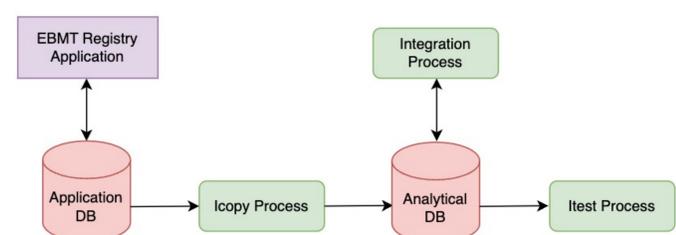


# Integrating a templated configurator into the ETL process

## Automated OMOP-CDM pipeline for the new EBMT Registry

**Background:** The configurator is a web application solution that templates the process of converting the EBMT Registry's complex data-entry forms base into OMOP-CDM. This solution allows for new configurations to be easily entered into the application through the user interface and applied to the Extract-Transform-Load (ETL).

The user enters the new configuration through the EBMT Registry's application user interface. The new configuration is saved to the application database. The nightly pipeline copies the relevant patient data and the configurations to the analytical database (icopy process), and uses the configurations during the ETL to map the patient data to OMOP-CDM format (integration process), and stored on the analytical database. The process is then tested for quality assurance (itest process).



## Methods

- 1 Configuration data for each field that is recorded is stored in the Field table (1a). The Data Service Option table (1b) stores configuration data for fields with pre-fixed options. Patient data is stored in the Field Response table (1c).

Field			
id	integrationConfig	eventTypeId	label
field1	{'references': {'omop_start_date': field1}, 'directs': {'concept_id': 4266367}}	event_type1	Date of diagnosis
field2	{'references': {'omop_start_date': field1}, 'data_sources': {'concept_id': data_service1}}	event_type1	Type of flu
field3	{'references': {'omop_start_date': field1, 'value_as_number': field3}, 'directs': {'concept_id': 4302666, 'unit_source_value': 'celsius'}}	event_type1	Body temperature

1a

Data Service Option			
id	dataServiceId	value	dataConfig
option1	data_service1	1	{'standard_concept_id': 1111}
option2	data_service1	2	{'standard_concept_id': 2222}

1b

Field Response				
id	patientId	patientEventId	fieldId	value
response1	patient1	patient_event1	field1	2008-08-08
response2	patient1	patient_event1	field2	1
response3	patient1	patient_event1	field3	38.7

1c

- 2 The ETL process first builds an intermediary STEM table with all the necessary data for each field. Using the configurations in the Field table, the process captures the date of each data point, even when it is not included in the Field Response.

- 3 The ETL process then joins the STEM and Concept tables on concept\_ids to retrieve the domain and populates the associated clinical table.

Measurement					
measurement_concept_id	measurement_date	value_as_number	unit_source_value	person_id	patient_event_id
4302666	2008-08-08	38.7	celsius	person1	patient_event1

Condition Occurrence			
condition_concept_id	condition_start_date	person_id	patient_event_id
4266367	2008-08-08	person1	patient_event1
1111	2008-08-08	person1	patient_event1

**Conclusion:** Templating mapping configurations via user interface keeps the OMOP-CDM at the forefront when expanding data collection and allows for seamless expansion of the application database without needing to update the code, pipeline, or ETL process. As a result, this pipeline can be maintained by any non-technical team member that is familiar with OMOP-CDM.



Shirah Cashriel<sup>1</sup>, Nir Assaraf<sup>1</sup>, Ignacio Garcia<sup>2</sup>, Fernando Cid<sup>2</sup>, Freija Descamps<sup>1</sup>

<sup>1</sup>edenceHealth  
<sup>2</sup>EBMT Registry

