Introducing the Disease and Phenotype OAEI Track*

Ian Harrow[1], Ernesto Jimenez-Ruiz[2], Andrea Splendiani[1],
Martin Romacker[1], Stefan Negru[1], Peter Woollard[1], Scott Markel[1],
Yasmin Alam-Faruque[1], Martin Koch[1], Erfan Younesi[1] and James Malone[1]

[1] Pistoia Alliance Ontologies Mapping Project, Pistoia Alliance Inc. USA
[2] Department of Computer Science, University of Oxford, UK

1 Introduction

The Pistoia Alliance Ontologies Mapping project[1] was set up to find or create better tools and services for mapping between ontologies (including controlled vocabularies) in the same domain and to establish best practices for ontology management in the Life Sciences. The project has developed a formal process to define and submit a request for information (RFI) from existing ontologies mapping tool providers to enable their evaluation.[2] A critical component of any Ontologies Mapping tool is the embedded ontology matching algorithm, therefore the project is supporting their development and evaluation through sponsorship and organisation of the new *Disease and Phenotype* track for the OAEI campaign[3] [1] which is described in this paper.

2 Datasets

The *Disease and Phenotype* track[4] comprises two tasks that will involve the pairwise alignment of the HPO, MP, DOID and ORDO ontologies (Table 1 shows the metrics of these ontologies):

- **Task 1:** matching of the Human Phenotype Ontology (HPO) to the Mammalian Phenotype Ontology (MP).
- **Task 2:** matching of the Human Disease Ontology (DOID) to the Orphanet and Rare Diseases Ontology (ORDO).

The first task is important for translational science where HPO includes inherited diseases and MP originated from rodents as a model mammalian organism for many laboratory studies, including gene knock out. The second task includes representation of rare human diseases in both ontologies which are of fundamental importance for understanding how genetic variation can cause disease. Currently, such mappings are mostly curated by bioinformatics and disease experts who would benefit from automation supported by implementation of ontology matching algorithms into their workflows.

We have extracted a "baseline" reference alignments for the track based on the available BioPortal mappings [2] which are considered as a baseline since they are incomplete and may contain errors.

* We have also submitted a 4-pages paper about the Pistoia Alliance Ontologies Mapping Project to the ISWC 2016 posters and demos track.

[1] http://www.pistoiaalliance.org/projects/ontologies-mapping

[2] https://pistoiaalliance.atlassian.net/wiki/display/PUB/Ontologies+Mapping+Resources

[3] http://oaei.ontologymatching.org/2016/

[4] http://oaei.ontologymatching.org/2016/phenotype/description.html

Table 1. Metrics of the track ontologies. Source: NCBI BioPortal on 19th Aug 2016

Ontology	Number of classes	Maximum depth	Avg. number of children
HPO	15,319	15	3
MP	11,720	Undisclosed	Undisclosed
DOID	10,905	12	3
ORDO	13,105	11	16

3 Evaluation process

The evaluation of the Disease and Phenotype Track will be run with support of the SEALS infrastructure.[5] Systems will be evaluated and ranked according to the following criteria:

- Precision and Recall with respect to a voted reference alignment that will be built automatically to generate consensus voting for the outputs of the participating systems.
- Recall with respect to manually generated mappings for three areas (carbohydrate, obesity and breast cancer).
- Manual assessment of a subset of the generated mappings, specially the ones that are not suggested by other systems.
- Performance in other tracks will also be taken into account, especially the OAEI interactive track [3] where the *Disease and Phenotype* dataset is also used.[6]

Additionally, systems able to discover complex logic relations in mappings beyond equivalence and subsumption will also be considered. The evaluation of these mappings will be in parallel to the evaluation of standard equivalence and subsumption mappings. Complex mappings should be provided in OWL 2 format.

Acknowledgements

This work was partially funded by the Pistoia Alliance Ontology Mappings project, the EU project Optique (FP7-ICT-318338), and the EPSRC projects ED3 and DBOnto.

References

1. Cheatham, M., Dragisic, Z., Euzenat, J., Faria, D., Ferrara, A., Flouris, G., Fundulaki, I., Granada, R., Ivanova, V., Jiménez-Ruiz, E., Lambrix, P., Montanelli, S., Pesquita, C., Saveta, T., Shvaiko, P., Solimando, A., dos Santos, C.T., Zamazal, O.: Results of the ontology alignment evaluation initiative 2015. In: Proceedings of the 10th International Workshop on Ontology Matching. (2015) 60–115
2. Fridman Noy, N., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.A.D., Chute, C.G., Musen, M.A.: BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Research **37**(Web-Server-Issue) (2009)
3. Dragisic, Z., Ivanova, V., Lambrix, P., Faria, D., Jimenez-Ruiz, E., Pesquita, C.: User validation in ontology alignment. In: International Semantic Web Conference. (2016)

[5] `http://oaei.ontologymatching.org/2016/seals-eval.html`

[6] `http://oaei.ontologymatching.org/2016/interactive/index.html`