

# Terminology Mappings for MSDW OMOP

**Timothy Quinn, PhD**

Scientific Computing

Icahn School of Medicine at Mount Sinai

November 29, 2022



Icahn  
School of  
Medicine at  
**Mount  
Sinai**

# Agenda Topics

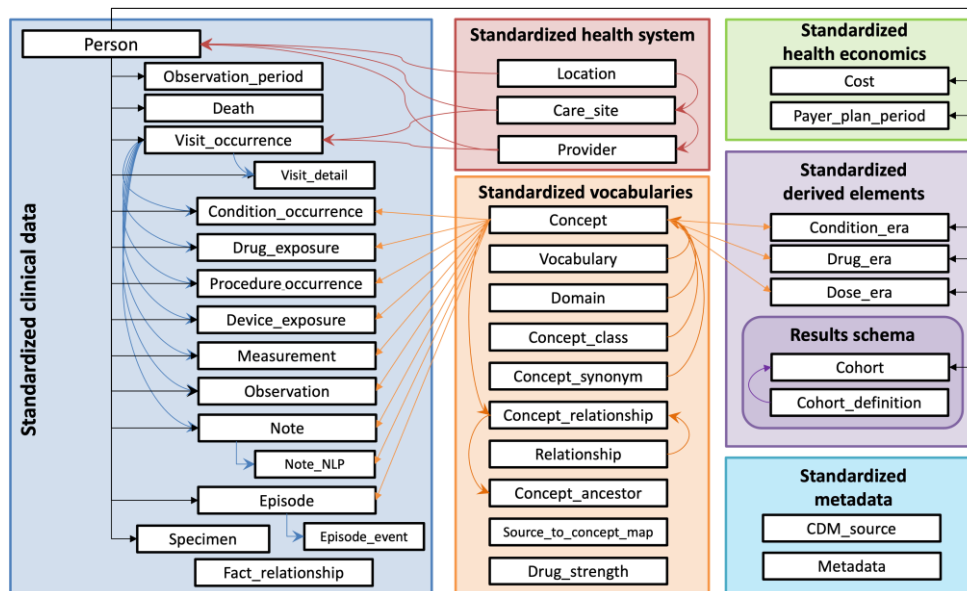
1. **Terminology Mappings for MSDW OMOP**
2. **Data Quality in MSDW OMOP**
3. **Enterprise Data Catalog**

# Terminology Mappings for MSDW OMOP

1. **MSDW uses the OMOP Common Data Model**
2. **OMOP Common Data Model Requirements**
  - Standardize data structure via common format
  - Standardize data content via mapping to standard healthcare vocabularies
3. **Vendor: Clinical Architecture (aka “Symedical”)**
4. **Mapping Epic’s 7 major master files**
  - Diagnosis (EDG)
  - Procedure (EAP)
  - Surgical Procedure (ORP)
  - Lab Component (LRR)
  - Medication (ERX)
  - Immunization (LIM)
  - Allergy (ELG)

# OMOP Common Data Model Requirements

## 1. Standardize **data structure** via common format



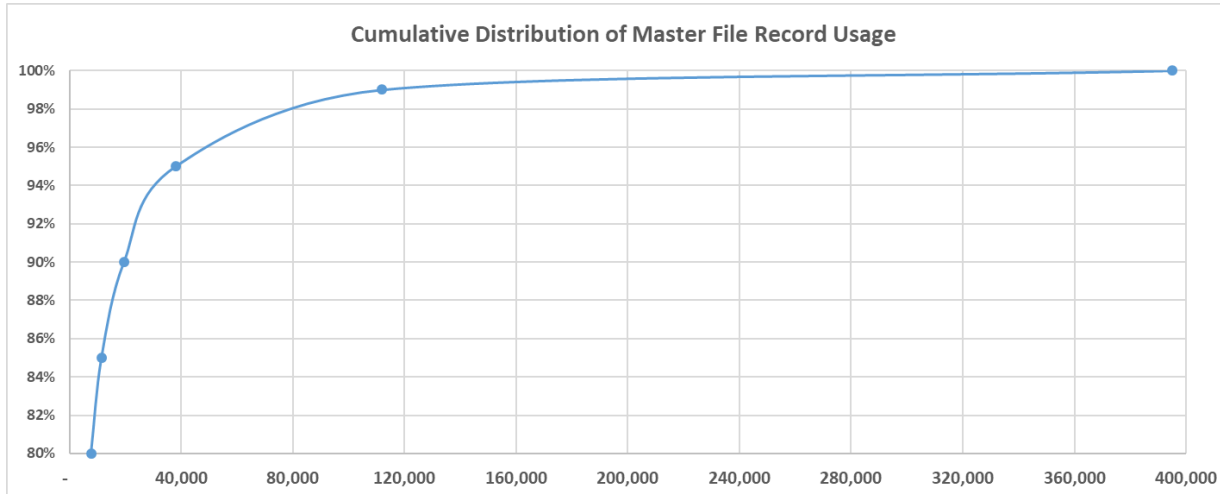
## 2. Standardize **data content** via mapping EHR codes to standard healthcare vocabularies

OMOP Domain	Standard Vocabularies	Non-standard Vocabularies
Condition	SNOMED-CT	ICD-10-CM, ICD-9-CM
Drug	RxNorm, CVX	ATC, NDC, Multum
Measurement	LOINC	SNOMED-CT, Nebraska Lexicon
Procedure	CPT4, HCPCS, ICD-10-PCS	ICD-9-Proc
Observation	SNOMED-CT, LOINC	ICD-10-CM, ICD-9-CM
Race, Ethnicity	OMOP Race, OMOP Ethnicity	SNOMED-CT, Nebraska Lexicon
Provider (Specialty)	NUCC, Medicare Specialty	SNOMED-CT, Nebraska Lexicon
Route	SNOMED-CT	Nebraska Lexicon
Unit	UCUM	SNOMED-CT, Nebraska Lexicon

## Epic Master Files – *as of May 4, 2022*

	Epic Master File	Target Vocabulary	Used Records	Total Records	“Gold Standard” Mappings
1.	Diagnosis (EDG)	SNOMED	324,721	1,566,543	1,942
2.	Diagnosis (EDG)	ICD10CM	324,721	1,566,543	–
3.	Procedure (EAP)	CPT / HCPCS	26,159	154,086	126
4.	Surgical Procedure (ORP)	CPT / HCPCS	4,712	31,486	486
5.	Lab Component (LRR)	LOINC	21,866	33,891	585
6.	Medication (ERX)	RxNorm	53,376	134,037	1,336
7.	Immunization (LIM)	CVX / RxNorm	333	399	–
8.	Allergy (ELG)	SNOMED / RxNorm	8,249	38,914	–

# Most Master File Records Used Infrequently (“Long Tail”)



Priority	Master File	Target Vocabulary	Rows for 80%	Rows for 85%	Rows for 90%	Rows for 95%	Rows for 99%	Used Rows	Total Rows
1	Diagnosis EDG	SNOMED-CT	6,000	9,000	16,000	32,000	96,000	324,721	1,566,543
2	Diagnosis EDG	ICD10CM	6,000	9,000	16,000	32,000	96,000	324,721	1,566,543
3	Lab Component LRR	LOINC / SNOMED	200	300	400	600	1,800	21,866	33,891
4	Medication ERX	RxNorm	800	1,000	1,400	2,400	6,200	53,376	134,037
5	Procedure EAP	CPT / HCPCS	200	400	600	1,000	3,000	26,159	154,086
6	Surgical Procedure ORP	CPT / HCPCS	400	600	800	1,400	2,600	4,712	31,486
7	Immunization LIM	CVX / RxNorm	30	40	50	80	130	333	399
8	Allergy ELG	SNOMED-CT / RxNorm	150	200	350	700	2,300	8,249	38,914
		<b>TOTAL:</b>	<b>7,780</b>	<b>11,540</b>	<b>19,600</b>	<b>38,180</b>	<b>112,030</b>	<b>439,416</b>	<b>1,959,356</b>
			80%	85%	90%	95%	99%	100%	

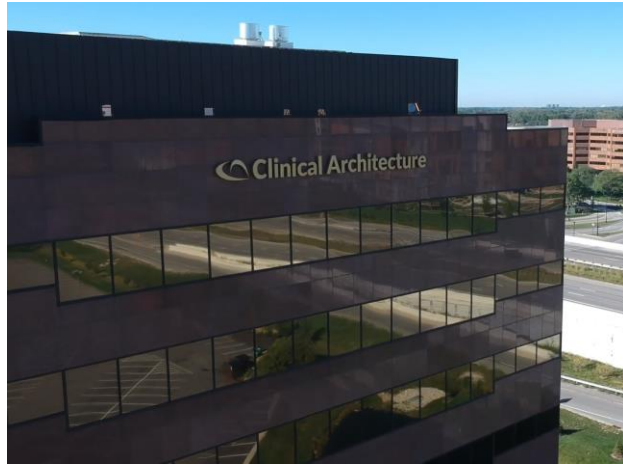
## MSDW's OMOP Mappings – *as of Oct 5, 2022*

OMOP Table	OMOP Column	Epic Fill Percent	MSDW Mapped Percent
person	ethnicity_concept_id	0.242 %	100.0 %
person	gender_concept_id	99.7 %	100.0 %
person	race_concept_id	24.5 %	100.0 %
provider	gender_concept_id	3.3 %	100.0 %
provider	specialty_concept_id	60.4 %	68.7 %
condition_occurrence	condition_concept_id	100.0 %	97.2 %
drug_exposure	drug_concept_id	100.0 %	87.2 %
measurement	measurement_concept_id	99.953 %	85.5 %
procedure_occurrence	procedure_concept_id	99.4 %	81.6 %
observation	observation_concept_id	99.996 %	100.0 %

# Terminology Mapping Vendor: Clinical Architecture (aka “Symedical”)



# Terminology Mapping Services Vendor



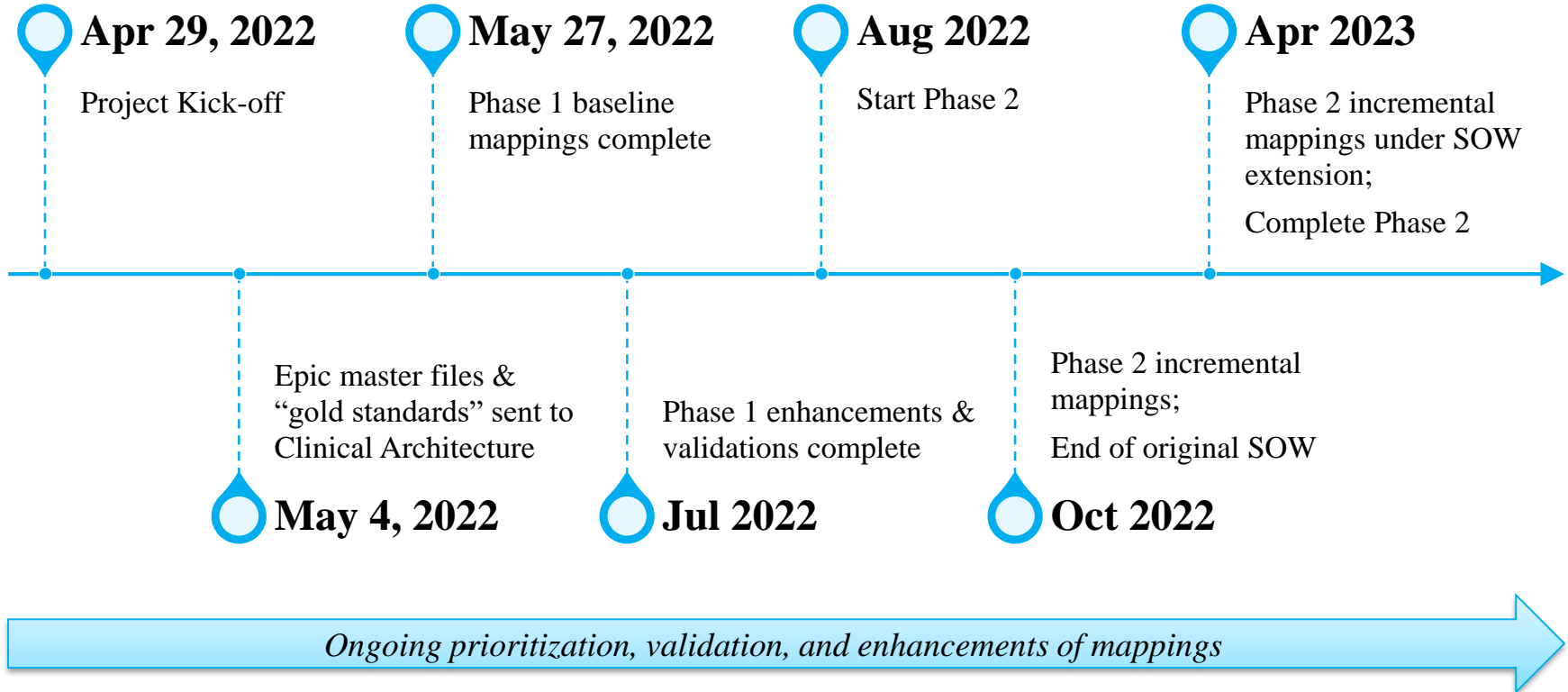
11611 North Meridian St  
Suite 500  
Carmel, IN 46032

**Leading provider of healthcare terminology mapping services**

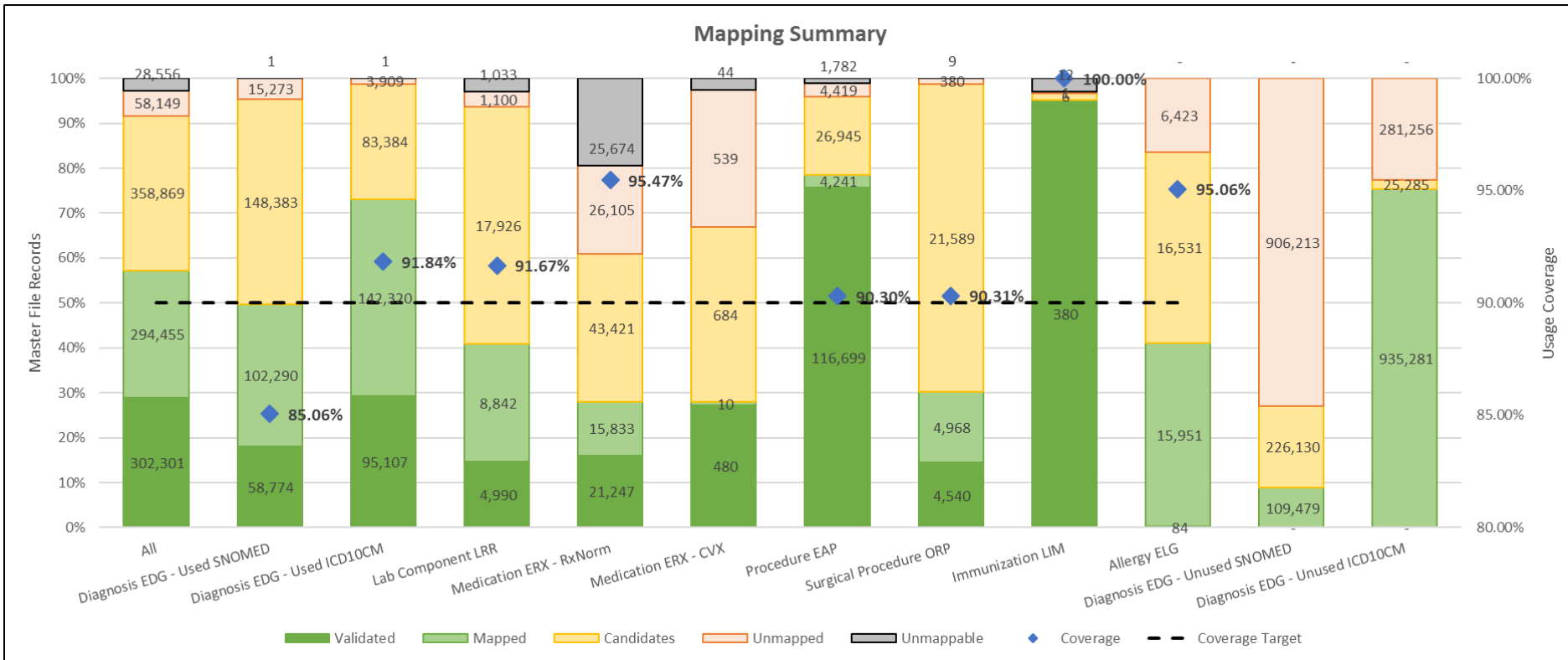
## **Symedical mapping software**

- ▶ Multiple matching algorithms
- ▶ Matching algorithms tailored to domain
- ▶ Each algorithm tuned for domain and customer data
- ▶ Underlying data model, dictionaries, thesauri for each domain

# Clinical Architecture Project Timeline

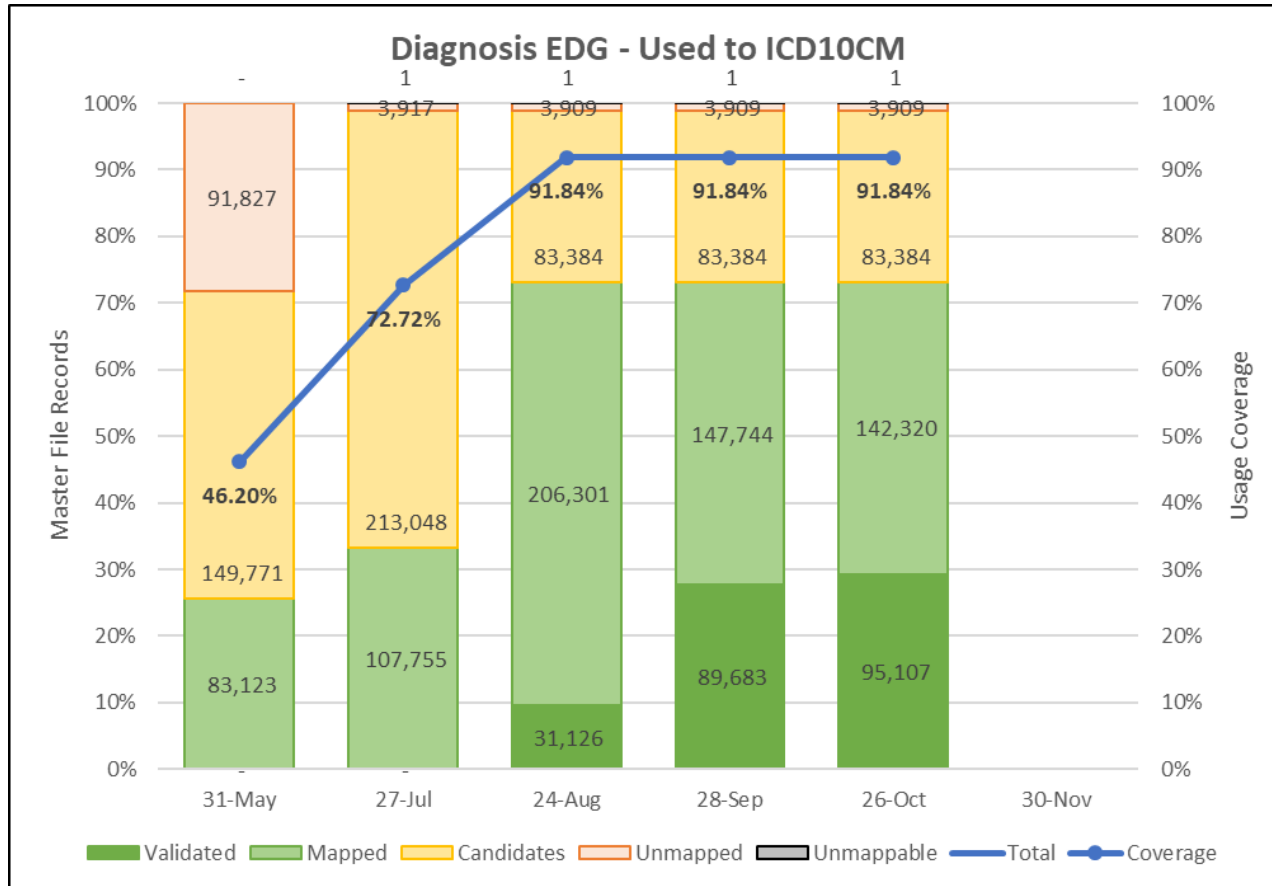


# Mapping Progress to 90% Usage Coverage Target – as of Oct 26

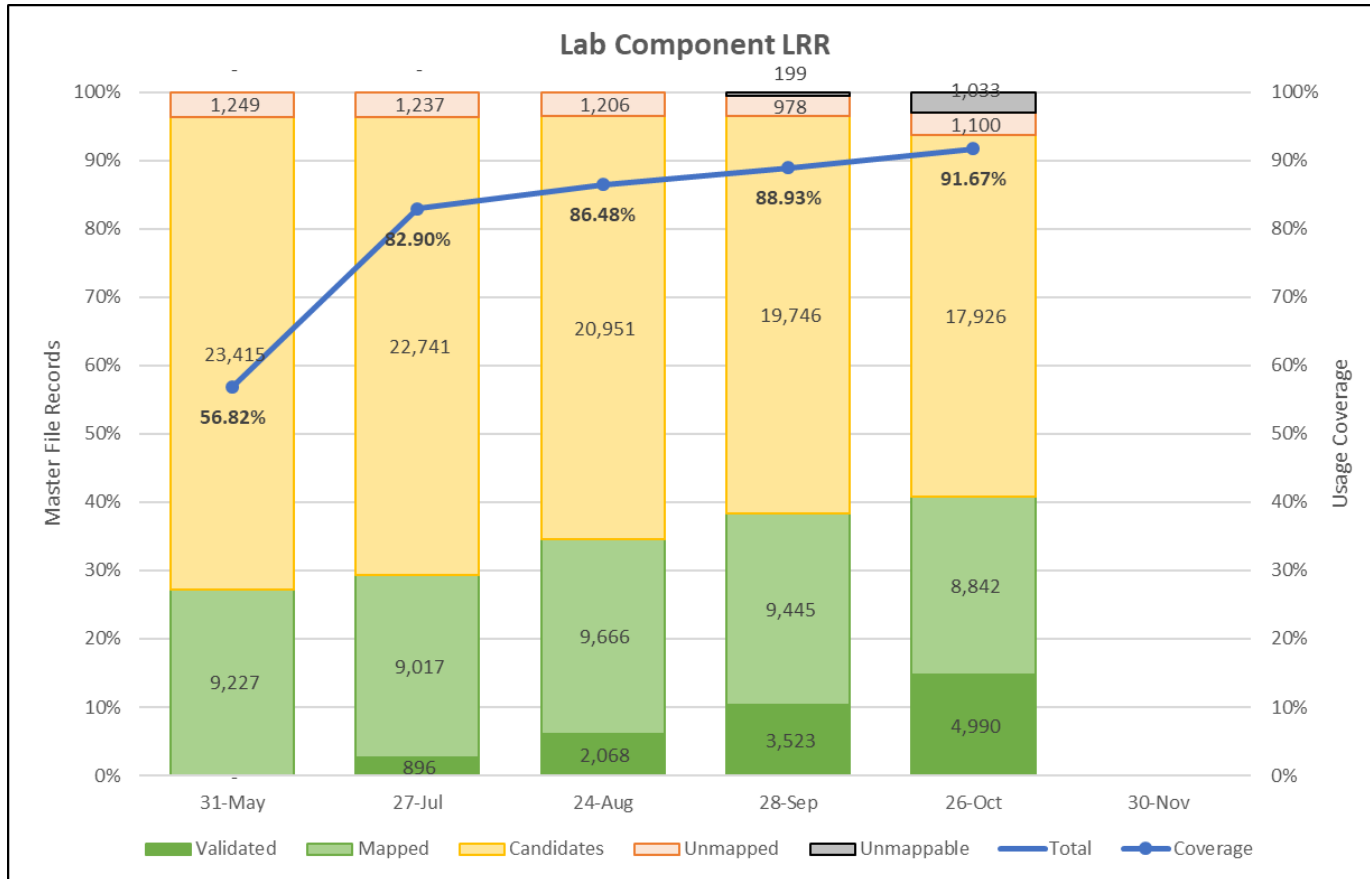


# Terminology Mapping Progress over Time May – Oct 2022

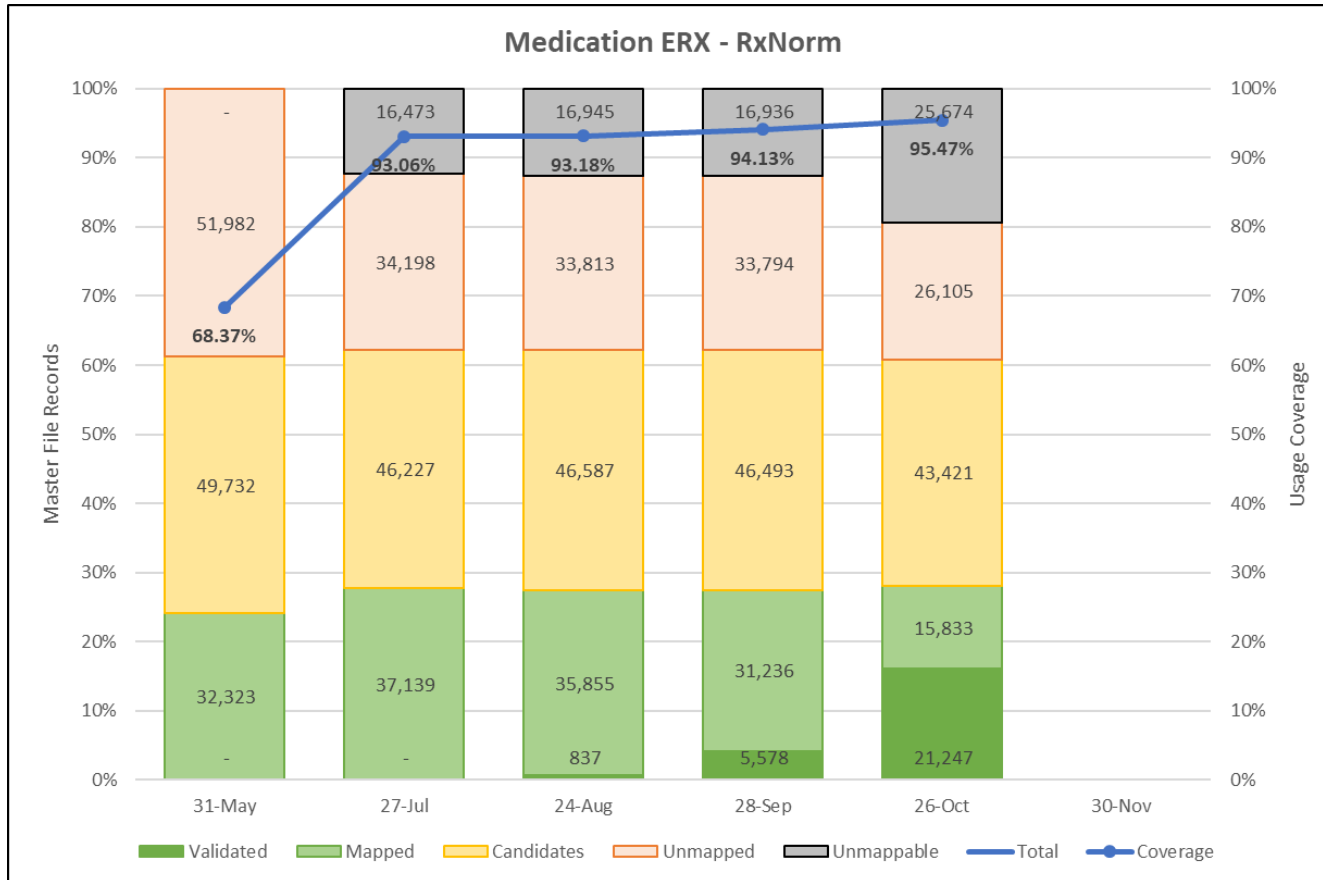
# Mapping Example 1: Epic Diagnosis (EDG) to ICD-10-CM



## Mapping Example 2: Epic Lab Component (LRR) to LOINC



# Mapping Example 3: Epic Medication (ERX) to RxNorm



# Data Quality in MSDW OMOP



# Methods for Ensuring Data Quality

1. **Data pipeline from Epic to OMOP**
  - Metadata-driven
  - Scanned & included in Enterprise Data Catalog
2. **Constraints in MSDW's OMOP database**
  - Prevent loading invalid values
3. **“Business Rules Engine”**
  - Automated data-quality checks
  - Built into MSDW, executed on a schedule
  - Currently have 306 rules, continually expanding
  - Many are automatically “generated” by metadata
4. **OHDSI tools ACHILLES and ACHILLES Heel**
  - Built into ATLAS cohort query tool

# Informatica Enterprise Data Catalog

# Data Catalog Automatically “Scans” Data Assets

