OHDSI Standardized Vocabularies Landscape Assessment

Version:	1.0
Prepared on:	April 23, 2023
Created by:	Anna Ostropolets

Acknowledgement: The report is based in part on work from the Observational Health Data Sciences and Informatics collaborative. OHDSI (http://ohdsi.org) is a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics. This project was funded through the Chan Zuckerberg Initiative Essential Open Source Software for Science grant.

Table of Contents

1. List of Abbreviations

Abbreviation	Phrase
CDM	Common Data Model
OHDSI	Observational Health Data Sciences & Informatics
OMOP	Observational Medical Outcomes Partnership

2. Responsible Parties

OHDSI's mission is to improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care. As a community, we strive to promote openness and inclusivity by creating an environment where all voices are heard.

Author	Affiliation	Email
Anna Ostropolets	Odysseus Data Services Columbia University Medical Center	ostropolets@ohdsi.org
Patrick Ryan	Janssen Research & Development Columbia University Medical Center	ryan@ohdsi.org
George Hripcsak	Columbia University Medical Center New York-Presbyterian Hospital	gh13@cumc.columbia.edu
Christian Reich	Northeastern University, Roux Institute Erasmus University Medical Center	reich@ohdsi.org
Alexander Davydov	Odysseus Data Services	alexander.davydov@odysseusinc.com
Mik Kallfelz	Odysseus Data Services	michael.kallfelz@odysseusinc.com

Study Timeline:

January 05, 2023:	Study initiation
January 24 - February 28, 2023:	Data collection
March 1 - March 23, 2023:	Data analysis and report preparation
March 23 - April 23, 2023:	Finalization of study report

3. Executive Summary

We received 183 responses from community members and information about 60 data sources around the globe and talked to package maintainers and working groups. Overall, 87% of the community feels confident about Vocabularies' integrity.

The most commonly used vocabularies were SNOMED, CPT4, HCPCS, LOINC, RxNorm, ATC, CVX, ICD family, and ICDO3. However, more efforts are needed to establish an external contribution pipeline for timely inclusion of source vocabularies and coding schemes into the Vocabularies.

Main errors in the Vocabularies content reported by the community include erroneous mappings, uphill mappings, unclear domain assignment and gaps in hierarchies.

The OHDSI community is interested in an ability to download a given version of the Vocabularies and project the impact of the changes across versions on research and ETL tasks. The common data and Vocabularies refresh in the community follows an annual or semi-annual cycle, but there is a high variation of the Vocabularies versions across the community, which may complicate network studies. The community also expressed a need for easily digestible documentation showcasing real-world Vocabularies' use cases, transparent roadmap, and impact of the releases and changes on ETL and research, and more details regarding specific vocabularies development and quality assurance.

4. Rationale & Background

This report summarizes the finding of the landscape assessment of the OHDSI community needs related to the Observational Health Data Sciences and Informatics (OHDSI) Standardized Vocabularies (“Vocabularies”) conducted in February 2023.

Vocabularies is a common reference ontology system mandatory to all data holders in the OHDSI network [1]. It consists of imported and de-novo created ontologies, terminologies and vocabularies that are used to harmonize the data in Observational Medical Outcomes Partnership (OMOP) common data model (CDM) and can be used for data extract-transform-load (ETL), research, software development and other purposes.

This document provides insights into the community's attitudes towards the Vocabularies and the main challenges associated with them. The landscape assessment will be used to prioritize activities in Vocabularies maintenance and improvement and as a benchmark for further assessment.

5. Methods

As a part of the landscape assessment, we distributed a two-part survey (a general survey and a database-specific survey) through various channels across the community. The survey contained questions about the use of the Vocabularies, its completeness, correctness, intuitiveness of use, recency and versioning as well as documentation. The database-specific part of the survey included questions about the vocabularies, ontologies and coding schemes used in the data, frequency of data and Vocabularies refresh, and the current Vocabularies and CDM versions.

We conducted interviews with the OHDSI package and tool maintainers (HADES, Atlas, Data Quality Dashboard) as well as collected feedback from the working groups (Oncology and Genomic, Phenotyping, GIS, Ophthalmology, Dentistry, Health Equity, Imaging, Psychiatry) and several individuals. Finally, we inspected Vocabulary v5.0 GitHub issues and the OHDSI forum.

6. Results

6.1. Overall assessment

We received 183 responses from the community members from 144 institutions across the US, UK, Europe, Asia, Africa. Our data-source specific part of the survey was filled for 60 data sources, which is the largest number of data sources covered in any of the OHDSI studies.

Deidentified responses can be found in supplementary materials. The data was de-identified and all sensitive information was removed to maintain participants' privacy.

We discovered that a substantial part of the community uses the Vocabularies for more than one task. 78% of responders use the Vocabularies for transforming the data into the OMOP CDM, 65% of responders use the Vocabularies for research (with characterization being the most common study type) and 28% of responders use the Vocabularies for software, tool and method development such as OHDSI stack tools and tools for ETL, NLP, mapping and clinical decision support. The other use cases for Vocabularies use include data and ontology manipulations outside of OMOP CDM.

6.1. Vocabulary use

As of March 2023, the Vocabularies contain 137 ontologies, terminologies and vocabularies [2]. Table 2 presents the list of the vocabularies most used in research and present in source data (denominator is non-empty responses, 60 for data and 119 for research).

Table 2. Vocabulary use in research and data. Green box represents use by >50% of the community, yellow – between 20% and 50% of the community, pink – between 10% and 20% of the community and dark pink – less than 10% of the community.

Vocabulary	Used in data	Used in research
ATC	45%	62%
CPT4	50%	40%
ICD-10(CM)	57%	<10%
ICD-9(CM)	62%	<10%
ICD-10PCS	43%	51%
ICD-9-Proc	48%	43%
LOINC	68%	25%
RxNorm	33%	79%
RxNorm Extension	<10%	53%
SNOMED	57%	86%
Cancer Modifier	<10%	25%
CVX	18%	13%
HCPCS	42%	33%
ICDO3	32%	35%
MedDRA	10%	24%
UCUM	<10%	26%
NDC	33%	<10%
ICD-9 (int. versions)	28%	<10%
ICD-10 (int. versions)	48%	<10%
NAACCR	17%	15%

Medicare Specialty	15%	<10%
Revenue Code (CMS)	15%	<10%
CMS Place of Service	13%	<10%
CDT	12%	<10%
DRG (CMS)	12%	<10%
Multum	12%	<10%
NCIt	12%	<10%
NUCC	12%	<10%
ABMS	10%	<10%
dm+d	10%	<10%
HemOnc	10%	14%
OncoKB	10%	<10%
OncoTree	10%	11%
ClinVar	<10%	10%
Nebraska Lexicon	<10%	10%
HGNC	<10%	10%

Overall across the responses, the most used vocabularies tend to be both present in the data and used in research: standard vocabularies (SNOMED, CPT4, HCPCS, LOINC, RxNorm), classificational (ATC, CVX) and source (ICD family, ICDO3). More than half of the community uses RxNorm Extension and ICD10PCS in research.

Other vocabularies that are used in less than 10% of the data sources include MeSH, NDFRT, Read, CTD (Comparative Toxicogenomic Database), Gemscript, HES Specialty, ICD-10(GM), ISBT, Korean Revenue Code, MDC (CMS), Nebraska Lexicon, OPCS4, PPI, SPL, VA Product, APC, CAP, CIViC, DPD, EDI, HGNC, ICD-9-Proc (CN), KCD-7, OPS, OXMIS, PCORNet vocabulary, VANDF, CGI, GGR, JAX, JMDC, NFC, SOPT, AMIS, AMT, BDPM, CCAM, CIEL, CIM10, DA France, Ephmra ATC, GGR, GPI, ICD-7, ICMP2, KNHIS, local code system, LPD Australia, LPD Belgium, MMI, Multilex, NCCD, OMOP Extension, OSM, PHDSC, Radiology AGFA Impax, Radlex, SMQ, SNOMED Veterinary, SUS and UK Biobank.

Vocabularies used by less of 10% of the responders engaged in research include ABMS, AMT, CCS, CIEL, ETC, Gemscript, GPI, HPO, ICD11, ICMP2 (+), ICPC, IMO, Indication, ISBT, JMDC, MDC, NDFRT, OMOP Extension, OMOP Genomic, OPCS4, OPS, Read, SNOMED Veterinary, SPL, VANDF and WHO Drug.

However, individuals expressed specific interest in the vocabularies that serve their use cases such as SNOMED Veterinary, oncology vocabularies and others.

Additionally, the community members reported having source vocabularies, coding schemes and terms not included in the Vocabularies. We will discuss them further in “Vocabularies completeness”.

6.2. Vocabulary integrity and main challenges

Overall, 87% of the community feels confident about Vocabularies’ integrity (Figure 1), with more than a quarter of responders feeling extremely confident.

While these numbers are encouraging on their own, they can also be used as a ballpark for further assessment of the progress and improvements in the Vocabularies over the years.

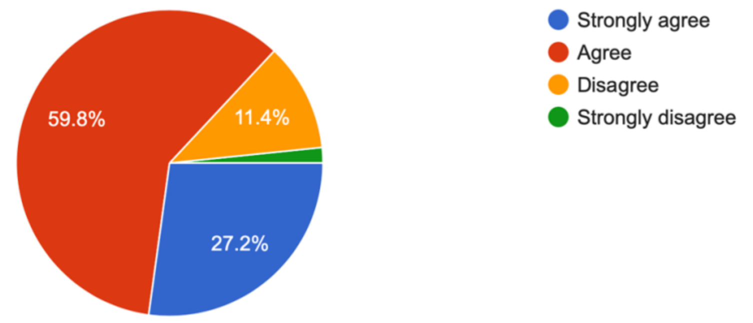

Figure 1. Confidence level in the integrity of the Vocabularies.

When looking at the *challenges that the individuals who strongly disagreed with the statement* faced, we found two main factors.

First, the researchers who strongly disagreed with the statement had previously dealt with the mappings from source to standard concepts in the Condition domain that did not meet their research needs or were erroneous (specifically ICD-10(CM) to SNOMED-CT mappings). As their studies required mapping change or improvement, the perception of the Vocabularies’ quality remained the same even if the errors were fixed later.

Second, we identified that there is a lack of educational materials on how to use vocabularies to construct concept sets and identify patients of interest or perform ETL refreshes. For example, domain changes or unclear domain assignment contributed to lower perceived quality.

When looking *at the overall challenges reported based on the totality of the community's responses*, mappings (Table 1) were one of the main concerns contributing to lower confidence. The table is organized around the axes of Vocabularies quality (conformance, completeness, recency, access) and documentation, which is separated into a standalone group.

Table 1. Main challenges with the Vocabularies and areas for improvement.

Challenges	**Areas for improvement**
Conformance	
Lack of transparent quality assurance procedures	- an ability to assess vocabulary maturity - publicly available quality checks - common environment for quality assurance
Erroneous relationships	- mapping improvement based on the previously reported issues and external validation - alignment of mappings within the ICD family - fix of gaps in hierarchies
Mappings associated with information loss	- upward (uphill) mappings, for example, mappings from ICD-10(CM) to SNOMED-CT - 1:many mappings
Completeness	
Lack of comprehensive hierarchies	- alignment of the standard vocabularies in Procedure and Measurement domains
Insufficient mapping coverage	- more mappings from non-standard to standard concepts - automated handling of mappings for deprecated standard concepts
Source vocabularies not included in the Vocabularies	- an ability to timely include source vocabularies and coding schemes - transparent pipeline for community requests
Recency and versioning	
Variability in release cadence	Clear roadmap and release schedule
Only the current version of the Vocabularies is available for download	Ability to download different versions
Unpredicted impact of Vocabularies changes on ETL, research and OHDSI tools	- better documentation for vocabulary changes (release notes) - tools to assess the impact of vocabulary changes on common tasks

	- proactive tracking of Vocabularies use in existing OHDSI tools
Vocabularies version variation in network studies	- better alignment of the community in terms of Vocabularies refresh - fewer releases
Vocabulary access	
Absence of automated pipeline for Vocabularies processing	- REST API for Athena - improve CPT4 injection (CPT4.jar)
Errors during Vocabularies upload into a relational database	- working sample scripts that deal with datatype mismatch and special characters
Documentation	
Lack of educational materials for common vocabulary tasks	Tutorials and guides on how to: - create concept sets - use classification concepts and vocabularies - handle Vocabularies changes during ETL refreshes
Lack of documentation on Vocabularies processing and quality assurance	- support mapping meta-data with an ability to distinguish different types of mapping precision - transparent and assessable quality checks
Lack of end-user documentation on specific vocabularies and topics	- documentation for separate vocabularies with links to source documentation - tutorials and easily digestible documentation on Vocabularies structure and principles
Challenges with custom mappings	- support mapping meta-data - guidance on how to perform custom mappings

We will discuss these groups in more detail.

6.3. Vocabularies conformance

Overall, there is a broad need for transparent quality assurance procedures that are publicly available. That includes an assessment of the vocabulary maturity, quality of the mappings, reports of the quality assurance (QA) tests performed for each release and other relevant reports that should be easily assessable to the community.

We observed several common topics related to Vocabularies quality. The largest concern about the mappings is connected to the mappings from a granular source concept to a broad target concept (so-called upward or uphill mappings). This type of mappings is mainly observed in Condition domain and is associated with a lack of granular concepts in SNOMED-CT that directly match the source codes (particularly the ICD family). They may interfere with accurate patient selection in phenotyping as it may lead to the inclusion of the patients with other disorders.

As the Vocabularies currently do not distinguish between exact mappings and upward mappings, the community would be interested in having meta-data elements that specify the precision of mappings.

Erroneous mappings were reported by a rather small number of community members and were commonly associated with existing GitHub issues. The community suggested, among, others, external validation, learning from previous use cases and checks for alignment in the ICD family as potential solutions to this problem. There were problems related to other types of relationships, particularly hierarchical, which results in the gaps in hierarchies (such as ingredients not connected to drug products).

Errors reported for domain assignment were mostly caused by unclear domain assignment procedures and require more detailed and comprehensive description of the current approaches to domain and concept class assignment. For example, the concepts that the users would expect to see in the Condition domain were assigned Observation or Measurement domain.

Takeaways:

Main errors in the Vocabularies content reported by the community include erroneous mappings, uphill mappings, unclear domain assignment and gaps in hierarchies.

6.4. Vocabularies completeness

We will discuss two aspects of completeness of the Vocabularies: a) the relationships that enable meaningful use of Vocabularies (‘Maps to’ and hierarchical relationships) and b) the vocabularies, ontologies and coding schemes that are not currently included in the Vocabularies.

6.4.1. Relationship completeness

The lack of mappings for non-standard concepts was a commonly reported problem. For example, that problem was especially evident for the concepts whose standard counterparts became non-valid, which, as a result, led to the loss of valid ‘Maps to’ links.

The lack of mappings among standard concepts was a commonly reported problem as well. Specifically, it was referred to as a lack of comprehensive hierarchies and existence of multiple standard concepts that represent similar meaning.

When multiple standard concepts with similar meaning exist, it is not necessarily clear which target concepts to choose for custom mappings in ETL or for concept sets in phenotyping. Examples include the domains that have not undergone deduplication (such as Device) or sporadic examples in other domains (such as Drug or Condition).

Standard concepts in other domains usually carry different nuances of meaning as they have different granularity, for example more broad SNOMED-CT procedures compared to more granular ICD10PCS procedures. These domains, as opposed to Condition and Drug, are only partially aligned. Absence of comprehensive hierarchies in these domains influences cohort creation due to the lack of vocabulary alignment. This specifically concerns Measurement (LOINC and SNOMED-CT) and Procedure domain (SNOMED-CT, HCPCS, CPT4, ICD10PCS, ICD9Proc) where only limited links and joint hierarchies exist. As the community's knowledge about existing links among these vocabularies may be limited, better documentation and guidance as well as dedicated efforts to create comprehensive hierarchies are required.

6.4.2. Vocabulary completeness

There is a large body of the vocabularies, ontologies and coding schemes that are used by the community but are not yet included in the OHDSI Vocabularies. The scope varies from several concepts to multiple vocabularies and is commonly dealt with by creating custom mappings to be stored in source_to_concept_map, creating 2 billion custom codes or leaving out the content not covered by the Vocabularies.

This body of entities can be loosely classified into two categories: a) structured external vocabularies (have semantic identifiers and are maintained by external organizations) and b) unstructured elements present in the data as source coding schemes or free text. The examples of the vocabularies not included in the Vocabularies include but are not limited to ICPC, Radlex, Z index, INDEPTH, NCSP and national ICD vocabularies.

Unstructured data-source specific elements not covered by the Vocabularies include race, ethnicity, specialty, care site, social determinants of health. Additionally, laboratory tests and values are commonly poorly structured and represent a high portion of the elements requiring custom mappings across the network. Similarly, surveys and questionnaires, electronic records flowsheets and data from the registries or clinical trials commonly require extensive harmonization.

There is an explicit need to establish an external contribution pipeline reported by several community members and an implicit need reflected in a high number of vocabularies not in the OMOP ecosystem. Such a pipeline would enable more timely inclusion of the source vocabularies into the Vocabularies and requires approaches and tools to a) propose new concepts and vocabularies, b) format the contribution according to the structure of the vocabulary tables to facilitate seamless incorporation and c) a comprehensive QA system to maintain consistency and integrity of the structure of Vocabularies. It must be accompanied by a transparent and clear process and guidelines for external contributors.

Takeaways:

Efforts are needed to create comprehensive hierarchies in certain domains and establish an external contribution pipeline for timely inclusion of source vocabularies into the Vocabularies. The inclusion of structured external vocabularies and unstructured elements will enhance the completeness of the Vocabularies, and better documentation, guidelines, and tools are required to achieve this goal.

6.5. Vocabularies recency: releases and roadmap

We observed high variation of the Vocabularies versions used across the community, spanning from 21-Aug-2020 to 23-Jan-2023 (median version 22-Jun-2022). Only 8% of data sources are on 2023 version of the Vocabularies.

On average, the lag between the data capture and OMOP ETL refresh is 5 months, which indicates that on average the new codes will be needed 5 month after they appear in the data. New drugs and COVID-19-related codes (such as codes for vaccines and infection) were the exception to this observation as explicitly reported by several members of the community.

More than half of the community update their vocabularies semi-annually or annually with only 5% of the community updating the Vocabularies with each new release (for reference, there were 8 releases made in 2022).

Several community members explicitly commented on the need for fewer releases to facilitate execution of the studies across the network as well as more seamless data refreshes.

Aside from releases, a need for a transparent roadmap indicating planned releases and refreshes is a common topic across the community. Refreshes of ATC, ICD-10(CM), SNOMED-CT and ICDO3 were among the most requested refreshes in the community (with more than 3 members of the community asking for or suggesting refreshing the corresponding vocabularies).

Takeaways:

The common data and Vocabularies refresh in the community follows annual or semi-annual cycle. There is a high variation of the Vocabularies versions across the community which may complicate network studies.

6.6. Vocabularies versioning

As of the moment, the community can only download the current version of the Vocabularies. An ability to download a given version appears to be of an interest to the community. Since different versions are not stored centrally, currently there is no easy way to compare the versions or estimate an effect of the vocabulary change on the common tasks performed in the community.

Longitudinal Vocabularies changes may impact ETL, cohort construction and software development and maintenance. The biggest challenge with vocabulary versioning is domain assignment changes. Such changes may impact ETL if the scripts are not designed to handle domain stages (for example, if all the records are not pulled together and then distributed to the corresponding tables based on the domains). Additionally, some of the changes present challenges to storing the values associated with records in those CDM tables that do not support them (for example, if measurements with associated values are moved to Condition domain). Similarly, domain change presents a problem with already existing concept sets that are executed against a set table in CDM.

Second, previously mentioned loss of mappings from source concepts to their standard counterparts leads to data loss during ETL refreshes as well as patient loss if previously created cohort definitions are executed on a new vocabulary version. ETL scripts should be designed to follow ‘Maps to’ relationships to automatically track the new mappings.

Third, there are parts of the packages that depend on the specific (hard-coded) concepts in the Vocabularies and therefore are prone to unexpected behavior if such concepts change.
The following packages currently have such concepts:

- CohortMethod (Table 1)
- FeatureExtraction (scores, handling of eras)
- Atlas (treemap, pruning of the top SNOMED-CT levels)
- Capr (multiple dependencies)
- Data Quality Dashboard (units, list of conditions for males/females)
- CirceR (depends on domains, invalidation with no replacement may a problem)
- CohortDiagnostics (index event breakdown, orphan codes)

While the packages may be refactored in future to eliminate dependencies, it is important to track a potential influence of vocabulary changes on them.

Some responders expressed a need for partitioning of the vocabulary downloads across refreshes (in other words an ability to only load new or changed concepts).

Takeaways:

OHDSI community is interested in an ability to download a given version of the Vocabularies and project the impact of the changes across versions on research and ETL tasks. The dependencies in the packages that may be influenced by the changes in the Vocabularies need to be proactively monitored.

6.7. Access to Vocabularies: download and upload

This section describes the feedback associated with the access to Vocabularies: vocabulary download which is performed through Athena and its subsequent upload to a relational database. While Athena is a distribution service, we noticed that there is confusion between the content of the vocabularies and the tool that distributes them, which may point at a need for better documentation of the Vocabularies parts.

Overall, approximately a quarter of the community experienced challenges with vocabulary download or upload. The main problem is CPT4.jar file (the speed of CPT4 vocabulary name retrieval or the overall inconvenience of using the file). Second most common issue is vocabulary upload into a relational database to be used in OMOP CDM. The issues mentioned include datatype mismatch, special characters, need to drop constraints and more.

Third problem is related to licensing. As OHDSI is distributing multiple vocabularies that have their own licensing practices, some of the vocabularies either require a separate license (such as CPT4) or are private (such as DA France). The process of obtaining a license is not always clear to the users and some licenses are hard to acquire. While it may be outside of the OHDSI control, the process of obtaining a license and potential pitfalls can be described better.

Finally, several members of the community expressed an interest in having automated pipelines for vocabulary upload and download, which can be facilitated by having an application programming interface for Athena. Similarly, some of the comments related to the lag between the Vocabularies version in Athena and local Atlas instances may point at the expectation that the pipelines for vocabulary download and upload are fully automated and enable instant updates.

Takeaways:

Challenges with obtaining CPT4 seem to be the major stumbling block in vocabulary download and upload, which may especially impact newcomers who do not have a great experience with OHDSI or technical capabilities in their organizations. There is a need for automated pipelines for Vocabularies download and upload.

6.8. Vocabularies documentation

About a third of the community would like to see more comprehensive or evolved documentation. Here, we will summarize all the documentation feedback.

First, there is a need for more educational materials. The responders noted the positive influence of the EHDEN Academy [3] and Book of OHDSI but would like to see more educational materials.
The topics that should be covered in such materials span from relatively simple concepts (such as reverse relationships) to more complex areas such as the use of the hierarchies.

More members would like to see easily digestible guides, tutorials and examples on the use of Vocabularies for ETL and phenotyping (standalone use and use in the tools such as Atlas). Specifically, hands-on tutorials or practical examples for concept set creation and custom mappings would be appreciated.

The examples for hierarchy use appeared to be a common request, partially because of the complexity of some of the hierarchies (such as ATC and RxNorm) and partially because of the gaps in completeness of the hierarchies (in Procedure and Measurement domains).

Second, the community would like to see more information on vocabulary changes and their impact on common tasks. The current release notes [4] are either unknown to common public or do not provide sufficient information to estimate how the changes in the Vocabularies may impact the common data harmonization and research tasks.

Moreover, there are no tools readily available to easily estimate the impact of the difference between two given versions of the Vocabularies on ETL or cohorts. While there are some existing solutions developed in the community [5,6], they are either not enforced for use by the broad community or lack needed features.

Third, we received comments about scattered documentation that can be found across multiple resources (Book of OHDSI, forums, OHDSI wiki, GitHub wiki). More work is needed to establish a single place for vocabulary documentation and aggregate all the information there.

Finally, there are requests for more comprehensive documentation when it comes to specific vocabularies. There should be links to the existing documentation that is provided by the original vocabulary developers (such as SNOMED-CT [7]) as well as more detailed documentation on the transformation procedures performed in OHDSI, quality assurance, maturity of mappings and details on who, how and when created them, comprehensiveness of hierarchies and more.

Commonly requested topics included explanation of how the domains are assigned, meaning of concept classes and more information on how standard versus non-standard concept determination is made (in other words, why some of the concepts are standard and the others are not).

Takeaways:

There is a need for easily digestible documentation showcasing a) real-world Vocabularies' use cases, b) a transparent roadmap and impact of the releases and changes on ETL and research, and c) more details regarding specific vocabularies development and quality assurance.

7. Strengths & Limitations

This is the first large-scale assessment of the community needs related to the use, maintenance, and distribution of the OHDSI Standardized Vocabularies. All the appropriate measures such as proactive outreach to the community, interval reminders and targeted outreach with the individuals and working groups were taken to ensure comprehensive community coverage. While the results suggest that we captured members with various levels of involvement and familiarity with OMOP CDM, it cannot be guaranteed that we covered all needs.

8. Protection of Human Subjects

Confidentiality of participating community members is maintained always. The study report contains aggregate data only and does not identify individuals. The supplementary material data was de-identified and all sensitive information was removed to maintain participants' privacy.

9. Plans for Disseminating & Communicating Study Results

This report will be disseminated in the OHDSI community and may be presented on the meetings and conferences.

10. Appendices

The file with deidentified survey responses can be found here.

11. References

1 Informatics OHDS and. *Chapter 5 Standardized Vocabularies | The Book of OHDSI*. https://ohdsi.github.io/TheBookOfOhdsi/ (accessed 16 Mar 2023).

2 Standardized Vocabularies. GitHub. https://github.com/OHDSI/Vocabulary-v5.0/wiki/Standardized-Vocabularies (accessed 16 Mar 2023).

3 EHDEN Academy. https://academy.ehden.eu/ (accessed 17 Mar 2023).

4 Releases · OHDSI/Vocabulary-v5.0. https://github.com/OHDSI/Vocabulary-v5.0/releases (accessed 16 Mar 2023).

5 Tantalus. https://mi-erasmusmc.shinyapps.io/Tantalus/ (accessed 18 Mar 2023).

6 DeFalco F. An Evaluation of the Impact of Vocabulary Evolution on Established Phenotypes. ohdsi.org/wp-content/uploads/2022/10/22-FDEFALCO-Evaluation-Impact-Vocabulary-Evolution-Frank-Defalco.pdf

7 Guides - SNOMED CT Document Library - SNOMED Confluence. https://confluence.ihtsdotools.org/display/doc/guides (accessed 16 Mar 2023).