

A general strategy for generating expert-guided, simplified views of ontologies

Anita R. Caron (1), Josef Hardi (2), Ellen M. Quardokus (3), James P. Balhoff (4), Bradley Varner (1), Paola Roncaglia (1), Bruce W. Herr II (3), Shawn Zheng Kai Tan (1), Helen Parkinson (1), Mark A. Musen (2), Katy Börner (2), David Osumi-Sutherland (1)

(1) EMBL-European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK (2) Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, 94305 USA
(3) Department of Intelligent Systems Engineering, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408, USA (4) Renaissance Computing Institute, University of North Carolina, Chapel Hill, NC US

Introduction

The use of common biomedical ontologies to annotate data within and across different communities improves data findability, integration and reusability. Ontologies do this not only by providing a standard set of terms for annotation, but via the use of ontology structure to group data in biologically meaningful ways. In order to meet the diverse requirements of users, and to conform to good engineering practices required for scalable development, biomedical ontologies inevitably become larger and more complex than the immediate requirements of individual communities and users. This complexity can often make ontologies daunting for non-experts, even with tooling that lowers the barriers to searching and browsing.

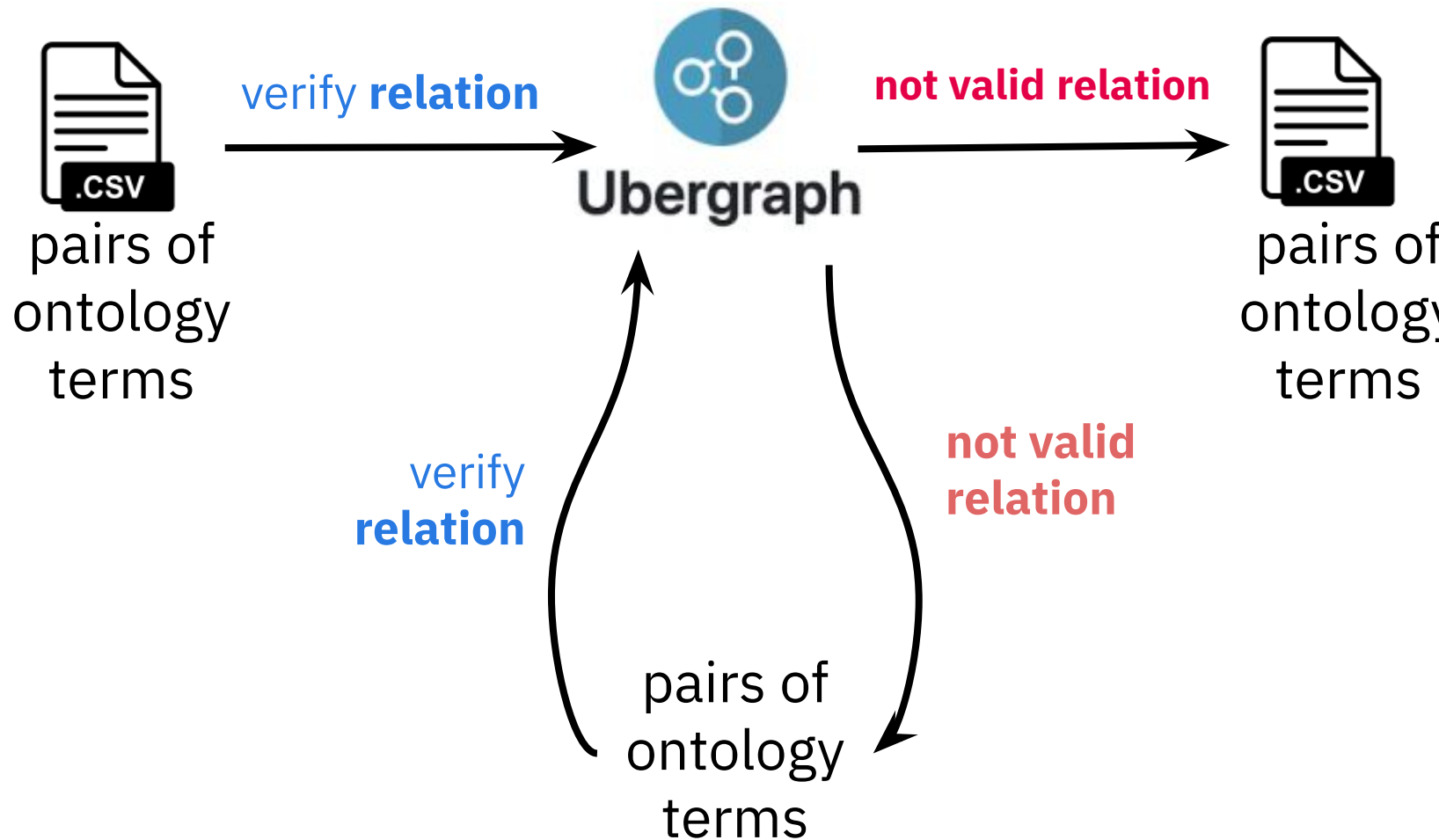
We have developed a suite of tools that take advantage of Ubergraph [1] to solve this problem for users that start from a simple list of terms mapped to a source ontology or for users who have already arranged terms in a draft hierarchy in order to drive browsing on their tools. This latter starting point is common among developers of anatomical and cell type atlases.

Methodology

Ubergraph

An RDF triplestore with 39 OBO ontologies merged, precomputed OWL classification and materialized class relationship from existential property restrictions. Ubergraph is the source of truth for the validation process.

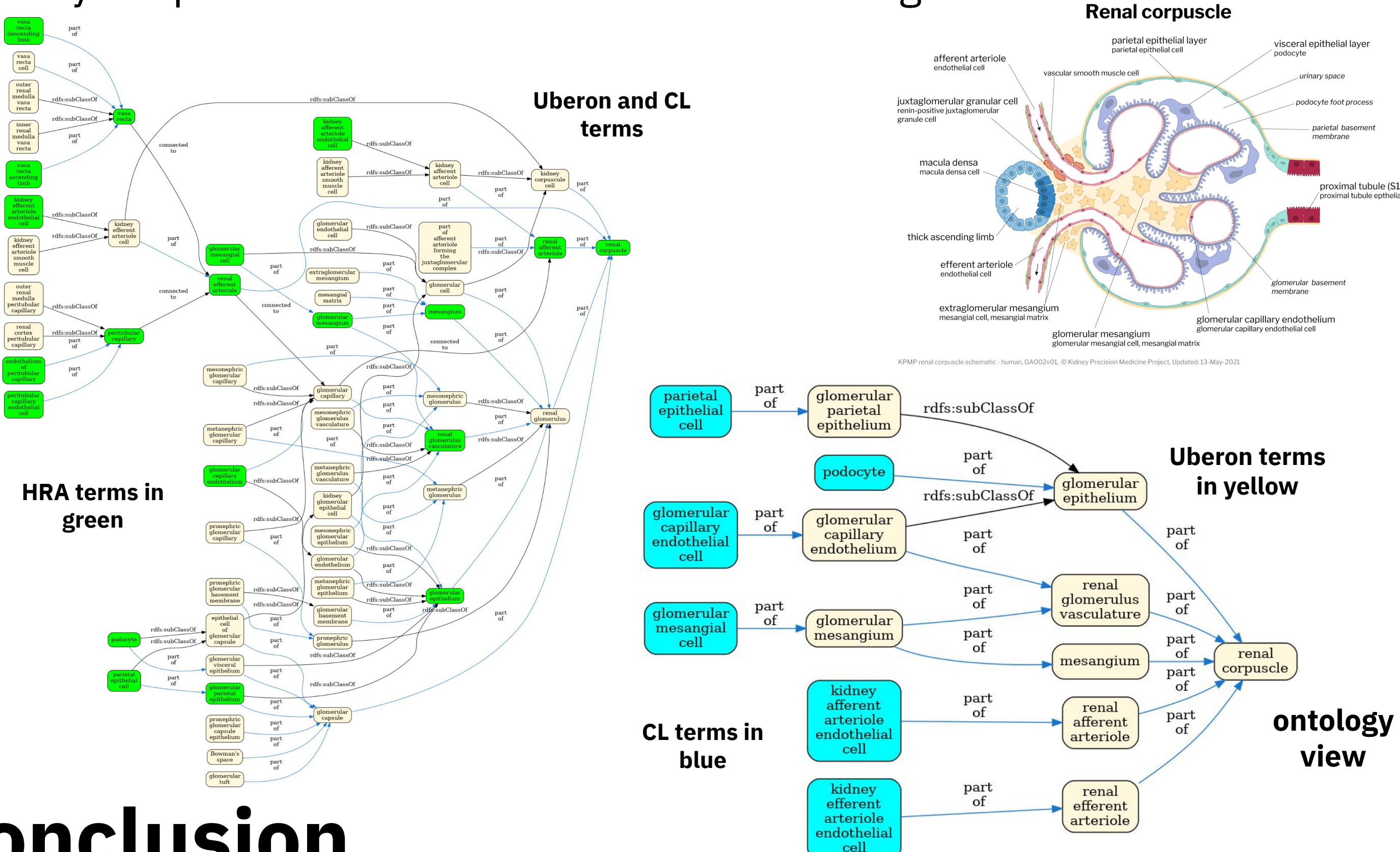
Validation using Ubergraph - relation-validator tool



Results

Ontology view

Relationships between terms for the anatomy of the renal corpuscle in Uberon and HuBMAP Human Reference Atlas (HRA) [3] illustrating the vastly simplified view of Cell and Uberon ontologies.



Conclusion

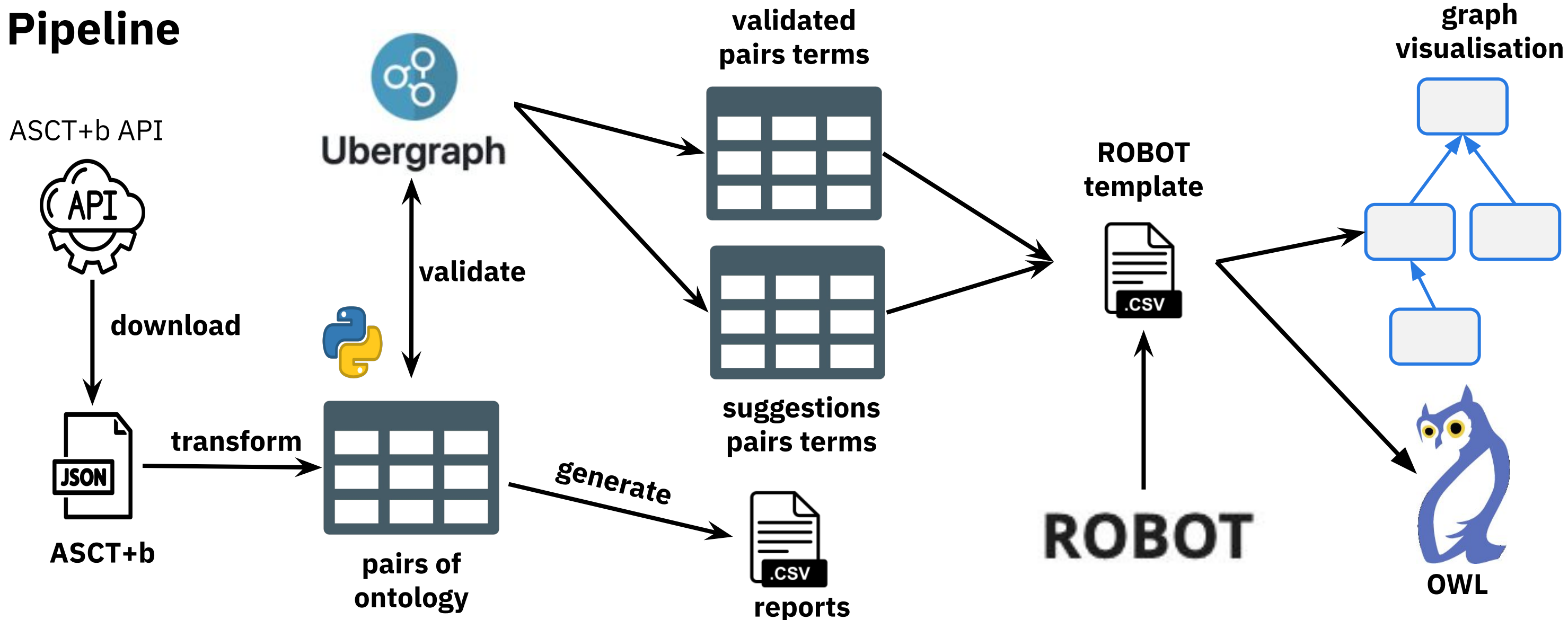
Here we described the view generation and hierarchy generation tools and illustrated their use in generating views and validation reports for the HuBMap Human Reference Atlas.

A view generation tool renders simple, tailored views of ontologies limited to a specified subset of classes and relationship types. These views accurately reflect the semantics of the source ontology, preserving its usefulness for grouping data in biologically meaningful ways. A hierarchy validation system validates these user-generated hierarchies against source ontologies, replacing unlabelled edges with formal ontology relationships which can be safely used to group content. A review of hierarchical relationships that do not validate against source ontologies provides potential corrections to hierarchies and source ontologies. A combination of validation and view generation can be used to generate ontology views based on the provided hierarchy.

The validation tool, *relation-validator*, is available at the PyPi repository <https://pypi.org/project/relation-validator/>.

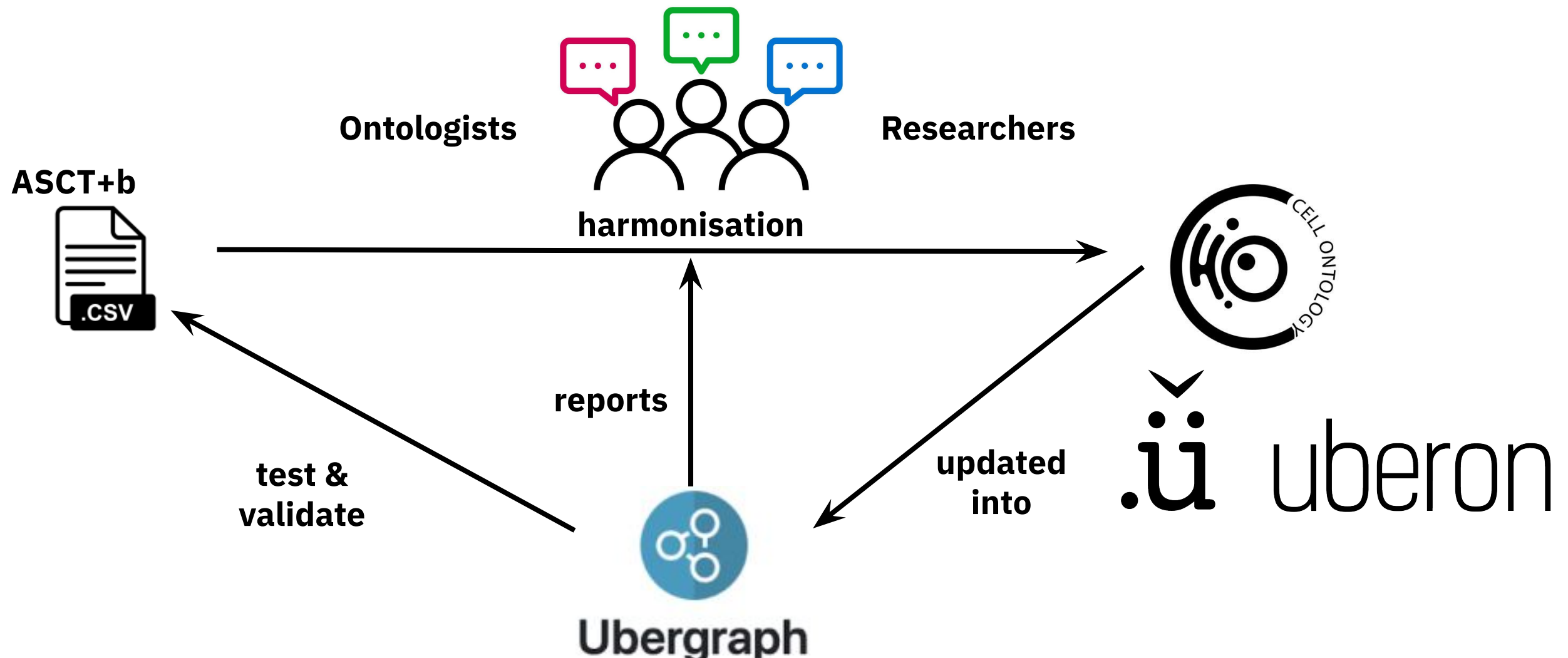
Anatomical Structures, Cell Types plus Biomarkers (ASCT+B) Tables [2]

CSV files representing the paratomy tree present relationships between various anatomical structures and substructures linked to their respective cell types and biomarkers.



Improvements on source ontologies

The validation reports can highlight possible improvements to the ontologies. Ontologists and researchers work together to harmonise their requirements and needs. Among the past corrections, the ontologies had changes in classification, relationship additions, and new term requests.



References

- [1] Balhoff, J. P., Bayindir, U., Caron, A. R. et al. Ubergraph: integrating OBO ontologies into a unified semantic graph <https://zenodo.org/record/7249759?c=Y7QE4QzP31c>
- [2] Börner, K., Teichmann, S.A., Quardokus, E.M. et al. Anatomical structures, cell types and biomarkers of the Human Reference Atlas. *Nat Cell Biol* 23, 1117–1128 (2021). <https://doi.org/10.1038/s41556-021-00788-6>
- [3] Herr, B.W., Hardi, J., Quardokus, E.M. et al. Specimen, biological structure, and spatial ontologies in support of a Human Reference Atlas. *Sci Data* 10, 171 (2023). <https://doi.org/10.1038/s41597-023-01993-8>

