Mapping of complex constructs in OMOP CDM

PRESENTER: **Alexander Davydov**

BACKGROUND

There is a growing need in the OHDSI community for the conversion of new types of data, different from the typical electronic health record systems or administrative claims data. This type of data is typically organized as entity-attribute-value (EAV) records, where the entity is either a question or a variable, the attribute is the link, and the value or answer is the value (table 1).

The ETL to convert them to OMOP CDM records is more complicated and requires specific solutions. In EAV, records often come in the form of variable/value or question/answer pairs. In the OMOP CDM, such data can be handled in the MEASUREMENT and OBSERVATION tables. The variable/question becomes the main concept, and the value/answer may be a value concept or not a concept at all.

Currently, there exist several approaches to relate different records in OMOP CDM: (i) through the SOURCE_TO_CONCEPT_MAP table; (ii) through the CONCEPT table in combination with the CONCEPT_RELATIONSHIP table and dedicated relationship IDs ("Maps to", "Maps to value", "Maps to unit", etc.). Both of these methods have certain limitations (table 2) that become relevant while converting EAV-modeled data.

METHODS AND RESULTS

In order to support and harmonize such data conversions, model changes in the structural organization of vocabulary mapping are required. It is crucial that mapping incorporates multiple source concepts, multiple target concepts of different domains, and data types (numeric, date, string). Here, we present two new solutions in this respect.

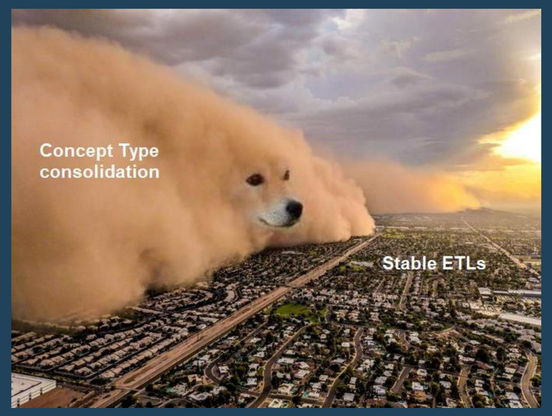
WIDE MAPPING table

The format is different in the way that mappings to multiple target entities will be performed by adding all the respective fields needed rather than creating the multiple rows and leveraging through the relationship_id/source_vocabulary_id (table 3).

Despite the promises, there are still some disadvantages and open questions:

- The table has a machine-readable format. Any attempts to add the descriptions will result in an excessive number of fields placed in one row so that users may have difficulties while building or looking up mappings.
- Source_concept_id key of the event table may be replaced by a foreign key back to the row in the wide mapping table, but then the table would become a reference for the source data. Currently, only the concept table plays this role.
- It's not clear how to represent text strings as a part of the source data.
- It forces users to create custom 2bil+ concepts out of the source data.
- Usually, the units of measure are separately coded in an additional field. The addition of the source_unit field to the wide mapping table gets us to a combinatorial explosion in most of the real-world data sources, even though it might be useful for controlled vocabularies and clean sources.
- The concept of the wide mapping table is to provide ETL with machine-readable instructions on how and where to extract the numeric value from. Additionally, it helps to differentiate the cases when there is no need to extract them (NULL numeric field). The ETL logic around it may be even more complex than the one that is currently used.

Remember Type Concept consolidation in 2020?



This time we really need **your** input.

Alexander Davydov, MD
Christian Reich, MD, PhD

ODYSSEUS DATA SERVICES

OHDSI

NUMERICAL RELATIONSHIP groups

Is an alternative solution that could be introduced into the concept_relationship table that indicate which attributes (value, unit, status, operator, modifier) belong to which "Maps to" relationship. Using the groups, the ETL scripts will automatically sort out one-to-many mappings and respective relationships into the target event records. In order to support the mappings of pre-coordinated pairs, the synthetic merged entity-value pairs still should be created. The mapping process to a field other than event_concept_id, will be leveraged using the new relationship_ids ("Maps to status", "Maps to operator", "Maps to modifier"). The most valuable benefit of this approach is that the mappings are still organized in many rows rather than in a single row and may be easily looked up by users. Also ETL logic and OMOP structure will not be affected that much. However, there are some points to be addressed still: mapping of (and to) ranges, numerics, strings and dates. The representation of these mappings probably still requires introduction of additional fields.

Type	Variable / Question	Value / Answer
Lab tests with the qualitative result	SARS-CoV-2 (COVID-19) IgA+IgM [Presence] in Serum or Plasma by Immunoassay	Equivocal / Negative / Positive
Historic facts	Family history of clinical finding	Myocardial infarction
Cancer stages and assessment measures	FIGO Stage (2018 FIGO Cancer Report)	I: Tumor confined to ovaries or fallopian tube(s)
	Circumferential Resection Margin (CRM)	100 mm or greater
Survey instruments created for specific projects (UK Biobank, All Of US PPI)	Has a doctor told you that you have any of the following problems with your eyes?	Macular degeneration
	How often did you use cannabis?	1-5 times per week
Surveys by itself (PhenX, PROMIS)	Because of your problem, do you feel frustrated	No / Sometimes / Yes
	Smoking helps me concentrate	Not at all / Somewhat / Very much

Table 1. Examples of EAV-structured data

Use case	Example	Issue
One-to-many "splitting" mappings through multiple relationships	"Maps to" and "Maps to value" pairs: "History of" + value of "COVID-19 vaccine" together with "SARS-COV2 PCR test" + value of "POS"	It is ambiguous which "Maps to" belongs to which "Maps to value", and the standard ETL process will inflate the records
Many-to-one "merging" or "pre-coordination" mapping	HHV-6B seropositivity for Human Herpesvirus-6: False	Only a single code can be an input for a map. As a result, the ETL needs to apply a workaround and first merge the entity/value codes to map them to the target concept
	EuroQol five dimension three level self-care score: 3 (I am unable to wash or dress myself)	
Separate mapping of entities and values, which is possible only if the values are entity agnostic	Generic "Yes", "No" answers to questions; drugs, conditions and other self-sufficient concepts	Now this is managed by splitting the source codes into separate synthetic source vocabularies
Mapping to numeric content or numeric with a unit	CS Tumor Size of 32 mm	Currently, ETL needs to extract the numeric values and units from the text
Mapping of a range	Blood alcohol level of 100-119 mg/100 ml	Ranges are currently not supported
Mapping to a string	White sliced bread eaten	Currently, ETL needs to extract the values from the text
Mapping to a date	Birthdate of a relative: "1988-Sep-17"	Currently, ETL needs to extract the dates from the source

Table 2. Limitations of existing system

Table 3. The structure of the wide mapping table and mapping examples

Source				Target								
Source Concept	**Question/Variable**	**Answer/Value**	**Range**	**Standard Concept**	**Numeric**	**Operator**	**Error**	**Unit Concept**	**Value Concept**	**String**	**Condition Status Concept**	**Visit Concept**
Ambulatory procedures - lithotripsy				Lithotripsy								Ambulatory Surgical Center
	CS Tumor Size	\d	001 - 988 millimeters (mm) |(Code exact size in mm)	Estimated Tumor Size	\d			millimeter				
Documentation of patients with primary headache diagnosis and imaging other than ct or mri obtained				Headache							Primary diagnosis	
				Imaging								
Evidence of alcohol involvement determined by blood alcohol level of 100-119 mg/100 ml				Ethanol [Mass/volume] in Blood	110		10	milligram per deciliter				
Home visit, phototherapy services (e.g., bili-lite), including equipment rental, nursing services, blood draw, supplies, and other services, per diem				Home visit, phototherapy services (e.g., bili-lite), including equipment rental, nursing services, blood draw, supplies, and other services, per diem								Home Visit
	Wears glasses or contact lenses	Yes		Abnormal vision								
				Uses visual aid								
	Age started wearing glasses or contact lenses	**\d (e.g. 15)**		History of event **longer than 10 years ago**					Uses visual aid			
	Type of sliced bread eaten	white		Food eaten						"white sliced bread"		