Presented by John Snow LABS

NLP SUMMIT

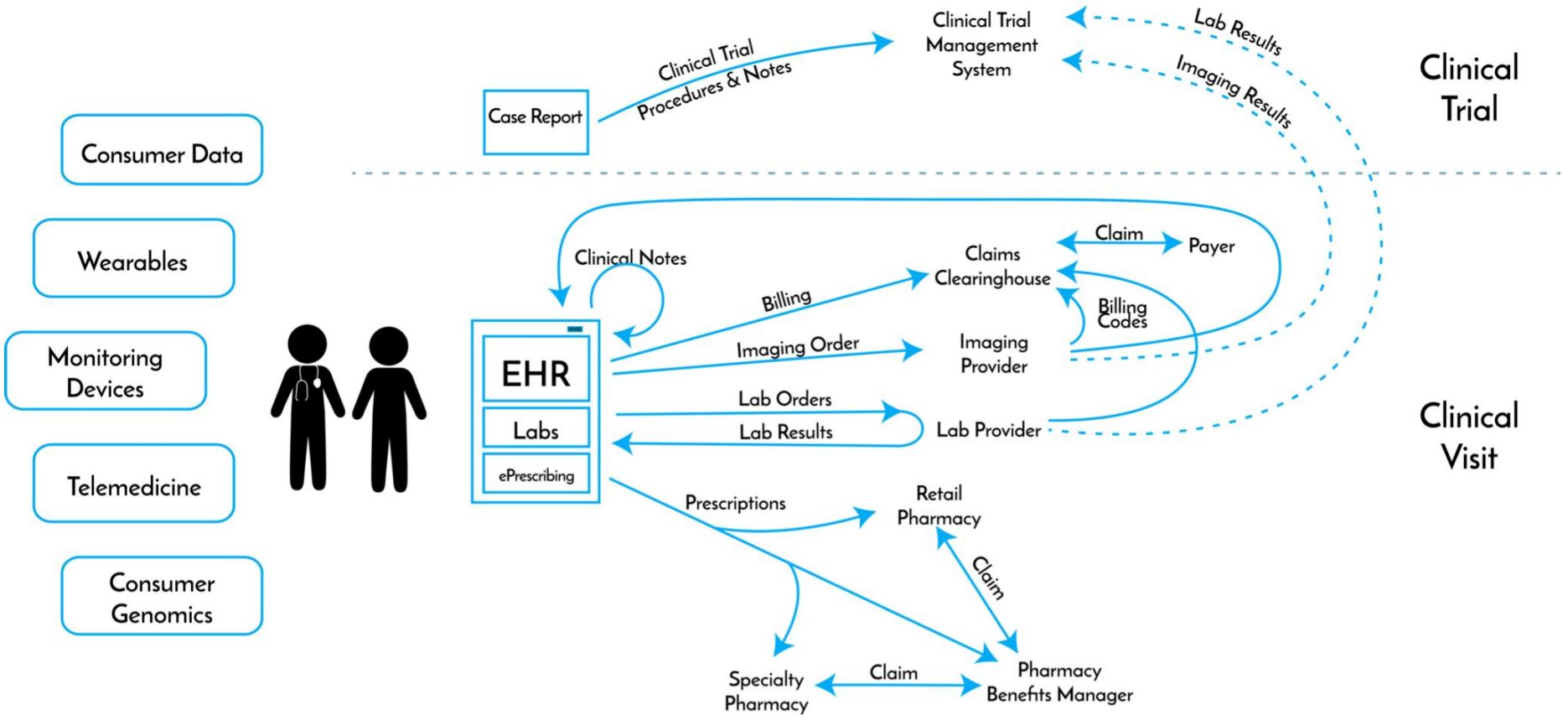
Building Patient Cohorts

Answering Patient Level Questions from Raw Clinical Data

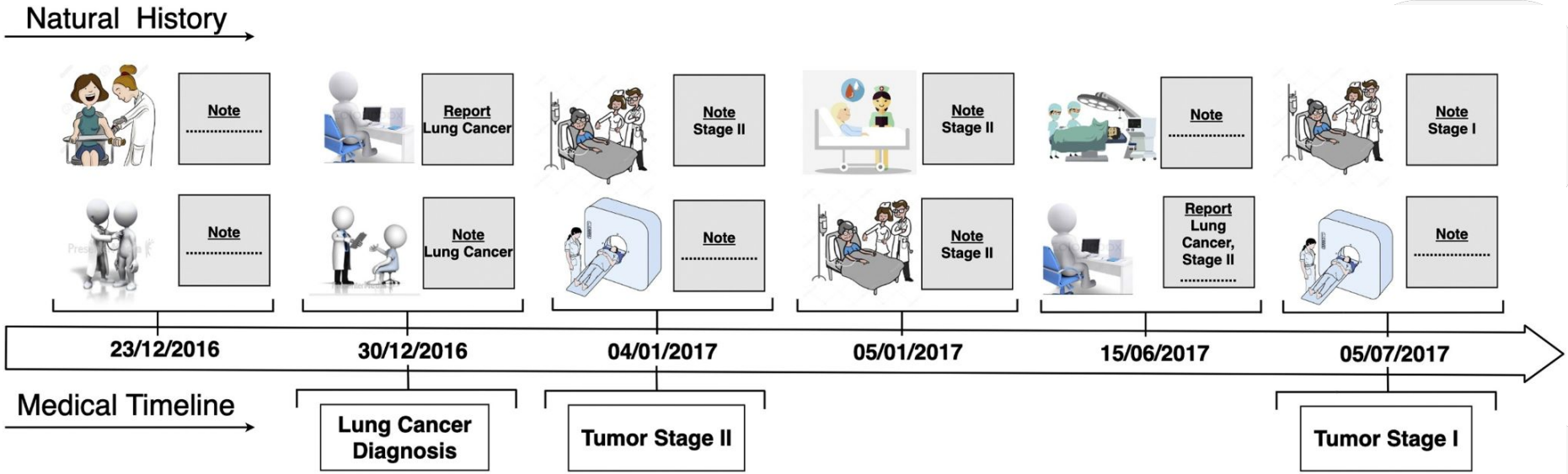
Certification Trainings,
John Snow Labs
July 18th, 2024

Veysel Kocaman, PhD
Head of Data Science
John Snow Labs

Data Origination and Exchange

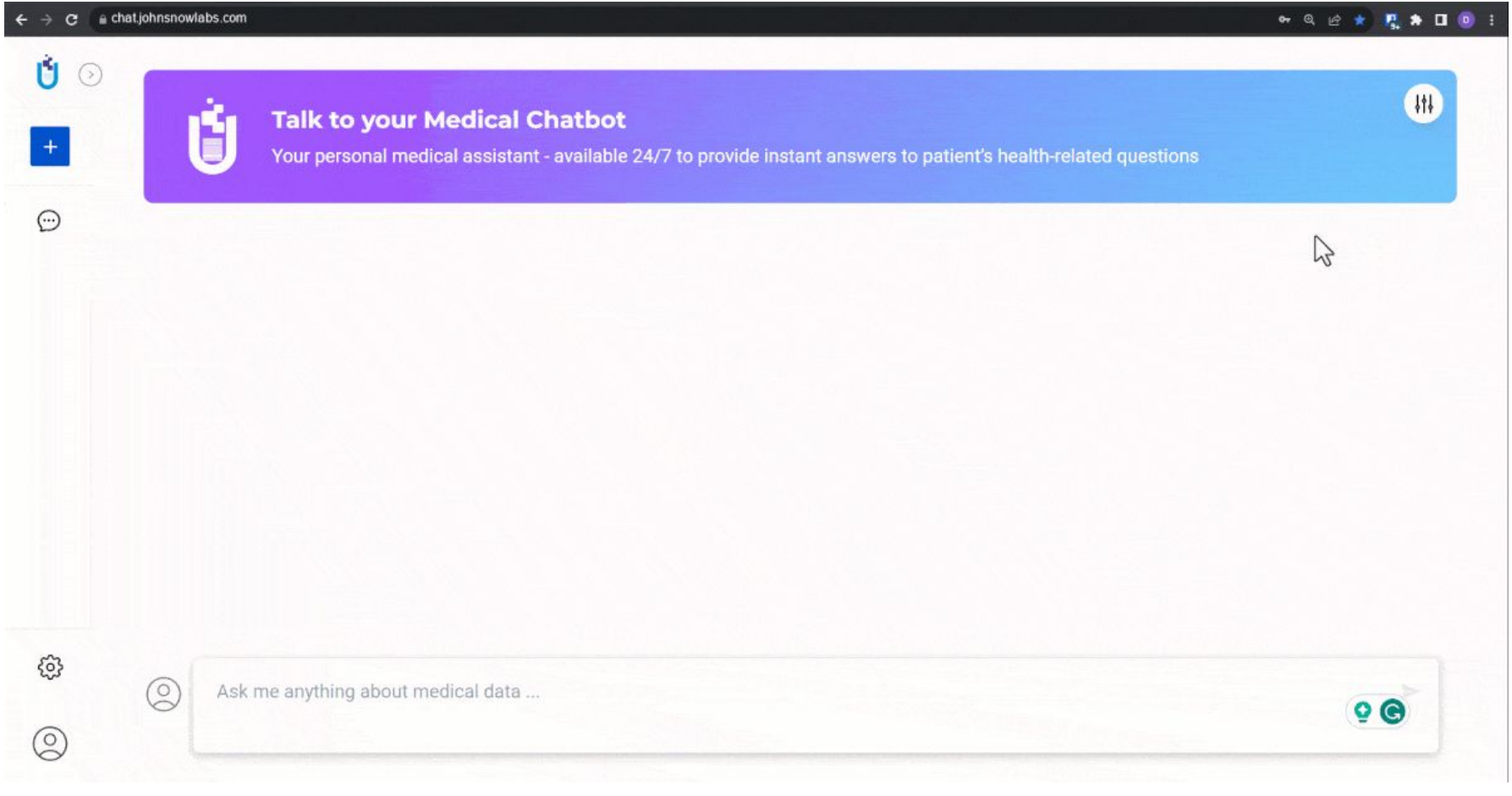


Putting the clinical facts on a timeline

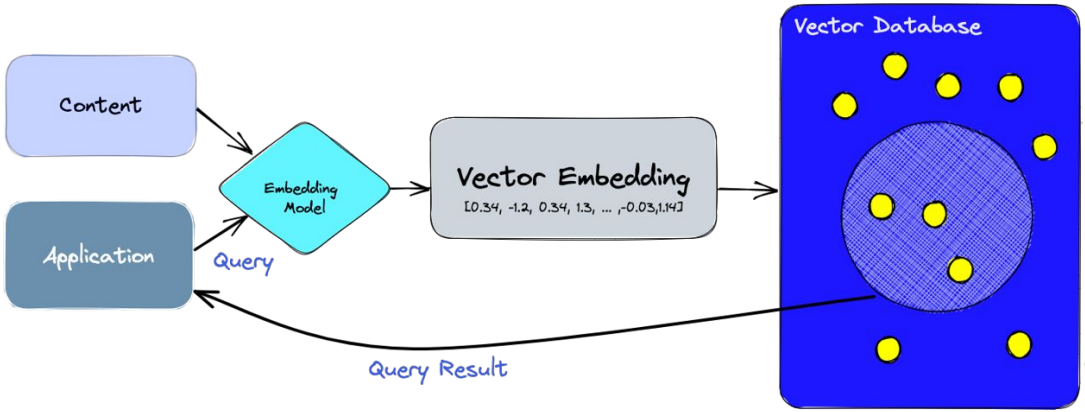


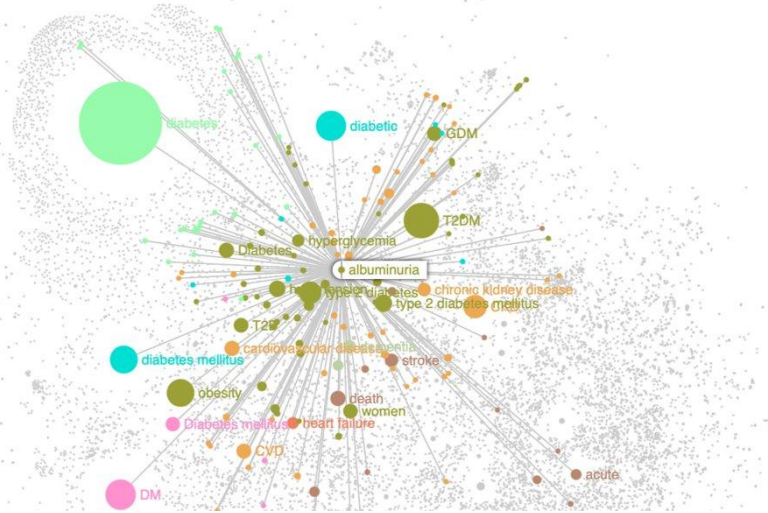
>> "Systems used to generate health data are designed for operations, not to organize data effectively for research or analytics."

Answering Patient Level Questions via Chatbot



Issues with RAG





Question: What are the concerns surrounding the AMOC?

Embedding Lookup

Continuous observation of the Atlantic meridional overturning circulation (AMOC) has improved the understanding of its variability (Frajka-Williams et al., 2019), but there is low confidence in the quantification of AMOC changes in the 20th century because of low agreement in quantitative reconstructed and simulated trends. Direct observational records since the mid-2000s remain too short to determine the relative contributions of internal variability, natural forcing and anthropogenic forcing to AMOC change (high confidence). Over the 21st century, AMOC will very likely decline for all SSP scenarios but will not involve an abrupt collapse before 2100. 3.2.2.4 Sea Ice Changes
Sea ice is a key driver of polar marine life, hosting unique ecosystems and affecting diverse marine organisms and food webs through its impact on light penetration and supplies of nutrients and organic matter (Arrigo, 2014).

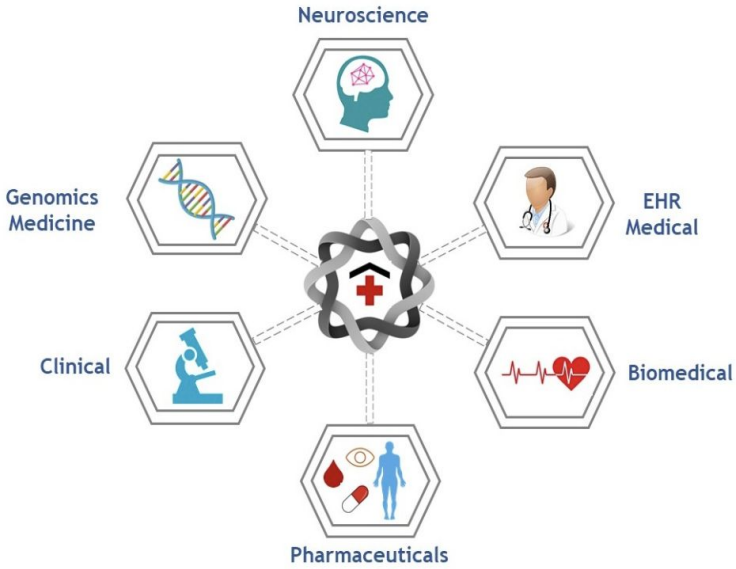
Retrieve Document Chunks for Synthesis

Document Chunks

"given the context splits, answer the question"

LLM

No LLM or RAG application can answer this question alone !

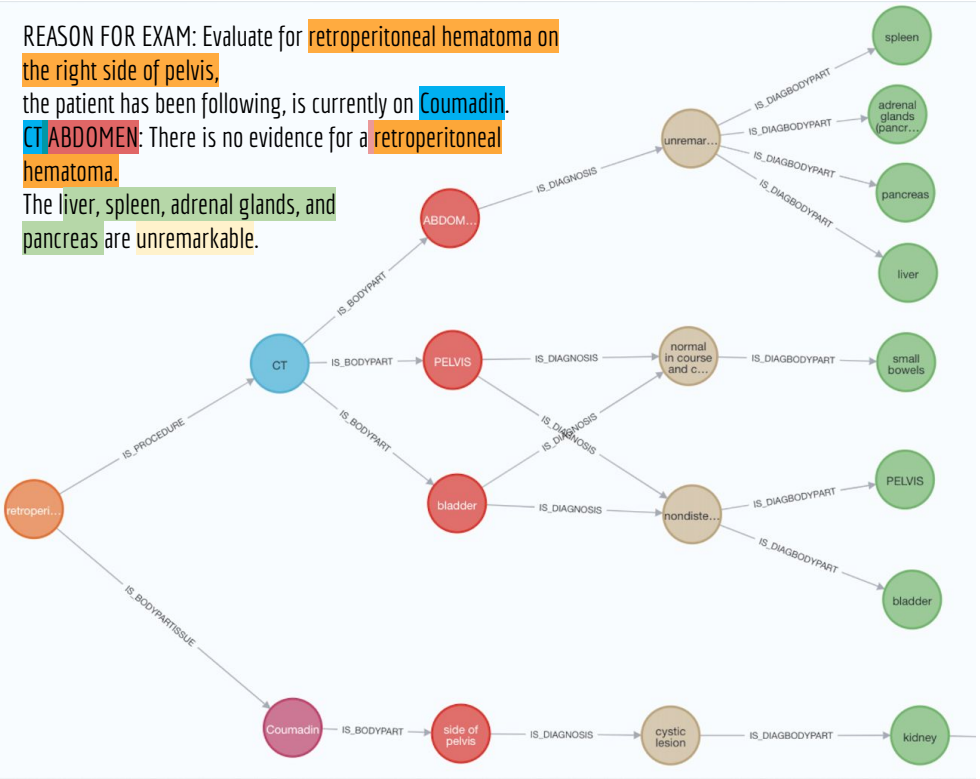


>> Give me all the patients who have type 2 diabetes, using metformin for the last 3 years, and also recently diagnosed stage-IV lung cancer?

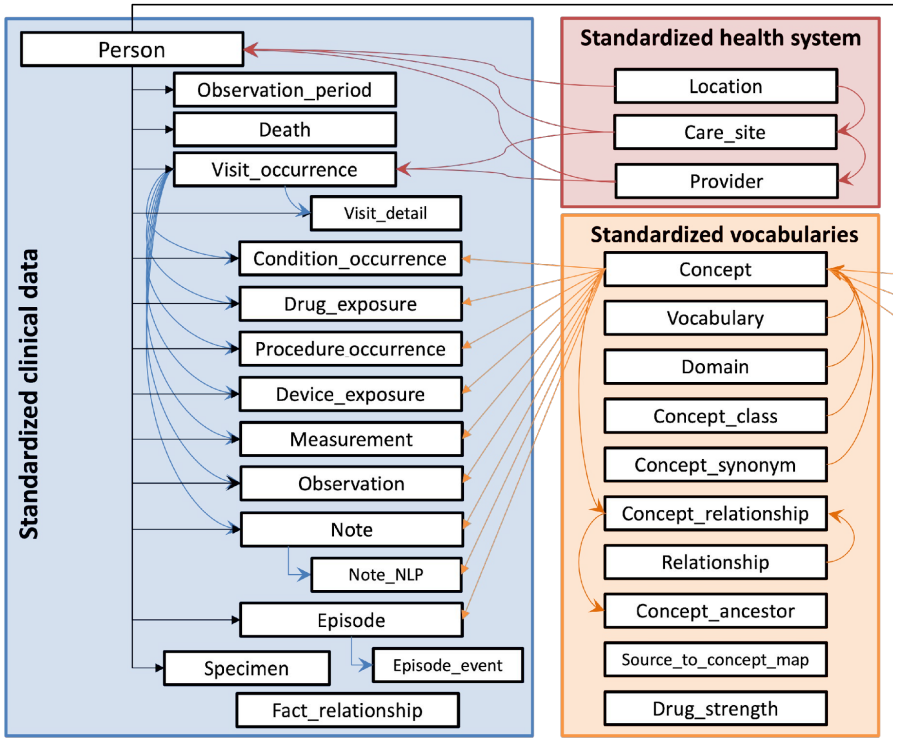
Unstructured EHR data

Answering Patient Level Questions from Raw Clinical Data

Knowledge Graph



Observational Medical Outcomes Partnership (OMOP)



Understanding OMOP CDM

(Observational Medical Outcomes Partnership - Common Data Model)

Enhancing Healthcare through Data, since 2009

Foundation: Part of the Observational Health Data Sciences and Informatics (OHDSI) initiative.

Objective: Utilize open-source data solutions to improve human health via large-scale analysis.

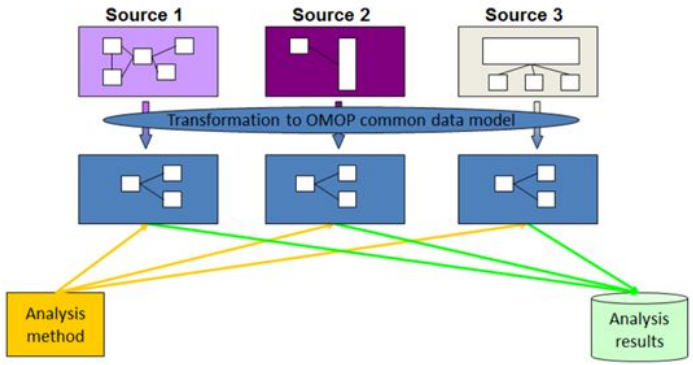
Purpose: Standardize the structure and content of observational healthcare data.

Methods:

- Through pseudonymisation and common data quality assessments, the OMOP-CDM provides a robust framework for converting complex EMR data into a standardised format.
- By securely sharing de-identified and aggregated data and conducting analyses across multiple OMOP-converted databases, patient-level data is securely firewalled within its respective local site.

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

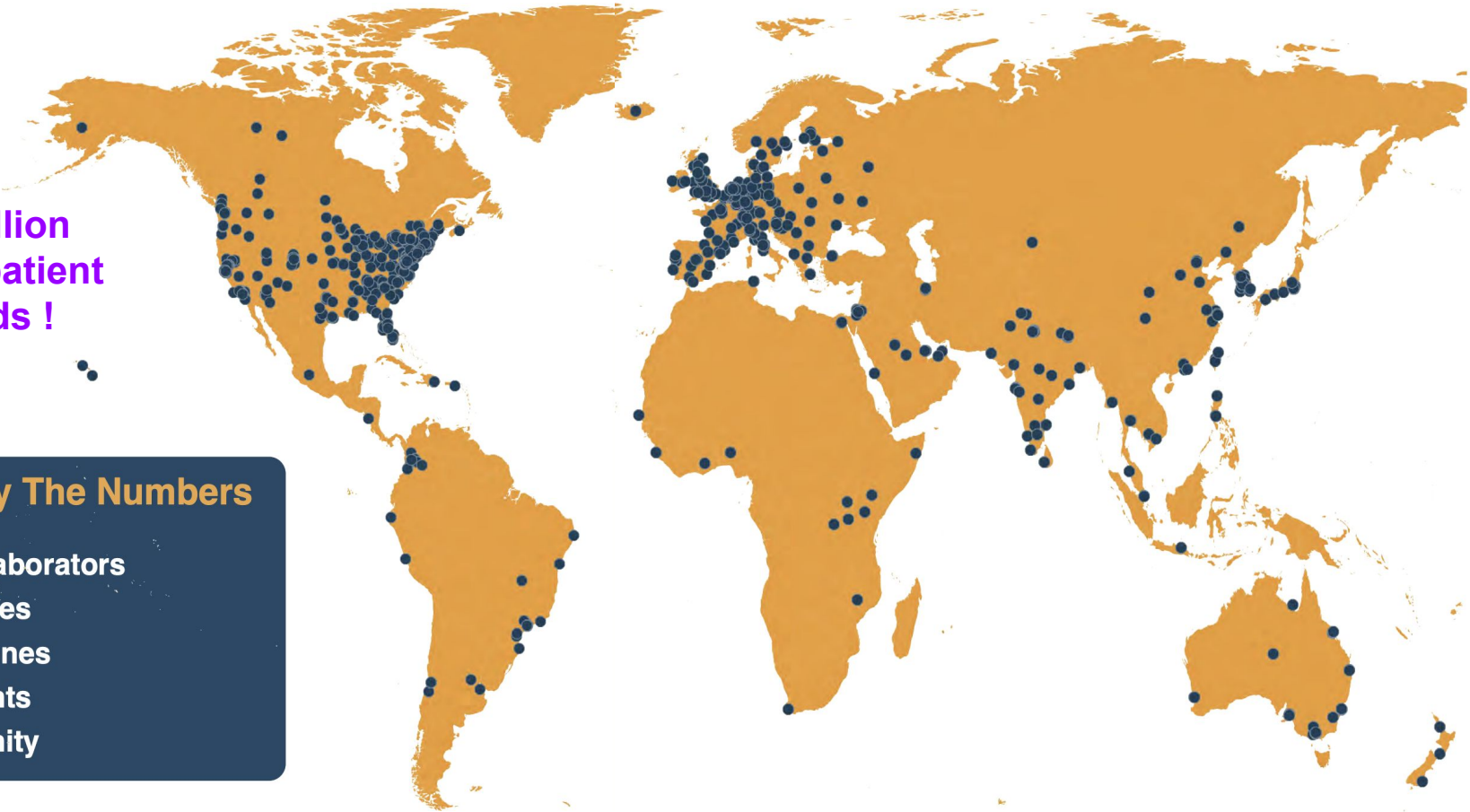
"an international collaborative whose goal is to create and apply open-source data analytic solutions to a large network of health databases to improve human health and wellbeing"



Seamless EMR data access: Integrated governance, digital health and the OMOP-CDM, Feb 2024.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10882353/

Adoption of OMOP Common Data Model

928 million unique patient records !



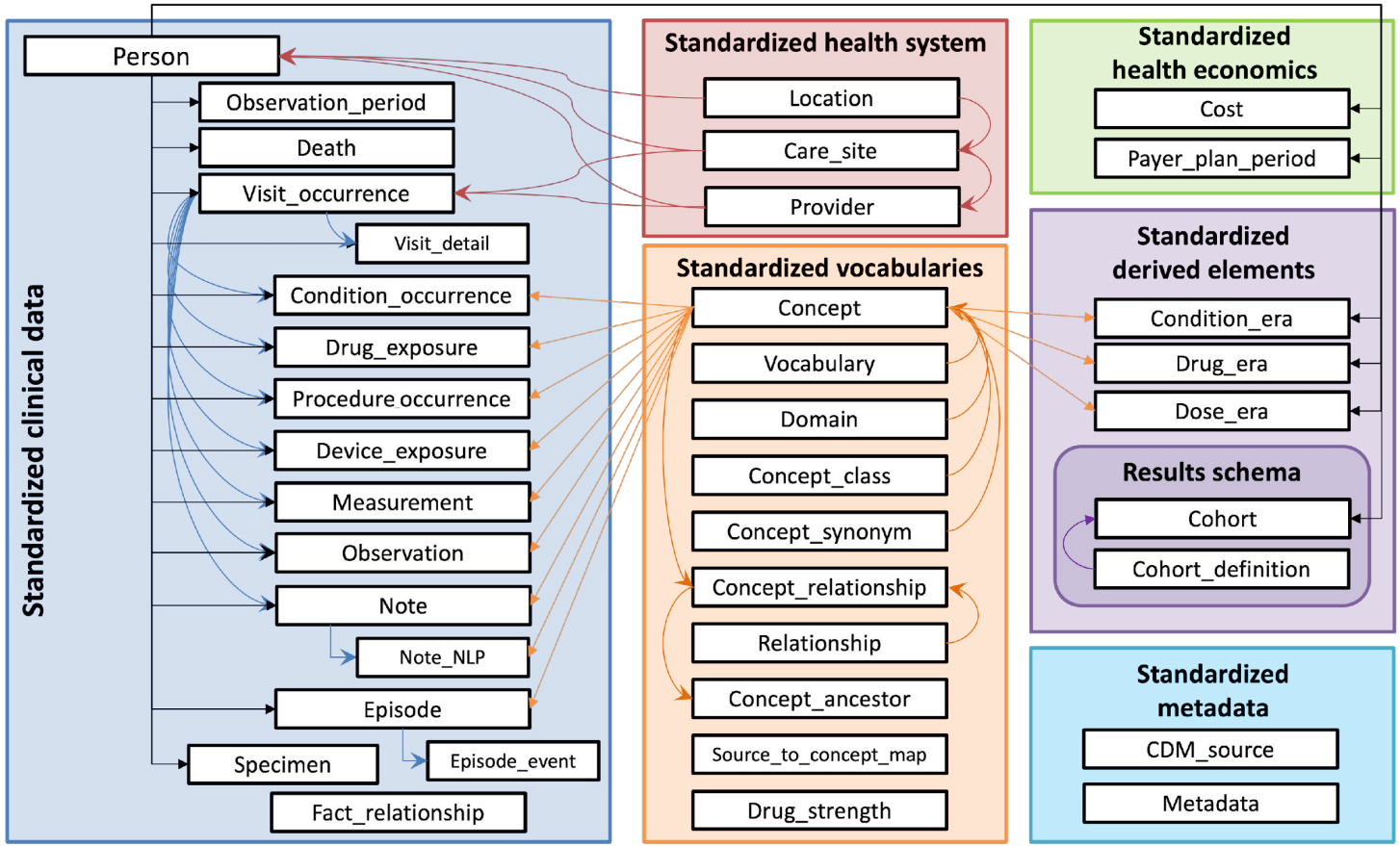
OHDSI By The Numbers

- 3,266 collaborators
- 80 countries
- 21 time zones
- 6 continents
- 1 community

https://www.ohdsi.org/wp-content/uploads/2022/10/OHDSI-OurJourney-2022.pdf

Understanding OMOP CDM

(Observational Medical Outcomes Partnership - Common Data Model)



39 tables

433 fields

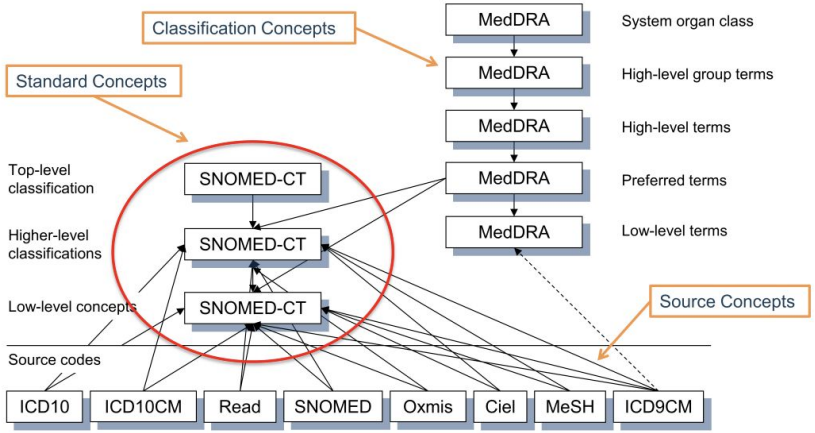
7 categories

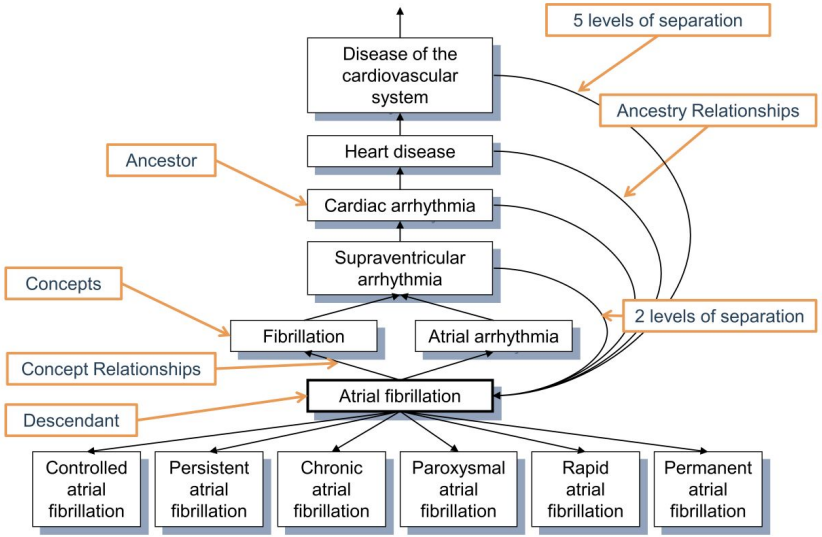
CDM v5.4

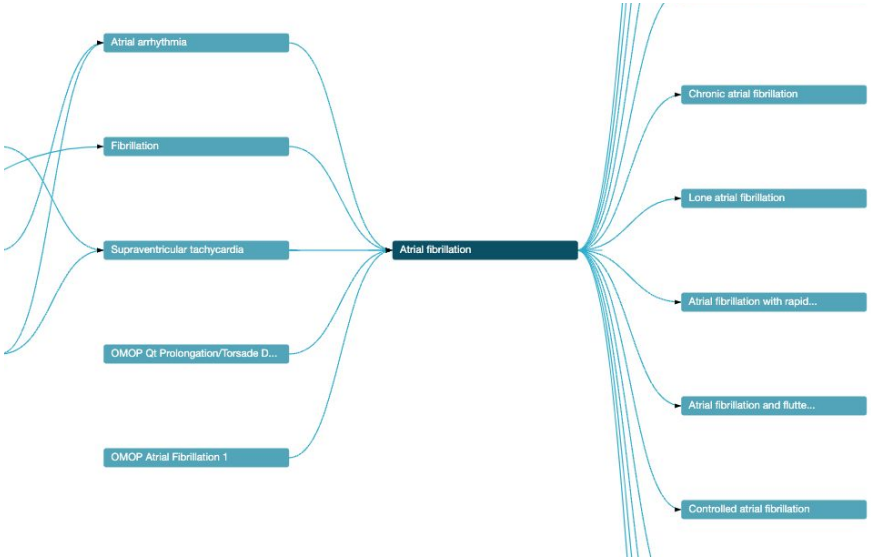
https://ohdsi.github.io/CommonDataModel/cdm54.html

Conversion to OMOP Common Data Model

Column	Explanation	Example
CONCEPT_ID	Primary Key	313217
CONCEPT_NAME	Description	Atrial Fibrillation
DOMAIN_ID	Domain	Condition
VOCABULARY_ID	Vocabulary	SNOMED
CONCEPT_CLASS_ID	Class in Vocabulary	Clinical Finding
STANDARD_CONCEPT	Standard, Source of Classification	S
CONCEPT_CODE	Code in Vocabulary	49436004
VALID_START_DATE	Valid during time interval	01-Jan-1970
VALID_END_DATE	Valid during time interval	31-Dec-2099
INVALID_REASON	Valid during time interval	







Conversion to OMOP Common Data Model

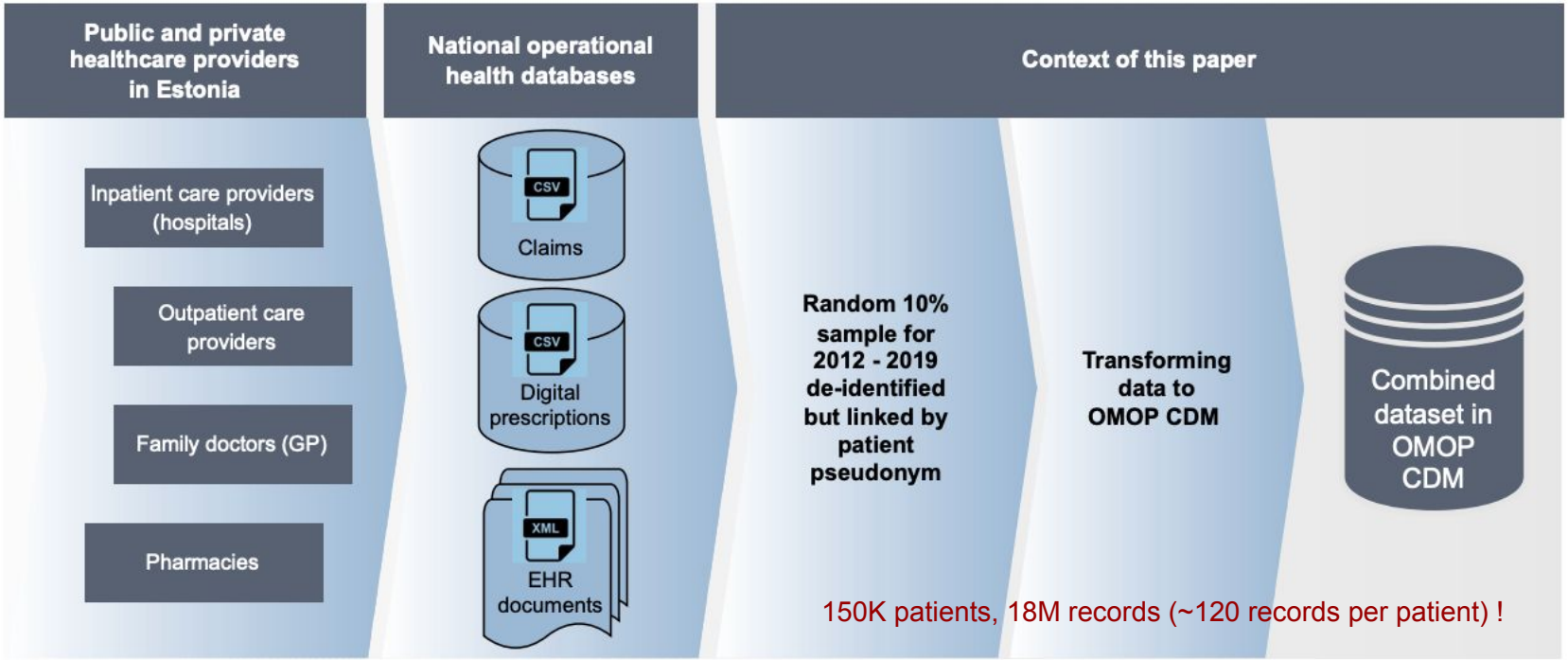


Figure 1. The data acquisition process of national health databases in Estonia and the context of this article.

Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned

https://pubmed.ncbi.nlm.nih.gov/38058679/, March 2023

Table 2. Number of entries in OMOP CDM tables together with mapping rates.

OMOP CDM table	Total number of entries in table	The number of entries mapped to OMOP standard concepts	Mapping rate
Location	1	Not applicable	Not applicable
care_site	1820	Not applicable	Not applicable
person	149 364	Not applicable	Not applicable
death	8277	Not applicable	Not applicable
observation_period	149 364	Not applicable	Not applicable
visit_occurrence	18 281 120	18 281 120	100.0%
visit_details	48 002	48 002	100.0%
condition_occurrence	20 351 014	20 333 065	99.9%
procedure_occurrence	6 956 568	5 392 192	77.5%
drug_exposure	7 945 992	7 842 231	98.7%
device_exposure	77 842	47 296	60.8%
observation	15 203 064	14 762 978	97.1%
measurement	32 230 620	29 250 571	90.8%

Table 3. Mapped concepts and number of entries according to source vocabularies.

Source vocabulary	Number of source concepts	Number of mapped concepts	Mapped concepts (%)	Number of source entries	Number of mapped entries	Mapped entries (%)
ICD-10	9752	9751	100.0	22 738 540	22 738 530	100.0
LOINC	3169	2652	83.7	20 640 878	20 488 392	99.3
Drug product code	3946	3589	91.0	7 562 932	7 502 917	99.2
ATC[a]	202	143	70.8	60 015	12 682	21.1
Local service codes	2518	979	38.9	29 252 643	24 757 917	84.6
NCSP	3960	602	15.2	842 504	420 140	49.9
Other[b]	1396	1396	100.0	709 268	709 268	100.0

[a] Only entries which were not mapped on drug product code level.
[b] Cancer-related codes, pathology codes, and body measurements.

Transformed over 100 million entries to standard concepts using standard OMOP vocabularies with the average mapping rate 95%.

https://pubmed.ncbi.nlm.nih.gov/38058679/, March 2023

Healthcare NLP by John Snow Labs

RxNorm Resolver

ICD10 Resolver

Snomed Resolver

MedDRA Resolver

Assertion Status Detection

Risk Adj. Module

Named Entity Recognition

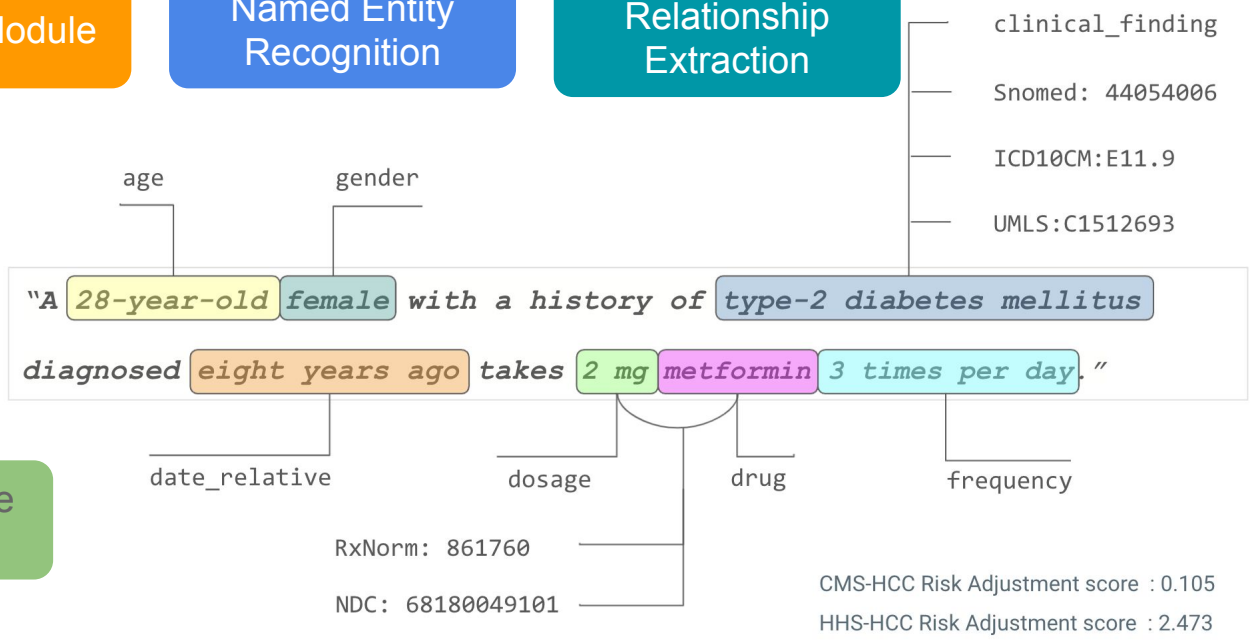
Relationship Extraction

Sentence Splitter

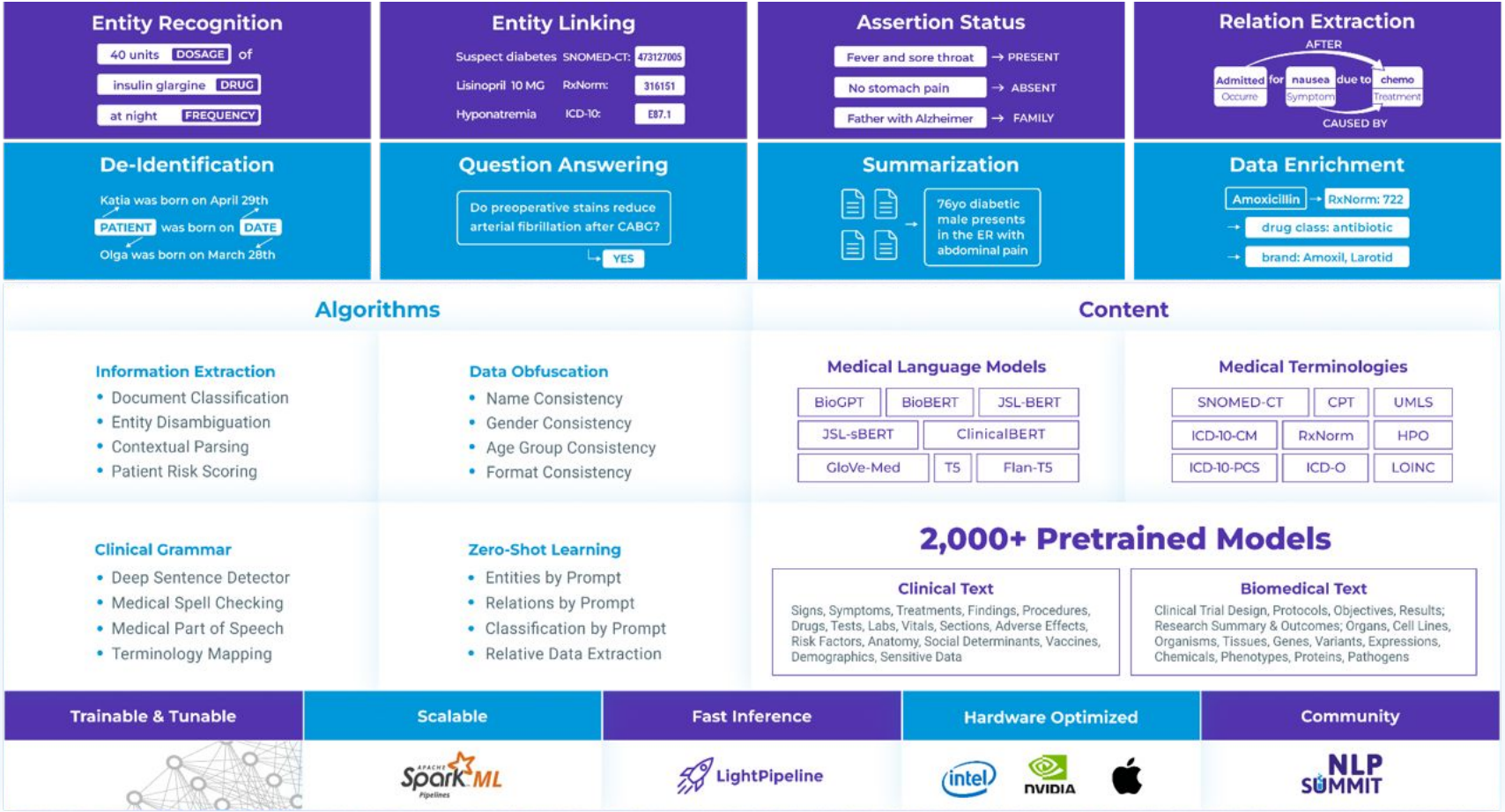
Tokenizer

Bert Embeddings

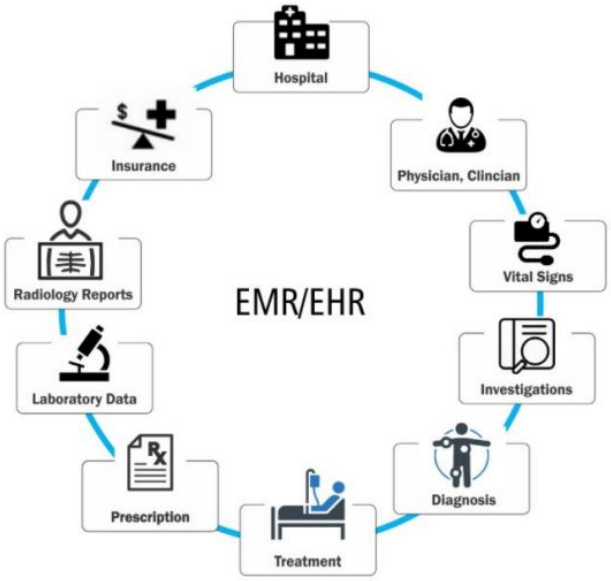
Small Language Models (SLM)

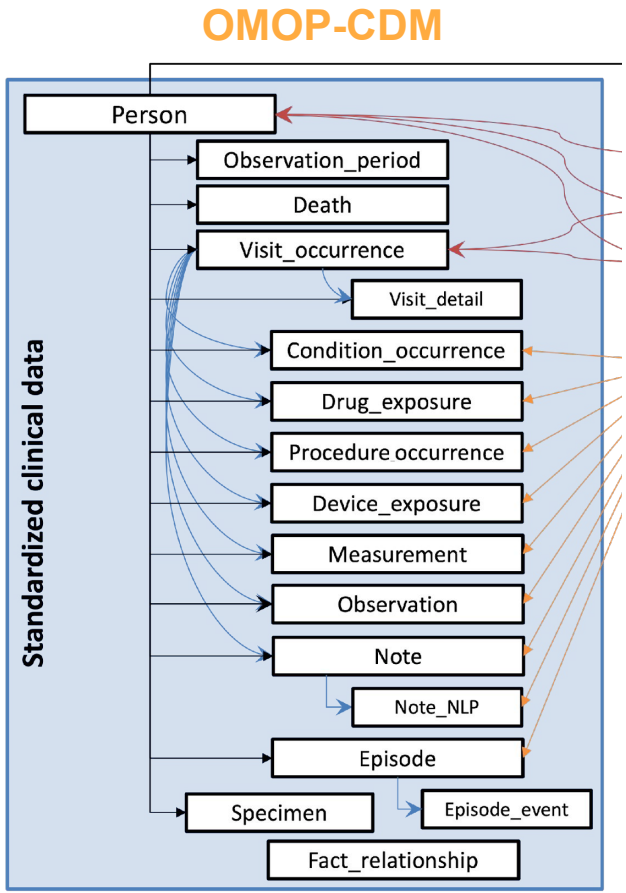


Healthcare NLP by John Snow Labs

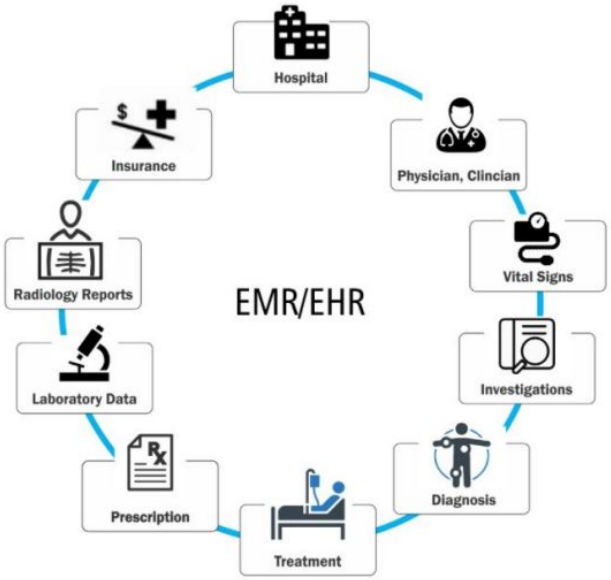


From EHR to OMOP Common Data Model



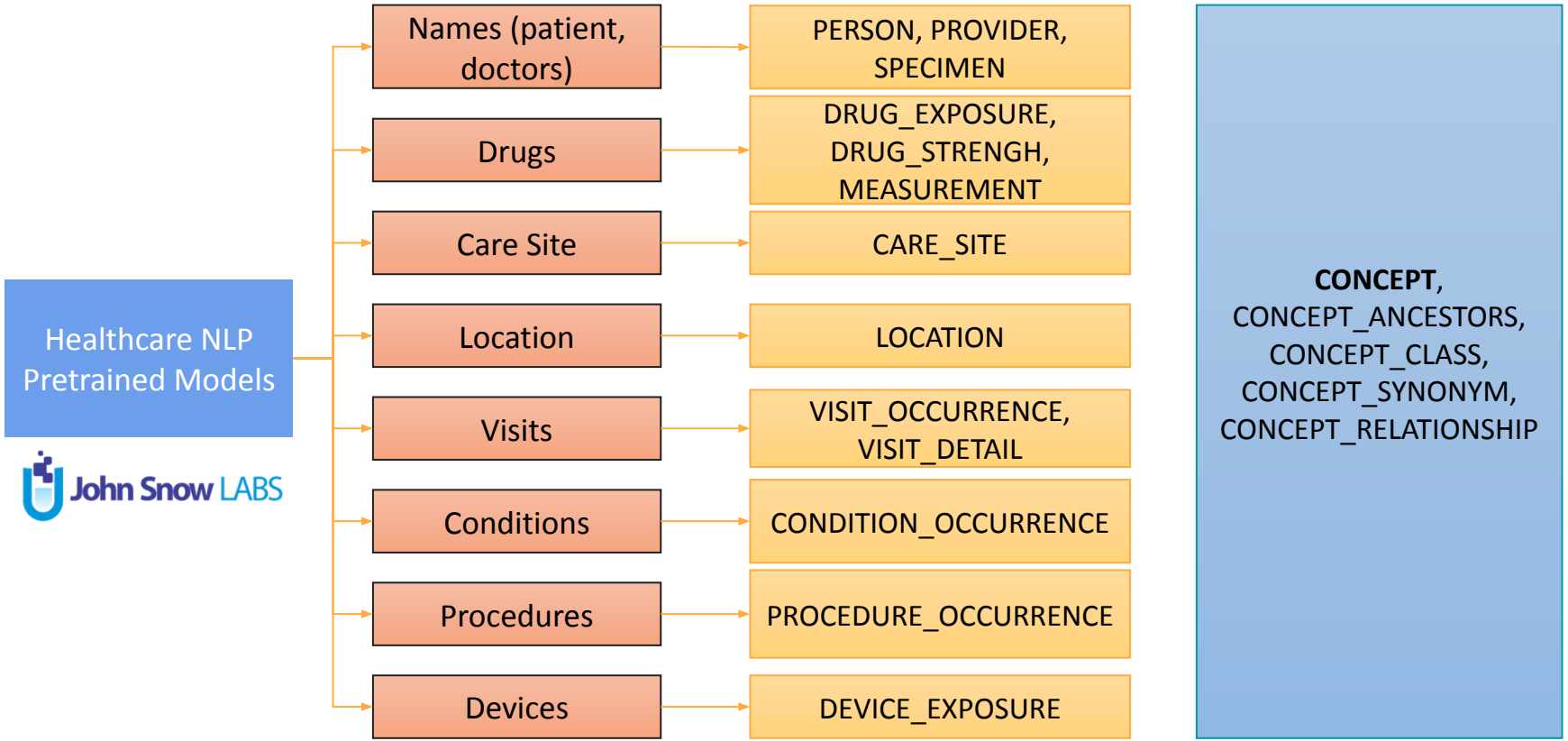


From EHR to OMOP Common Data Model



```
from pyspark.ml import Pipeline

document_assembler = DocumentAssembler()\
 .setInputCol("text")\
 .setOutputCol("document")

sentenceDetector = SentenceDetector()\
 .setInputCols(["document"])\
 .setOutputCol("sentences")

tokenizer = Tokenizer() \
 .setInputCols(["sentences"]) \
 .setOutputCol("token")

normalizer = Normalizer()\
 .setInputCols(["token"])\
 .setOutputCol("normal")

word_embeddings=WordEmbeddingsModel.pretrained()\
 .setInputCols(["document","normal"])\
 .setOutputCol("embeddings")

nlpPipeline = Pipeline(stages=[
 document_assembler,
 sentenceDetector,
 tokenizer,
 normalizer,
 word_embeddings,
 ])

nlpPipeline.fit(df).transform(df)
```

From EHR to OMOP Common Data Model



> Just a subset of entire process for illustrative purposes. The entire process is non-trivial !

From EHR to OMOP Common Data Model

Conditions

https://demo.johnsnowlabs.com/healthcare/OMOP_CDM/

subject_id	hadm_id	note_id	chunk	ner_label	ner_confidence	source	code	resolution	resolution_confidence	vocabul…
10001884	21268656	10001884-DS-24	Atypical chest pain	PROBLEM	0.9705	ner_clinical_chu…	R07.89	atypical chest…	0.4730	ICD10CM
10001884	21268656	10001884-DS-24	Pain	PROBLEM	0.9985	ner_clinical_chu…	R52	pain [pain, un…	0.9582	ICD10CM
10001884	21268656	10001884-DS-24	Asthma	PROBLEM	0.9991	ner_clinical_chu…	J45	asthma [asth…	0.9244	ICD10CM
10001884	21268656	10001884-DS-24	Diverticulitis	PROBLEM	0.9995	ner_clinical_chu…	K57.9	diverticulitis […	0.9881	ICD10CM
10001884	21268656	10001884-DS-24	HTN	PROBLEM	0.9979	ner_clinical_chu…	G60.0	hmsn [heredit…	0.2716	ICD10CM

Drugs

subject_id	hadm_id	note_id	chunk	ner_label	ner_confidence	source	code	resolution	resolution_conf…	vocabul…
10001884	21268656	10001884-DS-24	aspirin	DRUG	0.9989	ner_posology_chunks	1537020	aspirin Effervesce…	0.4929	RXNORM
10001884	21268656	10001884-DS-24	Calcium	DRUG	0.8857	ner_posology_chunks	1895	calcium [calcium]	0.9784	RXNORM
10001884	21268656	10001884-DS-24	hydrochlorothiazide	DRUG	0.8975	ner_posology_chunks	370635	hydrochlorothiazid…	0.4909	RXNORM
10001884	21268656	10001884-DS-24	omeprazole	DRUG	0.8263	ner_posology_chunks	1796429	omeprazole Oral …	0.4918	RXNORM
10001884	21268656	10001884-DS-24	simvastatin	DRUG	0.8802	ner_posology_chunks	36567	simvastatin [simva…	0.4950	RXNORM

Procedures

subject_id	hadm_id	note_id	chunk	ner_label	ner_confidence	source	code	resolution	resolution_confidence	vocabulary
10001884	25758848	10001884-DS-28	ATRIAL TACHYCARDIA…	PROCEDURE	0.17653334	ner_jsl_chunks	D0169YZ	High Dose Rate (HD…	0.0534	ICD10PCS
10001884	25758848	10001884-DS-28	BUNIONECTOMY HIP…	PROCEDURE	0.5097333	ner_jsl_chunks	0MRL4KZ	Replacement of Right…	0.0551	ICD10PCS
10001884	29675586	10001884-DS-29	FIBRILLATION ANXIE…	PROCEDURE	0.32776666	ner_jsl_chunks	04723Z6	Dilation of Gastric Art…	0.0872	ICD10PCS
10001884	29675586	10001884-DS-29	BUNIONECTOMY HIP …	PROCEDURE	0.5097333	ner_jsl_chunks	0MRL4KZ	Replacement of Right…	0.0551	ICD10PCS
10001884	29675586	10001884-DS-29	ADMISSION LABS	PROCEDURE	0.3921	ner_jsl_chunks	F01ZBZZ	Bed Mobility Assessm…	0.2956	ICD10PCS

From EHR to OMOP Common Data Model

Following thorough evaluation and diagnostic testing, the patient has been accurately diagnosed with AFib.

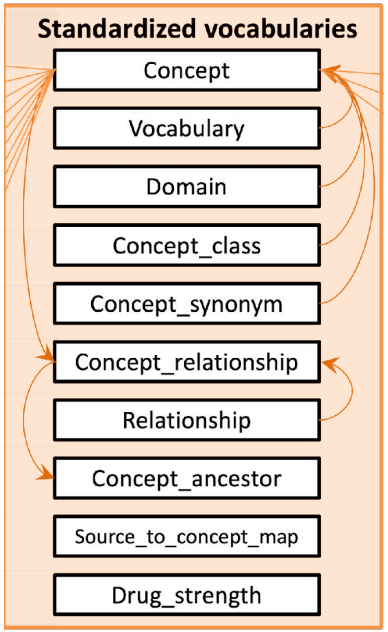
NER *condition*

"atrial fibrilation"

ATHENA

Atrial fibrillation

DETAILS	
Domain ID	Condition
Concept Class ID	Disorder
Vocabulary ID	SNOMED
Concept ID	313217
Concept code	49436004
Validity	Valid
Concept	Standard
Synonyms	Atrial fibrillation (disorder) AF - Atrial fibrillation
Valid start	31-Jan-2002
Valid end	31-Dec-2099



SNOMED Resolver

Concept_ID	CONCEPT_CODE
313217	49436004

From EHR to OMOP Common Data Model

Patient presented with Atypical chest pain, normal blood pressure; stress echo showed no issues. Prescribed aspirin 50 mg daily. Advised lifestyle adjustments, follow-up in 4 weeks. Stable upon discharge


```
"condition_annotations": [
  {
    "subject_id": 10001884,
    "hadm_id": 21268656,
    "note_id": "10001884-DS-24",
    "chunk": "Atypical chest pain",
    "ner_label": "PROBLEM",
    "ner_confidence": 0.9705,
    "source": "ner_clinical_chunks",
    "code": "102589003",
    "resolution": "atypical chest pain",
    "resolution_confidence": 0.9263,
    "vocabulary": "SNOMED_FINDINGS"
  },
```

```
"drug_annotations": [
  {
    "subject_id": 10001884,
    "hadm_id": 21268656,
    "note_id": "10001884-DS-24",
    "chunk": "aspirin",
    "ner_label": "DRUG",
    "ner_confidence": 0.9989,
    "source": "ner_posology_chunks",
    "code": 1537020,
    "resolution": "aspirin Effervescent Oral Tablet",
    "resolution_confidence": 0.4929,
    "vocabulary": "RXNORM"
  },
```

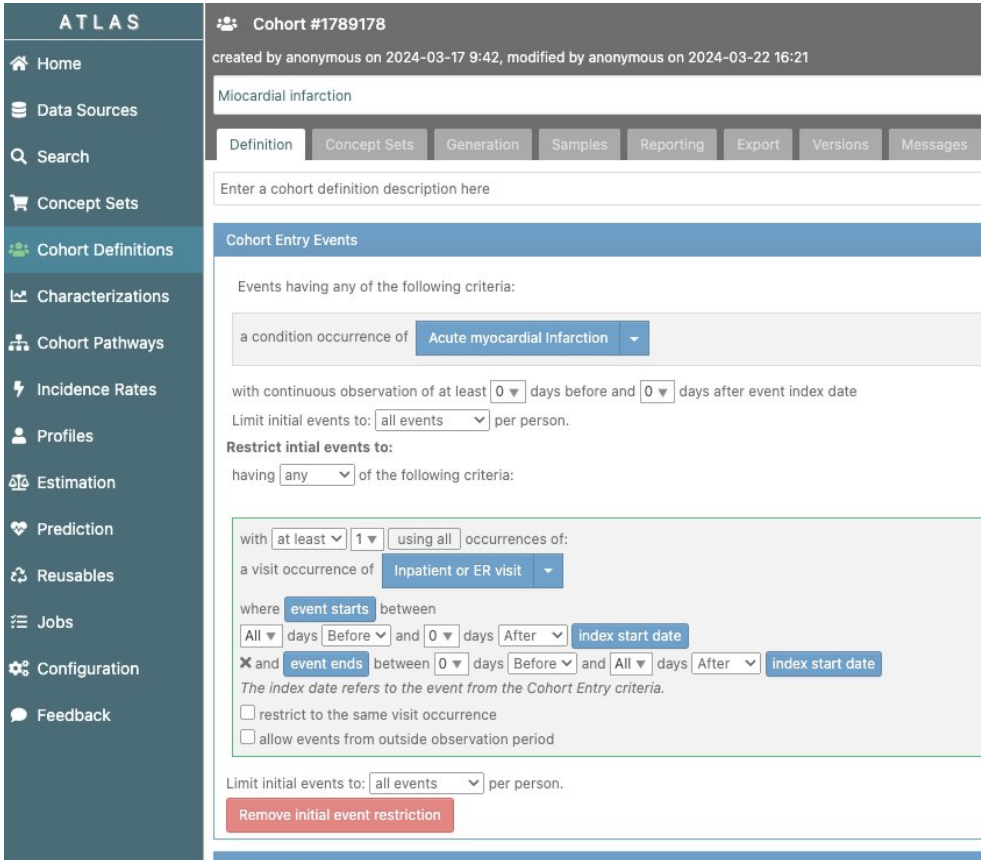
```
"measurement_annotations": [
  {
    "subject_id": 10001884,
    "hadm_id": 21268656,
    "note_id": "10001884-DS-24",
    "chunk": "normal blood pressure",
    "ner_label": "BLOOD_PRESSURE",
    "ner_confidence": 0.38863334,
    "source": "ner_jsl_chunks",
    "code": "LA4439-1",
    "resolution": "Within Normal Limits [Within Normal Limits]",
    "resolution_confidence": 0.1407,
    "vocabulary": "LOINC"
  },
```

```
"procedure_annotations": [
  {
    "subject_id": 10001884,
    "hadm_id": 26679629,
    "note_id": "10001884-DS-25",
    "chunk": "stress echo",
    "ner_label": "TEST",
    "ner_confidence": 0.4301,
    "source": "ner_jsl_chunks",
    "code": "816996009",
    "resolution": "stress echocardiography",
    "resolution_confidence": 0.363,
    "vocabulary": "SNOMED_PROCEDURES_MEASUREMENTS"
  },
```

```
"relations_annotations": [
  {
    "subject_id": 10001884,
    "hadm_id": 21268656,
    "note_id": "10001884-DS-24",
    "relation": "DRUG-STRENGTH",
    "re_confidence": 1.0,
    "direction": "both",
    "label1": "DRUG",
    "label2": "STRENGTH",
    "chunk1": "hydrochlorothiazide",
    "chunk2": "50 mg",
    "code1": "370635",
    "code2": "",
    "resolver_model": "RXNORM",
    "re_model": "re_posology"
  },
```

How to Query OMOP Common Data Model

- Query with plain SQL from scratch
- Using OHDSI **Query Library** for prebuilt queries or **ATLAS**, a unified interface to patient level data and analytics.
- Natural language to SQL libraries (nl2sql, nostos, etc)
- via LLMs (fine-tuned vs zero-shot vs few-shot)



https://github.com/OHDSI/QueryLibrary
a library of standard queries that run against the OMOP-CDM

https://atlas-demo.ohdsi.org/

How to Query OMOP Common Data Model

</> SQL

```sql
WITH statins AS (
SELECT descendant_concept_id AS concept_id
  FROM demo_cdm.concept_ancestor
WHERE ancestor_concept_id = 1539403
), diuretics AS (
SELECT descendant_concept_id AS concept_id
  FROM demo_cdm.concept_ancestor
WHERE ancestor_concept_id = 974166
)
SELECT COUNT(DISTINCT de1.person_id) AS num_statin_users,
       COUNT(DISTINCT de2.person_id) AS also_bp_users
  FROM demo_cdm.drug_exposure de1
  JOIN statins s
    ON de1.drug_concept_id = s.concept_id
  JOIN demo_cdm.drug_exposure de2
    ON de1.person_id = de2.person_id
  LEFT JOIN diuretics d
    ON de2.drug_concept_id = d.concept_id
   AND de2.drug_exposure_start_date < de1.drug_exposure_end_date
   AND de2.drug_exposure_end_date   > de1.drug_exposure_start_date;
```

How many patients on statins are also on blood pressure meds?

How to Query OMOP Common Data Model

```sql
SELECT DISTINCT
  CASE
    WHEN concept_name_1>concept_name_2 THEN concept_name_1
    ELSE concept_name_2
  END AS condition1,
  CASE
    WHEN concept_name_1>concept_name_2 THEN concept_name_2
    ELSE concept_name_1
  END AS condition2
FROM
  (SELECT
    concept_name_1  AS concept_name_1,
    concept_name    AS concept_name_2
   FROM
    (SELECT
      condition_concept_id_2,
      concept_name AS concept_name_1
     FROM
      (SELECT
        table1.condition_concept_id AS condition_concept_id_1,
        table2.condition_concept_id AS condition_concept_id_2
       FROM
          (SELECT * FROM cdm.condition_era
           WHERE person_id = 136931019 -- Input person_id
           ) AS table1,
          (SELECT * FROM cdm.condition_era
           WHERE person_id = 136931019 -- Input person_id
          ) AS table2
      WHERE table2.condition_era_start_date <= table1.condition_era_end_date

AND (table2.condition_era_end_date IS NULL OR table2.condition_era_end_date >= table1.co
ndition_era_start_date)
            AND table1.condition_concept_id<>table2.condition_concept_id
      ) AS comorb
    LEFT JOIN cdm.concept AS concept_list
    ON comorb.condition_concept_id_1=concept_list.concept_id
    ) AS comorb2
  LEFT JOIN cdm.concept AS concept_list
  ON comorb2.condition_concept_id_2=concept_list.concept_id
  ) AS condition_pairs;
```

What are a person's comorbidities?

How to Query OMOP Common Data Model

John Snow LABS

Fine tune SLM (<1B) with schema pruning

Query:
How many patients were born before the year 2060?

Ground Truth SQL:
```
Select count (distinct demographic."subject_id")
from demographic
where demographic."dob_year" < "2060"
```

Augmented version of SQL (illustrative):
```
Select count (distinct demographic."subject_id")
from demographic
where demographic."dob_year" less than "2060"
```

GPT 3.5-turbo:
```
Select count (distinct "subject_id")
from demographic
where "year_of_birth" <= 2060
```

GPT 4:
```
Select count (distinct "subject_id")
from demographic
where "year_of_birth" < "2060"
```

FlanT5-Base, fine-tuned (ours):
```
Select count (distinct demographic."subject_id")
from demographic
where demographic."dob_year" <= "2060"
```

LLaMA-2-7B:
```
SELECT COUNT(DISTINCT name)
FROM Demographic
WHERE date_of_birth <='2060';
```

FlanT5-Large, fine-tuned (ours):
```
Select count (distinct demographic."subject_id")
from demographic
where demographic."dob_year" < "2060"
```

Defog-SQLCoder:
```
SELECT COUNT(*)
FROM Demographic d
WHERE d.year_of_birth < 2060;
```

Breaking New Ground in Medical Text-to-SQL: The Potential of Compact, Fine-Tuned Language Models

Youssef Mellah, PhD, Veysel Kocaman, Hasham Ul Haq, David Talby, PhD

youssef,veysel,hasham,david@johnsnowlabs.com

John Snow Labs Inc., Delaware, USA

Artificial Intelligence in Health
Electronic ISSN: 3029-2387
Date of Foundation: 2024-01-19

Objective: This study focuses on the effectiveness of compact, fine tuned language models (LLMs) for Text-to-SQL tasks in the healthcare domain, particularly using the MIMICSQL dataset. The goal is to explore a 'schema-less' approach to Text-to-SQL conversion, which utilizes straightforward natural language questions without requiring prior knowledge of the database schema.

Materials and Methods: In this study, we compared the performance of various LLMs including GPT-3.5, GPT-4, DeFrog, and Llama under zero-shot settings. The TREQS model's performance was evaluated using its original test metrics as reported by its authors. For our approach, we modified and augmented the original MIMICSQL training set and fine-tuned compact language models (LLMs) on this enhanced dataset. The models were then tested on an intact test set from MIMIC SQL, emphasizing our schema-less approach in Text-to-SQL conversion.

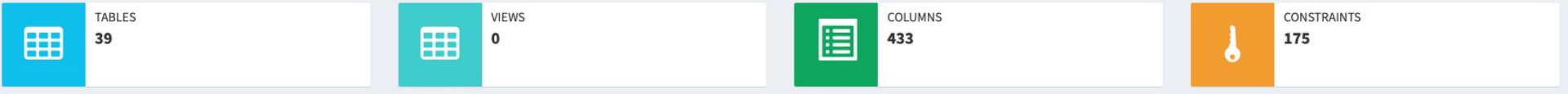
Model	Parameter	LFA
TREQS [26]	2.8M	0.48
TREQS + Recover [26]	2.8M	0.55
GPT-3.5-Turbo [2]	20B	0.60
LLaMA-2-7B [24]	7B	0.60
Defog-SQLCoder [1]	15B	0.65
GPT-4 [2]	-	0.70
FlanT5-Base (Ours) [5]	220M	0.56
FlanT5-Large (Ours) [5]	770M	**0.85**

On MIMIC-SQL test set

How to Query OMOP Common Data Model

Strategy for Managing Large OMOP Databases: Combining Language Models with Schema Reduction

- To navigate the OMOP schema's complexity, we've developed a streamlined method for isolating tables and columns crucial to user queries.
- We utilize cosine similarity to determine the relevance of schema elements to a user's query, calculating scores between the query's and schema elements' embeddings. This helps in highlighting the most pertinent parts of the schema.
- Incorporating a refined schema into LLM significantly surpasses GPT4 and GPT4-Turbo, achieving over 90% accuracy in generating SQL queries.
- Our selective methodology enhances the Text2SQL conversion process, ensuring efficient use of computational resources by focusing on key schema elements, thus boosting both efficiency and accuracy in query handling.
- Flexibility and Expandability: Our approach is designed to be adaptable, catering to a wide range of datasets and schemas. It supports easy integration of natural language queries across various data platforms.



How to Query OMOP Common Data Model

How many patients having angioplasty also used Plavix?

Healthcare NLP

John Snow LABS

Text2SQL LLM output

```
SELECT COUNT(DISTINCT e.person_id) AS
patient_count

FROM procedure_occurrence p

JOIN drug_exposure e ON p.person_id =
e.person_id

WHERE p.procedure_concept_id = 1478
  AND e.drug_concept_id = 1015
```

Angioplasty >> procedure with concept id 1478

Plavix >> drug with concept id 1015

How to Query OMOP Common Data Model

John Snow LABS

Fine tune SLM (<1B) with schema pruning

Query:

How many drug exposures were recorded for each person along with their gender?

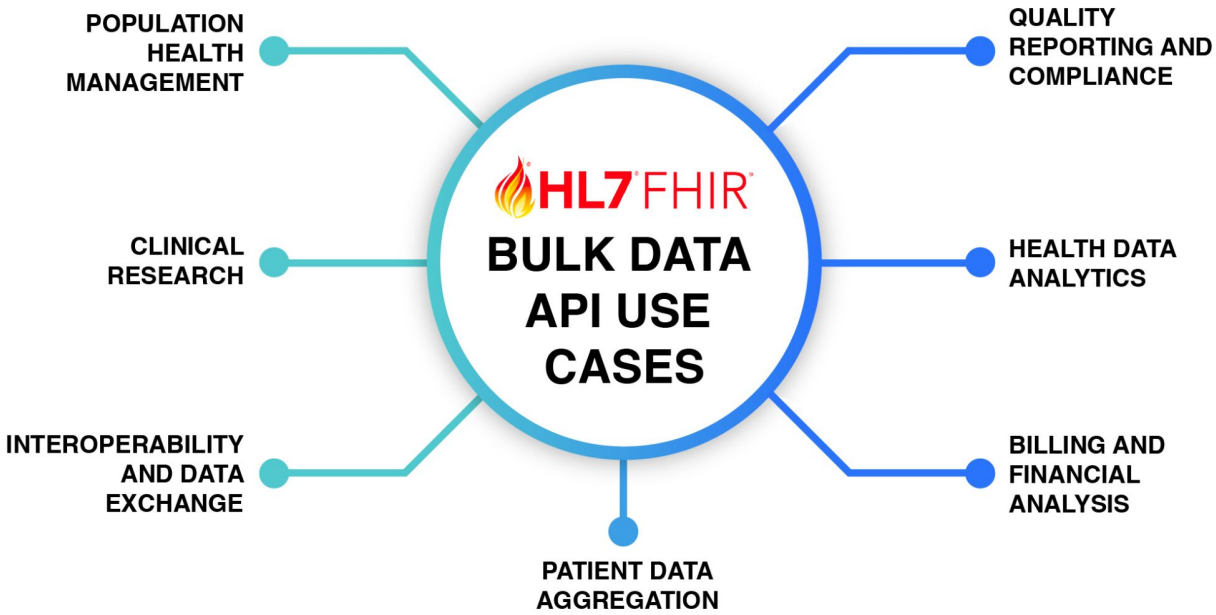
GPT 4 (with schema in the prompt):

```
SELECT p.person_id, p.gender_source_value, COUNT(de.person_id)
AS drug_exposure_count
FROM person p
LEFT JOIN drug_exposure de ON p.person_id = de.person_id
GROUP BY p.person_id, p.gender_source_value;
```

Ours:

```
SELECT p.person_id,
       p.gender_concept_id,
       COUNT(de.drug_exposure_id) AS drug_exposure_count
FROM drug_exposure de
JOIN person p ON de.person_id = p.person_id
GROUP BY p.person_id,
       p.gender_concept_id
ORDER BY drug_exposure_count DESC;
```

Text to FHIR

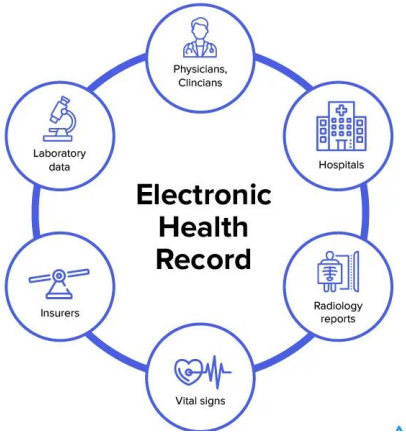


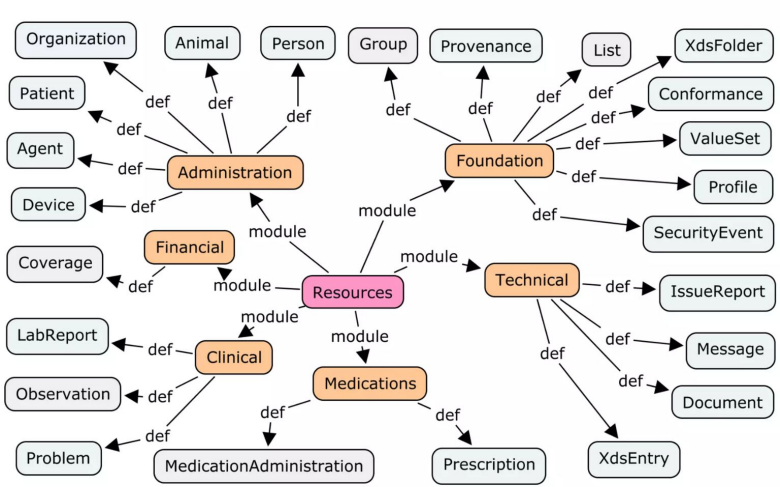
The **Fast Healthcare Interoperability Resources** (**FHIR**) standard is a set of rules and specifications for exchanging electronic health care data. The goal of FHIR is to enable the seamless and secure exchange of health care information (to transform our current state of messy, disparate backend data into a single, universal personal health record for each and every patient). It is essentially an API for healthcare. It is a free, vendor-neutral mechanism for promoting interoperability between any and all sources of healthcare data.

FHIR <-> OHDSI OMOP

- FHIR is a standard for exchanging healthcare information electronically, while OMOP is a common data model for representing clinical data for research purposes.
- FHIR is focused on enabling interoperability and data exchange between different healthcare systems and applications, such as electronic health records (EHRs), personal health records (PHRs), and medical devices. It defines a set of "resources" that represent clinical concepts and can be easily shared.
- In contrast, OMOP is designed to provide a standardized way of representing observational healthcare data from different sources in a common format, to enable large-scale data analysis and research.
- While FHIR and OMOP have some overlapping capabilities in representing clinical data, they serve different primary purposes. FHIR is more focused on data exchange and interoperability, while OMOP is focused on data integration and analysis for research.
- In practice, the two standards can be used together - FHIR can be used to exchange data, which can then be transformed into the OMOP common data model for analysis. Tools and mappings have been developed to translate between FHIR and OMOP representations of data.

Text to FHIR






```
<Patient xmlns="http://hl7.org/fhir">
  <id value="glossy"/>
  <meta>
    <lastUpdated value="2014-11-13T11:41:00+11:00"/>
  </meta>
  <text>
    <status value="generated"/>
    <div xmlns="http://www.w3.org/1999/xhtml">
      <p>Henry Levin the 7th</p>
      <p>MRN: 123456. Male, 24-Sept 1932</p>
    </div>
  </text>
  <extension url="http://example.org/StructureDefinition/trials">
    <valueCode value="renal"/>
  </extension>
  <identifier>
    <use value="usual"/>
    <type>
      <coding>
        <system value="http://hl7.org/fhir/v2/0203"/>
        <code value="MR"/>
      </coding>
    </type>
    <system value="http://www.goodhealth.org/identifiers/mrn"/>
    <value value="123456"/>
  </identifier>
  <active value="true"/>
  <name>
    <family value="Levin"/>
    <given value="Henry"/>
    <suffix value="The 7th"/>
  </name>
  <gender value="male"/>
  <birthDate value="1932-09-24"/>
  <careProvider>
    <reference value="Organization/2"/>
    <display value="Good Health Clinic"/>
  </careProvider>
</Patient>
```

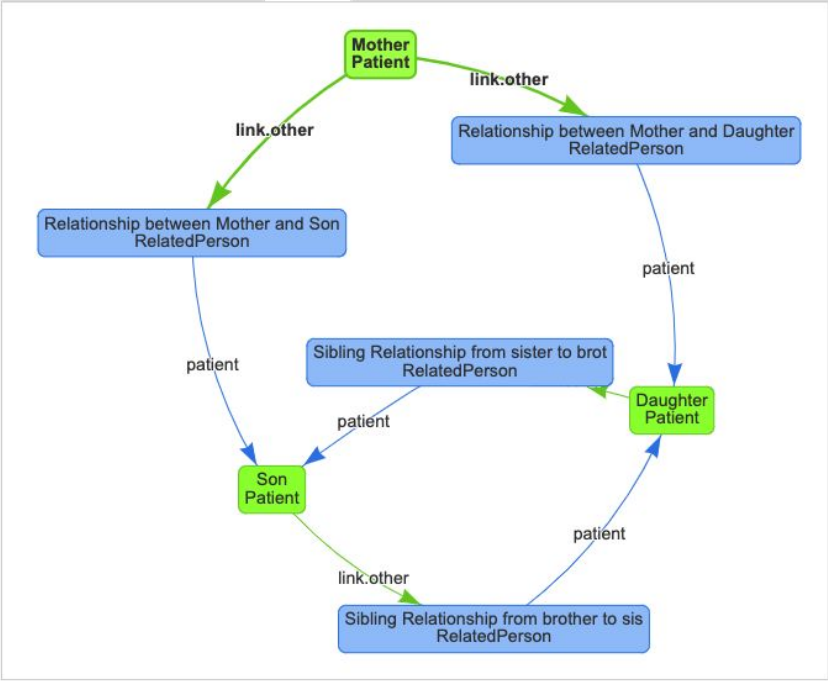
Resource Identity & Metadata

Human Readable Summary

Extension with URL to definition

Standard Data:
- MRN
- Name
- Gender
- Birth Date
- Provider

Text to FHIR




```
{
  "resource": {
    "resourceType": "Patient",
    "id": "mum",
    "text": {
      "div": "<div xmlns=\"http://www.w3.org/1999/xhtml\">Mother</div>",
      "status": "additional"
    },
    "name": [
      {
        "text": "Mary Mother"
      }
    ],
    "link": [
      {
        "other": {
          "reference": "RelatedPerson/sonRelation"
        },
        "type": "seealso"
      },
      {
        "other": {
          "reference": "RelatedPerson/daughterRelation"
        },
        "type": "seealso"
      }
    ]
  }
}
```

Text to FHIR

FHIR Resources in the Bundle

	Count
Medication	3149
Condition	1861
Procedure	1104
Observation	1011
BodyStructure	876
MolecularSequence	463
Patient	349
Encounter	113
Device	44
Bundle	15
ImagingStudy	9
Composition	7
AllergyIntolerance	1

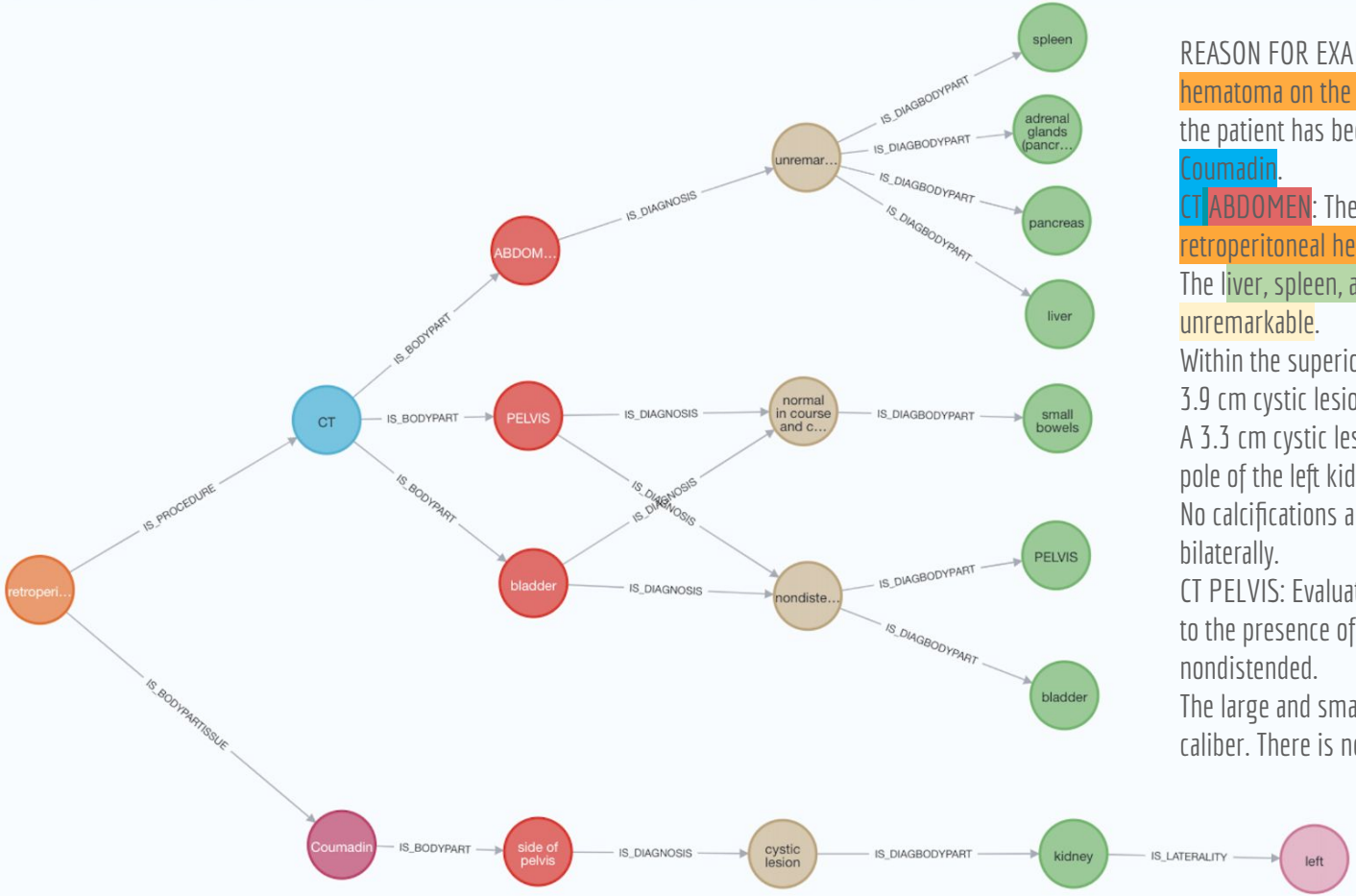
FHIR Bundle


```
{
  "resourceType" : "Bundle"
  "id" : "b43b03bf-456c-4377-bdc2-ca021b12da2f"
  "type" : {...}
  "entry" : [
    0 : {
      "resource" : {
        "resourceType" : "Patient"
        "id" : 10220991
        "identifier" : [
          0 : {
            "use" : "patient"
            "value" : 10220991
          }
        ]
        "active" : true
        "name" : {
          "use" : {
            "code" : "official"
            "display" : "Official"
```


https://demo.johnsnowlabs.com/healthcare/FHIR/

*(20) Exam(1) Procedure(1) BodyPart(4) Diagnosis(4) DiagnosisBodyPart(8) BodyPartIssue(1) Laterality(1)

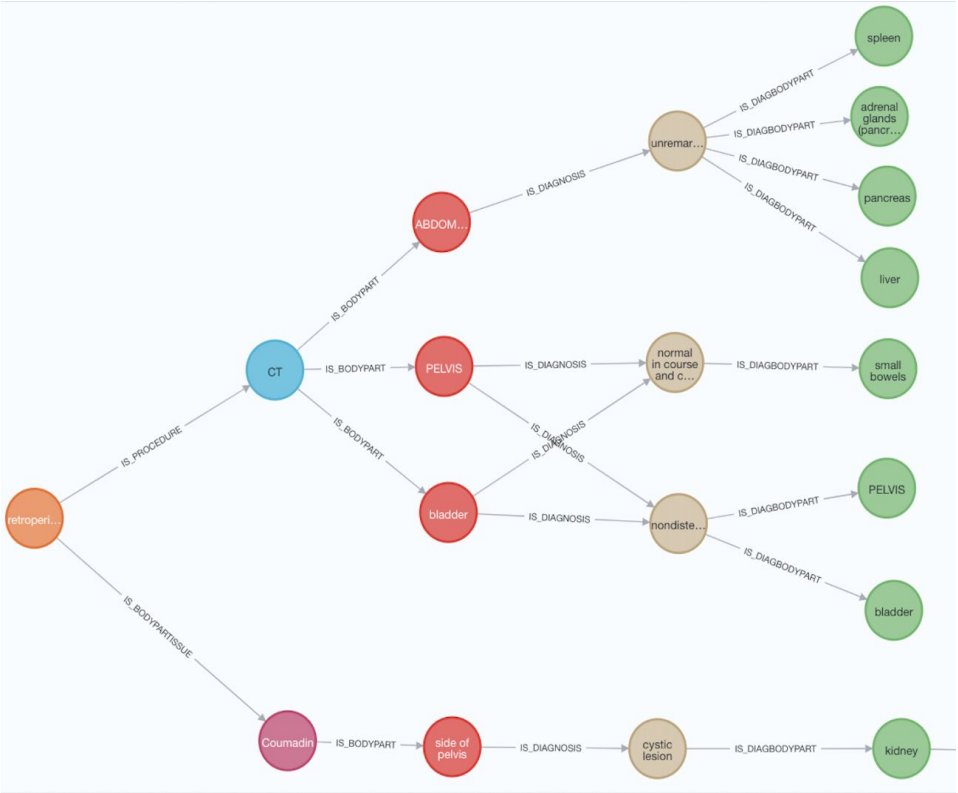
*(21) IS_BODYPARTISSUE(1) IS_PROCEDURE(1) IS_BODYPART(4) IS_DIAGNOSIS(6) IS_DIAGBODYPART(8) IS_LATERALITY(1)



REASON FOR EXAM: Evaluate for retroperitoneal hematoma on the right side of pelvis,
the patient has been following, is currently on Coumadin.
CT ABDOMEN: There is no evidence for a retroperitoneal hematoma.
The liver, spleen, adrenal glands, and pancreas are unremarkable.
Within the superior pole of the left kidney, there is a 3.9 cm cystic lesion.
A 3.3 cm cystic lesion is also seen within the inferior pole of the left kidney.
No calcifications are noted. The kidneys are small bilaterally.
CT PELVIS: Evaluation of the bladder is limited due to the presence of a Foley catheter, the bladder is nondistended.
The large and small bowels are normal in course and caliber. There is no obstruction.

neo4j

How to Query Knowledge Graph



Please write your question

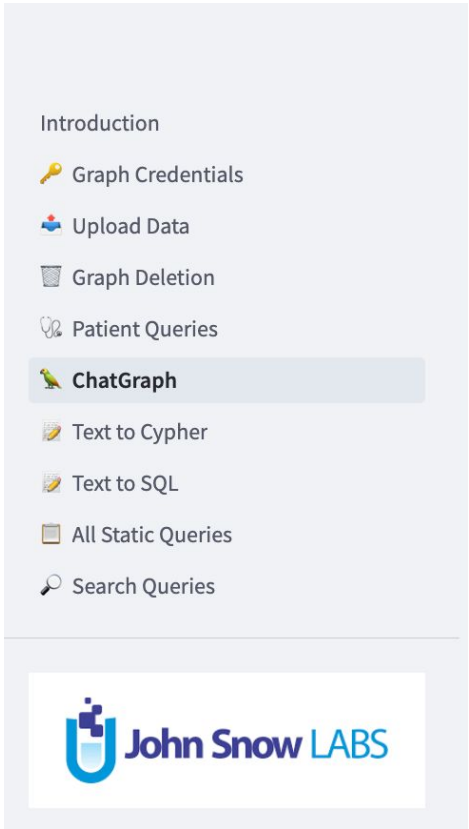
Find patients that have prescribed advil and relations status with all possibilities and return patient ids?

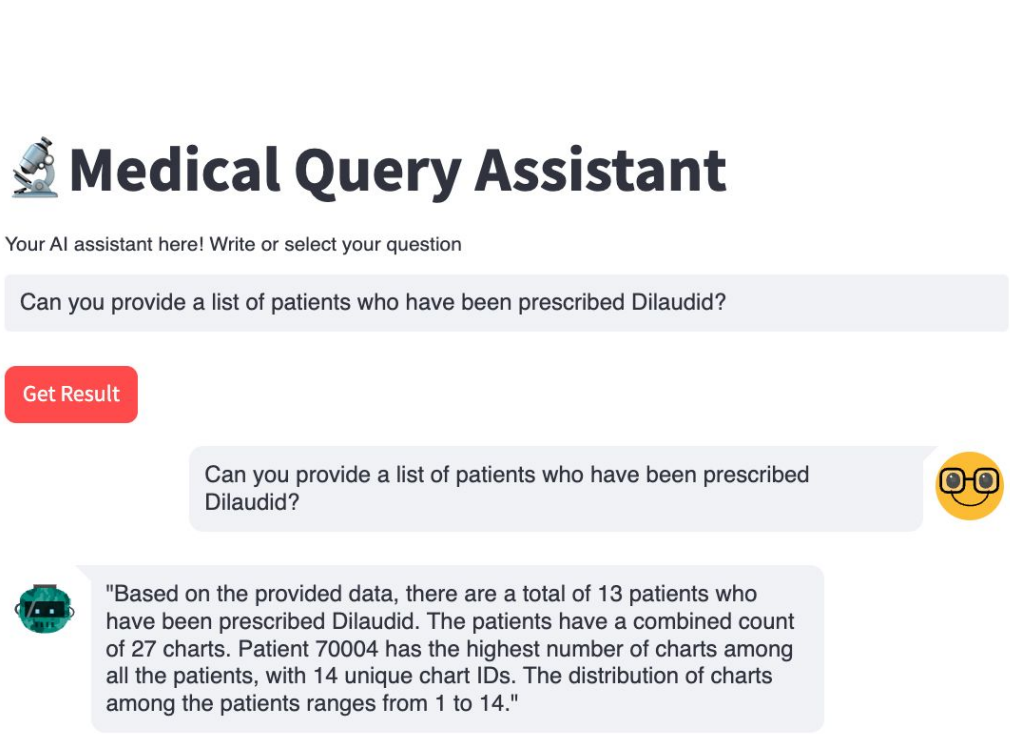
cypher query dict:


```
{
  "0" : NULL
  "stored_session" : []
  "summary_of_query" : [
    0 :
    "Based on the provided data, there are a total of two patients who have
    been prescribed advil. Patient 5062 has one chart associated with the
    prescription (chart ID 29106), while patient 20745 also has one chart
    linked to the advil prescription (chart ID 47600)."
  ]
  "0_user" : NULL
  "generated" : [...]
  "past" : [...]
  "payload" : {...}
  "input" : ""
  "cypher_code" : [
    0 :
    "MATCH (p:Patient)-[:HAS_RECORD]->(c:Chart)-[:MENTIONS]->(d:Drug_BrandName)
    WHERE d.name = "Advil"
    RETURN DISTINCT p.pId as patient_id, COLLECT(DISTINCT c.chartId) as
    chart_ids"
  ]
}
```


Cypher query

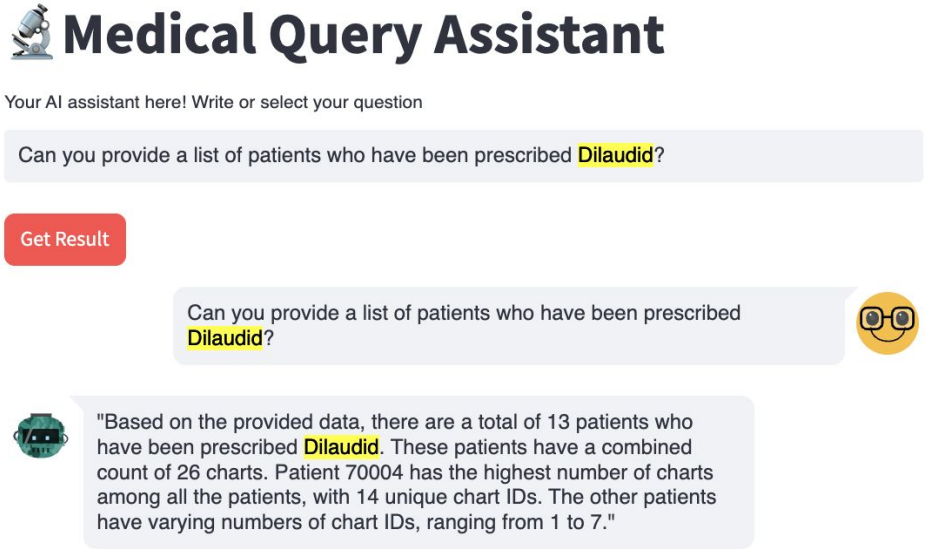
Querying unstructured notes via Knowledge Graph





Querying unstructured notes via Knowledge Graph

```
   Action:
          Lumbar drain kept clamped and opened hourly
(overnight
   started draining 15cc/hr from 2300 on)
          Discussed no CSF output with nsurg and per
PA to be expected
   d/t lge CSF output intraop
          Dilaudid IV/PO ordered (overnight using IV
d/t not eating a
   regular diet yet)
          IVF infusing for hydration
   Response:
          Tolerating PO meds does not want anything to
eat at this
   time
          Pt reports poor pain control
primary team notified
          SBP drops to 80
s after IV dilaudid
   Plan:
          Called out to floor with Q4hr neuro/vitals
          Lumbar drain to be clamped and opened hourly
just to drain
   15cc
          Dilaudid for pain transition to PO from IV
          Advance diet as tolerated
          SICU resident spoke with NSURG and OK to
raise HOB 25
   degrees
"
}
```



Presented by John Snow LABS

NLP SUMMIT

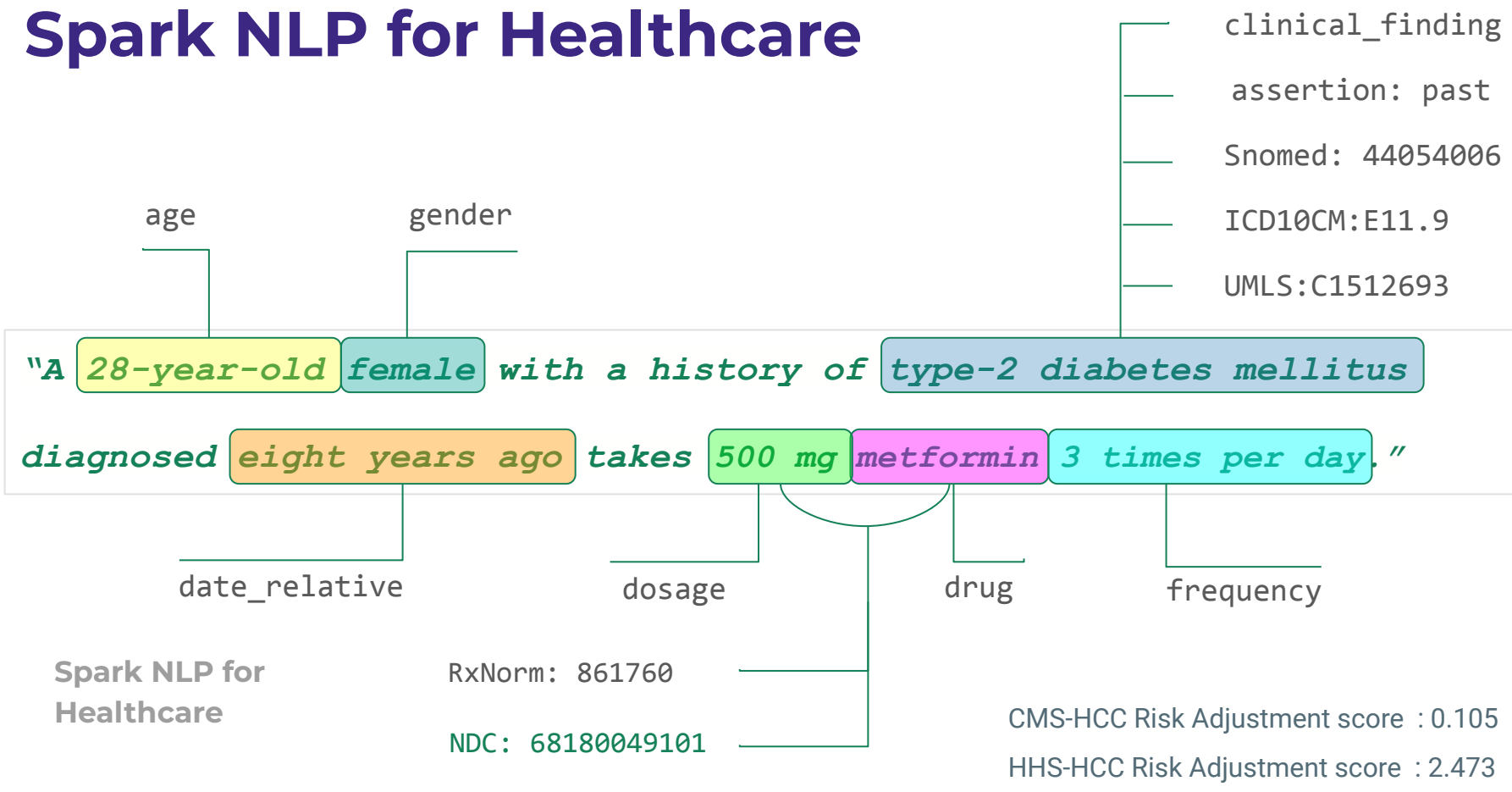
Thank you !

Veysel Kocaman, PhD

Head of Data Science

John Snow Labs

Spark NLP for Healthcare



Understanding OMOP CDM

(Observational Medical Outcomes Partnership - Common Data Model)

Enhancing Healthcare through Data

Foundation: Part of the Observational Health Data Sciences and Informatics (OHDSI) initiative.

Objective: Utilize open-source data solutions to improve human health via large-scale analysis.

Purpose: Standardize the structure and content of observational healthcare data.

Features:

- Enables efficient, reliable evidence production through analysis.
- Incorporates a common vocabulary and standards for clinical data management.

Focus: Centered on patient outcomes and includes recorded healthcare events.

Community: An open community data standard, fostering collaboration and innovation in healthcare data utilization.

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

OOP-CDM is a data model that allows clinical information to be presented in a standardized and reusable way for research