Aligning the Human Phenotype and Mammalian Phenotype Ontology using Dead Simple Ontology Design Patterns

Nicole Vasilevsky[1*], James P. Balhoff[2], Christopher J. Mungall[3], David Osumi-Sutherland[4], Sebastian Kohler[5], Susan Bello[6], Cynthia Smith[6], Peter Robinson[7], Melissa Haendel[1]

[1]Ontology Development Group, OHSU Library, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, OR, USA
[2]Renaissance Computing Institute, University of North Carolina, 100 Europa Drive, Suite 540, Chapel Hill, NC, USA
[33]Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, MS977 Berkeley, CA, USA
[4]School of Informatics, The University of Edinburgh, 10 Crichton Street, Edinburgh, Scotland, UK
[5]Charité - Universitätsmedizin Berlin, 10117 Berlin, Germany
[6]The Jackson Laboratory, 600 Main St, Bar Harbor, ME, USA
[7]The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT, USA

ABSTRACT

The Human Phenotype Ontology (HP) and Mammalian Phenotype (MP) Ontology represent information about abnormal phenotypes encountered in human diseases and mammalian organisms, respectively. It is a goal to co-develop these ontologies and align the terminology and logical axioms to increase interoperability of the two ontologies. Towards this end, we have worked to develop consistent design patterns for commonly used types of classes, such as morphological phenotypes or abnormal anatomical structures. We are currently implementing 'dead simple ontology design patterns' (DOSDP), which are design patterns that can be specified in a YAML text file and can be used to generate new terms, documentation, retrofit old terms, and allows for reuse of patterns. This paper describes the development and implementation of the DOSDP in the HP and MP.

1 INTRODUCTION

Genotype and phenotype information are commonly used for characterization and diagnosis of human diseases. The Human Phenotype (HP) Ontology (http://www.human-phenotype-ontology.org/) was developed as a standardized vocabulary describing phenotypic abnormalities encountered in human diseases (Köhler et al., 2017). Similarly, the Mammalian Phenotype (MP) Ontology (http://www.informatics.jax.org/vocab/mp_ontology) represents phenotypes encountered in mammalian organisms that are used as models of human disease (Smith et al., 2004). The HP and MP have differing and overlapping use cases and needs. One shared use case is the Monarch Initiative (www.monarchinitiative.org), which aims to use ontologies, such as the HP and MP, and semantic technologies to aggregate data to support disease diagnostics, for common and rare diseases, where use of large scale integrated data can inform disease diagnosis and decisions. To allow for integration of diverse data that is annotated to disparate ontologies, we developed unifying ontologies such as UberPheno (Koehler 2013), which integrates phenotype ontologies, including HP and MP, using OWL Definitions (Mungall 2010). We are working to align the OWL design patterns in these two ontologies to make them to be more interoperable, and allow reasoning across the two ontologies. We need a simple, light-weight standard for specifying these design patterns that can then be used for generating documentation, generating new terms and retrofitting old ones. An ideal solution should be readable and editable by anyone with a basic knowledge of OWL and the ability to read Manchester syntax. It must also be easy to use programmatically without the need for custom parsers—i.e. it should follow some existing data exchange standard. Human readability and editability requires that Manchester syntax be written using labels, but sustainability and consistency checking requires that the pattern make use of IDs. The approach we used was developed by David Osumi-Sutherland and is called "Dead Simple Ontology Design Patterns" (DOSDP) (https://github.com/dosumis/dead_simple_owl_design_patterns, Osumi-Sutherland 2017).

1 METHODS

To promote the alignment and consistency across the HP and MP ontologies (and with a view to extension to other ontologies), we recently created DOSDP templates for a number of common phenotype ontology patterns. These are

found in the UberPheno ontology repository (https://github.com/obophenotype/upheno). Each pattern is represented as YAML conforming to the DOSDP standard. Patterns were developed for commonly used classes in HP and MP, such as morphological abnormalities (see pattern: abnormalMorphology.yaml), for example, 'Abnormal heart morphology' (HP: 0001627 and MP:0000266) or a decreased level of a molecular entity in a location (see pattern: decreasedLevelOfMolecularEntityInLocation.yaml), for example, HP_0002902 'Hyponatremia' and MP_0005634 'decreased circulating sodium level'.

The DPs were generated through manual inspection of the ontologies combined with knowledge of the ontology curators. In many cases these patterns were already implicit if not formally documented (see Mungall 2010).

Once we generated the DPs, we used dosdp-tools (https://github.com/INCATools/dosdp-tools) to query OWL definitions in the ontology to determine which classes are defined according to which pattern.

2 RESULTS

To date, 43 patterns were created for UberPheno, which are available here (https://github.com/obophenotype/upheno/tree/master/src/patterns). These DOSDPs can be used to generate new terms, as the patterns can specify the label, synonyms (exact, broad, narrow, and related), the text definition, and the logical definition. An example pattern is displayed in Figure 1.

```
pattern_name: abnormal

classes:
  quality: PATO:0000001
  abnormal: PATO:0000460
  Thing: owl:Thing

relations:
  inheres_in: RO:0000052
  qualifier: RO:0002573
  has_part: BFO:0000051

vars:
  entity: Thing

name:
  text: "abnormal %s"
  vars:
    - entity

annotations:
  - annotationProperty: oio:hasExactSynonym
    text: "abnormality of %s"
    vars:
      - entity

def:
  text: "Abnormality of %s."
  vars:
    - entity

equivalentTo:
  text: "'has_part' some ('quality' and
('inheres_in' some %s) and ('qualifier' some
'abnormal'))"
  vars:
    - entity
```

Fig. 1. An example of a DOSDP for 'abnormal entity'. Any subclass of owl:Thing can be used in this pattern.

An initial analysis using dosdp-tools of the HP and MP was performed to identify the number of terms that currently match the design patterns, and the number of terms not matching any pattern. As shown in Table 1, in the HP, 47% of the terms have logical definitions, and 72% of the terms in the MP have logical definitions. 12% of the terms in the HP currently match a design pattern; terms with a logical definition that don't match any pattern yet defined make up 35% of all terms in HP. In the MP, 39% of the terms match a design pattern and 33% of terms have a logical definition but do not match a defined pattern.

Human Phenotype Ontology	Number	Percent of total
Total number of terms	12.358	
Total terms matching a pattern	1,462	12%
Total terms not matching a pattern	10,896	88%
Total terms that have logical definitions	5,785	47%
Total terms that have logical definitions but don't match a pattern	4,323	35%
Total terms matching a basic EQ pattern	2,958	24%
Total terms that have a logical definition and don't match a defined pattern but do match basic EQ	1,759	14%
Total terms matching IIPO EQ pattern	554	4%
Total terms that have a logical definition and don't match a defined pattern but do match IIPO EQ	536	4%

Mammalian Phenotype Ontology	Number	
Total number of terms	**11,909**	
Total terms matching a pattern	4,622	39%
Total terms not matching a pattern	7,289	61%
Total terms that have logical definitions	8,525	72%
Total terms that have logical definitions but don't match a pattern	3,903	33%
Total terms matching a basic EQ pattern	5,905	50%
Total terms that have a logical definition and don't match a defined pattern but do match basic EQ	1,690	14%
Total terms matching IIPO EQ pattern	573	5%
Total terms that have a logical definition and don't match a defined pattern but do match IIPO EQ	573	5%

Table 1. Analysis of patterns in Human Phenotype Ontology and Mammalian Phenotype Ontology. EQ = entity/quality. IIPO = inheres_in_part_of.

Next, we performed a query of all the quality (PATO) terms used in expressions matching a standard entity-quality (EQ) pattern in HP and MP, where the expression did not match any of the current DOSDP templates. The standard EQ pattern is as such:

```
"'has_part' some (%s and ('inheres_in' some %s)
and ('has_modifier' some 'abnormal'))"
```

The results showed 222 and 282 PATO terms were used by HP and MP, respectively, and we do not currently have a DOSDP template for these patterns. While a DOSDP will not be created for every pattern, we will aim to create DOSDPs for frequently used quality terms, such as PATO:0001509 functionality and PATO:0000645 hypoplastic.

3 CHALLENGES

While the DOSDP will be useful for aligning the design of the logical definitions between ontologies, of course, there are limitations. A DOSDP cannot be created and applied for every use case. Adding some additional annotations to the terms will still have to be done manually. For example, the HP uses tags for layperson synonyms or abbreviations, and these annotations may have to be added manually after the creation of the term.

4 FUTURE DIRECTIONS

Future work will aim to retrofit the logical definitions for HP and MP terms to align the logical axioms between the two ontologies. Additionally, once these design patterns are finalized, they can be applied other phenotype ontologies, like the Zebrafish Anatomy Ontology. The Cell Ontology (CL) plans to adopt these DOSDP as well. Ultimately, our hope is these design patterns can be used to develop new quality assurance methodology for ontologies.

The DOSDP will be used by the new Table Editor that is currently under development as part of the Monarch Initiative. The Table Editor enables domain-specific concept visualization, table-based editing, and to output semantically consistent computable artifacts for use in software applications and data analytics. The Table Editor, which is currently under development, can be used to view and edit ontologies, such as for generating new terms, within a lightweight spreadsheet-style web application (https://incatools.github.io/table-editor/settings).

ACKNOWLEDGEMENTS

Monarch is supported generously by a NIH Office of the Director Grant #5R24OD011883, as well as by NIH-UDP: HHSN268201300036C, HHSN268201400093P, NCI/Leidos #15X143. Table Editor and DOSDP tool development is supported by NIH Grant #1U01 HG009453-01.

REFERENCES

Sebastian Köhler, Nicole A. Vasilevsky, Mark Engelstad, Erin Foster, Julie McMurry, Ségolène Aymé, Gareth Baynam, Susan M. Bello, Cornelius F. Boerkoel, Kym M. Boycott, Michael Brudno, Orion J. Buske, Patrick F. Chinnery, Valentina Cipriani, Laureen E. Connell, Hugh J.S. Dawkins, Laura E. DeMare, Andrew D. Devereau, Bert B.A. de Vries, Helen V. Firth, Kathleen Freson, Daniel Greene, Ada Hamosh, Ingo Helbig, Courtney Hum, Johanna A. Jähn, Roger James, Roland Krause, Stanley J. F. Laulederkind, Hanns Lochmüller, Gholson J. Lyon, Soichi Ogishima, Annie Olry, Willem H. Ouwehand, Nikolas Pontikos, Ana Rath, Franz Schaefer, Richard H. Scott, Michael Segal, Panagiotis I. Sergouniotis, Richard Sever, Cynthia L. Smith, Volker Straub, Rachel Thompson, Catherine Turner, Ernest Turro, Marijcke W.M. Veltman, Tom Vulliamy, Jing Yu, Julie von Ziegenweidt, Andreas Zankl, Stephan Züchner, Tomasz Zemojtel, Julius O.B. Jacobsen, Tudor Groza, Damian Smedley, Christopher J. Mungall, Melissa Haendel, Peter N. Robinson; The Human Phenotype Ontology in 2017. Nucleic Acids Res 2017; 45 (D1): D865-D876. doi: 10.1093/nar/gkw1039

Cynthia L Smith, Carroll-Ann W Goldsmith and Janan T Eppig. The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. Genome Biology 2005 6:R7 DOI: 10.1186/gb-2004-6-1-r7

Köhler, S., Doelken, S. C., Ruef, B. J., Bauer, S., Washington, N., Westerfield, M., Gkoutos G, Schofield P, Smedley D, Lewis SE, Robinson1 PN, Mungall, C. J. (2013). Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research. F1000Research,1–12. http://doi.org/10.3410/f1000research.2-30.v1

Mungall, C. J., Gkoutos, G., Smith, C., Haendel, M., Lewis, S., & Ashburner, M. (2010). Integrating phenotype ontologies across multiple species. Genome Biology, 11(1), R2. http://doi.org/10.1186/gb-2010-11-1-r2

Osumi-Sutherland D, Courtot M, Balhoff JP, Mungall C. Dead simple OWL design patterns. J Biomed Semantics. 2017 Jun 5;8(1):18. doi: 10.1186/s13326-017-0126-0.