

Mapping OMOP Data to SDTM User Case for RWD in Registry Study

Joe Xi, Kevin Deng, Lorraine Fang

Mar 2025

Disclaimer

This presentation shares only the author's experience and opinions, and not necessarily those of Bristol Myers Squibb (BMS). BMS does not guarantee the accuracy or reliability of the information provided in the slides.

Outline

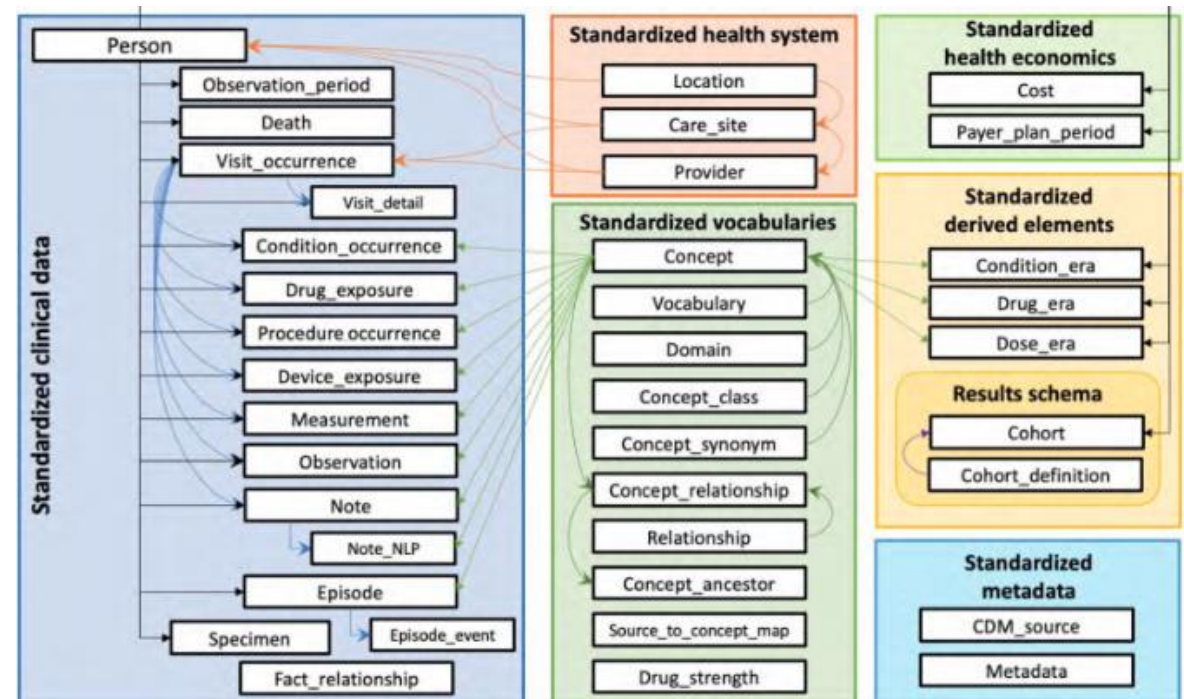
- **Using OMOP CDM in Pharma Industry**
- **How our OMOP Data are Organized**
- **How to Convert/Remap OMOP into SDTM-like**
- **Some Special Concerning**
- **Take-away**

Using OMOP CDM in Pharma Industry

- Multiple models (e.g. CDISC, FHIR, OMOP) proposed for RWD
- CDISC still preferred
- CDISC NOT convenient for RWD collection/mapping
- We plan to convert registry data from OMOP to SDTM-like
- Share experience/lesson-learned to contribute industry

How our OMOP Data Organized

- OMOP: open-source model to harmonize disparate healthcare data sources into one relational database with standardized tables and vocabularies
- Data split into tables by nature
- Decoding by linking **xx_concept_id** value in **Concept** or **OHDSI website**
- Non-standard table **form_option** used for supplemental information



How our OMOP Data Organized

Sample of spec and data for *Person* and *Observation* tables:

Person:	
Column	Description
person_id	A unique ID for every person in the database
gender_concept_id	A concept_id code for the gender or sex of the person at birth
year_of_birth	Year in which person was born
month_of_birth	Month in which person was born
day_of_birth	Day on which person was born
birth_datetime	Not populated
race_concept_id	A concept_id to code the race of the person
ethnicity_concept_id	A concept_id to code for the ethnicity of the person
location_id	ID for the location that links to the location_id in the location table
provider_id	Not populated
care_site_id	ID for the care site that links to the care_site_id in the care_site table
person_source_value	An ID that links back to a unique identifier in the source database from which OMOP is populated
gender_source_value	A value from the source database representing gender or sex at birth
race_source_value	Not populated
ethnicity_source_value	Not populated
ethnicity_source_concept_id	Not populated
donor	Not standard OMOP-CDM, implemented for the donor outcome registry. TRUE / FALSE indicating whether the person was a donor.

Observation:	
Column	Description
observation_id	This column contains the id by which this observation is known in database
observation_concept_id	This column contains the a concept_id coding for the observation
observation_date	This column contains the date of the observation, most likely the event date of the event the observation was recorded on
observation_datetime	Not populated
observation_type_concept_id	This column indicates where the data was extracted from.
value_as_number	The numerical value of the observation that was done, e.g. the chronological number of an HCT
value_as_string	This column can contain non-numerical data pertaining to the observation, providing flexibility to give extra information when the value_as_concept_id or value_as_number might benefit from additional context.
value_as_concept_id	The concept id of the observation if it was a categorical observation, e.g. the donor's sex at birth in an allogeneic HCT
qualifier_concept_id	Not populated
unit_concept_id	This column contains the unit of the observation coded as a concept_id, if applicable
observation_source_value	This column contains a reference to the form, field and value that created the particular row in the table
qualifier_source_value	Not populated
observation_event_id	Not populated
person_id	This column can be used to link the observation to the patient using the person table
visit_detail_id	Not populated
visit_occurrence_id	Not populated
provider_id	Not populated
form_option_id	Non-OMOP-CDM column. This column contains the form option ID, that can be used to identify the form option using the form_option table
patient_event_id	Non-OMOP-CDM column. This column contains the identifier of the patient event in the application
group_index	Non-OMOP-CDM column. Where the data row comes from a repeater group in the form, this will increment as repeater groups are added.

Person:

person_id	gender_concept_id	location_id	provider_id	care_site_id	gender_source_value	gender_source_concept_id	donor
10001	8532	1		601	1		FALSE
10002	8507	2		602	2		FALSE
10003	8532	3		603	1		FALSE
10004	8507	4		604	2		FALSE

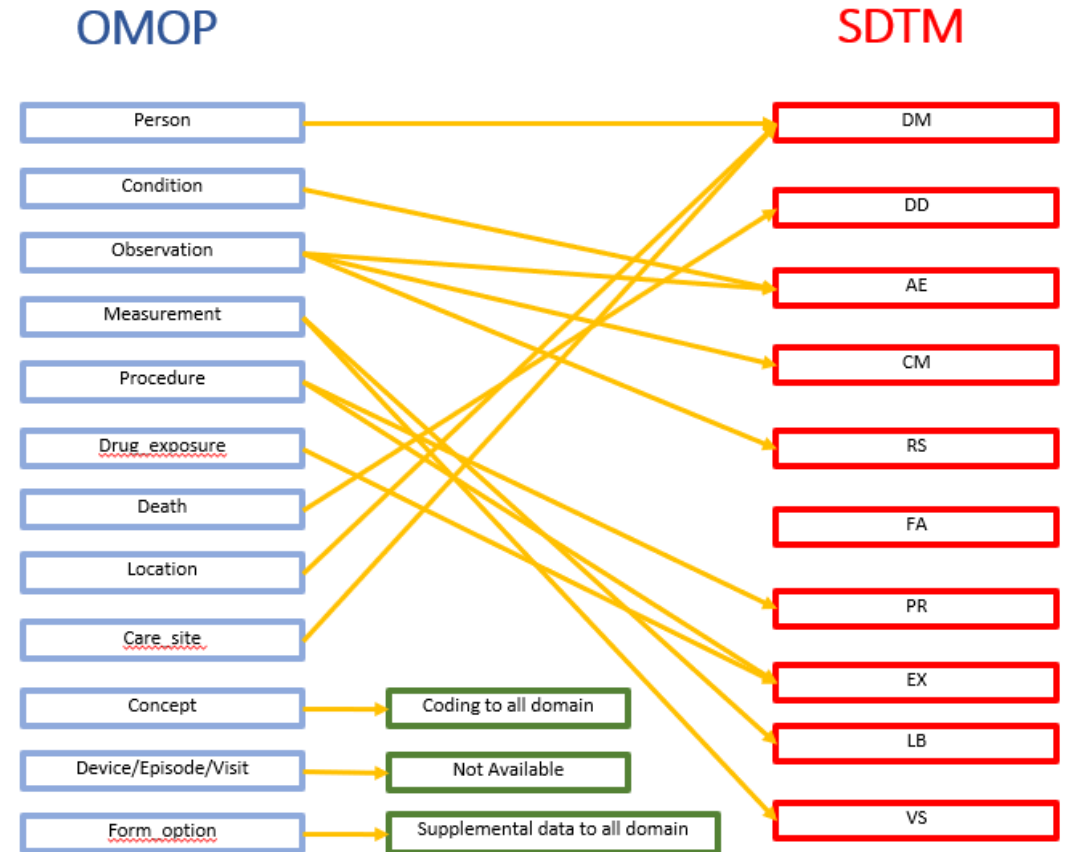
Observation:

observation_id	observation_concept_id	observation_date	observation_datetime	observation_type_concept_id	value_as_number	value_as_string	value_as_concept_id	qualifier_concept_id	unit_concept_id
50604889	20005000007	1/1/2024		32879	1				
24626896	2000200067	1/1/2024		32879			45878245		
14987021	36676867	1/1/2024		32879			739862		
52388325	0	1/1/2024		32879	73				9529

observation_source_value	unit_source_value	person_id	visit_detail_id	visit_occurrence_id	provider_id	form_option_id	patient_event_id	is_date_default	group_index
Therapy and cell infusion(s) Chronological number of cellular therapy treatment for this patient 1		10001				251612	9c67abe3-0997-4e1e-8000-000000000000	FALSE	
Lymphomas > 4 nodal sites involved 1		10001				238555	9de650ad-4209-41a1-b000-000000000000	FALSE	
Planned cellular infusion products Name of product 2		10001				240349	9c67abe3-0997-4e1e-8000-000000000000	FALSE	
Patient status Patient weight at initiation of HCT/CT/IST 73 kg		10001				252136	9de650ad-4209-41a1-b000-000000000000	FALSE	

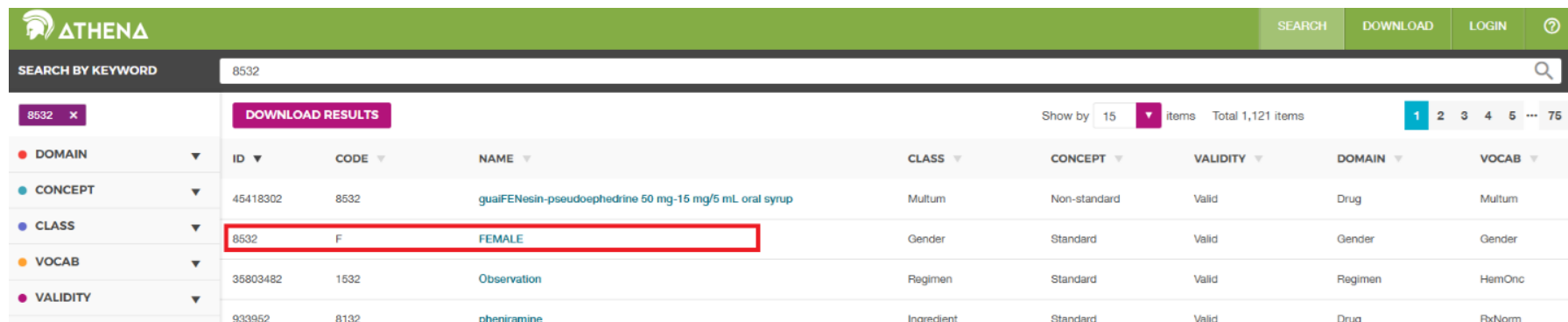
How to Convert or Remap OMOP to SDTM

- Person/Death → DM, DS in SDTM
- Observation/Condition → event-based domains such as AE, CM or efficacy (RS)
- Measurement → BDS domains such as LB, VS
- Other(e.g. site, visit/episode, exposure, and supplemental data) → DM/SV/EX/SUPPxx



Some Concerning - Coding

- Data can be provided as standardized code
- Value to be decoded by linking concept_id with Concept table or looking up at OHDSI website ([Athena](#)) such as:



The screenshot shows the Athena OHDSI website interface. The search bar contains the keyword '8532'. The results are displayed in a table with columns: ID, CODE, NAME, CLASS, CONCEPT, VALIDITY, DOMAIN, and VOCAB. The row for 'FEMALE' (concept ID 8532) is highlighted with a red box.

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
45418302	8532	guaifenesin-pseudoephedrine 50 mg-15 mg/5 mL oral syrup	Multum	Non-standard	Valid	Drug	Multum
8532	F	FEMALE	Gender	Standard	Valid	Gender	Gender
35803482	1532	Observation	Regimen	Standard	Valid	Regimen	HamOnc
933952	8132	pheniramine	Ingredient	Standard	Valid	Drug	RxNorm

- OMOP standard decodes vs CDISC standard terms.
- AE or CM verbatim terms may need to be re-coded for clinical analysis purpose.

Some Concerning – Traceability and Data Masking

Traceability:

- for RWD, CRF NOT available
- In our case, vendor-provided record ID (e.g. Observation_ID) → SDTM source ID (e.g. AE.AESPID)
- Difficult to maintain valid traceability to original source, and potential issue to data manage

Data Masking:

- Many RWD vendors have data privacy requirement based on HIPPA law
- Person information (e.g. subject ID, visit date) can be confidential and not share-able
- Data masking strategy required
- Industry-wide guideline needed

Take-away

- Some advantages about using OMOP for RWD collection:
 1. In vertical structure which is close to clinical data standard (e.g. CDISC)
 2. Uses basic categorization which facilitates further converting to SDTM model
 3. Flexible data structure to allow un-standardized data to be collected
 4. Relatively easy for hospital or other investigator to quickly fit health care data into one common data
- In conclusion, OMOP can be a good option for pharmaceutical industry to fill gap between RWD and clinical analysis

Reference

- (1) Abolafia, J, Ferko, S, & Holt, I. (2024). *“Future Clinical Data Submission Standards: CDISC, FHIR, OMOP, or Hybrid Model”*. Paper presented at the PHUSE US Connect 2024.
- (2) Shiralkar, P & Rutkowski, M. (2024). *“Usage of OMOP-CDM along with CDISC in Clinical Trials: Challenges and Opportunities”*. Paper presented at the PHUSE US Connect 2024.
- (3) Shanbhogue, A.Y., (2022). *“RWD (OMOP) to SDTM (CDISC): A primer for your ETL journey”*. Paper presented at the PharmaSUG Annual Conference 2022.
- (4) CDISC Guidance (2023) *“Considerations and Examples of Using SDTM for Observational Studies”*.
- (5) FDA Guidance (2021) *“Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry”* .
- (6) OHDSI <https://www.ohdsi.org/data-standardization/> (accessed 12/01/2024).

Acknowledgement

Great thanks to Joy Wang and Samar Noor

Thank you