

Computing for Medicine

Google Classroom Code: dnd5qkt5

Monsoon 2025

Lecture 3

Data Sources

Open Discussion (5 Minutes)

Who Benefits from Health Data?

Computing for Medicine Starts With Data!

Visits :

21 February 2006: dyspnea, coughing and fever. Dark defecation.

PE: BP 150/90, pulse 95/min, Fever: 39.3 °C.

Ronchi +, no abdominal tenderness.

Medications: 64 mg Aspirin/day.

Possible acute bronchitis and cardiac decompensation.

Possible bleeding due to Aspirin.

Rx: Amoxicilline 500 mg 2x1, Aspirin 32 mg/day.

4 March 2006: no cough, slight dyspnea, defecation normal.

PE: light rhonchi, BP 160/95, pulse 82/min.

Rx: Aspirin 32 mg/day.

Lab :

21 February 2006: ESR 25 mm, Hb 7.8, Fecal occult blood +.

4 March 2006: Hb 8.2, Fecal occult blood :-.

X-ray

21 February 2006: Chest x-ray: no atelectasis, light cardiac decompensation findings

Problem 1: Coughing

21 February 2006

S: dyspnea, coughing, fever.

O: pulse 95/min, Fever: 39.3 °C.
Rhonchi+. ESR 25 mm.

Chest x-ray: no atelectasis, light
cardiac decompensation
findings.

A: Acute bronchitis.

P: Amoxicilline 500 mg 2x1.

4 March 2006

S: no coughing, slight dyspnea.

O: pulse 82/min. Slight rhonchi.

A: minimal bronchitis findings.

Problem 2: Dyspnea

21 February 2006

S: Dyspnea.

O: Rhonchi+, BP 150/90 mmHg.
Chest x-ray: no atelectasis, slight
cardiac decompensation
findings.

A: Slight decompensation
findings.

4 March 2006

S: slight dyspnea.

O: BP: 160/95, pulse 82/min.

A: No decompensation.

Problem 3: Dark colored defecation

21 February 2006

S: Dark feces. Using Aspirin 64 mg/day.

O: No abdominal tenderness, rectal exam revealed no blood, Hb 7.8 mg/dl. Fecal occult blood +

A: Possible intestinal bleeding due to Aspirin.

P: Decrease Aspirin dose to 32 mg/day.

4 March 2006

S: Defecation normal.

O: Fecal occult blood -

A: No intestinal bleeding symptoms.

P: Continue Aspirin dosage 32 mg/day

Making Healthcare Data Science Reproducible and FAIR

To be **Findable**:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be **Accessible**:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

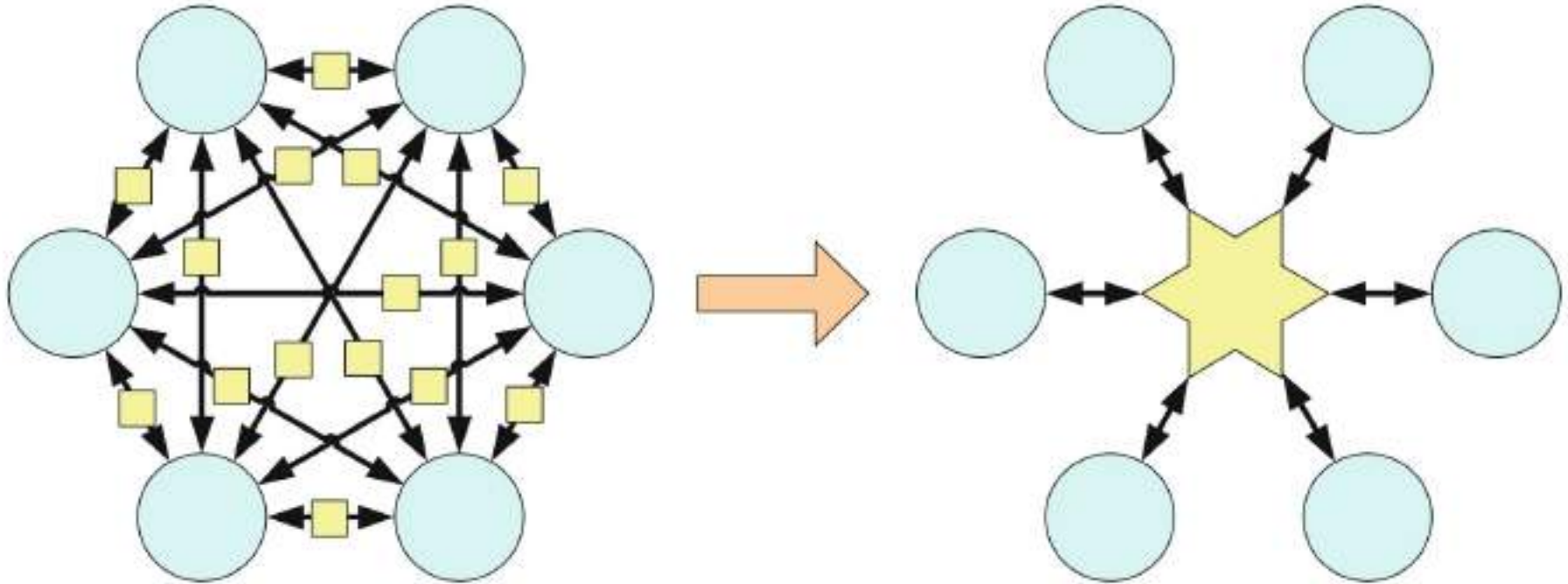
To be **Interoperable**:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be **Reusable**:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

Using Standards Allows the Same Vocabulary



Principles of Health Interoperability

How does a Computer Read a Medical Record?

At his first office visit, an obese 60-year-old male

99203 (CPT)

414915002 (SNOMED CT)

a (UCUM)

male (HL7)

reported increased thirst and frequent urination.

249477003 (SNOMED CT)

28442001 (SNOMED CT)

After a hemoglobin A1c test reported 8.2% HbA1c,

4548-4 (LOINC)

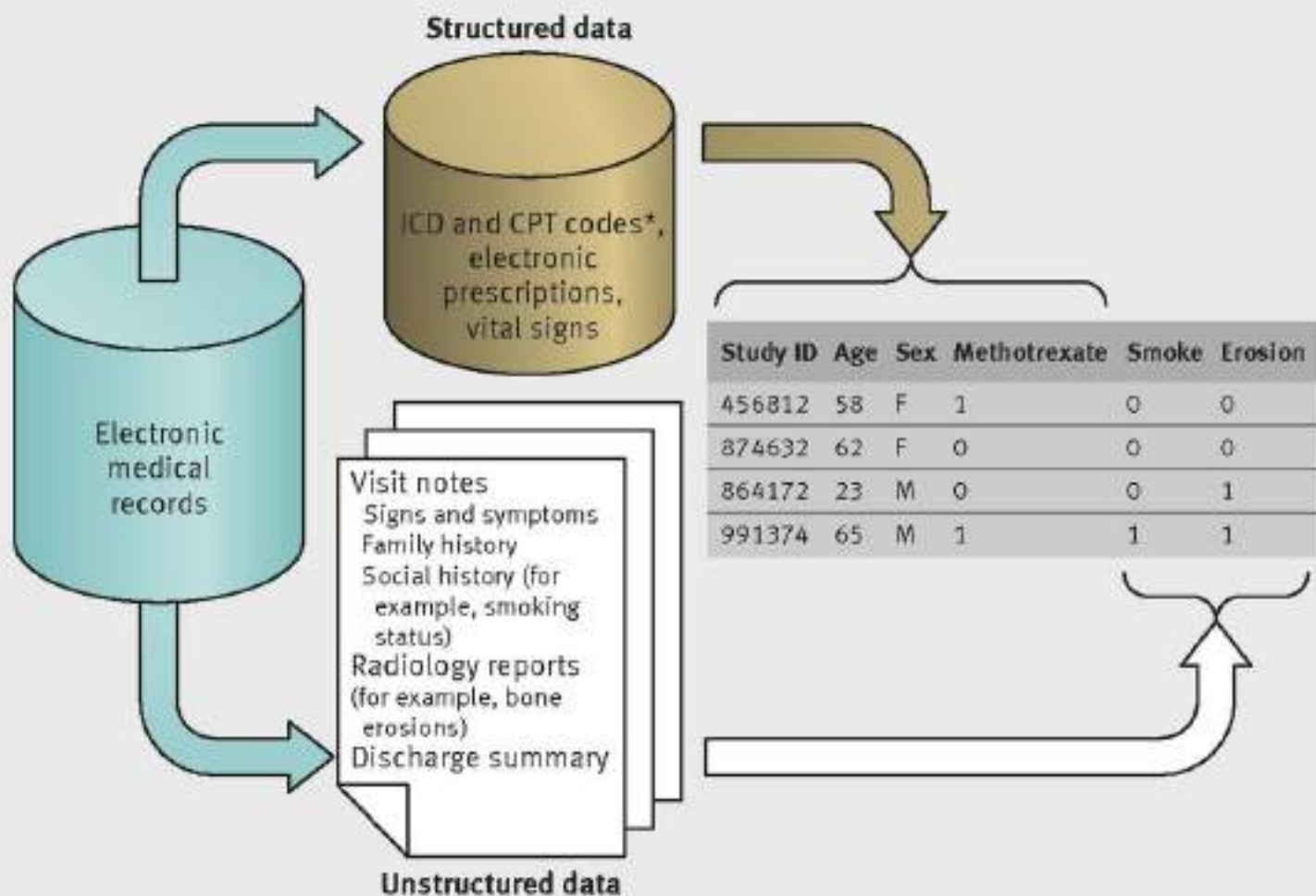
% (UCUM)

the patient was diagnosed with type 2 diabetes

E11.9 (ICD-10-CM)

and prescribed metformin.

6809 (RxNorm)

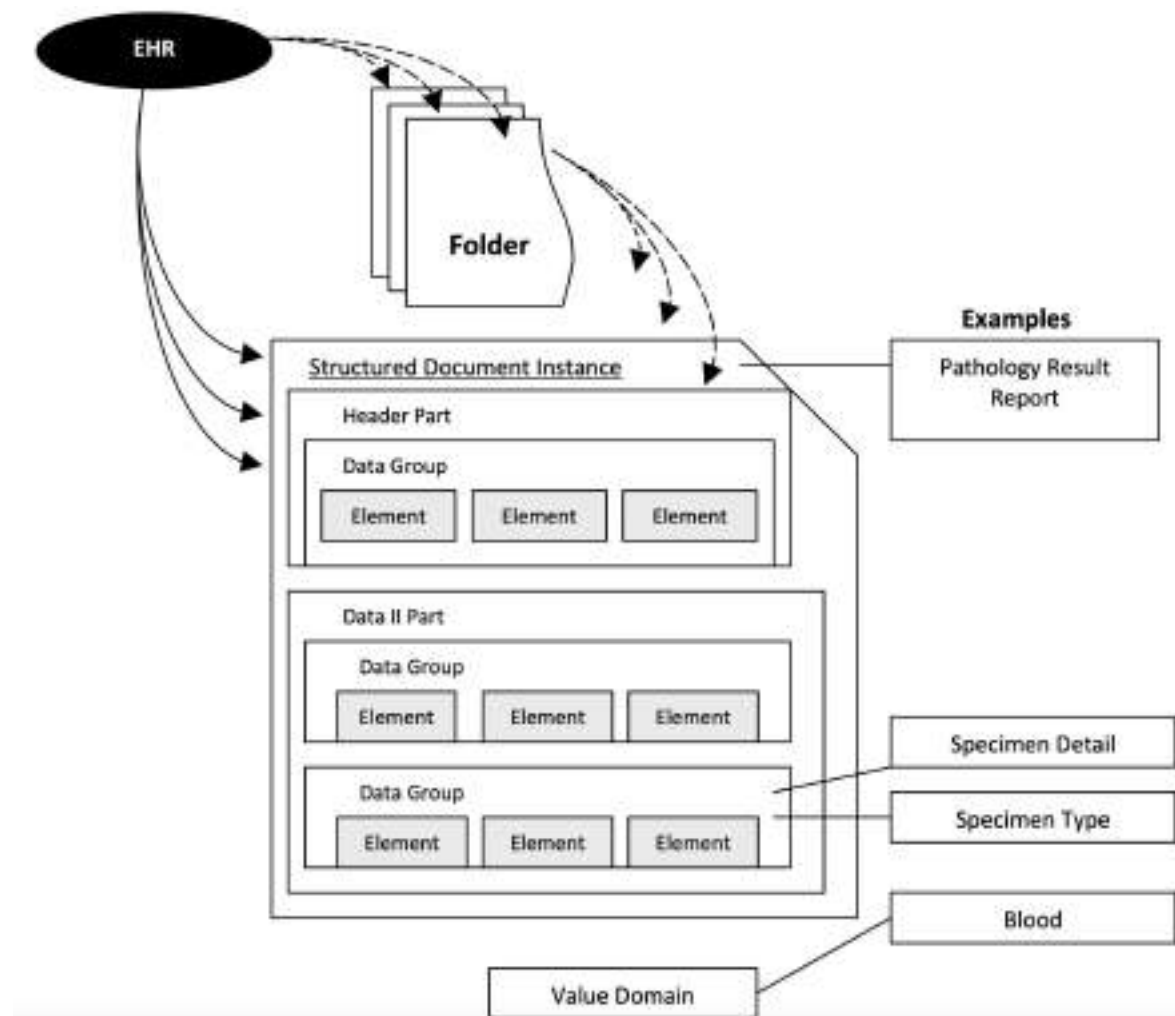


What is an Electronic Health Record

ISO/DTR 20514: “A repository of information regarding the health of a subject of care in **computer processable** form, stored and transmitted **securely**, and accessible by multiple authorized users. It has a **standardized** information **model**, which is independent of EHR systems.

Its primary purpose is the support of continuing efficient and quality integrated healthcare and it contains information, which is **retrospective, concurrent, and prospective**”

EHR



Purpose of EHR

Australia (HealthConnect)

Initial Health Profile	Pharmacy Provision
Hospital Discharge—Emergency	Community-Based Health Consultation
Medical Consultation—General Practitioner	Allied Health Consultation
Medical Consultation—Specialist	Referral
Diagnostic Investigation—Imaging	Event Notification (for example, admission to hospital)
Diagnostic Investigation—Pathology	Pharmacy Provision

Austria (ELGA)

Primary Goals	Secondary Goals
Patient care	Quality management
Patient management	Automating healthcare activities
Streamlining the treatment process	Financial and Administrative processes
A lifelong EHR	Science and Research

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

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](#)

12k Accesses | **33** Citations | **138** Altmetric | [Metrics](#)

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.

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>

Example of Syntactic Interoperability: HL7 Standard

```
##### Translate this function from HL7 V2.8 ADT_A01 to HL7 V2.4 ADT_A05  
### HL7 V2.8 ADT_A01
```

```
MSH|^~\&|ADT1|GOOD HEALTH HOSPITAL|GHH LAB, INC.|GOOD HEALTH  
HOSPITAL|198808181126|SECURITY|ADT^A01^ADT_A01|MSG00001|P|2.8||  
EVN|A01|200708181123||  
PID|1||PATID1234^5^M11^ADT1^MR^GOOD HEALTH  
HOSPITAL~123456789^^^USSSA^SS||EVERYMAN^ADAM^A^III||19610615|M||C|222  
2 HOME STREET^^GREENSBORO^NC^27401-1020|GL|(555) 555-2004|(555)555-  
2004||S||PATID12345001^2^M10^ADT1^AN^A|444333333|987654^NC|  
NK1|1|NUCLEAR^NELDA^W|SPO^SPOUSE|||NK^NEXT OF KIN  
PV1|1|I|2000^2012^01|||004777^ATTEND^AARON^A|||SUR|||ADM|A0|  
### HL7 V2.4 ADT_A05
```

Example of Syntactic Interoperability: FHIR Standard

```
```json
{
 "resourceType": "Bundle",
 "type": "message",
 "entry": [
 {
 "resource": {
 "resourceType": "MessageHeader",
 "eventCoding": {
 "system": "http://hl7.org/fhir/message-events",
 "code": "ADT_A01"
 },
 "destination": [
 {
 "endpoint": "http://localhost:8080/fhir/baseDstu3"
 }
],
 "source": {
 "name": "ADT1",
 "software": "GOOD HEALTH HOSPITAL",
 "endpoint": "http://localhost:8080/fhir/baseDstu3"
 }
 }
 }
]
}
```

## Subjects

- ☒ COVID-19
- ☐ Coronavirus
- ☐ SARS-CoV-2
- ☐ coronavirus
- ☐ covid-19
- ☐ Covid-19
- ☐ covid
- ☐ Coronaviridae
- ☐ pathogens
- ☐ CETAF-taskforce

## File type

- ☐ PDF
- ☐ BIN
- ☐ ZIP

Clear

March 16, 2020 (2020-03-01)

Dataset

Open

881

### COVID-19 Open Research Dataset (CORD-19)

Sebastian Kohlmeier; Kylie Lo; Lucy Lu Wang

352

Important: This dataset is updated regularly and the latest version for download can be found here. In response to the COVID-19 pandemic, the Allen Institute for AI has partnered with leading research groups to prepare and distribute the COVID-19 Open Research Dataset (CORD-19), a free resource of scholarly articles, including full text conten...

263

Uploaded on May 7, 2020

7 more versions exist for this record

13243

7422

180

148

April 11, 2020 (v1)

Preprint

Open

141

### Immunological mechanisms explaining the role of IgE, mast cells, histamine, elevating ferritin, IL-6, D-dimer, VEGF levels in COVID-19 and dengue, potential treatments such as mast cell stabilizers, antihistamines, Vitamin C, hydroxychloroquine, ivermectin and azithromycin

Arumugam, Vinu

119

A novel coronavirus, SARS-CoV-2 was identified in Wuhan, China. The disease caused by the virus can range in severity from asymptomatic to acute respiratory distress syndrome (ARDS) and death. Primary dengue infection results in IgE mediated sensitization against dengue virus (DENV) proteins. These IgE bind to receptors on mast cells. Upon sub...

118

Uploaded on April 11, 2020

13229

7445

4,293

November 13, 2020 (2020-10-20)

Journal article

Open

398

### COVIPENDIUM: information available to support the development of medical countermeasures and interventions against COVID-19

Martina DENIS; Valerie VANDEWEERD; Rein VERBEEKE

145

# Finding Datasets for Secondary Analysis

[About This Guide](#)[New to Hopkins?](#)[Research Data Repositories & Databases](#)[NIH Data Repositories](#)[Examples of NIH Data Repositories](#)[Other Data Repositories/Consortium](#)

## NIH Data Repositories

In general, NIH does not endorse or require sharing in any specific repository and encourages researchers to select the repository that is most appropriate for their data type and discipline (though such specification does exist for particular initiatives). To help researchers locate an appropriate resource for sharing their data, as well as to promote awareness of resources where datasets can be located for reuse, Trans\_NIH BioMedical Informatics Coordinating Committee (BMIC) maintains lists of several types of data sharing resources:

- Open NIH-supported domain-specific repositories that house data of a specific type or related to a specific discipline;
- Other NIH-supported domain-specific resources, including repositories and knowledgebases, that have limitations on submitting and/or accessing data; and
- Generalist repositories that house data regardless of type, format, content, or subject matter.



## 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](https://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 Repository
- 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

# Sources of Health Data

---

[Home](#)[About](#)[Members](#)[News & Events](#)[Resources](#)[Portals](#)[LOG IN](#)

## A global clinical research data sharing platform

The Vivli team is dedicated to helping researchers share and access data from clinical trials to advance science.

[SEARCH FOR STUDIES](#)[SUBMIT YOUR STUDY](#)



# Sources of Health Data



Sign up Log in

KEYWORD SEARCH

BROWSE

What are you looking for today?



Antimicrobials

Select Multiple

Organisms

Select Multiple

Time of Sample

Select Multiple

Data Contributor

Select Multiple

Resistance Grouping

Select Multiple

Country

Select Multiple

Region

Select Multiple

Years Data Collected

From

To

Select One

Select One

Sources of Samples

Select Multiple

☐ Includes Genotype Information

☐ Includes Pediatrics Information

9

Results

[How To Guide](#) [Privacy](#) [Cookie Policy](#) [EEA Disclosure Policy](#) [Contact Us](#)

© Copyright 2017 - 2020 Vivvi

# Sources of Health Data

Published May 30, 2018 | Version v1.0

Software

Open

## SAFE-ICU/AMRsteward: stone

Tavpritesh; shubham14101; Tavpritesh@SAFE-ICU

Release version of AMR\_Steward, the Bayesian Artificial Intelligence Dashboard for Stewarding Antimicrobial Stewardship at AIIMS, New Delhi, India.

### Files

SAFE-ICU/AMRsteward-v1.0.zip

SAFE-ICU/AMRsteward-v1.0.zip

SAFE-ICU-AMRsteward-b02c641

AMRsteward.Rproj	356 Bytes
DESCRIPTION	1.2 kB
LICENSE	35.1 kB



**Thanks for  
attending the class!**