ScHARe

Think-a-Thons

National Institutes of Health

Common Data Elements and Data Aggregation

Deborah Duran, PhD · NIMHD

Luca Calzoni, MD MS PhD Cand. · NIMHD

Denise Warzel, BBA MSc · NIH/NCI CBIIT

October 18, 2023

ScHARe

Science
collaborative for
Health disparities and
Artificial intelligence bias
Reduction

ScHARe

NIH
National Institute
on Minority Health
and Health Disparities
+
NIH
Office of
Data Science Strategy
+
NIH
National Institute
of Nursing Research

ScHARe

+

Dr. Deborah Duran NIH/NIMHD

Dr. Luca Calzoni NIH/NIMHD

Thank you

NIMHD

Dr. Eliseo Perez-Stable

ODSS

Dr. Susan Gregurick

NIH/OD

Dr. Larry Tabak

NINR

Dr. Shannon Zenk

NINR

Rebecca Hawes
Micheal Steele
John Grason

ORWH

OMH

NIMHD OCPL

Kelli Carrington
Thoko Kachipande
Corinne Baker

BioTeam

STRIDES

Terra

SIDEM

RLA

Broad Institute

CCDE Working Group

Deborah Duran
Luca Calzoni
Rebecca Hawes
Micheal Steele
Kelvin Choi
Paula Strassle
Michele Doose
Deborah Linares
Crystal Barksdale
Gneisha Dinwiddie
Jennifer Alvidrez
Matthew McAuliffe
Carolina Mendoza-Puccini
Simrann Sidhu
Tu Le

Experience poll

Please check your level of experience with the following:

	None	Some	Proficient	Expert
Python	☐	☐	☐	☐
R	☐	☐	☐	☐
Cloud computing	☐	☐	☐	☐
Terra	☐	☐	☐	☐
Health disparities research	☐	☐	☐	☐
Health outcomes research	☐	☐	☐	☐
Algorithmic bias mitigation	☐	☐	☐	☐

ScHARe

Overview

ScHARe

Phase I

Population Science and SDoH datasets
Tutorials and resources
Think-a-Thons

ScHARe

ScHARe is a cloud-based population science data platform designed to accelerate research in health disparities, health and healthcare delivery outcomes, and artificial intelligence (AI) bias mitigation strategies

ScHARe aims to fill three critical gaps:

- Increase participation of **women & underrepresented populations with health disparities** in data science through data science skills training, cross-discipline mentoring, and multi-career level collaborating on research
- Leverage population science, SDoH, and behavioral Big Data and cloud computing tools to foster a **paradigm shift** in healthy disparity, and health and healthcare delivery outcomes research
- **Advance AI bias mitigation and ethical inquiry** by developing innovative strategies and securing diverse perspectives

nimhd.nih.gov/schare

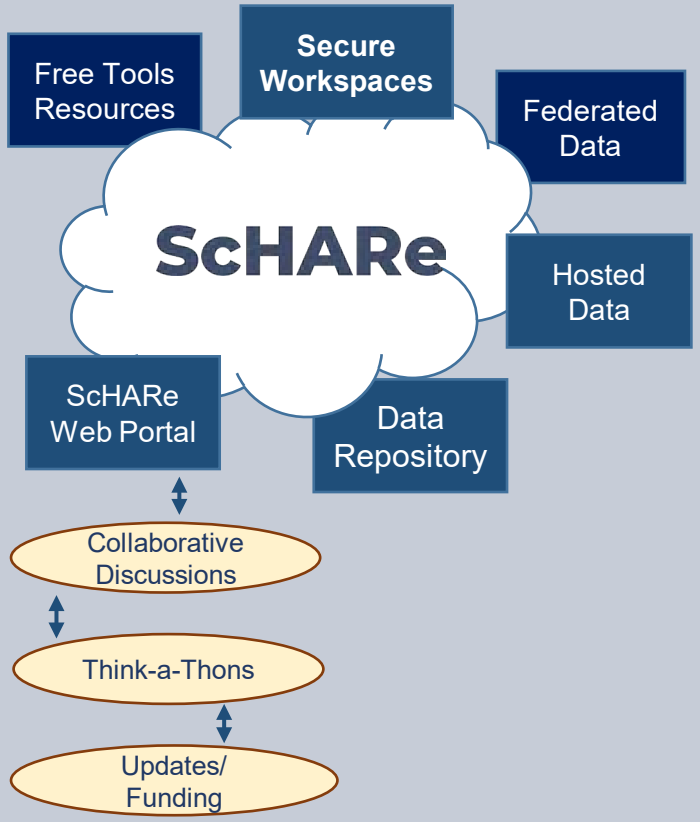
ScHARe Components

ScHARe co-localizes within the cloud:

- **Datasets** (including social determinants of health and social science data) relevant to minority health, health disparities, and health care outcomes research
- **Data repository** to comply with the required hosting, managing, and sharing of data from NIMHD- and NINR-funded research programs
- **Computational capabilities and secure, collaborative workspaces** for students and all career level researchers
- **Tools for collaboratively evaluating and mitigating biases** associated with datasets and algorithms utilized to inform healthcare and policy decisions

Frameworks: Google Platform, Terra, GitHub, NIMHD Web ScHARe Portal

Intramural & Extramural Resource



nimhd.nih.gov/schare

ScHARe Data Ecosystem

Researchers can access, link, analyze, and export **a wealth of datasets** within and across platforms relevant to research about health disparities, health care outcomes and bias mitigation, including:

- **Google Cloud Public Datasets:** publicly accessible, federated, de-identified datasets hosted by Google through the Google Cloud Public Dataset Program

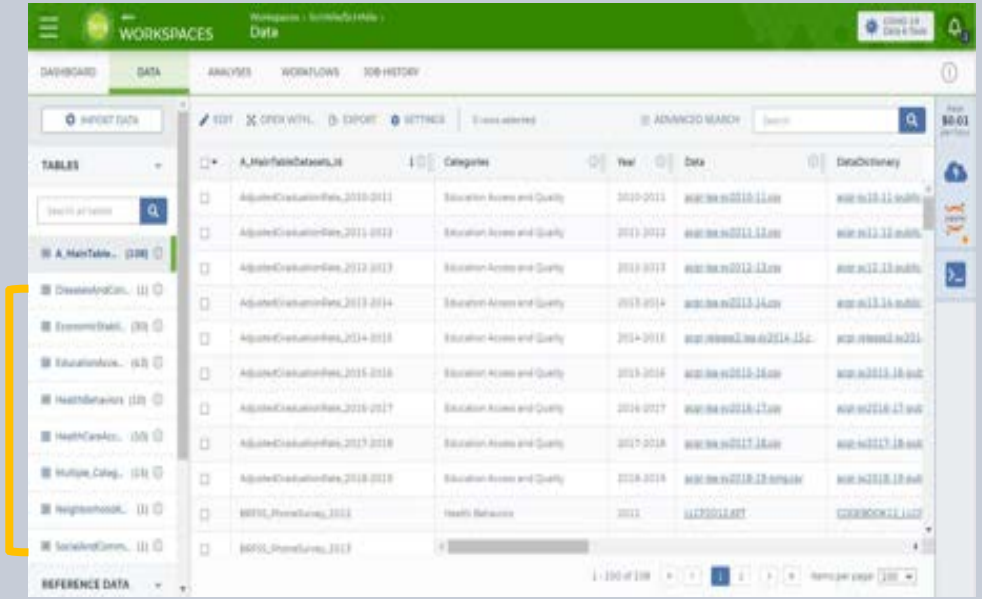
  Example: *American Community Survey (ACS)*

- **ScHARe Hosted Public Datasets:** publicly accessible, de-identified datasets hosted by ScHARe

  Example: *Behavioral Risk Factor Surveillance System (BRFSS)*

- **Funded Datasets on ScHARe:** publicly accessible and controlled-access, funded program/project datasets using Core Common Data Elements shared by NIH grantees and intramural investigators to comply with the NIH Data Sharing Policy

  Examples: *Jackson Heart Study (JHS); Extramural Grant Data; Intramural Project Data*



Datasets are categorized by content based on the CDC **Social Determinants of Health categories**:

1. Economic Stability
2. Education Access and Quality
3. Health Care Access and Quality
4. Neighborhood and Built Environment
5. Social and Community Context

with the addition of:

- **Health Behaviors**
- **Diseases and Conditions**

Users will be able to **map and link** across datasets

Access to Population Science datasets

ScHARe Data Ecosystem will offer access to **300+ datasets**, including:

- Google Cloud Public Datasets
- ScHARe Hosted Public Datasets:
  - American Community Survey
  - U.S. Census
  - Social Vulnerability Index
  - Food Access Research Atlas
  - Medical Expenditure Panel Survey
  - National Environmental Public Health Tracking Network
  - Behavioral Risk Factor Surveillance System

- **Coming Soon:** Repository for Funded Datasets on ScHARe, in compliance with NIH Data Sharing Policy

Cloud computing strategies

ScHARe

- Uses **workflows** in Workflow Description Language (**WDL**), a language easy for humans to read, for batch processing data
- **Python and R**, including most commonly used libraries
- Enables **customization** of computing environments to ensure everyone in your group is using the same software
- **Big Query** and **Tensorflow** access for advanced machine learning
- Enables researchers to create interactive **Jupyter notebooks** (documents that contain live code) and share data, analyses and results with their collaborators in real time
- For novice users, integration with **SAS** is planned

AI bias mitigation strategies

- Widespread use of AI raises a number of ethical, moral, and legal issues – likely not to go away
- Algorithms often are “black boxes”
- **Biases can result from:**
  - **social/cultural context not considered**
  - **design limitations**
  - **data missingness and quality problems**
  - **algorithm development and model training**
  - **Implementation**
- If not rectified, biases may result in decisions that lead to discrimination, unequitable healthcare, and/or health disparities
- **Lack of diverse perspectives:** populations with health disparities are underrepresented in data science
- **Guidelines** and recommendations emerging from HHS, NIST, White House, etc.

ScHARe

Critical thinking can rectify AI biases

ScHARe was created to:

- foster participation of **populations with health disparities in data science**
- promote the collaborative identification of **bias mitigation strategies** across the continuum
- create a **culture of ethical inquiry** and critical thinking whenever AI is utilized
- build **community confidence** in implementation approaches
- focus on **implementation of AI bias** guidelines and recommendations

ScHARe

Phase II

(in process)

Data ecosystem and repository

ScHARe Data Repository

CORE COMMON DATA ELEMENTS

NOVEL CDE FOCUSED REPOSITORY TO FOSTER INTEROPERABILITY

COMPLY WITH DATA SHARING POLICY - HOST PROJECT DATA

DATA ECOSYSTEM

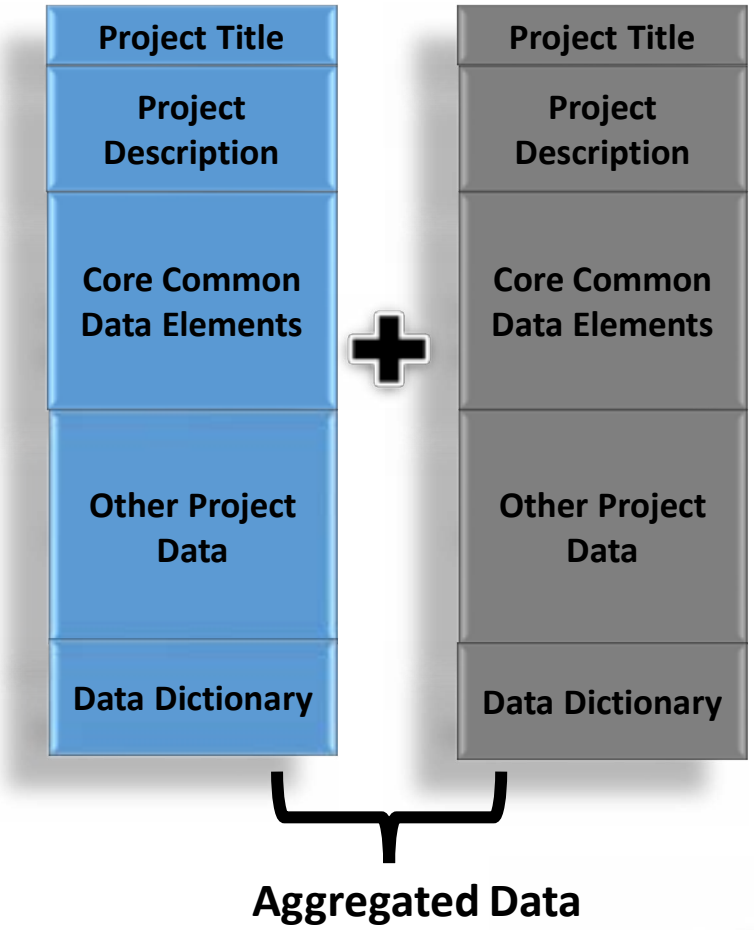
- Map across datasets
- Map across platforms

UPCOMING

ScHARe

Core Common Data Elements
Intramural and Extramural Project Repository

- Complies with **NIH Data Sharing Policy**
- Fosters dataset sharing and interoperability by using or mapping to **Core Common Data Elements**
- Provides resources for **intramural researchers** to work in a secure workspace and host data
- Centralizes **aggregated datasets** for repeat use



UPCOMING

ScHARe

Project & federated dataset mapping

- Project Title
- Project Description
- Core Common Data Elements
- Other Project Data
- Data Dictionary

\+ AMERICAN COMMUNITY SURVEY

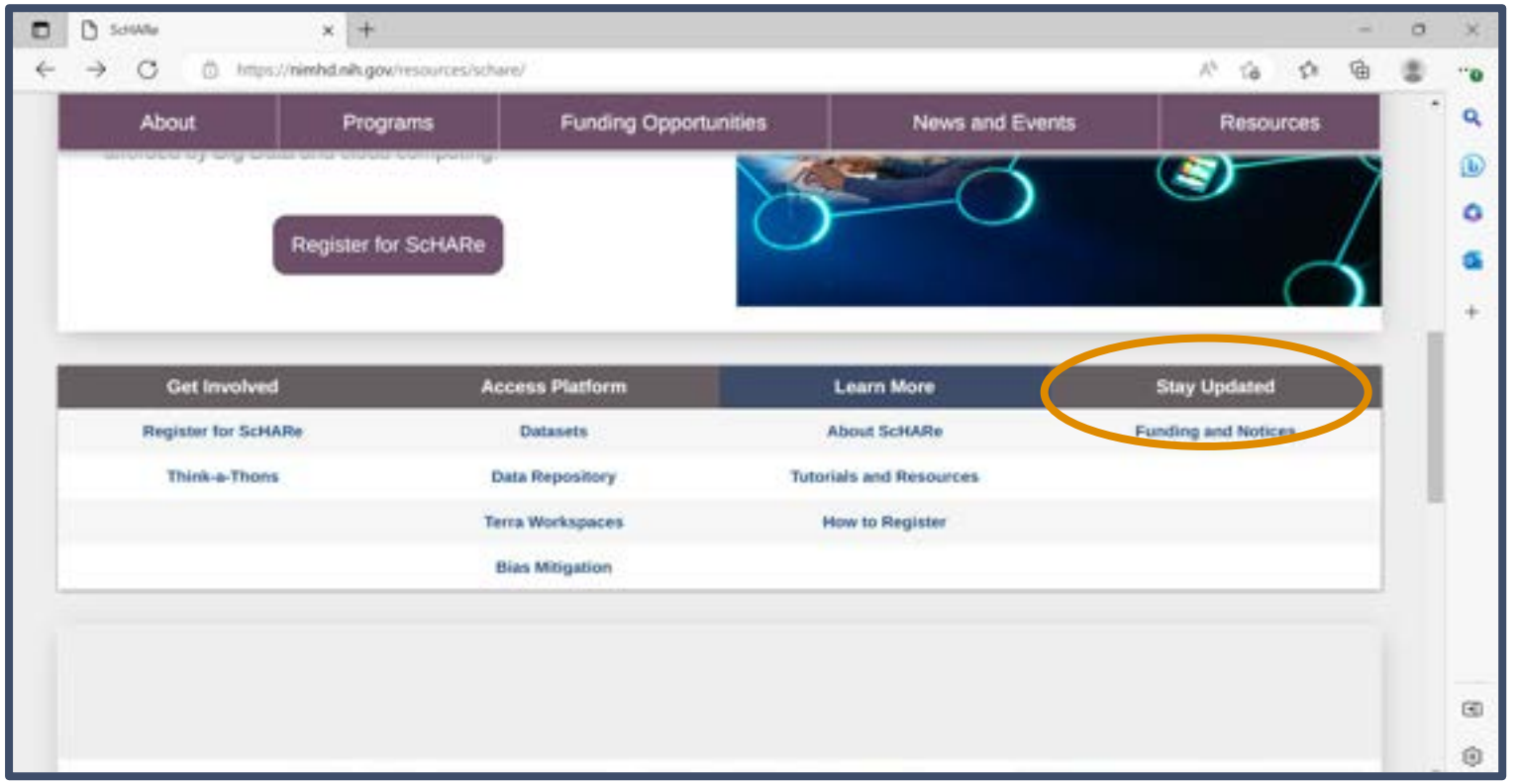
\+ Medical Expenditure Survey (MEPS)

\+ Pharmacy and health insurance databases

Mapping across cloud platforms

UPCOMING

Two ways to sign up for ScHARe news



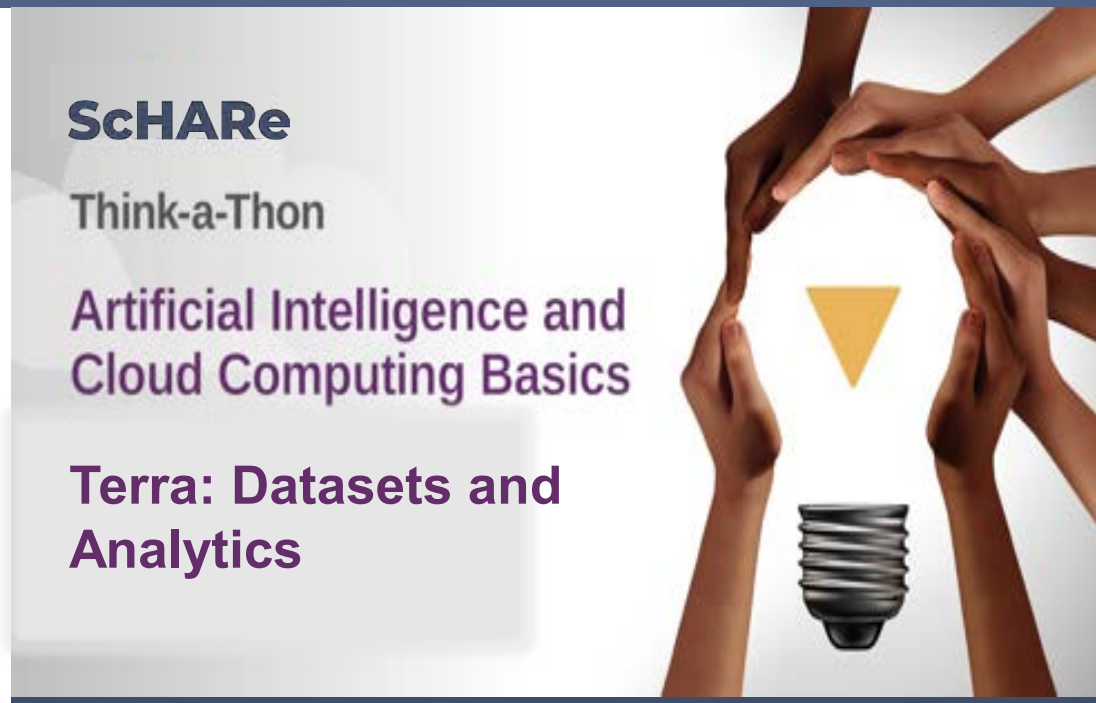
Scannable from your screen!

nimhd.nih.gov/schare

ScHARe Think-a-Thons (TaT)

- Monthly sessions (2 1/2 hours)
- Instructional/interactive
- Designed for new and experienced users
- Research & analytic teams to:
  - Conduct health disparities, health outcomes, bias mitigation research
  - Analyze/create tools for bias mitigation
- Publications from research team collaboration
- Networking
- Mentoring and coaching
- Focus:

✓ **Instructional**
✓ **Collaboration research teams**
✓ **Bias mitigation**



Register:

bit.ly/think-a-thons

Interest poll

I am interested in (check all that apply):

- ☐ Learning about Health Disparities and Health Outcomes research to apply my data science skills
- ☐ Conducting my own research using AI/cloud computing and publishing papers
- ☐ Connecting with new collaborators to conduct research using AI/cloud computing and publish papers
- ☐ Learning to use AI tools and cloud computing to gain new skills for research using Big Data
- ☐ Learning cloud computing resources to implement my own cloud
- ☐ Developing bias mitigation and ethical AI strategies
- ☐ Other

ScHARe

Guest expert

Denise Warzel

NIH/NCI CBIIT

About Denise

In 1997, doctors found that Denise Warzel's mother had lung cancer while she was working at IBM. Later that year, she went to NCI with a plan to use clinical trials data to improve research and patient care using computers. She realized that the data from different studies wasn't defined the same way, making it hard to use for computer analysis. So, she left IBM in 2000 and joined NCI to make sure all the data was defined in a common way, these are called common data elements or “CDEs”. NCI created something called the Cancer Data Standards Registry and Repository, which is now one of the biggest dictionaries of CDEs for cancer data. Many research groups use these CDEs to make sure they collect information in a similar way.

Denise played a big role in using special systems, like ontologies, to make data more meaningful. This helped people understand it better and use it correctly. She also made sure that different pieces of data worked well together and could be changed or found easily. Denise was a key part of making sure the ISO/IEC 11179 standard for CDE registries, which NCI, and NIH use, included this approach.

Since 2008, she's been the leader of the ISO Metadata Standards Working Group that takes care of this standard. Denise is very good at showing how to use these concepts to make data clear and easy to work with. She loves sharing what she knows to help others use CDEs better.

She led the NIH Covid working group to develop CDEs that laid a foundation for covid projects to accelerate the use of data. She sits on the NIH CDE governance group to share her expertise in using CDEs to advance research. Denise has an enthusiasm for CDE use because she knows the value for research.

She got her bachelor's degree from the University of Georgia, focusing on management and information systems. After that, she went to the University of Oxford and earned a Master of Science (MSc) degree, with Distinction, which is a big deal. She did well there, and her dissertation on structured data capture for healthcare and research using ISO/IEC 11179 played a big part in earning her that top honor.

ScHARe

Science **c**ollaborative for
Health disparities and
Artificial intelligence bias **Re**duction

Common Data Elements and Data Aggregation

Denise Warzel BBA, MSc

NIH/NCI CBIIT Semantic Infrastructure

October 18, 2023

National Institutes of Health

ScHARe

Think-a-Thon Goals:

- Understand the pivotal role of CDEs in the context of mapping and combining datasets (data aggregation)
- Learn how CDEs standardize and harmonize data across diverse datasets and studies, enabling research breakthroughs and innovations

Agenda

Before we Begin
- *Setting the stage*

Introduction to CDEs
- *What are CDEs and what is CDE Metadata?*

Increasing Awareness
- *Have you heard of…*

Terminologies and CDEs
- *What do you mean?*

Pulling it all Together
- *Standard and Structured*

Success Stories
- *CDEs in Action*

Before we begin…

- Data Sharing and Management Snafus in 3 short acts (length: 4:40)
  - NYU School of Medicine as part of their Health Sciences Library
    - 2012
  - Created by:
    - Karen Hanson
    - Alisa Surkis
    - Karen Yacobucci

Key Message

- Discover the problems created by not planning for data sharing

Knowledge Poll #1

Overall assessment

Before we begin…

1. I expect to encounter challenges similar to Panda's in my healthcare research.

   1. Very unlikely 2. Unlikely 3. Neutral 4. Likely 5. Very likely

2. It is important to my work to be able to reuse data from other sources.

   1. Very Unimportant 2. Unimportant 3. Neutral 4. Important 5. Very Important

3. I understand what CDEs are.

   1. Strongly disagree 2. Disagree 3. Neither agree or disagree 4. Agree 5. Strongly agree

Before we begin…

4. I understand the role of CDEs in mapping and aggregating data.

   1. Strongly disagree 2. Disagree 3. Neither agree or disagree 4. Agree 5. Strongly agree

5. I understand how to use CDEs.

   1. Strongly disagree 2. Disagree 3. Neither agree or disagree 4. Agree 5. Strongly agree

6. I understand when something is a CDE and when it is not.

   1. Strongly disagree 2. Disagree 3. Neither agree or disagree 4. Agree 5. Strongly agree

Before we begin...

7. I consider the following to be CDEs:
    7. OMOP - Agree Undecided Disagree
    8. Athena - Agree Undecided Disagree
    9. FHIR - Agree Undecided Disagree
    10. PhenX Toolkit - Agree Undecided Disagree
    11. REDCap Data Dictionary - Agree Undecided Disagree
    12. NIH CDE Repository - Agree Undecided Disagree

Success Story #1: CDEs helping reach underrepresented populations

CDEs in action

CDEs helping to reach underrepresented populations

NCI's Community Oncology Research Program (NCORP) Division of Cancer Prevention (DCP) Study-001

- Need to increase access to and enrollment of underrepresented populations in clinical trials
- Research Study included 46 Community Sites
  - 14 minority/underserved sites
  - 997 Affiliate sites - 43 states, Puerto Rico, and Guam
  - 19,373 responses
  - Expanded CDEs: demographic, including socioeconomic status (employment, income, education) and comorbidities
  - Analyzed after 43 months

Diane C. St. Germain, RN, MS; and Worta McCaskill-Stevens, MD, MS. ***Use of a Clinical Trial Screening Tool to Enhance Patient Accrual,*** *Cancer 2021;0:1-8.* *© 2021 American Cancer Society.*

NCORP DCP-001 – Screening CDEs

OMB Ethnicity question:

- CDE 2192217v2
  - Concept Term
  - C17459 Hispanic or Latino
  - C41222 Not Hispanic or Latino

BUT OMB Ethnicity = Hispanic or Latino

→ Additional question:

- CDE 7748817v1 Person Hispanic or Latino Ethnicity
  - **Concept** **Term**
  - C67118 Central American
  - C107608 Cuban
  - C67117 Dominican
  - C67113 Mexican
  - C17043 Puerto Rican
  - C26277 South American

AND…

NCORP DCP-001 – Screening CDEs

OMB Ethnicity = NOT Hispanic or Latino

Addition question:

- CDE 5216271v1 Patient Geographic Ethnic Group
  - 43 Additional Ethnic groupings – Sampling

Concept Codes	Permissible Value	VM Public ID	
C26272	Africa	2570074	The contine
C42331	African Descent	2794439	Denotes a
C42331:C26320	African Descent/Caribbean	6378835	Denotes a
C42331:C17649	African Descent/Other	6378833	Denotes a
C128457	Americas	5377233	A collection
C26273	Asia	5356418	The world's
C41260	Asian	2572233	A person ha
C41262	Indian	2577286	A person ha
C16311:C48928:C16914	Australia/New Zealand	5377232	The contine
C126535	Australian	5230637	Denotes a
C16352	African American	5021394	A person ha
C26320	Caribbean	5215959	A group of
C126528	Central African	5215962	Denotes a
C126534	Central Asian	5215955	Denotes a
C43391	Chinese	2570992	A person ha

NCORP DCP-001 Eligible but Declined

- CDE 3533849v1 Clinical Trial Participation Patient Declined Reason
- CDE 6369493v1 Patient Declined Participation Travel And Transportation Reason
- CDE 6369506v1 Patient Declined Participation Social Issue Reason
  - AND More!!

Most common reason:

No Desire to Participate in Research

Brenda Duggan, Angela Kummerow, Brenda Maeske, Dianne Reeves DCP-001 NCI Clinical Trials Case Report Form , 5049195v2, NCORP Clinical Trial Screening and Accrual Tool

Using CDEs Helps Improve Research!

- CDEs across sites confirmed known barriers:
  - Ineligibility based on comorbidities, **but also**:
    - Language barriers
    - Financial burdens
    - Logistical issues such as transportation
    - Time off from work
    - Lack of trust of the health care community
- Researchers better able to:
  - Identify characteristics of underserved communities
    - Lower income , large numbers of African Americans
  - Identify places and better ways to reach out
    - Beauty Parlors, Barber Shops, Churches
    - Lab Technicians and Nurses: Educate about Research and Build Trust

CDEs for Research

1.DSRMWS-2971

How Your Information Fuels Good Research

Scientific discoveries **help us better understand health issues** like COVID-19 and other diseases.

But, we can't do it without YOU.

Your information (data) helps us learn how to better help you and others.

Help Scientists, Help Your Community

When scientists add your protected data to information from thousands of others who participate in NIH studies, **powerful research** to find cures and treatments can happen.

Talk Data to Me: How is My Information Safe?

We value your privacy. Some health information (like your symptoms or medicines taken) is kept in a central research database. Information that identifies you is not used in the database—or used in any published studies or meetings. Additionally, data is never re-used without participants' consent.

The data we collect is:

Stored safely in a separate database

Cleaned up for **analyses**

Protected—because we follow HIPAA standards.

Common Data Elements (CDEs)

We can collect your information through **standardized questions**—which are questions that are asked the exact same way they've been asked before.

When questions are asked the same way they've been asked before, it's easier to put data together with other people's data. This creates something called a "dataset," which can be used to better understand diseases and disorders.

NIH Endorsed CDEs
Accelerate Research

Help Scientists, Help Your Community

If we want to know someone's age, we could ask the question several ways. If we ask "How old are you?" the answer is a number. If we ask "When is your birthday?", we can figure out your age from the date. However, when we ask questions in different ways, the information becomes difficult to combine.

With CDEs, researchers ask everyone the same question in the same way, each time. For example, we would ask everyone, "How old are you today?" Asking standardized questions makes it possible to combine data from different studies and use the information to improve public health!

EXAMPLE OF A STANDARDIZED QUESTION:

Q: What is your age?

A: 34 years old

Let Your Voice Be Heard: CDEs and Improving the Nation's Health

The Impact:

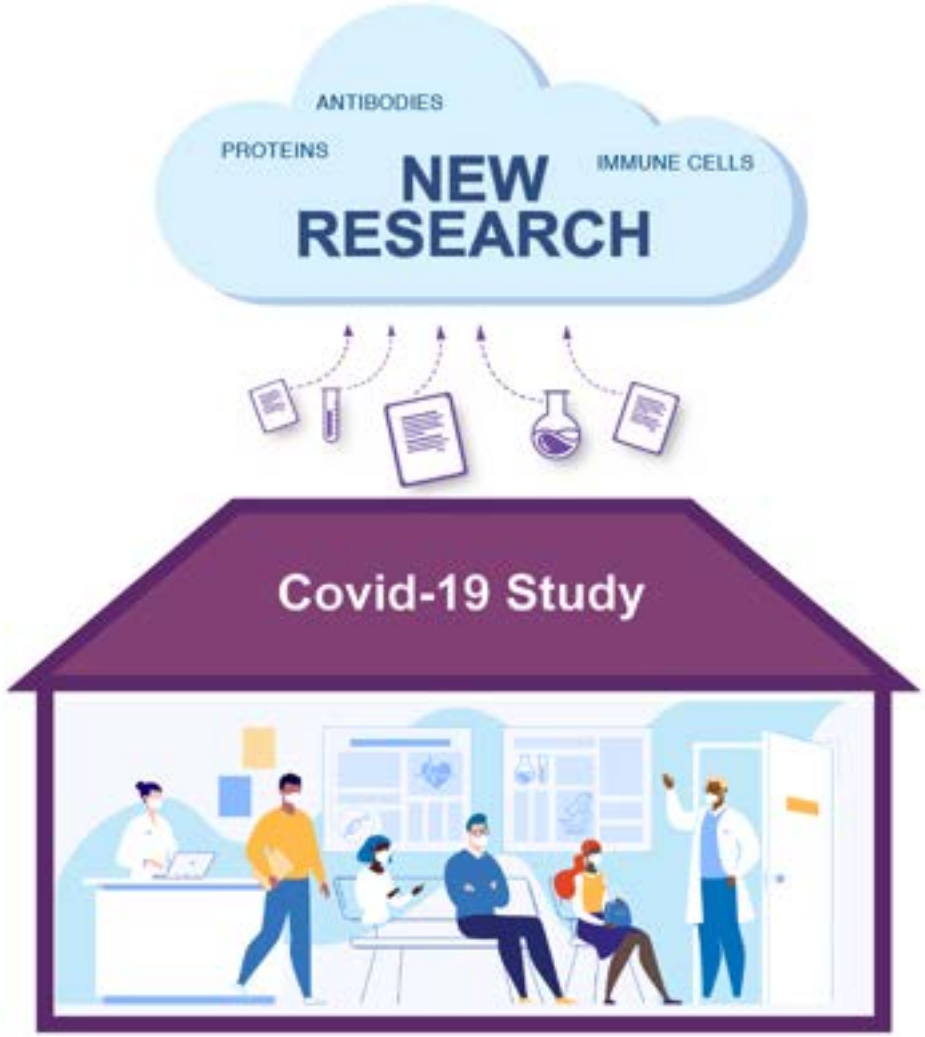
Researchers only request information that's relevant to the study, and unique identifiers (like the participant's name) are protected to ensure patient confidentiality and privacy. Additionally, any data that's collected is never re-used without participants' consent.

Your Data Helps Others!

Your anonymous information makes a big impact on public health—from small communities to large cities.

At the National Cancer Institute, researchers who study **cancer patients with COVID-19** will collect medical information using CDEs.

Future research using the anonymous information can help researchers better understand serious illnesses caused by the virus—and help doctors better treat people with cancer and COVID-19!



Because Researchers use CDEs...

they can more quickly share data and get results faster, which ultimately can help make a **meaningful difference to our nation's health.**

For more information about how CDEs accelerate research discoveries, visit: **cde.nlm.nih.gov/resources**

NIH Endorsed CDEs
Accelerate Research

Key Message

- Using CDEs Helps Improve Research!

Introduction to CDEs

What is a CDE and what is CDE Metadata anyway?

By the end of this section, you will be able to:

- Identify the difference between data and metadata.
- Discuss CDE Metadata
- Discuss how CDE metadata contributes to interoperability between systems.

What is Metadata?

ScHARe

Basic definition:

Metadata is data about data…

More robust definition:

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.

~National Information Standards Organization

CDE Metadata:

Standard and structured information that describes an **individual piece of data, a data dictionary entry**

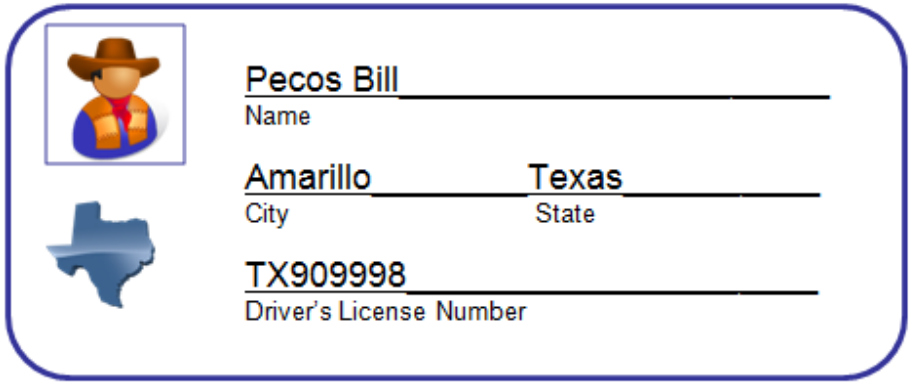
CDE Metadata addressing Data Science Needs

- Designed to:
  - Standardize data → people and computers
  - Enable data sharing across studies
  - Enhance data interpretation and analysis (knowledge)
  - Simplify collaboration
  - Speed study results → Healthcare Practice ⇧ Data available for analysis

- Address the critical problem facing data scientists:
  - Big Data and AI: Requires new approaches for collection, management, and analysis
- Let's look at some examples to clarify…

ScHARe

Metadata Example (1)



Simplified Texas Driver's License

What is the data?

We have a name "Bill Pecos", city "Amarillo", state "Texas", and a driver's license number, "TX909998".

*All are considered pieces of **data.***

What is the metadata?

WHAT are we referring to?

A person's driver's license

HOW we collect data, what data do we need?

Name

City and State

Driver's license number

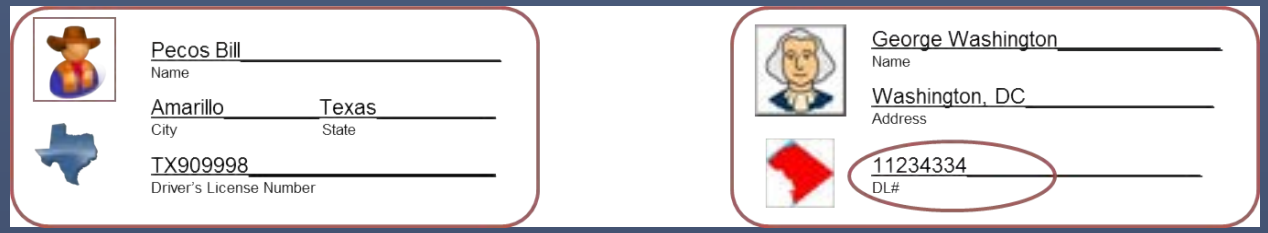
*The labels for the data are considered items of **metadata.***

The metadata is WHAT and HOW the DATA is defined and described.

Metadata Example (2)

Let's take it a step further…

- Different formats for different states
  - WHAT we are describing (driver's license) and the
  - HOW – what data we need (name, city, state, driver's license number)



- Labels may be different

→ What is being collected *means* the same thing.

The data elements are Semantically Equivalent.

Metadata Example (3)

Microsoft Excel Spreadsheet

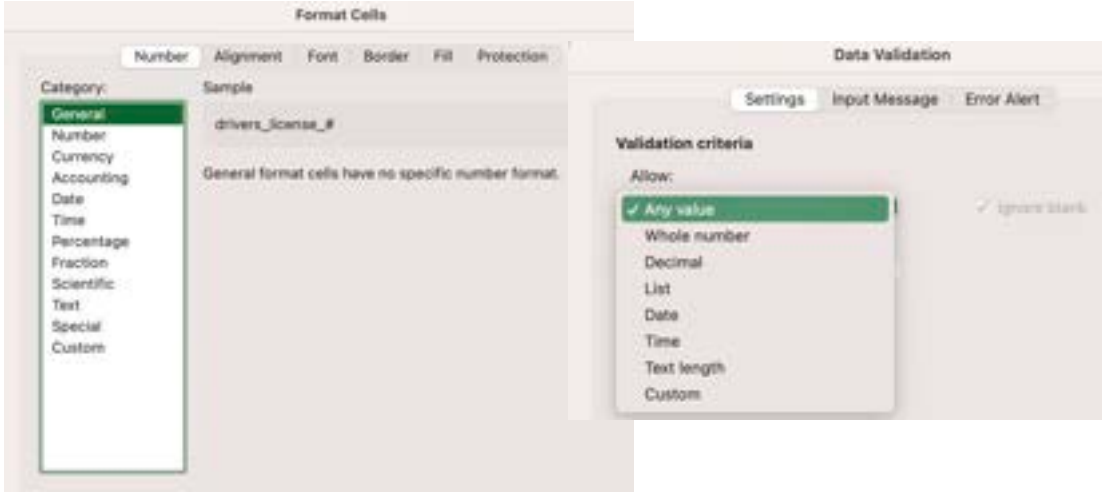
CDE Metadata

Driver's License Data

CDE Metadata

ScHARe

last name	first name	city	state	drivers_license_#
Pecos	Bill	Amarillo	Texas	TX909998



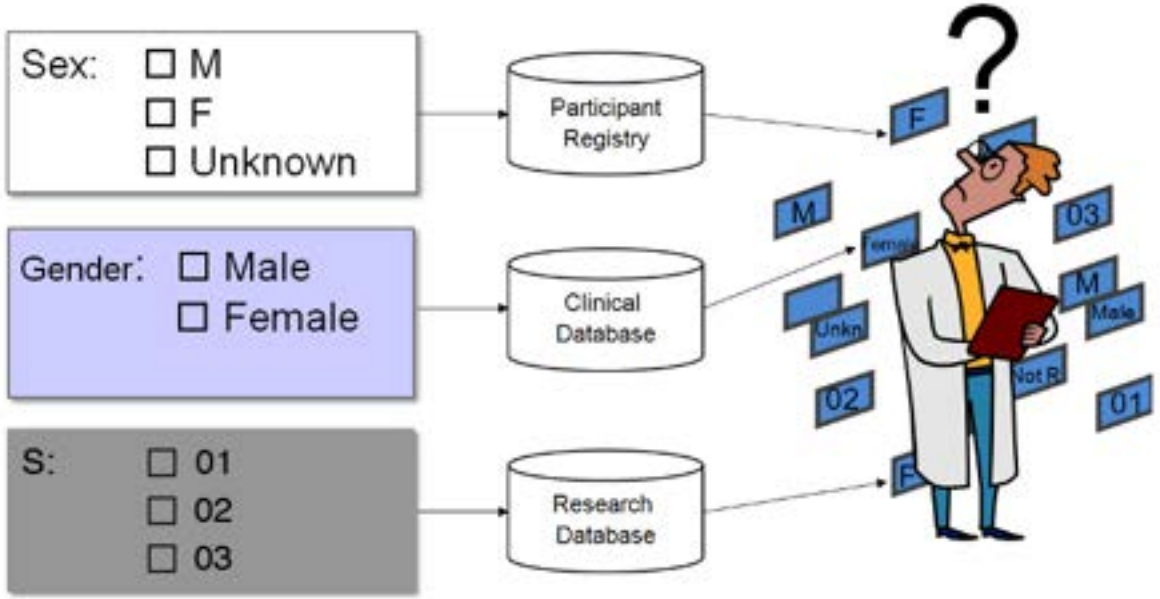
CDE Metadata Highlights

- CDE Metadata → "Data Dictionary"
  - Label or Question Text for data item
    - Name, City, State, Driver's License Number
  - Definition of the data item → "WHAT"
  - Format of the data item → "HOW"
    - E.g., Text, Number, List for Drop Down

CDEs are defined Independent
of any
System or Programming language

ScHARe

A World WITHOUT CDE Metadata



Puts the burden of data reuse on the data consumers.

Interoperability

- “the ability of computer systems or software to exchange and make use of information.”

- Interoperable data systems require several key elements:

  - Clearly expressed meaning → shared meaning

  - Understand the data → format and restrictions

Can CDEs Help?

- Problem → Make data interoperable for aggregation and analyses

- Solution → Metadata for humans and computers for data collection

Ensures Data is Understandable, Usable, and Interoperable

Primary CDE Metadata Elements

1. Standard Structure → predictable format
   - Makes CDEs easily used by humans and computers
2. Standard Terminology → shared meaning
   - Makes the meaning of data clear and more easily reused

These two (2) CDE elements facilitate interoperability and will be discussed in detail later

Key Messages

- CDEs are Metadata
- CDEs have Standard Structure
- CDEs use Standard Terminology

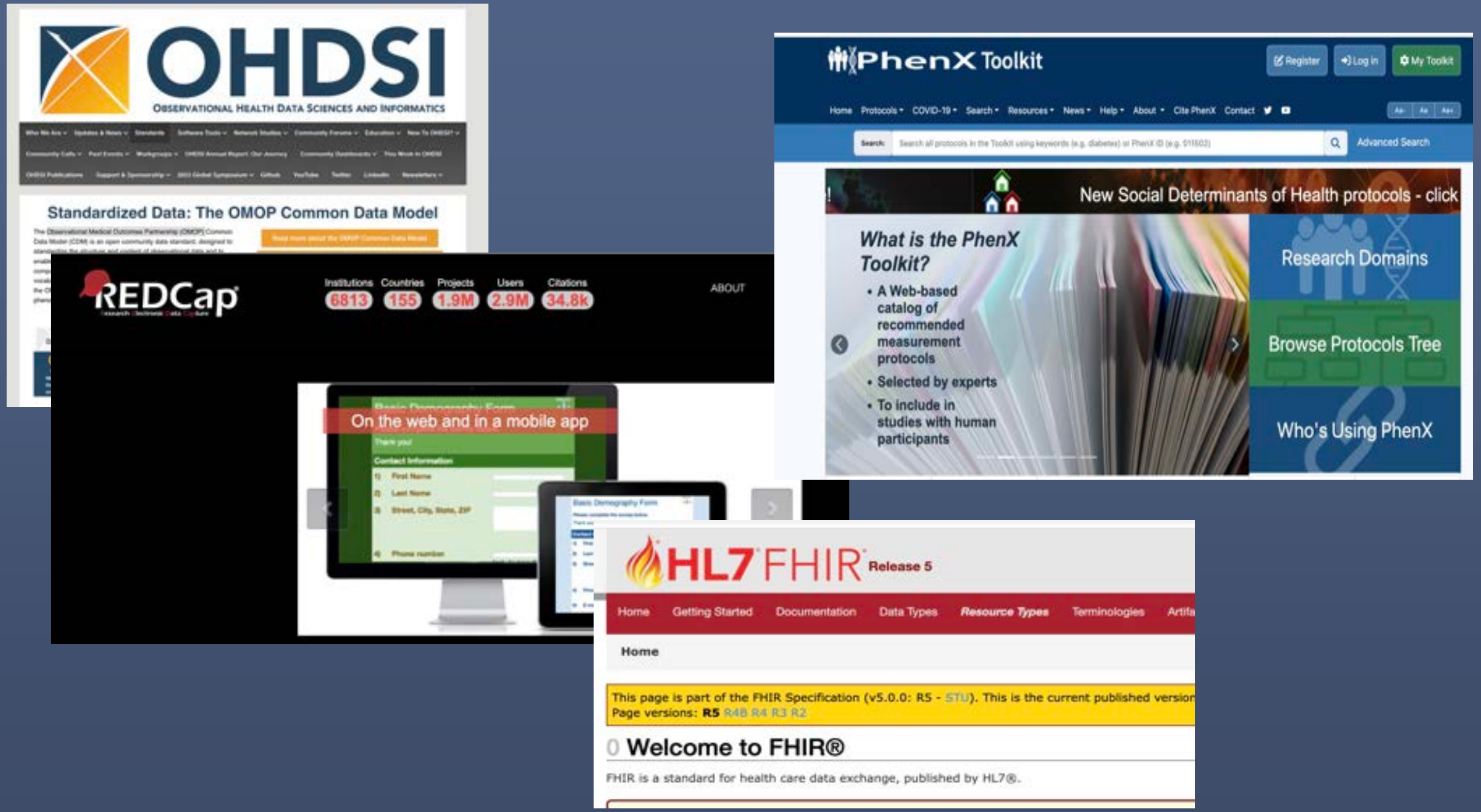
Increasing Awareness

Have you heard of…

By the end of this section, you will be able to:

- Demystify some widely recognized healthcare initiatives
- Understand the differences:
  - “Instruments”
  - Data Collection Tools
  - Data Models
  - Data Exchange Standards

These resources are all freely available
**May require joining the collaboration and licensing*


OHDSI
Observational Health Data Sciences and Informatics
Standardized Data: The OMOP Common Data Model
REDCap
Institutions 6813
Countries 155
Projects 1.9M
Users 2.9M
Citations 34.8k
ABOUT
On the web and in a mobile app
PhenX Toolkit
Register
Log in
My Toolkit
Home Protocols COVID-19 Search Resources News Help About Cite PhenX Contact
Advanced Search
New Social Determinants of Health protocols - click
What is the PhenX Toolkit?
• A Web-based catalog of recommended measurement protocols
• Selected by experts
• To include in studies with human participants
Research Domains
Browse Protocols Tree
Who's Using PhenX
HL7 FHIR Release 5
Home Getting Started Documentation Data Types Resource Types Terminologies
Home
This page is part of the FHIR Specification (v5.0.0: R5 - STU). This is the current published version
Page versions: R5 R4B R4 R3 R2
0 Welcome to FHIR®
FHIR is a standard for health care data exchange, published by HL7®.

PhenX Toolkit

- NIH National Human Genome Research Institute (NHGRI)
  - >500 Measures, 24 Research domains
- Protocols Created by Consensus
- "**Instruments**"
  - Assess or measure a condition or problem

- Questions asked together, as a "bundle"
- No definitions
- No concept annotations
- Labels human readable
- Designed to be easily analyzed

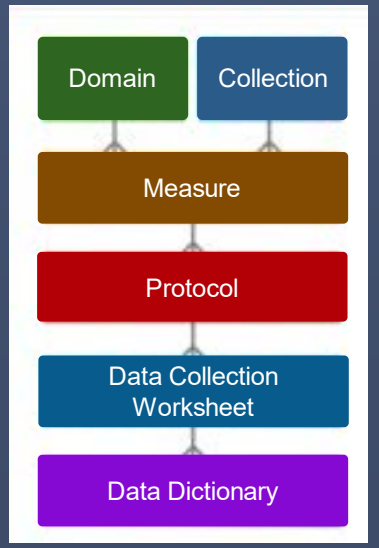
Substances Lifetime Use - 9 questions

Child's Lifetime Household Composition – 85 questions

"Collections" are a combination of Measures to Form a Protocols

Population Characteristics – Demography – 13 Protocols

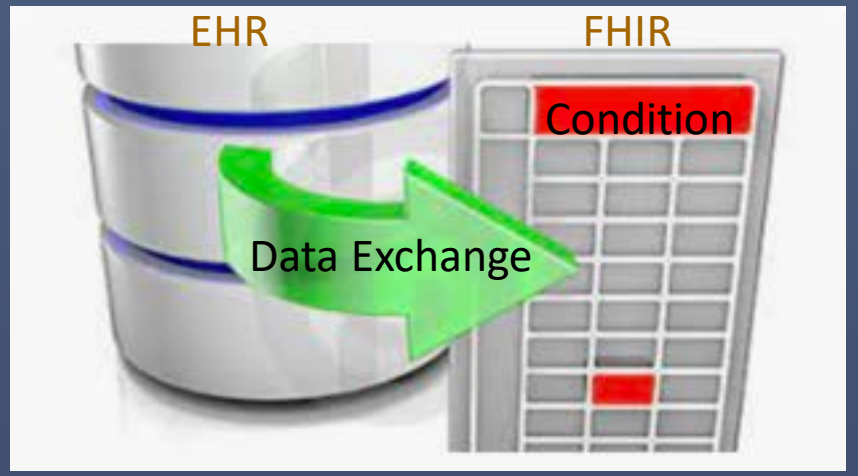
(many more than 13 questions)



About the PhenX Toolkit: https://www.phenxtoolkit.org/about

FHIR®

- Fast Healthcare Interoperability Resources (FHIR)
- Health Level Seven (HL7) format for **Healthcare Data Exchange**
- Common format for Electronic Healthcare Records (EHRs)
- Predefined "Resources" → buckets/fields for EHR data
  - Allery Tolerance (~25 fields)
  - Condition (~30 fields)
  - Person (~30 fields)
- Do not specify specifically HOW data is collected
  - "Whatever is in the EHR"

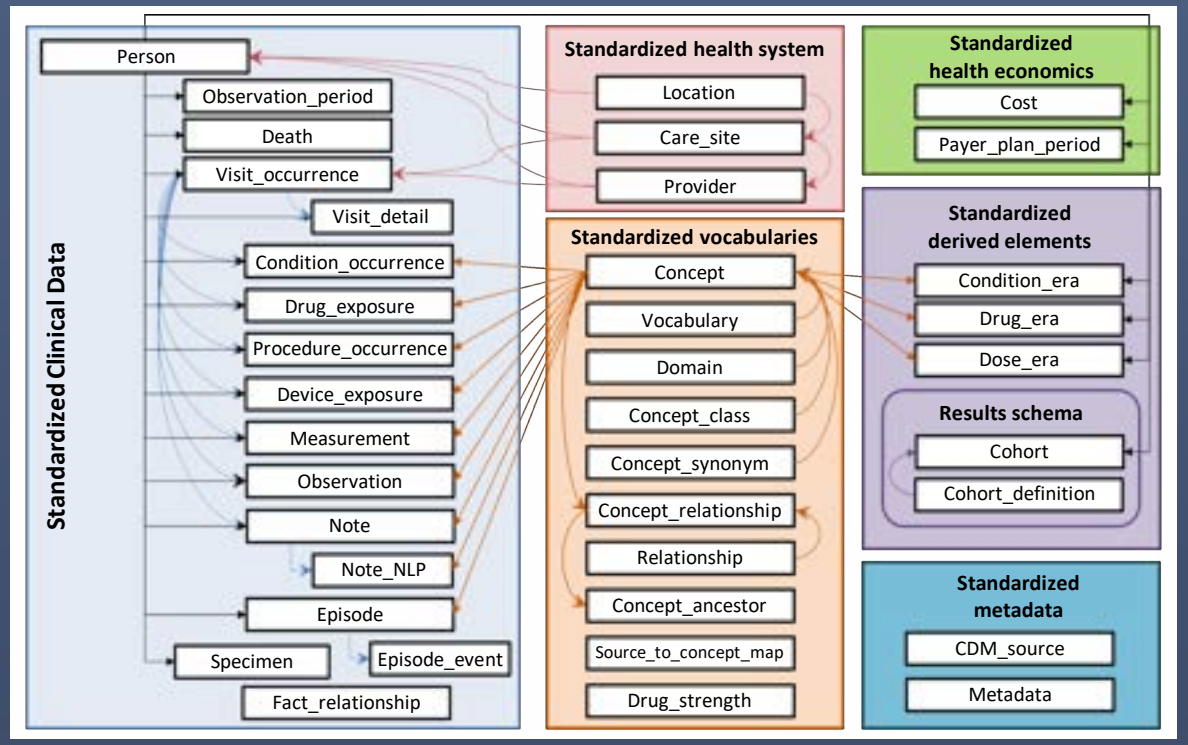


Used for Data Exchange

Welcome to FHIR! https://www.hl7.org/fhir/index.html

ODHSI/OMOP/Athena

- Observational Health Data Sciences and Informatics (ODHSI pronounced “Odyssey”)
  - A collaboration using a Common Data Model (CDM) and tools
- Observational Medical Outcomes Partnership (OMOP)
  - **Data Model** for data extracted from EHRs
- Athena
  - Sometimes referred to as the “OMOP terminology”
  - **Not a Standard Terminology**
  - Repository for terms and a search engine for the OMOP
  - OMOP Concept identifiers map to 100’s of standard terminologies



ODHSI: https://www.ohdsi.org/data-standardization/

REDCap®

- Research Electronic Data Capture (REDCap)
  - Software Developed by Vanderbilt University
  - Design and build forms → Manage on-line study **Data Collection**
  - Data entered into the system or imported from EHR
  - Questions/answers not individually reusable across studies
  - Whole studies can be copied, exported, and shared with other REDCap systems

  - *Note: CDEs can be imported into REDCap via DD format which is a CSV file

REDCap Learn More! https://projectredcap.org/software/

Key Message

- Don’t be Befuddled, Be Informed!

Knowledge Poll #2

Healthcare Initiatives

Knowledge Check

1. PhenX Toolkit Instruments
2. REDCap
3. FHIR
4. ODHSI/OMOP/Athena

_3_ Healthcare Data Exchange Format

_1_ Questions to be used as a “whole” to measure something

_4_ Data Model and Tools

_3_ Data Collection Tool

Terminologies and CDEs

What do you mean?

By the end of this section, you will be able to:

- Discuss the role of standard terminology and CDE metadata.
- Discuss the importance of standard terminologies in achieving interoperability.
- Identify NIH Terminology and CDE resources
- Discuss how terminology mappings enable harmonization, data transformation, and interoperability.

Semantics

ScHARe

Basic definition: The meaning of a word, phrase or sentence

More robust definition: A branch of linguistics and logic concerned with the meaning based on how a is sentence structured, including social and cultural context and relationships between words impact understanding.
Formal Semantics are expressed in a mathematical way using symbols.
Note: Ontologies are often expressed using formal semantics, which makes them different and often more precise than other kinds of terminologies.

CDE Semantics: The expression of meaning in a **Standard, Structured** way using terminology concepts.

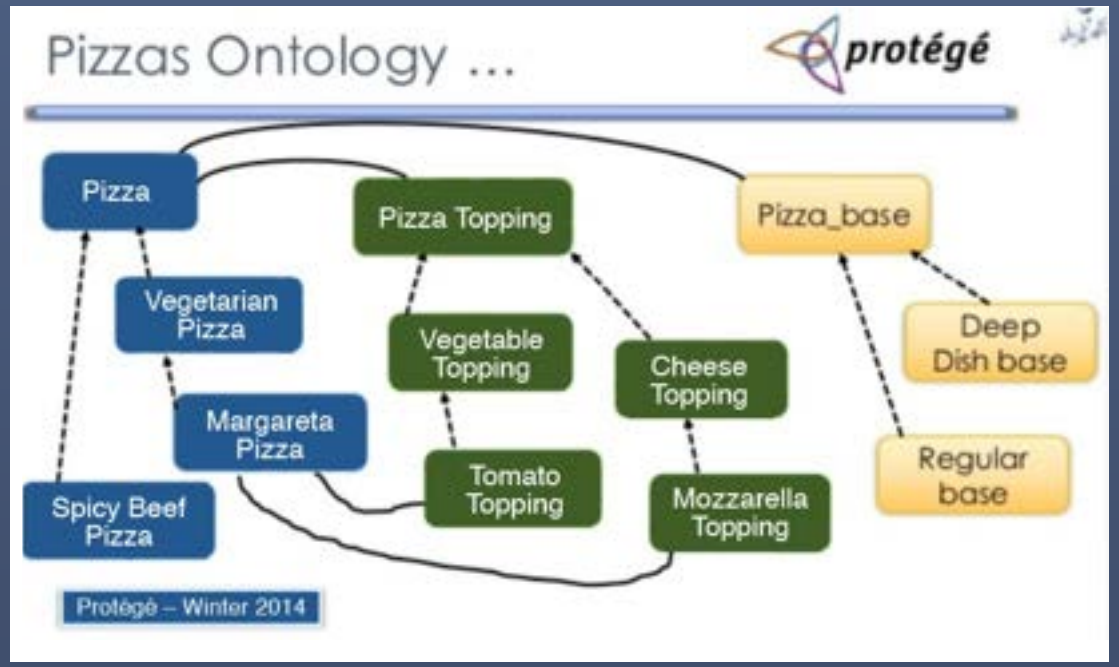
Standard Terminologies

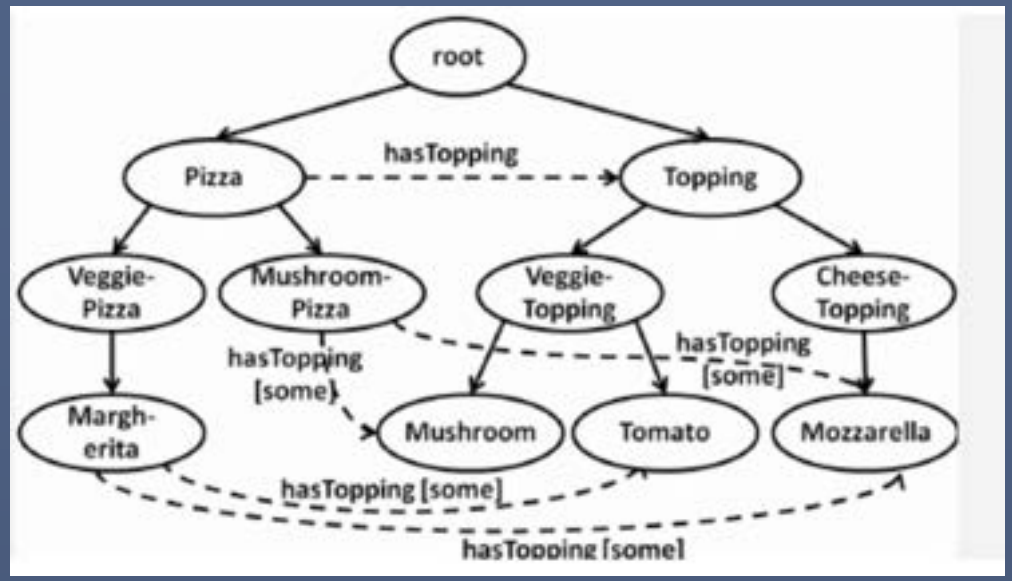
- Vocabulary for a specific field of study, or Context
  - Describes the terms used for a given profession
- Provide **language independent unique identifiers** (“codes”)
  - Ensures humans and computers → attach the same meaning
- Provide consistency and clarity
- Often include text definitions and synonyms

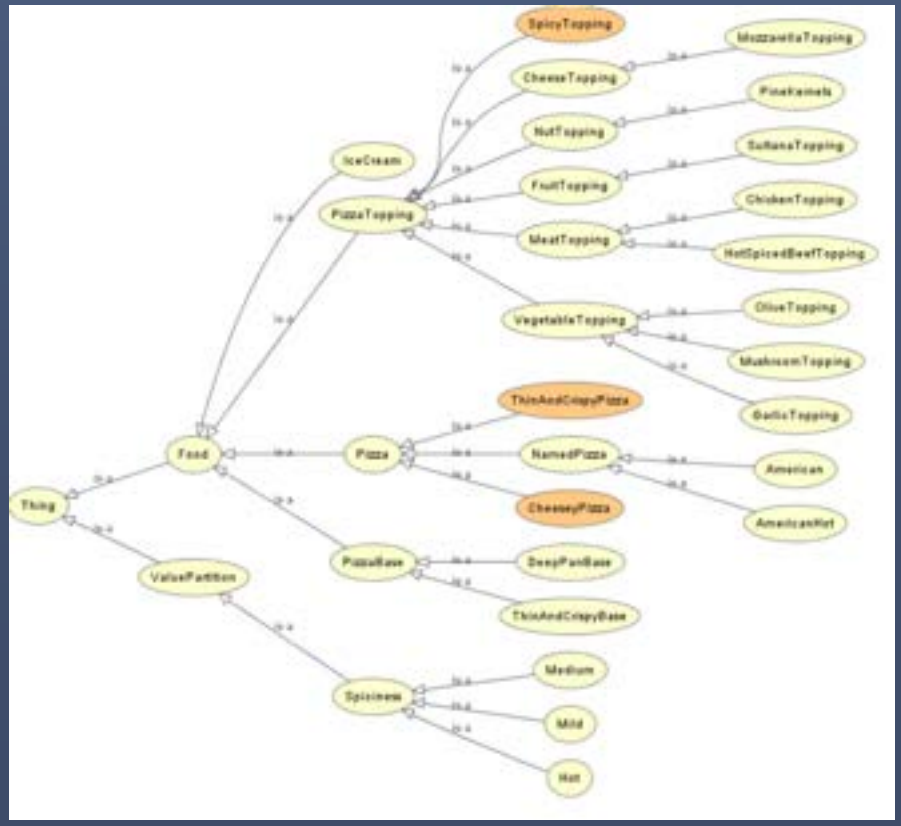
Ontologies

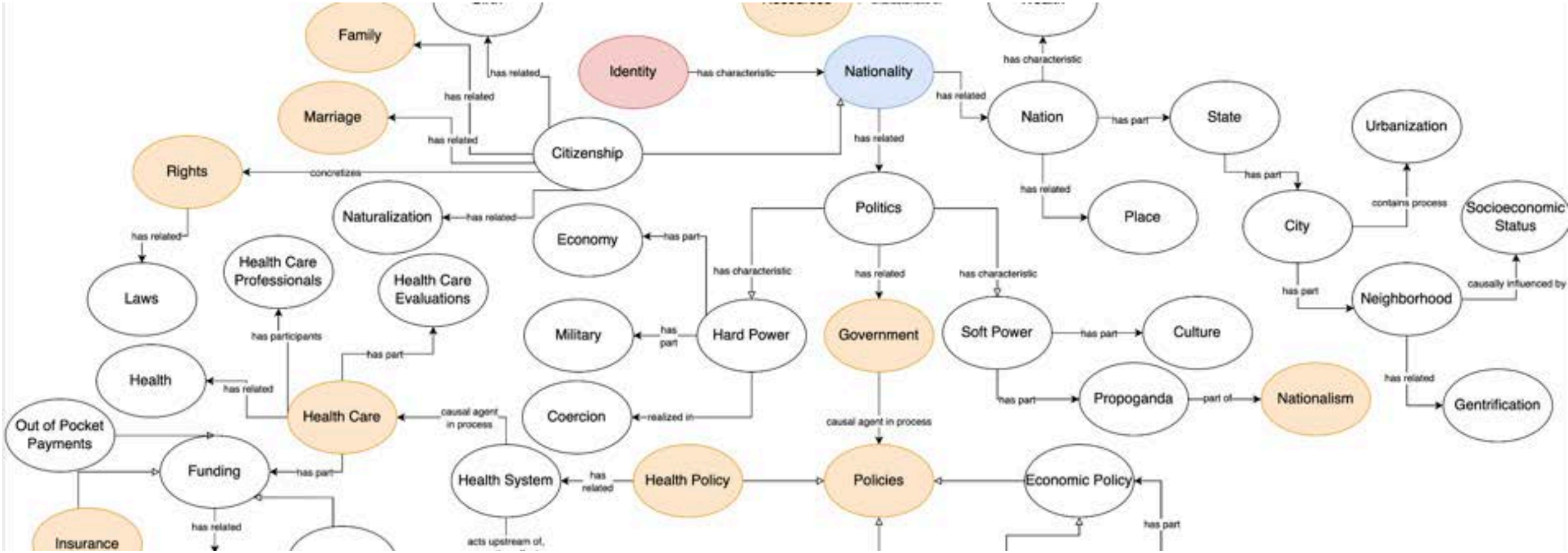
- Vocabulary that represents **Knowledge** (versus Terms)
  - Specialized languages and tools
    - Languages: OWL (Web Ontology Language), Description Logic (DL), Common Logic (CL)
    - Tools: Protégé, SWOOP, NeOn Toolkit
- Specific to a domain or subject areas
  - Pizza, Nationality, Disease
- Creates a shared understanding
  - Many ontologies for the same subject area
  - Varying levels of detail
- Focuses on relationships between concepts
  - Includes relationships such as "is a" "part of" "related to" "causes" ”plays a role in”
- Relationships support reasoning over data, or “figuring out” things
  - For example “Is this data a Pizza Topping or a Pizza Base?”
- Source of biomedical ontologies
  - BioPortal developed by Stanford
    - https://bioportal.bioontology.org/ontologies
  - Most of these ontologies share an “upper ontology” Basic Formal Ontology (BFO)

Examples of Pizza Ontologies









Example Nationality Ontology

- Talks about the characteristics to look at to help classify someone into an particular nationality
- Not a “list” of nationality terms

**UT Health https://github.com/UTHealth-Ontology/CRENO/blob/main/doc/nationality.ng*

What Do You Mean?

- Context is important in conveying meaning
  - Words have different meanings depending on words around it.
- Some examples:
  - **Agent:** chemical compound or government employee?
  - **Alcohol:** disinfecting or drinking?
  - **Colon:** sentence punctuation or biological organ?
  - **Mole:** animal, blemish, unit of measure, or spy?
  - **Probe:** examination, investigation, or instrument?

→ The above words are **SEMANTICALLY AMBIGUOUS**.

Words can mean different things in different contexts.

How do Standard Terminologies help Data Harmonization?

- Define meaning within healthcare and research contexts:
  - Accelerate common understanding → information sharing.
- Facilitate research:
  - Accelerates discovery → Big Data and AI
- Support integration of diverse data:
  - Validity of cross study data → comparison and analysis
  - Improve evaluation → efficiency, safety, and efficacy
- Improve link between Clinical Healthcare and Clinical Research:
  - Increase accuracy and speed of data collection
  - Enable querying across data from different settings

Key messages:

- Words can have different meanings
- Use Standard Terminologies for clear, shared meaning

Some NIH Terminology and CDE Resources

Freely Available

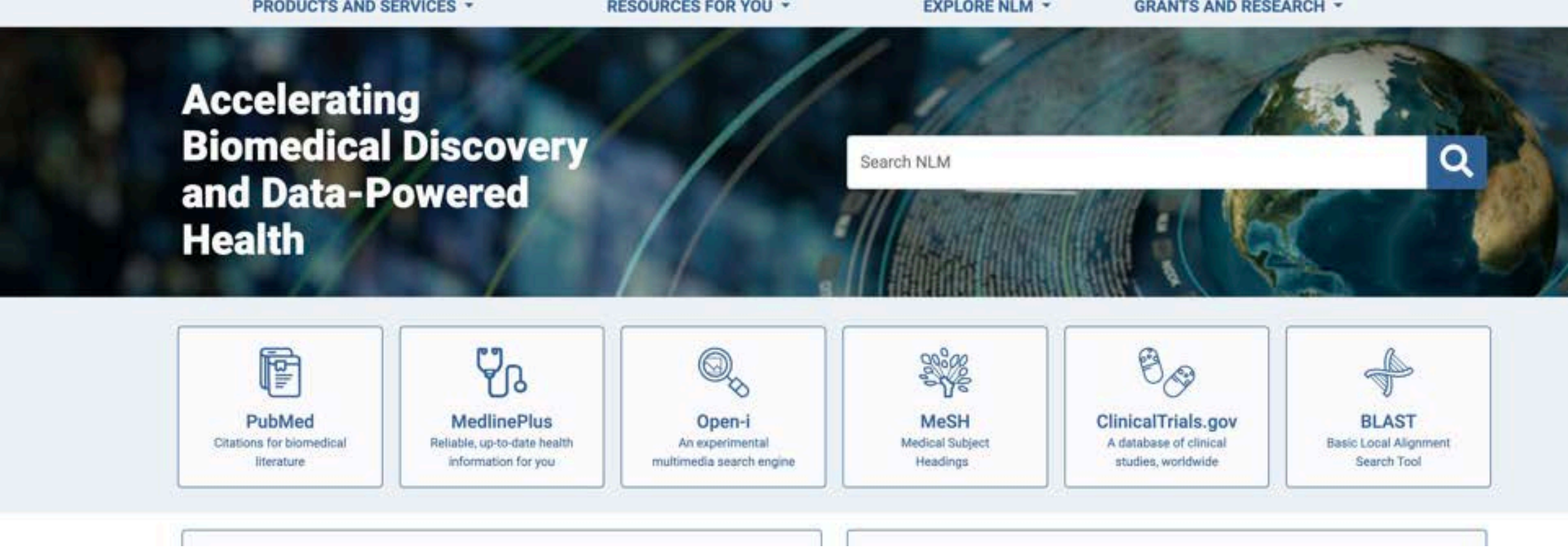
Terminology and Standardized CDEs for Interoperability

How do CDE metadata and Standard Terminologies support interoperability?

- Remember →CDE Metadata is “Standard” and “Structured”
- Based on a Standard – ISO/IEC 11179 Metadata Registry.
  - Specifies the Structure for CDEs
  - Specifies exactly ***how*** to use standard terminology for semantics
  - Makes CDEs usable by humans *a*nd computers.

The *Standard* CDE *Structure* for representing semantics is the foundation for semi-automated semantic interoperability.

- *A little more about this later*



National Library of Medicine (NLM)

- "World's largest biomedical Library"
  - Public resource
- Familiar resources including PubMed, MeSH AND….

NLM Unified Medical Language System (UMLS)

- Files and software that integrate and distribute health and biomedical terminologies and standards
- Database containing cross-terminology mappings (185 terminologies)
- **Assign Concept Unique Identifiers (CUIs) e.g. C0018681**

NLM UMLS: https://www.nlm.nih.gov/research/umls/index.html

NCI Enterprise Vocabulary Services (EVS)

- Terminology Services for NCI and other NIH Institutes and Centers (ICs)
  - Develop new concepts → unique identifiers and definitions
  - Record concept relationships → **scientific evidence**
  - Facilitates standardization across NCI and the larger biomedical community
- Two terminology products:
  - NCI Metathesaurus (NCIm)
  - NCI Thesaurus (NCIt)

Use of common terminologies are a key component of CDE Metadata.

NCI Thesaurus (NCIt)

- Cancer Terminology and Thesaurus
  - Definitions, synonyms, concept relationships, and terminology mappings
  - For some concepts, includes ontology-like relationships
- Broad coverage of cancer research and clinical terms
  - Clinical Terms are common across many diseases and disorders.
    - *Blood Pressure*
    - *Diagnosis*
    - *Laboratory Test*

Ontology-like Features

OBO Relation
Casually related to
Correlated with
Developmentally related to
Has phenotype or disease
Participates in
Related via evidence or inference to
Temporally related to

NCI Thesaurus (NCIt)

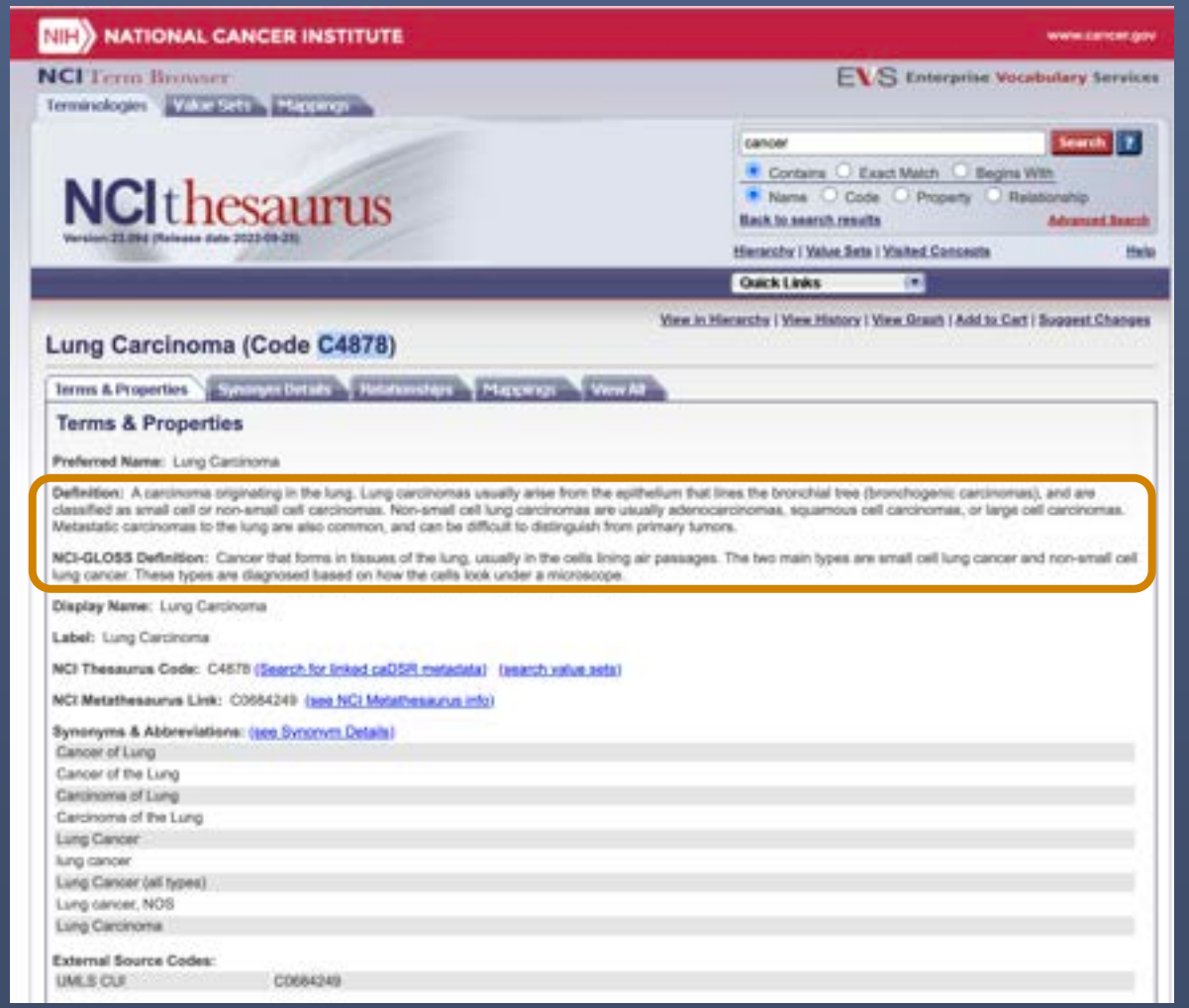
- Unique concept code → “C-code”
  - E.g. C16977
- Unique concepts → unambiguous meaning
- Linked to the NCI Metathesaurus via the UMLS CUI
- NCI Concept Codes provide critical linkage to a specific meaning
  - *Easily compared by computers to identify equivalent meaning regardless of the “words”*

NCI Metathesaurus (NCIm)

- Subset of the NLM's UMLS Metathesaurus
  - Terms needed/used by NCI and other NIH Institutes and Centers
  - Terminology Mappings validated by EVS specialists
- Additional cancer-specific terms
  - ➢Sent to NLM for inclusion in UMLS CUI
  - ➢Mapped to other terminologies

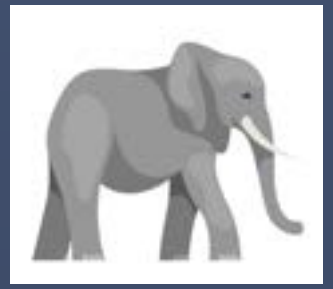
Example of NCI Thesaurus Concept Information – Multiple Definitions

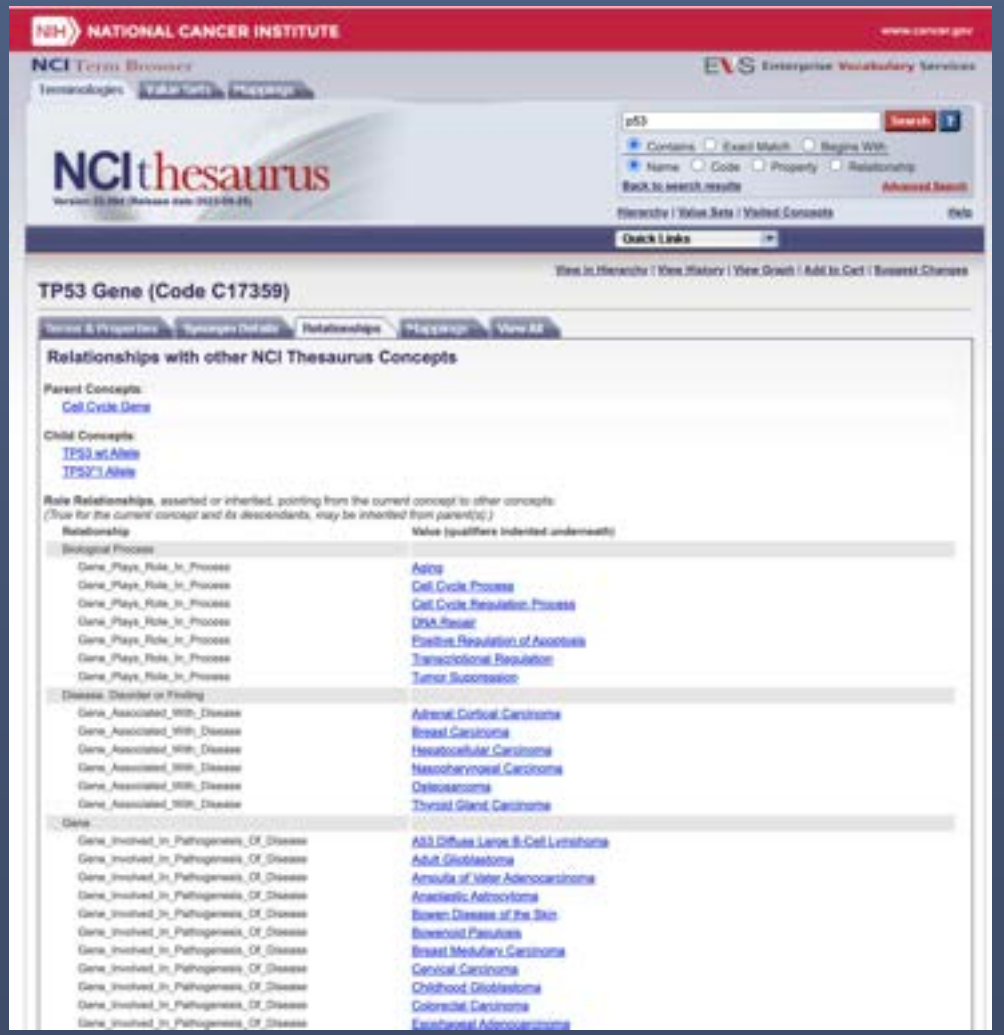
- "Lung Cancer" Code C4878
- From different NIH sources
  - Example of two kinds of definitions:
    - Definition – Technical → for Researchers
    - NCI-GLOSS Definition → Public



Example of NCI Thesaurus Concept Information – Ontology-like Features

- "TP53 Gene" Code C17359
- Concept Relationships
  - Gene_Plays_Role_In_Process
  - Gene_Associated_With_Disease
  - Gene_Involved_In_Pathogensis_Of_Disease
  - Gene_Has_Abnormality
  - Gene_Found_In_Organism





Example NCI Thesaurus Information - Mappings and Synonyms

- "TP53 Gene" Code C17359
  - Mapping Details
  - Synonym Details
    - Term and Source

NCI Metathesaurus (NCIm)

- Uses the UMLS Concept Unique Identifier (CUI) → Mapped to other standard terminologies

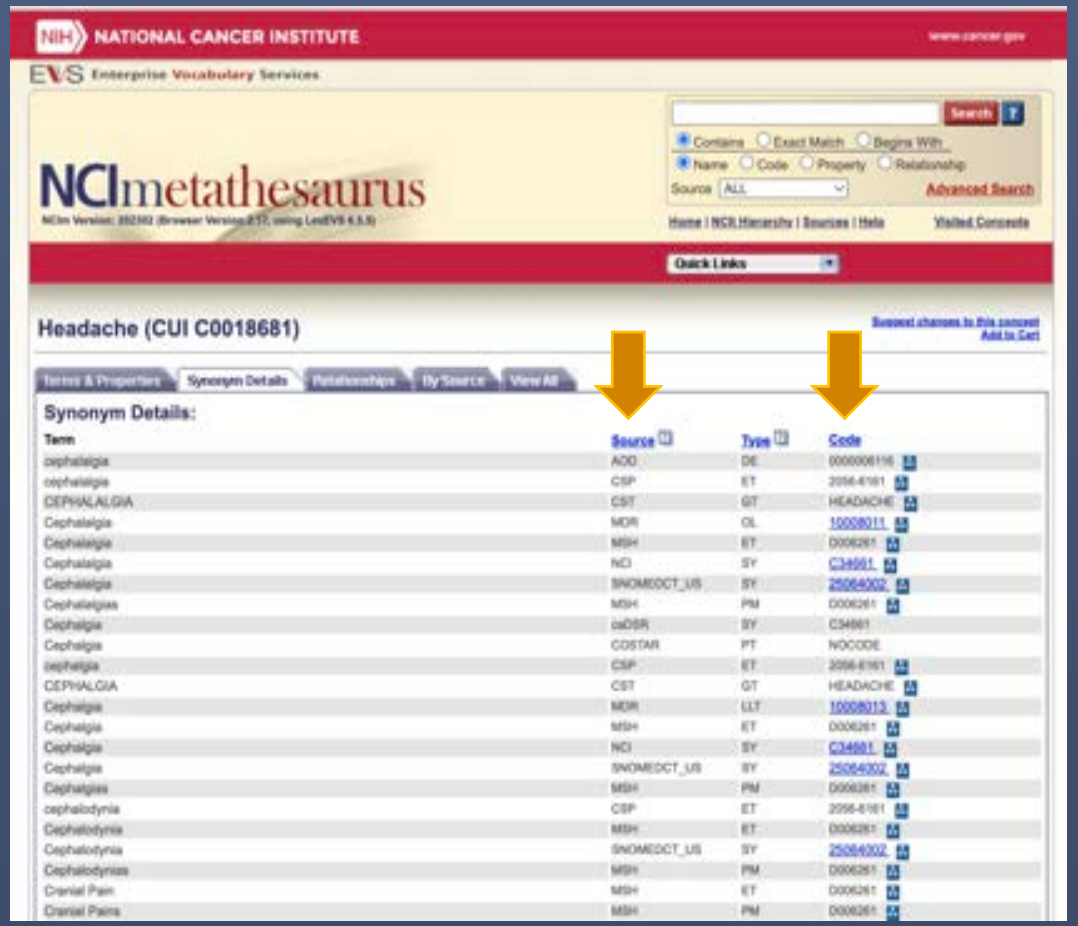
- Additional kinds of concept relationships like broader, narrower

- Provides access to additional definitions from different sources

Leveraging NCI Thesaurus Concepts and NCI Metathesaurus Mapped Information

- Computers can use this information to discover terminology mappings supporting data transformation and aggregation

How does this work?



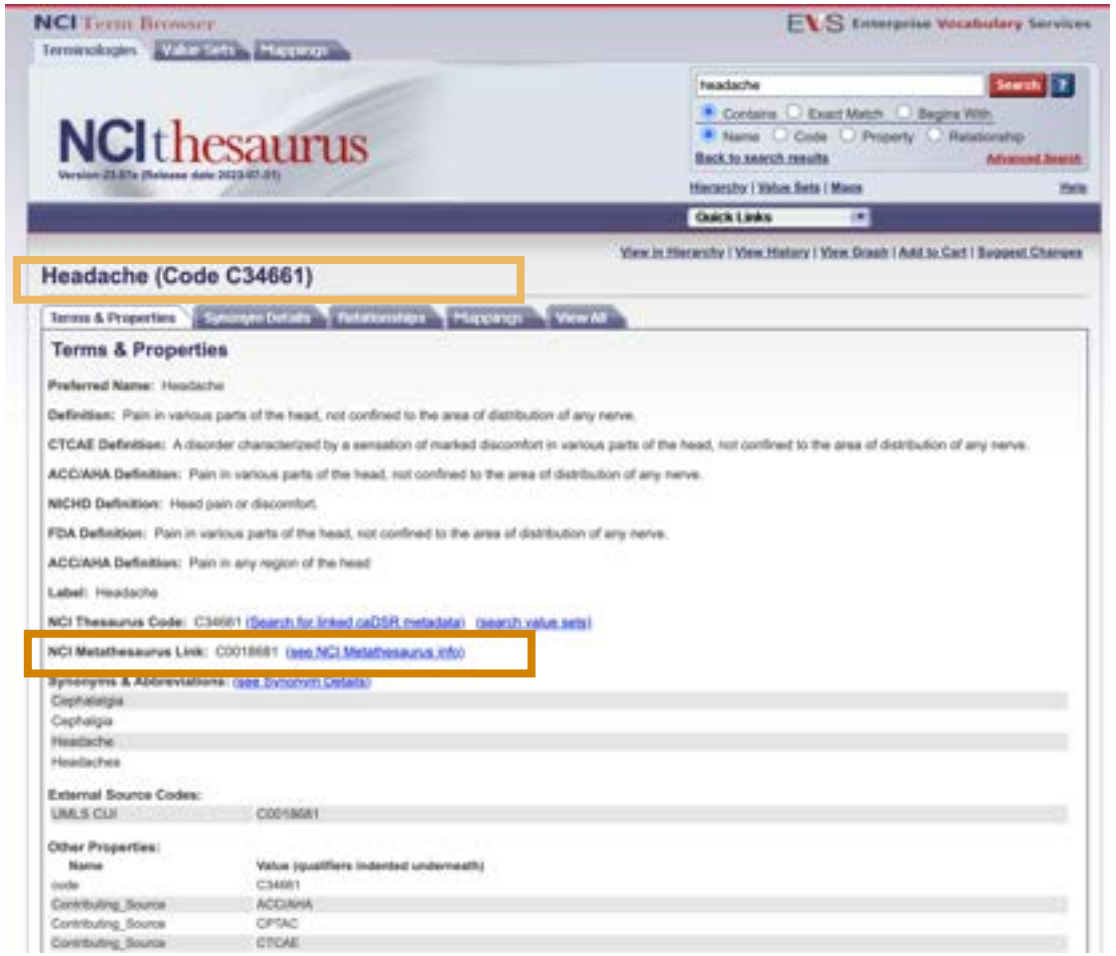
Example Using NCI Thesaurus Concepts and NCI Metathesaurus Mappings

1. NCI Thesaurus Concept: Headache

   NCI Code C34661

2. Click the link to the NCI Metathesaurus:

   UMLS CUI C0018681



NCI Metathesaurus Terminology Mappings

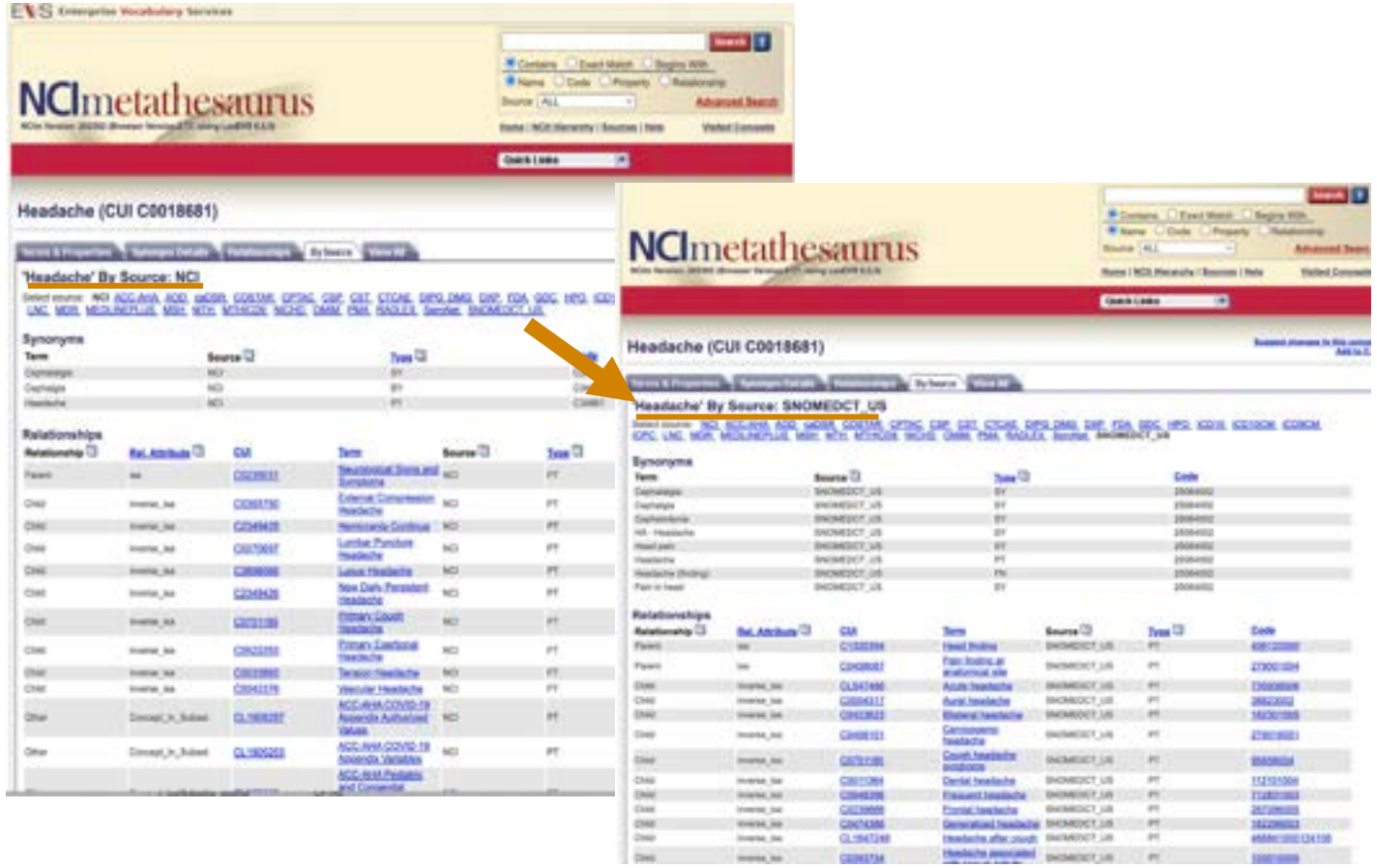
3. View “By Source” showing all the Source Terminologies that are mapped to CUI C0018681

   (Default is the NCI Source Concept Information)

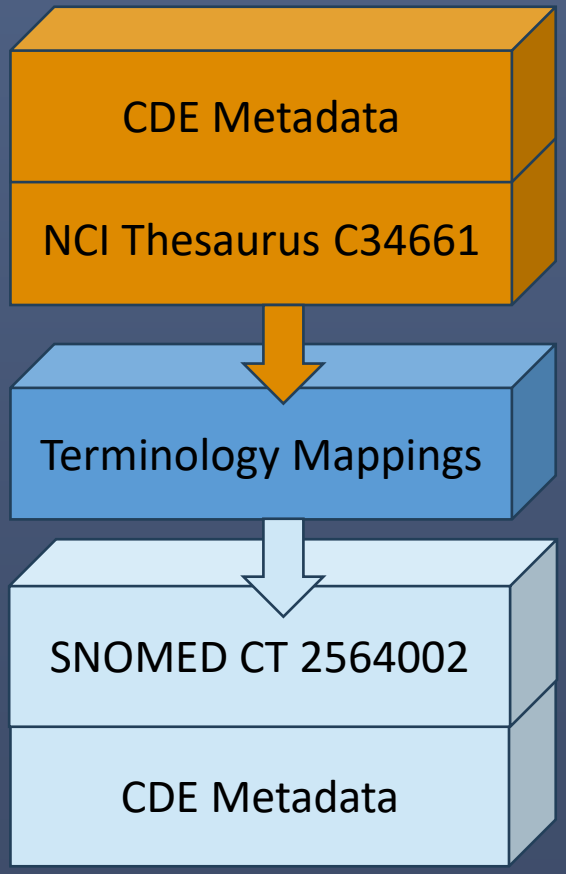
4. Select a new “Source”

   ✓ SNOMED CT US

5. View the SNOMED CT US Source Concept information



Importance of Concept Code Mapping and Interoperability



- CDE **unique concept codes** represent data semantics
- Mapping enables interoperability even if the same standard terminology was not used in another CDE
- CDE Metadata enables searching for concept codes across CDEs to compare data

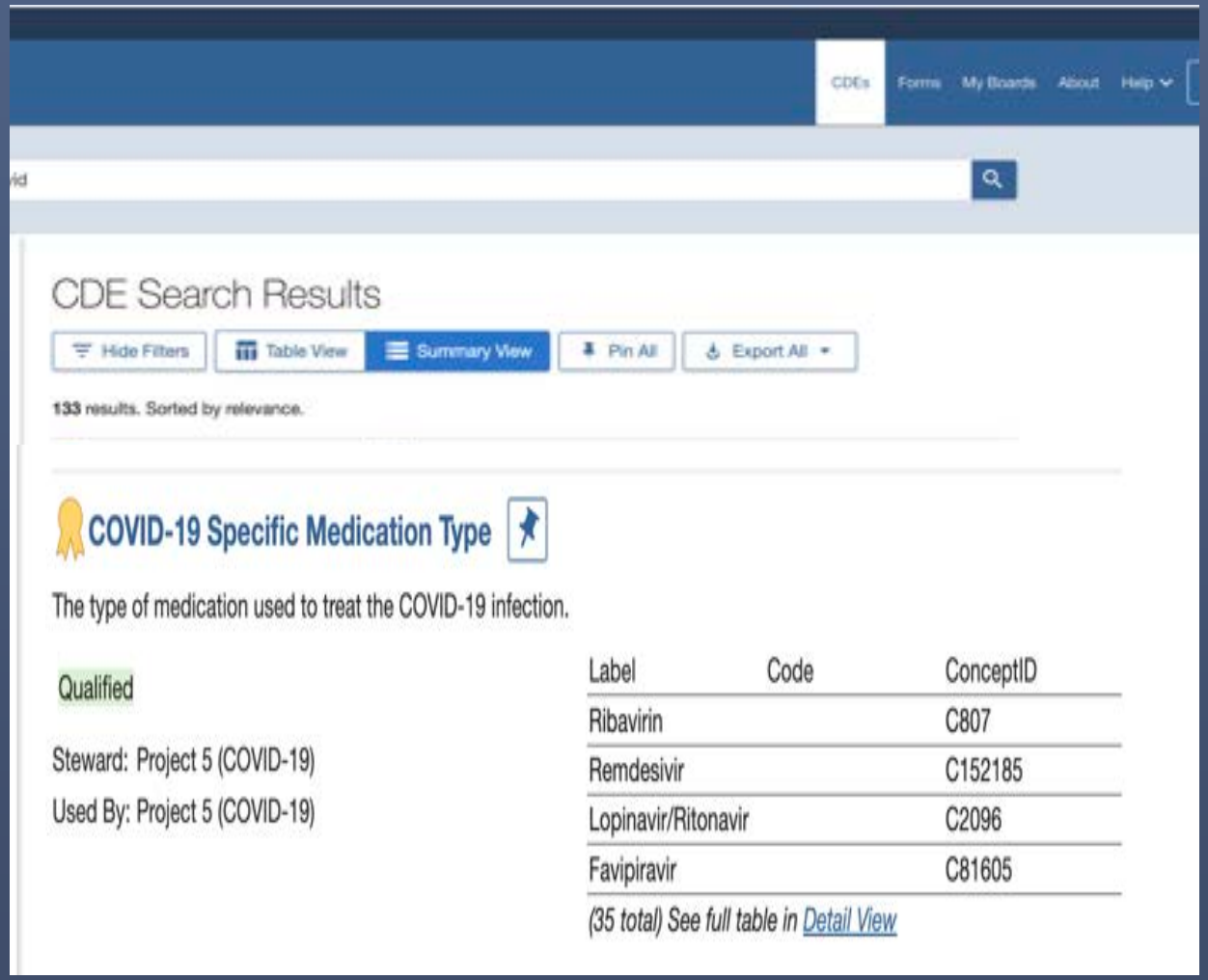
NIH CDE Repository NIH Recommended CDEs

- Reviewed and Approved
- Meet the criteria for Standard, Structured CDEs
  - ✓ Name and textual definition.
  - ✓ Concepts that define the meaning for the *What* and the *How* (for any drop-downs)
  - May include rules to use of specific terminology → LOINC or SNOMED-CT.
- **Designed to be used individually to create data collection forms or data base designs across studies.**

When the same CDEs is used in different studies the data is interoperable.

Example of NIH Endorsed CDE

What?



How?

Drop-Down

Key Messages

- Use Free NIH Standard Terminology Resources!

- NCI Thesaurus – Trusted source for Concepts
- NCI Metathesaurus – Trusted source for Mappings across Terminologies

Pulling It Together

Standard and Structured

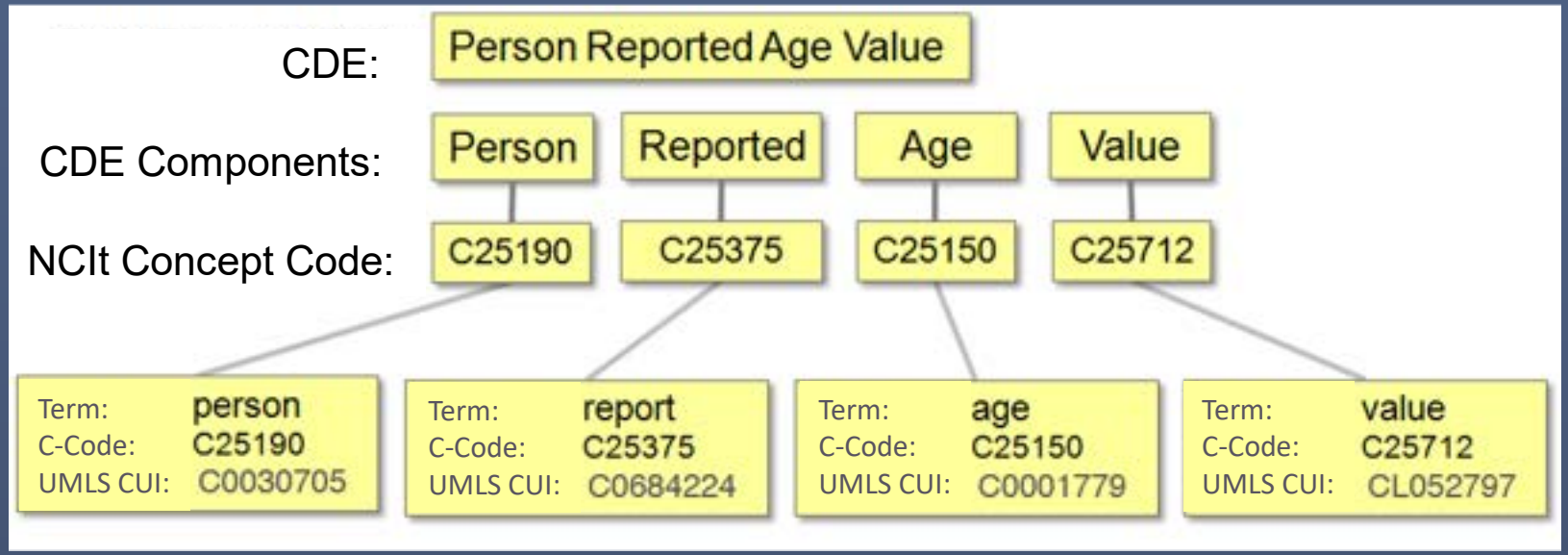
By the end of this section, you will be able to:

- Discuss the basic structure of CDEs
- Discuss the importance of standard, structured CDEs to support computable semantics.

Pulling It Together

- Standard, Structured CDEs using Standard Terminology →foundation of CDE Metadata and CDE semantics
- Enabling CDE Components:
  - Specific Standard Structure for the CDE (ISO/IEC 11179)
  - Unique Terminology Identifiers or Codes
- Why is this so important?
  - Unique Terminology Identifiers in a specific and standard structure makes the meaning independent of words, it makes the meaning *computable*.
  - In other words, computers will be able to find interpret the CDE metadata, rather than always inserting a person into the process.
- Let's take a last look at the CDE structure…

CDE Structure and Concepts - Example



- Term: the word in the terminology
- C-Code: in this case →NCI Thesaurus concept code
- UMLS CUI: the Unified Medical Language System Concept Unique Identifier

Organizing terminology concepts in a specific order is how we *structure* the CDE meaning

CDE Structure and Concepts

- The CDE on the previous slide is broken down into different terms in a specific order → it's "structure".
  - Each Term represents a concept in a standard terminology = *WHAT the data means*
  - Term names → Human readable
  - CUIs and Concept Codes → Comparable by Computers

CDE Structure and Concepts

- Labels can be specific to a community
  - Even if we MEAN the same thing
  - Concepts synonyms support label variation
- Using CUIs and Concept Codes ensure the same meaning
- Standard, Structured CDEs help computers establish when CDEs mean the same thing, regardless of the words or labels.

One more thing about Concepts and CDEs

Does it really matter?

Are These Data Fields Collecting the Same Data?

- Consider the following:
  - Form 1: Person Reported Age:____
  - Form 2: Participant Reported Age:____

- Note: Labels are different but are they asking for the same data?
- Answer: Yes and No.
  - Both ask for an individual's completed years since time of birth
  - Form 1 the individual → a human being
  - Form 2 the individual → someone who takes part in a study

  The difference in meaning of these concepts tells the data consumer more about the data

Why does it matter?

The more precise and understandable the CDE concepts are, the more useful the CDE metadata is to others

Key messages

- Foundation of Interoperability is CDEs
- Foundation of CDEs are:
  - Standard Structured format
    "What" and the "How"
  - Standard Terminology Concepts

Knowledge Poll #3

Pulling it Together

Knowledge Check

1. Which of the following is NOT a standard terminology?
    A. Wikipedia
    B. NCI Thesaurus
    C. NCI Metathesaurus
    D. Athena
    E. A and D (Correct Answer)
2. When concepts are used to define the What and How of data, they form a _____.
    A. Standard CDE (Correct Answer)
    B. Standard Term
    C. Concept Identifier
3. Which of the following is more important to consider when determining semantic equivalency?
    A. The label or what you call something
    B. The CDE concepts (Correct Answer)

Success Stories

CDEs in Action

Success Story #1: CDEs helping reach underrepresented population

CDEs in action

Success Story #2: CDEs help harmonize and transform data

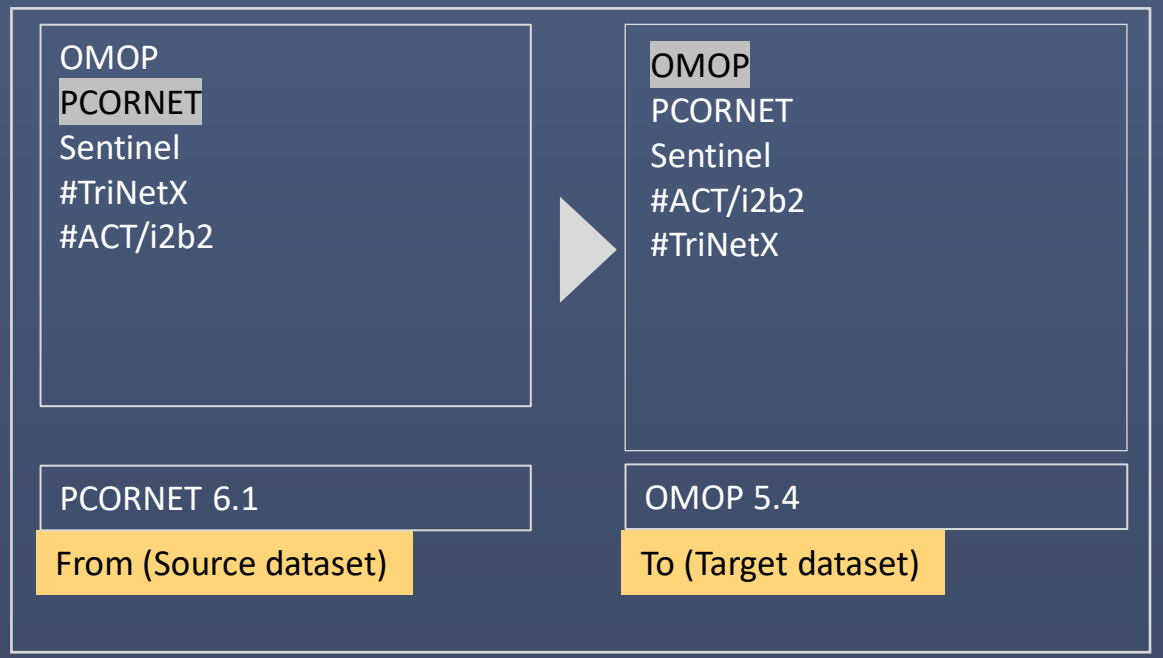
Semantic Mapping and Transformation

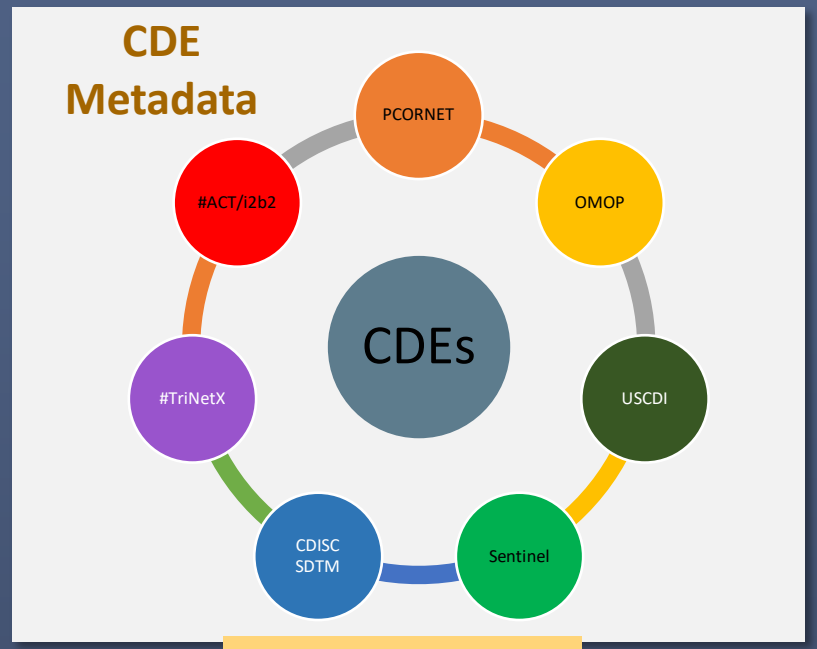
Code Map Services

- FDA, NIH/National Center for Advancing Translational Research (NCATS), NIH/National Cancer Institute Collaboration
- Funded by ONC/PCORI Trust Fund
  - Establish a “source of truth” within a sustainable ongoing government supported infrastructure
    - *-NCI’s Cancer Data Standards Registry and Repository (caDSR)*
  - Use CDE metadata semantics to establish where data is the same
  - Register national common data model-to-model mappings instead of redoing mappings from scratch

Code Map Services: Harmonization and Transformation of Disparate Data Models

Transformation Tool





PCORNet CDE
6153919v2
Term: Hispanic

Common
"What"

Different "Words"

Common
"How"

UN
NI
YN
OT
R

C17998 Unknown
C53269 No Information
C17459 Hispanic or Latino
C41222 Not Hispanic
C17649 Other
C51024 Refused

38003563
38003564

OMOP v5.4
6153918v1
Term: ethnic_concept_id

Standard, Structured CDE Metadata

WHAT?

PCORNet: Hispanic
OMOP: ethnic_concept_Id

Person Ethnic Group Code (What)

What is being described? Person

What characteristic? Ethnic Group

C25190 Concept

C16564 Concept

"Ordered "Concepts

HOW?

DropDown

PCORNet		OMOP
UN	C17998 Unknown	
NI	C53269 No Information	
Y	C17459 Hispanic or Latino	38003563
N	C41222 Not Hispanic or Latino	38003564
OT	C17649 Other	
R	C51024 Refused	

Permitted Value 38003563 (Data Value)

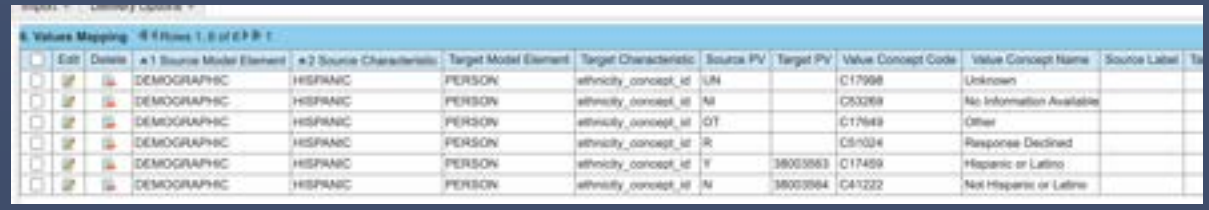
Permitted Value 38003564 (Data Value)

C17459 Concept

C41222 Concept

Automating Mapping Based On CDEs

1. Capture the data model and a CDE for each field
2. The system uses the CDE to determine which fields in the Source (PCORnet) and Target (OMOP) have the same meaning – the same “What”
3. Where the fields are the same, use the “How” concepts to figure out which data values are the same

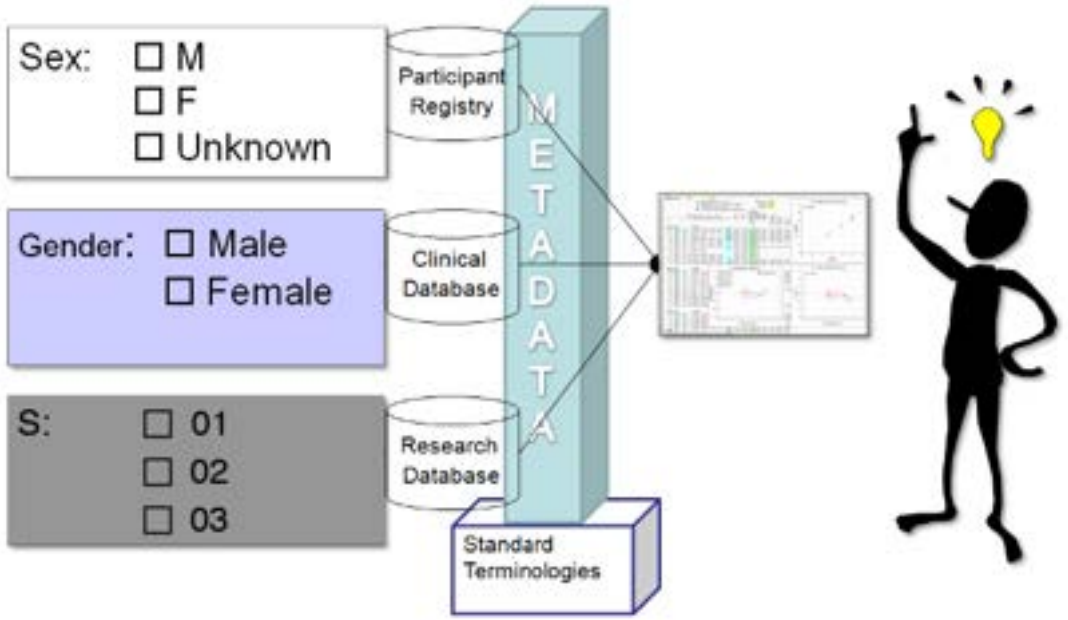


Putting it all together

- DCP-001
  - Studies across 43 states use CDEs to help researchers help underserved communities participate in cancer trials improving our understanding of how trials impact the broader community.
- Code Map Project
  - FDA, NCATS, NCI Collaboration
  - CDEs provide a roadmap for mapping and combining datasets (data aggregation).
  - CDEs standardize and harmonize data across diverse datasets and studies, enabling research breakthroughs and innovations.

ScHARe

A World WITH CDE Metadata



Frees data consumers from the burden of data reuse

Acknowledgements

- NYU Health Sciences Library – Panda Video

- Brenda R. Duggan, RN, BSN, NCI CBIIT Program Manager, Metadata Content
- Brenda Maeske, BS, SAIC, NCI Contractor, Sr. Health Metadata Analyst
- Gwen Dean, MS, BS, SAIC, NCI Contractor, Principal Health Metadata Analyst | Metadata Curator

ScHARe

Thank you!!

Please use CDEs in all your Studies!

Glossary

- Data – The information contained in a database or spreadsheet
- Metadata – Metadata is structured information that describes or explains an information resource, making it easier to use.
- CDE – Common Data Element – A specification that describes a question (What) and a list of choices for an answer (How) to collect a piece of data.
- CDE Metadata – Metadata describing Common Data Elements.
- Data Dictionary – A repository of information describing the contents, format, and structure of variable in a database.
- Interoperability - The ability of two or more computer systems to exchange and make use of the shared data. This process requires the systems to adopt standard data formats and structure and have a shared understanding of the meaning of data.
- Vocabulary - The body of words used in a particular language.
- Terminology – Vocabulary that describes the words or terms used for a given profession. An example is the NCI Thesaurus.
- Ontology – Describes knowledge about a particular domain or subject to aid shared understating. Uses terms and describes the meaning through relationships between the terms. Often includes a graphical depiction. An example is the Pizza Ontology. There can be more than one ontology describing the same domain.

Glossary

- Concept – An abstract idea or general notion. Concepts can be represented by terms in a terminology or ontology.
- CUI – Concept Unique Identifier. This term is used by the NLM UMLS to the terms in the UMLS Metathesaurus.
- Concept Code – A unique alphanumeric or numeric identify in a terminology.
- Semantics – The meaning of a word, phrase or sentence.
- Instrument – A tool or implement, especially one for scientific work used to measure or gauge the level of things related to a research question. Examples are a set of interview questions that are scales that can be scored. An example are the PhenX Toolkit questionnaires.
- Data Collection Tool – Specific software used to capture data for research, analysis, or any other purpose. An example is REDCap.
- Data Model – A visual representation of a group's data elements. Usually the data elements are arranged into higher level groups or classes of information and the model shows the relationships between them. An example is OMOP.
- Data Exchange Standard – A specific set of rules and structure for exchanging data between systems. An example is JSON or FHIR.

ScHARe

CDEs

ScHARe Data Repository

CORE COMMON DATA ELEMENTS

NOVEL CDE FOCUSED REPOSITORY TO FOSTER INTEROPERABILITY

COMPLY WITH DATA SHARING POLICY - HOST PROJECT DATA

DATA ECOSYSTEM

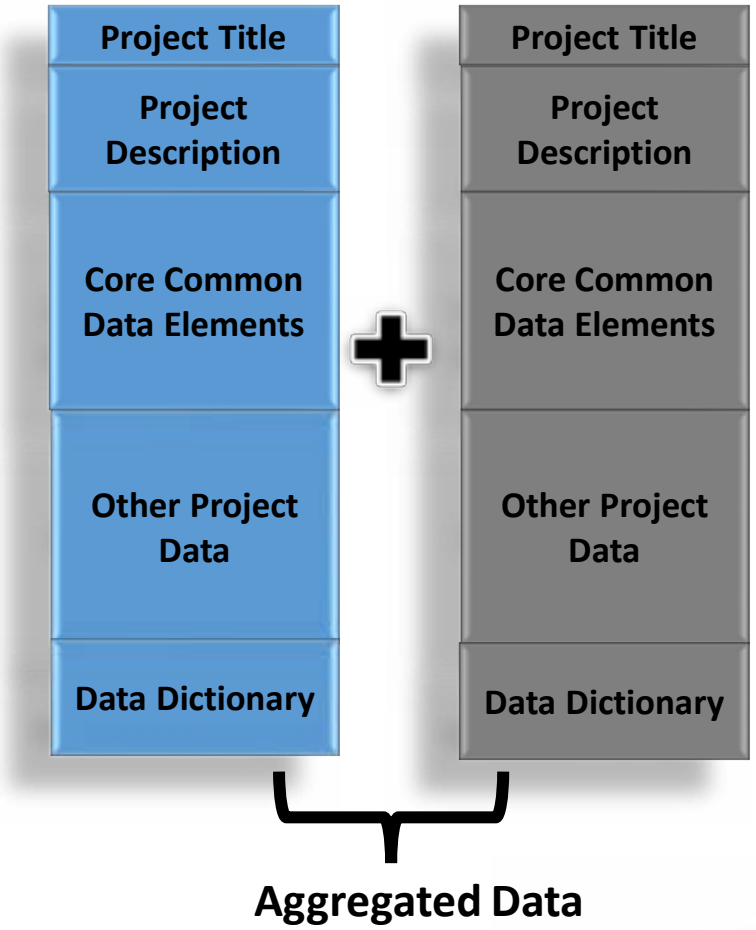
- Map across datasets
- Map across platforms

UPCOMING

ScHARe

Core Common Data Elements
Intramural and Extramural Project Repository

- Complies with **NIH Data Sharing Policy**
- Fosters dataset sharing and interoperability by using or mapping to **Core Common Data Elements**
- Provides resources for **intramural researchers** to work in a secure workspace and host data
- Centralizes **aggregated datasets** for repeat use



UPCOMING

Labels of draft Core Common Data Elements

- Age
- Birthplace
- Zip Code
- Race and Ethnicity
- Sex
- Gender
- Sexual Orientation
- Marital Status
- Education
- Annual Household Income
- Household Size
- English Proficiency
- Disabilities
- Health Insurance
- Employment Status
- Usual Place of Health Care
- Financial Security / Social Needs
- Self Reported Health
- Health Conditions (and Associated Medications)
- NIMHD Framework
- Health Disparity Outcomes

1. **Age**

What is the person's age? (collapse data over 89 yrs old / 2 yrs and under, report in months-does not exclude asking full birthdate)

_____ ☐ years ☐ months

Project 5 Covid-19 Age https://cde.nlm.nih.gov/cde/search?q=PROJECT%205&nihEndorsed=true

2. **Birthplace**

Where were you born?

☐ In the United States, including U.S. Territories (Puerto Rico, Guam, U.S. Virgin Islands, American Samoa and Northern Mariana Islands) (**Select from Drop Down-not doable on word doc**)

☐ Outside the United States (**Select from Drop Down-ISSO categories-not doable on word doc**)

PhenX – Birthplace https://www.phenxtoolkit.org/protocols/view/10201 *ADAPTED-Territoires with US; instead of seperate*

Source for PhenX : American Community Survey (ACS), 2008

3. **ZIP code (caveat collapse zip codes w less than 10)**

What is your current postal ZIP code? _____

Project 5 Covid-19 Address Postal Code https://cde.nlm.nih.gov/deView?tinyId=w_BHatIMoA

4. **Self-Identification (This question's intent is to get at bare minimum of identification, which will be determined by the changes proposed by OMB. Study can collect details of Race and Ethnicity as preferred. This does not supplant other required R/E reporting. Awaiting OMB.)**

Please select the racial category or categories with which you most closely identify. (*select all that apply***)** ***(this could be: Please select your heritage country)***

- ☐ American Indian or Alaska Native
- ☐ Asian or Asian American
- ☐ Black or African American
- ☐ Hispanic or Latino
- ☐ Native Hawaiian or Other Pacific Islander
- ☐ Middle Eastern or North African (in current reporting tables will be reported as white)
- ☐ White

ScHARe working group preference based on potential classifications in 2030 census https://www.npr.org/2021/09/30/1037352177/2020-census-results-by-race-some-other-latino-ethnicity-hispanic#:~:text=And%20under%20that%20combined%20question%2C%20the%20list%20of,federal%20agencies%20collect%20data%20on%20race%20and%20ethnicity.

5. Sex

What was your sex assigned at birth, on your original birth certificate?

- ☐ Female
- ☐ Male
- ☐ Intersex
- ☐ None of these describe me
- ☐ Prefer not to answer

PhenX Protocol - Biological Sex Assigned at Birth https://www.phenxtoolkit.org/protocols/view/11601

All of Us Research Program, Participant Provided Information (PPI), 2018

National Academies Sciences, Engineering, Medicine report: Measuring Sex, Gender Identity, and Sexual Orientation
https://www.nationalacademies.org/our-work/measuring-sex-gender-identity-and-sexual-orientation-for-the-national-institutes-of-health
and All of Us
chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://allofus.nih.gov/sites/default/files/aou_ppi_basics_version.pdf

6. Gender

What is your current gender? [Select only one]

- ☐ Man
- ☐ Woman
- ☐ Non-Binary
- ☐ Transgender
- ☐ None of these describe me-I would like to consider additional options

  Are any of these a closer description to your gender identity?

  [ ] Trans man/Transgender Man/FTM
  [ ] Trans woman/Transgender Woman/MTF
  [ ] Genderqueer
  [ ] Genderfluid
  [ ] Gender variant
  [ ] Questioning or unsure of your gender identity
  [ ] None of these describe me, and I want to specify _____

- ☐ Prefer not to answer

PhenX Protocol - Gender Identity https://www.phenxtoolkit.org/protocols/view/11801

All of Us Research Program, Participant Provided Information (PPI), 2018

National Academies Sciences, Engineering, Medicine report: Measuring Sex, Gender Identity, and Sexual Orientation
https://www.nationalacademies.org/our-work/measuring-sex-gender-identity-and-sexual-orientation-for-the-national-institutes-of-health
Adapted: Non Binary added

7. Sexual orientation

Which of the following best represents how you think of yourself? [Select only one]

☐ Lesbian
☐ Gay
☐ Straight, that is, not gay or lesbian, etc.
☐ Bisexual

If none of the above represents you, are any of these a closer description of how you think of yourself (drop down)

[ ] Queer
[ ] Polysexual, omnisexual, sapiosexual or pansexual
[ ] Asexual
[ ] Two-spirit
[ ] Have not figured out or are in the process of figuring out your sexuality
[ ] Mostly straight, but sometimes attracted to people of your own sex
[ ] Do not think of yourself as having sexuality
[ ] Do not use labels to identity yourself
[ ] Don't know the answer
[ ] No, I mean something else (optional free text) _____

☐ Prefer not to answer

Phen X Sexual Orientation Protocol. https://www.phenxtoolkit.org/protocols/view/11701?origin=subcollection

All of Us Research Program Participant Provided Information (PPI) Version: December 17, 2018

National Academies Sciences, Engineering, Medicine report: Measuring Sex, Gender Identity, and Sexual Orientation
https://www.nationalacademies.org/our-work/measuring-sex-gender-identity-and-sexual-orientation-for-the-national-institutes-of-health

8. **Marital status**

What is your current marital status?

- ☐ Married
- ☐ Living as married or living with a romantic partner
- ☐ Divorced
- ☐ Widowed
- ☐ Separated
- ☐ Single, never been married-not living with romantic partner
- ☐ Prefer not to answer

Hints 5 Cycle 4 (2020) https://hints.cancer.gov/view-questions-topics/question-details.aspx?qid=593 (BRFSS Questionnaire (2001), Section 13: Demographics modified)
All of Us
chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://allofus.nih.gov/sites/default/files/aou_ppi_basics_version.pdf

9. Education

What is the highest level of education you have completed?

- ☐ No formal schooling
- ☐ Primary/Grade/Elementary School (approximately grades 1st through 5th)
- ☐ Middle School/Lower Secondary Education (approximately grades 6th through 8th)
- ☐ Secondary/High School/Upper Secondary (grades 9th through 11th) without a high school diploma
- ☐ General Educational Diploma (GED)
- ☐ Secondary/High School/Upper Secondary (grades 9th through 12th) with a high school diploma
- ☐ Occupational/Technical/Vocational Programs/Short Cycle Tertiary Education - Associate's Degree (approximately 2 years)
- ☐ College/University/Bachelor's Degree/Equivalent Tertiary Education (approximately 3-5 years)
- ☐ Graduate/post-graduate degree/professional degree/ (JD, PhD, MD, EdD, Eng, Master's Degree, etc.)

International Standard Classification of Education (ISCED)
https://datatopics.worldbank.org/education/wRsc/classification#:~:text=The%20International%20Standard%20Classification%20of,revised%20in%201997%20and%202011
and
USA standards of Education https://www2.ed.gov/about/offices/list/ous/international/usnei/us/edlite-structure-us.html#:~:text=Early%20childhood%20education%20is%20followed,then%20postsecondary%20(tertiary)%20education.

10. Annual household income range

What is your annual household income from all sources within family, not including roommates?

- ☐ Less than $10,000
- ☐ $10,000-$24,999
- ☐ $25,000-$34,999
- ☐ $35,000-$49,999
- ☐ $50,000-$74,999
- ☐ $75,000-$99,999
- ☐ $100,000-$149,999
- ☐ $150,000-$199,999
- ☐ $200,000 or more

All of Us - Basic Information Survey https://allofus.nih.gov/sites/default/files/aou_ppi_basics_version.pdf

BRFSS = Behavioral Risk Factor Surveillance System (CDC)

11. Household family size

Approximately how many individuals (adult and children) does your household family income support?

☐ _____

Project 5 Covid-19 Shared Living Space Number of Individuals https://cde.nlm.nih.gov/cde/search?q=PROJECT%205&nihEndorsed=true

12. English proficiency

We are interested in your own opinion of how well you speak English. Would you say you speak English:

☐ Very well
☐ Well
☐ Not well
☐ Not at all
☐ Refused
☐ Don't Know

PhenX Toolkit - English Proficiency https://www.phenxtoolkit.org/protocols/view/270201

Regents of the University of California. (2019). CHIS 2018 Adult Questionnaire, question number "QA18_G8" is represented in this protocol as question 1. Retrieved from http://healthpolicy.ucla.edu/chis/design/Pages/questionnairesEnglish.aspx

13. Disabilities

Do you have a disability or have serious difficulty with any of the following? Select all that apply.

- ☐ Deafness or difficulty hearing
- ☐ Blindness or difficulty seeing
- ☐ Difficulty concentrating, remembering, and deciding
- ☐ Difficulty walking or climbing stairs
- ☐ Difficulty dressing or bathing
- ☐ Difficulty doing errands alone
- ☐ Not disabled

CDC Standard Disability Questions https://www.cdc.gov/ncbddd/disabilityandhealth/datasets.html (format adapted)

14. Health insurance

Are you currently covered by any of the following types of health insurance or health coverage plans?

- ☐ Insurance through a current or former employer or union (of yours or another family member's). This would include COBRA coverage.
- ☐ Insurance purchased directly from an insurance company (by you or another family member). This would include coverage purchased through an exchange or marketplace
- ☐ Medicare, for people 65 and older, or people with certain disabilities.
- ☐ Medicaid, Medical Assistance (MA), the Children's Health Insurance Program (CHIP), or any kind of state or government-sponsored assistance plan based on income or a disability.
- ☐ TRICARE or other military health care, including VA health care.
- ☐ Indian Health Service
- ☐ Any other type of health insurance. coverage or health coverage plan
- ☐ Uninsured

PhenX Health Insurance Coverage https://www.phenxtoolkit.org/protocols/view/11502

15. Employment status

We would like to know about what you do: are you working now, looking for work, retired, keeping house, a student, or what?

- ☐ Working now or paid sick leave/parental leave/family leave/administrative leave
- ☐ Only temporarily laid off, or unpaid sick leave/parental leave/family leave/administrative leave
- ☐ Looking for work, unemployed
- ☐ Retired
- ☐ Disabled, permanently or temporarily
- ☐ Raising children full-time, full-time caregiver, or keeping house
- ☐ Student
- ☐ Other/specify: _____

PhenX - Current Employment Status https://www.phenxtoolkit.org/protocols/view/11301 (Adapted-used parental instead of maternal, and family leave added with paid/unpaid)

16. Usual place of health care

Is there a place that you USUALLY go to when you are sick or need advice about your health? Select all that apply.

- ☐ A doctor's office or community health center, including Indian Health Service, or hospital-based clinics
- ☐ Walk-in clinic, urgent care center, or retail clinic in a pharmacy or grocery store
- ☐ Emergency room
- ☐ A VA Medical Center or VA outpatient clinic
- ☐ Some other place
- ☐ Does not go to one place most often
- ☐ Don't know

PhenX Protocol Access to Health Services Ques #5 https://www.phenxtoolkit.org/protocols/view/270101 (adapted with hospital-based clinics)

Project 5 Covid-19 Usual Place of Health Care Type https://cde.nlm.nih.gov/cde/search?q=PROJECT%205&nihEndorsed=true\ (adapted with hospital-based clinics)

17. Economic Stability – Social Needs

In the past year, have you or any family members you live with been unable to get any of the following when it was really needed? Select all that apply.

- ☐ Childcare
- ☐ Clothing
- ☐ Food
- ☐ Housing
- ☐ Internet/Broadband
- ☐ Phone (e.g., mobile or landline)
- ☐ Transportation (e.g., private or public)
- ☐ Utilities (e.g., gas, electric, propane, natural gas, etc.)
- ☐ Medicine or any health care (medical, dental, mental health, vision)
- ☐ Other/specify: _____

Source: Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences (PRAPARE) tool (Adapted-internet, housing, transportation added to question #14) Housing and transportation is included in survey. https://prapare.org/wp-content/uploads/2023/01/PRAPARE-English.pdf

U.S. Census Bureau, 2015 and 2016 American Community Survey – Internet/Broadband
https://www.census.gov/content/dam/Census/library/publications/2018/acs/ACS-39.pdf

18. Self-reported health

Would you say your health in general is excellent, very good, good, fair, or poor?

- ☐ Excellent
- ☐ Very good
- ☐ Good
- ☐ Fair
- ☐ Poor

Patient-Reported Outcomes Measurement Information (PROMIS)
https://www.healthmeasures.net/index.php?option=com_instruments&task=downloadComponentFile&file=PROMIS%20Scale%20v1.2%20-%20Global%20Health%20Physical%202a%2009062016.pdf

19. Health Conditions and Medications or other Treatments

Has a health care provider told you that you have any one or more of the following conditions? Select all that apply currently. Check the second box if you are taking medications or receiving some other treatment for the condition.

☐	☐	Cancer
☐	☐	Coronary heart disease
☐	☐	Heart failure
☐	☐	High blood pressure/hypertension
☐	☐	Stroke
☐	☐	Thrombotic disorders
☐	☐	High cholesterol
☐	☐	Diabetes (type I)
☐	☐	Diabetes (type II)
☐	☐	Obesity
☐	☐	Hepatitis
☐	☐	Other chronic liver disease

- ☐ ☐ Asthma
- ☐ ☐ Other chronic respiratory disease (e.g., COPD, emphysema)
- ☐ ☐ Chronic kidney disease
- ☐ ☐ Psychological and/or psychiatric disease or disorder (e.g., anxiety, depression, bipolar disorder)
- ☐ ☐ Alzheimer's disease
- ☐ ☐ Dementia
- ☐ ☐ Epilepsy
- ☐ ☐ Multiple sclerosis
- ☐ ☐ Other chronic neurological condition (e.g., Parkinson's disease, migraine)
- ☐ ☐ Immunodepression
- ☐ ☐ HIV/AIDS
- ☐ ☐ Autoimmune condition (e.g., rheumatoid arthritis, systemic lupus erythematosus, vasculitis)
- ☐ ☐ Chronic musculoskeletal condition (e.g., back pain, osteoarthritis, osteoporosis)

☐ ☐ Sickle cell disease

☐ ☐ Sleep disorder (e.g., insomnia, sleep apnea, narcolepsy)

☐ ☐ Solid organ transplant

☐ ☐ Smoking

☐ ☐ Other substance use disorder (e.g., drugs and/or alcohol dependence)

☐ ☐ Long Covid (also known as long-haul COVID, long-term effects of COVID, chronic COVID, post-acute COVID-19, and PASC - post-acute sequelae of SARS-CoV-2)

☐ ☐ Chronic fatigue

☐ ☐ Dental diseases and conditions (e.g., caries, periodontal disease, oral and pharyngeal cancer)

☐ ☐ Eye diseases and conditions (e.g., cataract, glaucoma, amblyopia, myopia and other refractive errors, age-related macular degeneration, diabetic retinopathy, ocular trauma, uveitis, keratoconus)

☐ ☐ Other chronic disease/specify:

☐ None of the above

Project 5 Covid-19 Comorbidity or Underlying conditions
https://cde.nlm.nih.gov/cde/search?q=PROJECT%205&nihEndorsed=true\ (Adapted for Medications-Added Chronic musculoskeletal conditions, High Cholesterol, Sleep Disorders and Stroke)

20. Minority Health and Health disparities research content area

Which of the following content areas of research is this study addressing, if any? Select all that apply.

- ☐ Minority health study focused on a one race or ethnic population and not addressing a health disparity.
- ☐ Health Disparity Outcome (select the focus area)
  - ☐ Higher incidence and/or prevalence of disease, including earlier onset or more aggressive progression of disease
  - ☐ Premature or excessive mortality from specific health conditions
  - ☐ Greater global burden of disease, such as Disability Adjusted Life Years (DALY), as measured by population health metrics
  - ☐ Poorer health behaviors and/or clinical outcomes using established measures
  - ☐ Worse outcomes on validated self-reported measures that reflect daily functioning or symptoms from specific conditions
- ☐ Other Health Outcomes / Healthcare Delivery

Duran D, Perez-Stable, E. Novel Approaches to Advance Minority Health and Health Disparities; Am J Public Health. 2019, Jan;109(S1):S8-S10. doi:10.2105/AJPH. 2019.304952. PMID: 30699026; PMCID:PMC6356133. *ADAPTED with Other health outcomes delivery/care*

21. NIMHD Framework

What NIMHD Research framework levels and domains of influence is your study targeting? (Select all that apply)

Levels of Influence

- ☐ Individual
- ☐ Interpersonal
- ☐ Community
- ☐ Societal

Domains of Influence

- ☐ Biological
- ☐ Behavioral
- ☐ Physical/Built Environments
- ☐ Sociocultural Environment
- ☐ Health Care Systems and Clinical Care

NIMHD Research Framework. https://www.nimhd.nih.gov/about/overview/research-framework/nimhd-framework.html

ScHARe

Thank you

Think-a-Thon poll

1. **Rate how useful this session was:**

- ☐ Very useful
- ☐ Useful
- ☐ Somewhat useful
- ☐ Not at all useful

Think-a-Thon poll

2. Rate the pace of the instruction for yourself:

- ☐ Too fast
- ☐ Adequate for me
- ☐ Too slow

Think-a-Thon poll

3. **How likely will you participate in the next Think-a-Thon?**

- ☐ Very interested, will definitely attend
- ☐ Interested, likely will attend
- ☐ Interested, but not available
- ☐ Not interested in attending any others

Terra tutorials and resources

If you are new to Terra, we recommend exploring the following resources:

- Overview Articles: Review high-level docs that outline what you can do in Terra, how to set up an account and account billing, and how to access, manage, and analyze data in the cloud
- Video Guides: Watch live demos of the Terra platform's useful features
- Terra Courses: Learn about Terra with free modules on the Leanpub online learning platform
- Data Tables QuickStart Tutorial: Learn what data tables are and how to create, modify, and use them in analyses
- Notebooks QuickStart Tutorial**:** Learn how to access and visualize data using a notebook
- Machine Learning Advanced Tutorial: Learn how Terra can support machine learning-based analysis

ScHARe

Next Think-a-Thons:

bit.ly/think-a-thons

Register for ScHARe:

bit.ly/join-schare

schare@mail.nih.gov