A strategy to create a comprehensive disease ontology for human disease curation

Nicole Vasilevsky, James Balhoff, Matthew Brush, Sherri de Coronado, Gilberto Fragoso, Lawrence W. Wright, Laura Christopherson, Kimberly Robasky, Chris Mungall, Melissa Haendel

The Monarch Initiative is an international consortium that uses ontologies to integrate data from diverse sources in support of disease diagnostics and mechanism discovery. However, the lack of interoperability of ontologies across the basic–clinical divide is a challenge.

A large number of disease ontologies exist that represent different classification strategies or disease areas. The NCIt is a cancer ontology and used extensively in the clinical community, such as in the Genomic Data Commons, for drug applications, and for federal reporting. The NCIt is less well adopted in basic biomedical research in part due to its lack of interoperability with the OBO ontologies that are used more often by this community.

The Monarch Merged Disease Ontology (MONDO) integrates multiple disease vocabularies into a single coherent ontology. It was initialized via a semi-automatic method and has been iteratively enhanced with manual curation efforts. MONDO includes NCIt, the Online Mendelian Inheritance of Man (OMIM), which encompasses Mendelian diseases, Orphanet, which focuses on rare diseases, the Experimental Factor Ontology (EFO) used for drug discovery, the Disease Ontology (DO), which broadly classifies human diseases, and a number of other disease resources.

MONDO IDs were assigned to integrated class cliques based upon historical cross-referencing within existing ontologies, using the kBOOM algorithm to determine equivalency, subclass relations, or other relationships. This new merged ontology will be maintained using this strategy, but is also being curated for completeness and clinical relevancy. For NCIt, we largely accepted axioms as-is, except we weakened the defining equivalent axioms to subClassOf and added defining axioms using intuitive design patterns. With this strategy, NCIt and MONDO are merged coherently with either IDs being available as clique leaders. Both of the MONDO and NCIt-OBO version ontologies are available to the community for biocuration of cancers and other human diseases on the OBO Foundry site.