# LOINC Document Ontology

## Clinical Note Mapping

April 25, 2023

# OVERVIEW

- **LOINC Document Ontology**
  - What is it?
  - What have people done with it?
- **LOINC Code Mapping**
  - Data Staging
  - Bag of Words Approach
  - Improvements
- **Results & Analysis + Discussion**

# A. LOINC Document Ontology

# A.1 LOINC Document Ontology (LOINC DO)

## Logical Observation Identifiers, Names, and Codes

**Assign a LOINC Code to all clinical documents**

- Ease of organization
- Gives a broad view the content
- Makes documents Query-able

**Documents are assigned LOINC Code based on FIVE Dimensions**

**Each dimension requires a Part Number and a Part Name**

**It is not necessary for every LOINC Coded document to have all FIVE dimensions mapped**

# A.2 Document Ontology Dimensions

## LOINC DO Dimensions

- **Subject Matter Domain**
  - Dentistry, Allergy, etc…
- **Kind of Document**
  - Note, Report, Prescription list, etc…
- **Type of Service**
  - Consultation, Procedure, Counseling, etc…
- **Setting**
  - Outpatient, Telehealth, Pharmacy, etc…
- **Role**
  - Patient, Nurse, Physician, etc…

| PARTNAME | PARTNUMBER |
|---|---|
| {Role} | LP187187-2 |
| Physician | LP173084-7 |
| Attending | LP269965-2 |
| Social worker | LP269801-9 |
| Consultant | LP269966-0 |
| Nurse | LP173075-5 |
| Medical student | LP173092-0 |
| Hygienist | LP173071-4 |
| Interdisciplinary | LP173072-2 |
| Patient | LP173083-9 |
| Pharmacist | LP181523-4 |
| Licensed practical nurse | LP173081-3 |
| Resident | LP269969-4 |
| Team | LP173073-0 |
| Rapid response team | LP203036-1 |
| Midwife | LP221282-9 |
| Clinical nurse specialist | LP173078-9 |
| Intern | LP269968-6 |
| Technician | LP173094-6 |
| Therapist | LP173095-3 |

# A.3 Value-Set Mapping

## LOINC Document Ontology Manual Mapping

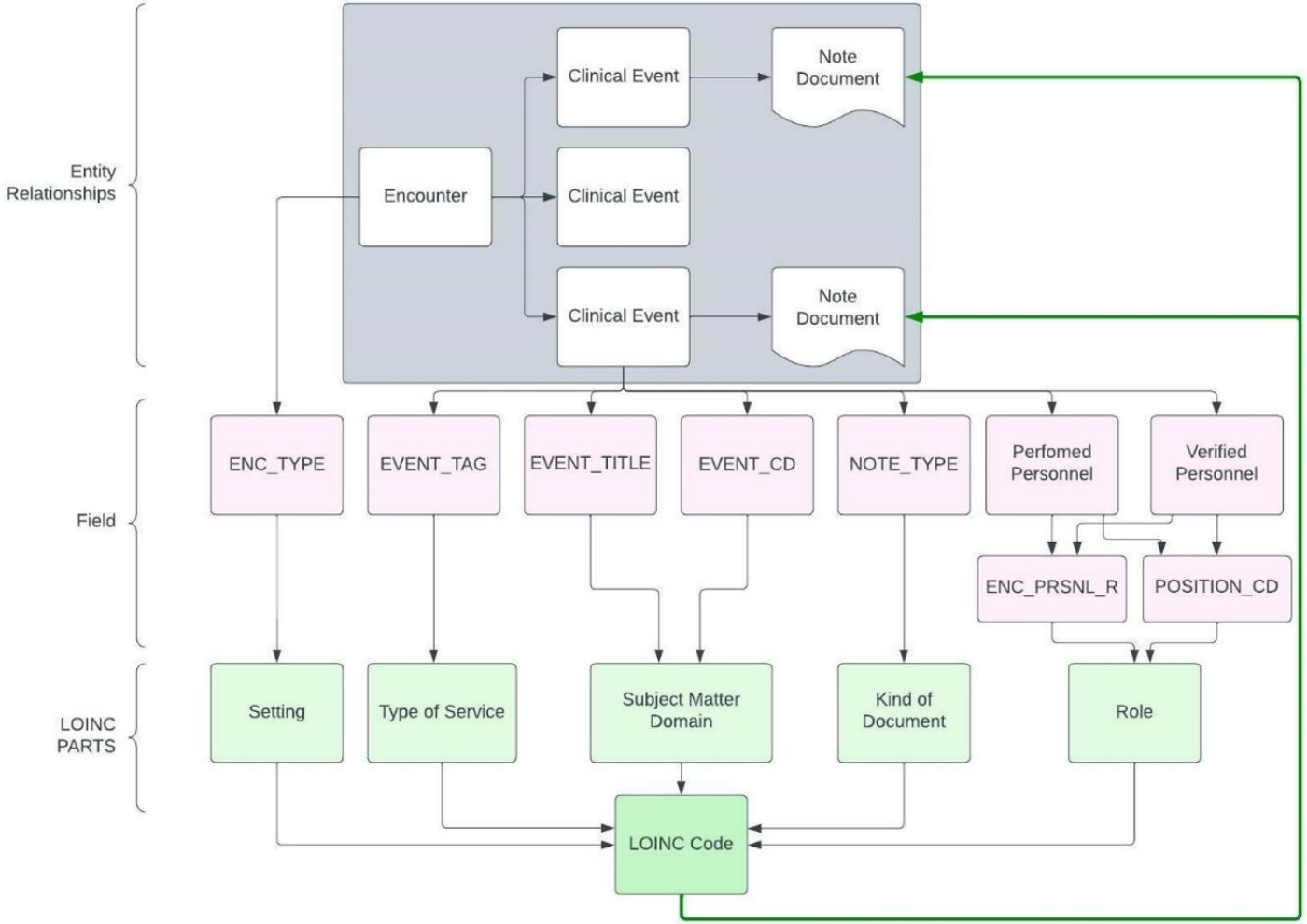
- **Subject Matter Domain**
  - EVENT_CD Display Value (5671 Distinct Values, mapped **56 (30%)** of top 100 (90%))
- **Kind of Document**
  - NOTE_TYPE_DESCRIPTION (1007 Distinct Values, mapped **85 (50%)** of top 100 (94%))
- **Type of Service**
  - EVENT_TAG (45,439 Distinct Values, mapped **44 (40%)** of top 100 (92%))
- **Setting**
  - ENCNTR_TYPE_CD Display Value (125 Distinct Values, mapped **82 (80%)**)
- **Role Mapping**
  - ENCNTR_PRSNL_R_CD Display Value (84 Distinct Values mapped **50 (35%)**)

- **LOINC Code**
  - Even though around 45% of our notes have at least 3 Dimensions mapped, only around 25% of our notes have a valid LOINC Code assigned to them.

- **Programmatic Mapping**
  - Will give us more coverage
  - Will be repeatable and Extensible
  - Bag of Words

| Num of Parts | Number of Notes | % |
|---|---|---|
| 0 | 7,225,678 | 5% |
| 1 | 25,015,785 | 19% |
| 2 | 41,216,686 | 31% |
| 3 | 33,005,516 | 25% |
| 4 | 21,267,870 | 16% |
| 5 | 4,899,873 | 4% |
| Total | 132,631,408 | |

# A.3 Value-Set Mapping Diagram




## Summary

- The diagram visualizes the columns used for value-set mapping.
- Specific fields (columns) were chosen for each of the LOINC dimensions, as shown by the arrows between the fields and the LOINC PARTS

# B. LOINC Code Mapping

# B.1 Data Staging Details (Cerner)

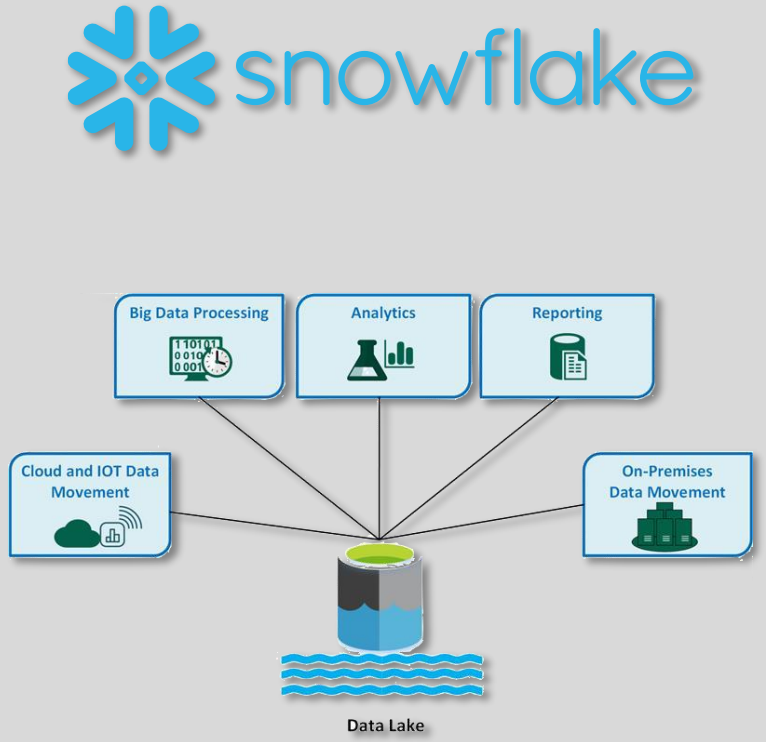
## Transferring to Snowflake

- We stage our data from Cerner Oracle Database into Snowflake data-lake
- Used AWS Database Migration Service (DMS)
  - Oracle to AWS S3 bucket using DMS
  - Create appropriate tables in Snowflake
  - Transfer data from AWS S3 bucket to Snowflake
  - Automatic nightly refresh pipeline for metadata using AWS EventBridge and AWS Lambda functions.

## Notes Decompression

- Cerner keeps note data in BLOBs (Binary Large Objects) format
- We use Snowflake Java UDF (User Defined Functions) to decompress
- Separate AWS Pipeline for files larger than 16MB, which is a snowflake Limit.

Reach out to us and we can provide more details for each step!



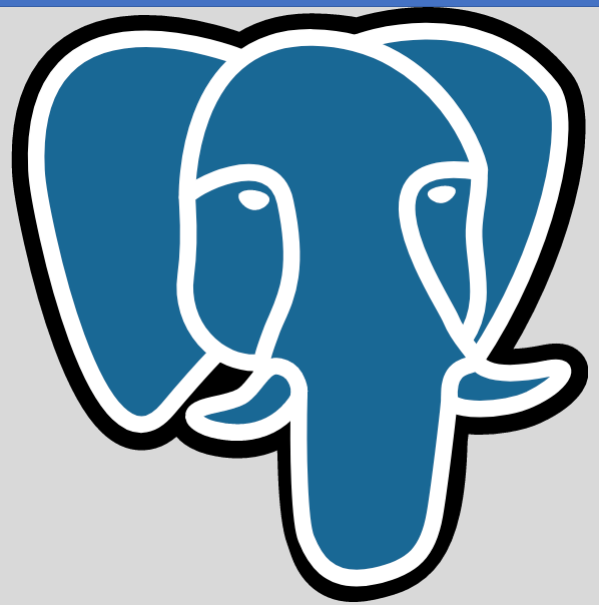
# B.2 Data Staging Details (Epic)

## Extracting Metadata

- Our ETL -> postgres server
- Using only HNO notes
- Extracting metadata
  - Assumptions/Filters
    - Valid patients
    - Duplicates
  - Speeding up joins

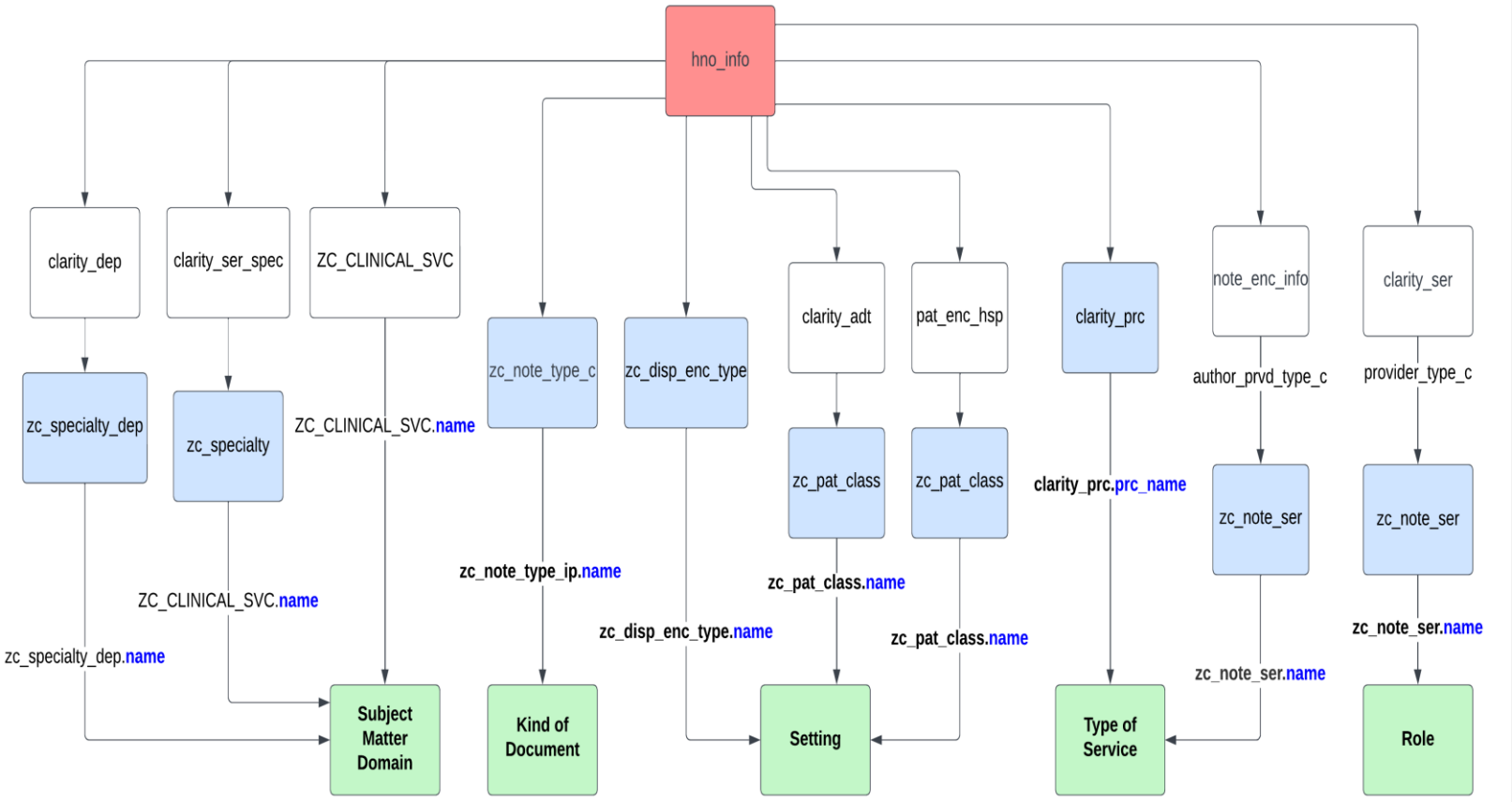
## Post extraction filtering

- Document types we don't want
  - AVS snapshot
  - Scanned document
  - nulls



Epic

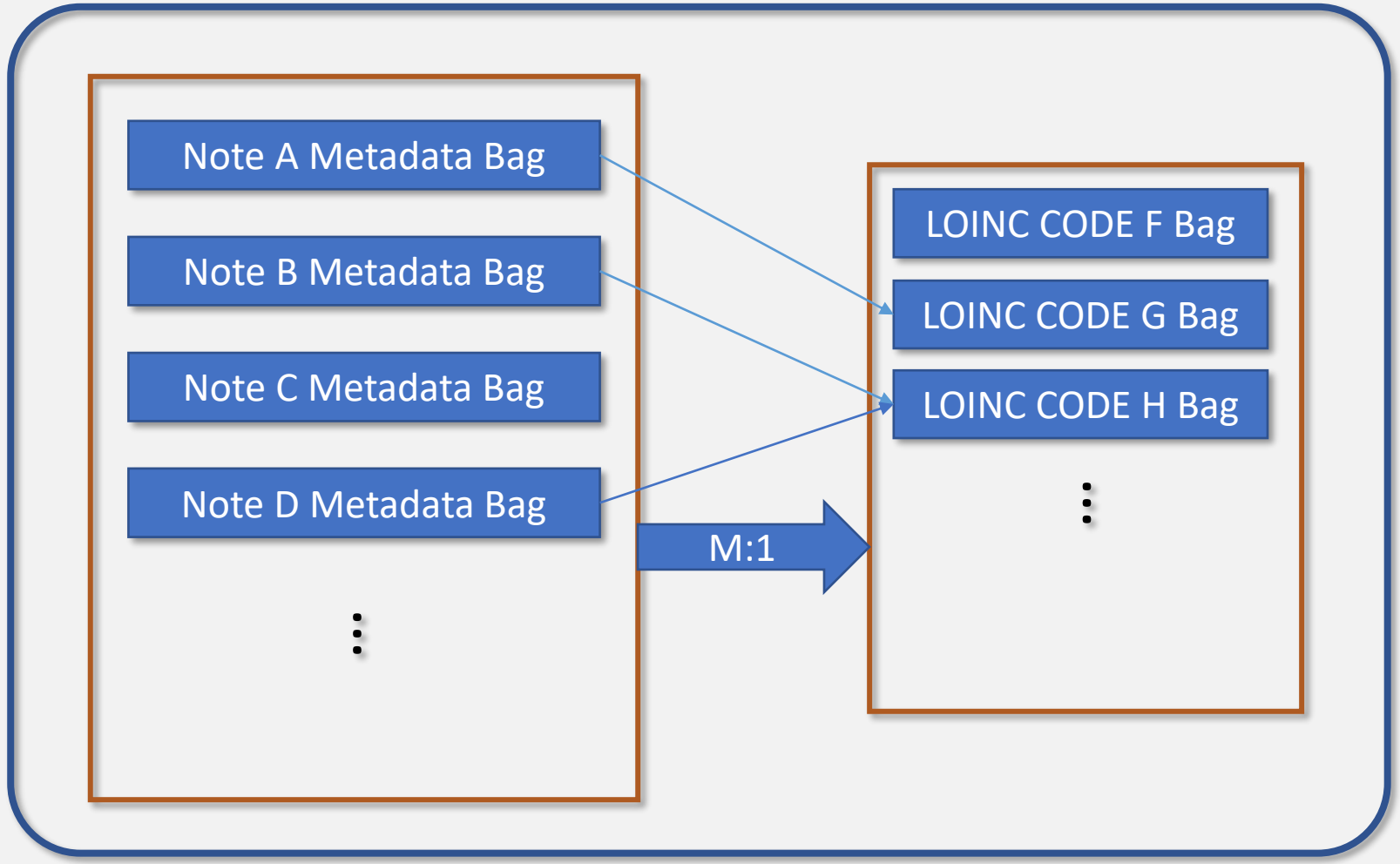
# B.3 LOINC Mapping Diagram (EPIC)




## Summary

- This diagram visualizes the columns used for the LOINC Mapping
- Multiple sources are used for domains
- HNO notes are the only source

# B.3 Bag of Words




**Mapping Notes Metadata to LOINC Codes**

- A Bag of words for all LOINC Code values
- A Bag of words for each Note
- Find single best match for every Note
- This way we don't need to map part numbers individually
- Synonymy List for Coverage Improvement

# B.4 Bag of Words (LOINC Codes)

| | LoincNum | PartNumb | PartTypeName | | PartName |
|---|---|---|---|---|---|
| 38 | 100446-4 | LP173051-6 | Document.Setting | . | Outpatient |
| 39 | 100446-4 | LP420041-8 | Document.SubjectMat... | . | Breastfeeding |
| 40 | 100447-2 | LP173418-7 | Document.Kind | . | Note |
| 41 | 100447-2 | LP173213-2 | Document.TypeOfService | . | Progress |
| 42 | 100447-2 | LP173051-6 | Document.Setting | . | Outpatient |
| 43 | 100447-2 | LP268363-1 | Document.SubjectMat... | . | Burn management |
| 44 | 100448-0 | LP173418-7 | Document.Kind | . | Note |
| 45 | 100448-0 | LP173213-2 | Document.TypeOfService | . | Progress |
| 46 | 100448-0 | LP173051-6 | Document.Setting | . | Outpatient |
| 47 | 100448-0 | LP207300-7 | Document.SubjectMat... | . | Cardiac surgery |
| 48 | 100449-8 | LP173418-7 | Document.Kind | . | Note |
| 49 | 100449-8 | LP173213-2 | Document.TypeOfService | . | Progress |

## LOINC Code Bag of Words

- Each LOINC Code/Number has individual parts between 2 to 6
- Example code **100447-2** has 4 Part Numbers, each represented by a row.
- A Bag of words for **100447-2** would be represented as

**{“note”, “progress”, “outpatient”, “burn”, “management”}**

# B.5 Bag of Words (Notes Metadata)

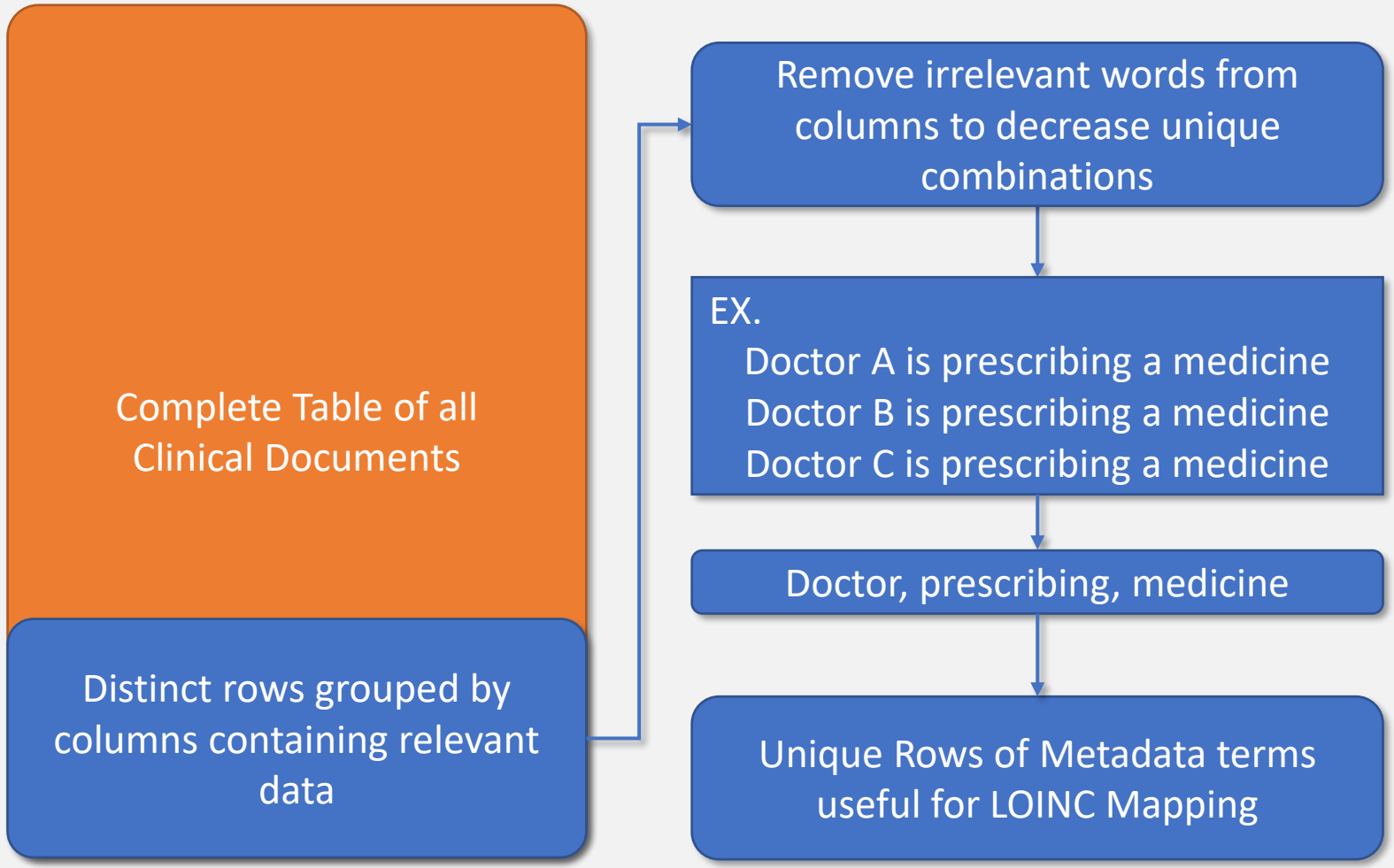
| | EVENT_TITLE_TEXT | EVENT_TAG | EVENT_CD_DESC | LOC_FACILITY_CD_DESC | VERIFIED_ENCNTR_P |
|---|---|---|---|---|---|
| 0 | Care Coordination Narrative | Care Coordination Narrative | Care Coordination Narrative | University Hospital | Care Delivery |
| 0 | Care Coordination Narrative | Care Coordination Narrative | Care Coordination Narrative | MU Ambulatory Care Management | Assigned Care Ma... |
| 0 | | Results Notification to Patient | Results Notification to Patient | CRP-Versailles | NaN |
| 0 | | Discharge Information | Discharge Information | MIZZOU URGENT CARE | Clerical Support |
| 0 | Addendum by Thompson LPNSR, Bre... | General Message | General Message | UP-FAYETTE FAMILY MEDICINE ... | NaN |
| 0 | Clinic Visit Narrative | Clinic Visit Narrative | Clinic Visit Narrative | LRHS Lake Regional Heart & ... | Care Delivery |
| 0 | | Prelim | Prelim | WOMENS AND CHILDRENS HOSPITAL | NaN |
| 0 | Report | Report | Report | UP-MISSOURI ORTHOPAEDIC INS... | NaN |
| 0 | | Prescription Note | Prescription Note | UP-GM FAMILY MEDICINE CLINI... | NaN |
| 0 | | Progress Note | Progress Note | University Hospital | Care Delivery |
| 0 | Report | Report | Report | University Hospital | NaN |
| 0 | Addendum by Tuggle LPN, Mary M ... | General Message | General Message | UP-OB/GYN ASSOCIATES | NaN |
| 0 | | Dermatology Clinic Note | Dermatology Clinic Note | UP-JEFF CITY DERMATOLOGY | Attending Physician |
| 0 | Respiratory Therapy Narrative | Respiratory Therapy Narrative | Respiratory Therapy Narrative | University Hospital | Care Delivery |
| 0 | Care Coordination Narrative | Care Coordination Narrative | Care Coordination Narrative | UP-SPMB Family Medicine | Care Delivery |
| 0 | ED Patient Depart Summary | ED Patient Depart Summary | ED Patient Depart Summary | WOMENS AND CHILDRENS HOSPITAL | Care Delivery |
| 0 | | General Message | General Message | University Hospital | Other |
| 0 | Intra-Operative Record | Perioperative SurgiNet Note | Perioperative SurgiNet Note | University Hospital | NaN |
| 0 | | Care Plan Note | Care Plan Note | Capital Region Medical Center | Care Delivery |
| 0 | Addendum by Sohl MD, Kristin Am... | General Message | General Message | Mizzou Therapy Children's H... | NaN |
| 0 | oncology clinic note | Result In Error | Hem Oncology IM Clinic Note | ELLIS FISCHEL CANCER CLINICS | NaN |
| 0 | Assets/Strengths Related to Tx ... | Assets/Strengths Related to T... | Assets/Strengths Related to T... | University Hospital | Care Delivery |
| 0 | Ambulatory Patient Depart Summary | Ambulatory Patient Depart Sum... | Ambulatory Patient Depart Sum... | ELLIS FISCHEL CANCER CLINICS | Care Delivery |
| 0 | Addendum by Ahmed MD, Ramia on ... | General Med IM Clinic Note | General Med IM Clinic Note | UP-FAIRVIEW GENERAL INTERNA... | Other |
| N | 70486 | CT SINUSES | 70486-26 HX | University Hospital | NaN |
| 0 | Addendum by Wilbers RN, Chelsey... | Phone Msg | Phone Msg | CRP-Internal Medicine | NaN |
| 0 | Addendum by Beckmann MD, Joseph | Phone Msg | Phone Msg | UP-SMILEY LANE CLINICS | NaN |

* Example Table from Cerner

## Notes Bag of Words

- The Note Bag of Words are made from Associated Metadata
- Currently used columns are **EVENT_TITLE_TEXT, EVENT_TAG, EVENT_CD_DESC, LOC_FACILITY_CD_DESC, etc…**
- Similarly, each word is individually added into the bag of words.

# B.6 Notes Metadata Consolidation




**Mizzou**

- Mizzou notes number in at around **132.6 Million**.
- Distinct rows and removal of irrelevant words takes us to **289,000 Rows (99.8% reduction)**

**MCW**

- The notes number in around **193.8 Million**.
- Distinct rows and removal of irrelevant words takes us to **864,625 Rows (99.5% reduction)**

# B.7 Data Table Preparation

Original Table

| A | B | C | D | E | F |
|---|---|---|---|---|---|
| ** | ** | ** | ** | ** | ** |
| ** | | ** | ** | ** | ** |
| ** | *** | | | | ** |
| ** | ** | ** | ** | ** | ** |
| ** | ** | ** | ** | | ** |
| ** | *** | ** | ** | * | ** |
| ** | *** | ** | | ** | ** |
| ** | *** | ** | | * | ** |
| ** | ** | ** | ** | ** | ** |

Intermediary table of Distinct Values

| B | D | ID | Code |
|---|---|---|---|
| ** | ** | 1 | |
| | ** | 2 | |
| *** | | 3 | |
| *** | ** | 4 | |

Table with ID Column

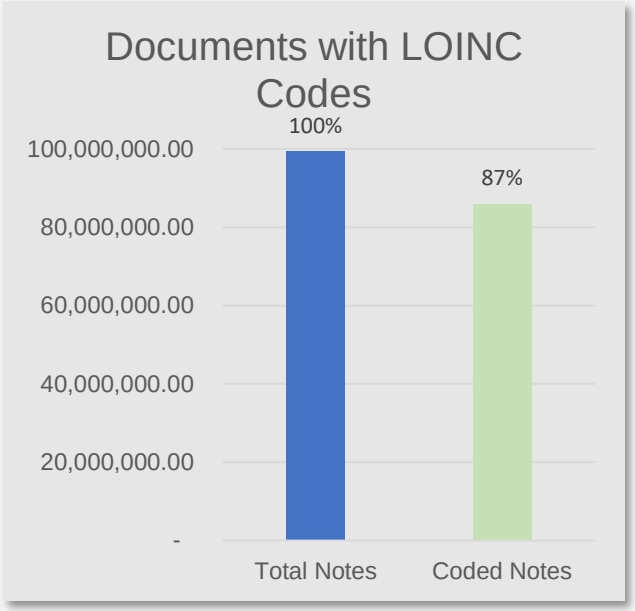
| A | B | C | D | E | F | ID |
|---|---|---|---|---|---|---|
| ** | ** | ** | ** | ** | ** | 1 |
| ** | | ** | ** | ** | ** | 2 |
| ** | *** | | | | ** | 3 |
| ** | ** | ** | ** | ** | ** | 1 |
| ** | ** | ** | ** | | ** | 1 |
| ** | *** | ** | ** | * | ** | 4 |
| ** | *** | ** | | ** | ** | 3 |
| ** | *** | ** | | * | ** | 3 |
| ** | ** | ** | ** | ** | ** | 1 |

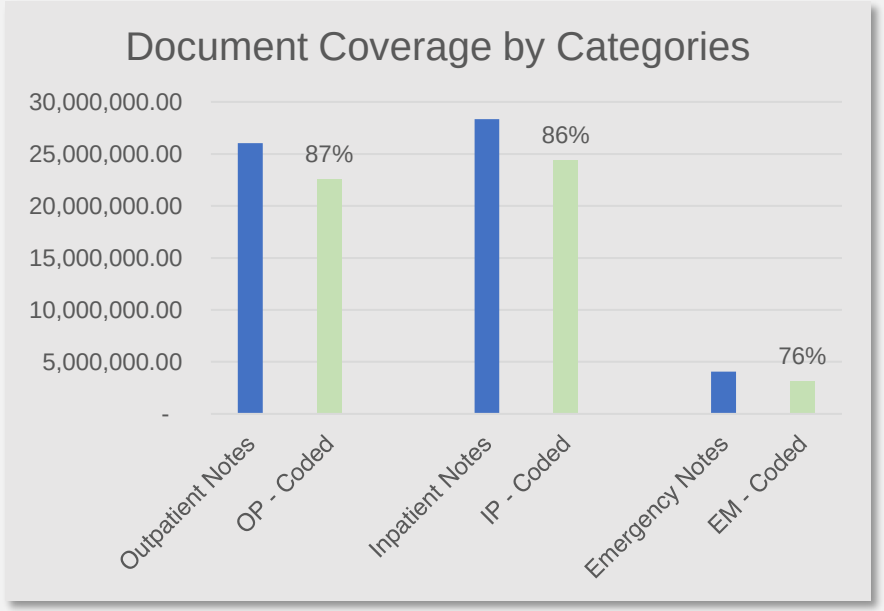
## Summary

- The ID field will act as the primary key, allowing for efficient joining of tables with correlated LOINC Codes.
- We will operate on the significantly smaller **Intermediary Table** and join back to the individual notes at the end.

# C. Results & Analysis + Discussion

# C.1 Coverage Results (Cerner)

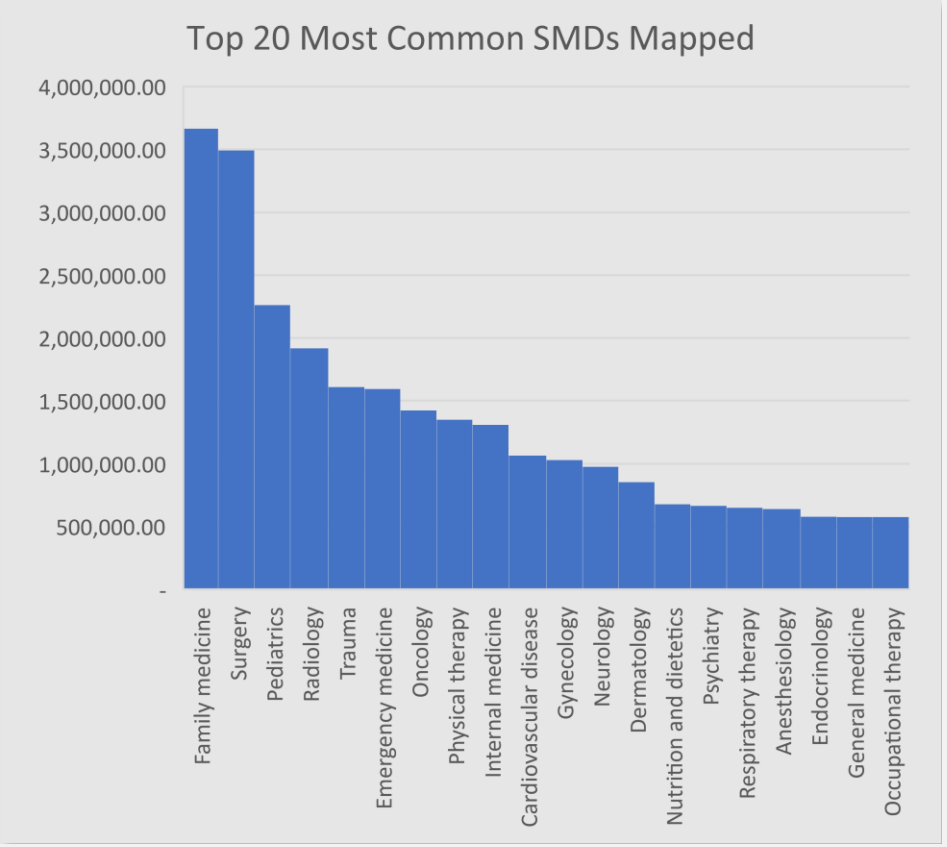


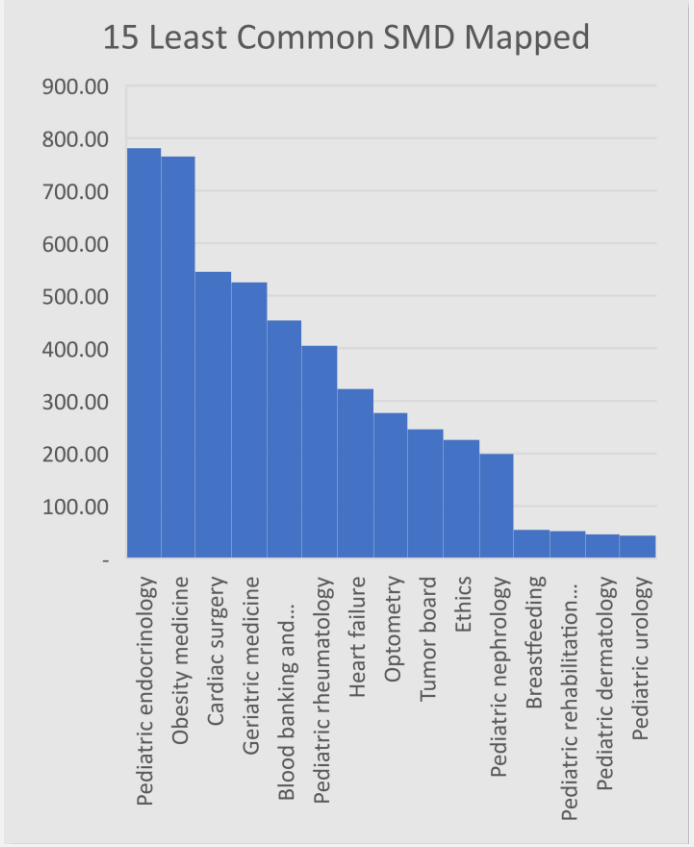




## Summary

- Around **87%** of our documents were mapped to LOINC Codes.
- Broken down into broad categories, the numbers range from **76-87%** coverage.
- We started with **35%** coverage and slowly improved our coverage using the Synonymy List.

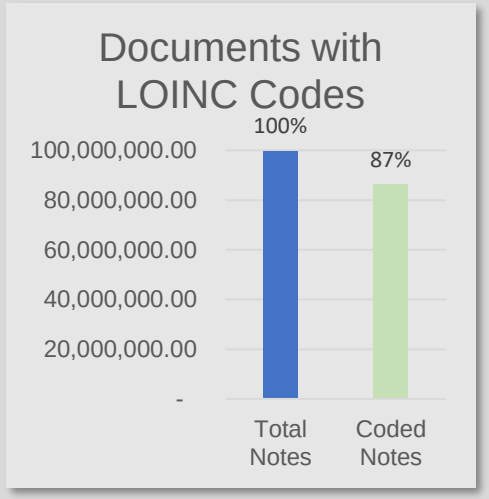
# C.2 SMD Mapping Distribution (Cerner)



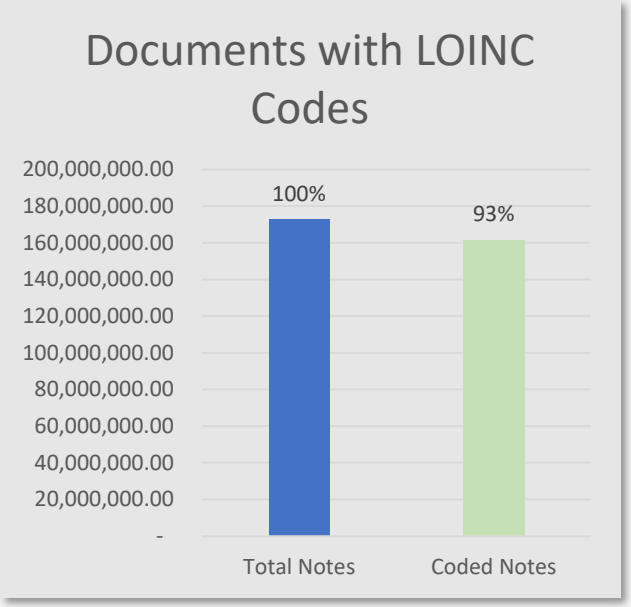



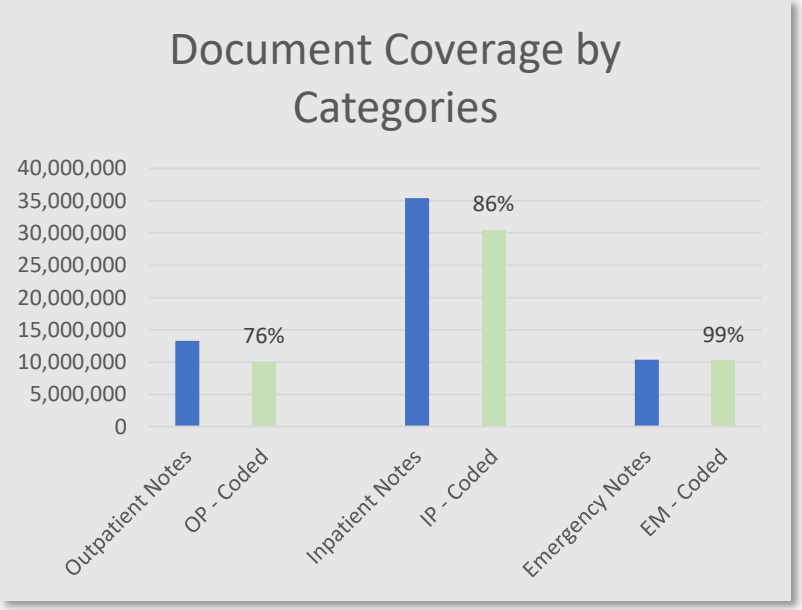

## Summary

- 82/155 different SMDs Mapped
- 18/74 different KOD Mapped
- 8/25 different Settings Mapped
- 35/147 different TOS Mapped
- 17/26 unique Roles Mapped



# C.3 Coverage Results (EPIC)

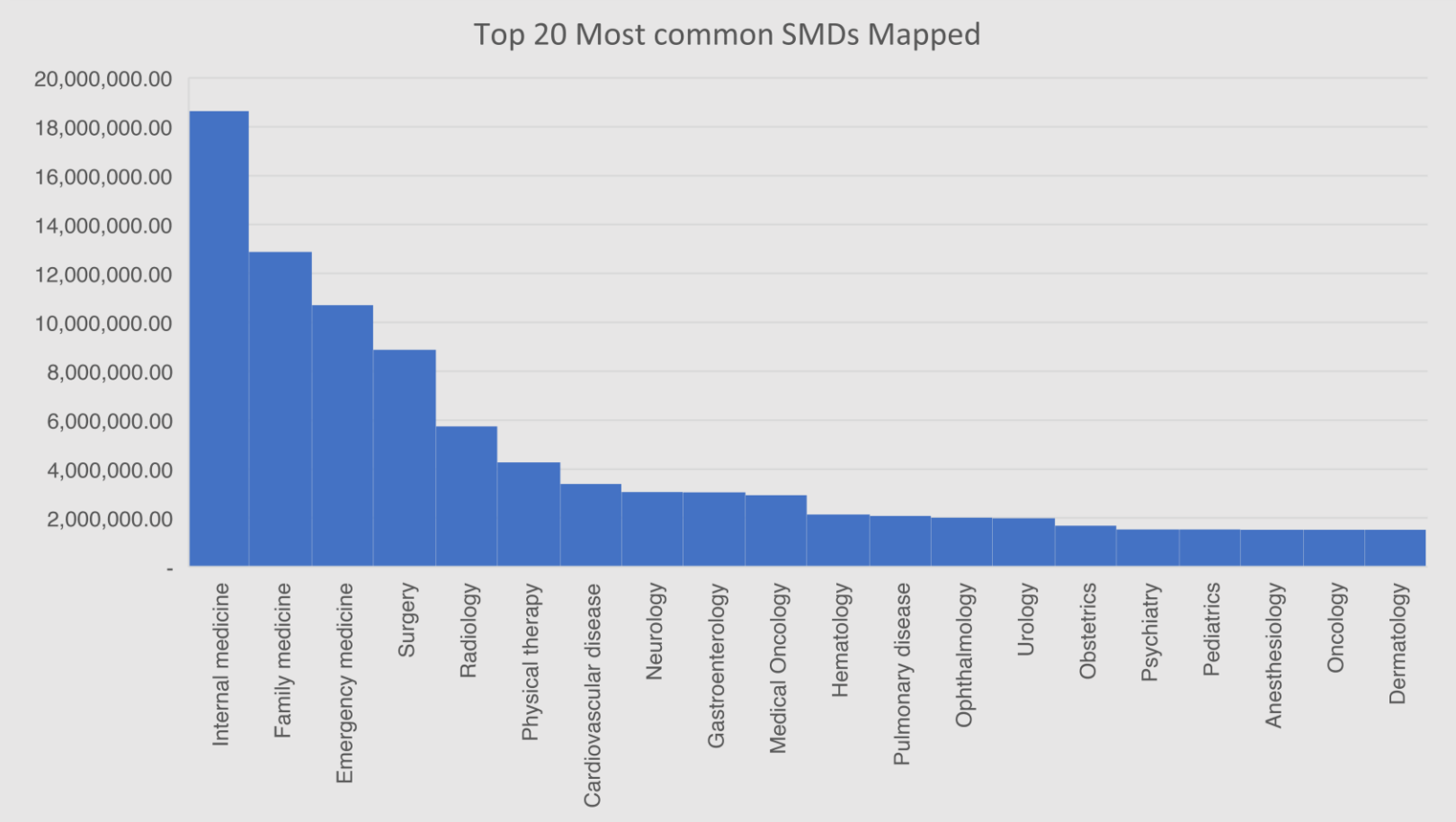



**Summary**

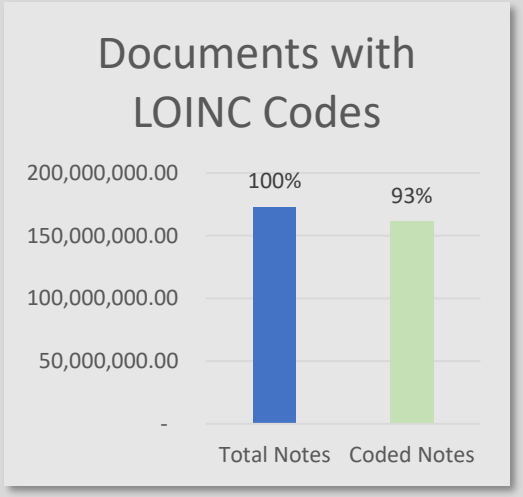
- The pipeline was tested with **193M Documents**.
- Around **93%** were mapped to LOINC Codes
- Broken down into categories, the numbers range from **76-99%** coverage.
- We started with **55%** coverage and jumped to 80% using the Synonymy List.

# C.2 SMD Mapping Distribution (EPIC)




## Summary

- 99/155 different SMDs Mapped
- 12/74 different KOD Mapped
- 8/25 different Settings Mapped
- 24/147 different TOS Mapped
- 22/26 unique Roles Mapped



# C.4 Final Thoughts

**Framework Summary**

- Lightweight and easily deployable
- Efficient and designed with scalability in mind.
- Synonym list allows ease of improvement of coverage.

**Future Work**

- Program needs official validation
- While it may make sense to share synonymy lists between sites, some synonyms may incorrectly improve coverage as false positives.
- Testing the framework at different sites may illuminate further flaws to fix

# D. END