



# **APAC Scientific Forum: Integration Local vocabulary into OMOP CDM**

Yiju Park, Seng Chan You

July 3, 2025

---



# Publication Progress

- Preprint published on medRxiv
- Currently under review at HIR (Healthcare Informatics Research)

The screenshot shows the medRxiv preprint page for the article "Integrating Local Vocabulary into OMOP CDM: A Step-by-Step Tutorial". The page includes the medRxiv logo, navigation links (HOME | SUBMIT | FAQ | BLOG | ALERTS / RSS | RESOURCES | ABOUT), a search bar, and a "Follow this preprint" button. The article title is "Integrating Local Vocabulary into OMOP CDM: A Step-by-Step Tutorial", posted on May 08, 2025. The authors listed are Yiju Park, Jinwoo Yoon, Aleh Zhuk, Anna Ostroplets, and Seng Chan You. The DOI is https://doi.org/10.1101/2025.05.07.25327200. A warning message states: "This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice." The page offers options to download the PDF, print/save options, author declarations, supplementary material, and data/code. It also features a "Subject Area" section for "Health Informatics" and a "Reviews and Context" sidebar with options for comment, TRIP Peer Reviews, community reviews, automated services, blogs/media, and author videos.



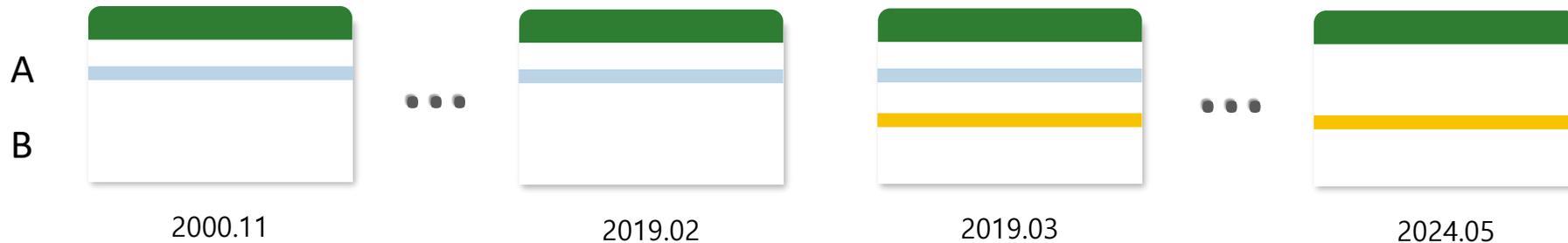
# Topics

- Introduce the **Korean EDI** Vocabulary
  - **Process** for Data Integration in ATHENA
    - Create a Semi-automated process : **SYNC package**
    - **Mapping to Standard Concepts**
    - **Community Contribution Process**
-



# Introduction of EDI

- **EDI (Electronic Data Interchange) is a widespread code system for the reimbursement or claim data in Korea**
- EDI is developed and maintained by **HIRA** (Health Insurance Review & Assessment Service)
- EDI concepts are divided into **drugs, devices, and medical services**
- EDI is **updated every month** and is a **longitudinal data**



Example of EDI data from HIRA (each line represents one EDI code)



# Introduction of EDI

- Despite widespread adoption in Korean EHR systems, **limitations** still persist.
  - Validity dates are not recorded in the official monthly announcement.
  - There are expired or replaced EDI code and outdated EDI can be assigned to new concepts.
  - EDI has duplicated identifiers due to the lack of a unified encoding system across domains.

## OUR GOAL WAS ...

To enhance EDI vocabulary for a controlled and standardized vocabulary system

- For this purpose, we incorporated the EDI into OMOP vocabulary **using a semi-automated process.**



# SYNC package

- SYNC package is a **semi-automated process** we made.
- This package makes EDI vocabulary as a Source Concepts using OMOP structure.
- There are **four main steps** to incorporate EDI into OHDSI standardized vocabulary.

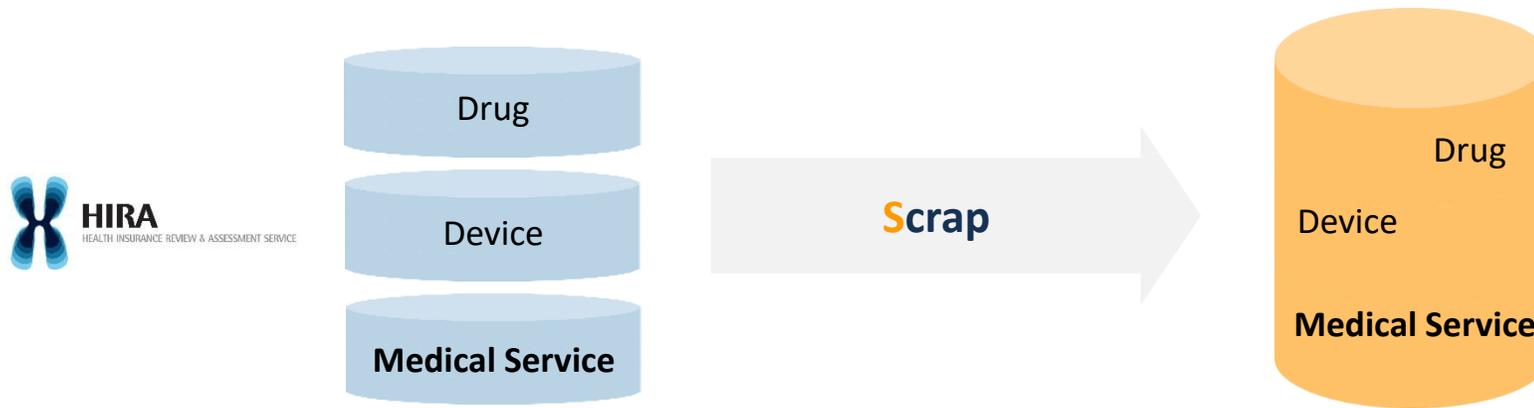
- 1 Scrap: Crawl EDI data from HIRA website**
- 2 Yield: Categorize codes into domains (Device, Drug, Procedure, Measurement)**
- 3 Nest: Built a vertical hierarchy between EDI concepts**
- 4 Convert: Added an English definition for each EDI concept using Google Translation**



# SYNC package

## 1 Scrap: Crawl EDI data from HIRA website

- Through web crawling, all necessary EDI data from the HIRA website is downloaded
- There are three types of data provided by HIRA: Device, Drug, Medical Service
- We have collected all EDI data from November 2000 to May 2024





# SYNC package

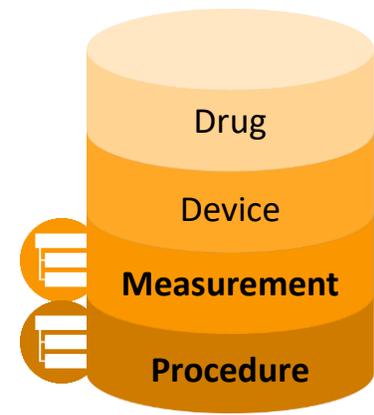
## 2 Yield: Categorize codes into domains (Device, Drug, Procedure, Measurement)

- As mentioned, EDI concepts are divided into **Device**, **Drug** and **Medical Service**.
- The scope of **Medical Service** is **too broad** for the OHDSI standardized vocabularies.
- So, we subclassified Medical Services into **Procedure** and **Measurement** to match the OHDSI domains.





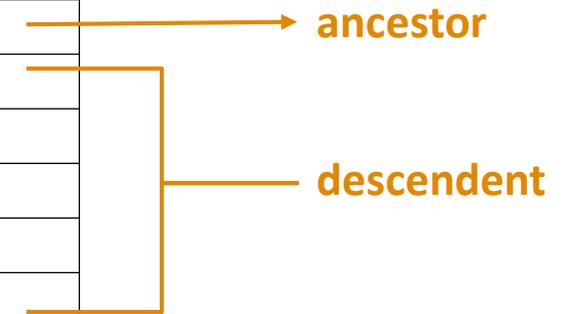
# SYNC package



## 3 Nest: Built a vertical hierarchy between EDI concepts

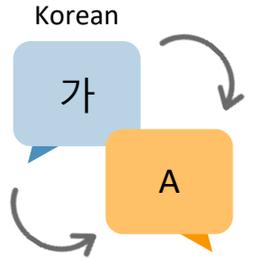
- We built a formal vertical hierarchy for EDI concepts as ICD-9 or ICD-10 code system.
- The **first five digits** of the EDI code in the medical service domain (procedure, measurement) **represent the ancestor terms** for longer descendent EDI code.

Concept code	Concept name	Ancestor concept code
HC281	Whole Body Scan	
HC281006	Whole Body Scan, Nuclear Medicine and other physician reading	HC281
HC281300	Whole Body Scan, Under 8 years old	HC281
HC281306	Whole Body Scan, Under 8 years old, read by nuclear medicine physician	HC281
HC281600	Whole Body Scan, Under 72 months	HC281
HC281606	Whole Body Scan, Under 72 months, Nuclear Medicine physician reading	HC281





# SYNC package



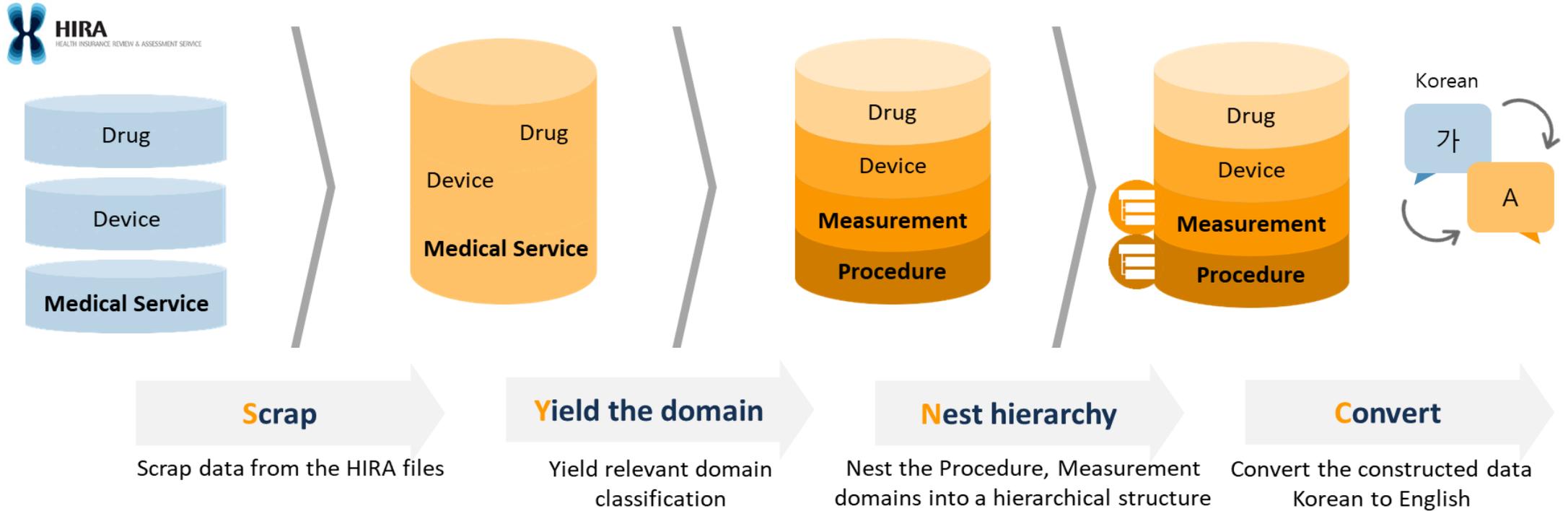
## 4 Convert: Added an English definition for each EDI concept using Google Translation

- We have to add an English definition for each EDI term.
- We utilized **Google Cloud Translation API** for the initial translation.
- Inaccurately translated words underwent **review by experts** and were retranslated.

Korean definition	Using Google Translation API	Using Google Translation API With glossary
M핵산증폭-정량그룹1_B형 감염바이러스 [중합효소연쇄반응교잡반응법]	Nucleic acid amplification-quantitative group 1 hepatitis B virus [polymerase chain reaction hybridization method]	Nucleic acid amplification-quantitative group 1_HBV [PCR-Hybridization]
단기사용담관용튜브·카테터	Short-term use bile duct tube and catheter	Cahteter, bile duct short-term use



# SYNC package





# SYNC package

concept	concept_name	concept_synonym	domain_id	vocab	concept_class_id	valid_start	valid_end	inv	ances	mate	comp	dosag	dosag	sanju	name
O2055	Thrombectomy(Artery),Neck	혈전제거술(동맥-경부)	Procedure	EDI	Proc Hierarchy	2008-01-01	2099-12-31								
O2055010	Thrombectomy(Artery),Neck,Nighttime	혈전제거술(동맥-경부),야간	Procedure	EDI	Procedure	2008-01-01	2099-12-31		O2055					야간	
O2055020	Thrombectomy(Artery),Neck,emergency	혈전제거술(동맥-경부),응급	Procedure	EDI	Procedure	2016-01-01	2023-12-31	D	O2055					응급	
D3030	1.5-Anhydro-D-Glucitol	1.5-Anhydro-D-Glucitol[화학빈	Measurement	EDI	Proc Hierarchy	2018-01-01	2099-12-31								
D3030001	1.5-Anhydro-D-Glucitol,Diagnostic and laboratory te	1.5-Anhydro-D-Glucitol[화학빈	Measurement	EDI	Proc Hierarchy	2018-01-01	2099-12-31		D3030					진단검사	질가산(4%)
D3030002	1.5-Anhydro-D-Glucitol,Diagnostic and laboratory te	1.5-Anhydro-D-Glucitol[화학빈	Measurement	EDI	Proc Hierarchy	2018-01-01	2099-12-31		D3030					진단검사	질가산(3%)
646902600	acetaminophen(encapsulated) 0.65g	타이레놀8시간이알서방정(C	Drug	EDI	Drug Product	2022-07-01	2022-12-31	U	101430ATR	(주)한국1		정			
672300240	acetaminophen(encapsulated) 0.65g	타이레놀8시간이알서방정(C	Drug	EDI	Drug Product	2023-01-01	2099-12-31		101430ATR	한국존슨1		정			
G2201003	COREVALVE EVOLUT SYSTEM		Device	EDI	Device	2015-06-01	2099-12-31			PORCINE	메드트로	전규격	1KIT		
G2201002	EDWARDS SAPIEN 3 AND SAPIEN 3 ULTRA TRANSCA		Device	EDI	Device	2015-06-01	2099-12-31			BOVINE	에드워즈	전규격	1KIT		

- After completing all processes in the SYNC package, we generate a merged vocabulary tables as shown.
- These tables contain comprehensive information(material, company names, dosage) about each EDI code.



# Mapping to Standard Concepts

- EDI concepts were mapped to standard concepts following the OHDSI-Korea community guidelines
- Every concepts should be **standard** based on the ATHENA

ohdsi-korea / OmopVocabularyKorea

<> Code Issues 3 Pull requests Actions Projects Wiki Security Insights

OmopVocabularyKorea Public Watch

master 1 Branch Tags

Go to file Add file Code

chandryou Merge pull request #15 from hhipch/master f611f09 · last year 39 Commits

- device\_guideline total procedure 3 years ago
- guidelines Add files via upload 5 years ago
- mapping table add UTF-8 encoded files last year
- measurement\_guideline Adding source code concept id 4 years ago
- procedure\_guideline Publishing procedure mapping guidelines(pdf) 4 years ago

## < Priority of Standard Concepts in Mapping >

	Non-standard Concepts	Standard Concepts
Procedure	EDI	SNOMED
Measurement	EDI	1) LOINC 2) SNOMED
Drug	EDI	RxNorm RxNorm Extension
Device	EDI	SNOMED



# 1. Procedure Mapping

1. <http://athena.ohdsi.org/>

2. Filter

- Domain: **Procedure**
- Standard Concept: **Standard**
- Vocabulary: **SNOMED**

The screenshot shows the Athena OHDSI search interface. The search bar contains the keyword 'aspirin'. The search results are displayed in a table with columns: ID, CODE, NAME, CLASS, CONCEPT, VALIDITY, DOMAIN, and VOCAB. The results are filtered by SNOMED, Procedure, and Standard. The table shows 15 items per page, with a total of 48,211 items. The first five items are listed below:

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
4017324	171425002	"Section" examination - approved doctor	Procedure	Standard	Valid	Procedure	SNOMED
4017323	171424003	"Section" examination - patient's GP	Procedure	Standard	Valid	Procedure	SNOMED
4017322	171426001	"Section" examination - social worker	Procedure	Standard	Valid	Procedure	SNOMED
4098808	252865007	+3 diopter spheres test	Procedure	Standard	Valid	Procedure	SNOMED
4145308	268400002	12 lead ECG	Procedure	Standard	Valid	Procedure	SNOMED
40491312	447113005	12 lead electrocardiogram at rest	Procedure	Standard	Valid	Procedure	SNOMED



# 1. Procedure Mapping

If you want to map 'Hemiarthroplasty of hip'

1. Search your procedure name in ATHENA

**ATHENA** SEARCH DOWNLOAD LOGIN ?

SEARCH BY KEYWORD  ?

Hemiarthroplasty o... x

Procedure x

Standard x

SNOMED x

DOMAIN ▼

CONCEPT ▼

CLASS ▼

VOCAB ▲

**DOWNLOAD RESULTS** Show by 15 items Total 31,692 items 1 2 3 4 5 ... 2113 >

ID ▼	CODE ▼	NAME ▼	CLASS ▼	CONCEPT ▼	VALIDITY ▼	DOMAIN ▼	VOCAB ▼
4079261	179321008	Revision cemented hemiarthroplasty of hip	Procedure	Standard	Valid	Procedure	SNOMED
4079263	179328002	Revision uncemented hemiarthroplasty of hip	Procedure	Standard	Valid	Procedure	SNOMED
4205235	309458002	Cemented Thompson hemiarthroplasty of hip joint	Procedure	Standard	Valid	Procedure	SNOMED
44807070	794781000000101	Attention to prosthetic hemiarthroplasty of hip	Procedure	Standard			

No appropriate concept name



# 1. Procedure Mapping

If you want to map 'Hemiarthroplasty of hip'

2. Modify Search Terms and Re-search

e.g., Hemiarthroplasty → partial joint replacement

The screenshot shows the Athena search interface. The search bar contains the text "partial joint replacement of hip". The results are displayed in a table with columns: ID, CODE, NAME, CLASS, CONCEPT, VALIDITY, DOMAIN, and VOCAB. The table shows four results, with the last one highlighted by a red box.

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
37153323	1217139005	Partial replacement of joint of left hip with bipolar prosthesis	Procedure	Standard	Valid	Procedure	SNOMED
37153324	1217140007	Partial replacement of joint of right hip with bipolar prosthesis	Procedure	Standard	Valid	Procedure	SNOMED
4225800	340922009	Partial hip replacement by cup	Procedure	Standard	Valid	Procedure	SNOMED
4297365	386649003	Partial hip replacement by prosthesis	Procedure	Standard	Valid	Procedure	SNOMED



# 1. Procedure Mapping

If you want to map 'Hemiarthroplasty of hip'

3. Move up to broader concepts, and check subsumes for better matches

SEARCH DOWNLOAD LOGIN ?

← Prosthetic hemiarthroplasty of articulation of bone

DETAILS	
Domain ID	Procedure
Concept Class ID	Procedure
Vocabulary ID	SNOMED <span data-bbox="1039 896 1065 916">?</span>
Concept ID	40490972
Concept code	448649009
Validity	Valid
Concept	Standard
LANGUAGE	SYNONYM CONCEPT
English	Prosthetic hemiarthroplasty of joint

TERM CONNECTIONS (95) ☐ HIERARCHY RELATED CONCEPTS

RELATIONSHIP	RELATES TO	CONCEPT ID	VOCABULARY
Subsumes	Bilateral reverse prosthetic total arthroplasty of shoulders	619172	SNOMED
	Cemented resurfacing hemiarthroplasty of head of humerus	4330789	SNOMED
	Conversion to cemented hemiarthroplasty of hip	4079521	SNOMED
	Conversion to prosthetic replacement of head of radius	44783175	SNOMED
	Hemiarthroplasty of head of right humerus	764226	SNOMED
	<b>Partial hip replacement by prosthesis</b>	4297365	SNOMED
	Partial left hip replacement by prosthesis	37155390	SNOMED
	Partial right hip replacement by prosthesis	37155391	SNOMED



## 2. Measurement Mapping

1. <http://athena.ohdsi.org/>

2. Filter

- Domain: **Measurement**
- Standard Concept: **Standard**
- Vocabulary: **LOINC(mostly)**

The screenshot shows the Athena OHDSI search interface. The search bar contains the keyword 'aspirin'. The search results are displayed in a table with columns: ID, CODE, NAME, CLASS, CONCEPT, VALIDITY, DOMAIN, and VOCAB. The results are filtered to show only 'Procedure' concepts from the 'LOINC' vocabulary. The table shows 6,671 total items, with 15 items displayed per page. The first few rows of results are:

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
1761911	100225-2	Bone density quantitative ultrasound study	Clinical Observation	Standard	Valid	Procedure	LOINC
1989694	99633-0	Cone beam CT Teeth	Clinical Observation	Standard	Valid	Procedure	LOINC
1989690	99631-4	Cone beam CT Temporomandibular joint - bilateral	Clinical Observation	Standard	Valid	Procedure	LOINC
1989681	99632-2	Cone beam CT Temporomandibular joint WO and W contrast IV	Clinical Observation	Standard	Valid	Procedure	LOINC
3032013	41806-1	CT Abdomen	Clinical Observation	Standard	Valid	Procedure	LOINC
21492095	79066-7	CT Abdomen 3D post processing WO contrast	Clinical Observation	Standard	Valid	Procedure	LOINC



# Measurement Mapping

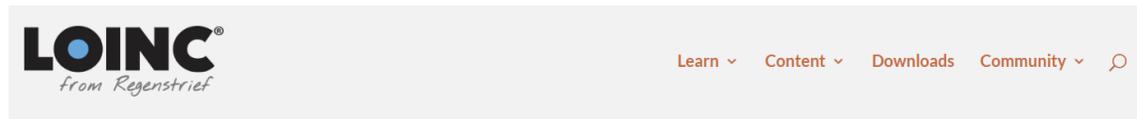
- Search with **the unit** and **the specimen type**
  1. Unit: mg/dL, Specimen: serum
    - mg/dL → [mass/volume]
    - Search for '[mass/volume] in serum' in ATHENA
  2. Unit: g/dL, Specimen: whole blood
    - g/dL → [mass/volume]
    - Search for '[mass/volume] in whole blood' in ATHENA

dL	1/10 L	Volume
mL	1/1,000 L	Volume
mg	1/1,000 g	Mass
µg	1/1,000,000 g	Mass
Ng	1/1,000,000,000 g	Mass
Pg	1/1,000,000,000,000 g	Mass
U	Standardized Units for Each Test	Units Enzymatic activity
mIU	1/1,000 U	Units
µU	1/1,000,000 U	Units
mol	Quantity of Atoms and Molecules	Moles
mmol	1/1,000 mol	Moles
µmol	1/1,000,000 mol	Moles
nmol	1/1,000,000,000 mol	Moles



# 2. Measurement Mapping

- When it's difficult to choose between similar concepts, Go to <https://loinc.org/>



HOME > CONTENT

## Download LOINC

### Other ways to access LOINC content

A LOINC login is required to access each of these resources. You can [create one for free](#).

#### SearchLOINC

This web-based tool allows you to find and export LOINC codes using an intuitive interface.

#### LOINC API using HL7® FHIR®

Access LOINC content programmatically using the HL7's popular FHIR standard.

#### RELMA®

Download and install this program for Windows to search for LOINC terms and map them to your local codes.

#### Archive of past LOINC releases

A complete archive of previous releases is available for your review.

HOME > CONTENT

## LOINC Mapping Guides

### LOINC Mapping Guide - Allergy

Published September 2022.

DOWNLOAD Version 1.0

### LOINC Mapping Guide - Cell Markers

Published September 2022.

DOWNLOAD Version 1.0

### LOINC Mapping Guide - Chemistry

Published September 2022.

DOWNLOAD Version 1.0

### LOINC Mapping Guide - Drug and Toxicology

Published September 2022.

DOWNLOAD Version 1.0

### LOINC Mapping Guide - Hematology and Serology

Published September 2022.

DOWNLOAD Version 1.0

### LOINC Mapping Guide - Molecular Pathology

Published September 2022.

DOWNLOAD Version 1.0

### Guide for Using LOINC Microbiology Terms

Originally published August 2018. Version 1.1 released in August 2019.

DOWNLOAD Version 1.1



#### SearchLOINC

Find LOINC concepts in your browser with this powerful web app. A free LOINC login is required.

#### Hierarchy Browser

Instead of searching, peruse LOINC codes via numerous hierarchies. A free LOINC login is required.

#### Community Forum

Ask questions and chime in on conversations with your fellow LOINCers.

#### Request a LOINC

Cannot find an appropriate LOINC concept? Follow instructions for requesting a new code.

#### Submission Queue

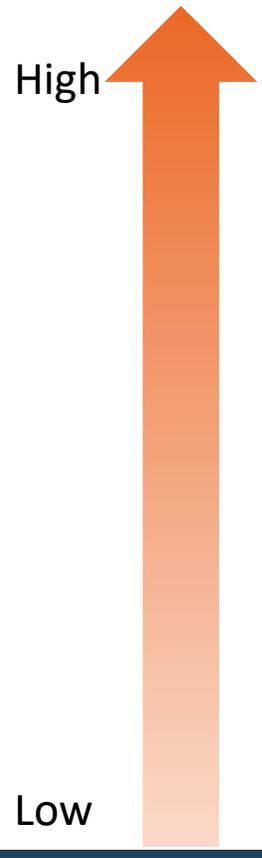
See all the new concepts we are working on now, including any submissions you may have made.





## 3. Drug Mapping

- Mapping to **RxNorm** or **RxNorm Extension**
- **Mapping Priority**



1. Ingredient + Dosage + Dosage Form + [Drug Name] + by Company Name

Acetaminophen + 160 mg + chewable tablet + [TAIFEN] + by Youngpoong

Concept id: **2071554**

2. Ingredient + Dosage + Dosage Form + [Drug Name]

Acetaminophen + 160 mg + chewable tablet + [TAIFEN]

Concept id: **2071553**

3. Ingredient + Dosage + Dosage Form

Acetaminophen + 160 mg + chewable tablet

Concept id: **19079924**

4. Ingredient + Dosage Form

Acetaminophen + chewable tablet

Concept id: **40005631**

5. Ingredient only

Acetaminophen

Concept id: **1125315**



## 3. Drug Mapping

- We need to distinguish between **Injection** and **Injectable solution**
  - MG → Injection
  - MG/ML → Injectable solution

- **Injection**

- A pharmaceutical preparation that is administered directly into the body through the skin or mucous membranes, available as solutions, suspensions, emulsions, or as powder forms that need to be dissolved or suspended before use
- **E.g., Powder injections in MG, G forms**

- **Injectable Solution**

- A ready-to-use injection containing one or more drugs in a suitable solvent or mixture for injection
- **E.g., Injections in MG/ML forms**



## 4. Device Mapping

- Mapping to **SNOMED**
- Utilize company name or materials
- (Only for EDI) Recommended to refer to the 'Medical Device Insurance Guide Map'  
published by HIRA



# Community Contribution Process

- For the integration, you need to adhere to the **Community Contribution pipeline**.
- If you are interested in loading the vocabulary, please contact the Vocabulary Team.

The screenshot shows the GitHub Wiki page for OHDSI Vocabulary-v5.0. The page title is "Community contribution" and it was last edited by Anna Ostropolets on Jan 7. The page content includes:

- A heading "Part 1: simple use-cases" with a sub-heading "Part 1 guidance covers:" followed by a numbered list of 6 items: 1. adding new simple non-standard vocabulary with mappings but no other internal relationships; 2. adding new non-standard concept(s) to vocabularies in Athena; 3. adding new mappings for existing concepts (non-standard to standard and standard to standard or deduplication); 4. changing existing concept attributes (such as domain); 5. changing existing mappings; 6. promoting non-standard concept(s) to standard.
- A heading "Part 2: complex use-cases" with a sub-heading "Part 2 covers:" followed by a numbered list of 1 item: 1. Adding new relationships other than 'Maps to' such as hierarchies - *under construction, please come to the Vocabulary WG meeting*.

On the right side of the page, there is a "Quick access" sidebar with a list of links: Home, News, Introduction, Glossary, The Vocabulary Team, Roadmap, Release Notes, Upcoming Changes, Community Contribution, General Structure, Download and Use, Vocabularies, Vocabulary Statistics, Vocabulary Development Process, Quality Assurance and Control, Known Issues in Vocabularies, Articles, and COVID-19 Vocabulary/ETL Instructions.

Contact the Vocabulary Team

or

[Vocabulary WG meeting.](#)



# Community Contribution Process

- **Prerequisites**

- Preparation Schemas : *sources, dev\_xyz, devv5*
- Copies of tables, fully indexed (downloaded from Athena and put into Schema *dev\_xyz, devv5*)

- **Process**

1. Run *load\_stage.sql* in the *dev\_xyz* schema
2. QA/QC part 1
3. Generic update
4. QA/QC part 2 : semi-automatic process



# Community Contribution Process

- **Prerequisites**

- Preparation Schemas : *sources, dev\_xyz, devv5*
- Copies of tables, fully indexed (downloaded from Athena and put into Schema *dev\_xyz, devv5*)

- **Process**

1. Run *load\_stage.sql* in the *dev\_xyz* schema
2. QA/QC part 1
3. Generic update
4. QA/QC part 2 : semi-automatic process



# Community Contribution Process

- **Prerequisites**
  - **Preparation Schemas : *sources, dev\_xyz, devv5***
  - Copies of tables, fully indexed (downloaded from Athena and put into Schema *dev\_xyz, devv5*)

## Schema *sources*

- Put your **source vocabulary**

## Schema *dev\_xyz*

- Set as **Working directory**
- Run *DevV5\_DDL.sql* to create empty tables
- Put copy of vocabularies downloaded from Athena

## Schema *devv5*

- **Reference Schema** of vocabularies from Athena
- Run *DevV5\_DDL.sql* to create empty tables
- Put copy of vocabularies downloaded from Athena



# Community Contribution Process

- Prerequisites

- Preparation Schemas : *sources, dev\_xyz, devv5*
- Copies of tables, fully indexed (downloaded from Athena and put into Schema *dev\_xyz, devv5*)

ATHENA SEARCH DOWNLOAD LOGIN ?

Search

aspirin Search

1. Usage of quotation marks forces an exact-match search  
2. In case of a typo, or if there is a similar spelling of the word, the most similar result will be presented

Explore domains

- Drugs 5,613,135
- Conditions 675,961
- Procedures 738,383
- Devices 518,229
- Observations 973,354
- Measurements 561,032

① Go to download

② Check the corresponding vocabulary

ATHENA SEARCH DOWNLOAD Park yijoo ?

Show all SHOW HISTORY DOWNLOAD VOCABULARIES

<input type="checkbox"/>	ID (CDM V4.5)	CODE (CDM V4.5)	NAME	REQUIRED	LATEST UPDA
<input checked="" type="checkbox"/>	1	SNOMED	Systematic Nomenclature of Medicine - Clinical Terms (IHTSDO)		27-Sep-2023
<input type="checkbox"/>	2	ICD9CM	International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 1 and 2 (NCHS)		01-Oct-2014
<input type="checkbox"/>	3	ICD9Proc	International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 3 (NCHS)		01-Oct-2014
<input type="checkbox"/>	4	CPT4	Current Procedural Terminology version 4 (AMA)	EULA required	01-May-2023
<input type="checkbox"/>	5	HCPCS	Healthcare Common Procedure Coding System (CMS)		01-Jan-2024

③



# Community Contribution Process

- **Prerequisites**

- Preparation Schemas : *sources, dev\_xyz, devv5*
- **Copies of tables, fully indexed (downloaded from Athena and put into Schema *dev\_xyz, devv5*)**

```
Vocabulary-v5.0 / working / DevV5_DDL.sql
Code Blame 314 lines (288 loc) · 14.1 KB
19
20 --Main DDL
21
22 DROP TABLE IF EXISTS concept CASCADE;
23 CREATE TABLE concept (
24     concept_id int4 NOT NULL,
25     concept_name VARCHAR (255) NOT NULL,
26     domain_id VARCHAR (20) NOT NULL,
27     vocabulary_id VARCHAR (20) NOT NULL,
28     concept_class_id VARCHAR (20) NOT NULL,
29     standard_concept VARCHAR (1),
30     concept_code VARCHAR (50) NOT NULL,
31     valid_start_date DATE NOT NULL,
32     valid_end_date DATE NOT NULL,
33     invalid_reason VARCHAR (1)
34 );
```

④ Run --Main DDL in postgresQL

Tables (21)

- > concept
- > concept\_ancestor
- > concept\_class
- > concept\_class\_conversion
- > concept\_manual
- > concept\_relationship
- > concept\_relationship\_manual
- > concept\_relationship\_stage
- > concept\_stage
- > concept\_synonym
- > concept\_synonym\_manual
- > concept\_synonym\_stage
- > domain
- > drug\_strength
- > drug\_strength\_stage
- > pack\_content
- > pack\_content\_stage
- > relationship
- > relationship\_conversion
- > vocabulary
- > vocabulary\_conversion

⑤ Upload the corresponding tables from the vocabulary bundle in Athena

```
Vocabulary-v5.0 / working / DevV5_DDL.sql
Code Blame 314 lines (288 loc) · 14.1 KB
243 --Create PKs
244 ALTER TABLE concept ADD CONSTRAINT xpk_concept PRIMARY KEY (concept_id);
245 ALTER TABLE vocabulary ADD CONSTRAINT xpk_vocabulary PRIMARY KEY (vocabulary_id);
246 ALTER TABLE domain ADD CONSTRAINT xpk_domain PRIMARY KEY (domain_id);
247 ALTER TABLE concept_class ADD CONSTRAINT xpk_concept_class PRIMARY KEY (concept_class_id)
248 ALTER TABLE concept_relationship ADD CONSTRAINT xpk_concept_relationship PRIMARY KEY (con
249 ALTER TABLE relationship ADD CONSTRAINT xpk_relationship PRIMARY KEY (relationship_id);
250 ALTER TABLE concept_ancestor ADD CONSTRAINT xpkconcept_ancestor PRIMARY KEY (ancestor_con
251 ALTER TABLE drug_strength ADD CONSTRAINT xpk_drug_strength PRIMARY KEY (drug_concept_id,
252
253 --Create external keys
254 ALTER TABLE concept ADD CONSTRAINT fpk_concept_domain FOREIGN KEY (domain_id) REFERENCES
255 ALTER TABLE concept ADD CONSTRAINT fpk_concept_class FOREIGN KEY (concept_class_id) REFER
256 ALTER TABLE concept ADD CONSTRAINT fpk_concept_vocabulary FOREIGN KEY (vocabulary_id) REF
```

⑥ Create PKs, external keys, indexes and checks



# Community Contribution Process

- **Prerequisites**

- Preparation Schemas : *sources, dev\_xyz, devv5*
- Copies of tables, fully indexed (downloaded from Athena and put into Schema *dev\_xyz, devv5*)

- **Process**

1. Run *load\_stage.sql* in the *dev\_xyz* schema
2. QA/QC part 1
3. Generic update
4. QA/QC part 2 : semi-automatic process



# Community Contribution Process

- **Prerequisites**

- Preparation Schemas : *sources, dev\_xyz, devv5*
- Copies of tables, fully indexed (downloaded from Athena and put into Schema *dev\_xyz, devv5*)

- **Process**

1. Run *load\_stage.sql* in the *dev\_xyz* schema
2. QA/QC part 1
3. Generic update
4. QA/QC part 2 : semi-automatic process



# Community Contribution Process

- **Process**
  1. **Run *load\_stage.sql* in the *dev\_xyz* schema**
    - **If *load\_stage* references *devv5* or *sources*, replace them with the names of your schema**



# Community Contribution Process

- **Process**

1. **Run *load\_stage.sql* in the *dev\_xyz* schema**

- **If *load\_stage* references *devv5* or *sources*, replace them with the names of your schema**
- **Run function *Vocabulary\_pack.SetLatestUpdate.sql***

```
DO $_$
BEGIN
    PERFORM VOCABULARY_PACK.SetLatestUpdate(
        pVocabularyName      => 'vocabulary_id of your vocabulary as in Vocabulary table',
        pVocabularyDate      => 'date of new vocabulary version',
        pVocabularyVersion   => 'name of the vocabulary version, if none use date of the version',
        pVocabularyDevSchema => 'name of your development schema'
    );
END $_$;
```



# Community Contribution Process

- **Process**

1. **Run *load\_stage.sql* in the *dev\_xyz* schema**

- **If *load\_stage* references *devv5* or *sources*, replace them with the names of your schema**
- **Run function *Vocabulary\_pack.SetLatestUpdate.sql***
- **Run supporting functions to stage tables**

Name	Last commit message
..	
ATCPostprocessing.sql	added ATC postprocessing [AVOF-2548]
AddFreshMAPSTO.sql	improved description
AddFreshMapsToValue.sql	fix bug with incorrect processing of mapping...
BuildRxE.sql	refactoring
CheckManualConcepts.sql	first commit
CheckManualRelationships.sql	first commit

Use these support functions:

```
VOCABULARY_PACK.ProcessManualRelationships()  
VOCABULARY_PACK.AddFreshMAPSTO()  
VOCABULARY_PACK.DeprecateWrongMAPSTO()  
VOCABULARY_PACK.DeleteAmbiguousMAPSTO()
```

① You should download function queries you need

③ Run query

④ If you've done it correctly, it should be contained within the *vocabulary\_pack* schema functions



# Community Contribution Process

- **Process**

- 2. **QA/QC part 1**

- **As a result of previous step, you will populate stage tables**  
(*concept\_stage*, *concept\_relationship\_stage*, *concept\_synonym\_stage*, etc)
    - **Run *qa\_tests.check\_stage\_tables ()* in *create\_qa\_tests.sql***

```
CREATE OR REPLACE FUNCTION qa_tests.check_stage_tables ()
RETURNS TABLE (
    error_text TEXT,
    rows_count BIGINT
) AS
$BODY$
BEGIN
    RETURN QUERY
    SELECT reason, COUNT(*) FROM (
        --concept_relationship_stage
        SELECT
            CASE WHEN v1.vocabulary_id IS NOT NULL AND
```

This function performs conformance checks and should return no errors.

- If the concept is valid, check *valid\_end\_date* = 12/31/2099
      - Filed length does not exceed limits in the standard DDL
        - There are no duplicates
      - *Vocabulary\_id* exist in *VOCABULARY* table



# Community Contribution Process

- **Process**

- 3. Generic update**

- **This function integrates the content of the *\_stage* tables into basic tables and assigns *concept\_ids*.**
    - **If you execute this function, you can inspect *CONCEPT* and *CONCEPT\_RELATIONSHIP***

```
CREATE OR REPLACE FUNCTION devv5.GenericUpdate (  
)  
RETURNS void AS  
$BODY$  
BEGIN  
    --1. Prerequisites:  
    --1.1 Check vocabulary table, at least one vocabulary must  
    PERFORM FROM vocabulary WHERE latest_update IS NOT NULL LI  
    IF NOT FOUND THEN  
        RAISE EXCEPTION 'At least one vocabulary must ha  
        USING HINT = 'Forgot to execute Setlat  
    END IF;  
  
    --1.2 Check stage tables for incorrect rows  
    DO $$
```

If you need to modify your scripts or stage tables,  
you will need to clean stage tables and revert basic tables.

You can easily erase all changes using *fast\_recreate\_schema.sql*



# Community Contribution Process

- Process

- 4. QA/QC part 2 : semi-automatic process

- Execute *get\_checks* to check the compliance of resulting basic tables to the OMOP rules
    - Run *manual\_checks\_after\_generic* to review manually

```
CREATE OR REPLACE FUNCTION qa_tests.get_checks (checkid IN INT DEFAULT NULL)
RETURNS TABLE
(
    check_id int4,
    check_name VARCHAR(1000),
    concept_id_1 int4,
    concept_id_2 int4,
    relationship_id VARCHAR(20),
    valid_start_date DATE,
    valid_end_date DATE,
    invalid_reason VARCHAR(1)
)
AS $BODY$
--relationships cycle
```

\* This query is very heavy  
It may take a long time to execute.



# Community Contribution Process

1. Send the source file + stage tables to Vocabulary Team
2. Create a pull request in Vocabulary GitHub

 concept_relationship_manual	2024-06-19 오후 11:00	Microsoft Excel Co...	24,546 KB
 concept_relationship_stage	2024-06-19 오후 11:01	Microsoft Excel Co...	64,568 KB
 concept_stage	2024-06-19 오후 11:00	Microsoft Excel Co...	58,724 KB
 concept_synonym_stage	2024-06-19 오후 11:00	Microsoft Excel Co...	42,964 KB
 sources.edi_data	2024-06-19 오후 11:00	Microsoft Excel Co...	118,135 KB

Commits

History for Vocabulary-v5.0 / EDI on `master`

All users All time

Commits on Sep 12, 2024

edi\_mapped moved to he manual\_work folder, update of source schema removed  
olegzhuik committed on Sep 12 603c319

Commits on Aug 21, 2024

Formatting changed  
olegzhuik committed on Aug 21 4ba20a1

Minor changes  
olegzhuik committed on Aug 21 aeeabc1

Commits on Aug 12, 2024

Update revision requests  
parkyjoo committed on Aug 12 5e42eb5

Commits on Aug 8, 2024

EDI update  
parkyjoo committed on Aug 8 948f95d



# Korean EDI Updates in ATHENA

**ATHENA** SEARCH DOWNLOAD LOGIN ?

SEARCH BY KEYWORD aspirin

EDI x

DOMAIN

filter

- Device (36196)
- Drug (62064)
- Measurement (20602)
- Procedure (318945)
- Condition (0)
- Condition Status (0)
- Condition/Device (0)
- Condition/Meas (0)
- Condition/Obs (0)

DOWNLOAD RESULTS Show by 1... items Total 437,807 items

Download Results	Show by	1...	items	Total	437,807	items
42086580C5467480	(MULTI-FIX) CALCANEAL LOCKING PLATE	Device	Non-standard	Valid	Device	EDI
42086581C3111180	(MULTIFIX) NAIL SYSTEM - FEMORAL TYPE	Device	Non-standard	Valid	Device	EDI
42086582C3191180	(MULTIFIX) NAIL SYSTEM - LOCKING SCREW	Device	Non-standard	Valid	Device	EDI
42086583C3115080	(MULTIFIX) NAIL SYSTEM - PROXIMAL TYPE (PFN)	Device	Non-standard	Valid	Device	EDI
42086584C3112180	(MULTIFIX) NAIL SYSTEM - TIBIAL TYPE	Device	Non-standard	Valid	Device	EDI
42105220M1004120	(REGULATOR)OLEFUSION	Device	Non-standard	Valid	Device	EDI
42086592J3022001	0.035 FIBERED PLATINUM COILS	Device	Non-standard	Valid	Device	EDI

**DETAILS**

Domain ID	Drug
Concept Class ID	Drug Product
Vocabulary ID	EDI
Concept ID	42239803
Concept code	643104140
Validity	Valid
Concept	Non-standard

**TERM CONNECTIONS (1)**

RELATIONSHIP	RELATES TO	CONCEPT ID	VOCABULARY
Non-standard to Standard map (OMOP)	ambroxol-theophylline-7-acetate 100 MG Oral Capsule [PHYLLICOL]	42925779	RxNorm Extension

**LANGUAGE** **SYNONYM CONCEPT**

Korean	필리콜캡슐(아세브로필린)_ (0.1g/1캡슐)
--------	---------------------------

Valid start: 01-Sep-2016

Valid end: 31-Dec-2099

We integrated an updated version of the Korean EDI in September 2024



**Thank You!**

Yiju Park (yijoo0320@yuhs.ac)  
Seng Chan You (seng.chan.you@ohdsi.org)