# Community Contribution to the OHDSI Vocabularies: moving towards collaborative shared resource

Oleg Zhuk[1], Anna Ostropolets[1], Alexander Davydov[1], Christian Reich[2]

[1] Odysseus Data Services Inc., Cambridge MA, USA

[2] Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, Netherlands

## Background

OHDSI Vocabularies – the backbone of OMOP CDM - has grown over the years with more than 130 vocabularies currently incorporated. With growing number of use cases and countries adopting OMOP CDM, there is an increased need for transparent, robust, and seamless approach for incorporating new content and improving the quality of existing vocabularies. As a part of process improvement initiative, we are rolling out community contribution process, which lets the community maintain content and submit it for inclusion into the release.

## Methods

Community contribution pipeline and guidelines is rolled out in two parts.

The first part was released in the first half of 2023 and covers community contribution that can be done through templates.[1] The use cases currently covered can be divided into two distinct activities: adding new content and modifying the existing content. Adding new content includes adding a vocabulary as non-standard with or without mappings to existing OHDSI standard concepts, non-standard concept(s), synonyms, or mappings for existing concepts. Adding content does not require Vocabulary Team review and ensuring the quality of submission is the contributor's responsibility. As a tradeoff, additions can be performed fast. Modification of the existing content (changing concept attributes, mappings or promoting non-standard concepts to standard) requires Vocabulary Team and other working groups review. If the community wants to add new standard terms, they first need to add them as non-standard and then promote them to standard. This pipeline ensures the quality of the content that is used to populate OMOP CDM and, subsequently, in research across the community. Similarly, if a vocabulary contributed by the community needs a refresh, one would submit an addition or modification request depending on what the refresh entails.

Community can submit their requests through GitHub issues.[2] The issues contain the links to templates that need to be filled and placed on shared Google Drive. Templates contain the tables to populate, submission metadata and the checklist (quality checks) that must be filled in before submission.

Second part is being released in the second half of 2023. It includes processing more complex use cases: hierarchies, vocabularies processed through machinery and others. Such use cases require technical knowledge on the contributor's side and a common development environment on the Vocabularies side. Additionally, we are actively soliciting stewards who would support vocabularies not on the roadmap. Please reach out to the Vocabulary Team if you have a complex use case.

The contributions are incorporated in the next release if they are submitted and approved at least 2 months before the target release date (for example, November 2023 for the February 2024 release).

**Results**

We incorporated all part one contributions approved in time in the August 2023 release: two new non-standard vocabularies, modification of ICD10CM-SNOMED mappings, and addition of RxNorm Extension terms to support international drug studies. Contributions required multiple iterations and additional quality assurance, which points to the need for a more robust pipelines. Currently, we have twelve more contributions, six of which are done through the templates and are candidates for the next release.

We received four requests for external vocabulary support for vocabularies not on the roadmap and are working with stewards on their refreshes. To support this, we published vocabulary development guidelines on GitHub[3] and are setting up a common development environment.

**Conclusion**

Community contribution enables more transparent and collaborative approach to Vocabulary maintenance. It has a potential to greatly improve Vocabularies' quality and expand their content to meet community need. We encourage everybody to submit their requests on GitHub.

**References**

1. OHDSI Standardized Vocabularies. *Wiki· OHDSI/vocabulary-v5.0*, *GitHub*. Available at: https://github.com/OHDSI/Vocabulary-v5.0/wiki/Community-contribution (Accessed: 16 June 2023).
2. OHDSI Standardized Vocabularies. *GitHub Issues· OHDSI/vocabulary-v5.0*, *GitHub*. Available at: https://github.com/OHDSI/Vocabulary-v5.0/issues (Accessed: 16 June 2023).
3. OHDSI Standardized Vocabularies. *Vocabulary Development Process· OHDSI/vocabulary-v5.0*, *GitHub*. Available at: https://github.com/OHDSI/Vocabulary-v5.0/wiki/Vocabulary-development-process (Accessed: 4 October 2023).