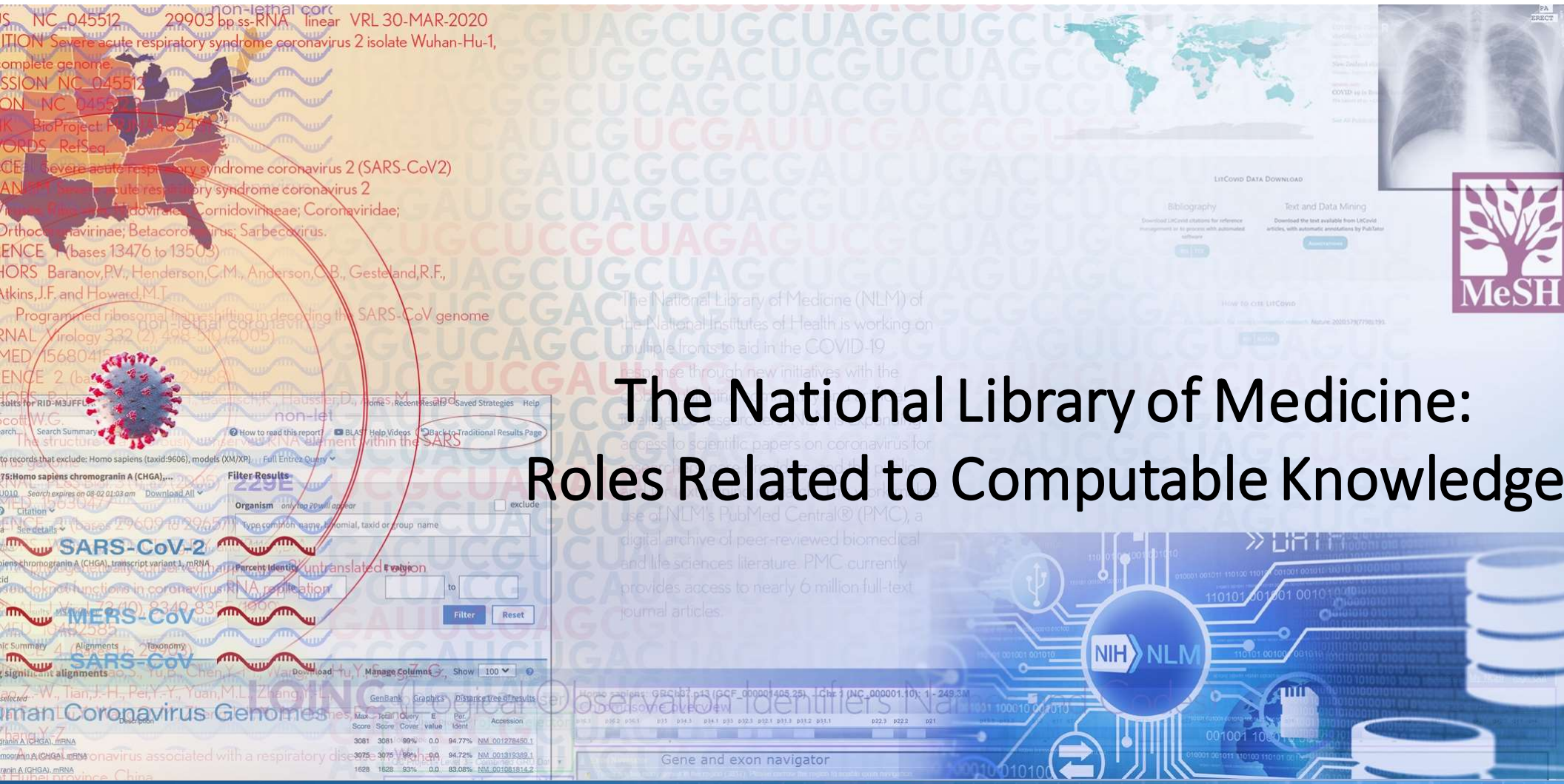
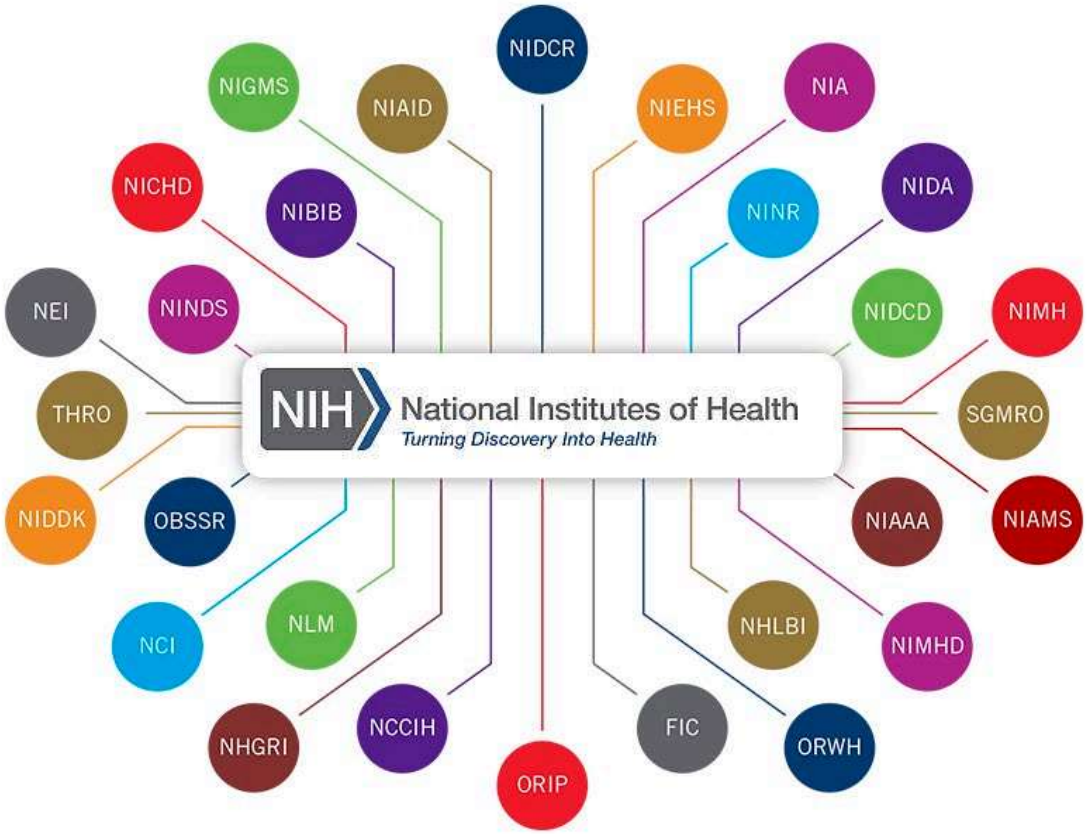
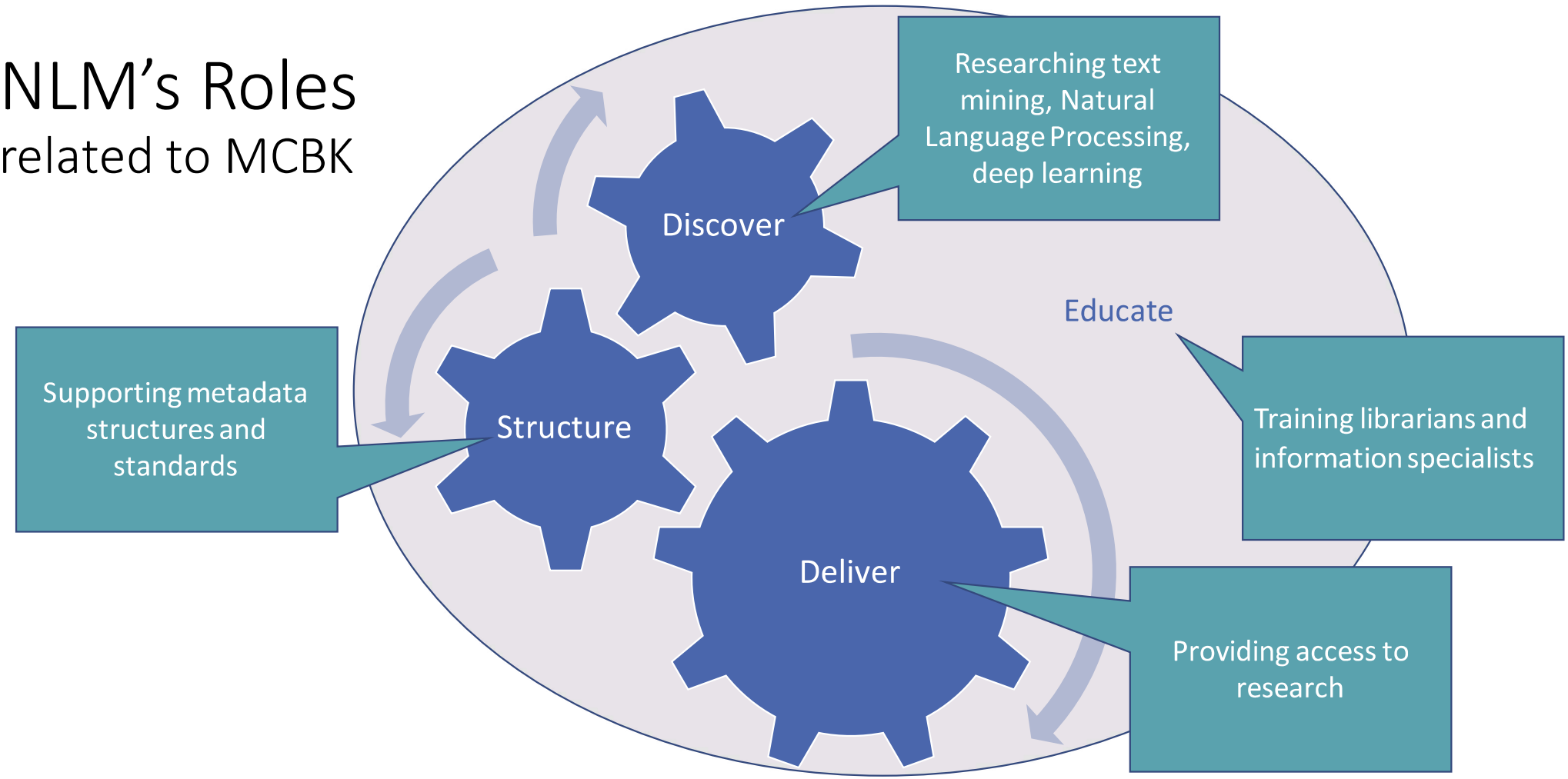


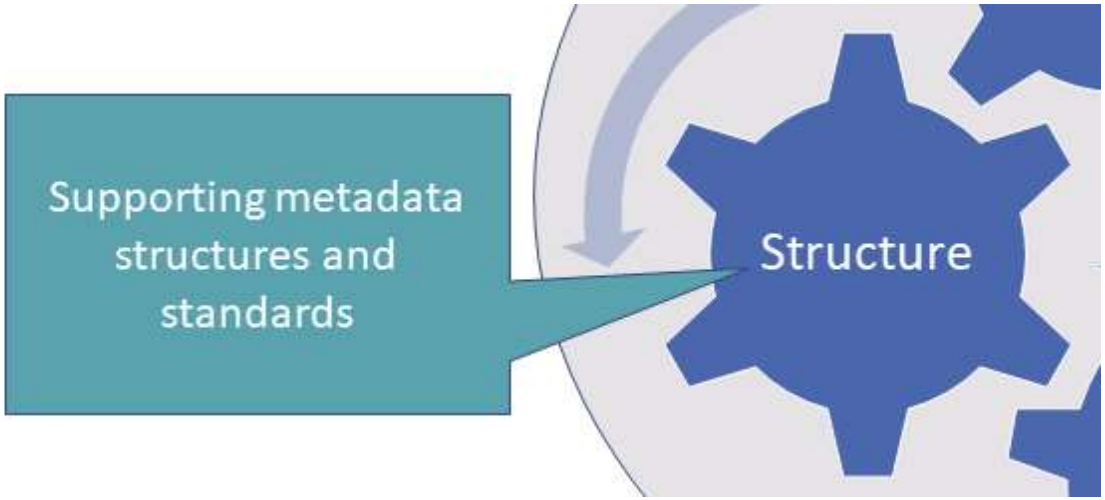
The National Library of Medicine: Roles Related to Computable Knowledge

National Institutes of Health



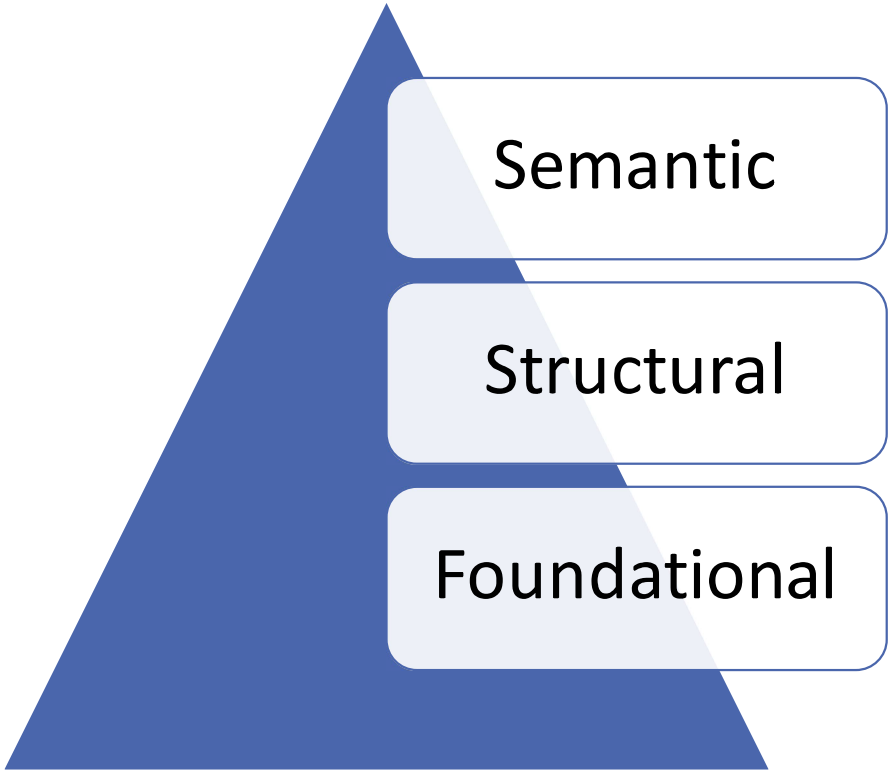
NLM's Roles
related to MCBK





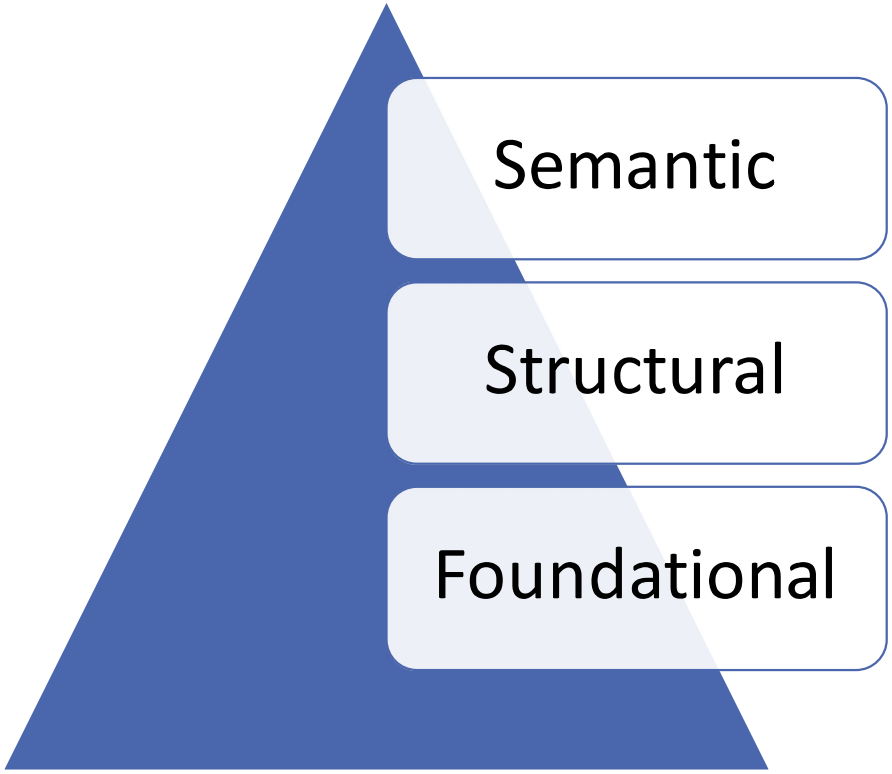
Supporting Health Data and Metadata Standards

Levels of Interoperability



-- Adapted from HIMSS and ONC

Levels of Interoperability



- Vocabulary
  - code sets, ICD-10-CM, SNOMED CT, LOINC, RxNorm, etc.
- 2 structural interoperability projects
- Content exchange
  - HL7 (v2, v3, FHIR, CDA)

Who makes and sets the standards?

Organizations make them

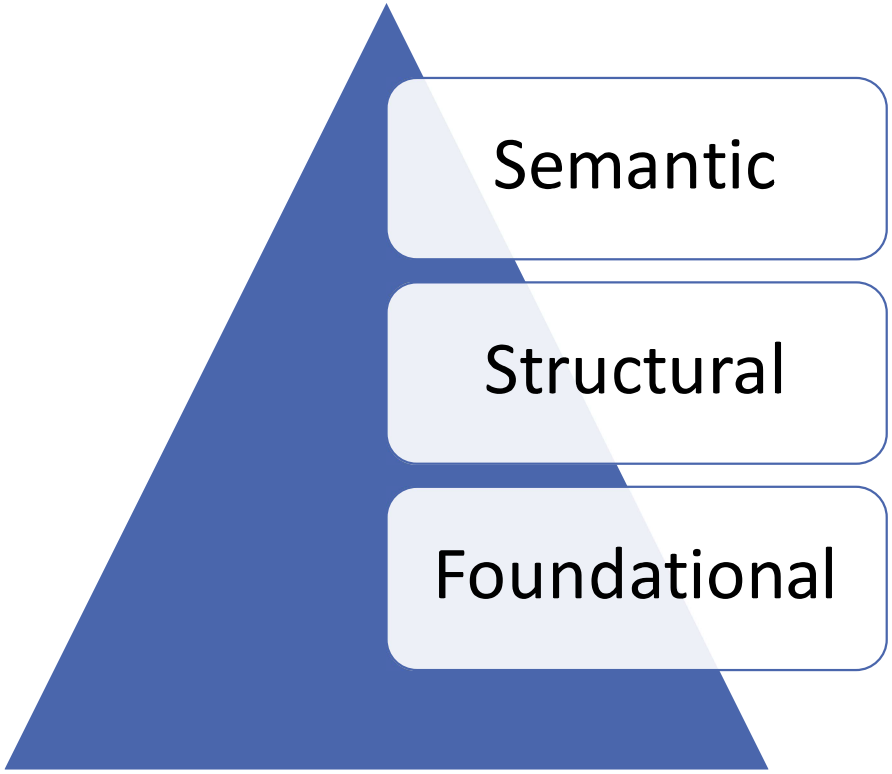
- Health Level Seven International (HL7)
- SNOMED International, WHO, etc.

Federal government sets them

- Congress
- Office of the National Coordinator for Health Information Technology (ONC)
- Centers for Medicare & Medicaid Services (CMS)
- NLM

https://www.nlm.nih.gov/healthit/

Levels of Interoperability



- Vocabulary
  - code sets, ICD-10-CM, SNOMED CT, LOINC, RxNorm, etc.
- 2 structural interoperability projects
- Content exchange
  - HL7 (v2, v3, FHIR, CDA)

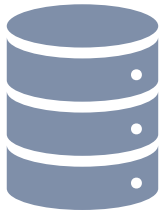
-- Adapted from HIMSS and ONC

Exchange Standards

Coordinated primarily by HL7

Document structure

Clinical Document Architecture (CDA)



Data transfer

Version 2

Version 3

FHIR

Health Level Seven (HL7)

- International non-profit
- Over 1,600 members from corporations, government agencies, universities, and others
- Focused on developing standards for exchanging health information

Clinical Document Architecture (CDA)

Consolidated-Clinical Document Architecture (C-CDA)

- Standard that specifies the structure of "clinical documents"
- Example: "Consultation Note" and "Discharge Summary" have distinct document codes and structures
- Documents can include text, images, sounds, and other multimedia content

HL7 data transfer standards

V2

- Easy to use and understand
- Most widely used data exchange standard
- Works well within a single enterprise but not really interoperable generally

V3

- Not an update of version 2!
- Broader applications but harder to use
- NLM: MedlinePlus Connect

- Fast Healthcare Interoperability Resources:
  a standard for health information exchange

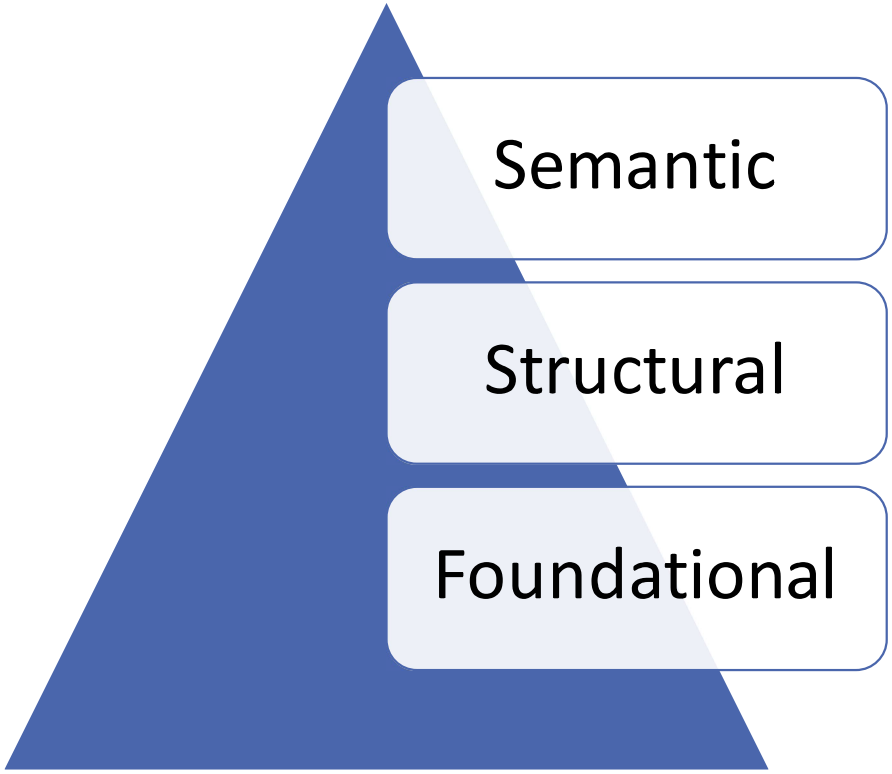
- Uses common Web standards (API, HTTP, HTML, CSS, Atom)
- Describes data formats and elements as **resources** (in JSON, XML or RDF)
- All resources have:
  - A common way to define and represent them
  - A common set of metadata
  - A human readable part

FHIR Resource Example: Patient

```
<Patient xmlns="http://hl7.org/fhir">…

<identifier><!-- 0..* Identifier An identifier for this patient --></identifier>
 <active value="[boolean]"/><!-- 0..1 Whether this patient's record is in active use -->
 <name><!-- 0..* HumanName A name associated with the patient --></name>
 <telecom><!-- 0..* ContactPoint A contact detail for the individual --></telecom>
…
```

Levels of Interoperability



- Vocabulary
  - code sets, ICD-10-CM, SNOMED CT, LOINC, RxNorm, etc.
- 2 structural interoperability projects
- Content exchange
  - HL7 (v2, v3, FHIR, CDA)


An official website of the United States government Here's how you know
National Library of Medicine
Login
NIH CDE Repository
CDEs
Forms
My Boards
About
Help
NIH CDE Repository
Search CDEs
Search
The NIH Common Data Elements (CDE) Repository has been designed to provide access to structured human and machine-readable definitions of data elements that have been recommended or required by NIH Institutes and Centers and other organizations for use in research and for other purposes.
Follow NLM
National Library of Medicine
8600 Rockville Pike, Bethesda, MD 20894
Copyright
FOIA
Privacy
Accessibility
Help
Careers
NLM | NIH | HHS | USA.gov

Common Data Element (CDE):
a fixed representation of a variable including a specific question and a fixed set of permissible answers

CDE Example

- Name: Ethnicity OMB.1997
- Definition: The ethnicity of a person based on the Office of Management and Budget (OMB), Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity (Oct. 20, 1997)

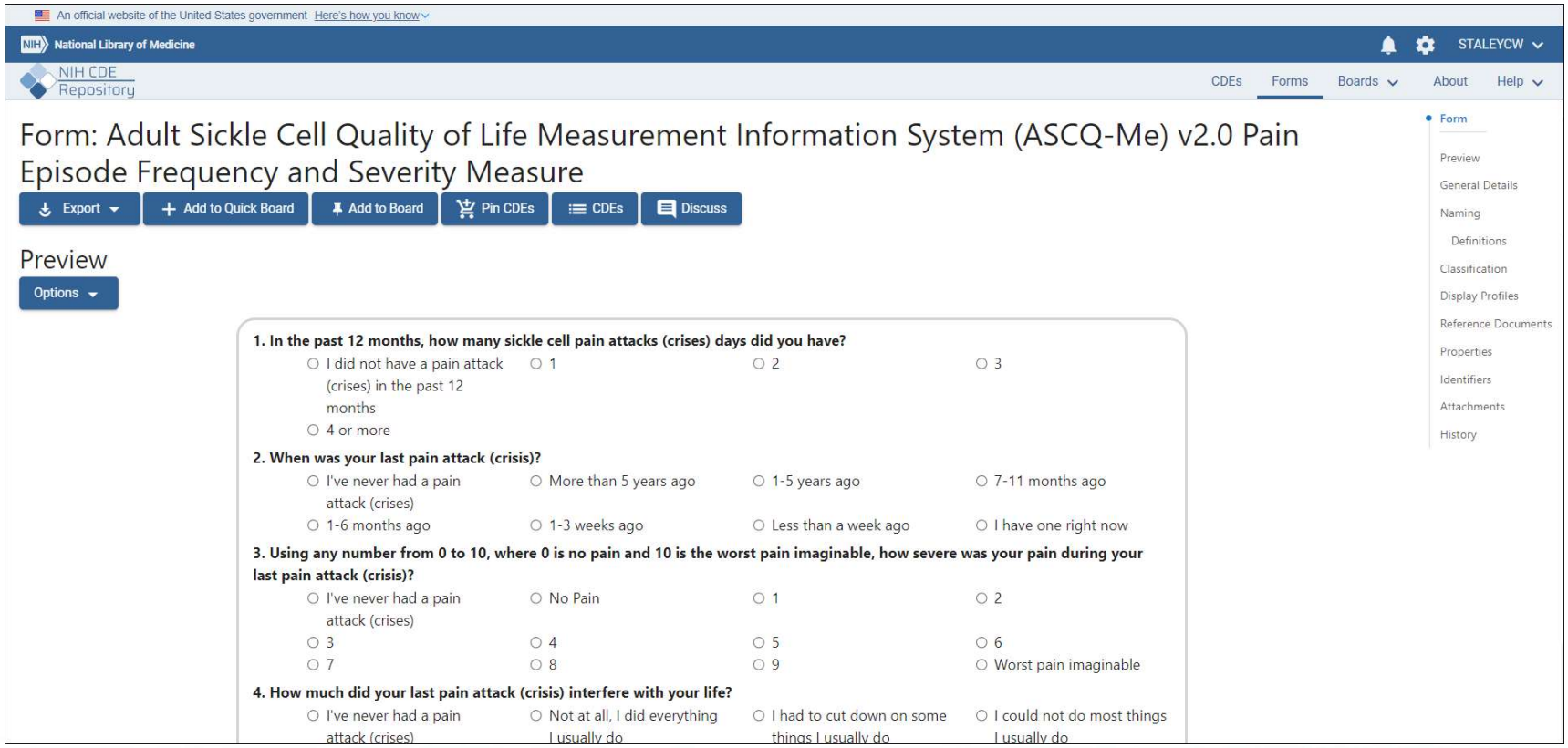
Permissible Value

Value	Code Name	Code	Code System	Code Description
Hispanic or Latino	Hispanic or Latino	2135-2	CDCREC	N/A
Not Hispanic or Latino	Not Hispanic or Latino	2186-5	CDCREC	N/A

Types of CDEs

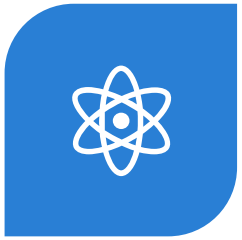
- Demographic
- Disease-agnostic
- Disease-specific

Form with CDEs



Use CDEs to. . .

SAVE
RESOURCES



ADVANCE
SCIENCE

FOLLOW
POLICIES

NIH National Library of Medicine

Common Data Elements: Standardizing Data Collection

https://go.usa.gov/xtCQZ

Training & Outreach > Learning Resources

- Course Home
- Why the Need for Common Data Elements?
- FAIR Data
- Common Data Elements (CDEs)
- NLM and Common Data Elements
- Closing

- Tools and Resources
- How to Navigate This Course

Common Data Elements: Standardizing Data Collection

Have you written a survey or designed a protocol for a research study? If so, you are familiar with the challenges of gathering consistent and standardized data. This course from the National Library of Medicine® is about making data collection and data sharing easier using common data elements (CDEs).

This course includes:

- Scenarios
- Exercises

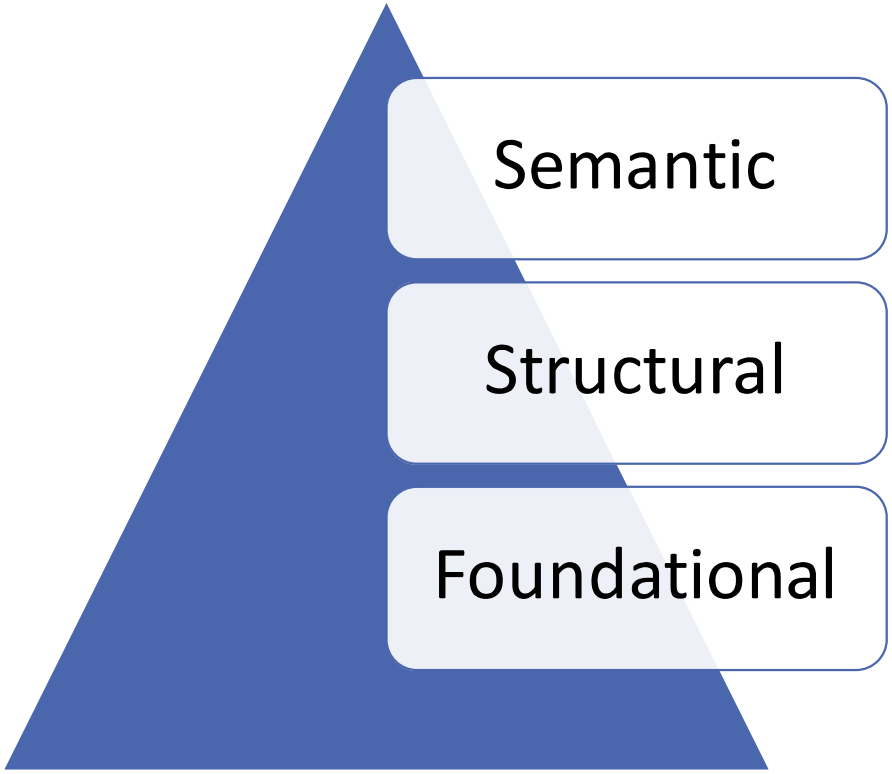
Click for Learning Objectives

Continuing Education Credits

If you are taking this course for Medical Library Association credits, please **register and complete the class** through the

Journal Article Tag Suite (JATS)

Levels of Interoperability



- Vocabulary
  - code sets, ICD-10-CM, SNOMED CT, LOINC, RxNorm, etc.
- 2 structural interoperability projects
- Content exchange
  - HL7 (v2, v3, FHIR, CDA)

What is a medical terminology?

A set of specialized terms that facilitate precise communication by minimizing or eliminating ambiguity.

Features of a terminology

- Unique Identifier (often a code)
- Official Name
- Synonyms (in many cases)

Examples

- CPT, ICD-10, SNOMED-CT, LOINC, RxNorm

Why do we need medical terminologies?

- To facilitate precise and clear communication by minimizing or eliminating ambiguity.
- To cross-walk or map between other terminologies
  - Examples: UMLS and RxNorm

Some terminology problems for biomedical researchers:

- There are many biomedical terminologies in use.
- Terminologies come in a variety of formats and a variety of data models.
- Terminologies do not necessarily link to each other.
- Terminologies are not always freely available.

Solution: Unify the medical language in one system.

The National Library of Medicine has accomplished this with the
Unified Medical Language System (UMLS)

What is the UMLS?

Unified

Medical

Language

System

The UMLS provides:

Access to Terminology Data

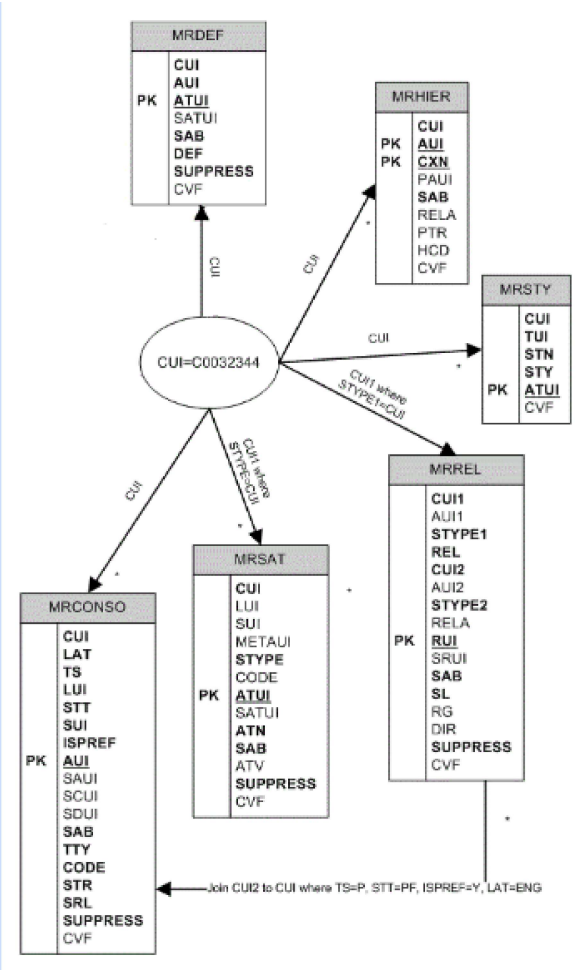
A Common Data Model for Terminologies

Interoperability through Synonymy

The UMLS:
A Common Data Model

Problem: Terminology data is represented in a variety of formats according to a variety of data models.

Solution: The UMLS represents all terminology data according to a standard data model.



Interoperability through Synonymy

Problem: Terminologies are not always linked to each other.

Solution: UMLS asserts synonymy by grouping names from different terminologies into concepts. This can be used as a starting point for crosswalking from one terminology to another.

A UMLS Concept: Addison Disease

Name	Source Terminology	Code	Atom Identifier	Concept Identifier
Addison Disease	MeSH	D000224	A6954527	C0001403
Addison's disease	ICD-10-CM	E27.1	A17799651	C0001403
Addison's Disease	MeSH	D000224	A26597849	C0001403
Primary adrenal insufficiency	MedDRA	S2164152	A2018590	C0001403
Primary adrenocortical insufficiency	ICD-10-CM	E27.1	A17786892	C0001403
Insufficiency, Primary Adrenocortical	MeSH	D000224	A6970512	C0001403
Primary adrenocortical insufficiency (disorder)	SNOMED CT	373662000	A3644299	C0001403
Primary hypoadrenalism	MedDRA	S0718109	A25720215	C0001403
Primary hypoadrenalism	SNOMED CT	373662000	A3060485	C0001403
Enfermedad de Addison	MeSH Spanish	D000224	A9175691	C0001403

UMLS Case Study: Cleveland Clinic

Problem: How do you make patient data ready for researchers to discover and use?

Solution: Integrate identifiers, hierarchies, and relationships from UMLS.

Milinovich A, Kattan MW. Extracting and utilizing electronic health data from Epic for research. Ann Transl Med. 2018;6(3):42. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5879514/

UMLS Case Study: Cleveland Clinic

- Cleveland Clinic converts electronic health record data from 185 tables to 18 research-ready tables annotated with UMLS identifiers. Data is updated on a weekly basis.
- **Result:** "Cleveland Clinic can do live population exploration as well as produce datasets for analysis faster than it takes most organizations to simply identify their base population."

Milinovich A, Kattan MW. Extracting and utilizing electronic health data from Epic for research. Ann Transl Med. 2018;6(3):42. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5879514/

Identifying Patient Populations

Diabetic patients on a GLP-1 medication with an HbA1c >10

2,666 different "diabetic" diagnoses

Get all descendants of:

Diabetes Mellitus (UMLS CUI: C0011847)

- Acidosis due to type 1 diabetes mellitus
- Acidosis due to type 2 diabetes mellitus
- Acute complication with diabetes mellitus
- Diabetic dyslipidemia associated with type 2 diabetes mellitus
- Diabetic hyperosmolar non-ketotic state
- Diabetic lumbosacral radiculoplexus neuropathy
- Diabetic mastopathy
- Diabetic severe hyperglycemia
- Disorder of nerve co-occurrent and due to type 1 diabetes mellitus
- Dyslipidemia due to type 1 diabetes mellitus
- Hyperglycemia due to type 1 diabetes mellitus
- Hyperglycemia due to type 2 diabetes mellitus
- Hyperglycemic crisis in diabetes mellitus
- Hyperlipidemia due to type 1 diabetes mellitus
- Hyperlipidemia due to type 2 diabetes mellitus
- Hyperosmolality due to uncontrolled type 1 diabetes mellitus
- Hyperosmolar coma associated with diabetes mellitus
- Hyperosmolarity co-occurrent and due to drug induced diabetes mellitus
- Hypoglycemia due to type 1 diabetes mellitus
- Hypoglycemia due to type 2 diabetes mellitus
- Hypoglycemic coma in diabetes mellitus
- Hypoglycemic state in diabetes
- Ketoacidosis due to secondary diabetes mellitus
- Lactic acidosis with diabetes mellitus
- Malnutrition-related diabetes mellitus with multiple complications
- Metabolic acidosis with diabetes mellitus
- Mixed hyperlipidemia due to type 1 diabetes mellitus
- Mixed hyperlipidemia due to type 2 diabetes mellitus
- Multiple complications due to diabetes mellitus
- Peripheral neuropathy due to type 1 diabetes mellitus
- Radiculoplexoneuropathy due to diabetes mellitus
- Type 1 diabetes mellitus with hyperosmolar coma
- Type 2 diabetes mellitus with hyperosmolar coma
- Diabetic acute painful polyneuropathy
- Coma associated with malnutrition-related diabetes mellitus
- Diabetic coma with ketoacidosis
- Hyperosmolar coma associated with diabetes mellitus
- Hypoglycemic coma co-occurrent and due to diabetes mellitus type II
- ETC

Identifying Patient Populations

Diabetic patients on a GLP-1 medication with an HbA1c >10

69 different GLP-1 medications

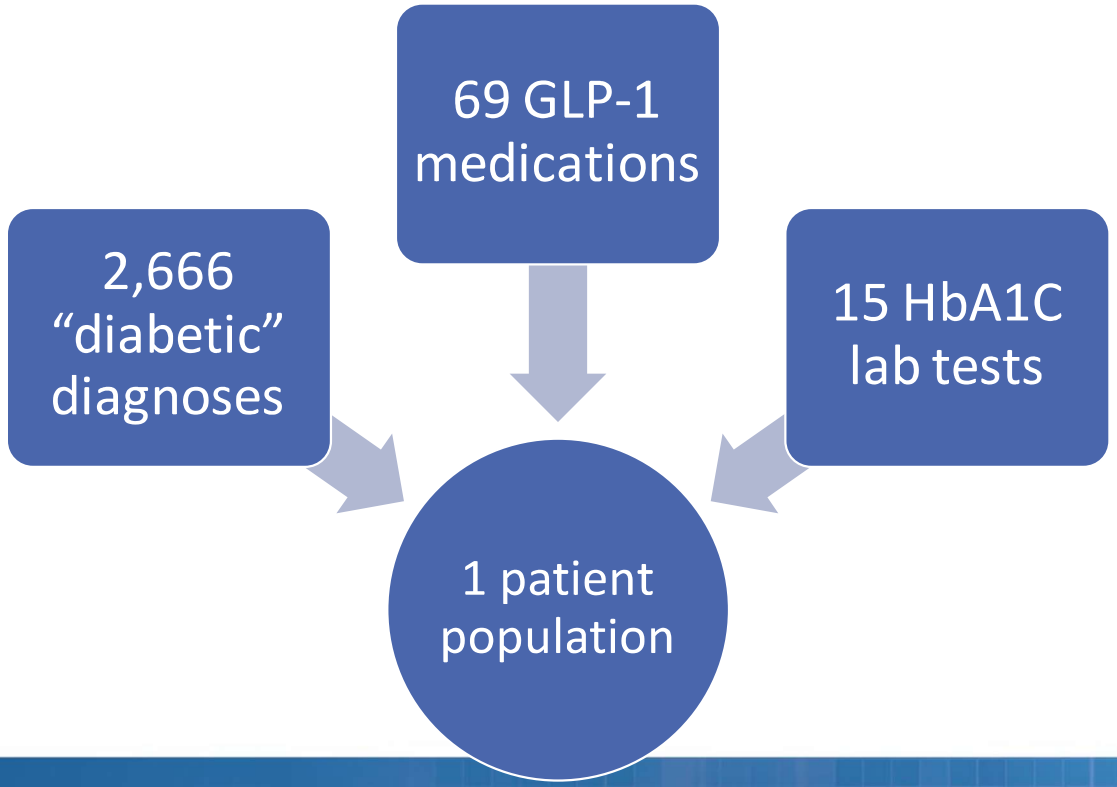
Get all medications associated with a drug class:
Glucagon-like Peptide-1 (GLP-1) Agonists [MoA] (UMLS CUI: C2916791)

15 different HbA1c lab tests

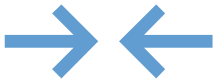
Map lab tests in EHR to:
Hemoglobin A1c/Hemoglobin.total:Mass Fraction:Point in time:Whole blood:Quantitative (UMLS CUI: C0366781)

Identifying Patient Populations

Diabetic patients on a GLP-1 medication with an HbA1c >10



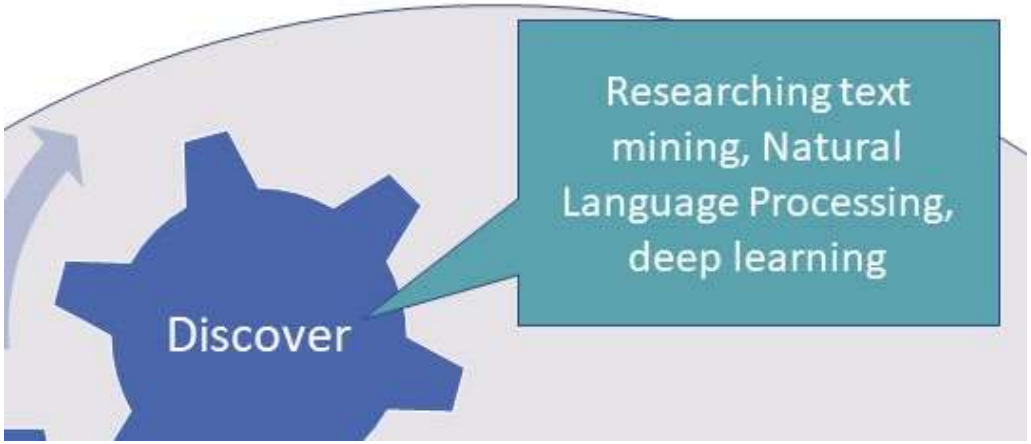
Synonymy Applied



Crosswalking between Terminologies

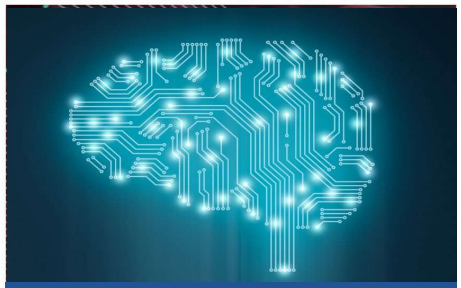
Search and Retrieval

Identifying Meaning in Text



Researching Methods in Informatics

NLM Research Portfolio



Machine Learning Applied to Electronic Health Record Data

Health Information Standards and Discovery

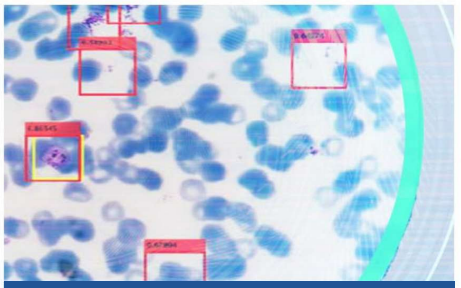
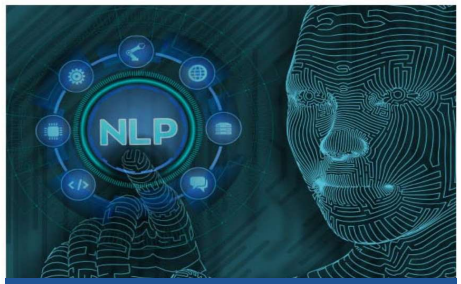


Image Processing



Natural Language Processing

Statistical Methods

Networks, Gene Regulation and Chromatin

Identifying Meaning in Text (Natural Language Processing)

Problem: Medical data in the real world is often unstructured. Examples: clinical notes or the biomedical literature / abstracts.

Solution: The UMLS can help identify meaning in text in combination with various tools.

Why Identify Meaning in Text?

- Improve search and retrieval by annotating records in a research database
- Find co-occurrences of concepts in text
- Annotate clinical text on the fly
- Identify a patient population

Text Processing Tools

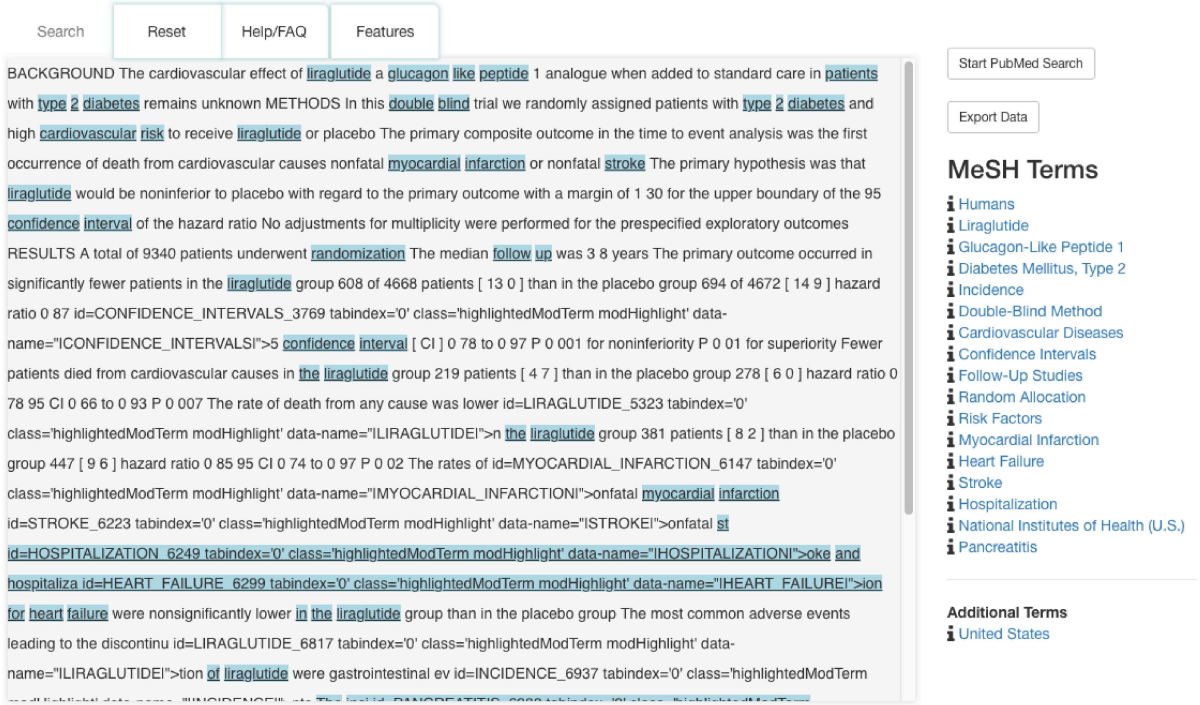
NLM Tools

- **MetaMap** – A tool for recognizing UMLS concepts in text
- **Medical Text Indexer** – A tool for automated indexing of the biomedical literature
- **MeSH on Demand** – Uses the Medical Text Indexer to extract MeSH terms via UMLS Concepts.
- **SemRep –** A tool for extracting assertions from sentences in biomedical text

Third-Party Tools that include UMLS

- **Apache cTakes –** "…a natural language processing system for extraction of information from electronic medical record clinical free-text… Originally developed by the Mayo Clinic…"
- **CLAMP (Clinical Language Annotation, Modeling, and Processing Toolkit) –** "…a comprehensive clinical Natural Language Processing (NLP) software that enables recognition and automatic encoding of clinical information in narrative patient reports." (University of Texas Health Science Center at Houston)
- **And many many more.**

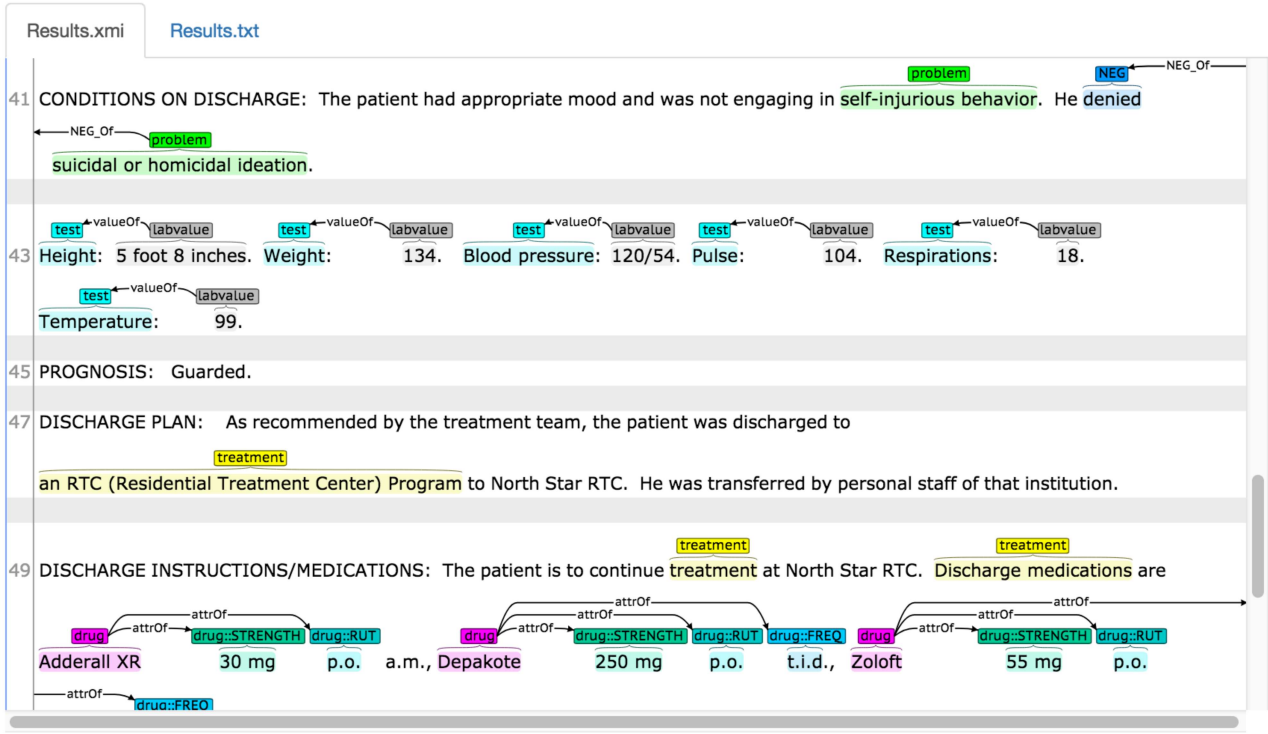
NLM MeSH on Demand



https://meshb.nlm.nih.gov/MeSHonDemand

CLAMP

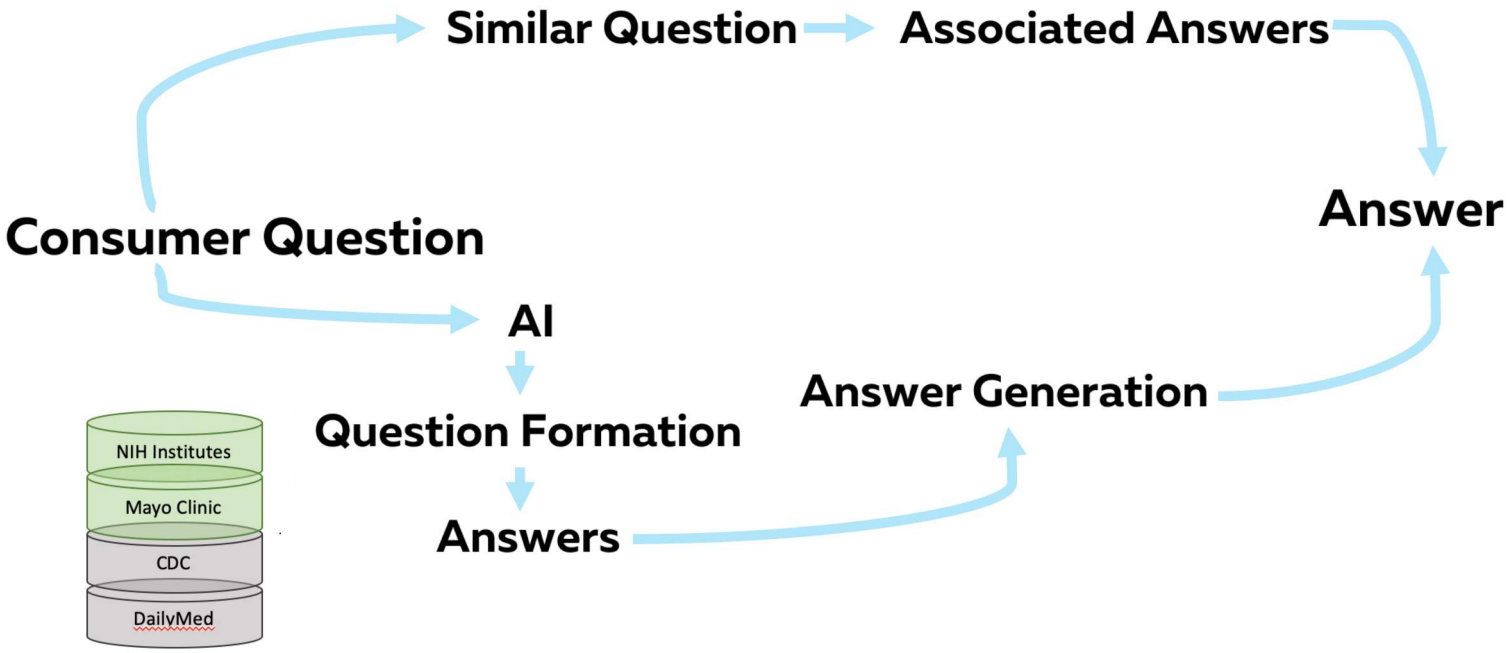
(Clinical Language Annotation, Modeling, and Processing Toolkit)



https://clamp.uth.edu/clampdemo.php

AI-Enabled Access to Reliable Health Information

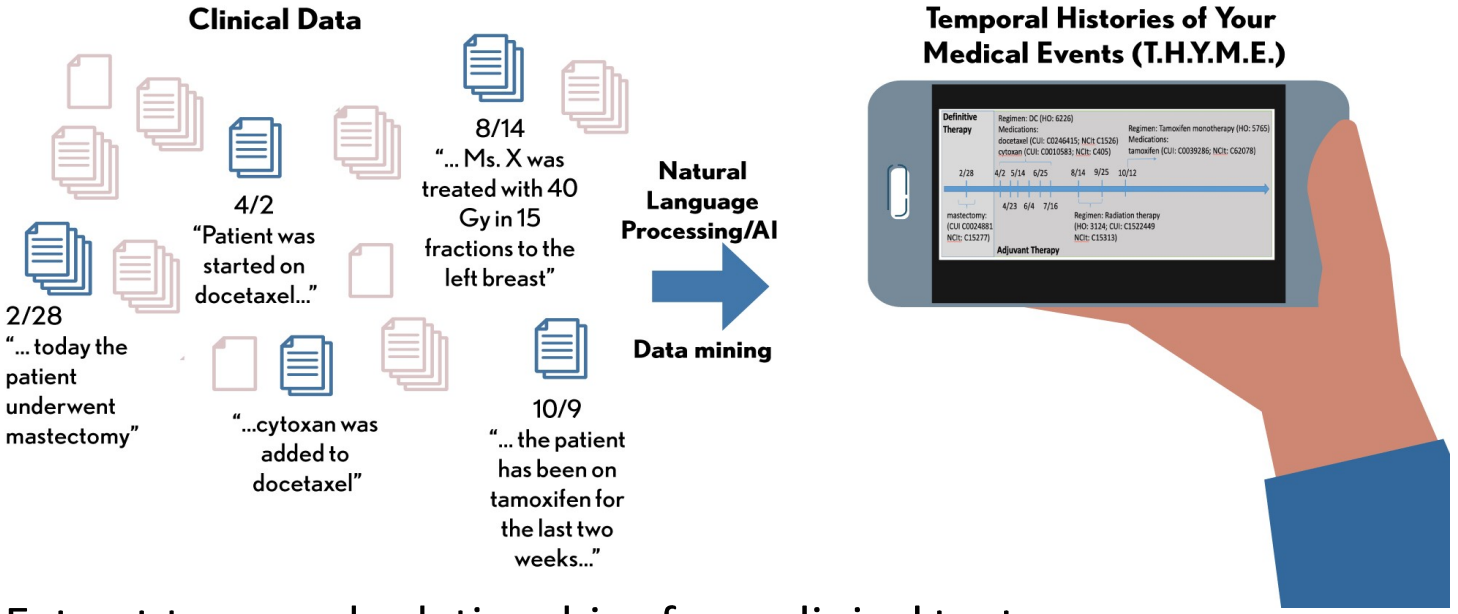
NLM's Consumer Health Question Answering Program



Consumer Health Information and Question Answering: Helping consumers find answers to their health-related information needs. Demner-Fushman, D., Mrabet, Y., and Ben Abacha, A. JAMIA 2020.

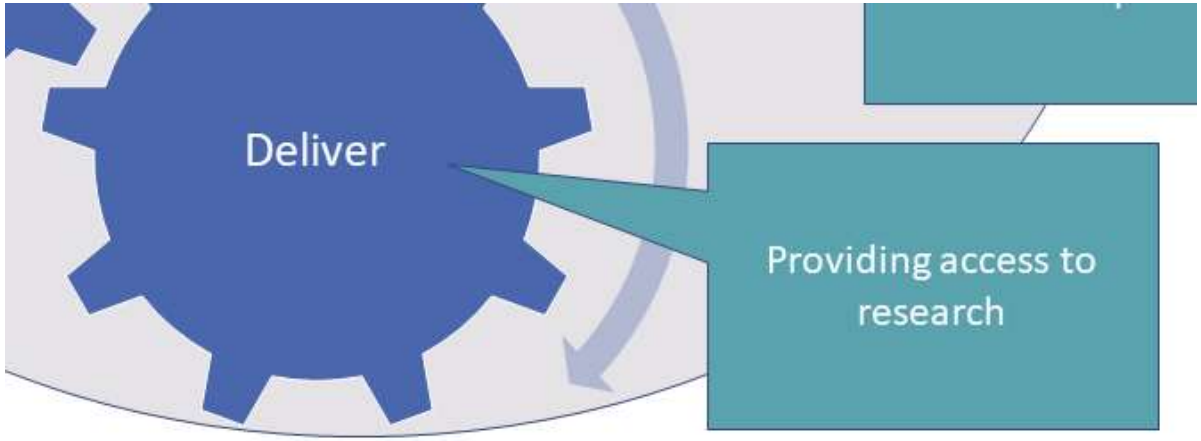
Discovery: Natural Language Processing

Temporal Histories of Your Medical Events (T.H.Y.M.E) Project



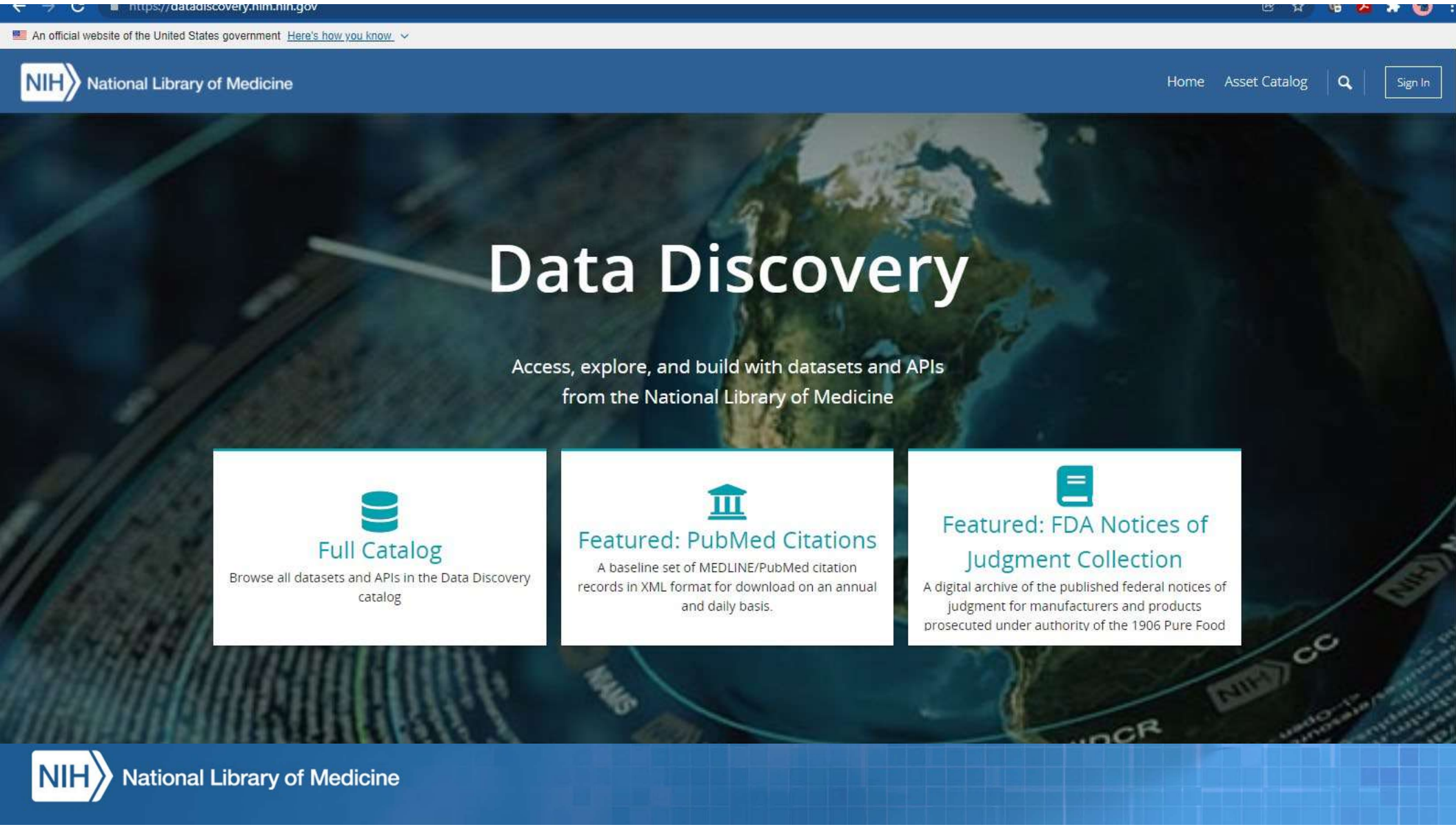
Extract temporal relationships from clinical text

Guergana Savova, PhD [U54LM008748, R01LM010090]



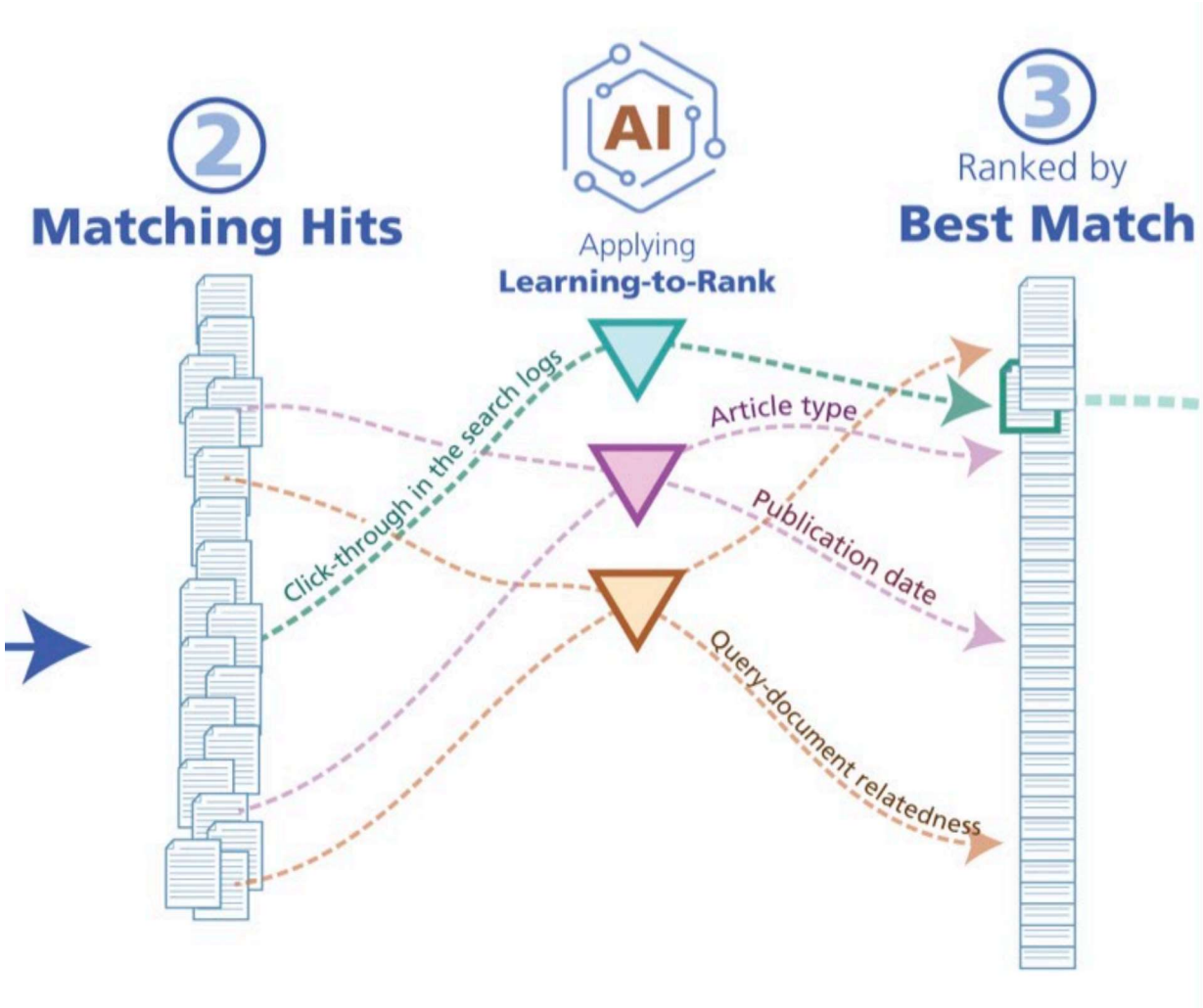
Providing Access to Research

Research Data Services


https://datadiscovery.nlm.nih.gov
An official website of the United States government Here's how you know
NIH National Library of Medicine
Home
Asset Catalog
Sign In
Data Discovery
Access, explore, and build with datasets and APIs
from the National Library of Medicine
Full Catalog
Browse all datasets and APIs in the Data Discovery catalog
Featured: PubMed Citations
A baseline set of MEDLINE/PubMed citation records in XML format for download on an annual and daily basis.
Featured: FDA Notices of Judgment Collection
A digital archive of the published federal notices of judgment for manufacturers and products prosecuted under authority of the 1906 Pure Food
NIH National Library of Medicine


PubMed


PMC



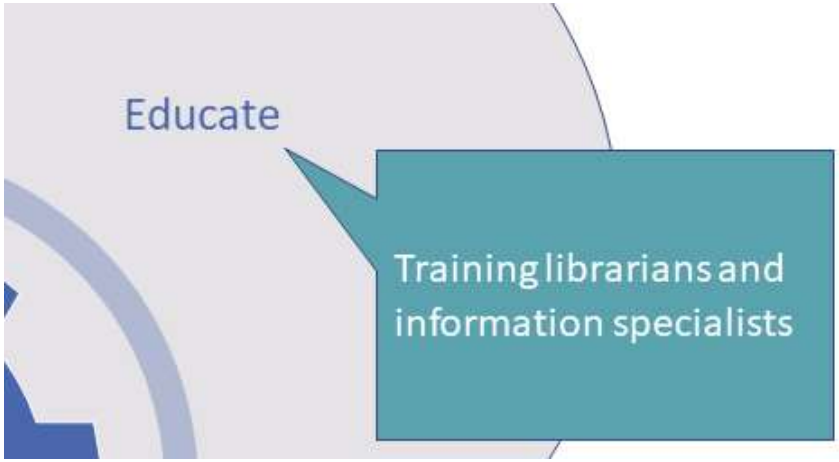
PubMed's Best Match Sorting

A **24-hour response time** for MeSH-indexed citations to appear in PubMed

Expanded chemical and gene curation by subject matter experts

Continuous improvement of the automated indexing algorithm

MEDLINE Automated Indexing



Training Information Specialists

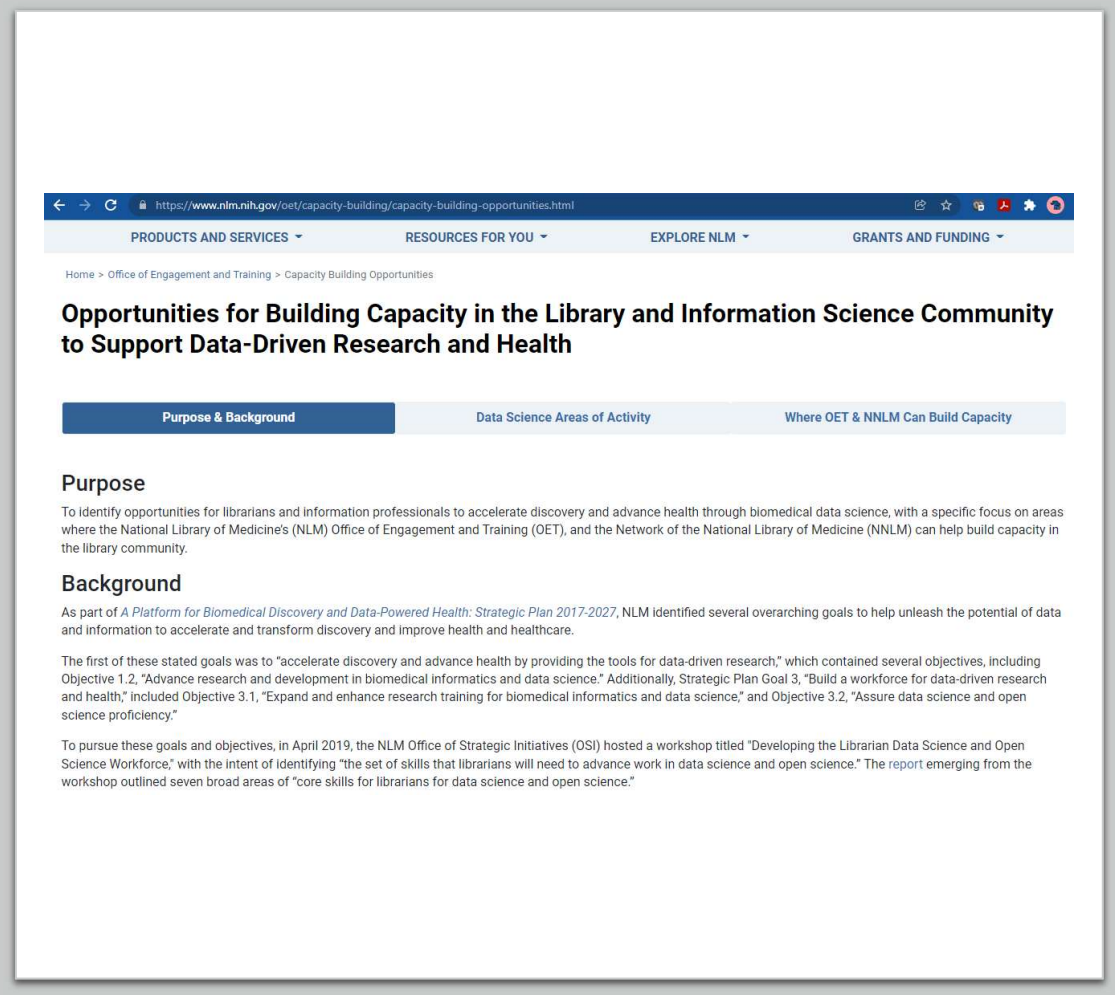
Facing New Challenges

Establishing the informatics- informed workforce

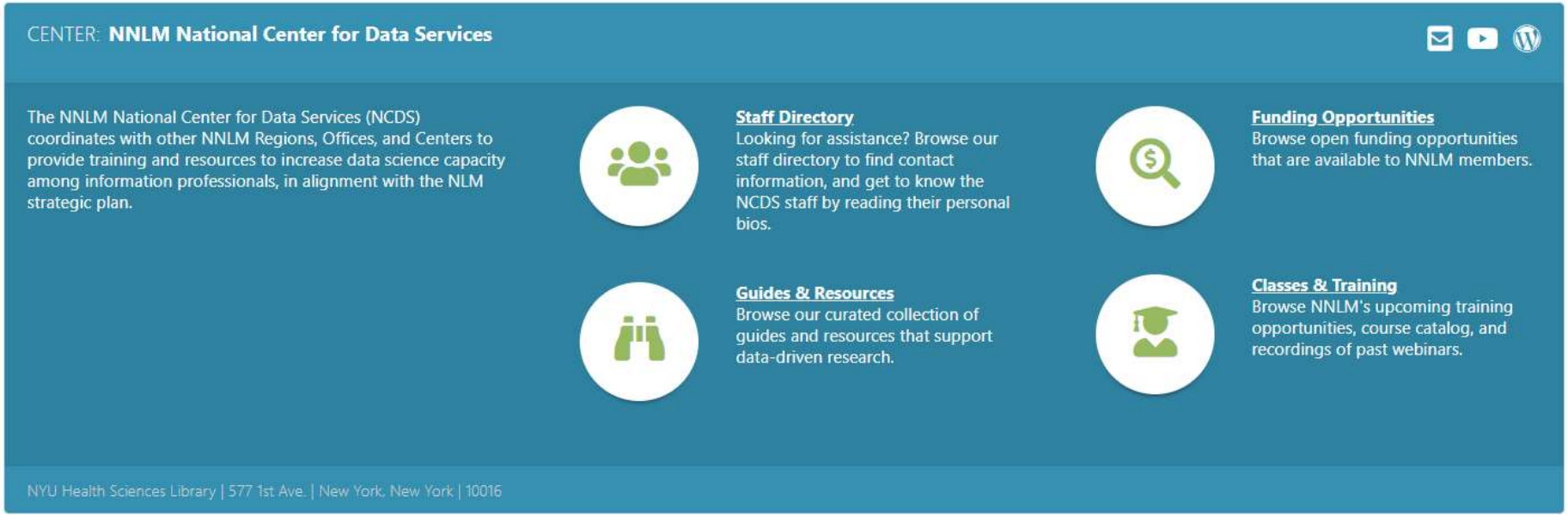
NIH National Library of Medicine

Building Capacity in Data Science

https://go.usa.gov/xt3vw

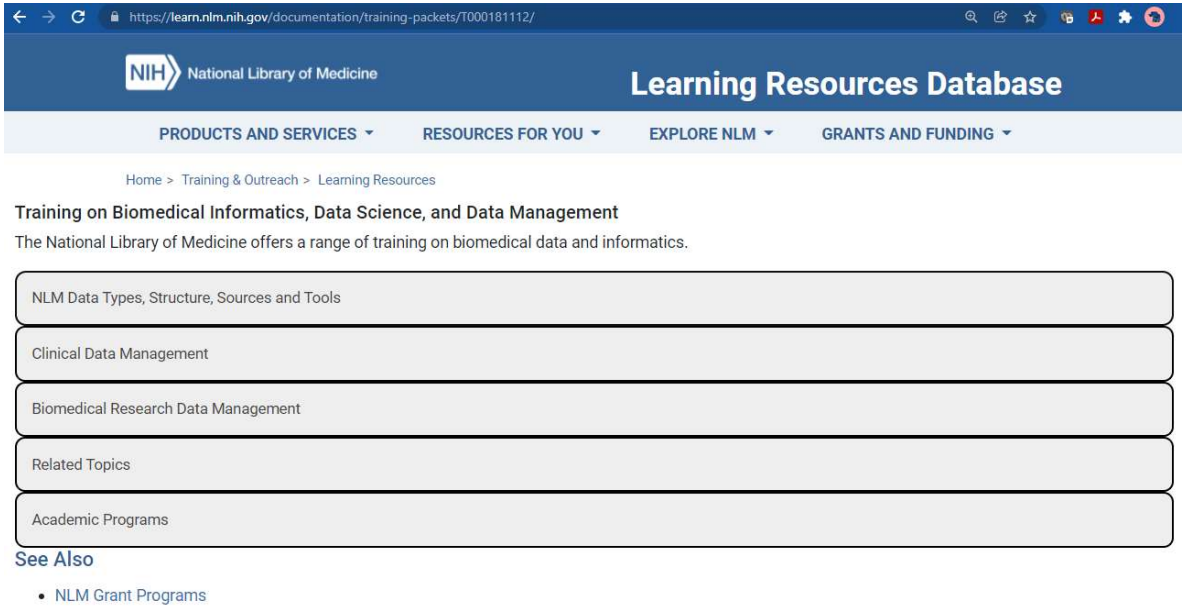


NNLM National Center for Data Services



Kick-off tomorrow at 1pm ET: https://go.usa.gov/xtCEa

Training for Information Specialists



https://go.usa.gov/xt3p8

Connect with Us

NLM Homepage
nlm.nih.gov

NLM Technical Bulletin
https://go.usa.gov/xt3vZ

@NLM_news