

Community Contribution to the OHDSI Vocabularies, User-Level QC and a New Entity Mapping System SSSOM

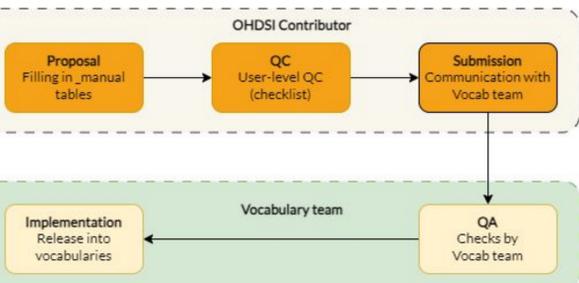
PRESENTER: **Oleg Zhuk**

BACKGROUND
OHDSI Vocabularies as of the beginning of 2023 encompass more than 130 vocabularies that are imported, manipulated and released by the OHDSI Vocabulary Team, which is a part of the OHDSI CDM Working Group.

METHODS
We split the use cases of community contribution into two groups: relatively simple, such as the addition or modification of relationships, and more complex, such as the modification of standard vocabularies and hierarchies, and then developed system to support them. Published guidelines represent the developed system and are currently only attributable to the first group of use cases.

RESULTS
According to published guidelines, every contributor needs to submit proposed changes in standardized format (*_manual* tables). User-level QC in the form of a population of a checklist, unique for every use case, is required. After filling out the checklist, proposed changes must be submitted to the vocabulary team. At that stage, no more input is expected from the OHDSI collaborator. The vocabulary team continues with additional QA and incorporates proposed changes into the vocabulary ecosystem (Figure 1).

Figure 1. Community contribution dataflow



Fix your own Vocabulary:

Templates and guidelines let you

- add and modify
- concepts and relationships

QC checklists help you don't do anything wrong

SSSOM metadata save your contribution



Published guidelines cover the following use cases (Table 1).

Table 1. Supported contribution use cases

#	Type
T1	Adding new non-standard concept(s) to an existing vocabulary
T2	Adding new synonym(s) to an existing concept(s)
T3	Adding a mapping to an existing concept
T4	Adding a new vocabulary as non-standard with mappings (full or partial) to a standard vocabulary
T5	Modifying attributes of an existing concept(s)
T6	Modifying mapping for an existing concept
T7	Promoting non-standard concepts to standard

Depending on the use case, the contributor needs to submit the following tables: *concept_manual*, *concept_relationship_manual*, *concept_synonym_manual*, *metadata*, *checklist*

The Vocabulary team stores metadata about the contribution. Their format is compatible with SSSOM. As of now, **date of submission, license status and contact details (email, name, organisation)** are collected per contribution and other parameters are collected per relationship (Table 2).

Table 2. Metadata collected per relationship

confidence	confidence in the new mapping (0-1)
predicate_id	type of matching (exactMatch, narrowMatch, broadMatch)
mapping_source	if applicable
mapping_justification	how the mapping was arrived at: ManualMappingCuration, LexicalMatching, etc.
mapping_tool	if applicable

Oleg Zhuk,
Anna Ostroplets,
Nicolas Matentzoglu,
Melissa Haendel,
Davera Gabriel,
Alexander Davydov,
Christian Reich

