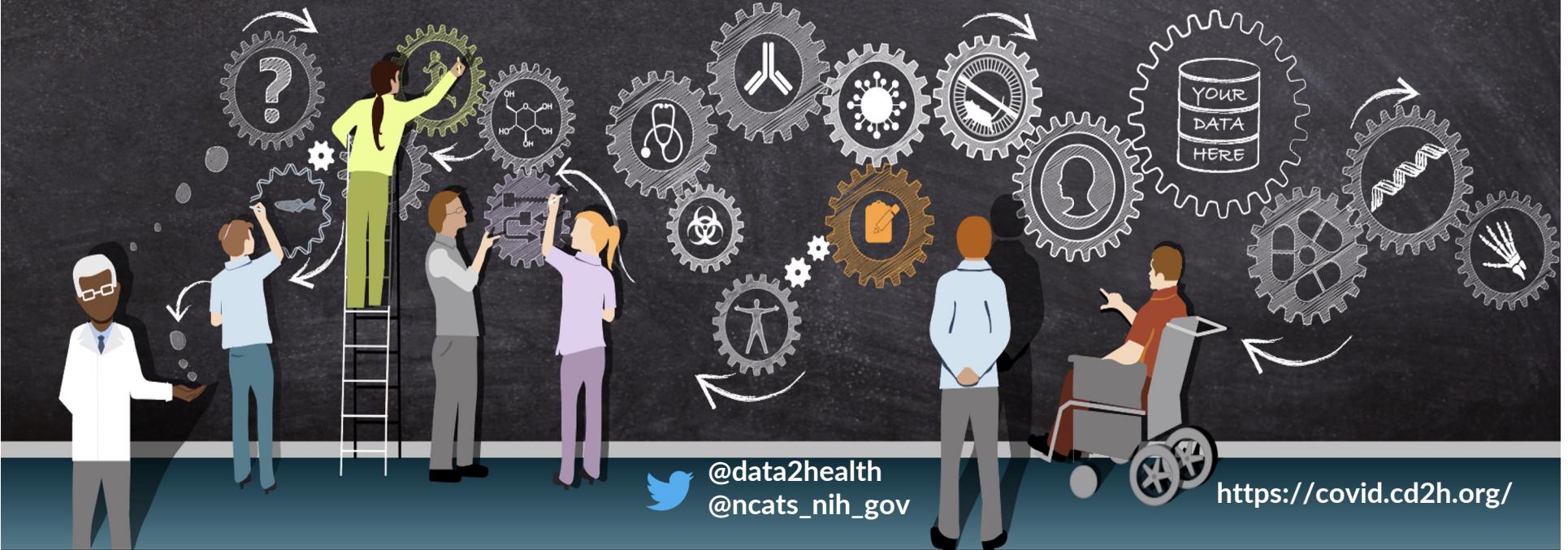


Observational Research in OHDSI, the OMOP CDM, and Standard Vocabularies

Education and Training Domain Team Short Course, Spring 2024

Jerrod Anzalone, University of Nebraska



@data2health
@ncats_nih_gov

<https://covid.cd2h.org/>



National
COVID
Cohort
Collaborative

Conventions in Use



NATIONAL CENTER
FOR DATA TO HEALTH

A note on source material and references. If I provide a reference for content in this lecture as source material, that's just a reference. If I provide a citation or link, I'm doing so because I think it's worth looking into if you want to know more.

Exhibit A: Not important

Source:

[https://google.com/
images](https://google.com/images)

Exhibit B: Maybe important or at least worth digging into if you want to know more!

Haendel MA, Chute CG, Bennett TD, et al. The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. *J Am Med Inform Assoc.* 2021;28(3):427-443.
doi:10.1093/jamia/ocaa196



National
COVID
Cohort
Collaborative

OHDSI Network Research



NATIONAL CENTER
FOR DATA TO HEALTH



National
COVID
Cohort
Collaborative

Observational Research in OHDSI



NATIONAL CENTER
FOR DATA TO HEALTH

OHDSI COLLABORATORS

Map of Collaborators

The OHDSI community brings together volunteers from around the world to establish open community data standards, develop open-source software, conduct methodological research, and apply scientific best practices to both answer public health questions and generate reliable clinical evidence.

OHDSI COLLABORATORS

Our community is ALWAYS seeking new collaborators. Do you want to focus on data standards or methodological research? Are you passionate about open-source development or clinical applications? Do you have data that you want to be part of global network studies? Do you want to be part of a global community that truly values the benefits of open science? Add a dot to the map below and JOIN THE JOURNEY!

OHDSI By The Numbers

- 3,266 collaborators
- 80 countries
- 21 time zones
- 6 continents
- 1 community

Source:

<https://www.ohdsi.org/who-we-are/collaborators/>

History of OHDSI



NATIONAL CENTER
FOR DATA TO HEALTH

Foundation and Objectives

- Established in 2014
- Focused on improving health outcomes

Evolution from OMOP

- Originated from the Observational Medical Outcomes Partnership (OMOP)
- Transition to a broader scope

Global Collaboration

- Diverse international participation
- Emphasis on open science

Technological Advancements

- Development of open-source tools to support the OMOP CDM
- Integration and maintenance of standard terminologies into a unified vocabulary

Current Focus and Future

Directions

- Emphasis on real-world evidence
- Ongoing expansions and updates

Source: <https://www.ohdsi.org/wp-content/uploads/2021/09/OHDSI-OurJourney2021-Final.pdf>

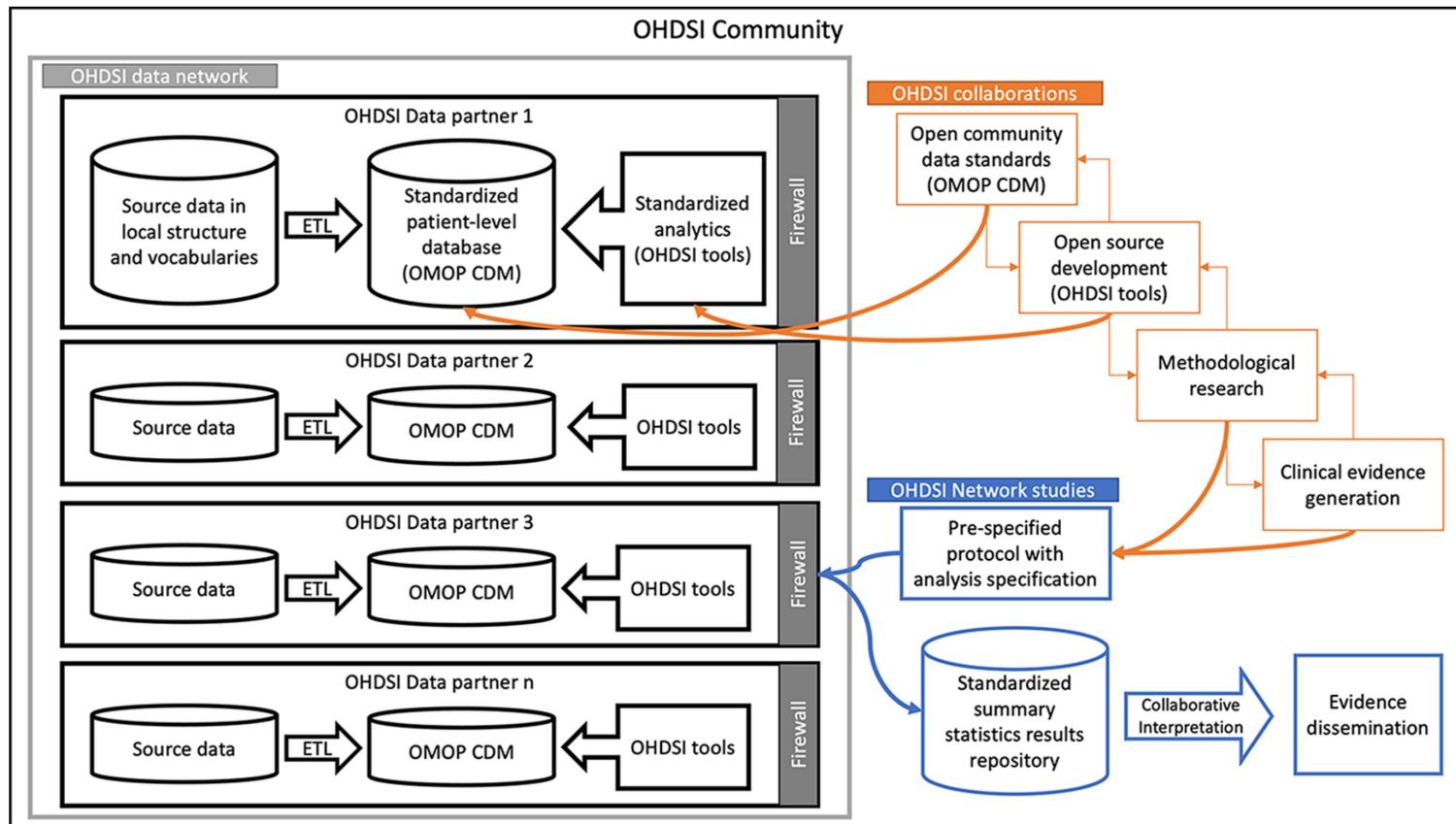


National
COVID
Cohort
Collaborative

OHDSI Research Flow



NATIONAL CENTER
FOR DATA TO HEALTH



Hripcak G, Schuemie MJ, Madigan D, Ryan PB, Suchard MA. Drawing Reproducible Conclusions from Observational Clinical Data with OHDSI. Yearb Med Inform. 2021;30(1):283-289. doi:10.1055/s-0041-1726481



National
COVID
Cohort
Collaborative

Federated vs. Centralized



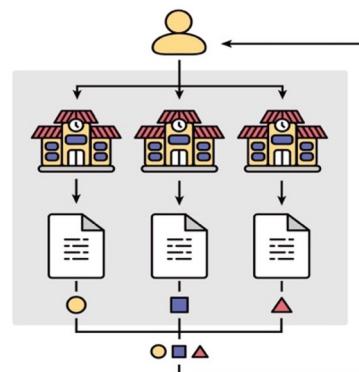
NATIONAL CENTER
FOR DATA TO HEALTH

Federated Query

Questions are
Sent to the Network

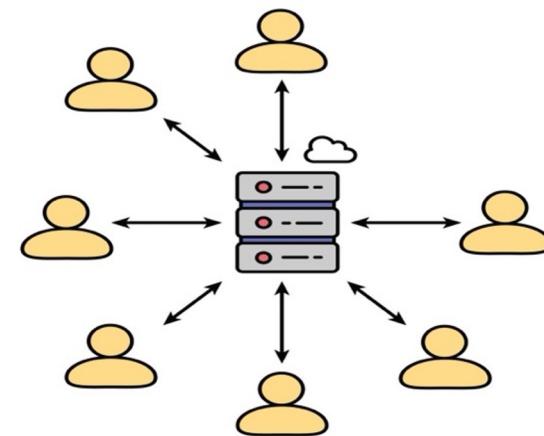


Aggregated Results are
Returned



Centralized Analytics

Data Resides Centrally in a Secure enclave





National
COVID
Cohort
Collaborative

The Debate Continues!



NATIONAL CENTER
FOR DATA TO HEALTH



OHDSI Debate #1

In order to efficiently and reliably generate robust real-world evidence across multiple data sources, observational studies are best conducted as a distributed network analysis and not a centralized data repository.



Affirmative

Kristin Kostka



Negative

Andrew Williams



@OHDSI

www.ohdsi.org

#JoinTheJourney



linkedin.com/company/ohdsi

<https://youtu.be/7bkGPd1MpCY?si=V1u3DNXL7nQeo8o>



Observational Medical Outcomes Partnership (OMOP) CDM



NATIONAL CENTER
FOR DATA TO HEALTH

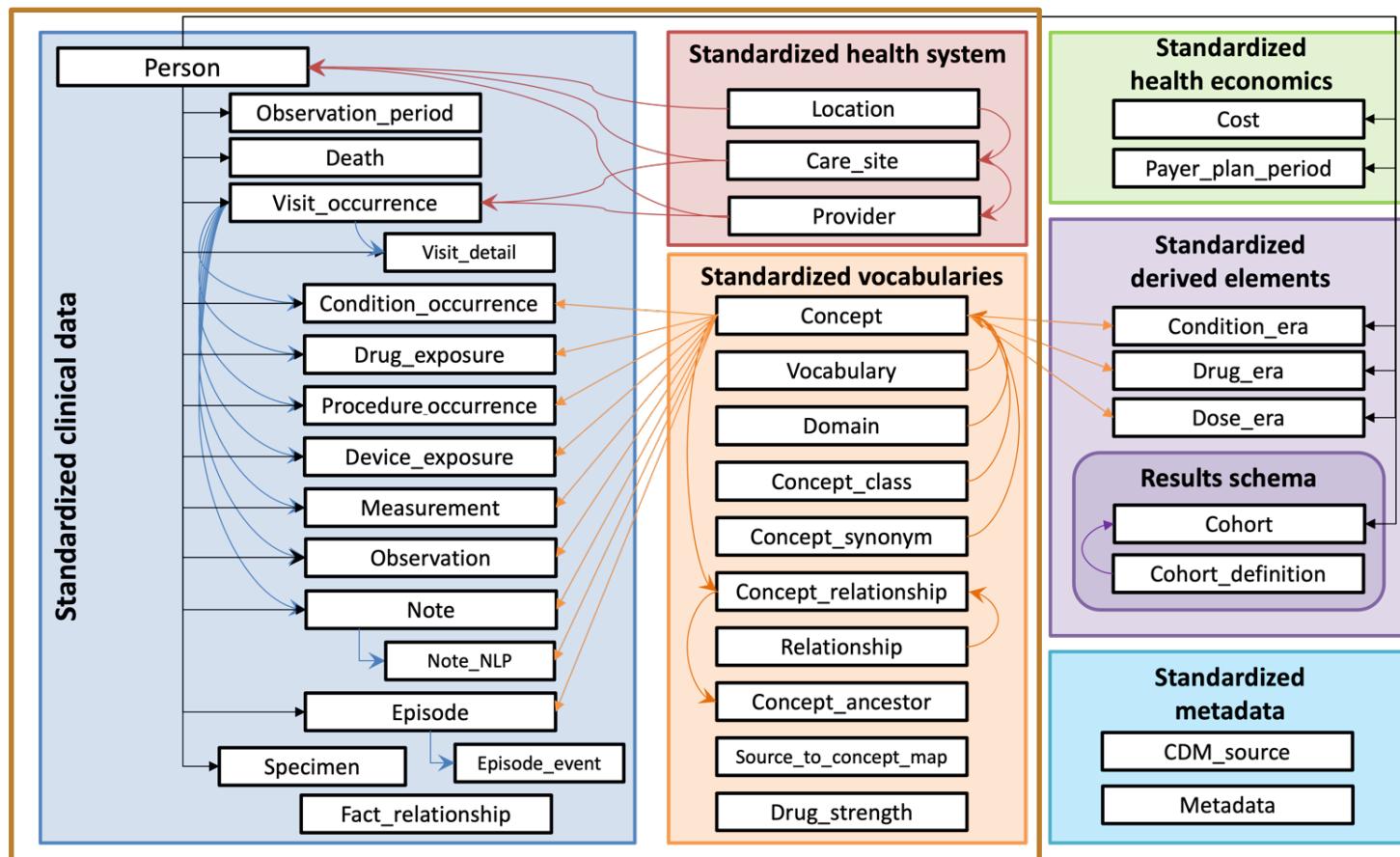


National
COVID
Cohort
Collaborative

The OMOP CDM v5.3.1



NATIONAL CENTER
FOR DATA TO HEALTH





National
COVID
Cohort
Collaborative

Helpful Resources



NATIONAL CENTER
FOR DATA TO HEALTH

OMOP CDM v5.3.1 Documentation

<https://ohdsi.github.io/CommonDataModel/cdm53.html>

Jerrod's Shiny App Containing an Interactive E-R Diagram

<https://anzalone23.shinyapps.io/n3c-short-course/>

More detailed E-R Diagram

<https://omop-erd.surge.sh/index.html>

Hopefully, the app doesn't crash! If it does, feel free to load a local copy in RStudio by cloning this repository:

<https://github.com/>

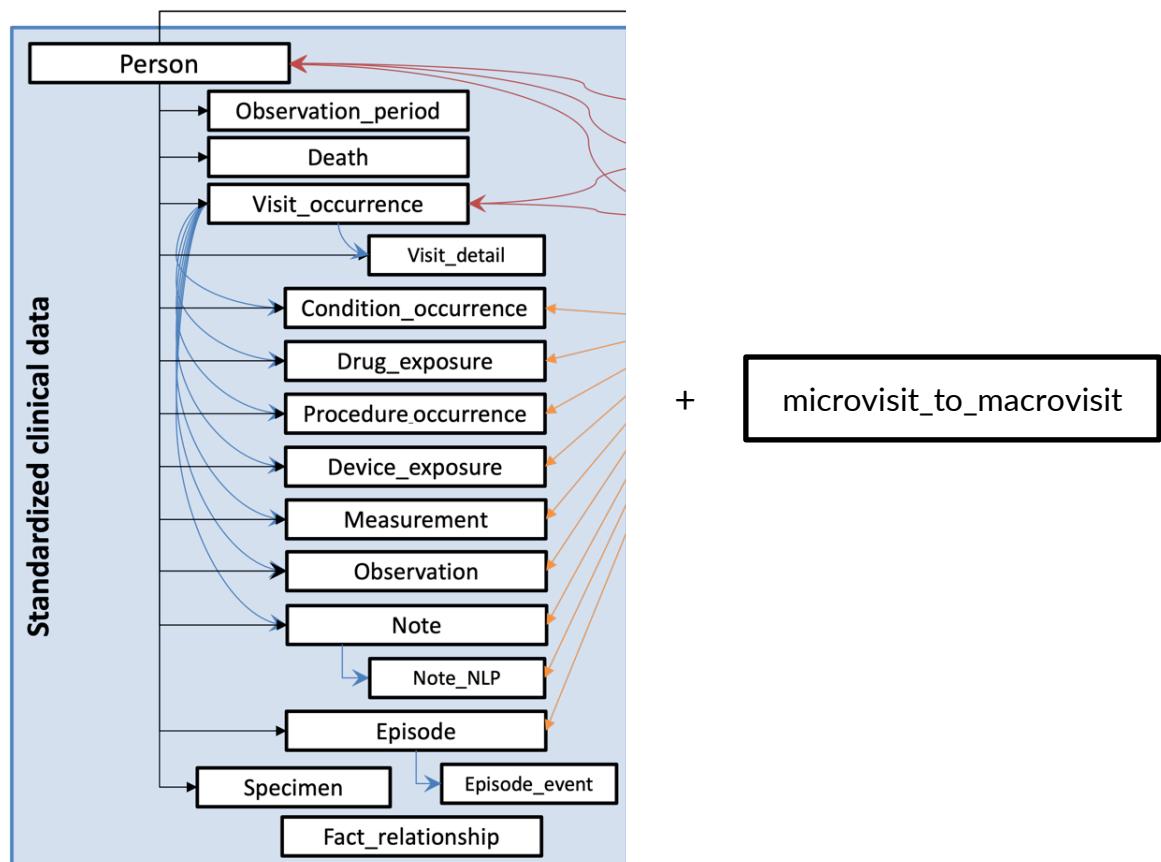


National
COVID
Cohort
Collaborative

Standardized Clinical Data



NATIONAL CENTER
FOR DATA TO HEALTH





person



PERSON		[table]
🔑	person_id	bigint[8]
🔑	gender_concept_id	int[4]
	year_of_birth	int[4]
	month_of_birth	int[4]
	day_of_birth	int[4]
	birth_datetime	datetime[16,3]
🔑	race_concept_id	int[4]
🔑	ethnicity_concept_id	int[4]
🔑	location_id	int[4]
🔑	provider_id	int[4]
🔑	care_site_id	int[4]
	person_source_value	varchar[50]
	gender_source_value	varchar[50]
🔑	gender_source_concept_id	int[4]
	race_source_value	varchar[50]
🔑	race_source_concept_id	int[4]
	ethnicity_source_value	varchar[50]
🔑	ethnicity_source_concept_id	int[4]
< 9	0 rows	18 >

Person Table

- Overview:** The Person table is the foundational table in the OMOP CDM, providing demographic information about each patient.
- Use:** Stores data like birth date, gender, race, and ethnicity, crucial for patient identification and demographic analysis.
- Standard Vocabulary:** Uses vocabularies like SNOMED CT for race and ethnicity, and custom OMOP vocabularies for gender.



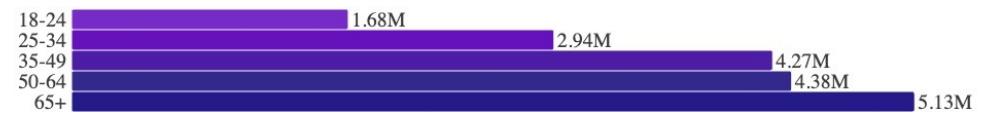
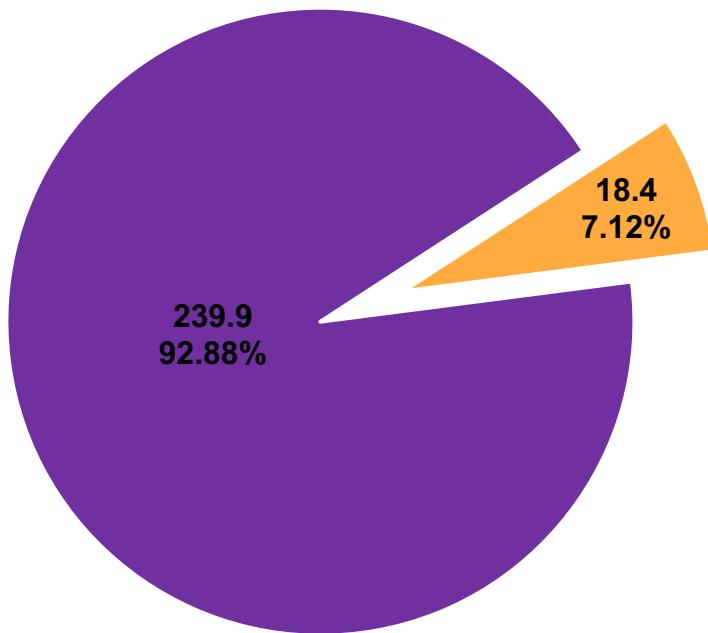
National
COVID
Cohort
Collaborative

person



NATIONAL CENTER
FOR DATA TO HEALTH

Adult Population (in million persons / %)



- N3C
- Non-N3C



visit_occurrence



VISIT_OCCURRENCE		[table]
visit_occurrence_id	int[4]	
person_id	bigint[8]	
visit_concept_id	int[4]	
visit_start_date	date[6]	
visit_start_datetime	datetime[16,3]	
visit_end_date	date[6]	
visit_end_datetime	datetime[16,3]	
visit_type_concept_id	int[4]	
provider_id	int[4]	
care_site_id	int[4]	
visit_source_value	varchar[50]	
visit_source_concept_id	int[4]	
admitted_from_concept_id	int[4]	
admitted_from_source_value	varchar[50]	
discharged_to_concept_id	int[4]	
discharged_to_source_value	varchar[50]	
preceding_visit_occurrence_id	int[4]	
< 9	0 rows	9 >

Visit Occurrence Table

- Overview:** This table logs each instance of healthcare service utilization or encounter by a patient.
- Use:** Details include visit type, dates, and location, essential for tracking patient healthcare interactions.
- Standard Vocabulary:** Utilizes vocabularies such as SNOMED CT and CPT-4 for visit types.



condition_occurrence

CONDITION_OCCURRENCE		[table]
🔑	condition_occurrence_id	int[4]
🔑	person_id	bigint[8]
🔑	condition_concept_id	int[4]
	condition_start_date	date[6]
	condition_start_datetime	datetime[16,3]
	condition_end_date	date[6]
	condition_end_datetime	datetime[16,3]
🔑	condition_type_concept_id	int[4]
🔑	condition_status_concept_id	int[4]
	stop_reason	varchar[20]
🔑	provider_id	int[4]
🔑	visit_occurrence_id	int[4]
🔑	visit_detail_id	int[4]
	condition_source_value	varchar[50]
🔑	condition_source_concept_id	int[4]
	condition_status_source_value	varchar[50]
< 8	0 rows	0 >

Condition Occurrence Table

- Overview:** This table records all diagnosed conditions and illnesses of a patient.
- Use:** Includes condition type, dates, and standard concepts, key for understanding patient health status.
- Standard Vocabulary:** Primarily uses SNOMED CT for medical conditions.



National
COVID
Cohort
Collaborative

procedure_occurrence



NATIONAL CENTER
FOR DATA TO HEALTH

PROCEDURE_OCCURRENCE [table]	
procedure_occurrence_id	int[4]
person_id	bigint[8]
procedure_concept_id	int[4]
procedure_date	date[6]
procedure_datetime	datetime[16,3]
procedure_end_date	date[6]
procedure_end_datetime	datetime[16,3]
procedure_type_concept_id	int[4]
modifier_concept_id	int[4]
quantity	int[4]
provider_id	int[4]
visit_occurrence_id	int[4]
visit_detail_id	int[4]
procedure_source_value	varchar[50]
procedure_source_concept_id	int[4]
modifier_source_value	varchar[50]
< 8	0 rows
	0 >

Procedure Occurrence Table

- Overview:** Documents all medical procedures performed on a patient.
- Use:** Captures details like procedure type and date, important for treatment and intervention analysis.
- Standard Vocabulary:** Employs vocabularies like CPT-4, SNOMED CT, and HCPCS for procedures.



device_exposure



DEVICE_EXPOSURE		[table]
device_exposure_id	int[4]	
person_id	bigint[8]	
device_concept_id	int[4]	
device_exposure_start_date	date[6]	
device_exposure_start_datetime	datetime[16,3]	
device_exposure_end_date	date[6]	
device_exposure_end_datetime	datetime[16,3]	
device_type_concept_id	int[4]	
unique_device_id	varchar[255]	
production_id	varchar[255]	
quantity	int[4]	
provider_id	int[4]	
visit_occurrence_id	int[4]	
visit_detail_id	int[4]	
device_source_value	varchar[50]	
device_source_concept_id	int[4]	
unit_concept_id	int[4]	
unit_source_value	varchar[50]	
unit_source_concept_id	int[4]	
< 9	0 rows	0 >

Device Exposure Table

- Overview:** Records instances of medical device usage for patients.
- Use:** Captures details such as the type of device, the period of use, and the quantity for analyzing the impact and patterns of medical device utilization in patient care.
- Standard Vocabulary:** For identifying devices, this table predominantly utilizes standard vocabularies like SNOMED CT and occasionally HCPCS.



drug_exposure



DRUG_EXPOSURE		[table]
drug_exposure_id	int[4]	
person_id	bigint[8]	
drug_concept_id	int[4]	
drug_exposure_start_date	date[6]	
drug_exposure_start_datetime	datetime[16,3]	
drug_exposure_end_date	date[6]	
drug_exposure_end_datetime	datetime[16,3]	
verbatim_end_date	date[6]	
drug_type_concept_id	int[4]	
stop_reason	varchar[20]	
refills	int[4]	
quantity	float[8]	
days_supply	int[4]	
sig	varchar[2147483647]	
route_concept_id	int[4]	
lot_number	varchar[50]	
provider_id	int[4]	
visit_occurrence_id	int[4]	
visit_detail_id	int[4]	
drug_source_value	varchar[50]	
drug_source_concept_id	int[4]	
route_source_value	varchar[50]	
dose_unit_source_value	varchar[50]	
< 8	0 rows	0 >

Drug Exposure Table

- Overview:** Tracks the exposure of patients to various drugs and medications.
- Use:** Details include drug type, exposure period, and dosage, crucial for medication effect studies.
- Standard Vocabulary:** Primarily uses RxNorm for drug names and types.



measurement



MEASUREMENT		[table]
measurement_id	int[4]	
person_id	bigint[8]	
measurement_concept_id	int[4]	
measurement_date	date[6]	
measurement_datetime	datetime[16,3]	
measurement_time	varchar[10]	
measurement_type_concept_id	int[4]	
operator_concept_id	int[4]	
value_as_number	float[8]	
value_as_concept_id	int[4]	
unit_concept_id	int[4]	
range_low	float[8]	
range_high	float[8]	
provider_id	int[4]	
visit_occurrence_id	int[4]	
visit_detail_id	int[4]	
measurement_source_value	varchar[50]	
measurement_source_concept_id	int[4]	
unit_source_value	varchar[50]	
unit_source_concept_id	int[4]	
value_source_value	varchar[50]	
measurement_event_id	bigint[8]	
meas_event_field_concept_id	int[4]	
< 12	0 rows	0 >

Measurement Table

- Overview:** Contains quantitative measurements and lab test results for patients.
- Use:** Includes measurement type, value, and date, vital for tracking patient health metrics.
- Standard Vocabulary:** Utilizes LOINC for lab tests and measurements.



observation

OBSERVATION		[table]
observation_id	int[4]	
person_id	bigint[8]	
observation_concept_id	int[4]	
observation_date	date[6]	
observation_datetime	datetime[16,3]	
observation_type_concept_id	int[4]	
value_as_number	float[8]	
value_as_string	varchar[60]	
value_as_concept_id	int[4]	
qualifier_concept_id	int[4]	
unit_concept_id	int[4]	
provider_id	int[4]	
visit_occurrence_id	int[4]	
visit_detail_id	int[4]	
observation_source_value	varchar[50]	
observation_source_concept_id	int[4]	
unit_source_value	varchar[50]	
qualifier_source_value	varchar[50]	
value_source_value	varchar[50]	
observation_event_id	bigint[8]	
obs_event_field_concept_id	int[4]	
< 11	0 rows	0 >

Observation Table

- Overview:** A versatile (messy) table capturing a variety of clinical and non-clinical observations.
- Use:** Stores diverse data like symptoms, lifestyle factors, and clinician observations.
- Standard Vocabulary:** Uses vocabularies like SNOMED CT for diverse clinical observations, but it's all over the place.



National
COVID
Cohort
Collaborative

Special Table: observation_period



NATIONAL CENTER
FOR DATA TO HEALTH

OBSERVATION_PERIOD		[table]
observation_period_id		int[4]
person_id		bigint[8]
observation_period_start_date		date[6]
observation_period_end_date		date[6]
period_type_concept_id		int[4]
< 2	0 rows	0 >

Observation Table

- Overview:** The Observation Period table in the OMOP CDM is dedicated to defining the timeframes during which a patient's healthcare data is actively captured and monitored.
- Use:** It records the start and end dates of observation for each patient, delineating the specific period for which the patient's data is relevant and can be reliably used for analysis. This is crucial for ensuring that analyses and studies only utilize data from periods where complete information is available.

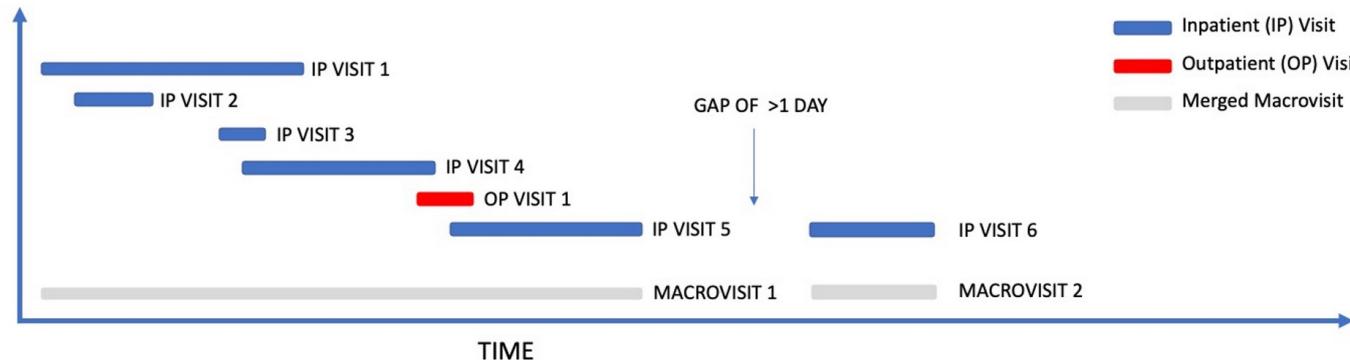


National
COVID
Cohort
Collaborative

Special Table: macrovisits!



NATIONAL CENTER
FOR DATA TO HEALTH



Journal of the American Medical Informatics Association, 30(6), 2023, 1125–1136
https://doi.org/10.1093/jamia/ocad057
Advance Access Publication Date: 22 April 2023
Research and Applications



Read more here:

<https://pubmed.ncbi.nlm.nih.gov/37087110/>

Research and Applications

Clinical encounter heterogeneity and methods for resolving in networked EHR data: a study from N3C and RECOVER programs

Peter Leesse ¹, Adit Anand ¹, Andrew Girvin ¹, Amin Manna ¹, Saaya Patel ¹, Yun Jae Yoo ¹, Rachel Wong ¹, Melissa Haendel ¹, Christopher G. Chute ², Tellen Bennett ³, Janos Hajagos ¹, Emily Pfaff ¹, and Richard Moffitt ^{1,2,3}

¹NC TraCS Institute, UNC-School of Medicine, Chapel Hill, North Carolina, USA, ²Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY, USA, ³Department of Biostatistics, School of Public Health, University of North Carolina at Chapel Hill, NC, USA



National
COVID
Cohort
Collaborative

Standardized Health System Data



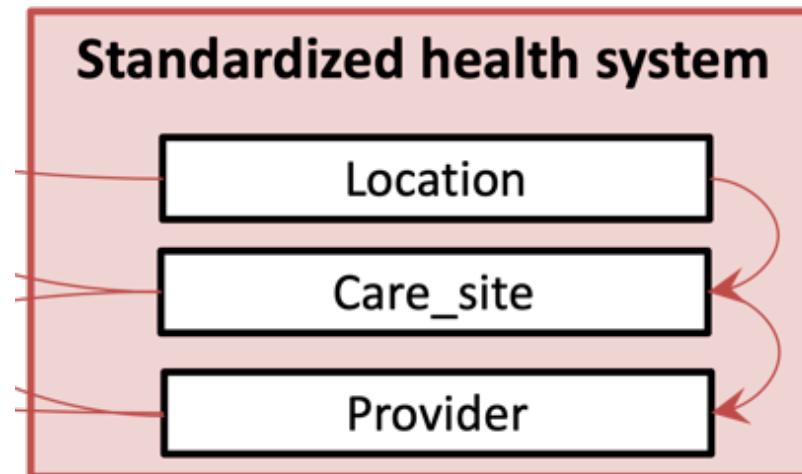
NATIONAL CENTER
FOR DATA TO HEALTH

Standardized health system

Location

Care_site

Provider





Standardized Health System Data



NATIONAL CENTER
FOR DATA TO HEALTH



location



LOCATION [table]	
📍 location_id	int[4]
address_1	varchar[50]
address_2	varchar[50]
city	varchar[50]
state	varchar[2]
zip	varchar[9]
county	varchar[20]
location_source_value	varchar[50]
🔑 country_concept_id	int[4]
country_source_value	varchar[80]
latitude	float[8]
longitude	float[8]
< 1	0 rows
	2 >

*Not in N3C

Location Table

- Overview:** Stores information about geographical locations for Persons and Care Sites.
- Use:** Includes data such as address, city, state, and ZIP code. This table is key for demographic studies and understanding geographical variations in healthcare delivery and outcomes.



National
COVID
Cohort
Collaborative

care_site



NATIONAL CENTER
FOR DATA TO HEALTH

CARE_SITE		[table]
🔑 care_site_id		int[4]
care_site_name		varchar[255]
🔑 place_of_service_concept_id		int[4]
🔑 location_id		int[4]
care_site_source_value		varchar[50]
place_of_service_source_value		varchar[50]
< 2	0 rows	4 >

Care Site Table:

- **Overview:** This table records information about the locations where care is provided.
- **Use:** It includes details like the care site's name, type, and relationship to other entities in the healthcare system. It's vital for understanding where patients receive care and for linking care events to specific locations.
- Note that a lot of the general utility of this table is not available in N3C due to site privacy requirements built into the N3C DTA.



National
COVID
Cohort
Collaborative

provider



NATIONAL CENTER
FOR DATA TO HEALTH

PROVIDER		[table]
provider_id	int[4]	
provider_name	varchar[255]	
npi	varchar[20]	
dca	varchar[20]	
specialty_concept_id	int[4]	
care_site_id	int[4]	
year_of_birth	int[4]	
gender_concept_id	int[4]	
provider_source_value	varchar[50]	
specialty_source_value	varchar[50]	
specialty_source_concept_id	int[4]	
gender_source_value	varchar[50]	
gender_source_concept_id	int[4]	
< 5	0 rows	10 >

Provider Table:

- **Overview:** Documents information about individual healthcare providers.
- **Use:** Captures details like provider IDs, names, specialties, and affiliations. This table is crucial for associating healthcare events and services with specific providers, offering insights into who is delivering care and their qualifications.
- Note that a lot of the general utility of this table is not available in N3C due to site privacy requirements built into the N3C DTA.

*Not in N3C



National
COVID
Cohort
Collaborative

Special Table: manifest!



NATIONAL CENTER
FOR DATA TO HEALTH

“This table, found in the folder along with the OMOP tables, provides data about each data partner that analysts should find helpful when trying to understand variation in the data: CDM type (ACT, OMOP, PCORNet, TriNetX) and version, whether dates are shifted (and how much), and date of last contribution.”

[~Guide to N3C, Chapter 7](#)



National
COVID
Cohort
Collaborative

Standardized Vocabularies



NATIONAL CENTER
FOR DATA TO HEALTH



concept



CONCEPT [table]	
concept_id	int[4]
concept_name	varchar[255]
domain_id	varchar[20]
vocabulary_id	varchar[20]
concept_class_id	varchar[20]
standard_concept	varchar[1]
concept_code	varchar[50]
valid_start_date	date[6]
valid_end_date	date[6]
invalid_reason	varchar[1]
< 3	0 rows
	118 >

Concept Table:

- **Overview:** This is the central table in the vocabulary component of the OMOP CDM. It stores standardized definitions for a wide range of medical terms, concepts, and classifications.
- **Use:** It includes concepts for diseases, drugs, procedures, devices, and more, each assigned a unique concept ID. This table is essential for mapping disparate data sources to a common language.



National
COVID
Cohort
Collaborative

concept



NATIONAL CENTER
FOR DATA TO HEALTH

DETAILS

Domain ID	Condition
Concept Class ID	Clinical Finding
Vocabulary ID	SNOMED
Concept ID	201820
Concept code	73211009
Validity	Valid
Concept	Standard
Synonyms	DM - Diabetes mellitus Diabetes mellitus (disorder)
Valid start	31-Jan-2002
Valid end	31-Dec-2099

This will be found in the
condition_occurrence
table



National
COVID
Cohort
Collaborative

concept



NATIONAL CENTER
FOR DATA TO HEALTH

DETAILS

Domain ID	Condition
Concept Class ID	Clinical Finding
Vocabulary ID	SNOMED
Concept ID	201820 ←
Concept code	73211009
Validity	Valid
Concept	Standard
Synonyms	DM - Diabetes mellitus Diabetes mellitus (disorder)
Valid start	31-Jan-2002
Valid end	31-Dec-2099

This is a unique id
assigned to this concept
in the OMOP vocabulary



National
COVID
Cohort
Collaborative

concept



NATIONAL CENTER
FOR DATA TO HEALTH

DETAILS

Domain ID	Condition
Concept Class ID	Clinical Finding
Vocabulary ID	SNOMED
Concept ID	201820
Concept code	73211009
Validity	Valid
Concept	Standard
Synonyms	DM - Diabetes mellitus Diabetes mellitus (disorder)
Valid start	31-Jan-2002
Valid end	31-Dec-2099

This is the unique identifier for the concept in the source vocabulary (e.g., SNOMED CT)



National
COVID
Cohort
Collaborative

concept



NATIONAL CENTER
FOR DATA TO HEALTH

DETAILS

Domain ID	Condition
Concept Class ID	Clinical Finding
Vocabulary ID	SNOMED
Concept ID	201820
Concept code	73211009
Validity	Valid
Concept	Standard
Synonyms	DM - Diabetes mellitus Diabetes mellitus (disorder)
Valid start	31-Jan-2002
Valid end	31-Dec-2099

This indicates whether
this concept is a standard
concept or not



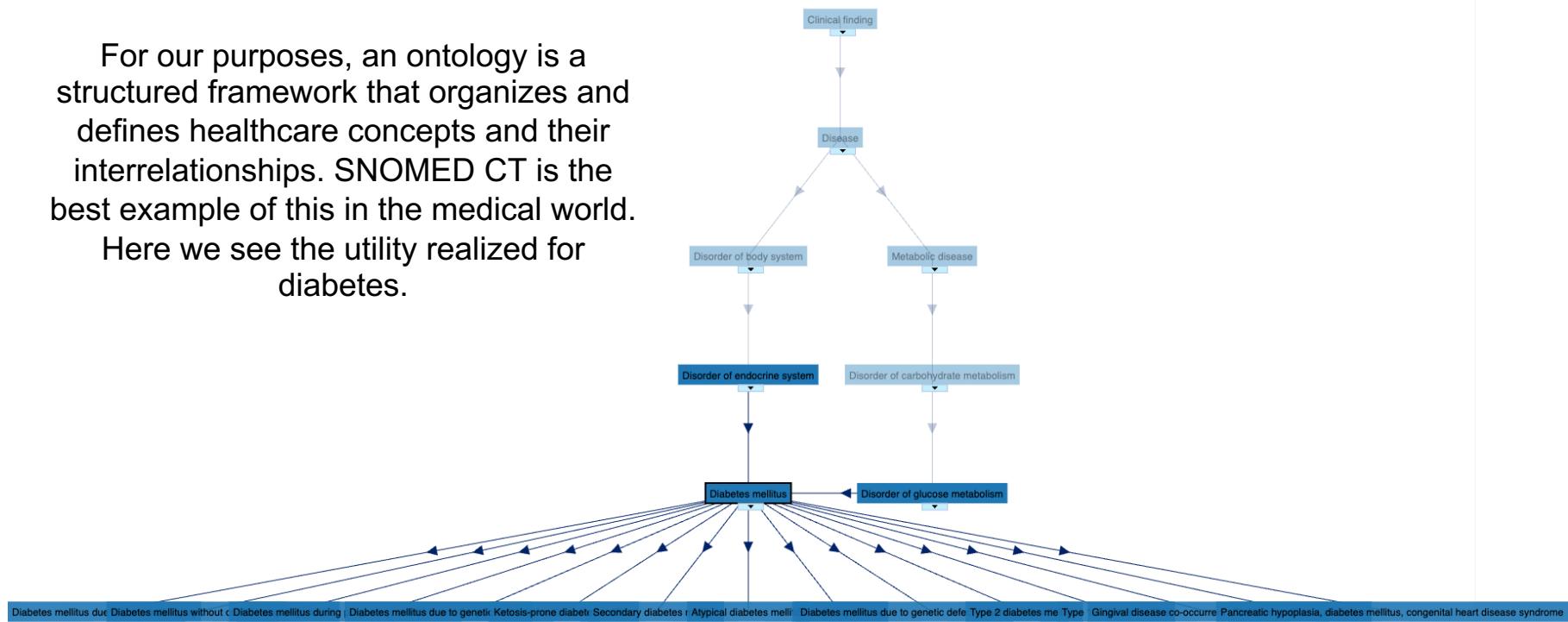
National
COVID
Cohort
Collaborative

concept



NATIONAL CENTER
FOR DATA TO HEALTH

For our purposes, an ontology is a structured framework that organizes and defines healthcare concepts and their interrelationships. SNOMED CT is the best example of this in the medical world.
Here we see the utility realized for diabetes.



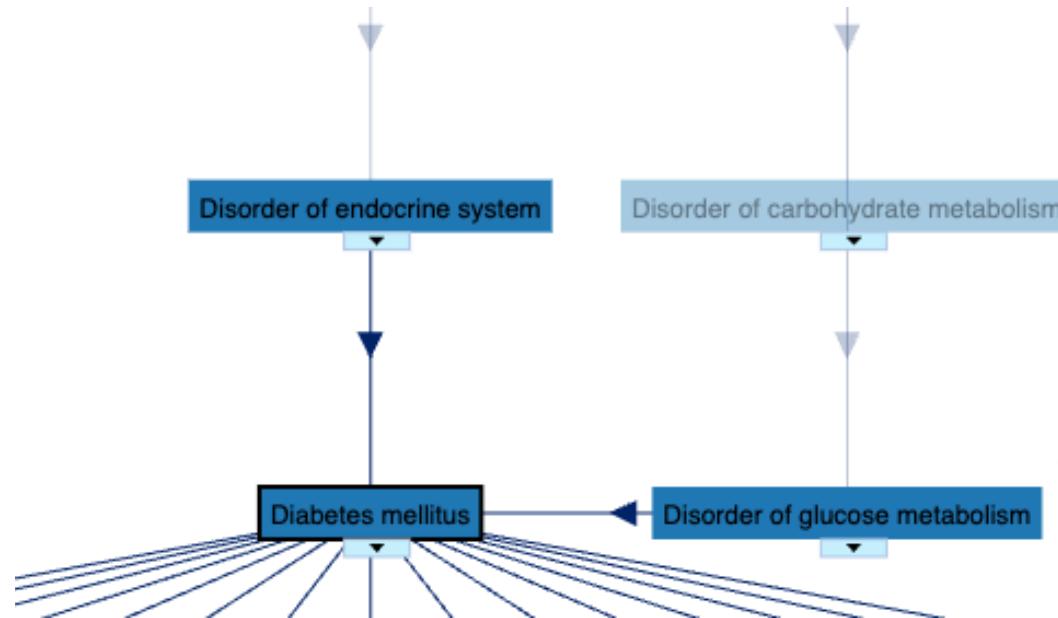


National
COVID
Cohort
Collaborative

concept



NATIONAL CENTER
FOR DATA TO HEALTH



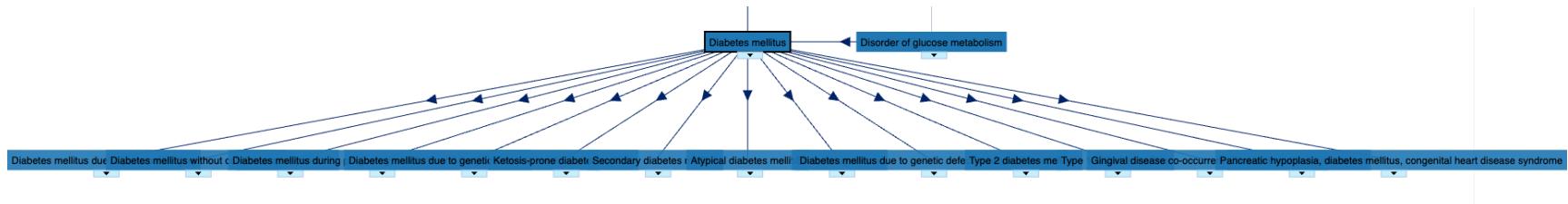


National
COVID
Cohort
Collaborative

concept



NATIONAL CENTER
FOR DATA TO HEALTH



There are 124 descendant concepts in total
from the ancestor diabetes mellitus

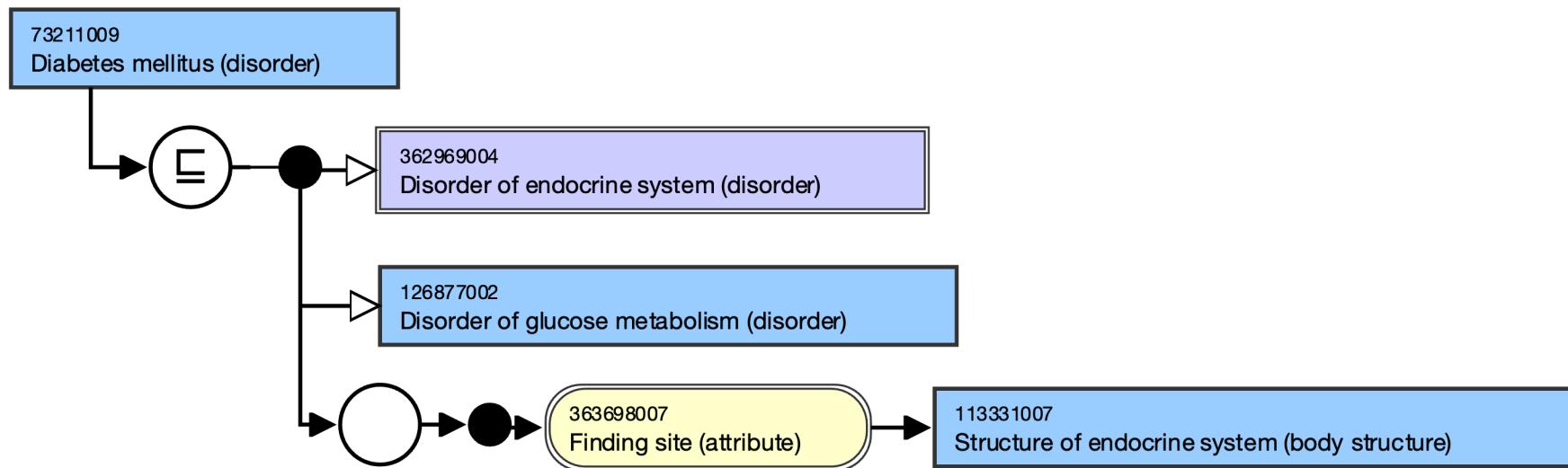


National
COVID
Cohort
Collaborative

concept



NATIONAL CENTER
FOR DATA TO HEALTH





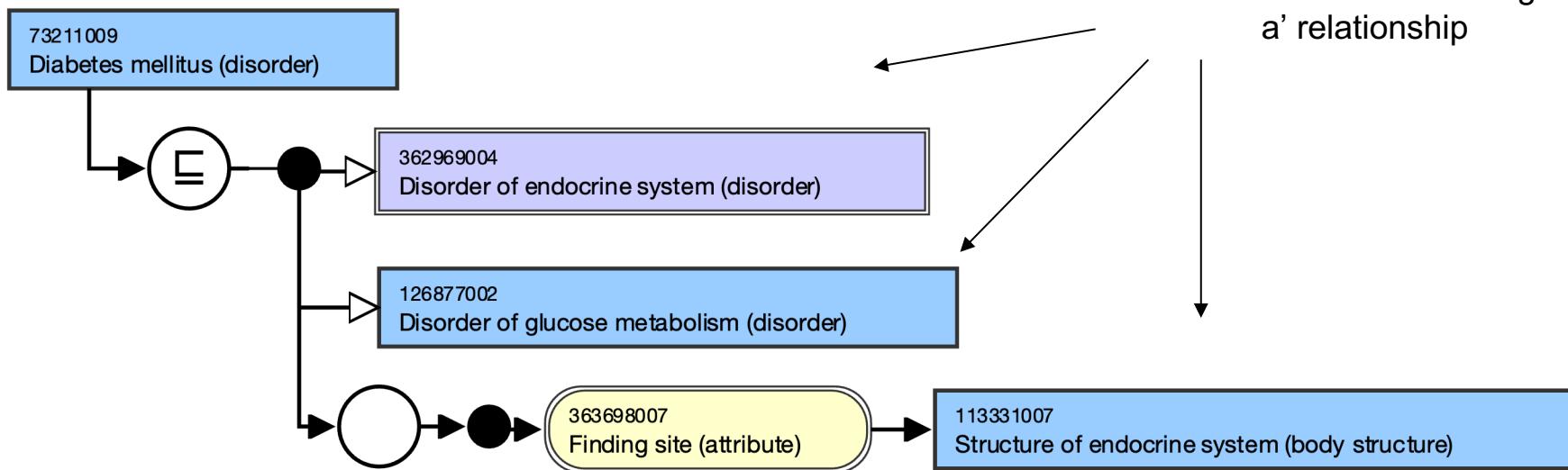
National
COVID
Cohort
Collaborative

concept



NATIONAL CENTER
FOR DATA TO HEALTH

You can query any of the metadata to find related disorders using the 'is a' relationship





National
COVID
Cohort
Collaborative

concept_class



NATIONAL CENTER
FOR DATA TO HEALTH

CONCEPT_CLASS [table]		
concept_class_id	varchar[20]	
concept_class_name	varchar[255]	
concept_class_concept_id	int[4]	
< 1	0 rows	1 >

Concept Class Table

- Overview:** This table categorizes concepts into classes, grouping similar types of concepts together.
- Use:** Helps in organizing and filtering concepts based on their class, like distinguishing between medication and procedure codes.



National
COVID
Cohort
Collaborative

domain



NATIONAL CENTER
FOR DATA TO HEALTH

DOMAIN		[table]
domain_id	varchar[20]	
domain_name	varchar[255]	
domain_concept_id	int[4]	
< 1	0 rows	2 >

Domain Table:

- **Overview:** It assigns each concept to a domain, indicating the broad category of healthcare data it belongs to, such as 'Drug' or 'Condition'.
- **Use:** Helps determining the context in which a concept is used so you can effectively identify which table the concept will be used in when querying against clinical tables

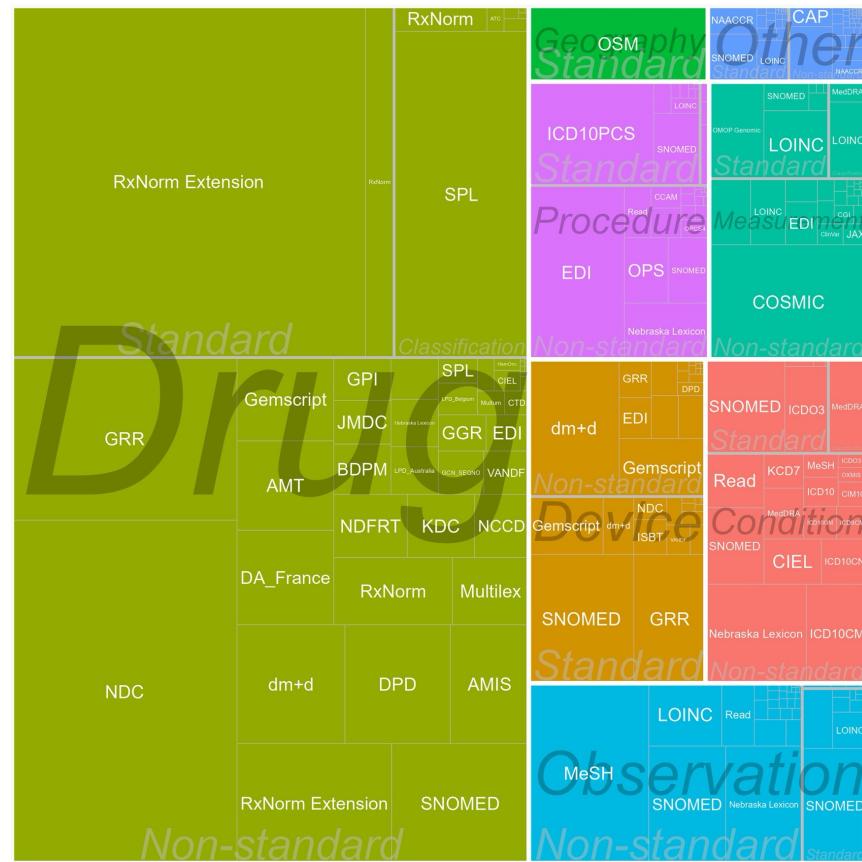


National
COVID
Cohort
Collaborative

domain



NATIONAL CENTER
FOR DATA TO HEALTH



Reich C, Ostropolets A, Ryan P, et al. OHDSI Standardized Vocabularies-a large-scale centralized reference ontology for international data harmonization. *J Am Med Inform Assoc.* Published online January 4, 2024.
doi:10.1093/jamia/ocad247



National
COVID
Cohort
Collaborative



NATIONAL CENTER
FOR DATA TO HEALTH

vocabulary

VOCABULARY [table]	
vocabulary_id	varchar[20]
vocabulary_name	varchar[255]
vocabulary_reference	varchar[255]
vocabulary_version	varchar[255]
vocabulary_concept_id	int[4]
< 1	0 rows
	2 >

Vocabulary Table:

- **Overview:** This table lists the vocabularies or coding systems used in the healthcare data, like ICD-10, SNOMED CT, or RxNorm.
- **Use:** It helps to identify which coding system a particular concept comes from, facilitating the understanding and management of various medical terminologies used across datasets.



National
COVID
Cohort
Collaborative



NATIONAL CENTER
FOR DATA TO HEALTH

Standard Vocabularies

Standard Terminology	Associated Domain(s)	Maintaining Organization	Update Frequency	Overview
SNOMED CT	Conditions, Observations, Devices	SNOMED International	Biannually	A comprehensive clinical terminology, providing a standardized way to represent clinical phrases captured by healthcare providers.
RxNorm	Drugs	National Library of Medicine (NLM)	Monthly	A normalized naming system for generic and branded drugs, centralizing drug information from various US sources.
LOINC	Measurements, Observations	Regenstrief Institute	Biannually	A common language for clinical and laboratory observations, widely used for lab tests and other measurements.
ICD-10	Procedures, Conditions (non-standard)	World Health Organization (WHO)	Annually	The 10th revision of the International Classification of Diseases, used for morbidity and mortality statistics and billing.
CPT-4	Procedures, Observations	American Medical Association (AMA)	Annually	A set of codes used by healthcare professionals to report medical, surgical, and diagnostic procedures.
HCPCS	Procedures, Devices, Observations	Centers for Medicare & Medicaid Services (CMS)	Annually	The Healthcare Common Procedure Coding System, used for processing outpatient claims by Medicare and other health insurers.
CMS Place of Service	Visit, Provider	Centers for Medicare & Medicaid Services (CMS)	As needed	A set of codes used to visit location (e.g., ambulatory surgical center), physician specialty (e.g., anesthesiology), or provider type for non-physicians (e.g., social worker)



National
COVID
Cohort
Collaborative



NATIONAL CENTER
FOR DATA TO HEALTH

Standard Vocabularies

ATHENA				SEARCH	DOWNLOAD	Alfred Anzalone ▾	?
<input type="checkbox"/>	ID (CDM V4.5)	CODE (CDM V5)	NAME	REQUIRED	SHOW HISTORY	DOWNLOAD VOCABULARIES	
<input checked="" type="checkbox"/>	1	SNOMED	Systematic Nomenclature of Medicine - Clinical Terms (IHTSDO)			27-Jan-2022	
<input checked="" type="checkbox"/>	2	ICD9CM	International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 1 and 2 (NCHS)			30-Sep-2014	
<input checked="" type="checkbox"/>	3	ICD9Proc	International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 3 (NCHS)			30-Sep-2014	
<input checked="" type="checkbox"/>	4	CPT4	Current Procedural Terminology version 4 (AMA)	EULA required		30-Apr-2023	
<input checked="" type="checkbox"/>	5	HCPGS	Healthcare Common Procedure Coding System (CMS)			30-Jun-2023	
<input checked="" type="checkbox"/>	6	LOINC	Logical Observation Identifiers Names and Codes (Regenstrief Institute)			14-Aug-2023	
<input type="checkbox"/>	7	NDFRT	National Drug File - Reference Terminology (VA)			05-Aug-2018	
<input checked="" type="checkbox"/>	8	RxNorm	RxNorm (NLM)			02-Jul-2023	
<input checked="" type="checkbox"/>	9	NDC	National Drug Code (FDA and manufacturers)			26-Aug-2023	
	10	GPI	Medi-Span Generic Product Identifier (Wolters Kluwer Health)	License required		13-Dec-2017	
<input checked="" type="checkbox"/>	12	Gender	OMOP Gender				
<input checked="" type="checkbox"/>	13	Race	Race and Ethnicity Code Set (USBC)				
<input checked="" type="checkbox"/>	14	CMS Place of Service	Place of Service Codes for Professional Claims (CMS)				
	15	MedDRA	Medical Dictionary for Regulatory Activities (MSSO)	EULA required		28-Feb-2022	
<input type="checkbox"/>	16	Multum	Cerner Multum (Cerner)				
<input type="checkbox"/>	17	Read	NHS UK Read Codes Version 2 (HSCIC)			02-Apr-2018	
<input type="checkbox"/>	18	OXMIS	Oxford Medical Information System (OCHP)			26-Apr-2015	
	19	Indication	Indications and Contraindications (FDB)	License required		18-Nov-2015	
	20	ETC	Enhanced Therapeutic Classification (FDB)	License required		18-Nov-2015	
<input checked="" type="checkbox"/>	21	ATC	WHO Anatomic Therapeutic Chemical Classification			06-Sep-2021	



National
COVID
Cohort
Collaborative

concept_relationship



NATIONAL CENTER
FOR DATA TO HEALTH

CONCEPT_RELATIONSHIP		[table]
concept_id_1	int[4]	
concept_id_2	int[4]	
relationship_id	varchar[20]	
valid_start_date	date[6]	
valid_end_date	date[6]	
invalid_reason	varchar[1]	
< 3	0 rows	0 >

Concept Relationship Table

- **Overview:** It contains records that define relationships between concepts, showing how different concepts are connected or related.
- **Use:** This is useful for understanding the hierarchy or linkage between concepts, like how a specific drug is related to its broader drug class.



National
COVID
Cohort
Collaborative

concept_ancestor



NATIONAL CENTER
FOR DATA TO HEALTH

CONCEPT_ANCESTOR		[table]
ancestor_concept_id		int[4]
descendant_concept_id		int[4]
min_levels_of_separation		int[4]
max_levels_of_separation		int[4]
< 2	0 rows	0 >

Concept Ancestor Table

- **Overview:** Designed to provide historical relationships between concepts over time, showing how they evolve or get replaced.
- **Use:** This is especially important for longitudinal studies where concepts might change, ensuring continuity in data interpretation.

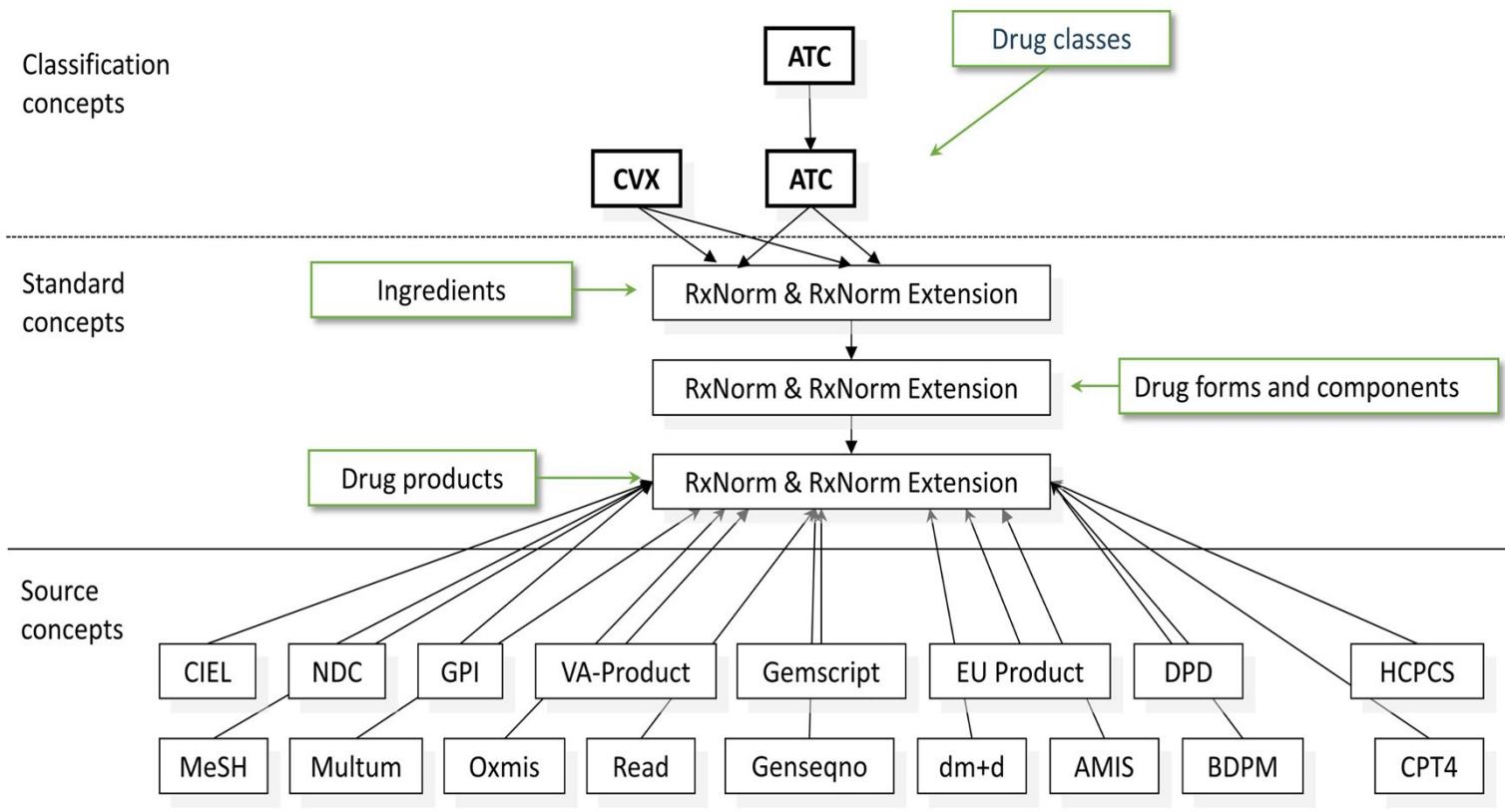


National
COVID
Cohort
Collaborative

Medication Example



NATIONAL CENTER
FOR DATA TO HEALTH



Reich C,
Ostropolets A,
Ryan P, et al.
OHDSI
Standardized
Vocabularies-a
large-scale
centralized
reference
ontology for
international data
harmonization. *J
Am Med Inform
Assoc.* Published
online January 4,
2024.
doi:10.1093/jami
a/ocad247



National
COVID
Cohort
Collaborative

Let's Try it Out!



NATIONAL CENTER
FOR DATA TO HEALTH

Jerrod's Shiny app containing a small (~0.1%) subset of
the OMOP vocabulary tables

<https://anzalone23.shinyapps.io/n3c-short-course/>

Exercises (do in class or offline depending on time):

https://github.com/National-COVID-Cohort-Collaborative/short-course-2024-january/tree/main/sessions/session-2/session2_exercises.docx