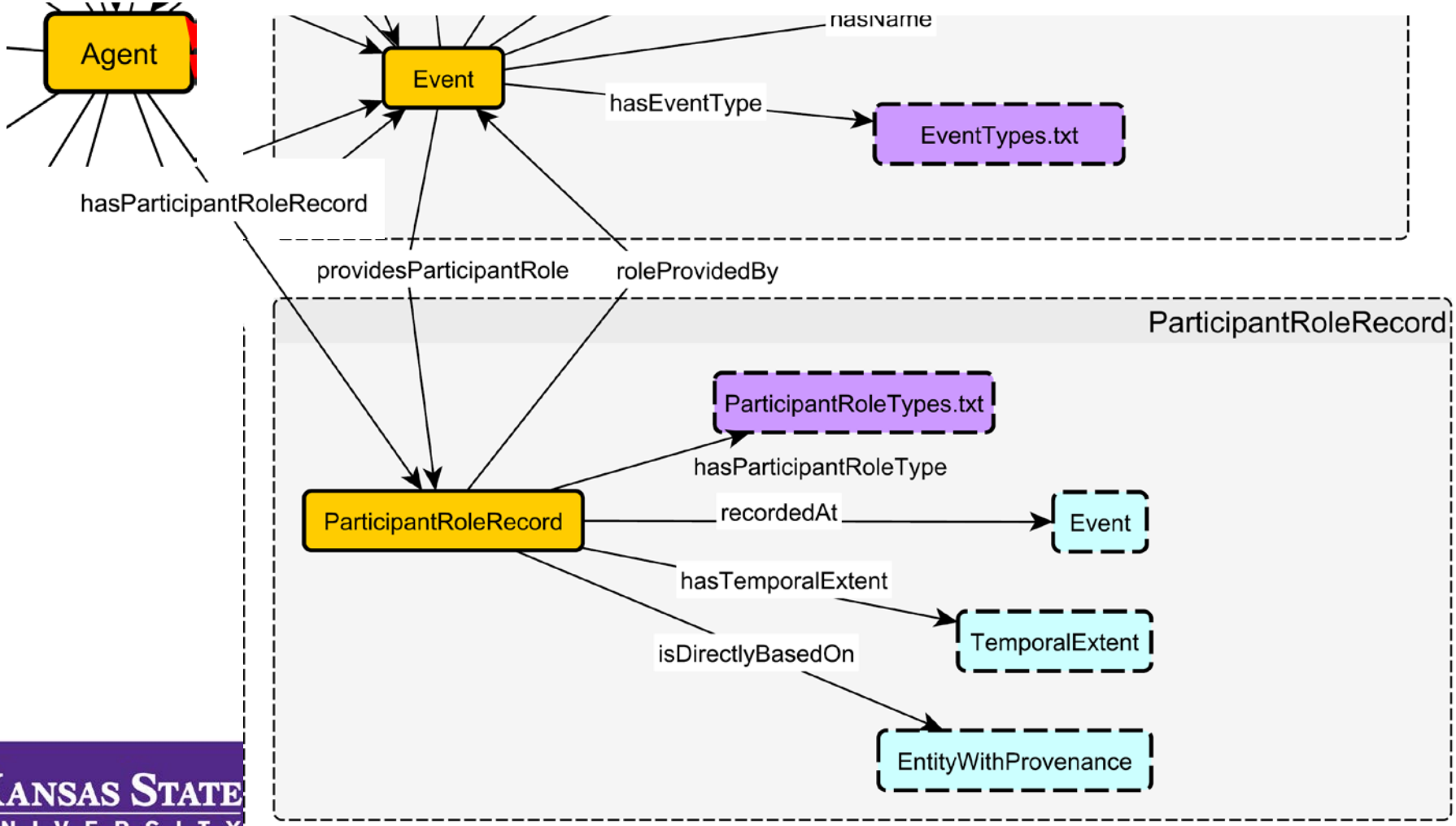
Ontology

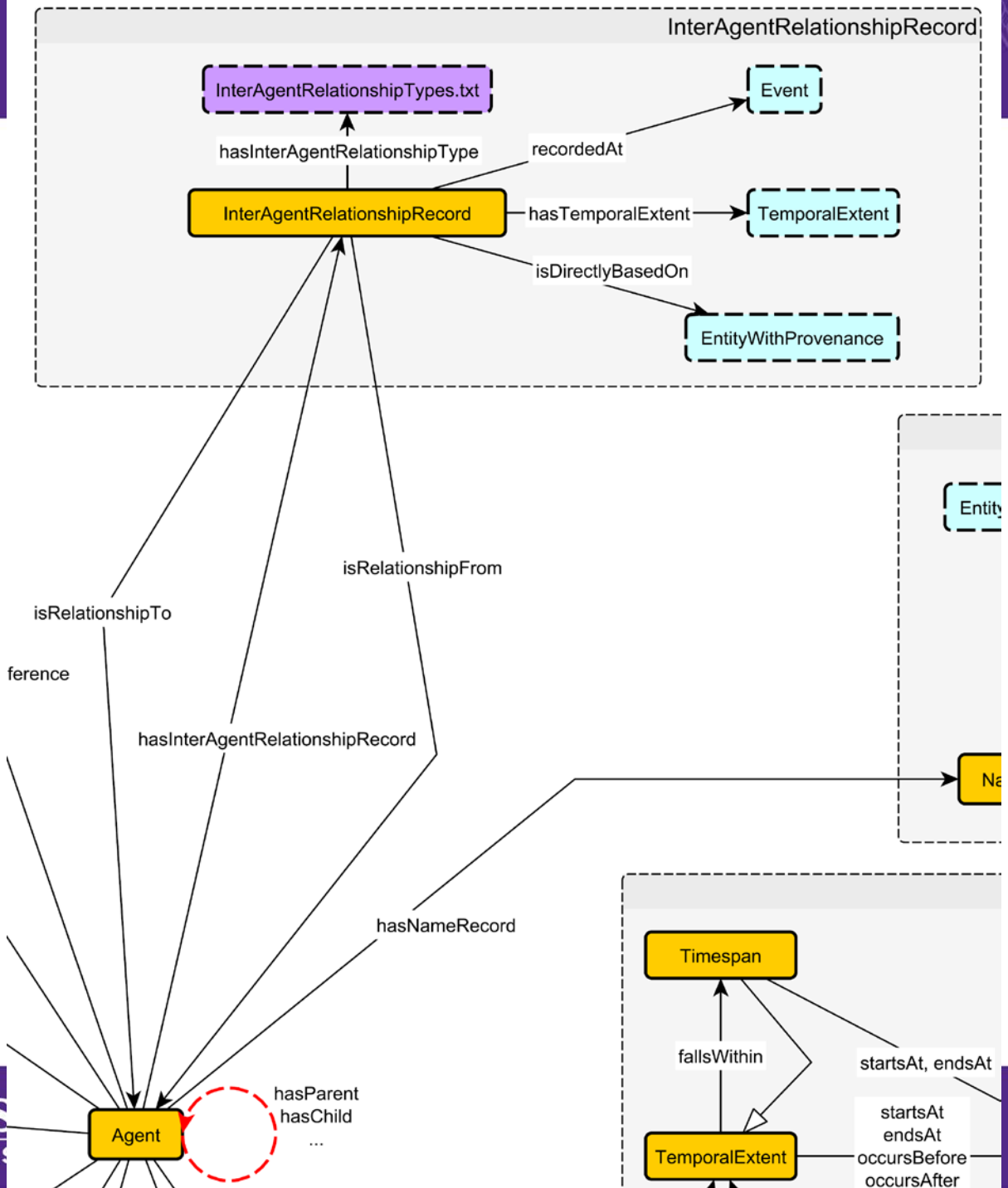
- Modular ontology modeling process to obtain a high-quality knowledge graph schema. [Shimizu et al, 2021]
- Focus on Persons, Events, Places.
- Developed using an ontology design patterns approach.
- Primary concern: quality of the ontology for later re-use and/or extension.
- In particular: We ignored Wikibase constraints.

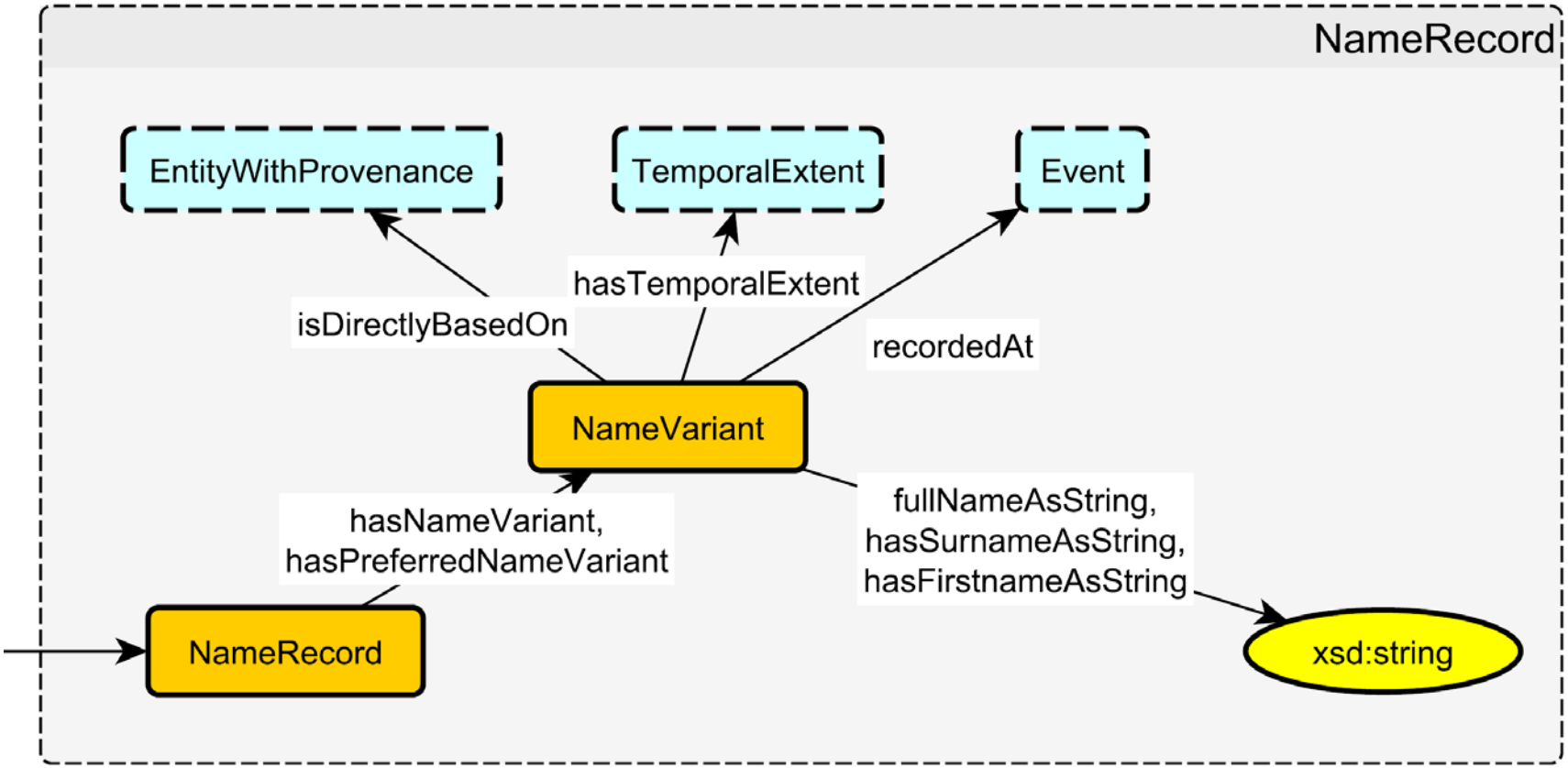
Enslaved Ontology [JWS 2020]




OriginRecord
Place
refersToPlaceOfOrigin
ECVO
hasECVO
OriginRecord
isDirectlyBasedOn
hasTemporalExtent
recordedAt
EntityWithProvenance
TemporalExtent
Event
RaceRecord
xsd:string
asText
RaceRecord
isDirectlyBasedOn
hasTemporalExtent
recordedAt
EntityWithProvenance
TemporalExtent
Event
AgeRecord
xsd:double
hasAgeValue
AgeCategory.txt
hasValue
AgeRecord
isDirectlyBasedOn
hasTemporalExtent
recordedAt
EntityWithProvenance
TemporalExtent
Event
isRelationshipTo
hasExternalReference
hasInterAgentRelationshipRe
hasDescription
hasOriginRecord
hasRaceRecord
Agent
hasParent
hasChild
...
hasAgeRecord
hasSexRecord
hasOccupationRecord
hasPersonStatusRecord
has


Agent
Event
hasName
hasEventType
EventTypes.txt
hasParticipantRoleRecord
providesParticipantRole
roleProvidedBy
ParticipantRoleRecord
ParticipantRoleTypes.txt
hasParticipantRoleType
ParticipantRoleRecord
recordedAt
Event
hasTemporalExtent
TemporalExtent
isDirectlyBasedOn
EntityWithProvenance


InterAgentRelationshipRecord
InterAgentRelationshipTypes.txt
Event
hasInterAgentRelationshipType
recordedAt
InterAgentRelationshipRecord
hasTemporalExtent
TemporalExtent
isDirectlyBasedOn
EntityWithProvenance
isRelationshipFrom
isRelationshipTo
ference
hasInterAgentRelationshipRecord
hasNameRecord
Timespan
fallsWithin
startsAt, endsAt
hasParent
hasChild
...
Agent
TemporalExtent
startsAt
endsAt
occursBefore
occursAfter


NameRecord
EntityWithProvenance
TemporalExtent
Event
isDirectlyBasedOn
hasTemporalExtent
recordedAt
NameVariant
hasNameVariant,
hasPreferredNameVariant
fullNameAsString,
hasSurnameAsString,
hasFirstnameAsString
NameRecord
xsd:string

Ontology into Wikibase: Issues

- The Wikibase realization is similar but not identical to the ontology.
- The RDF export is based on an ontology (graph schema) that is inferred from the Wikibase realization.

- How difficult is it to map between these two ontologies (and thus, between these to graphs?

- We used this as basis for an OAEI “complex ontology alignment” benchmark, published at CIKM 2020
  - This realistic benchmark is currently way beyond capabilities of automated ontology alignment systems.

Enslaved benchmark example mapping rules

$$enslaved : Person(x) \wedge enslaved : hasRaceRecord(x, y) \wedge enslaved : RaceRecord(y) \wedge enslaved : asText(y, z) \leftrightarrow ed : Q410(Person)(x) \wedge ep : P32(hasRaceorColor)(x, y) \wedge wikibase : Statement(y) \wedge eps : P32(hasRaceorColor)(y, z)$$

$$enslaved : Person(w) \wedge enslaved : hasAgeRecord(w, x) \wedge enslaved : AgeRecord(x) \wedge enslaved : hasAgeValue(x, z) \leftrightarrow ed : Q410(Person)(w) \wedge ep : P42(hasAge)(w, x) \wedge wikibase : Statement(x) \wedge eps : P42(hasAge)(x, y) \wedge ed : Q424(AgeRecord)(y) \wedge edt : P3(hasAgeValue)(y, z)$$

$$enslaved : Person(x) \wedge enslaved : hasInterAgentRelationshipRecord(x, y) \wedge enslaved : InterAgentRelationshipRecord(y) \wedge enslaved : hasInterAgentRelationshipType(y, z) \wedge enslaved : InterAgentRelationshipTypes(z) \leftrightarrow ed : Q410(Person)(x) \wedge ep : P39(hasInterAgentRelationshipTypeTo)(x, y) \wedge wikibase : Statement(y) \wedge eps : P39(hasInterAgentRelationshipTypeTo)(y, z) \wedge ed : Q463(InteragentRelationship)(z)$$

$$enslaved : Person(x) \wedge enslaved : hasParticipantRoleRecord(x, y) \wedge enslaved : ParticipantRoleRecord(y) \wedge enslaved : roleProvidedBy(y, z) \wedge enslaved : Event(z) \leftrightarrow ed : Q410(Person)(x) \wedge ep : P17(hasParticipantRole)(x, y) \wedge wikibase : Statement(y) \wedge epq : P19(roleProvidedBy)(y, z) \wedge ed : Q238(Event)(z)$$

$$enslaved : Person(w) \wedge enslaved : hasNameRecord(w, x) \wedge enslaved : NameRecord(x) \wedge enslaved : hasPreferredNameVariant(x, y) \wedge enslaved : NameVariant(y) \wedge enslaved : fullNameAsString(y, z) \leftrightarrow ed : Q410(Person)(w) \wedge ep : P20(hasName)(w, x) \wedge wikibase : Statement(x) \wedge eps : P20(hasName)(x, z)$$

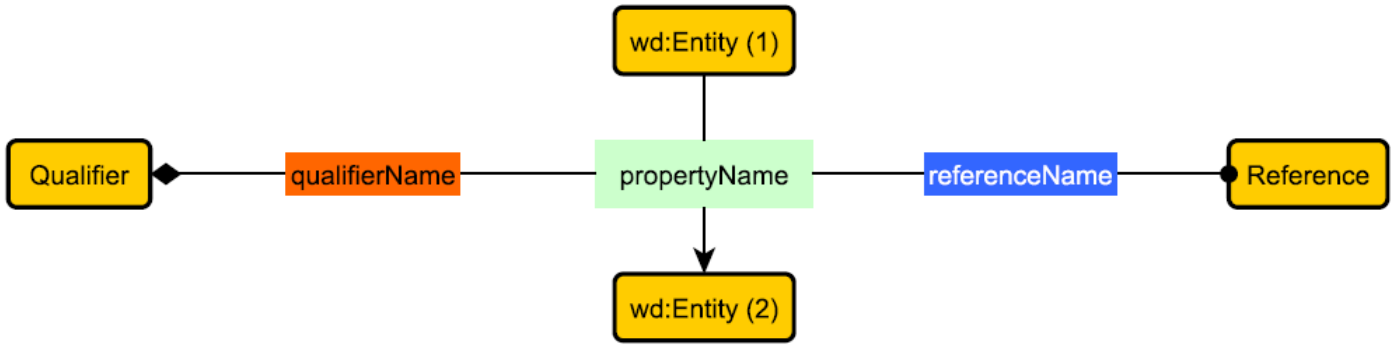
Ontology modeling for Wikibse

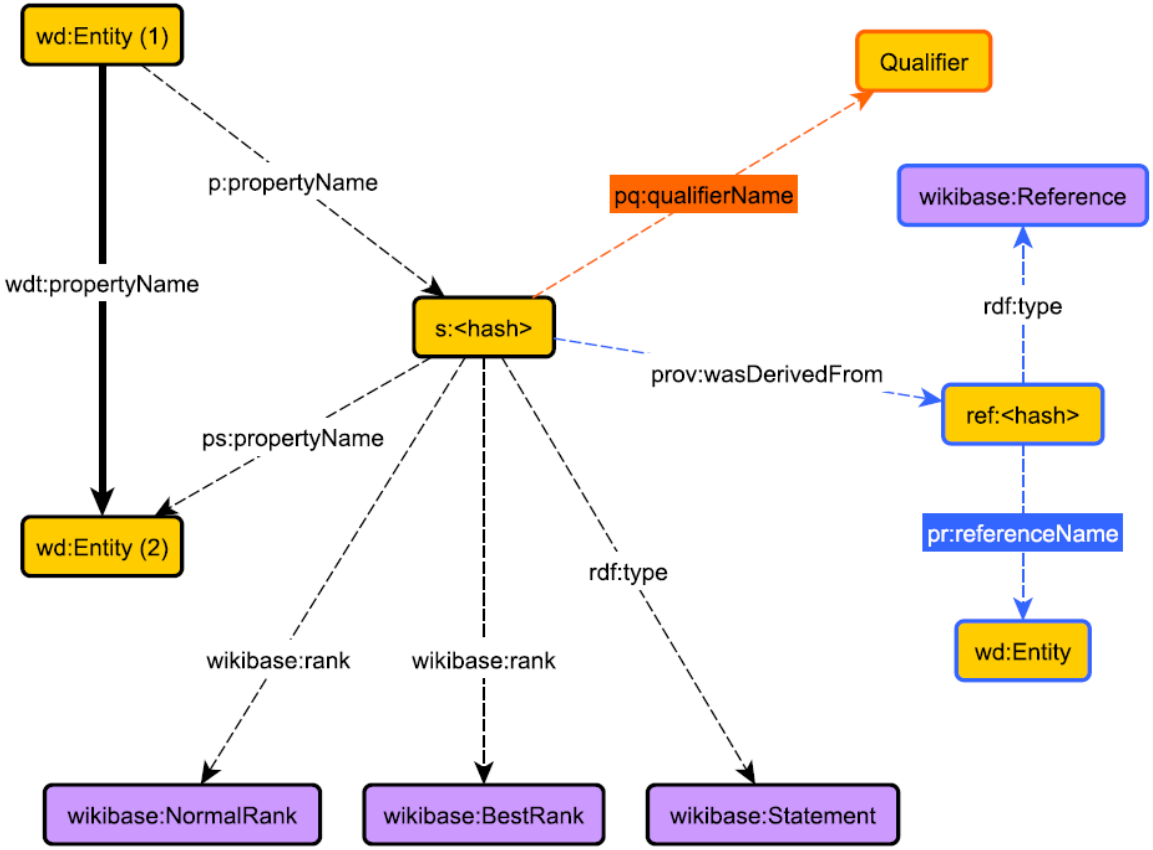
- **Provide “patterns” (Ontology design patterns) that bridge the gap between modular ontology modeling and Wikibase use for the corresponding graph.**

- **The patterns shall give a traditional modeling look-and-feel,**
- **while at the same time restricting the modeling to seamless transfer to Wikibase.**

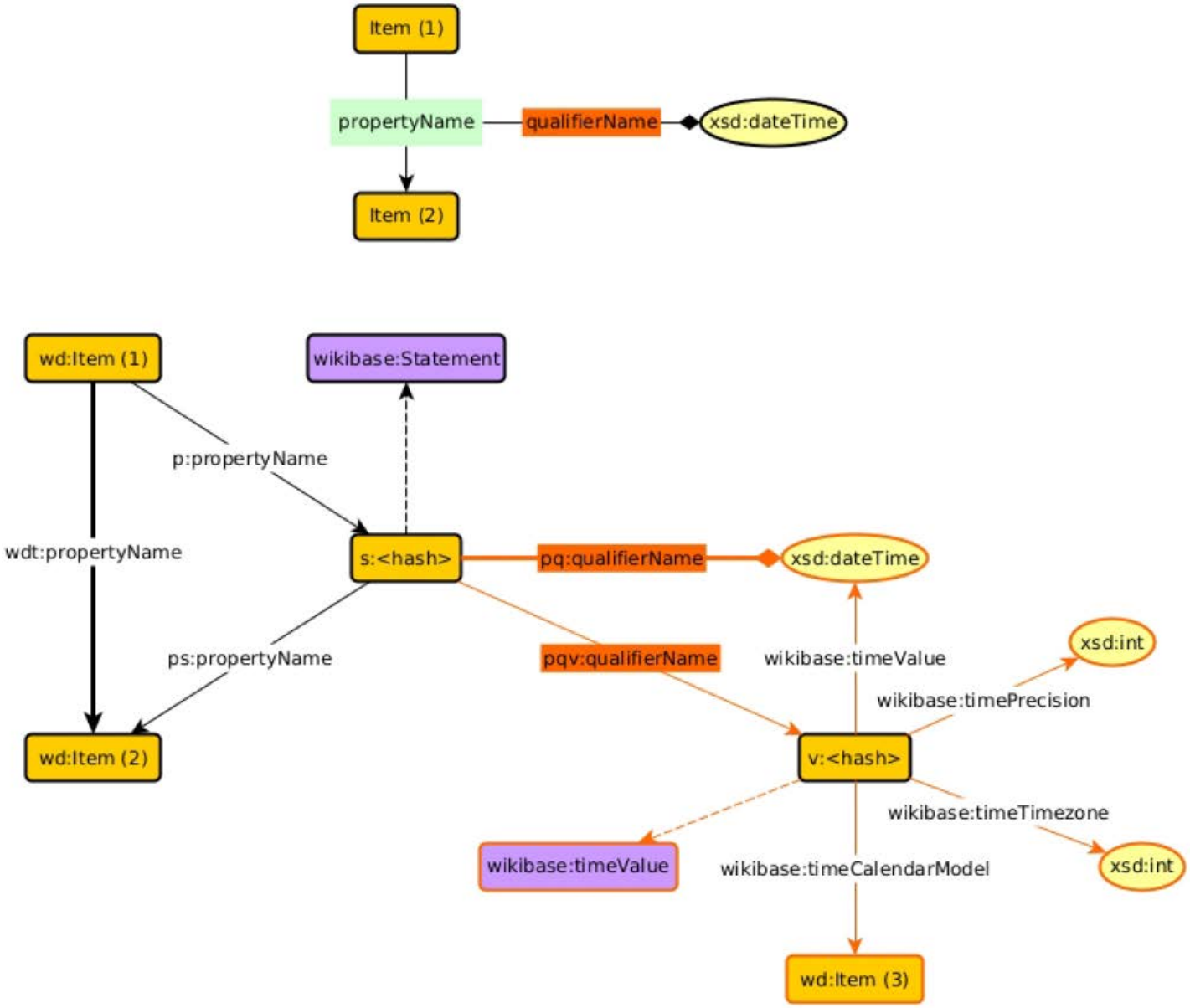
- **Modeling with domain experts done mostly via schema diagrams**
  - **technique taken from Modular Ontology Modeling**
  - **there is always an underlying (OWL) axiomatization of course.**

Wikibase ODPs: basic

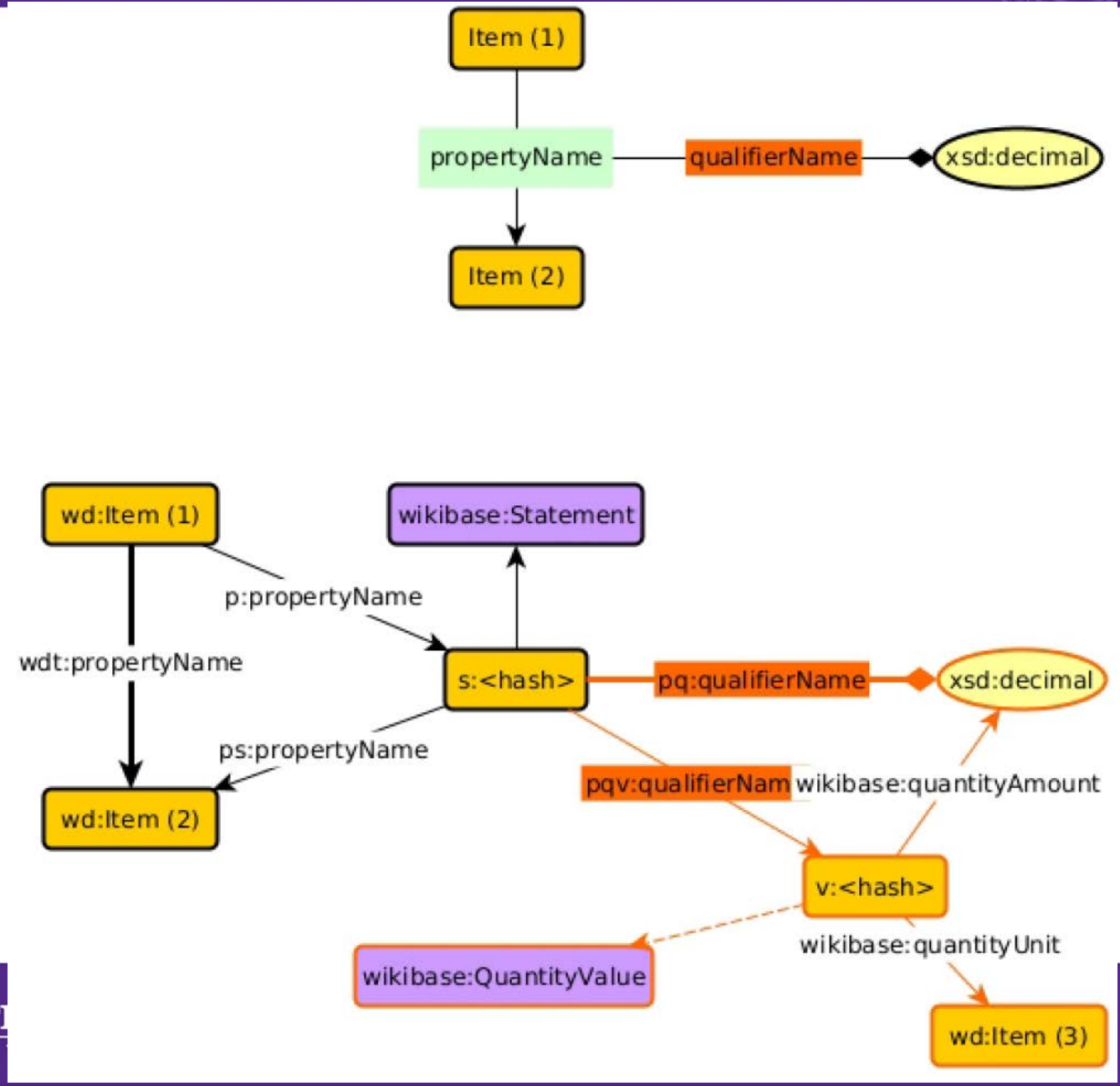




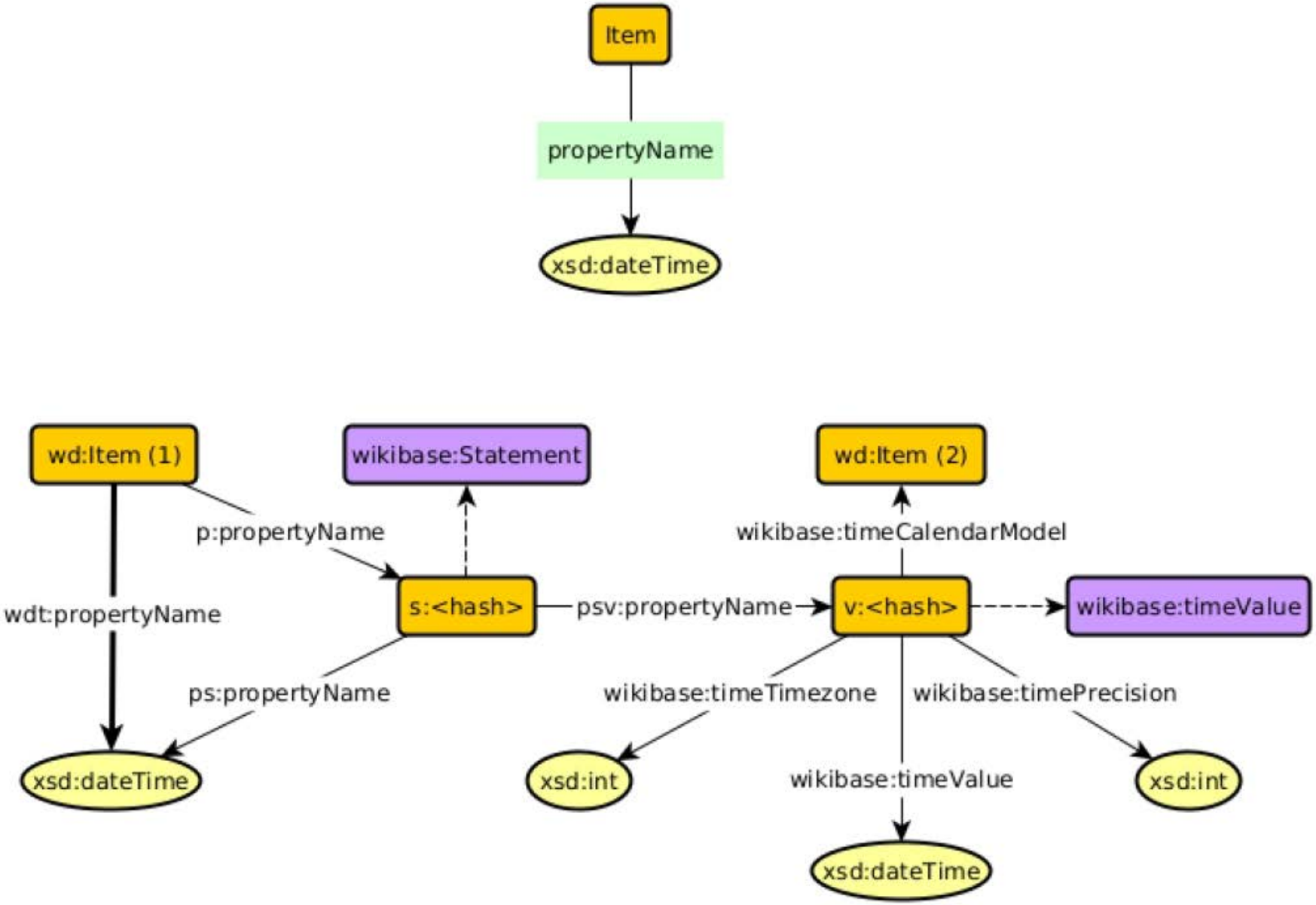
Wikibase ODPs: dateTime



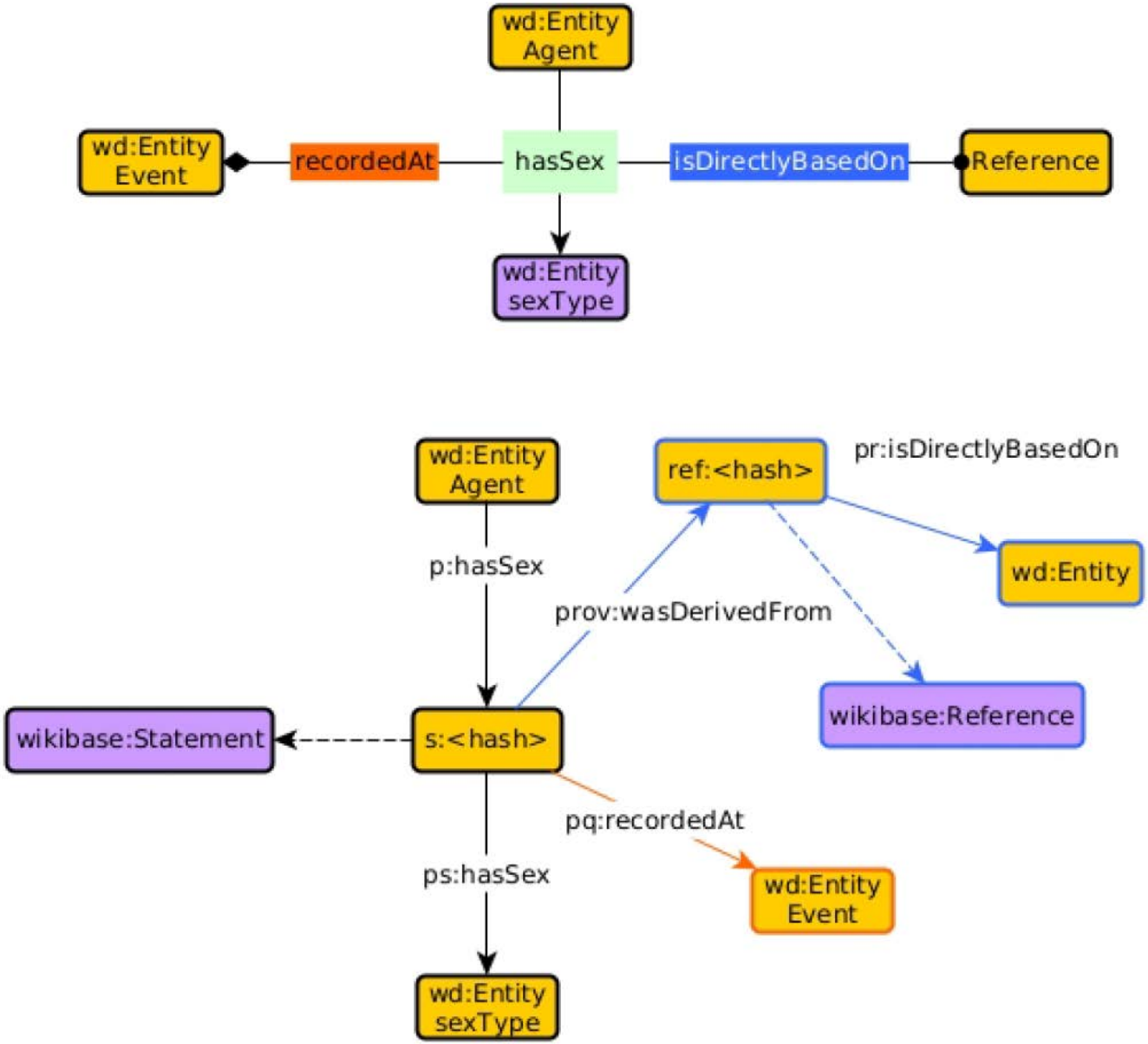
Wikibase ODPs: decimal



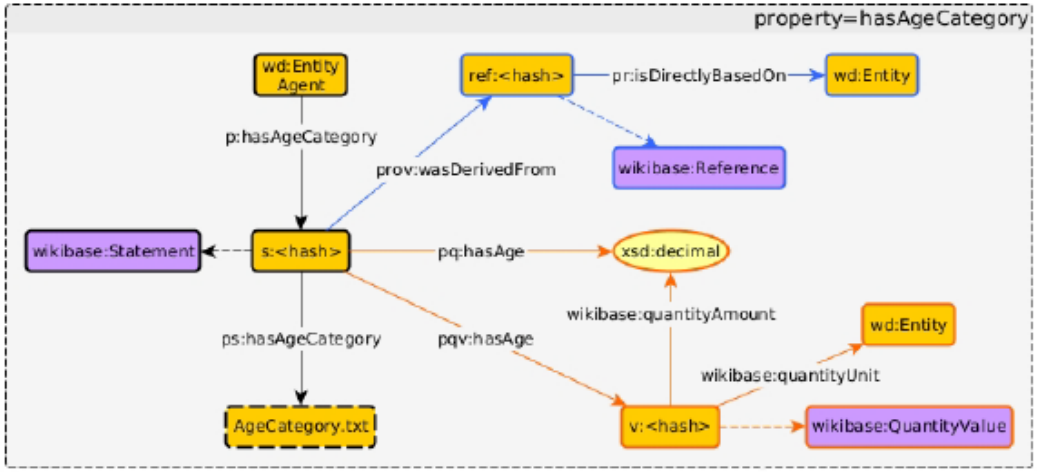
Wikibase ODPs: datatype properties

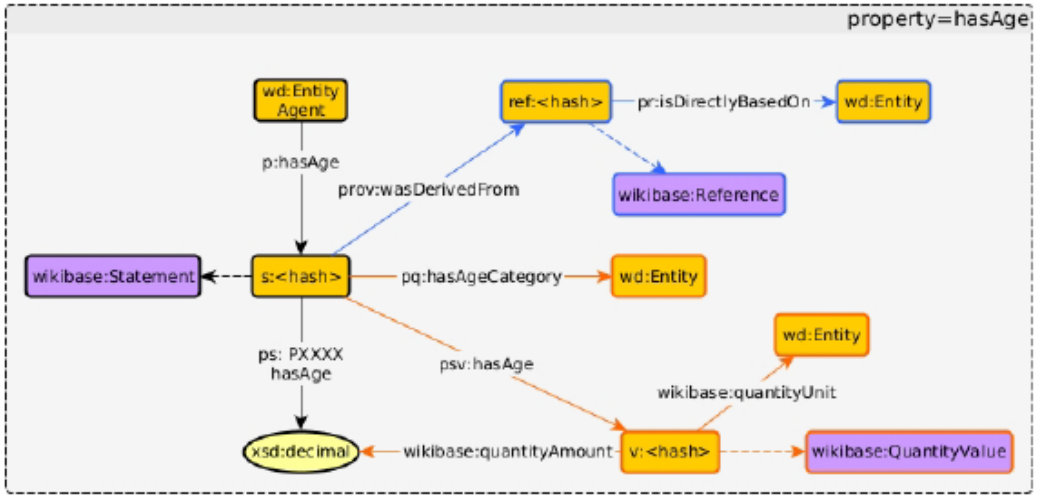


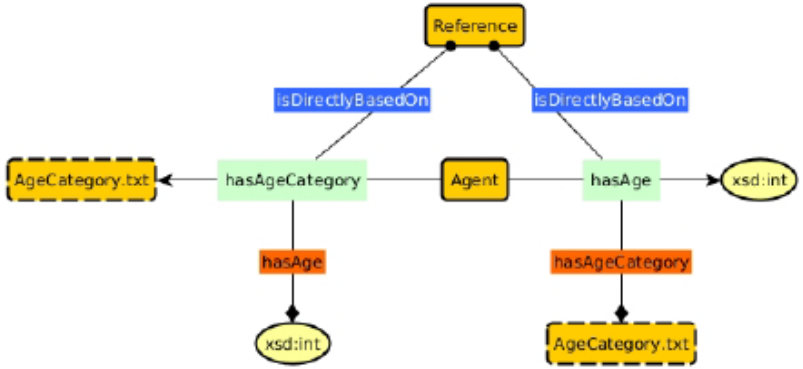
Enslaved redone



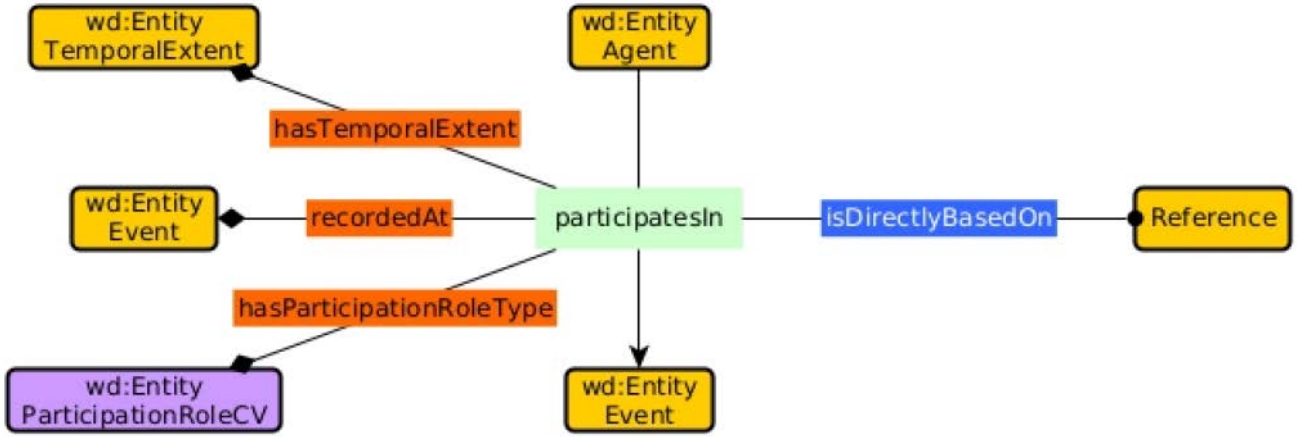
Enslaved redone

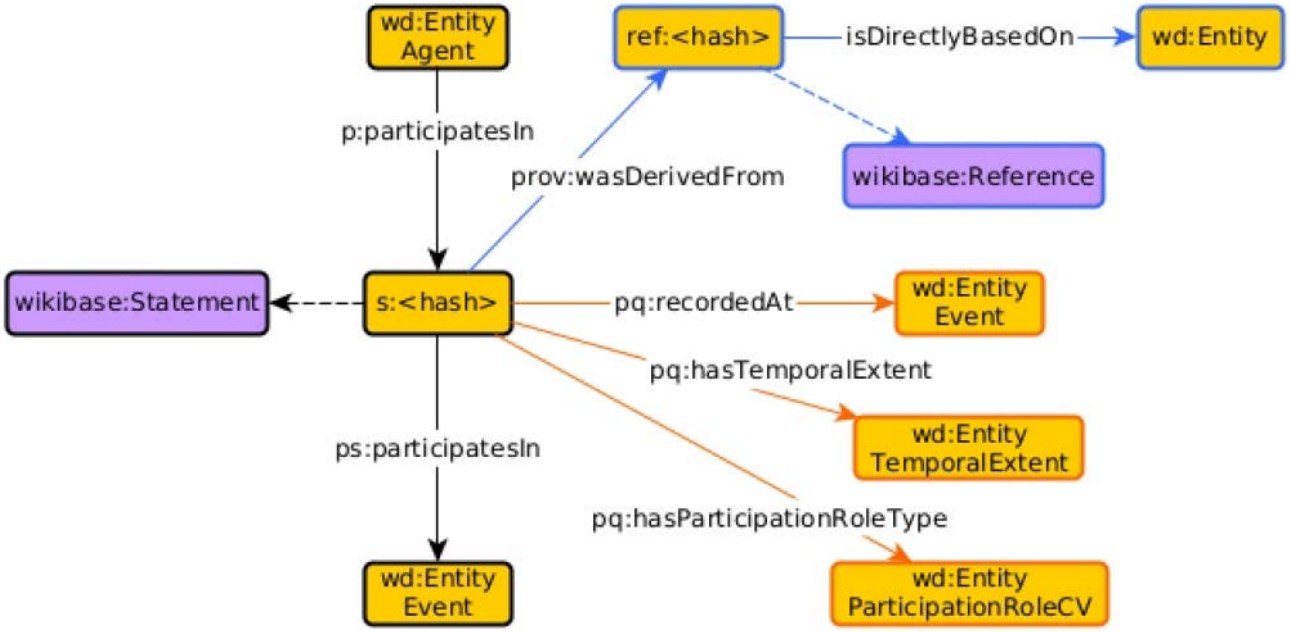




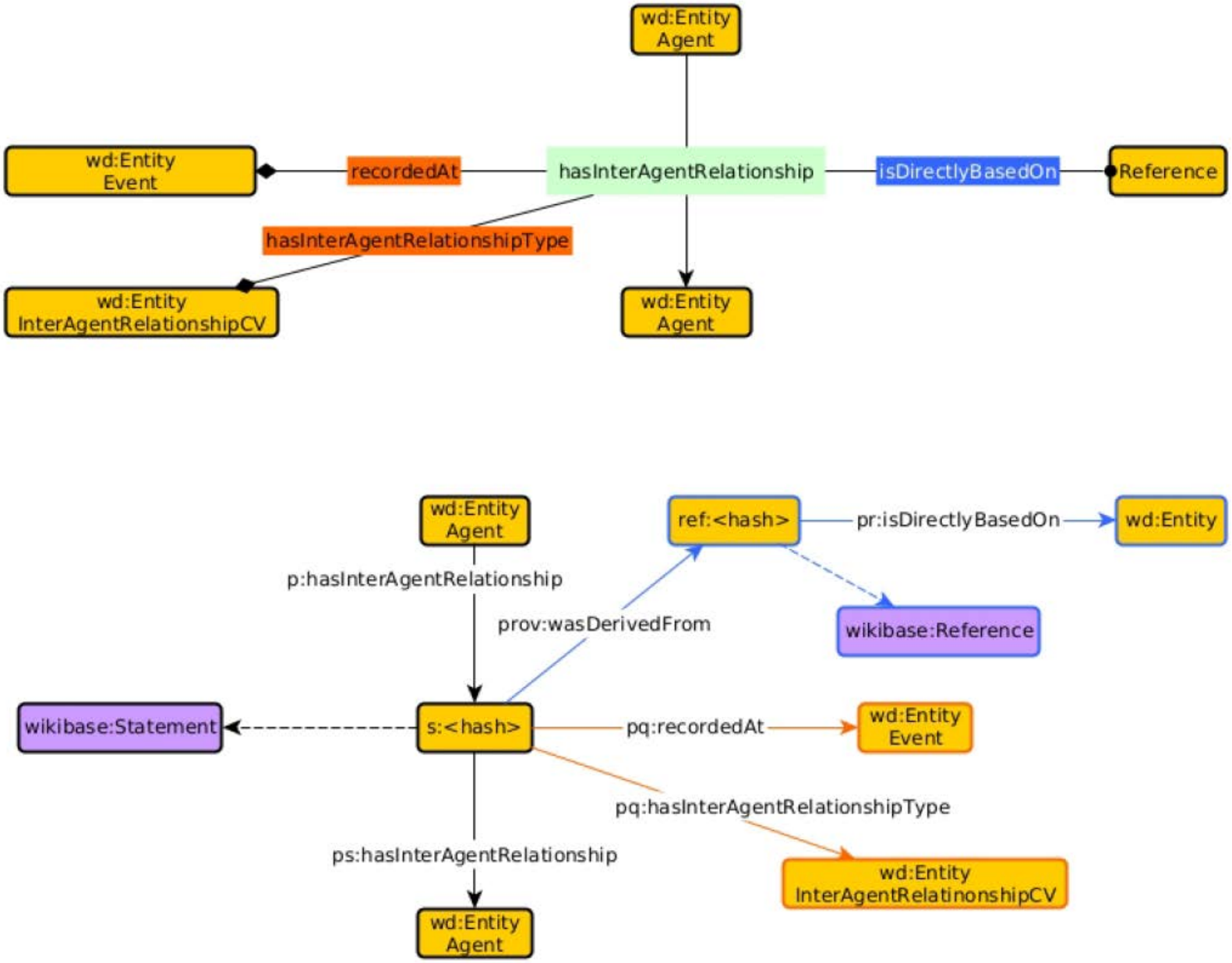


Enslaved redone

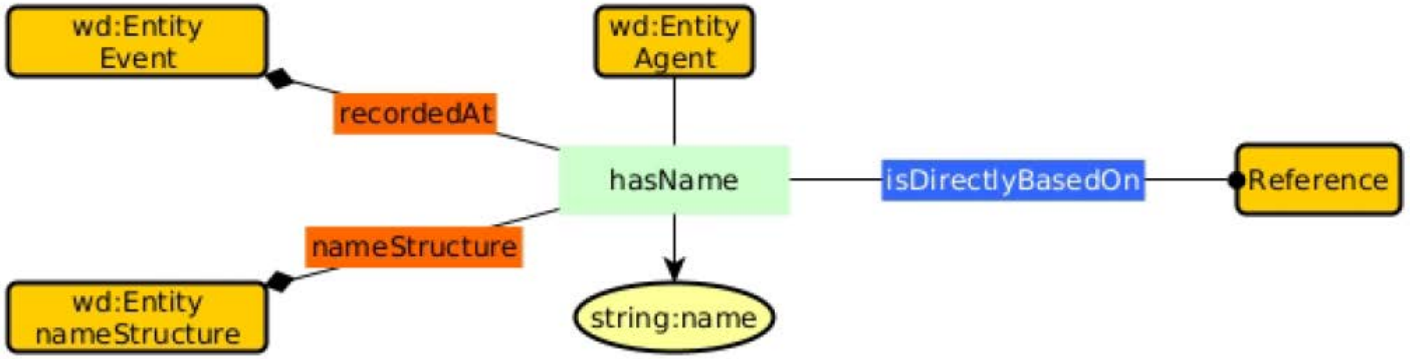


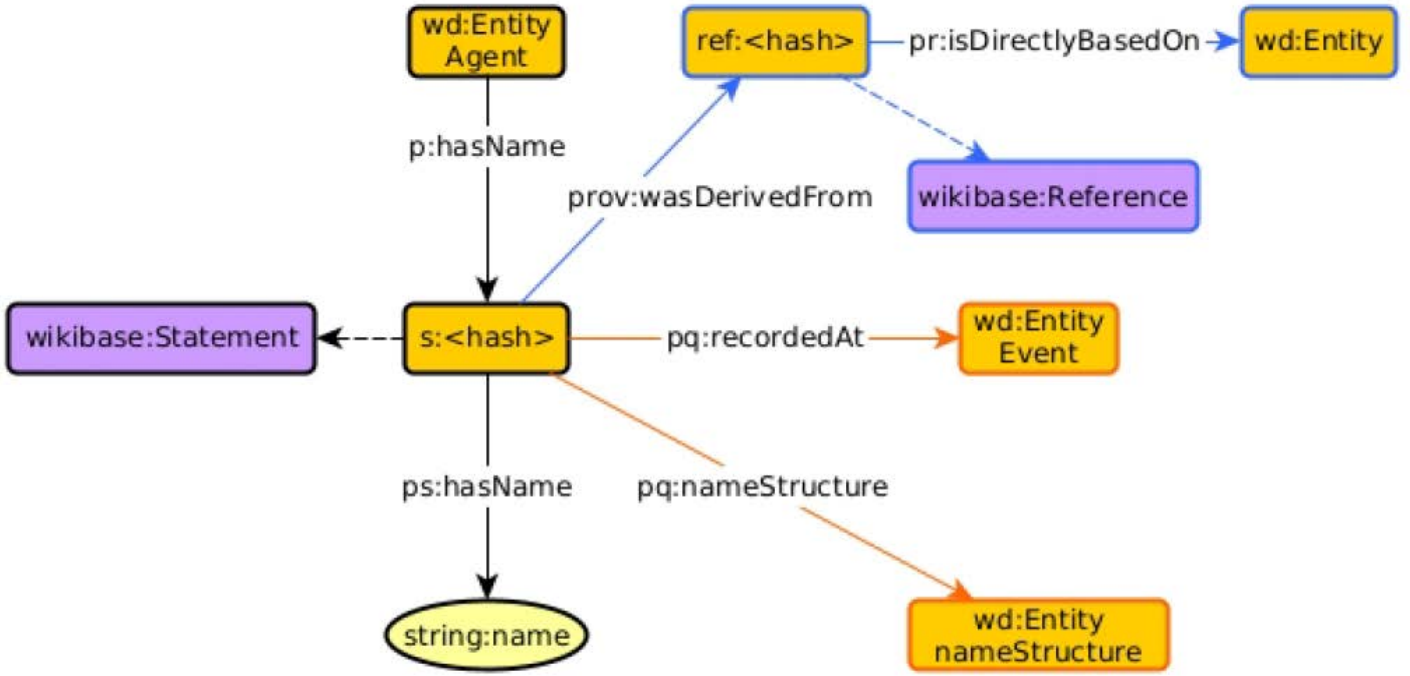


Enslaved redone



Enslaved redone





Forthcoming?

Certainly:

- **Axiomatization, ShEx**

Possibly:

- **a variety of tool support**

Contribution Summary

- **With the patterns we now have, we could have saved a lot of time earlier.**

- **Patterns and approach should be adoptable by anybody who is interested in bridging**
  - **traditional ontology modeling and**
  - **use of Wikibase as graph platform.**

- **Perhaps it can even help stimulate discussions on recommended patterns for Wikibase?**

Thanks!

Collaborators:

Cogan Shimizu, Andrew Eells, Lu Zhou
DaSeLab, Kansas State University
Seila Gonzalez, Alicia Sheill, Catherine Foley, Dean Rehberger
MATRIX, Michigan State University

References

- **Pascal Hitzler, Semantic Web: A Review of the Field. Communications of the ACM 64 (2), 76-82, 2021.**
- **Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P., and Rudolph, S. (Eds.). *OWL 2 Web Ontology Language: Primer (2nd Ed.)*. W3C Recommendation 11 (Dec. 2012); http://www.w3.org/TR/owl2-primer/.**
- **Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, Foundations of Semantic Web Technologies. Textbooks in Computing, Chapman and Hall/CRC Press, 2010.**
- **Vrandecic, D. and Krötzsch, M. Wikidata: A free collaborative knowledgebase. *Commun. ACM 57*, 10 (Oct. 2014), 78–85.**

References

- Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., and Taylor, J. Industry-scale knowledge graphs: lessons and challenges. *Commun. ACM 62*, 8 (Aug. 2019), 36–43.
- Cogan Shimizu, Pascal Hitzler, Quinn Hirt, Dean Rehberger, Seila Gonzalez Estrecha, Catherine Foley, Alicia M. Sheill, Walter Hawthorne, Jeff Mixter, Ethan Watrall, Ryan Carty, Duncan Tarr: The Enslaved ontology: Peoples of the historic slave trade. J. Web Semant. 63: 100567 (2020)
- C. Shimizu, K. Hammar, P. Hitzler, Modular Ontology Modeling, 2021, under review.
- A. Eells, Shimizu, C., Zhou, L., Hitzler, P., Estrecha, S. Gonzalez, and Rehberger, D., Aligning Patterns to the Wikibase Model, Workshop on Ontology Design and Patterns. 2021.
- Lu Zhou, Cogan Shimizu, Pascal Hitzler, Alicia M. Sheill, Seila Gonzalez Estrecha, Catherine Foley, Duncan Tarr, Dean Rehberger: The Enslaved Dataset: A Real-world Complex Ontology Alignment Benchmark using Wikibase. CIKM 2020: 3197-3204