A Complex Alignment Benchmark: GeoLink Dataset

Lu Zhou, Michelle Cheatham, Adila Krisnadhi, Pascal Hitzler
Data Semantics Laboratory
Wright State University
ISWC 2018 Resources Track, Date: 10/10/2018

WRIGHT STATE UNIVERSITY
Department of Computer Science and Engineering
Data Semantics Laboratory
Dayton OH, 45324, USA
https://dase.cs.wright.edu

Ontology Alignment

DaSe Lab

Ontology alignment (OA) is about mapping one ontology onto another.
Think data integration.

Until recently, OA was heavily pre-occupied with just finding ways to map classes onto classes.

[Cheatham & H, ISWC 2013] demonstrated that the state of the art does hardly better than clever string matching.

Now there is growing awareness that we really need to study complex mapping rules.

What “rules” do we need?

DaSe Lab

Even for one-one class-to-class mappings it’s difficult to come up with a good benchmark, as humans disagree what the correct mapping would be.

This is even more tricky when it’s about complex mapping rules.

As a consequence, we have not yet really understood what “types” of mapping rules are in fact needed.

Creating synthetic benchmarks (to drive OA research on complex alignment) runs the risk of falling far from reality.

The GeoLink Setting

DaSe Lab

GeoLink: Completed NSF EarthCube Building Blocks project concerning the integration of ocean science data repositories.

Integration was done via an overarching ontology (schema), which was made for the purpose, the GeoLink Modular Ontology (GMO) [Krisnadhi+, ISWC 2015].

GMO followed ontology modeling best practices using ontology design patterns. It is a rather complicated ontology.

GMO and GBO

To make things easier for data providers and application developers, we developed a simplified version of the GMO, called the GeoLink Base Ontology (GBO).

GBO is smaller and straightforward.

Since it was constructed from the GMO, we know exactly how they align.

In fact, by design, the alignment was to be used to populate the GMO through the GBO.

GeoLink benchmark

We thus have:

- GMO – complex ontology
- GBO – simplified version
- A complex alignment between them which is by design *the* reference alignment.

We did not initially plan to propose this as a complex OA benchmark.

However it turns out that this natural alignment contains mappings which look rather odd from the perspective of the OA field's preconceptions.

GeoLink benchmark

The original reference alignment has 111 axioms.

- Some of them are not expressible in OWL.
- Some of them are not expressible in EDOAL (tool-supported OA language).
- Some of them were left out of this years OAEI benchmark as they were deemed too much out of scope for the current state of investigations.

Consequently, the OAEI benchmark has only 66 axioms.

Dataset

- Size of the ontologies

Ontology	Classes	Object Properties	Data Properties
GeoLink Base Ontology	40	149	49
GeoLink Modular Ontology	156	124	46

Dataset

DaSe Lab

12 types of correspondences

Pattern	Occurrences	Category
Class Equivalence	10	1:1
Class Subsumption	2	1:1
Property Equivalence	7	1:1
Property Equivalence Inverse	5	1:1
Class Typecasting Equivalence	4	1:n
Class Typecasting Subsumption	1	1:n
Property Typecasting Subsumption	5	1:n
Property Typecasting Subsumption Inverse	5	1:n
Typed Property Chain Equivalence	26	m:n
Typed Property Chain Equivalence Inverse	17	m:n
Typed Property Chain Subsumption	17	m:n
Typed Property Chain Subsumption Inverse	12	m:n

Simple

Complex

WRIGHT STATE UNIVERSITY
Department of Computer Science and Engineering
Data Semantics Laboratory
Dayton OH, 45324, USA
https://dase.cs.wright.edu

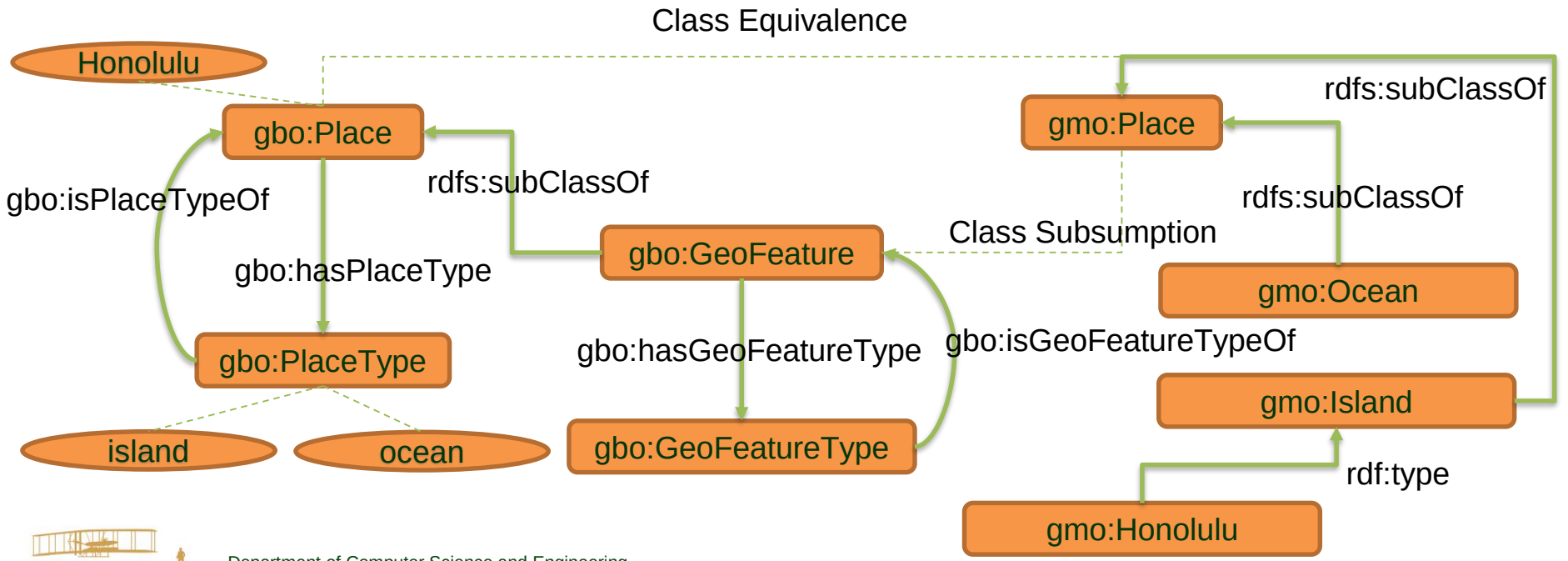
Dataset

- 1:1 Simple Correspondences
  - Class Equivalence → Formal Pattern: $C_1(x) \leftrightarrow C_2(x)$ Example: $\mathsf{Award}(x) \leftrightarrow \mathsf{FundingAward}(x)$
  - Class Subsumption (next slide) → Formal Pattern: $C_1(x) \rightarrow C_2(x)$ Example: $\mathsf{GeoFeature}(x) \rightarrow \mathsf{Place}(x)$
  - Property Equivalence → Formal Pattern: $p_1(x, y) \leftrightarrow p_2(x, y)$ Example: $\mathsf{hasAward}(x, y) \leftrightarrow \mathsf{fundedBy}(x, y)$
  - Property Equivalence Inverse → Formal Pattern: $p_1(x, y) \leftrightarrow p_2(y, x)$ Example: $\mathsf{isAwardOf}(x, y) \leftrightarrow \mathsf{fundedBy}(y, x)$

WRIGHT STATE UNIVERSITY
Department of Computer Science and Engineering
Data Semantics Laboratory
Dayton OH, 45324, USA
https://dase.cs.wright.edu

Dataset

- 1:n Complex Correspondences
  - Class Typecasting Equivalence → Formal Pattern: $C_1(x) \leftrightarrow \text{rdfs:subclassOf}(x, C_2)$; Example: $\text{PlaceType}(x) \leftrightarrow \text{rdfs:subclassOf}(x, \text{Place})$
  - Class Typecasting Subsumption → Formal Pattern: $C_1(x) \rightarrow \text{rdfs:subclassOf}(x, C_2)$; Example: $\text{GeoFeatureType}(x) \rightarrow \text{rdfs:subclassOf}(x, \text{Place})$
  - Property Typecasting Subsumption → Formal Pattern: $p_1(x, y) \rightarrow \text{rdf:type}(x, y) \wedge \text{rdfs:subclassOf}(y, C_2)$; Example: $\text{hasPlaceType}(x, y) \rightarrow \text{rdf:type}(x, y) \wedge \text{rdfs:subclassOf}(y, \text{Place})$
  - Property Typecasting Subsumption Inverse → Formal Pattern: $p_1(x, y) \rightarrow \text{rdf:type}(y, x) \wedge \text{rdfs:subclassOf}(x, C_2)$; Example: $\text{isGeoFeatureTypeOf}(x, y) \rightarrow \text{rdf:type}(y, x) \wedge \text{rdfs:subclassOf}(x, \text{Place})$



WRIGHT STATE UNIVERSITY
Department of Computer Science and Engineering
Data Semantics Laboratory
Dayton OH, 45324, USA
https://dase.cs.wright.edu

Dataset

- m:n Complex Correspondences
  - Typed Property Chain Equivalence
  - Typed Property Chain Equivalence Inverse

Formal Pattern:

$$D_1(x_1) \wedge r(x_1, x_{n+1}) \wedge D_2(x_{n+1}) \leftrightarrow C_1(x_1) \wedge p_1(x_1, x_2) \wedge C_2(x_2) \wedge \cdots \wedge p_n(x_n, x_{n+1}) \wedge C_{n+1}(x_{n+1})$$

Example[11]:

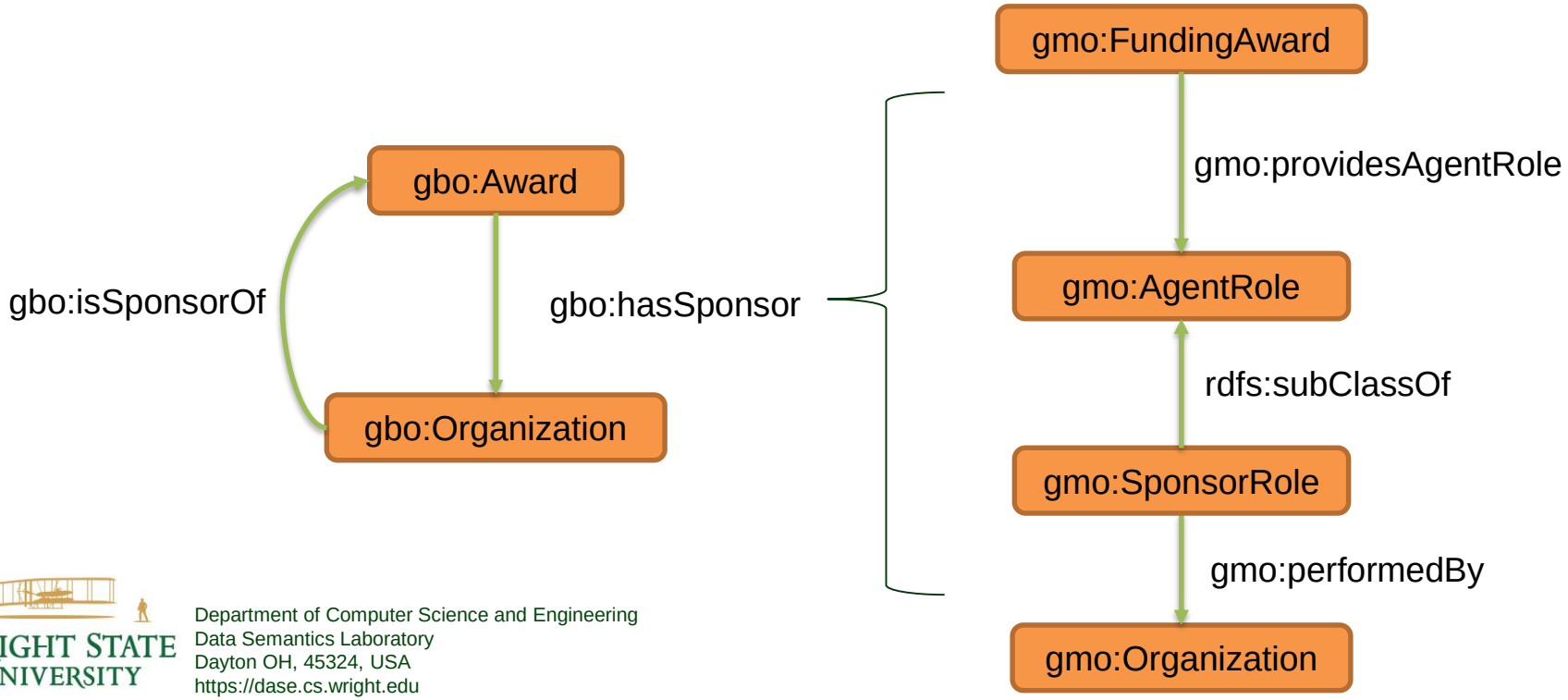
$$\text{Award}(x) \wedge \text{hasSponsor}(x, z) \leftrightarrow \text{FundingAward}(x) \wedge \text{providesAgentRole}(x, y) \wedge \text{SponsorRole}(y) \wedge \text{performedBy}(y, z)$$

Formal Pattern:

$$D_1(x_1) \wedge r(x_1, x_{n+1}) \wedge D_2(x_{n+1}) \leftrightarrow C_1(x_{n+1}) \wedge p_1(x_{n+1}, x_n) \wedge C_2(x_n) \wedge \cdots \wedge p_n(x_2, x_1) \wedge C_{n+1}(x_1)$$

Example:

$$\text{Award}(z) \wedge \text{isSponsorOf}(x, z) \leftrightarrow \text{FundingAward}(z) \wedge \text{provideAgentRole}(z, y) \wedge \text{SponsorRole}(y) \wedge \text{performedBy}(y, x)$$



WRIGHT STATE UNIVERSITY
Department of Computer Science and Engineering
Data Semantics Laboratory
Dayton OH, 45324, USA
https://dase.cs.wright.edu

Dataset

- m:n Complex Correspondences
  - Typed Property Chain Subsumption
  - Typed Property Chain Subsumption Inverse

Formal Pattern:

$$D_1(x_1) \wedge r(x_1, x_{n+1}) \wedge D_2(x_{n+1}) \rightarrow C_1(x_1) \wedge p_1(x_1, x_2) \wedge C_2(x_2) \wedge \cdots \wedge p_n(x_n, x_{n+1}) \wedge C_{n+1}(x_{n+1})$$

Example:

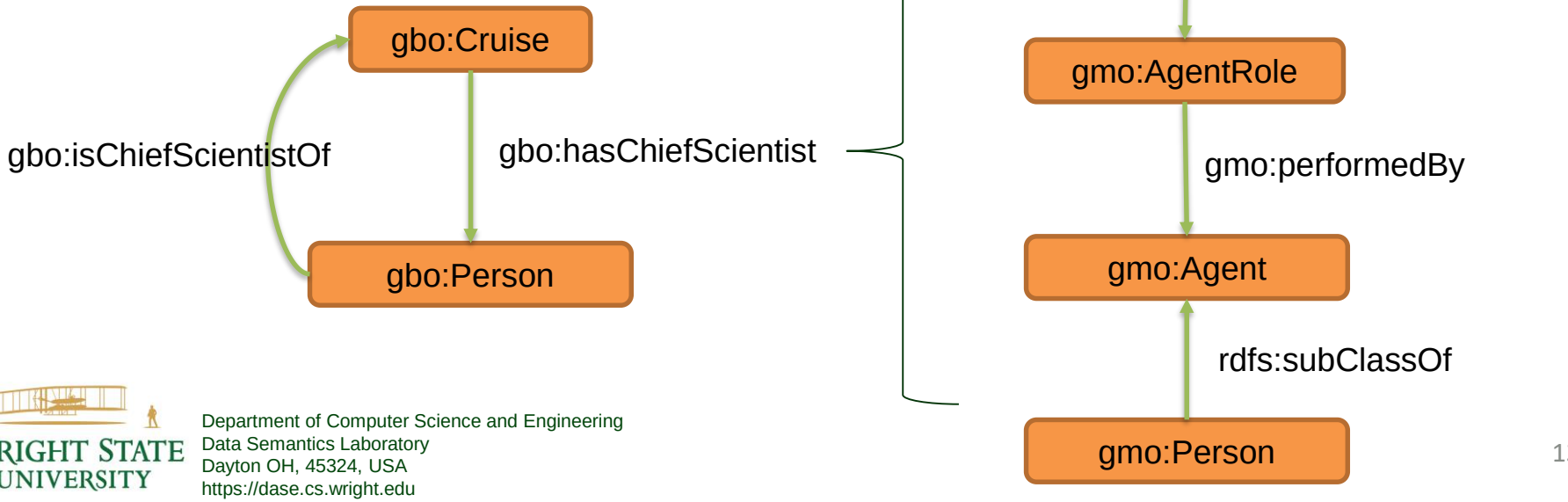
$$\text{Cruise}(x) \wedge \text{hasChiefScientist}(x, z) \rightarrow \text{Cruise}(x) \wedge \text{providesAgentRole}(x, y) \wedge \text{AgentRole}(y) \wedge \text{performedBy}(y, z)$$

Formal Pattern:

$$D_1(x_1) \wedge r(x_1, x_{n+1}) \wedge D_2(x_{n+1}) \rightarrow C_1(x_{n+1}) \wedge p_1(x_{n+1}, x_n) \wedge C_2(x_n) \wedge \cdots \wedge p_n(x_2, x_1) \wedge C_{n+1}(x_1)$$

Example:

$$\text{Cruise}(z) \wedge \text{isChiefScientistOf}(x, z) \rightarrow \text{Cruise}(z) \wedge \text{providesAgentRole}(z, y) \wedge \text{AgentRole}(y) \wedge \text{performedBy}(y, x)$$



WRIGHT STATE UNIVERSITY
Department of Computer Science and Engineering
Data Semantics Laboratory
Dayton OH, 45324, USA
https://dase.cs.wright.edu

Dataset

- Representation
  - EDOAL
    - More expressive than standard alignment format
  - Rule Syntax
    - Easy for humans to understand

- Data publication
  - Download link (CC-BY): http://doi.org/10.6084/m9.figshare.5907172
  - OAEI complex track website

- HermiT reasoner has been applied to the ontologies independently to check the satisfiability

Conclusion and Future Work

DaSe Lab

- Conclusion
  - Presentation of two ontologies from real-world case to support data representation, sharing, integration, and discovery.
  - Creation of an alignment between these two ontologies includes simple and complex alignment.
  - Publication of the benchmark alignment in both rule syntax and EDOAL format at a persistent URL under a CC-BY license.

- Future Work
  - Complex track in OAEI (ongoing)
  - Create an automated complex alignment to detect the complex correspondences and evaluate the performance.

Thank you, Questions?

WRIGHT STATE UNIVERSITY
Department of Computer Science and Engineering
Data Semantics Laboratory
Dayton OH, 45324, USA
https://dase.cs.wright.edu