

# Reconciling ontology definitions using the Ontology Pattern Reconciliation Workbench and the DOSDP framework

Nicolas Matentzoglou<sup>1\*</sup> and David Osumi-Sutherland<sup>1</sup>

<sup>1</sup>European Bioinformatics Institute (EMBL-EBI), Hinxton, UK

## ABSTRACT

Many bio-ontologies use formal, logical definitions to automate multiple inheritance classification and drive cross-ontology inference. This requires the use of standardised design patterns: shared patterns of axiomatisation using common reference ontologies. Developing, managing and implementing a suitably consistent set of design patterns can be challenging. In many phenotype ontologies, for example, a large proportion of class terms have formal definitions following a general framework, known as entity/quality (EQ), with common relations and reference ontologies. Despite this, the formal definitions used are often too divergent to drive classification and cross-ontology inference. Here we present software tools for improving and managing formalisation using design patterns. The Ontology Pattern Reconciliation Workbench helps users prioritise patterns for reconciliation between two related ontologies based on the impact pattern reconciliation will have on cross-ontology mapping. An extension to the ontology starter kit provides a practical workflow for developing ontologies using formally specified design patterns.

## 1 BACKGROUND

Pattern-based ontology development is increasingly used to automate and improve classification within ontologies (Osumi-Sutherland *et al.*, 2017; Rocca-Serra *et al.*, 2011). It also has great potential for improving cross-ontology mappings, for example between different species-specific phenotype ontologies. Some degree of standardisation in phenotype ontologies has been achieved through the widespread use of entity/quality-based (EQ) logical definitions (Mungall *et al.*, 2010) such as the Human Phenotype Ontology (HP), Mammalian Phenotype Ontology (MP), Drosophila Phenotype Ontology (Osumi-Sutherland *et al.*, 2013), Fission Yeast Phenotype Ontology (Harris *et al.*, 2013). EQ definitions reference an entity, typically an independent continuant such as ‘finger’ or ‘cell’ or an occurrent such as ‘gastrulation’, and a quality that inheres in that entity, such as ‘long’ or ‘increased pressure’.

While EQ provides a basic framework for standardising phenotype definitions, it is flexible enough that groups using it independently tend to come up with divergent, often incompatible definitions for similar phenotypes. HP, for example, defines ‘unilateral deafness’ as follows:

### Unilateral deafness (HP:0009900)

‘has part’ some (‘lacking processual parts’ and (‘towards’ some ‘sensory perception of sound’) and ‘has modifier’ some ‘abnormal’ and ‘has modifier’ some ‘unilateral’)

The phenotype is defined in terms of a PATO quality ‘lacking processual parts’ and there is no associated entity that bears it. With respect to the resulting classification, this is no problem, because HP:‘unilateral deafness’ is inferred to be a subclass of ‘abnormal ear physiology’, which is defined to ‘**inhere in**’ some ear. (Note that, unless ‘**has part**’ is made functional, it is not the ‘lacking processual parts’ quality that ‘**inheres in**’ the ear.) The corresponding definition in MP is as follows:

### unilateral deafness (MP:0004699)

‘has part’ some (‘unilateral’ and ‘inheres in’ some ‘deafness’ and ‘has modifier’ some ‘abnormal’)

Here, we have the PATO ‘spatial pattern’ quality of ‘unilateral’ being the central quality by which the phenotype is defined. This quality is being held by the entity ‘deafness’ (a phenotype rather than an occurrent or independent continuant).

Not all definitions diverge to such a high degree. An example of a class that has only slightly different definitions, ‘hypertension’, can be seen in Figure 1. In both definitions, ‘increased pressure’ is the abnormal quality (modified by ‘abnormal’ and ‘chronic’) and the relevant entity (the ‘bearer’ of the quality) is ‘blood’.

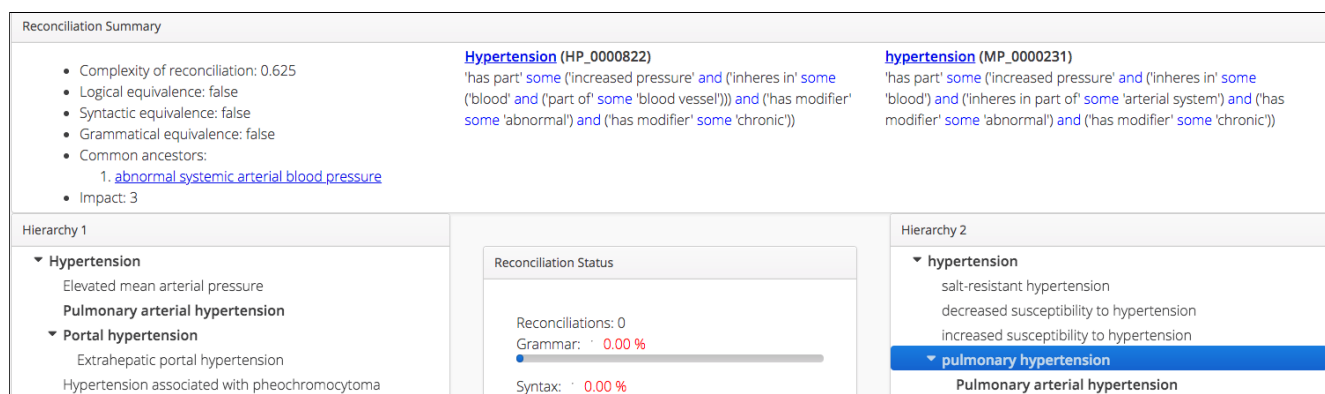
However, rather than using separate clauses for the anatomical entity ‘blood’ and ‘part of’ some ‘arterial system’, the HP definition only defines a single entity: ‘blood’, restricted to be ‘part of’ some ‘blood vessel’. These hypertension definitions may not look significantly different, but they nevertheless result in different inferred classifications. For example, MP:‘hypertension’ is classified as ‘increased systemic arterial blood pressure’, while HP:‘hypertension’ is not. Both, however, are classified as ‘abnormal systemic arterial blood pressure’.

To reach an appropriate level of semantic integration, definitions such as the ones presented here should be reconciled. In the following, we introduce a prototype of the Ontology Pattern Reconciliation Workbench, a tool suite that enables the identification, collection and exploration of reconciliation candidates across multiple ontologies. We also describe extensions to the the ontology starter kit that make it easy for ontology developers to incorporate DOSDP-based definition generation into the general release workflow of their ontologies.

## 2 THE ONTOLOGY PATTERN RECONCILIATION WORKBENCH

A reconciliation candidate is a set of terms that should correspond to a common design pattern, for example {HP:0009900, MP:0004699} (terms referring to unilateral deafness) or {HP:0000822, MP:0000231} (terms referring to hypertension). The Ontology

\*To whom correspondence should be addressed: matentzn@ebi.ac.uk



**Fig. 1.** Comparing MP:hypertension and HP:Hypertension using the Ontology Pattern Reconciliation Workbench. The top panel shows an overview of the reconciliation state; in this case, the definitions are neither logically, syntactically nor grammatically equivalent, have 1 most specific common ancestor and subsume 3 other defined classes (impact). The reconciliation complexity in this case is the proportion of shared sub-expressions. The lower panel shows the two expandable class hierarchies under the hypertension classes, and in the central panel an analysis of the state of all patterns that classify underneath them.

Pattern Reconciliation Workbench caters to the following use cases across multiple ontologies: The user wants to

- search and collect reconciliation candidates
- explore collected reconciliation candidates
- test the impact of a set of externally specified or generated patterns
- monitor the reconciliation degree across (parts of the) ontologies

To realise these use cases, the workbench comes with three main views. The ‘CandIdent’ view allows users to select any number of ontologies, search them, individually or at once, for syntactic patterns and group them into reconciliation candidates (search and collect). For example, a user might search for patterns relating to mating behaviour across MP, HP and DPO, and group them together to form a reconciliation candidate. Candidates can be exported for use in future sessions.

The ‘Quick impact’ view allows the user to select a set of ontologies, classify them against a set of patterns and browse the resulting class hierarchy (test pattern impact). Patterns can be sorted by impact, which allows the user to get a quick sense of which patterns might have instances across which ontologies.

The ‘Reconciliation’ view allows users to explore reconciliation candidates based on a pre-defined mapping. Reconciliation candidates can be browsed, sorted and filtered through a table or through the tree browser view, which presents the merged class hierarchy of the selected input ontologies. By browsing down the class hierarchy, the user can inspect particular ‘branches’ of the merged ontology (e.g. sub-classes of ‘behavioural phenotype’), their level of semantic integration and their corresponding reconciliation candidates. Individual reconciliation candidates can be inspected along with the class hierarchies of the terms they related to, see Figure 1.

### 3 PATTERN DEVELOPMENT AND INSTANTIATION

The ontology starter kit (OSK) provides a standard toolkit and repository structure for developing ontologies. We have extended the OSK<sup>1</sup> to incorporate pattern-based ontology development using

the DOSDP framework (Osumi-Sutherland *et al.*, 2017). Once reconciliation candidates are identified (previous section), experts of the phenotype ontologies involved try to identify a common pattern for representing definitions, for example ‘**has part**’ *some* (‘increased amount’ and ‘**inheres in**’ *some* \$A and ‘**towards**’ *some* \$B and ‘**has modifier**’ *some* ‘abnormal’), where \$A and \$B are instances that can be instantiated by concrete concept from, for example, UBERON and the Gene Ontology (GO). This pattern can then be represented as a DOSDP. Our extended OSK incorporates pattern generation into the general release workflow by first compiling all patterns and their instance files into OWL definitions and then importing or merging them into the main release file.

The workbench is available on GitHub<sup>2</sup> and as a docker container.<sup>3</sup>

### ACKNOWLEDGEMENTS

This work is funded by a NIH Office of the Director Grant (5R24OD011883) to the Monarch Initiative.

### REFERENCES

- Harris, M. A., Lock, A., Bhler, J., Oliver, S. G., and Wood, V. (2013). FYPO: the fission yeast phenotype ontology. *Bioinformatics*, **29**(13), 1671–1678.
- Mungall, C. J., Gkoutos, G. V., Smith, C. L., Haendel, M. A., Lewis, S. E., and Ashburner, M. (2010). Integrating phenotype ontologies across multiple species. *Genome Biology*, **11**(1), R2.
- Osumi-Sutherland, D., Marygold, S. J., Millburn, G. H., McQuilton, P. A., Ponting, L., Stefancsik, R., Falls, K., Brown, N. H., and Gkoutos, G. V. (2013). The Drosophila phenotype ontology. *Journal of Biomedical Semantics*, **4**(1), 30.
- Osumi-Sutherland, D., Courtot, M., Balhoff, J. P., and Mungall, C. (2017). Dead simple OWL design patterns. *Journal of Biomedical Semantics*, **8**, 18.
- Rocca-Serra, P., Rutenberg, A., O’Connor, M., Whetzel, P., Schober, D., Greenbaum, J., Courtot, M., Brinkman, R., Sansone, S., Scheuermann, R., and Peters, B. (2011). Overcoming the ontology enrichment bottleneck with quick term templates. *Applied ontology*, **6**(1), 1322.

<sup>1</sup> <https://bit.ly/2IYSnZX>

<sup>2</sup> <https://bit.ly/2L4ejjm>

<sup>3</sup> <https://bit.ly/2kC2eX6>