Matching life sciences ontologies in the Ontology Alignment Evaluation Initiative (OAEI)

Ernesto Jiménez-Ruiz, Thomas Liener and Ian Harrow.

† City, University of London, UK. ‡ Pistoia Alliance, USA

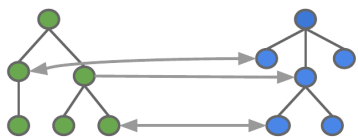
Outline

- Ontology Alignment
- Ontology Alignment Evaluation Initiative
- Pistoia Alliance

Ontology Alignment

Ontology alignment: Nomenclature

- Knowledge graph alignment as a type of **ontology alignment** or **ontology matching**.
- **To match or align or map**: the process that produces an alignment or mapping set.
- **An alignment or mapping set**: the final output of matching or aligning.
- **A mapping or match**: a single link between related entities; also called a cross-reference.



Ontology alignment: definition (atomic mappings)

- Basic definition in the OM community.
- An **ontology alignment** $\mathcal{M}$ (or $\mathcal{A}$) is a set of tuples $\langle e_1, e_2, n, \rho \rangle$
  - e_1, e_2 are **entities** in the input ontologies ($e_1 \in \mathcal{O}_1$ and $e_2 \in \mathcal{O}_2$)
  - n a **confidence** value between 0 and 1
  - ρ is the **semantic relationship** between e_1 and e_2
    - OM: subsumption, equivalence, disjointness
    - LS: broadMatch, narrowMatch, closeMath, relatedMatch, exactMatch.

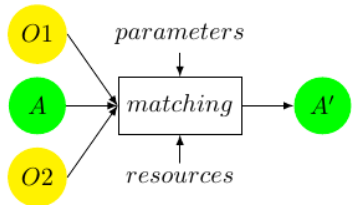
P. Shvaiko, J. Euzenat. Ontology matching: state of the art and future challenges. IEEE Transactions on Knowledge and Data Engineering 2013

Ontology alignment: (exchange) formats

- **RDF Alignment** format (OM Community)
- A Simple Standard for Sharing Ontology Mappings (**SSSOM**)
- **OWL 2 axioms**
  - Where the semantic relationship ρ is one of $\{\equiv, \sqsubseteq, \sqsupseteq, \perp\}$
  - Confidence values n are represented as axiom annotations
  - Enables OWL 2 reasoning.
  - $\mathcal{O}_1{:}Joint\_structure \equiv \mathcal{O}_2{:}Joint$

Alignment systems

- Given two input ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$ **generate an alignment** $\mathcal{A}'$ as output.
- In addition a system can get as input a **partial alignment** $\mathcal{A}$, **matching parameters** and **external resources**.



Ontology Alignment Evaluation Initiative (OAEI)

Ontology Alignment Evaluation Initiative (OAEI)

- **Annual Campaign** since 2004: `http://oaei.ontologymatching.org/`
- **De facto benchmark** for the OM community and driving force for tool improvement
- Collocated with the **Ontology Matching workshop** and the **International Semantic Web Conference**
- **Driven by academia**
- **Supported by industry** (*e.g.*, IBM research, Pistoia Alliance, SIRIUS)

Virtual workshop and conference: `http://om2021.ontologymatching.org/` & `https://iswc2021.semanticweb.org/`

OAEI Objectives

Common tasks and framework for the **systematic evaluation** of ontology alignment systems.

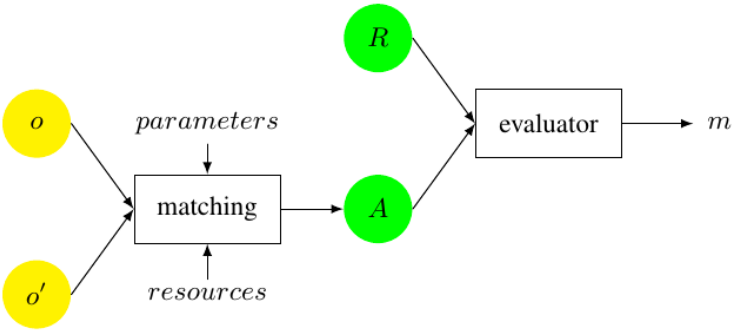
- Assessing **strengths** and **weaknesses** of alignment/matching systems
- **Comparing** performance of techniques
- Increasing **communication** among algorithm developers
- Helping **improve** the work on **ontology alignment**.
- **Improving evaluation** techniques

OAEI schedule

- **Preparation**: June 1st—July 15th (datasets ready)
- **Execution**:
  July 31st (participants register their tools)
  August 31st (participants submit final systems)
- **Evaluation**: September–October
- **OM workshop** (ISWC conference): October/November
- **Closing**: November/December

OPEN COMMUNITY: CALL FOR NEW ORGANISERS/DATASETS/TASKS OPEN ALL YEAR ROUND!!

OAEI evaluation platform



MELT (since 2021): `http://oaei.ontologymatching.org/2021/melt/`

OAEI metrics

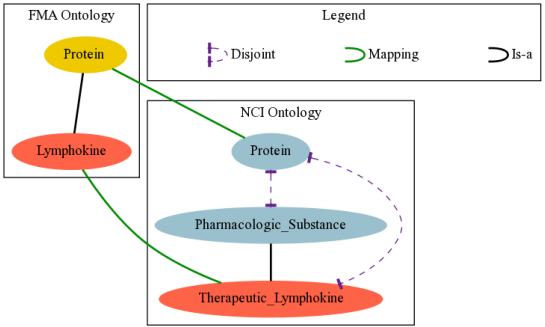
- **Precision** and **recall** wrt reference alignment or gold standard $|\mathcal{R}|$
  - Precision (Pre) = $|\mathcal{A} \cap \mathcal{R}|/|\mathcal{A}|$
  - Recall (Rec) = $|\mathcal{A} \cap \mathcal{R}|/|\mathcal{R}|$
  - F-score (F) = $(2 \times \text{Pre} \times \text{Rec})/(\text{Pre} + \text{Rec})$.
- **Logical errors** of $\mathcal{A}$ wrt $\mathcal{O}_1$ and $\mathcal{O}_2$.
- Computation **times** are also considered.

OAEI metrics: logical errors

The integration of different ontologies via (OWL 2) mappings ($\mathcal{O}_1 \cup \mathcal{O}_2 \cup \mathcal{A}$) can cause **unsatisfiabilities**.

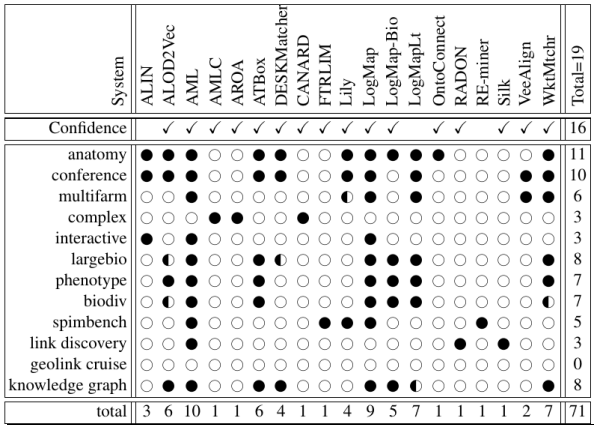
Possible solutions:

- Repair/remove mappings.
- Modify ontologies
- Be aware of the logical incompatibilities.



Ernesto Jiménez-Ruiz et al. Evaluating Mapping Repair Systems with Large Biomedical Ontologies. Description Logics 2013
Ernesto Jiménez-Ruiz et al. Logic-based assessment of the compatibility of UMLS ontology sources. J. Biomed. Semant. 2011
Daniel Faria, Ernesto Jiménez-Ruiz, et al. Towards Annotating Potential Incoherences in BioPortal Mappings. ISWC 2014

OAEI 2020: summary of tasks and participants



System	ALIN	ALOD2Vec	AML	AMLC	AROA	ATBox	DESKMatcher	CANARD	FTRLIM	Lily	LogMap	LogMap-Bio	LogMapLt	OntoConnect	RADON	RE-miner	Silk	VeeAlign	WtkMtchr	Total=19
Confidence		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	16
anatomy	●	●	●	○	○	●	●	○	○	●	●	●	●	●	○	○	○	○	●	11
conference	●	●	●	○	○	●	●	○	○	●	●	○	●	○	○	○	○	●	●	10
multifarm	○	○	●	○	○	○	○	○	○	◐	●	○	●	○	○	○	○	●	●	6
complex	○	○	○	●	●	○	○	●	○	○	○	○	○	○	○	○	○	○	○	3
interactive	●	○	●	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	3
largebio	○	◐	●	○	○	●	◐	○	○	○	●	●	●	○	○	○	○	○	●	8
phenotype	○	●	●	○	○	●	○	○	○	○	●	●	●	○	○	○	○	○	●	7
biodiv	○	◐	●	○	○	●	○	○	○	○	●	●	●	○	○	○	○	○	◐	7
spimbench	○	○	●	○	○	○	○	○	●	●	●	○	○	○	○	●	○	○	○	5
link discovery	○	○	●	○	○	○	○	○	○	○	○	○	○	○	●	○	●	○	○	3
geolink cruise	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	0
knowledge graph	○	●	●	○	○	●	●	○	○	○	●	●	◐	○	○	○	○	○	●	8
total	3	6	10	1	1	6	4	1	1	4	9	5	7	1	1	1	1	2	7	71

OAEI 2020:

- 19 systems
- 12 Tracks
- 4 Bio-related tracks

OAEI Bio tracks (i): Anatomy

- Running since 2007.
- **Adult Mouse Anatomy** (2744 classes) vs human anatomy portion of **NCI Thesaurus** (3304 classes).
- Manually curated reference alignment.

Zlatan Dragisic et al. Experiences from the anatomy track in the ontology alignment evaluation initiative. J Biomedical Semantics 2017

OAEI Bio tracks (ii): Largebio

- Running since 2012
- 3 very large and semantically rich ontologies;
- UMLS as the basis for the reference alignments;

6 tasks	**FMA-NCI**	**FMA-SnoMed**	**SnoMed-NCI**
small	3,696 6,488	10,157 13,412	51,128 23,958
large	78,989 66,724	78,989 122,464	122,464 66,724
reference	3,024	9,008	18,844

OAEI Bio tracks (iii): Biodiversity

- Running since 2018
- Tasks:
  - Environment Ontology (**ENVO**) vs. Semantic Web for Earth and Environment Technology Ontology (**SWEET**)
  - Flora Phenotype Ontology (**FLOPO**) vs. Plant Trait Ontology (**PTO**)

Naouel Karam, et al. Matching biodiversity and ecology ontologies: challenges and evaluation results. KER 2020

OAEI Bio tracks (iv): Disease-Phenotype

- Running since 2016 in collaboration with the **Pistoia Alliance**.
- Tasks:
  - Human Phenotype Ontology (**HPO**) vs. Mammalian Phenotype Ontology (**MP**)
  - Human Disease Ontology (**DOID**) vs. Orphanet and Rare Diseases Ontology (**ORDO**)
- Evaluation with 3-vote consensus alignment from participating tools, grouped by family, since 2016

Ian Harrow, Ernesto Jiménez-Ruiz, et al. Matching disease and phenotype ontologies in the ontology alignment evaluation initiative. J Biomedical Semantics 2017

OAEI Systems

LogMap

- Open-source Java maven project:
  `https://github.com/ernestojimenezruiz/logmap-matcher`
- (Proof-of-concept) web interface
  `http://krrwebtools.cs.ox.ac.uk/logmap/`

AML

- Open-source Java tool with graphical user interface:
- `https://github.com/AgreementMakerLight/AML-Project`

Other systems: `http://oaei.ontologymatching.org/`

OAEI Challenges

- ✓ Large ontology size
- ✓ Rich and complex vocabularies
- ✓ Use of background knowledge

OAEI Challenges

- ✓ Large ontology size
- ✓ Rich and complex vocabularies
- ✓ Use of background knowledge
- ✓ Different modelling views leading to logical errors
- ✓ User involvement

OAEI Challenges

- ✓ Large ontology size
- ✓ Rich and complex vocabularies
- ✓ Use of background knowledge
- ✓ Different modelling views leading to logical errors
- ✓ User involvement
- ✗ Need for complex mappings beyond atomic equivalence/subsumption
- ✗ Combination with ML techniques
- ✗ Better connection with real-world needs.

Pistoia Alliance

What is the Pistoia Alliance

Pistoia Alliance is a not-for-profit members' organization working to lower barriers to innovation in life science and healthcare R&D through pre-competitive collaboration (`https://www.pistoiaalliance.org`)

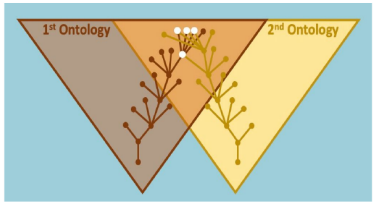
Ontology mapping project. Mappings in the life science domain (Phenotype disease and laboratory analytics)

Pistoia's Ontology mapping project

- **Paxo**: Lightweight algorithm using EBI's Ontology Lookup service (OLS) and Ontology Xref Service (OxO) as backbone
- **OAEI Phenotype track** since 2016.
- **Validation**: OAEI consensus mappings as silver standard and manual validation of subsets.

I. Harrow et al. Ontology Matching for theLaboratory Analytics Domain. OM workshop 2020.
I. Harrow et al. Matching Disease and Phenotype Ontologies in the Ontology Alignment Evaluation Initiative. J. Biomedical Semantics 2018
I. Harrow et al. Ontology mapping for semantically enabled applications. Drug Discovery Today, 2019

Pistoia Alliance partners and collaborators



PM: Ian Harrow

Current funders, partners & collaborators

The ontology mapping project was merged into Pistoia's *FAIR implementation project*, managed by Thomas.Liener@pistoiaalliance.org

https://www.pistoiaalliance.org/projects/current-projects/fair-implementation/

Acknowledgements

Acknowledgements

- Co-organisers of the SWAT4(HC)LS 2019 tutorial on **Ontology Matching in the Biomedical Domain**.
  - `https://tinyurl.com/tutorial-ontology-alignment`
- Ontology Matching workshop and OAEI organisers.
  `http://ontologymatching.org/`
- Pistoia Alliance.
  `https://www.pistoiaalliance.org/projects/current-projects/ontologies-mapping/`
  `https://www.pistoiaalliance.org/projects/current-projects/fair-implementation/`