

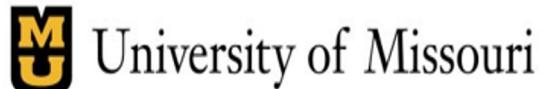
Scalable Mapping of Clinical Notes to LOINC Document Ontology via Bag of Words

A Greater Plains Collaborative (GPC) Effort

University of Missouri, Columbia (MU)
Huzaifa Khan
Vyshnavi Pakka
Abu Saleh Mohammad Mosa
Lemuel R. Waitman

Medical College of Wisconsin, Milwaukee (MCW)
Dave Bell
Kristen Osinski
Alex Stoddard
Brad Taylor

October 19th , 2023



Clinical & Translational Science Institute
of Southeast Wisconsin



Overview

LOINC Document Ontology

LOINC Code Mapping

Results & Analysis + Validation

LOINC Document Ontology (DO)

Logical Observation Identifiers, Names, and Codes

- Our goal: assign a LOINC Code to all clinical documents
 - Ease of organization
 - Gives a broad view of the content
 - Makes documents Query-able

Documents are assigned LOINC Codes 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



Document Ontology Dimensions

- 1. Subject Matter Domain (SMD)**
 - Dentistry, Allergy, etc...
- 2. Kind of Document (KOD)**
 - Note, Report, Prescription list, etc...
- 3. Type of Service (TOS)**
 - Consultation, Procedure, Counseling, etc...
- 4. Setting**
 - Outpatient, Telehealth, Pharmacy, etc...
- 5. Role**
 - Patient, Nurse, Physician, etc...

Examples

- 68478-7: Pulmonary Attending Hospital Progress Note
- 97711-6: Heart failure Outpatient Note
- 34112-3: Hospital Note

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

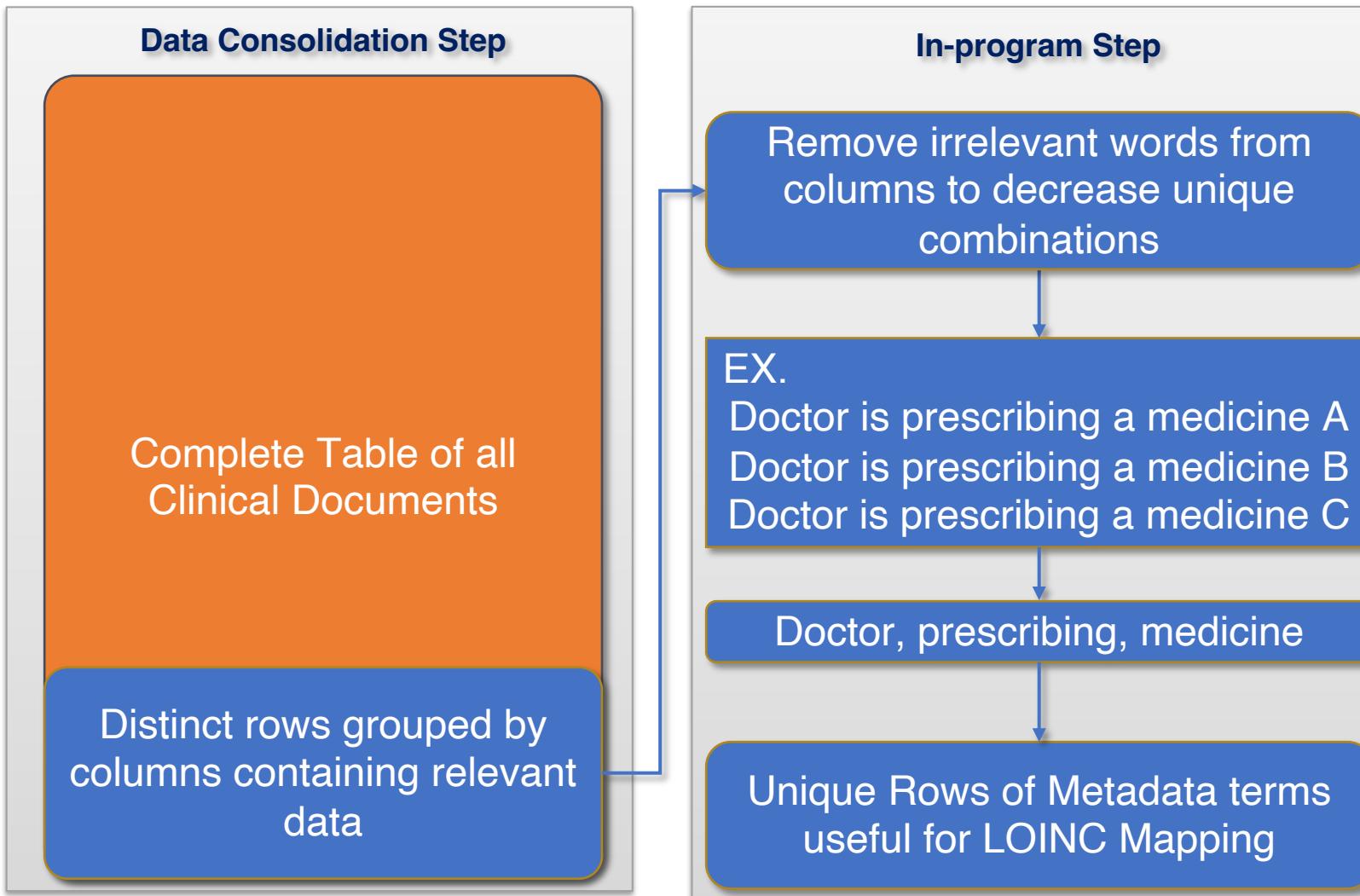
Manual Value-Set Mapping

- **Subject Matter Domain – Note_Title**
 - Mapped **30%** of all notes to a SMD
- **Kind of Document – Note_Type**
 - Mapped **50%** of all notes to a KOD
- **Type of Service – Event_Description**
 - Mapped **40%** of all notes to a TOS
- **Setting – Encounter_Type**
 - Mapped **80%** of all notes to a Setting
- **Role – Encounter_Note_Author**
 - Mapped **35%** of all notes to a Role
- **Result: only around 25% of notes were mapped with Valid LOINC Code**
- **Programmatic Mapping**
 - Will give us more coverage
 - Will be repeatable and extensible
 - **Bag of Words**

Number of Parts	Number of Documents	% values
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	

Bag of Words Mapping Implementation

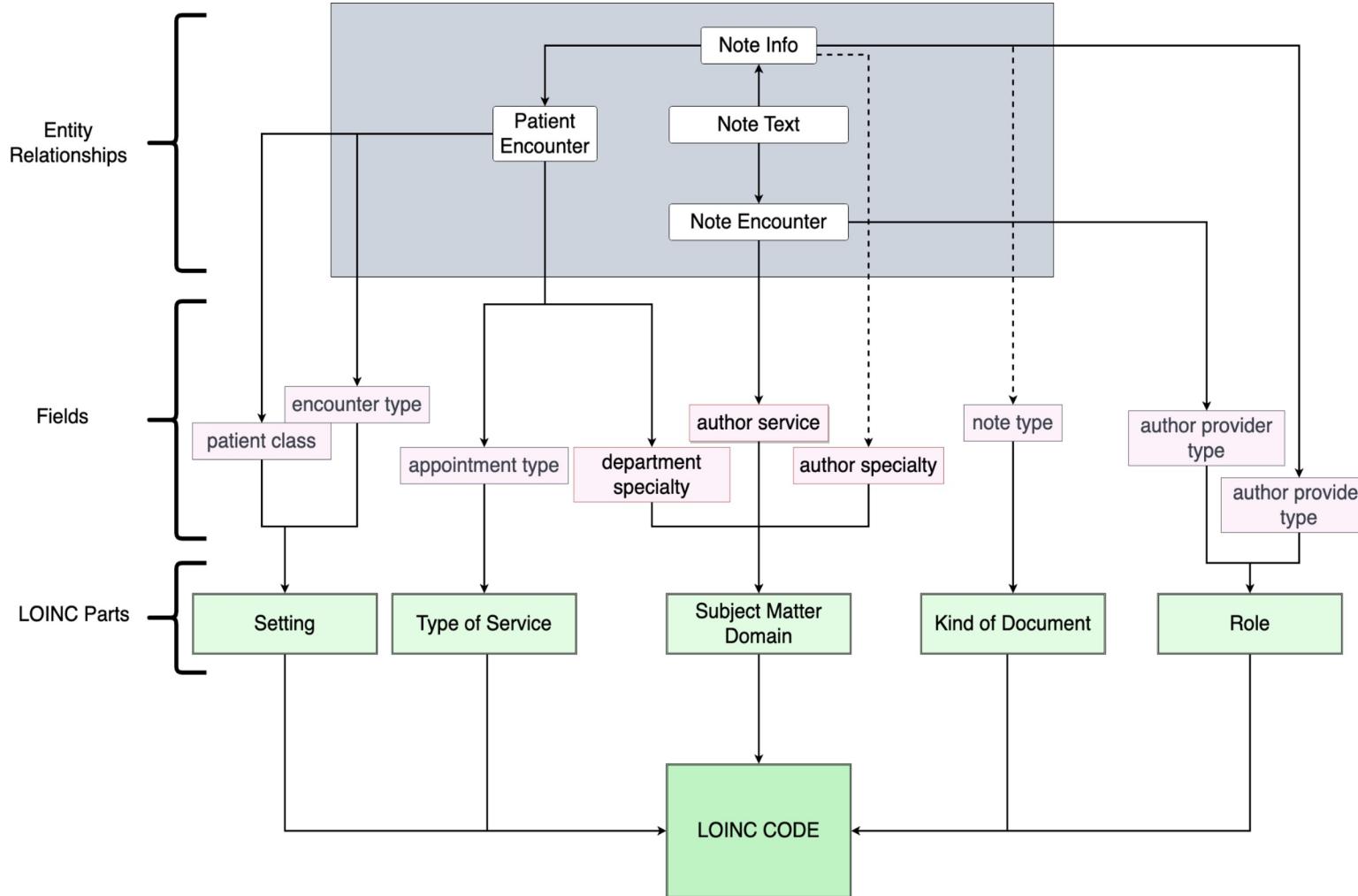
Pipeline Overview



Summary

- After Data staging and gather clinical notes, Metadata needs to be consolidated
- Grouping by Metadata should significantly reduce size of unique rows
- Program takes metadata and further reduces the unique rows by removing unhelpful keywords
- Results in a fraction of original problem space

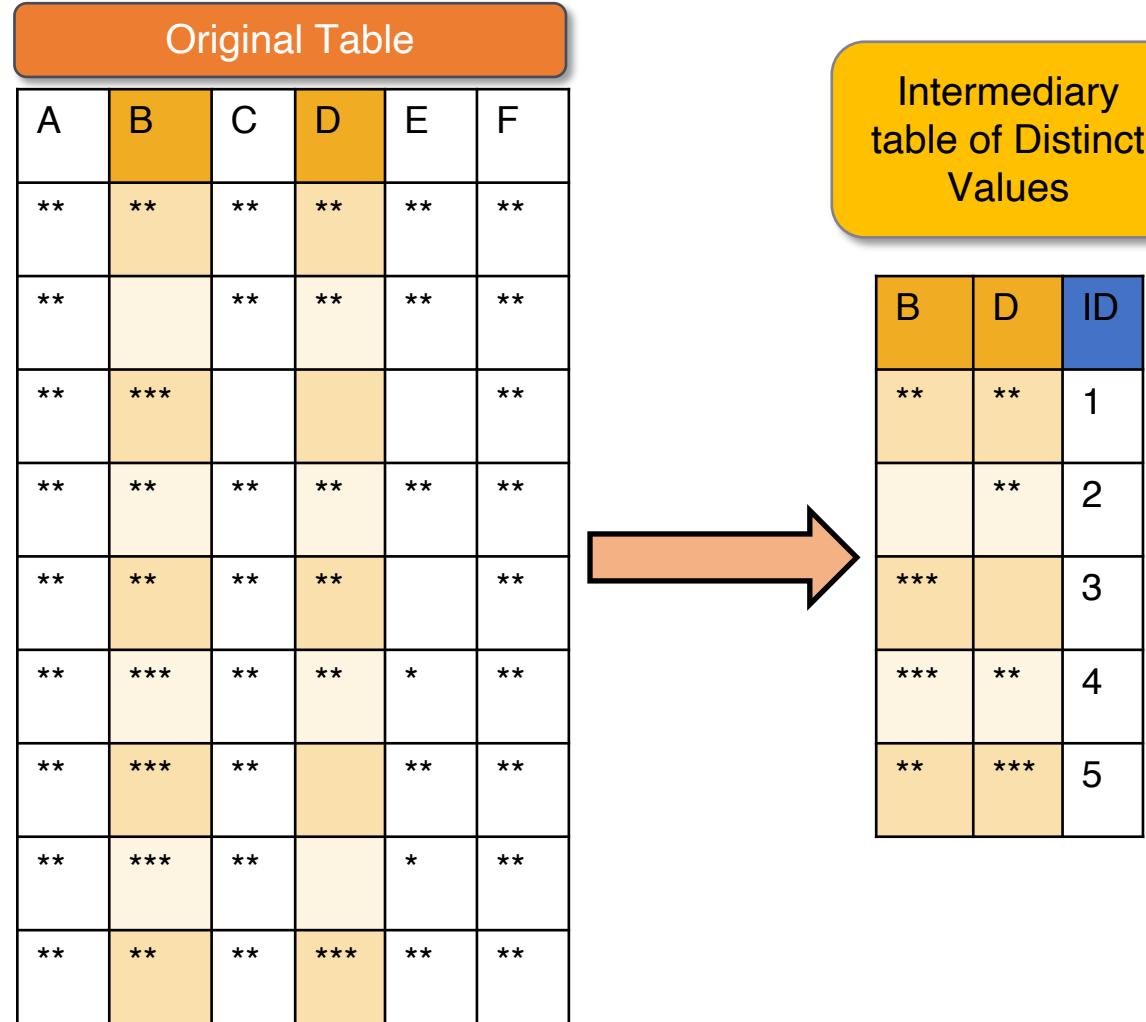
LOINC Mapping Diagram



Summary

- This diagram visualizes the columns used for the LOINC Mapping
- Multiple sources are used for domains
- **Actual column names may differ depending on your site's EHR vendor**

Data Table Preparation



MU

- **132.6 Million notes.**
- **289,000 Rows** distinct metadata sets

MCW

- **193.8 Million notes**
- **267,151 Rows** distinct metadata sets

The program will operate on the significantly smaller **Intermediary Table** and join back to the individual notes at the end

Bag of Words Generation

	LoincNum	PartNumb	PartTypeName	PartName
38	100446-4	LP173051-6	Document.Settings.	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.Settings.	Outpatient
43	100447-2	LP268363-1	Document.SubjectMat...	Burn management
44	100448-0	LP173418-7	Document.Kind.	Note

LOINC Code Bag of Words

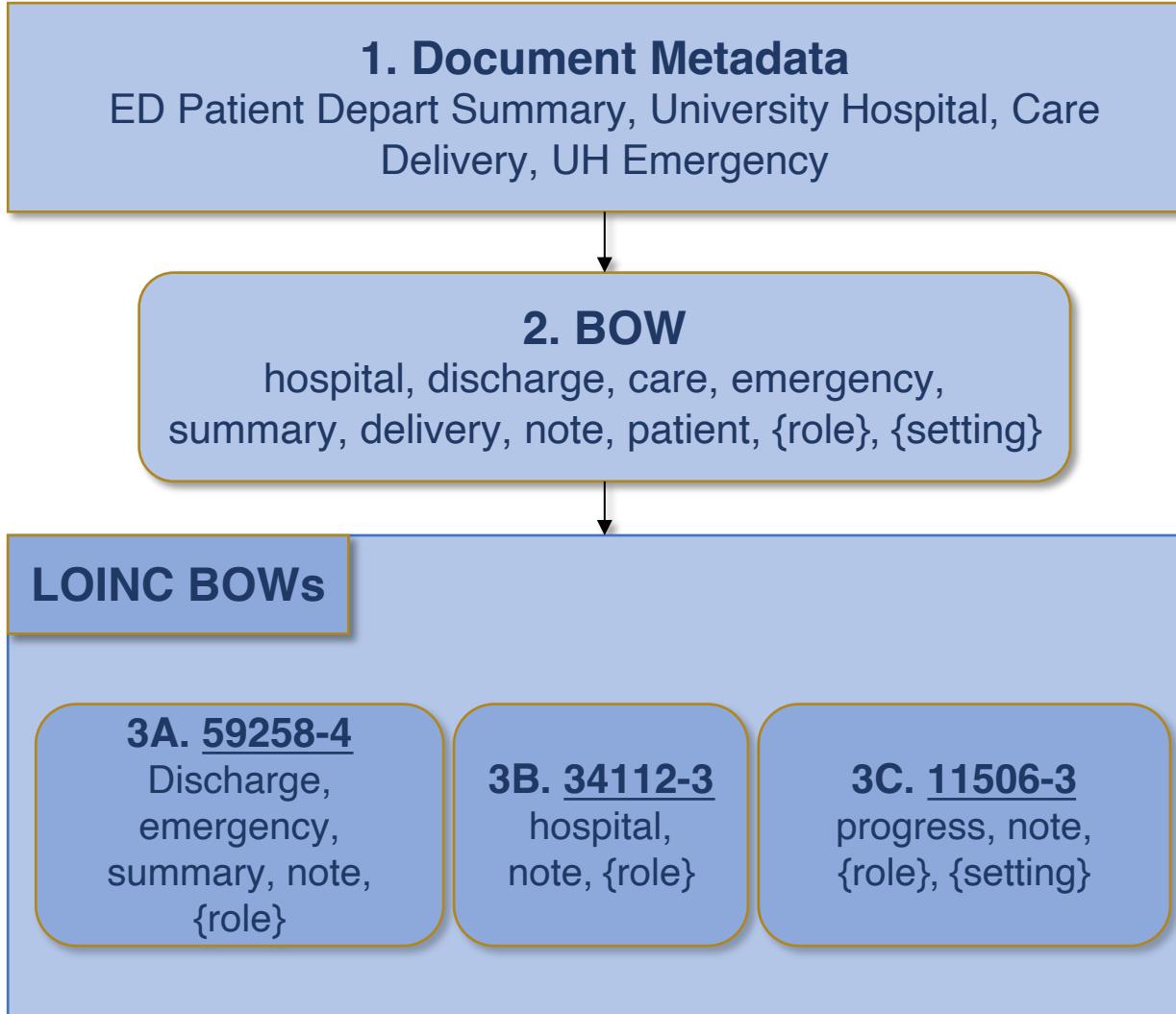
- Example code **100447-2** has 4 Part Numbers, each represented by a row
- {"note", "progress", "outpatient", "burn", "management"}

Metadata Bag of Words

- {"registered", "nurse", "internal", "medicine", "refill", "telephone", "encounter"}

Role	SMD	TOS	Setting	KOD
Registered Nurse	Internal Medicine	Registered Nurse	Refill	Telephone Encounter

Bag of Words (BOW) Example



Mapping Summary

- Word order is lost in BOW
- “{role}” and “{setting}” wildcards are added.
 - Priority is given to real axis, if available (e.g. Emergency)

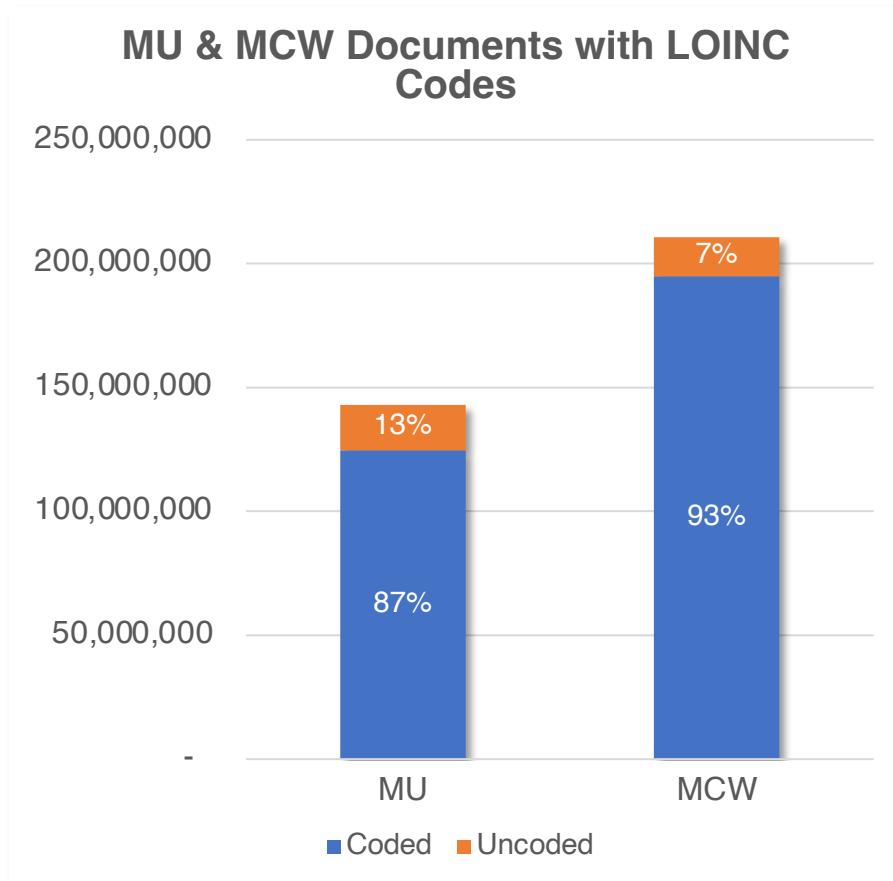
Synonymy List

- Depart -> Discharge
- Summary -> Note

Chosen LOINC Code is **59258-4**

Results

Coverage Results



MU

- ~ 87% of all clinical documents were mapped to LOINC Codes
- 35% initial coverage and slowly improved using the Synonymy List

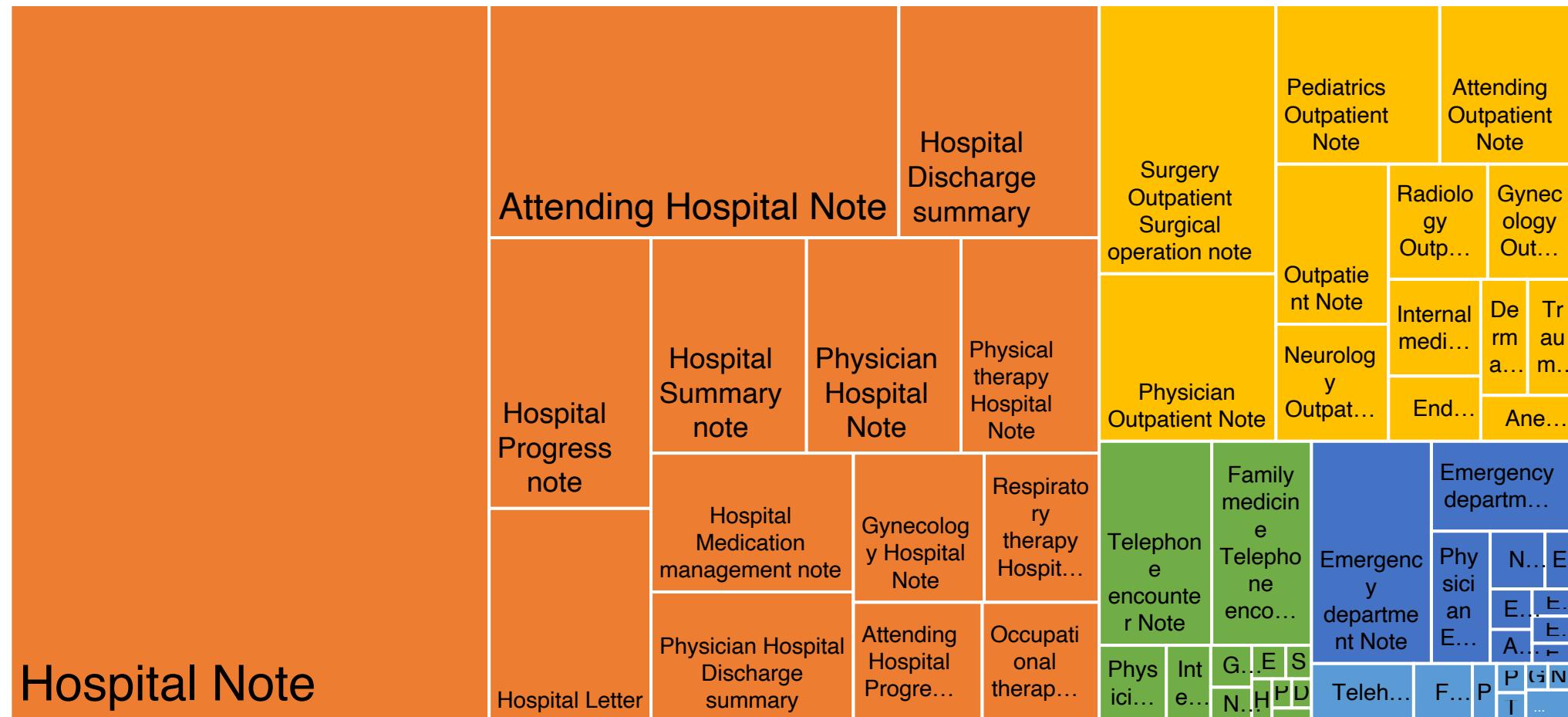
MCW

- ~ 93% of all clinical documents were mapped to LOINC Codes
- 55% initial coverage and slowly improved using the Synonymy List

Notes Distribution Based on Setting

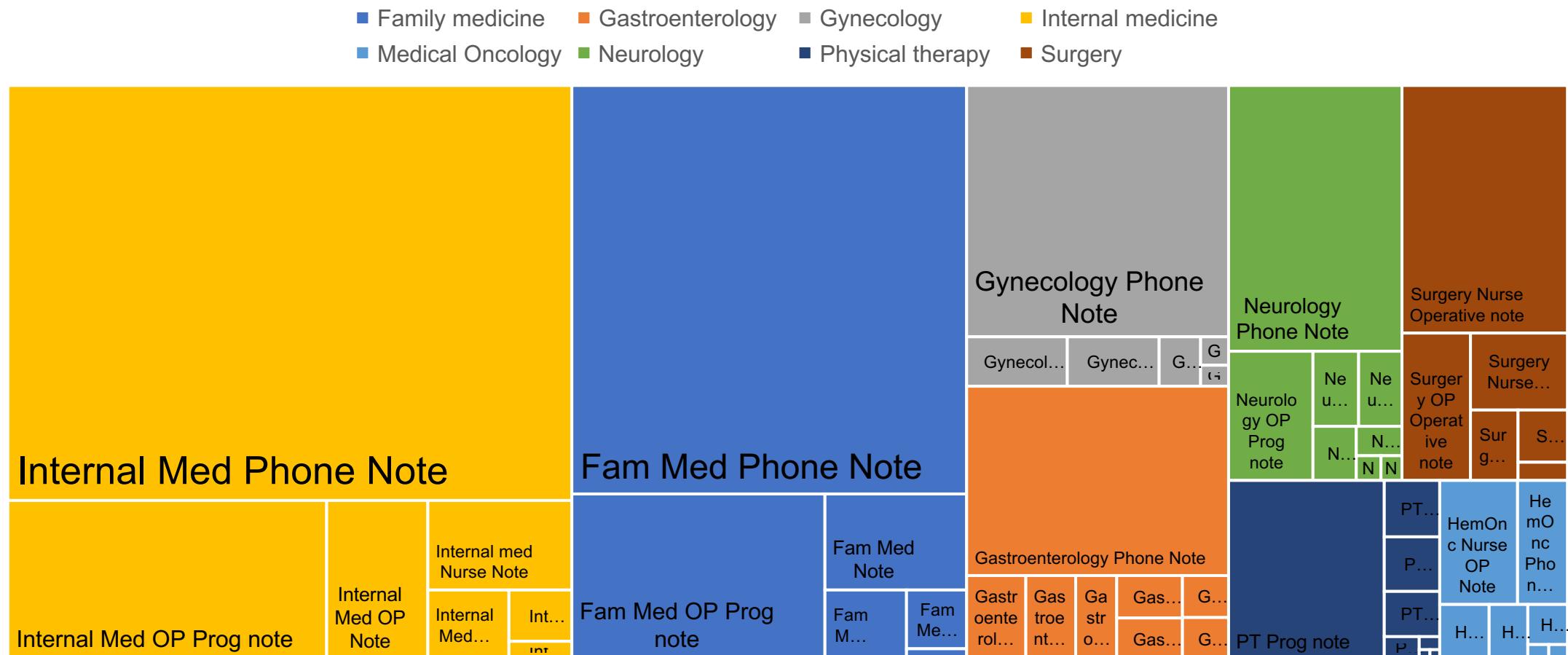
MU Data: Categorized by Setting

- Emergency ■ Hospital ■ Observation unit ■ Outpatient ■ Telehealth ■ Telephone encounter



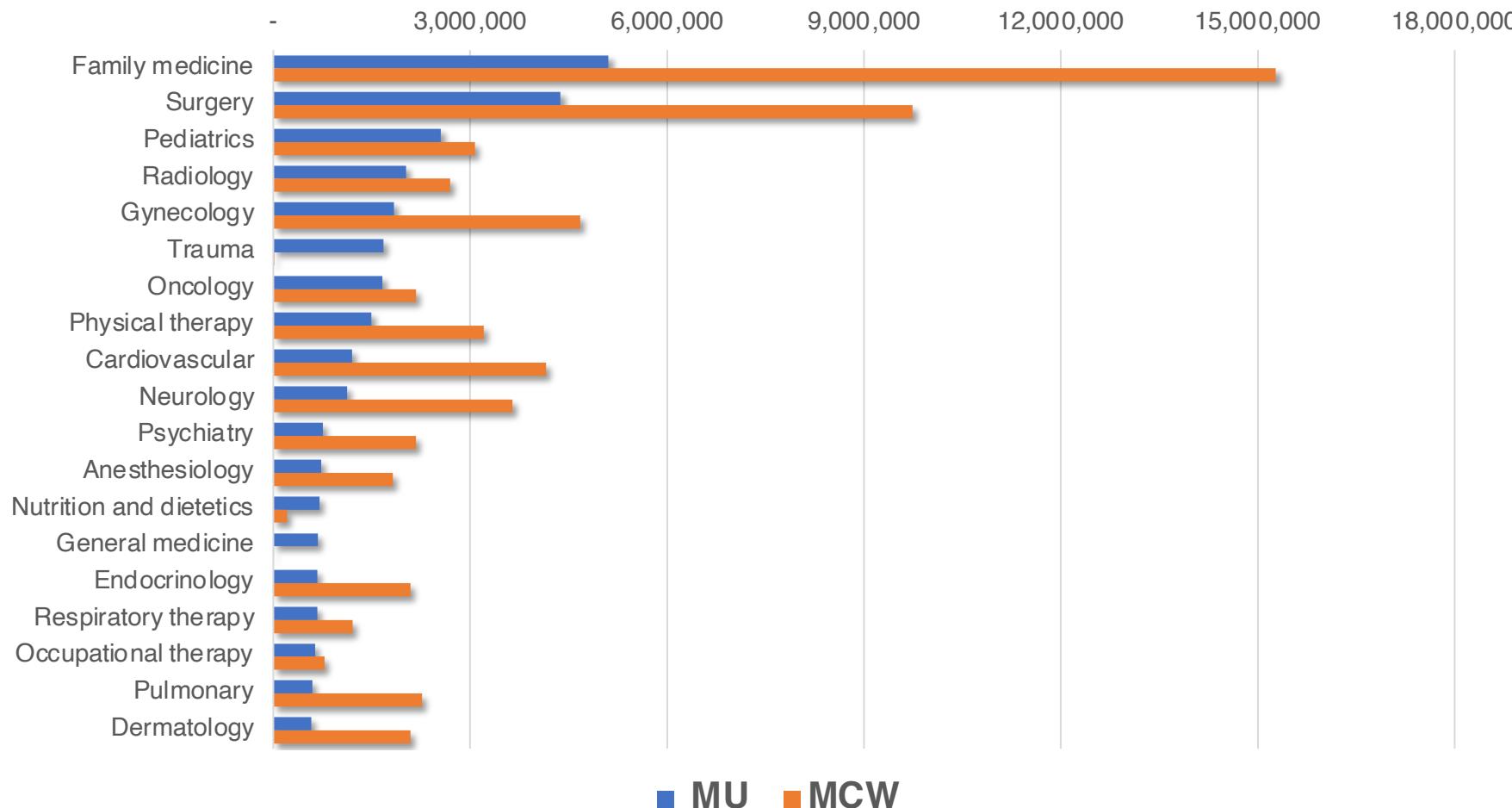
Notes Distribution Based on SMD

MCW Data: Categorized by SMD



SMD Distribution between sites

Most Common SMDs' Document Distribution in MU vs MCW



Validation

Validation

Note Selection

- 200 rows randomly selected from 1000 most common combinations of metadata

MU:

- 200 rows or **~15 million** notes, **~10%** of total
- Basic accuracy in terms of note count: **~91%**

MCW:

- 200 rows or **~25 million** notes, **~8%** of total
- Basic accuracy in terms of note count: **~85%**

<u>Row Count</u>	Valid	Invalid
MU	167 (83.5%)	33
MCW	157 (78.5%)	43

<u>Note Count</u>	Valid	Invalid
MU	13.93 M (91%)	1.27M
MCW	21.78 M (85%)	3.91M

Validation

Iterative Mapping Improvements (From MCW Notes)

- Initial validation highlighted issues:
 - Undesirable document types
 - “scanned documents”
 - Missing synonym list entries
 - “OB/GYN” -> “Obstetrics & Gynecology”
 - Some phrases cause consistent, incorrect labelling
 - “surgery” -> “operative note”

<u>Row Count</u>	<u>Valid</u>	<u>Invalid</u>
Round 1	116 (58%)	84
Round 2	138 (69%)	62
Round 3	157 (78.5%)	43

role	SMD	TOS	setting	KOD	bow	loinc name
Registered Nurse	Transplant Surgery		Telephone	Telephone Encounter	encounter registered {role} {setting} note surgery nurse transplant telephone	Surgery OP Operative note

Closing Remarks

Future Work

- Synonymy list sharing issues
- UMLS Integration
- Mapping to multiple LOINC codes
- Testing the framework at different sites may illuminate further flaws to fix

Framework Summary

- Made in Python, using the *Bag of Words* approach
- Synonymy list allows ease of improvement of mapping coverage
- Lightweight and easily deployable
- Efficient and designed with scalability in mind



Acknowledgements

Source code is available on GitHub

❖ GitHub Repositories

- Public: https://github.com/gpcnetwork/gpc_doc_ontology
- Private: https://github.com/gpcnetwork/gpc_document_ontology

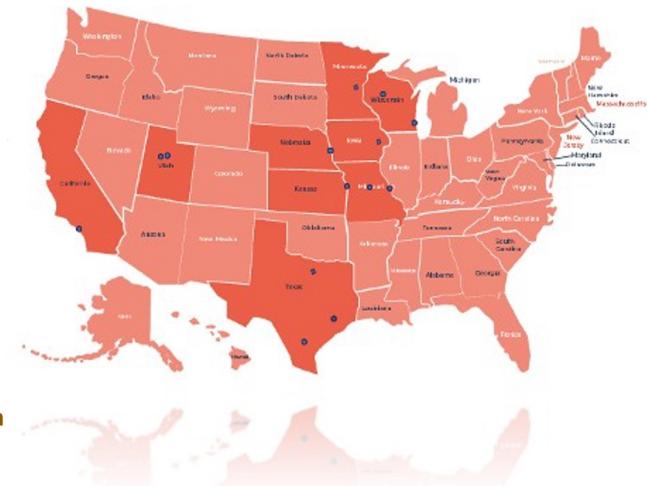
❖ GPC Website Info - <https://gpcnetwork.org/>

❖ Contact

- Dave Bell
 - dabell@mcw.edu
 - GitHub: [dabell4](#)
- Huzaifa Khan
 - hkg8b@umsystem.edu
 - GitHub: [ghkkhan](#)

GPC Member Institutions

- Allina Health (MN)
- Intermountain Health (UT)
- University of Kansas Medical Center (KS)
- Marshfield Clinic Research Institute (WI)
- Medical College of Wisconsin (WI)
- University of Missouri (MO)
- University of Iowa Healthcare (IA)
- University of Nebraska Medical Center (NE)
- University of Texas Southwestern Medical Center (TX)
- University of Utah (UT)
- University of Texas Health Sciences Center at San Antonio (TX)
- University of Texas Health Houston (TX)
- University of California Los Angeles (CA)
- Washington University (MO)



Questions?