

# Common Data Models, CDEs, and BRIDG

*Denise Warzel, NCI*

# CDMs, CDEs and BRIDG

- FAIR Data Sharing: The Roles of Common Data Elements (CDEs) and Harmonization
  - Findable, Accessible, Interoperable and Reusable (FORCE 11, Wilkinson et al., 2016)
  - CDEs = ‘discrete, clearly defined and reusable data collection units’ (ISO/IEC 11179)
  - CDEs: not always ‘common’, not always well-constructed, not usually unique
  - NCI caDSR (established 2000) – linked CDEs to controlled terminology (NCI Enterprise Vocabulary Services – NCI Thesaurus)
  - NIH CDE repository (established ~ 2010), contains CDEs from a subset of NIH Centers, including NCI and NINDS; “currently there are no ... specifications governing their ... use”\* (Sheehan et al., 2016)

\* “Although CDEs are a step in the direction of ‘FAIR Data Sharing’, the lack of **standardization** and **governance to harmonize and elevate specific CDEs** to standards, **limited engagement in [CDE Development by] the research community**, **inconsistent enforcement efforts**, **difficulty in implementation and use**, and **issues with harmonization or curation** remain barriers and deterrents to broader deployment.”

# Common Data Model Harmonization I (CDMH)



## Goal:

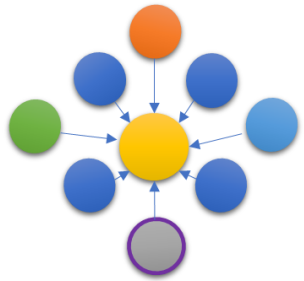
Build a data infrastructure for conducting patient-centered outcomes research using Real World Data (RWD) derived from the delivery of health care in routine clinical settings.







## Objective:

Develop the method to harmonize the Common Data Models (CDM) of various networks, allowing researchers to simply ask research questions on much larger amounts of Real World Data than currently possible, leveraging open standards and controlled terminologies to advance Patient-Centered Outcomes Research.

# CDMH Project Process: CDM, CDE, BRIDG



- Mapped Disparate CDMs to BRIDG
- Extended BRIDG clinical content
- Balloted BRIDG w/Extensions
- Registered/mapped CDMs as Metadata in caDSR

Model	Version	Model Tables	Model Elements	Elements Found In BRIDG	Elements Not Found In BRIDG	Percent Elements Not Found In BRIDG	Elements Not Mapped
 Sentinel	6.0.2	12	126	112	11	9%	3
 PCORNET CDM	3.1	14	151	87	17	11%	46
 OHDSI/OMOP	5.2	16	163	136	27	17%	7 tbls
 i2b2	1.4	6	53	33	5	9%	15
Totals		48	493	368	60	12%	64

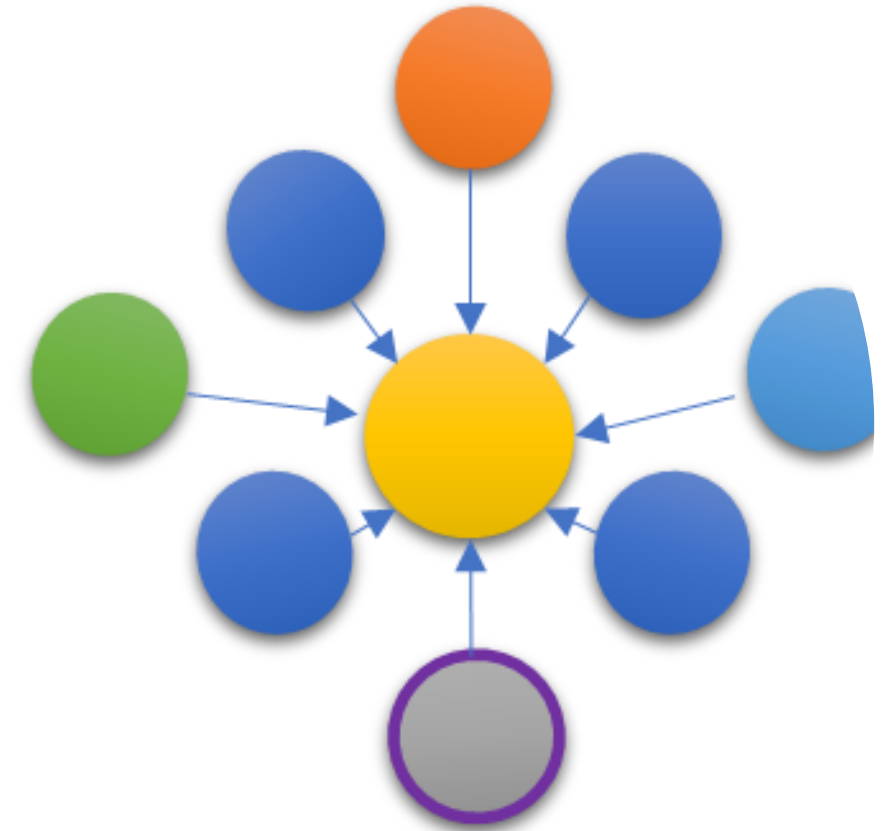
## Metadata Annotation and Registration

Metadata makes the meaning of data FAIR

- Findable: Data specification represented in **human and machine** readable format
- Accessible: Metadata Repository with Web UI and API interfaces
- Interoperable: **Normalizes the meaning** of the fields and the data values using standard terminology
- Reusable: Persistent, unique identifier including version, semantic concept annotations supporting data validation and data transformation

# Why use BRIDG as a “SEMANTIC HUB”?

- BRIDG (Biomedical Research Integrated Domain Group Model)
  - Sponsored by NCI, FDA, CDISC, HL7
    - **CDISC, HL7 and ISO Standard**
  - Model: Relationships, Hierarchy
  - Conceptual: No data specifications - **not bound to any coding system**
    - Allows linkage of fields in different systems that are semantically the same but coded differently
  - Standardized Semantics
    - Meaning of classes and attributes are annotated and registered as metadata using **NCIt standard terminology concepts**
    - NCIt:
      - Source of truth for CDISC terminology
      - Source of truth for FDA terminology
      - Linked to UMLS CUIs
      - Preferred terminology for NCI
  - **Serves as basis for CDMH physical data model**
    - **Standard data values/codings for mapping between the 4 CDMs**
  - **Take advantage of cross-walk to other models** that have been mapped to BRIDG
    - CDISC
    - NCI Clinical Trials Reporting System (CTRP)
    - CIBMTR/NMDP
    - .....



# caDSR Example: Semantic mapping of Hispanic/Ethnicity

1 Field Label	<i>Ethnicity</i>	<i>hispanic</i>	<i>Hispanic</i>	<i>Hispanic</i>	<i>ethnic_concept_id</i>
2 Public ID	6153917	6153919	6153920	6153918	6153921
3 BRIDG Name	Person Biological Entity Ethnic Group	Person Biological Entity Ethnic Group	Person Biological Entity Ethnic Group	Person Biological Entity Ethnic Group	Person Biological Entity Ethnic Group
4 BRIDG Concepts	C25190:C28226:C51070	C25190:C28226:C51070	C25190:C28226:C51070	C25190:C28226:C51070	C25190:C28226:C51070
5 CDM Value Set	CDMH HL7 FHIR v3 Ethnicity Category Code	PCORnet CDM Hispanic Code	Sentinel CDM Hispanic Indicator	ACT I2B2 CDM Hispanic Indicator	OMOP CDM Ethnicity Category Code
Permissible Value(s)	6 Permissible Value(s)	6 Permissible Value(s)	3 Permissible Value(s)	3 Permissible Value(s)	2 Permissible Value(s)
	<b>Data Value</b> Value	<b>Data Value</b> Value	<b>Data Value</b> Value	<b>Data Value</b> Value	<b>Data Value</b> Value
	<b>Data Value</b> Concept	<b>Data Value</b> Concept	<b>Data Value</b> Concept	<b>Data Value</b> Concept	<b>Data Value</b> Concept
	UNK C17998	UN C17998	U C17998		
	NI C53269	NI C53269		NI C53269	
6	2135-2 C17459	Y C17459	Y C17459	Y C17459	38003563 C17459
	2186-5 C41222	N C41222	N C41222	N C41222	38003564 C41222
	OTH C17649	OT C17649			
	ASKU C79729	R C79729			

1 - Field labels: CDM fields

2 - Public ID: persistent caDSR identifier for the CDE

3 - BRIDG Name: Standardized BRIDG name  
in caDSR linked to CDE

4 - BRIDG NCI Thesaurus Concept Annotations (NCIt)

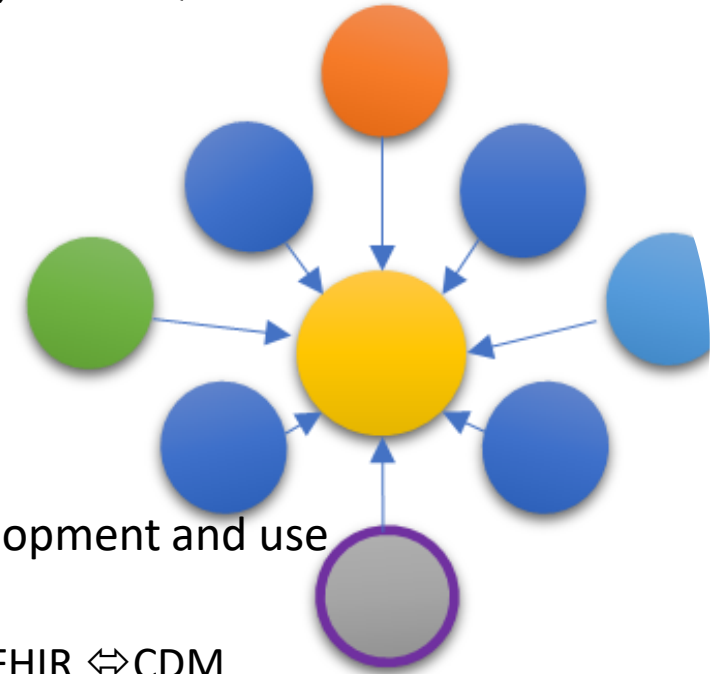
5 - CDM Value Set: Details of how each CDM collects the responses

6 – Concept Annotations

- Red boxed row shows codes of 2135-2, Y, Y, Y, and 38003563
- C17459 concept annotation enables detection of the same meaning of Hispanic or Latino

# CDMH Phase II: CDC, NCATS, NCI, FDA, CD2H

- Universal Query Tool (Write Once Query All) – Clinical Quality Language (CQL) , FHIR SQL
  - FHIR SQL – very promising but not ready at this time
- FHIR Repository – BULK FHIR Repository of De-identified patient data
  - Evaluation of VONK, SMILE, MS, Google and Bunsen
  - SMILE Repository was the most mature, with eMPI, Apache Integration
- Common Data Model Adapter FHIR ↔ BRIDG, FHIR ↔ CDM
  - CDM Networks are the front door to success
  - FHIR US Core Value proposition being tested
- Terminology Services – A tool that support interoperable, terminological development and use
  - NCI and CD2H
  - NCI CDEs with NCI annotations as the gold standard for FHIR ↔ BRIDG, FHIR ↔ CDM



# Recommendations for FAIR Data Sharing

- Encourage the Acceptance and Re-use of Harmonized and Preferred CDEs and Global Data Standards
- Remove Political and Social Barriers to Data Sharing
  - Includes removing political barriers, aligning incentives and encouraging broad collaboration across multiple types of organizations
  - CDE governance should seize the opportunity to broaden CDE impact beyond niche implementation and by engaging the broader research and healthcare community
- Build Better 'Bridges' Between Research and Healthcare

Overcoming these barriers is essential to efficiently and responsibly share meaningful data that can ultimately evolve into learning health systems through which research more rapidly informs care decisions for all of us as patients.





**NATIONAL  
CANCER  
INSTITUTE**

[www.cancer.gov](http://www.cancer.gov)

[www.cancer.gov/espanol](http://www.cancer.gov/espanol)