

OHDSI Goals for 2023 Pillar #1: Standardized vocabularies

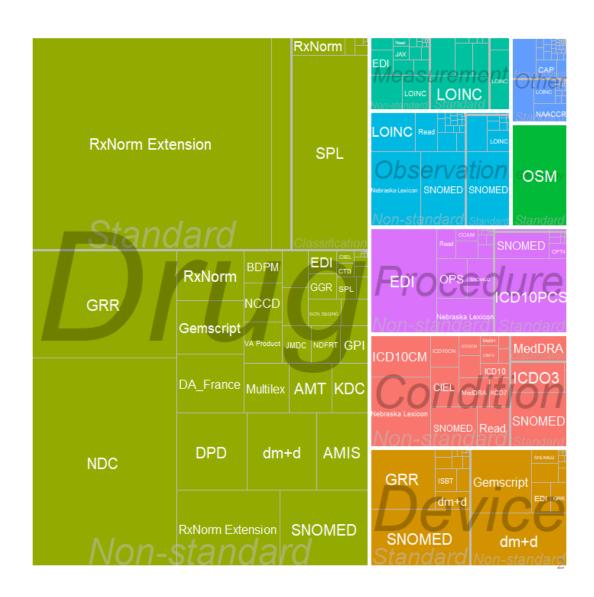
- Opportunity: Increase transparency and maturity with vocabulary development and evaluation process
- Proposed solutions:
 - Conduct landscape assessment to understand community needs
 - Develop code of conduct and developer guidelines
 - Disseminate vocabulary process and end-user documentation and roadmap
 - Establish centralized development infrastructure
 - Create standardized test development
 - Build vocabulary version release distribution service





OHDSI Standardized Vocabularies

- 136 vocabularies
- >8 million valid concepts
- >28 million valid relationships





OHDSI Vocabulary Team



Alex



Dmitry



Oleg



Christian



Vlad



Irina



Varvara



Anna

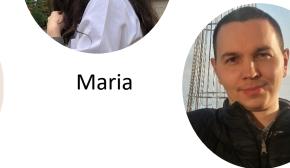


Masha



Mik

Dmitry



Timur



What we did in 2022

- Regularly updated >25 external vocabularies (such as SNOMED or RxNorm)
- Created/maintained our own mappings (incl. manual) for more than 20 vocabularies (such as NDC)
- Added 31 new vocabulary in 2020-22
- Addressed 37 issues, bugs and requests on GitHub in 2022
- Made 9 releases in 2022, which were downloaded >48,000 times



What we did in 2022

- Regularly updated >25 external vocabularies (such as SNOMED or RxNorm)

 But our documentation lags
- Created/maintained our own mappings (incl. manual) for more than 20 vocabularies (such as NDC)

But mapping coverage is not 100%

- Added 31 new vocabulary in 2020-22
 But more to add
- Addressed 37 issues, bugs and requests on GitHub in 2022
 But there are 169 outstanding issues
- Made 9 releases in 2022, which were downloaded 48,000 times

But we don't know what you do with them, and which problems are the most serious



Landscape assessment of community needs

1. Community survey

We are looking for feedback from:

- ETLers
- Developers
- Researchers

and others!

2. Interviews and focus group discussions with the working groups



What we will ask about

• Which vocabularies you use in ETL, research or development

What standard and source vocabularies do you use or have in your source data? Do you have vocabularies that are not in the OHDSI Vocabularies?

 Problems you encountered with Vocabularies completeness and correctness Have you encountered missing mappings to standard concepts? Wrong mappings or domain assignment?

• Problems you encountered with Vocabularies recency and updates

Have you had problems with Vocabularies download from Athena or upload into your database?

• What you like to see improved

Have you had problems with delayed Vocabularies release or when doing research on multiple Vocabularies versions?

What is needed to improve your confidence in Vocabularies content and processes?



What we will do with it

• Which vocabularies you use in ETL, research or development

• Determine how to allocate the resources across the vocabularies to prioritize more important content

 Problems you encountered with Vocabularies completeness and correctness

• Problems you encountered with Vocabularies recency and updates

• Prioritize process improvement activities

• What you like to see improved



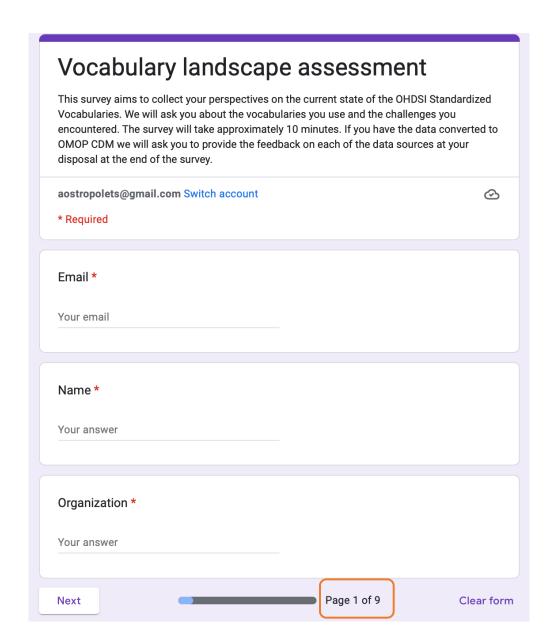
Establish a better way for community contribution

• Publish the report



Takes $\sim 10 \text{ mins}$

Second survey for those who have a data source converted to OMOP CDM





Please help us and fill the survey!

bit.ly/3iTnyco



Due February 23rd