Computing for Medicine

Google Classroom Code: dnd5qkt5

Monsoon 2025

Lecture 4

Semantic Interoperability

What is Interoperability?

Interoperability is ability of two or more systems or components to exchange information and to use the information that has been exchanged (IEEE 1990)

Levels of Interoperability

- Technical - exchange of information
- Semantic - capability of the recipient to “use” that information
- Process - actual use of information
- Human - clinical interoperability

Perspective | Open Access | Published: 20 August 2019

Why digital medicine depends on interoperability

Moritz Lehne ✉, Julian Sass, Andrea Essenwanger, Josef Schepers & Sylvia Thun

npj Digital Medicine **2**, Article number: 79 (2019) | Cite this article

Abstract

Digital data are anticipated to transform medicine. However, most of today's medical data lack interoperability: hidden in isolated databases, incompatible systems and proprietary software, the data are difficult to exchange, analyze, and interpret. This slows down medical progress, as technologies that rely on these data – artificial intelligence, big data or mobile applications – cannot be used to their full potential. In this article, we argue that interoperability is a prerequisite for the digital innovations envisioned for future medicine. We focus on four areas where interoperable data and IT systems are particularly important: (1) artificial intelligence and big data; (2) medical communication; (3) research; and (4) international cooperation. We discuss how interoperability can facilitate digital transformation in these areas to improve the health and well-being of patients worldwide.

https://www.nature.com/articles/s41746-019-0158-1

HSR Health Services Research

RESEARCH ARTICLE

Sharing information electronically with other hospitals is associated with increased sharing of patients

Jordan Everson PhD MPP ✉ Julia Adler-Milstein PhD

First published: 12 November 2019 | https://doi.org/10.1111/1475-6773.13240 | Citations: 8

Read the full text > PDF TOOLS SHARE

Abstract

Objective

One potential benefit of greater electronic health information exchange is a reduction in the effort required for patients to switch between providers. We therefore assessed whether hospital participation in health information organizations (HIOs) led to increased patient sharing.

Data Sources

Secondary data from 2010 to 2016.

Study Design

https://onlinelibrary.wiley.com/doi/10.1111/1475-6773.13240

Semantics: Interoperating with text

```
> dat[1,2]
[1] "A 23-year-old white female presents with complaint of allergies."
> dat[1,5]
[1] "SUBJECTIVE:,  This 23-year-old white female presents with complaint of allergies.  She used to have allergies when s
he lived in Seattle but she thinks they are worse here.  In the past, she has tried Claritin, and Zyrtec.  Both worked fo
r short time but then seemed to lose effectiveness.  She has used Allegra also.  She used that last summer and she began
using it again two weeks ago.  It does not appear to be working very well.  She has used over-the-counter sprays but no p
rescription nasal sprays.  She does have asthma but doest not require daily medication for this and does not think it is
flaring up.,MEDICATIONS: , Her only medication currently is Ortho Tri-Cyclen and the Allegra.,ALLERGIES: , She has no kno
wn medicine allergies.,OBJECTIVE:,Vitals:  Weight was 130 pounds and blood pressure 124/78.,HEENT:  Her throat was mildly
erythematous without exudate.  Nasal mucosa was erythematous and swollen.  Only clear drainage was seen.  TMs were clear.
,Neck:  Supple without adenopathy.,Lungs:  Clear.,ASSESSMENT:,  Allergic rhinitis.,PLAN:,1.  She will try Zyrtec instead
of Allegra again.  Another option will be to use loratadine.  She does not think she has prescription coverage so that mi
ght be cheaper.,2.  Samples of Nasonex two sprays in each nostril given for three weeks.  A prescription was written as w
ell."
```

Table 1. Types of Terminologies Used for Computational Analysis.*

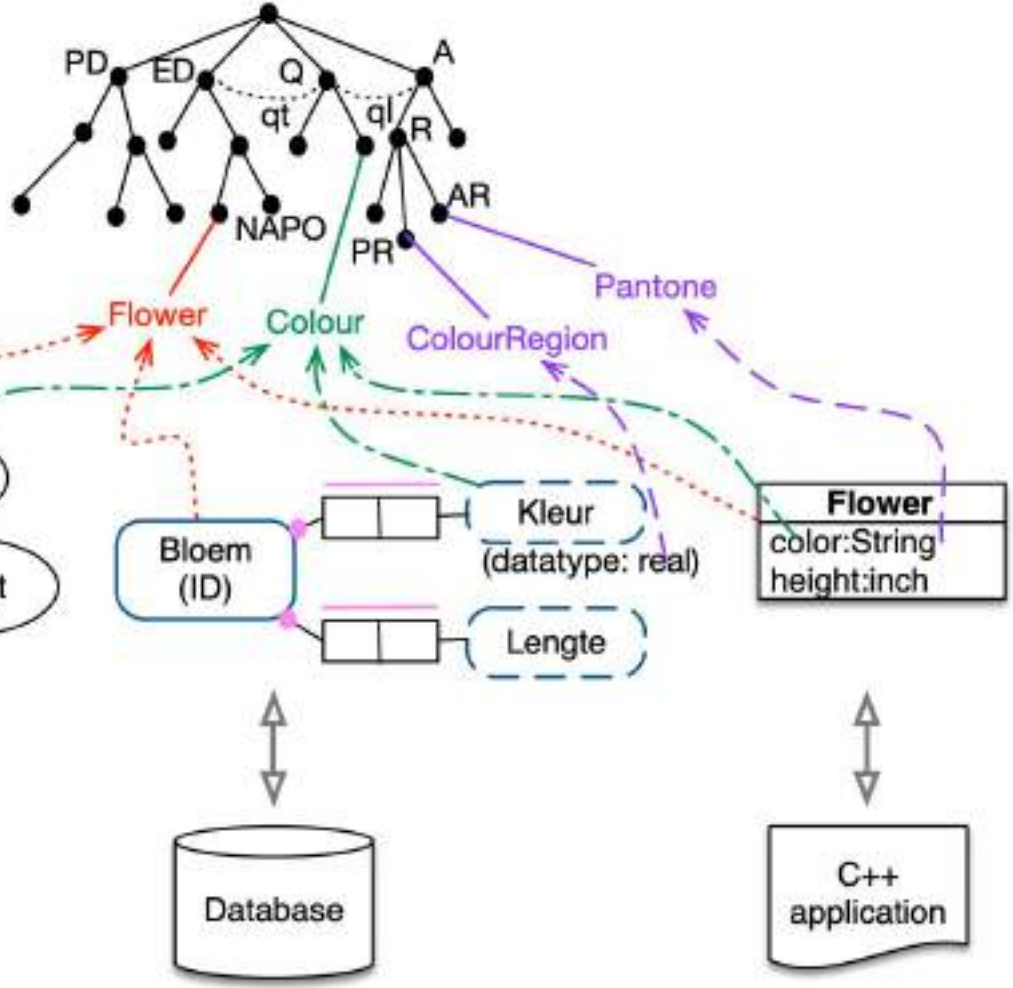
Term	Definition
Index	List of relevant terms pulled directly from a body of unstructured or semistructured text. An index is produced to improve the speed and relevance of search results.
Terminology	Set of preferred or official terms in a domain. A terminology may be a systematic nomenclature supported by a centralized body or as simple as the common usage that arises in a specific community of practice.
Thesaurus	Terminology that clusters synonyms and plesionyms (near synonyms) into categories.
Controlled vocabulary	Set of preferred terms created specifically for a domain or body of text.
Classification	Controlled vocabulary that is intended to comprehensively describe a topic or domain from a conceptual perspective and is not developed solely from a text corpus that it is meant to describe.
Statistical classification	Classification in which all concepts are mutually exclusive to avoid counting anything twice. This is typically achieved with the use of a monohierarchy, in which each concept has one and only one parent, such as the *International Classification of Diseases* (ICD).[7] A statistical classification is exhaustive because it includes residual categories such as "unspecified" or "not elsewhere classified."
Ontology	Controlled terminology invoking formal semantic relationships between and among concepts, manifested as a type of description logic, which is a subset of first-order predicate logic, chosen to accommodate computational tractability.[8] A common example is OWL (Web Ontology Language; www.w3.org/OWL/).

Table 2. Standards, Terminologies, and Ontologies Widely Used in Clinical Medicine.*

Type of Data	Ontology	Example of Term
Diagnoses	Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) ICD Orphanet Rare Disease Ontology (ORDO) National Cancer Institute Thesaurus (NCIT)	Triple-negative breast carcinoma (NCIT:C71732)
Phenotypic abnormalities	Human Phenotype Ontology (HPO)	Bronchopulmonary sequestration (HP:0010960)
Medications	RxNorm DrugBank ChEMBL	Panobinostat (CHEMBL483254)
Adverse reactions	Ontology of Adverse Events (OAE)	Injection-site induration (OAE:0000323)
Procedures	Medical Dictionary for Regulatory Activities (MedDRA)	Cardiac aneurysm repair (MEDDRA/10007514)
Laboratory examinations	Logical Observation Identifiers Names and Codes (LOINC)	Creatinine in serum or plasma (LOINC:2160-0)
Imaging data	Digital Imaging and Communications in Medicine (DICOM) RadLex	Periosteal cortical thinning (RID45761)

What are Ontologies?

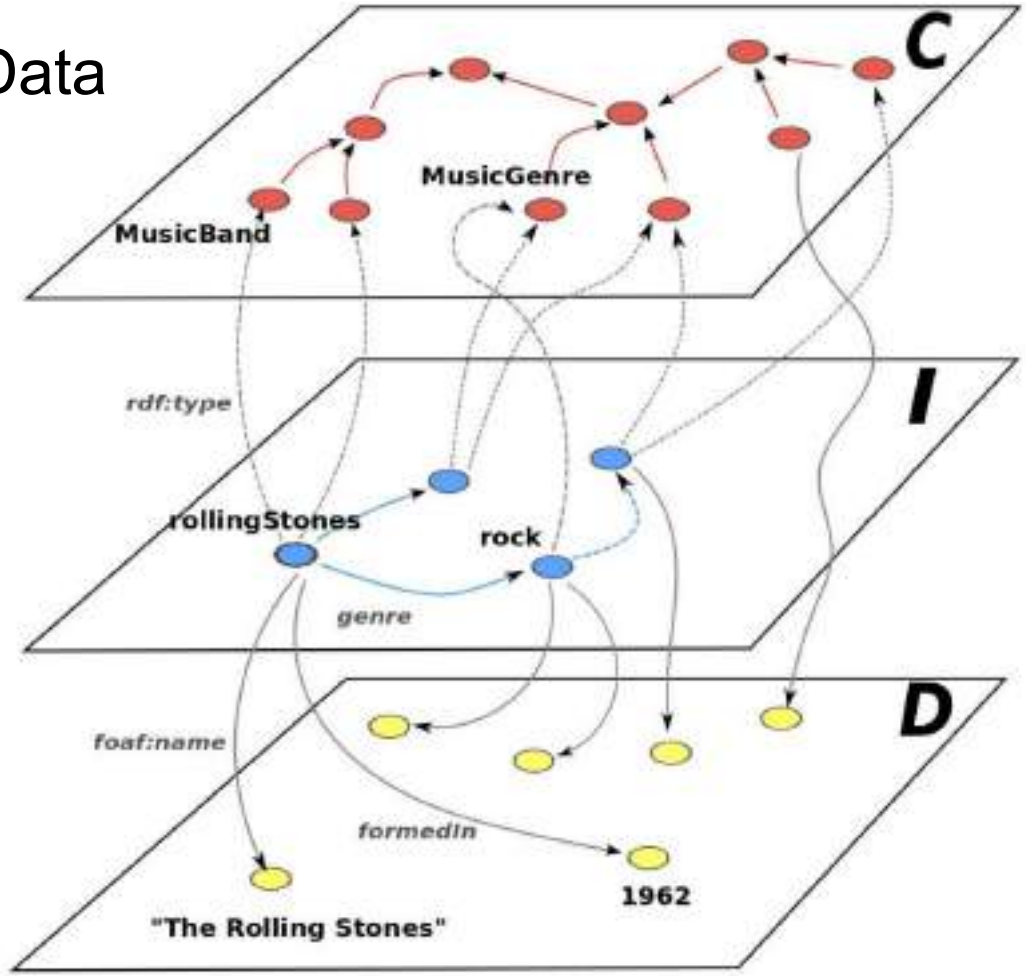
Ontology
provides the common vocabulary and constraints that hold across the applications

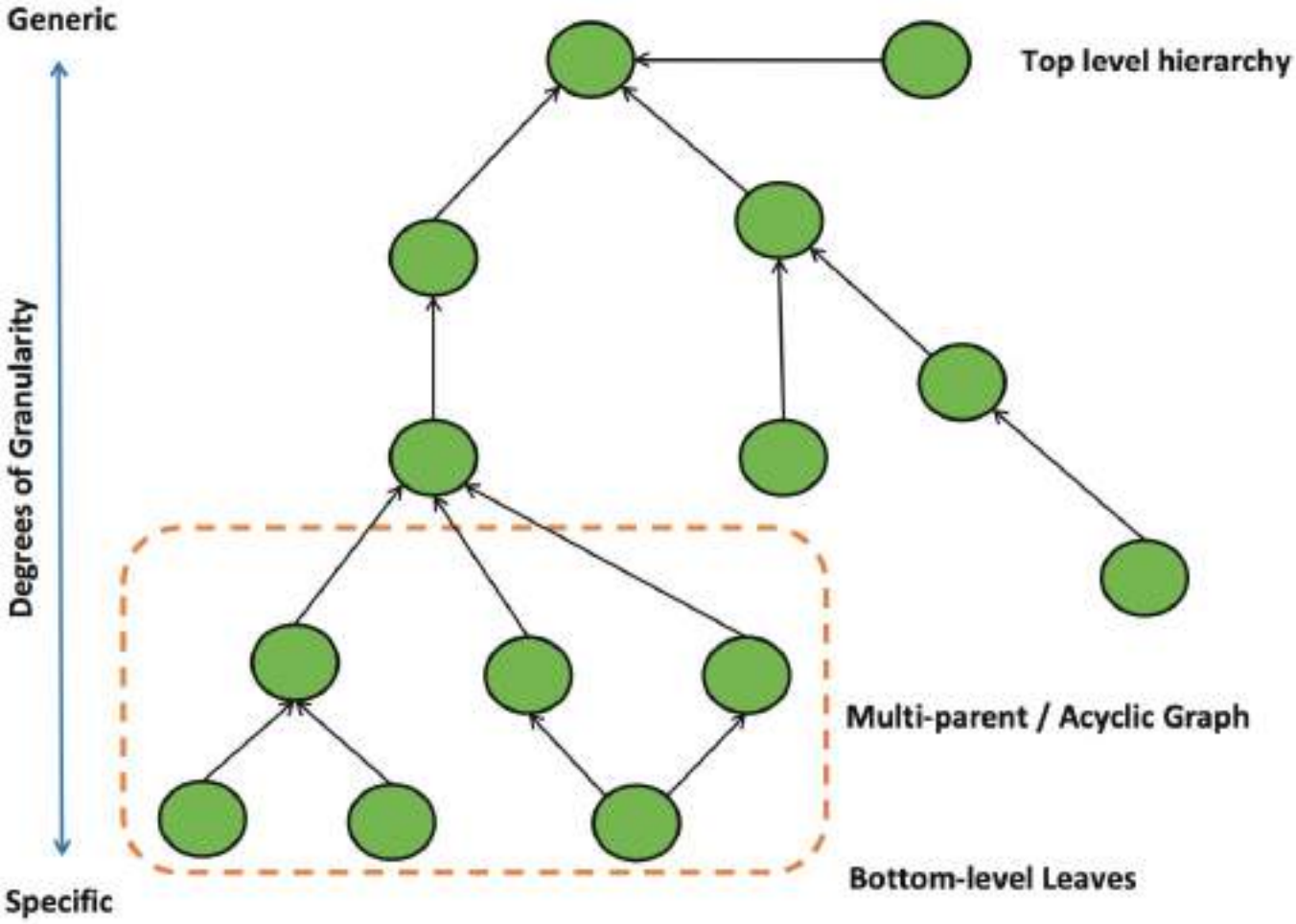


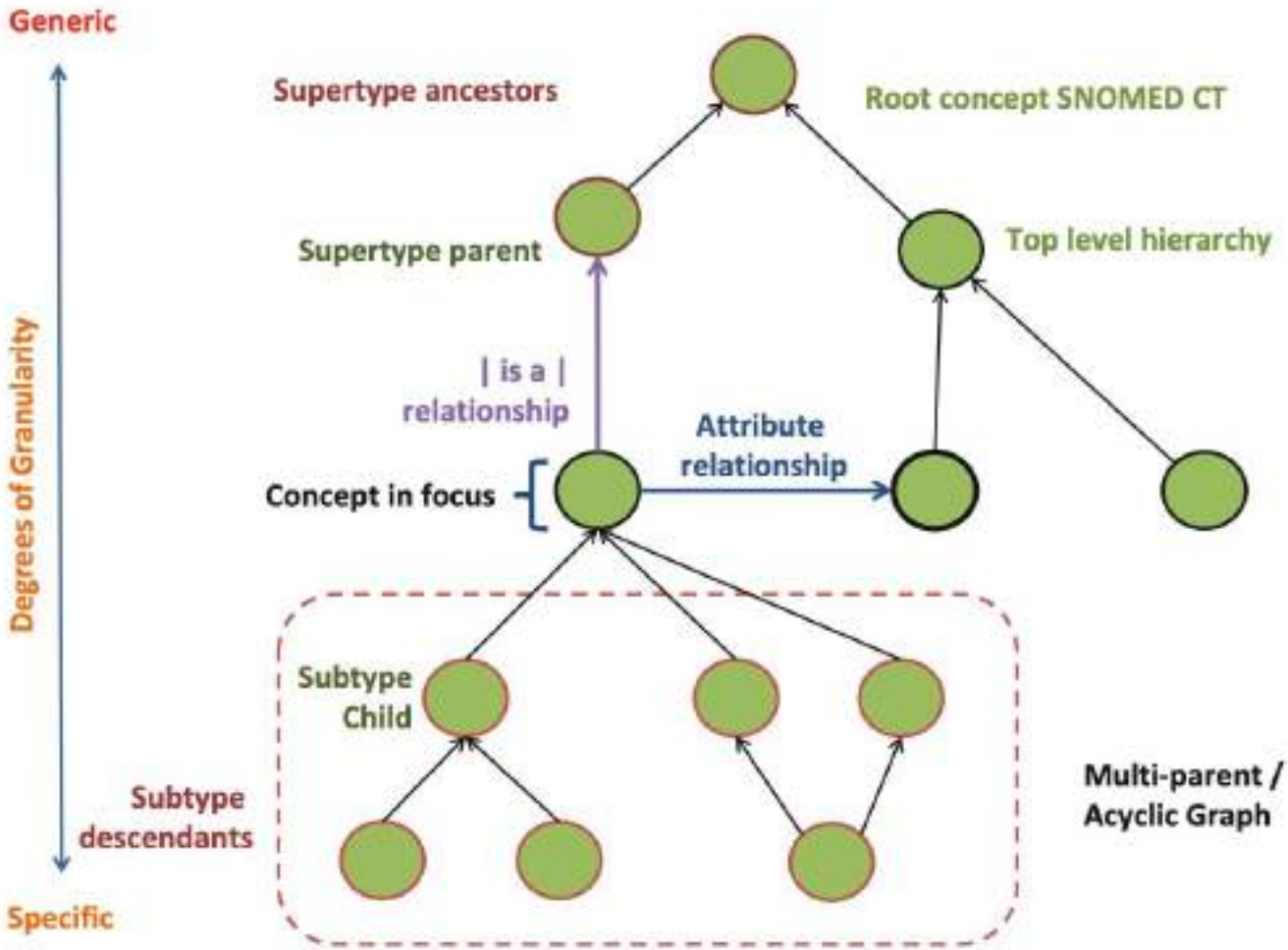
Conceptual model
shows what is stored in that particular application

Implementation
the actual information system that stores and manipulates the data

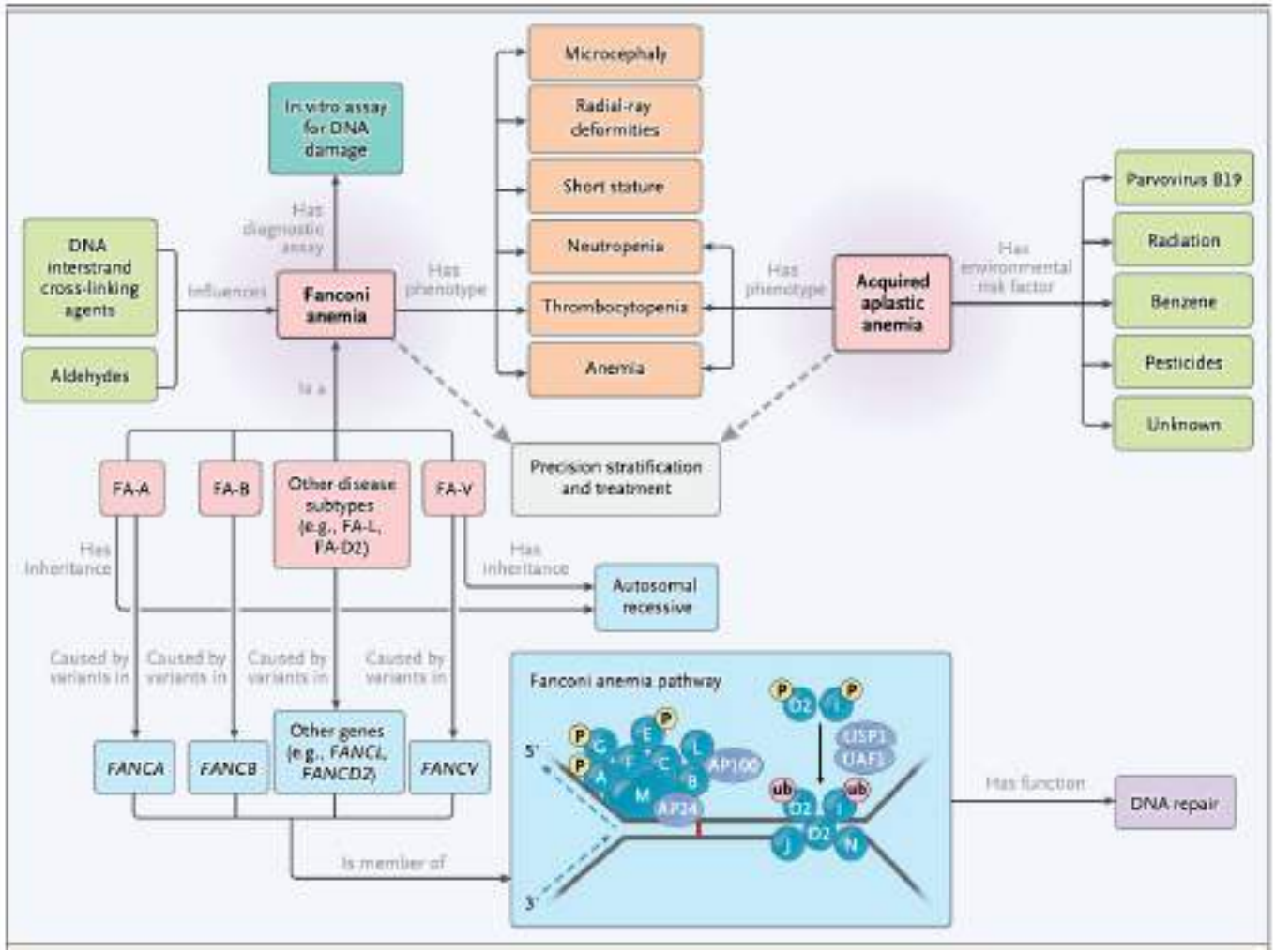
Concepts Instances Data




Generic
Degrees of Granularity
Specific
Top level hierarchy
Multi-parent / Acyclic Graph
Bottom-level Leaves


Generic
Degrees of Granularity
Specific
Supertype ancestors
Root concept SNOMED CT
Supertype parent
Top level hierarchy
| is a |
relationship
Attribute
relationship
Concept in focus
Subtype
Child
Subtype
descendants
Multi-parent /
Acyclic Graph

Automated Reasoning



SNOMED CT

- Most comprehensive multilingual clinical healthcare terminology and ontology
- > 80 countries use it in EHR systems for reporting and documentation
- Translated into 11 languages
- Concepts are related to each other, grouped and analysed
- Simple application: A Coding System for machine reading clinical documents
- Sophisticated applications: Compositional grammar with complex expressions

Origin

- 1955- American College of Pathologists started SNOP (Systematized Nomenclature of Pathology)
- 1965- SNOP published, describes pathology findings using four axes:
  - Topography (anatomic site affected)
  - Morphology (structural changes associated with disease)
  - Etiology (the cause of disease) including organisms
  - Function (physiologic alterations associated with disease)
- SNOP was the first multi-axial coding system used in healthcare

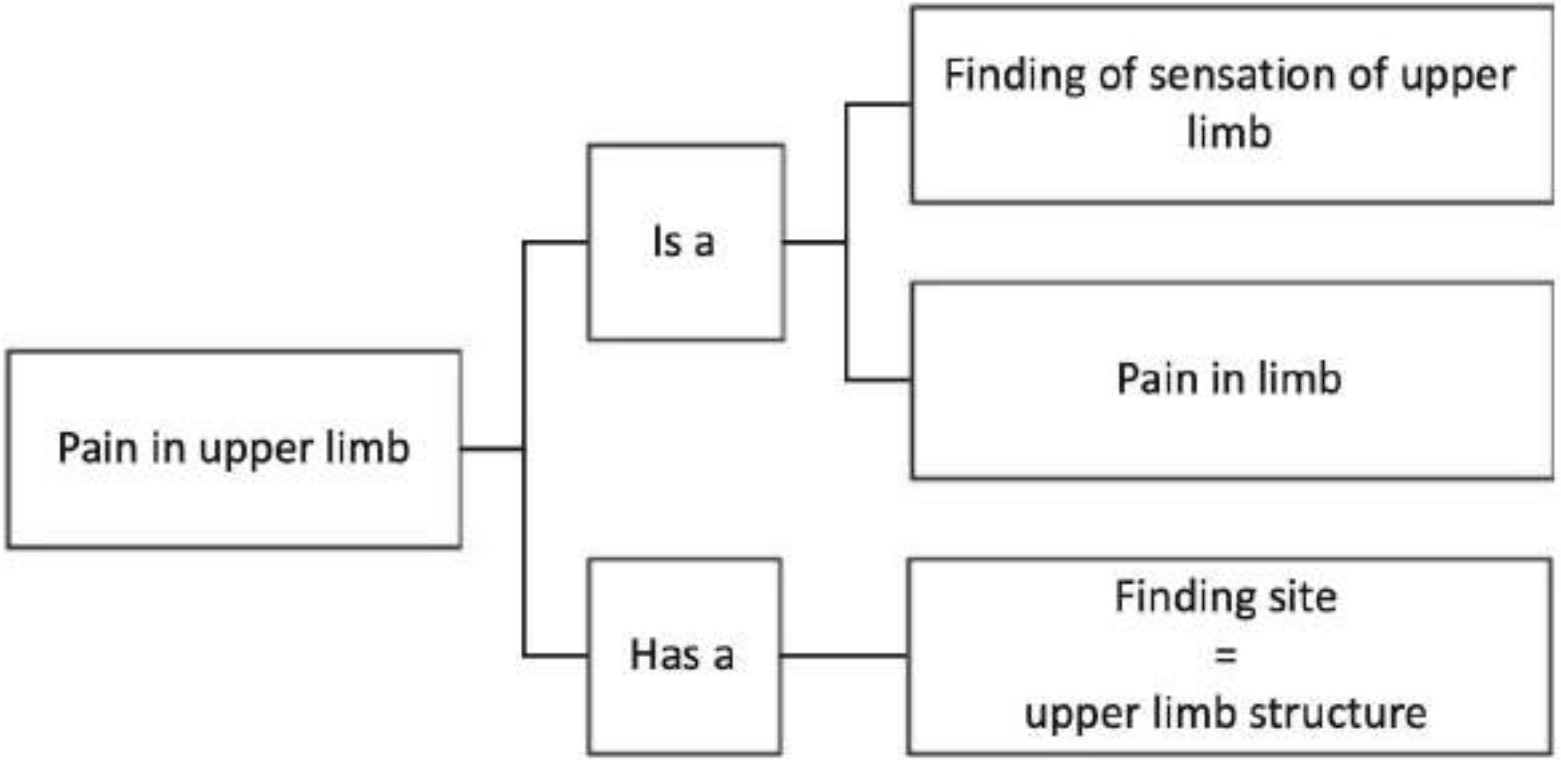
- urine and glucose are *substances*
- urine glucose concentration is an *observable entity*
- a urine glucose test is a *procedure*
- a urine glucose test result is a *clinical finding*
- if a urine glucose test is not done it is a *situation with explicit context.*

SNOMED 3.5 Axes

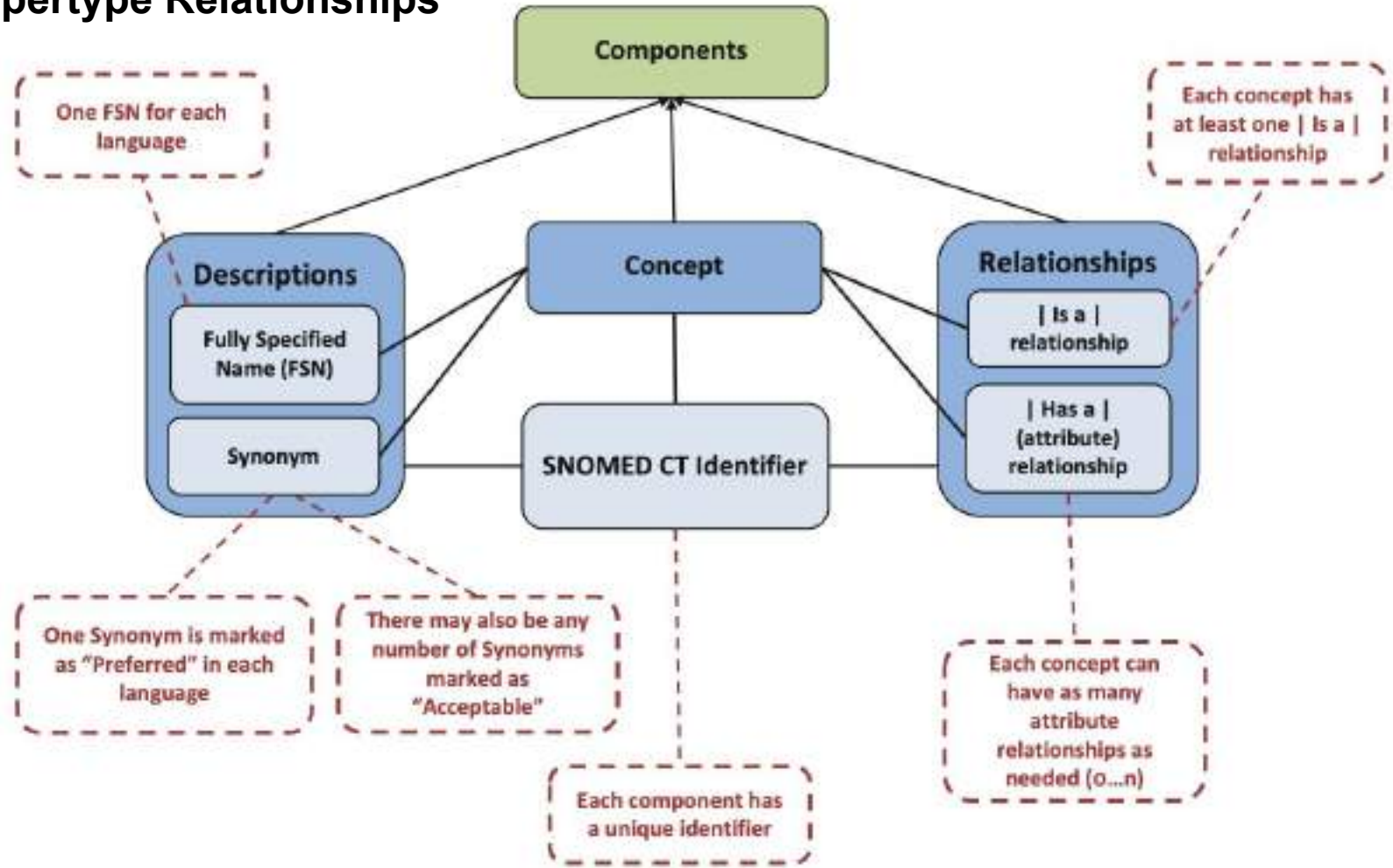
Axis	Description
T	Topography—Anatomic terms (13,000 records)
M	Morphology—Changes found in cells, tissues and organs (6000 records)
L	Living organisms—Bacteria and viruses (25,000 records)
C	Chemical—Drugs (15,000 records)
F	Function—Signs and symptoms (19,000 records)
J	Jobs Terms that describe the occupation (1900 records)
D	Diagnosis—Diagnostic terms (42,000 records)
P	Procedure—Administrative, diagnostic and therapeutic procedures (31,000 records)
A	Agents—Devices, physical agents and activities associated with disease (1600 records)
S	Social context—Social conditions and important relationships in medicine (500 records)
G	General—Syntactic linkages and qualifiers (1800 records)

Concepts, Descriptions

- **Concepts** are clinical meanings that do not change.
- SNOMED CT is concept-oriented
- Concept IDs are machine readable
- Each concept is associated text **descriptions**,
- Descriptions are human readable form of the concept
- Every concept has at least **two descriptions**
    - Fully Specified Name (FSN) - unique and unambiguous, not meant for end users
    - Display term in the language being used
- FSN has a suffix in () indicating its primary hierarchy e.g. myocardial infarction **(disorder)**.
- Display term is often FSN without suffix, e.g. myocardial infarction
- All other descriptions are synonyms
- Synonyms are marked as **Preferred** or **Acceptable**


Pain in upper limb
Is a
Finding of sensation of upper limb
Pain in limb
Has a
Finding site
=
upper limb structure

Supertype Relationships



Defining Relationships

- (1) Supertype Relationships and (2) Defining Attributes
- Everything is a concept
- E.g. |is a| Relationship -> Concept ID: 116680003
- Every concept except the SNOMED CT root has a supertype
- A concept is **sufficiently defined** if its defining relationships are sufficient to distinguish it from all its supertype and sibling concepts.
- Another concept represented as a combination of the same defining characteristics, is equivalent to it or a subtype of it.
- Large parts of SNOMED CT are not yet sufficiently defined.
- Primitive concepts: not fully defined and do not have the unique relationships needed to distinguish them from their parent or sibling concepts.
- E.g. Pneumonia - unless defining characteristics are specified

Expressions

- Pre-coordinated: when a single concept identifier is used to represent a clinical idea
- Post-coordinated: to represent a meaning using a combination of two or more concept identifiers
- ID |Name|
- Refinement : (colon)
  - Attribute (Can be more than one, separated by comma)
  - Value

```
80146002|appendectomy|:260870009|priority|=25876001|emergency|
```

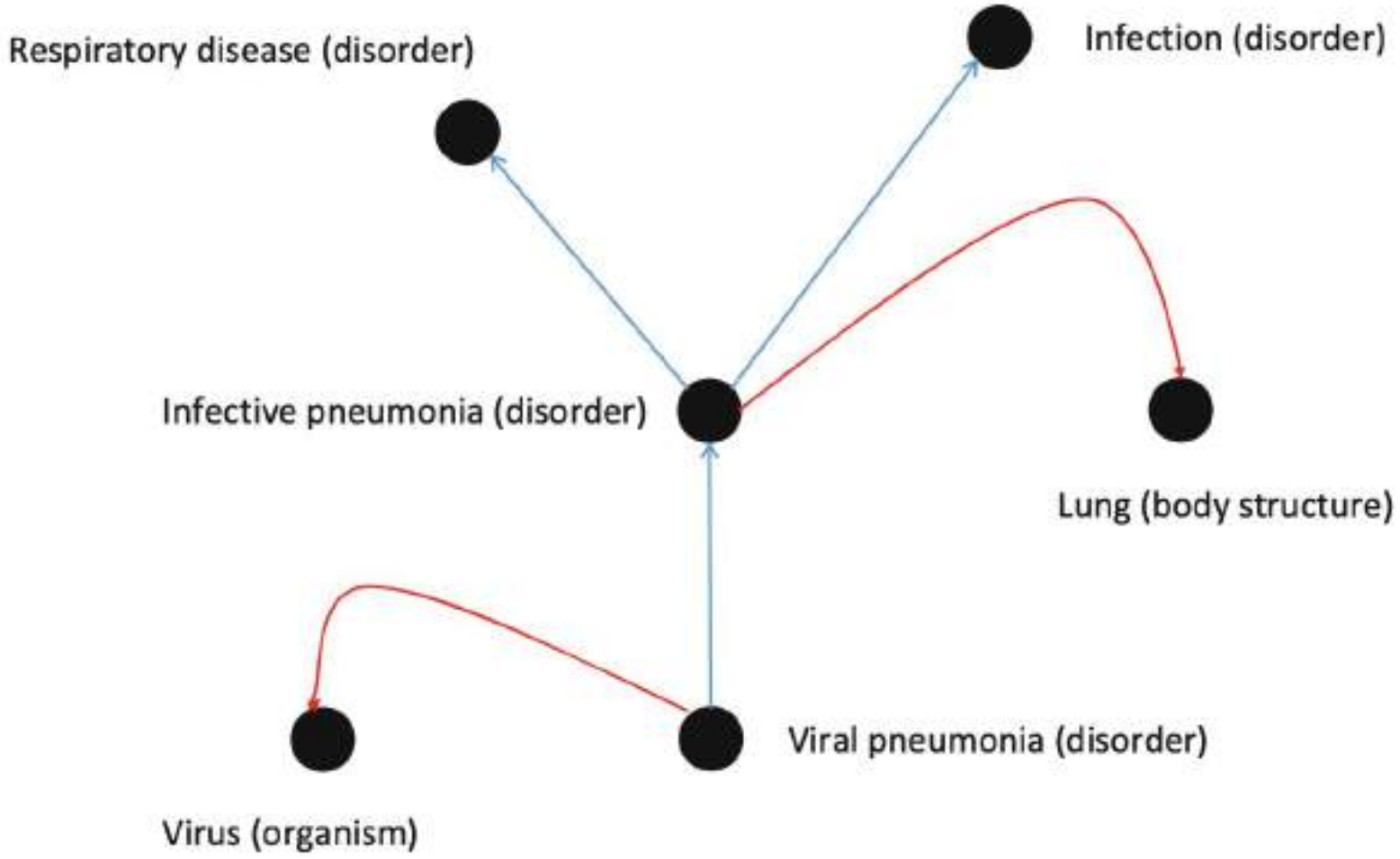
```
80146002|appendectomy|:260870009|priority|=25876001|emergency|,
425391005|using access device|=86174004|laparoscope|
```

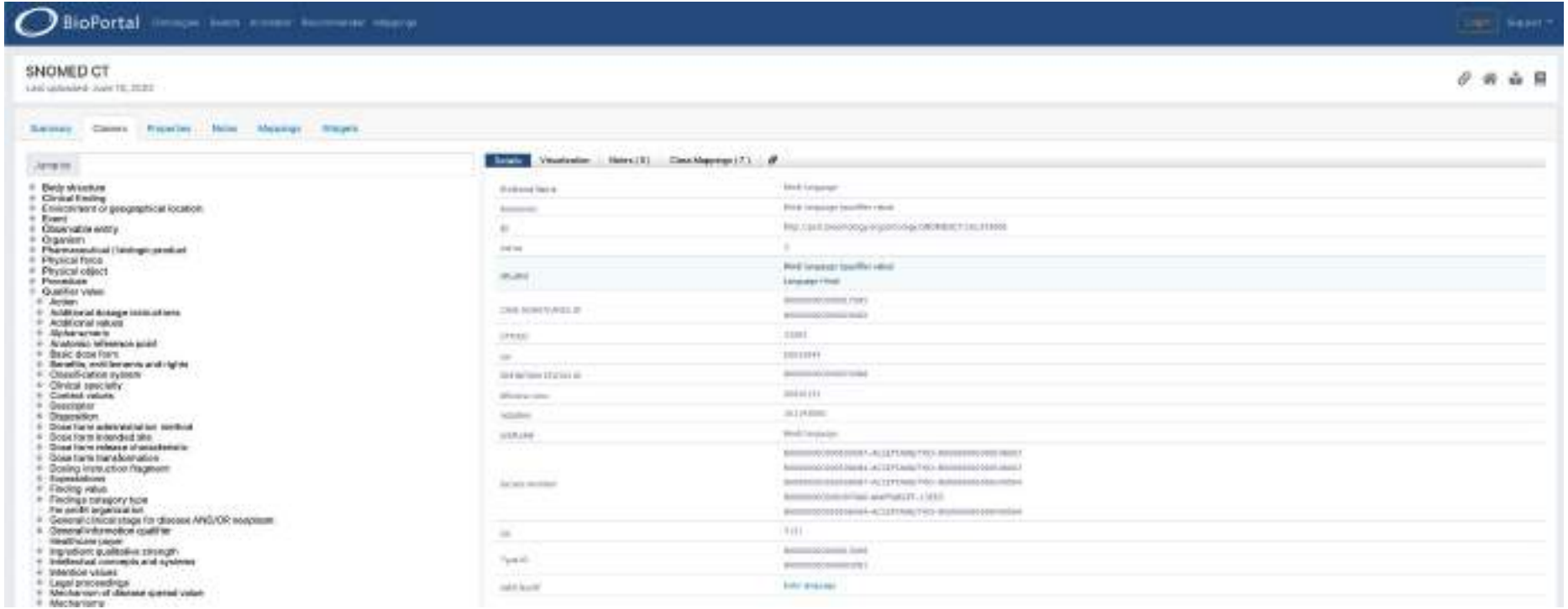
Description Logic

|concept|:|attribute|=|value|

| Viral pneumonia (disorder) | | is a | | Infective pneumonia (disorder) |

| Viral pneumonia (disorder) | "has a" | Causative agent (attribute) |
= | Virus (organism) |
| Infective pneumonia (disorder) | | is a | | Respiratory disease (disorder) |
| Infective pneumonia (disorder) | | is a | | Infection (disorder)
| Infective pneumonia (disorder) | "has a" | Finding site (attribute) |
= | Lung (body structure) |


Respiratory disease (disorder)
Infection (disorder)
Infective pneumonia (disorder)
Lung (body structure)
Viral pneumonia (disorder)
Virus (organism)



https://bioportal.bioontology.org/ontologies/SNOMEDCT?p=classes&conceptid=161143006

ramses-antibiotics/snomedizer
R Interface to the SNOMED-CT Terminology Server REST API
Package index

Search the ramses-antibiotics/snomedizer package

Vignettes
README.md

Functions 72

Source code 10

Man pages 12

api_operations: SNOMED CT Terminology Se
concepts_descendants: Fetch descendants
concepts_descriptions: Fetch descriptions c
concepts_find: Find SNOMED CT concepts
release_version: Fetch SNOMED CT RF2 rel
result_completeness: Check that results are
result_flatten: Flatten results from a server re
snomed_endpoint_test: Test a SNOMED CT
snomedizer_options: Set SNOMED CT endp

snomedizer-package: snomedizer: R Interface to the SNOMED-CT Terminology Server...

In ramses-antibiotics/snomedizer: R Interface to the SNOMED-CT Terminology Server REST API

Description | SNOMED CT licensing | Licensing for reference use | Author(s) | See Also

Description

Manipulate the SNOMED CT clinical ontology using the SNOMED International Terminology Server REST API <https://github.com/IHTSDO/snowstorm>. Jogo

SNOMED CT licensing

In order to use SNOMED CT, a licence is required which depends both on the country you are based in, and the purpose of your work. Although SNOMED International allows access to the SNOMED CT terminology server hosted by IHSTO to anyone for reference purpose, uses for data analysis or health care production systems is subject to other licences.

SNOMED International allows access to public SNOMED-CT terminology servers to anyone for strict 'reference purposes' under the SNOMED International SNOMED CT Browser License Agreement (see section below).

Use of SNOMED CT terminology for data analysis or health care production systems is subject to other licences. Some users are eligible for free licences:

- UK-based users can obtain a licence free of charge on the NHS TRUD website
- residents of other Member Countries and low-income countries are also eligible. More information can be found on SNOMED International's website.

Licensing for reference use

`snomedizer` contains examples and documentation that makes uses of an IHSTO public API endpoint restricted to reference uses such as testing or browsing the terminology. It is subject to the SNOMED International Browser License Agreement reproduced below:

1. The meaning of the terms "Affiliate", or "Data Analysis System", "Data Creation System", "Derivative", "End User", "Extension", "Member", "Non-Member Territory", "SNOMED CT" and "SNOMED CT Content" are as defined in the SNOMED International Affiliate License Agreement (see http://snomed.org/license-affiliate).
2. Information about Affiliate Licensing is available at https://www.snomed.org/snomed-ct/get-snomed. Individuals or organizations wishing to register as SNOMED International Affiliates can register at mlds.ihtsdotools.org, subject to acceptance of the Affiliate License Agreement (see http://snomed.org/license-affiliate).
3. The current list of SNOMED International Member Territories can be viewed at www.snomed.org/members. Countries not included in that list are "Non-Member Territories".
4. End Users, that do not hold an SNOMED International Affiliate License, may access SNOMED CT® using SNOMED International SNOMED CT Browser subject to acceptance of and adherence to the following sub-license limitations:

    - a) The sub-licensee is only permitted to access SNOMED CT® using this software (or service) for the purpose of exploring and evaluating the terminology.
    - b) The sub-licensee is not permitted the use of this software as part of a system that constitutes a SNOMED CT "Data Creation System" or "Data Analysis System", as defined in the SNOMED International Affiliate License. This means that the sub-licensee must not use SNOMED International SNOMED CT Browser to add or copy SNOMED CT identifiers into any type of record system, database or document.
    - c) The sub-licensee is not permitted to translate or modify SNOMED CT Content or Derivatives.
    - d) The sub-licensee is not permitted to distribute or share SNOMED CT Content or Derivatives.
5. SNOMED International Affiliates may use SNOMED International SNOMED CT Browser as part of a "Data Creation System" or "Data Analysis System" subject to the following conditions:

    - a) The SNOMED International Affiliate, using SNOMED International SNOMED CT Browser must accept full responsibility for any reporting and fees due for use or deployment of such a system in a Non-Member Territory.
    - b) The SNOMED International Affiliate must not use SNOMED International SNOMED CT Browser to access or interact with SNOMED CT in any way that is not permitted by the Affiliate License Agreement.
    - c) In the event of termination of the Affiliate License Agreement, the use of SNOMED International SNOMED CT Browser will be subject to the End User limitations noted in 4.

Development of phenotype algorithms using electronic medical records and incorporating natural language processing

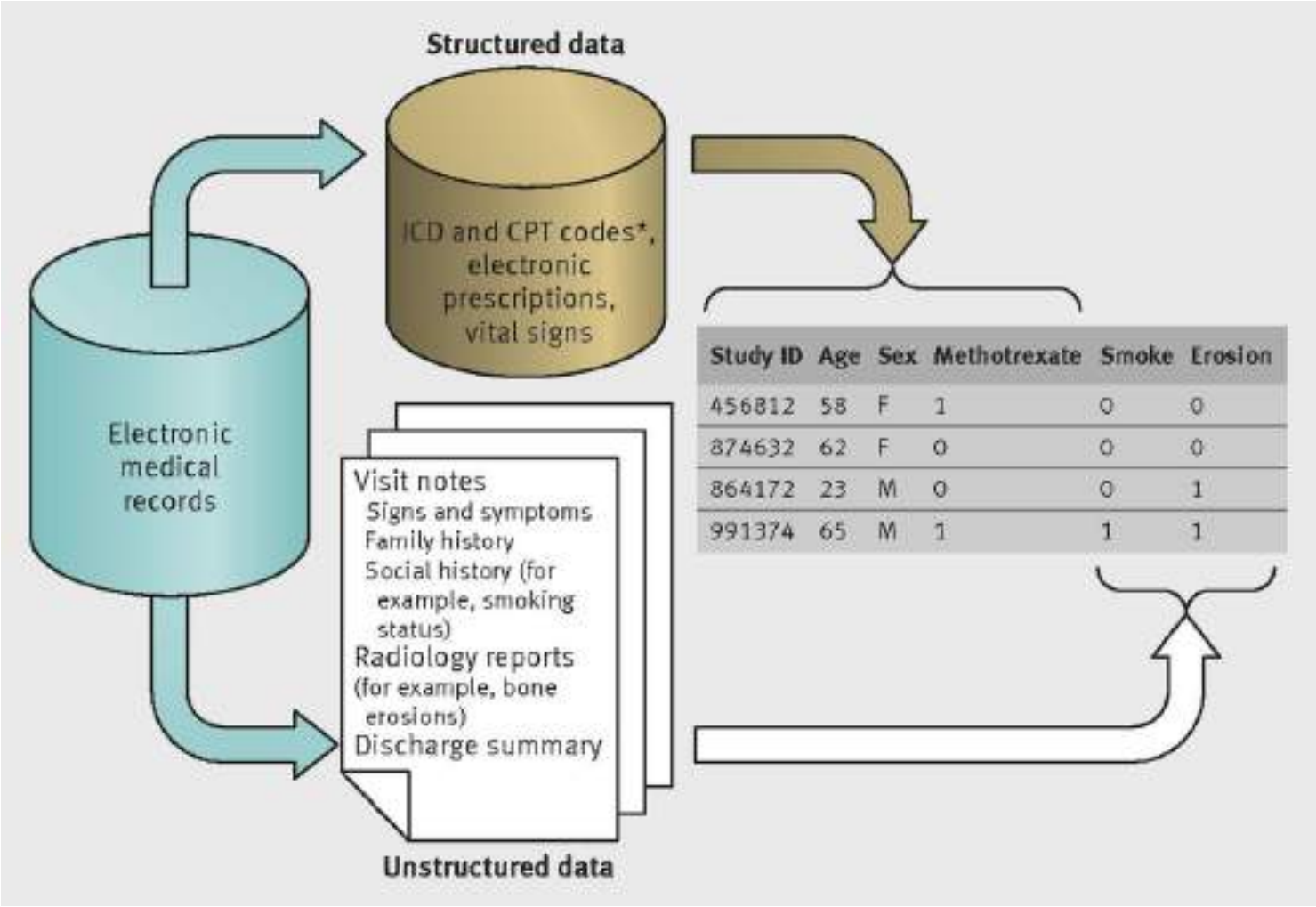
BMJ 2015 ; 350 doi: https://doi.org/10.1136/bmj.h1885 (Published 24 April 2015)
Cite this as: *BMJ* 2015;350:h1885

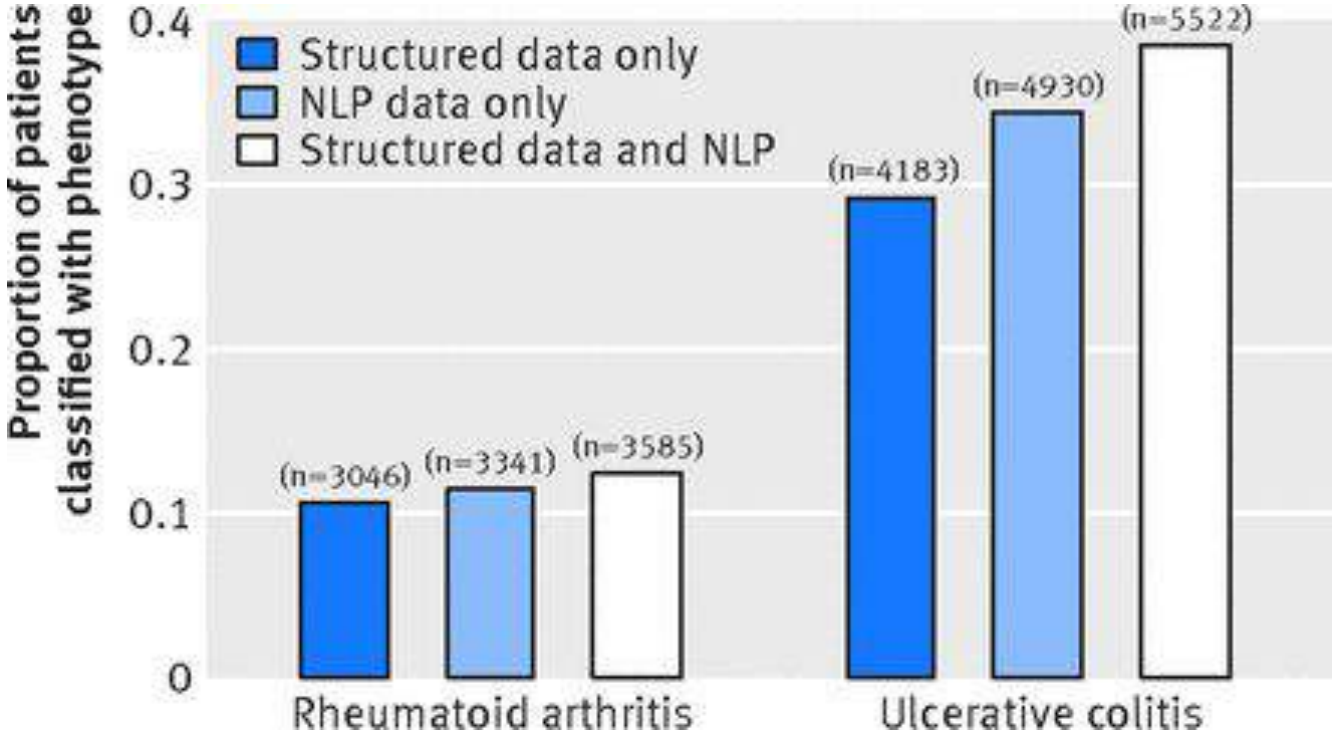
Katherine P Liao, assistant professor [1 2], Tianxi Cai, professor [3], Guergana K Savova, associate professor [4], Shawn N Murphy, associate professor [5], Elizabeth W Karlson, associate professor [1 2], Ashwin N Ananthakrishnan, assistant professor [6], Vivian S Gainer, senior analyst [7], Stanley Y Shaw, assistant professor [2 8], Zongqi Xia, assistant professor [2 9], Peter Szolovits, professor [10], Susanne Churchill, executive director [2], Isaac Kohane, professor [2 5]

Author affiliations

Correspondence to: K P Liao kliao@partners.org

Accepted 2 February 2015


Structured data
ICD and CPT codes*, electronic prescriptions, vital signs
Electronic medical records
Visit notes
Signs and symptoms
Family history
Social history (for example, smoking status)
Radiology reports (for example, bone erosions)
Discharge summary
Unstructured data
Study ID Age Sex Methotrexate Smoke Erosion
456812 58 F 1 0 0
874632 62 F 0 0 0
864172 23 M 0 0 1
991374 65 M 1 1 1


Proportion of patients classified with phenotype
0.4
0.3
0.2
0.1
0
Structured data only
NLP data only
Structured data and NLP
(n=3046)
(n=3341)
(n=3585)
(n=4183)
(n=4930)
(n=5522)
Rheumatoid arthritis
Ulcerative colitis

https://browse.welch.jhmi.edu/datasets/nih-data-repositories

Hopkins Initiatives

- COVID-19 Precision Medicine Analytics Platform Registry (JH-CROWN)
  The main data source is Johns Hopkins' electronic medical record, Epic. The registry is refreshed weekly with new and updated data and is available for Johns Hopkins investigators to analyze subsets of the COVID-19 patient population for retrospective analyses. CHSOR members Dr. Jodi Segal and Dr. Caleb Alexander have been using these data.
- COVID-19-specific Common Data Model
  PCORnet®, the National Patient-Centered Clinical Research Network, is creating a COVID-19-specific Common Data Model that will allow the use of information from patients across PCORnet's network At Johns Hopkins, Dr. Harold Lehmann leads these activities.
- Johns Hopkins COVID-19 Collaboration Platform
  Over 400 such trials have been registered on clinicaltrials.gov with dozens being added each day. Many of them are designed to answer similar questions and combining data or aggregating evidence could dramatically increase their efficiency and precision, getting answers to doctors faster and more reliably. *more...*
- National COVID Cohort Collaborative
  The National COVID Cohort Collaborative is the partnership among the NCATS-supported Clinical and Translational Science Awards (CTSA) Program hubs and the National Center for Data to Health (CD2H). At Johns Hopkins, Dr. Christopher Chute has taken the lead on this activity.

The Healthcare Data Revolution -> Transformation

- 4CE: Consortium for Clinical Characterization of COVID-19 by EHR
- Figshare: COVID-19 open data
- GitHub: COVID-19 Open Repo Data
- Harvard Dataverse: COVID-19 Data
- ICPSR: COVID-19 Data Repoasitory
- ImmPort: COVID 19
- Mendeley Data: Elsevier COVID-19 Research Environment
- National COVID Cohort Collaborative (N3C)
  OHDSI: Characterizing Health Associated Risks, and Your Baseline Disease In SARS-COV-2 (CHARYBDIS)
- Open-Access Data and Computational Resources to Address COVID-19.
- OpenSAFELY
  Tableau: COVID-19 Data Hub
- Vivli: Covid data
  Zenodo: Coronavirus Disease Research Community - COVID-19
- COVID-19 Research Database

MIMIC (Medical Information Mart for Intensive Care) Database

Medical Information Mart for Intensive Care III (MIMIC-III) is a large, freely-available database comprising de-identified health-related data associated with over 40,000 patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012.

The database includes information such as demographics, vital sign measurements made at the bedside (~1 data point per hour), laboratory test results, procedures, medications, caregiver notes, imaging reports, and mortality (both in and out of hospital).

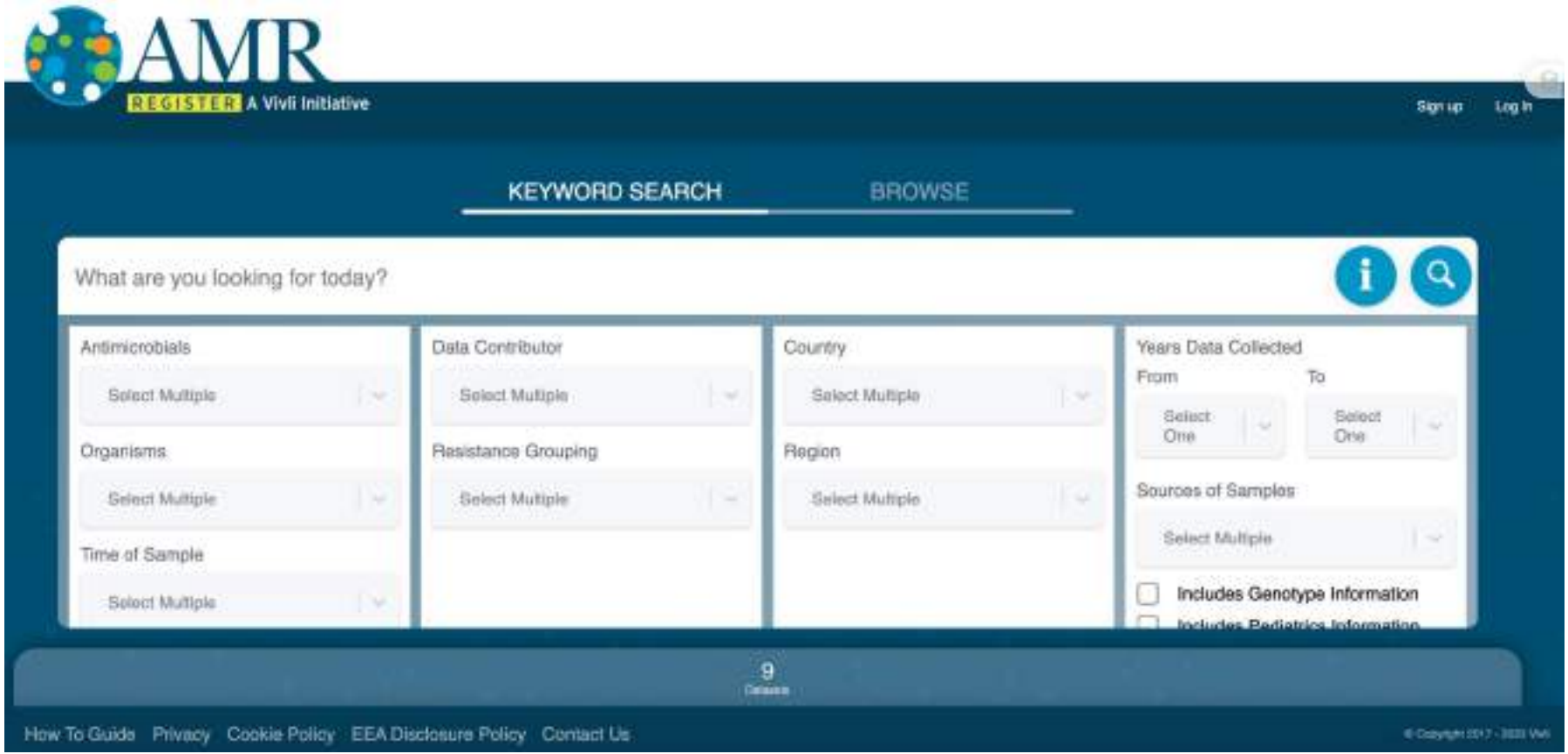
MIMIC-III supports a diverse range of analytic studies spanning epidemiology, clinical decision-rule improvement, and electronic tool development. It is notable for three factors:

- it is freely available to researchers worldwide
- it encompasses a diverse and very large population of ICU patients
- it contains high temporal resolution data including lab results, electronic documentation, and bedside monitor trends and waveforms

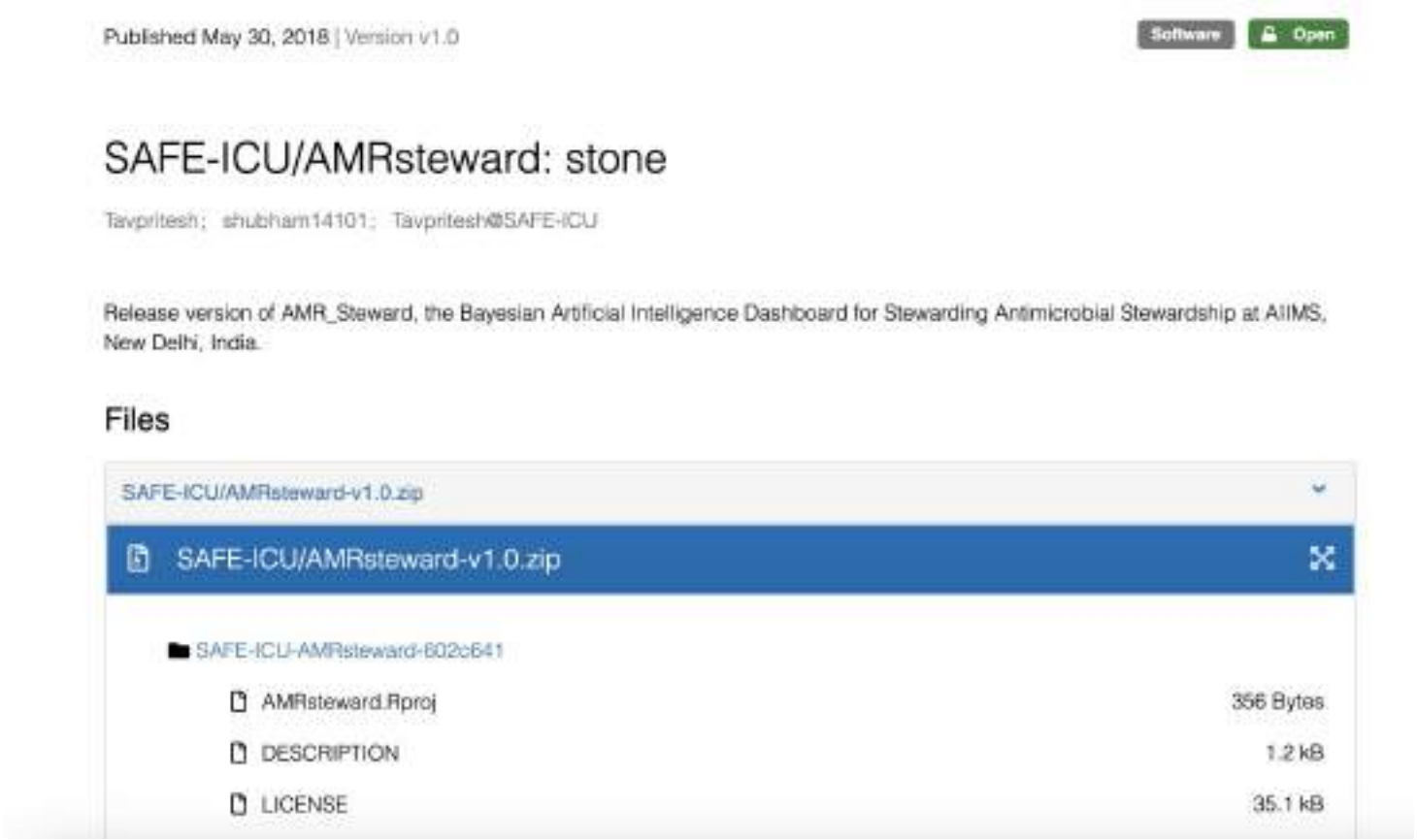
https://browse.welch.jhmi.edu/datasets/ehr-databases

Sources of Health Data

Sources of Health Data



Sources of Health Data



https://zenodo.org/records/1255584

Thanks for attending the class!