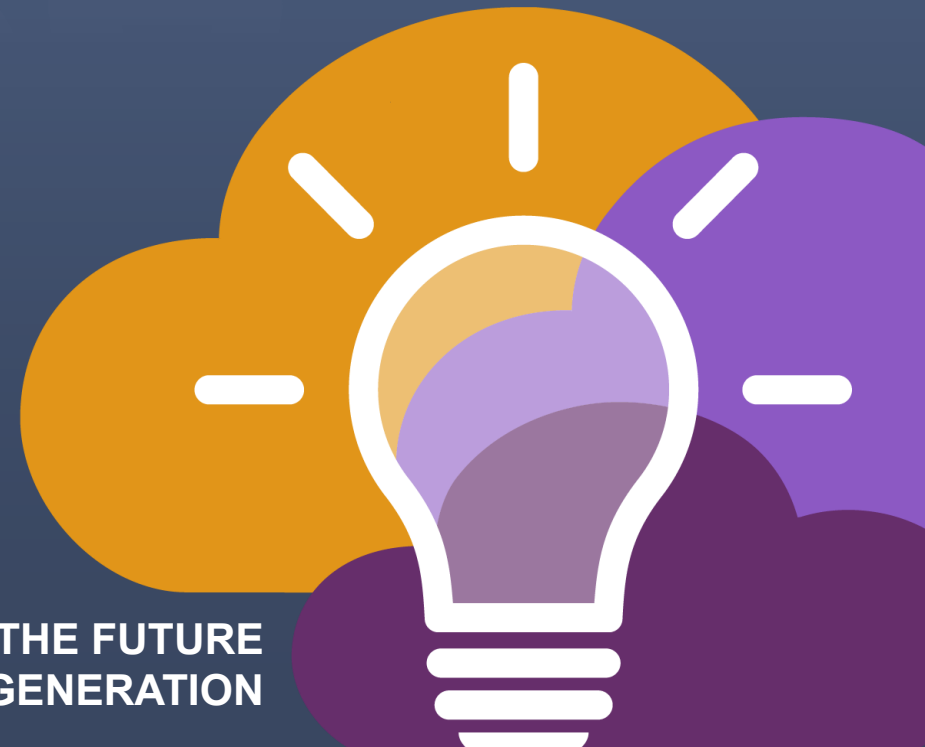


ScHARe

CDE Adoption

- Making Data Interoperable (URI Approach)
- Concept Code Mapping (Data Harmonization)

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION

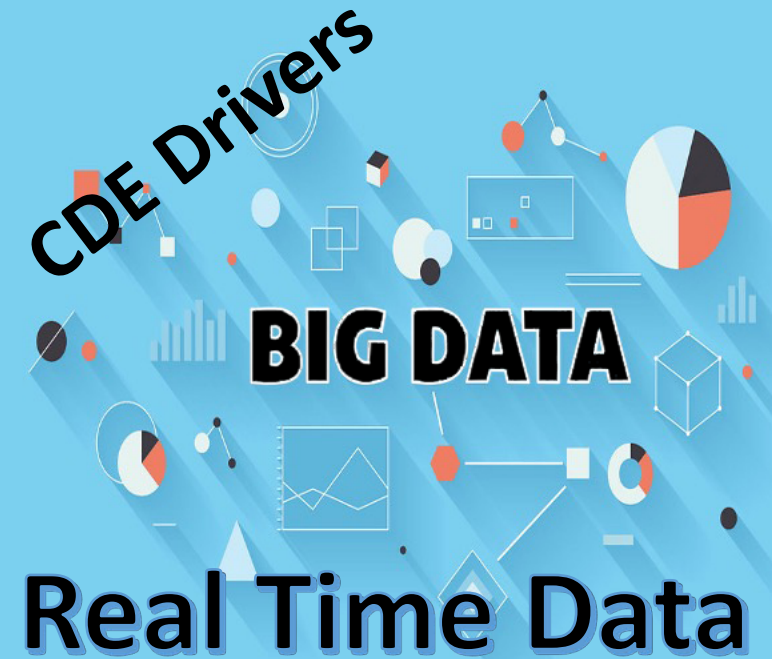




Adopted CDEs to:

- Standardize data for people & computers (human and machine readable)
- Enable data sharing across studies (data interoperability)
- Enhance data interpretation & analysis (semantically defined and standardized coded)
- Simplify collaboration
- Reduces project start-up & results time

BIG DATA AND AI: REQUIRES NEW APPROACHES FOR COLLECTION, MANAGEMENT, ANALYSIS



Covid revealed the need to have real time data



ScHARe

“CORE” CDE Development

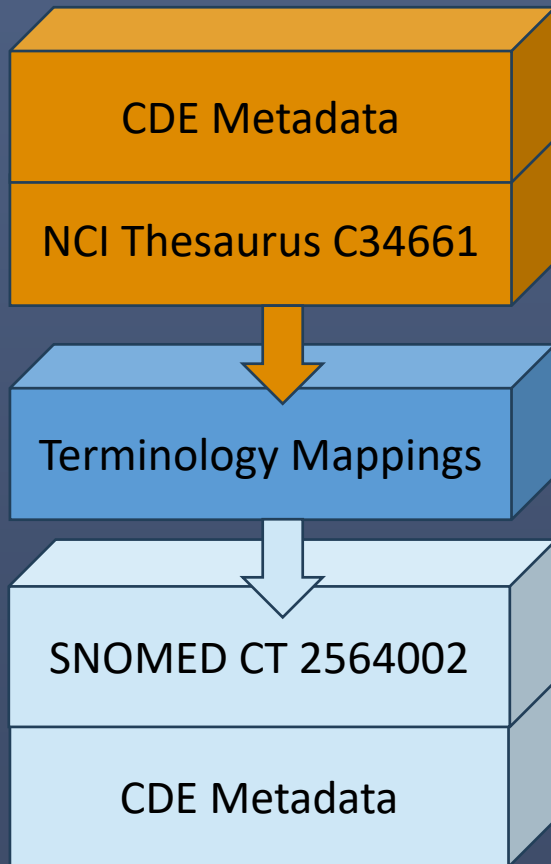
Core Set:

- Few critical questions required from all studies/sites
- Minimal burden
- Allows for questions to be asked in any way, but reported in a standardized format
- Allows for any number of other questions to be collected as collector chooses

Criteria:

- PhenX Toolkit first
- Validated source
- Adaptation of a validated source
- Generate new gap area CDE

Importance of Concept Code Mapping and Interoperability (Uniform Resource Identifier (URI))



- CDE unique **CONCEPT CODES** represent data semantics
 - Human readable
 - Machine readable format
- **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

Questions become CDEs When Defined and Coded

Education

What is the highest level of education you have completed?

Shared Semantics and Concept Code:

An indication of the years of schooling completed in graded public, private, or parochial schools, and in colleges, universities, or professional schools. **C17953**

URI approach in data repository uses codes to harmonize data rather than semantics (words).

Human Readable w Shared Meaning

Codes facilitate Machine Readable



CDES: Words Precisely Defined-Shared Meaning

Words can be **SEMANTICALLY AMBIGUOUS**.

- Context is important in conveying meaning when using CDEs
 - Words have different meanings depending on words around it and context.
- Some examples:
 - **Seizure:** uncontrolled electrical activity between brain cells / spiritual experience?
 - **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?

Words can mean different things in different contexts



Questions become CDEs When Defined and Coded

Education

What is the highest level of education you have completed?

Shared Semantics and Concept Code:

An indication of the years of schooling completed in graded public, private, or parochial schools, and in colleges, universities, or professional schools. **C1795**

URI approach in data repository uses codes to harmonize data rather than semantics (words).

Human Readable w Shared Meaning

Machine-Readable format—excel spreadsheet: codes increase interoperability and use of pipes to separate concepts & codes

Permissible Value (PV) Labels	PV Definitions	PV Concept Identifiers
No formal Schooling	Indicates that a person has never attended an educational program or formal schooling.	C67122
Primary/Grade/Elementary School (approximately grades 1st through 5th)	Indicates that 5th grade potentially is the highest level of educational achievement.	C67127
Middle School/Lower Secondary Education (approximately grades 6th through 8th)	Indicates that 8th grade potentially is the highest level of educational achievement.	C67130

Some Concept Coding Systems One NOT Better Than the Other

General use....

LOINC Laboratory and Clinical Research

ULMS (CUI) Biomedical

FHIR Electronic Health Records

*NCIt Cancer

*ScHARe used NCIt because it has several population concepts

How a Survey Question Became a CDE

Please select the racial category or categories with which you most closely identify. *(select all that apply)*

- 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

Survey Questions become CDEs when they are:

- **semantically defined by a standardized coding system for shared meaning**
- **in a format that is human and machine readable for ease of reuse**

Making of a CDE from a Protocol/Question

Need a standardized defined concept and related code.

Source: NCI Thesaurus

Race/Ethnicity Self-Identification

A textual description of a person's race. C17049 | The ethnicity of a person. C16564 | An individual's perspective or subjective interpretation of an event or information. C74528

- American Indian or Alaska Native |
- Asian or Asian American |
- Black or African American |
- Hispanic, Latino, or Spanish |
- Native Hawaiian or Other Pacific Islander |
- Middle Eastern or North African |
- White

URI approach in data repository uses codes to harmonize data rather than semantics. This improves data interoperability.

Making of a CDE from a Protocol/Question

- A person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment. (OMB) **C41259** |
- A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent, including for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. (OMB) **C41260** |
- A person having origins in any of the Black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American". (OMB) **C16352** |
- A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. The term, "Spanish origin" can be used in addition to "Hispanic or Latino". (OMB) **C17459** |
- A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. (OMB) **C41219** |
- Denotes a person having origins in the region of southwest Asia, between the India subcontinent and Europe, including Kuwait, Turkey, Lebanon, Israel, Iraq, Iran, Jordan, Saudi Arabia, lands east of Pakistan or the other countries of the Arabian Peninsula. Also includes people of Jewish ethnicity including Sephardic and Ashkenazic. **C77820** :
- Denotes a person whose ancestry is in any of the countries of the northern part of the African continent: Algeria, Egypt, Libya, Morocco, Sudan, Tunisia, and Western Sahara. **C126529** |
- A person having origins in any of the original peoples of Europe, the Middle East, or North Africa. (OMB) **C41261**

Making of a CDE from a Protocol/Question

Need a standardized defined concept and related code. Source: NCI Thesaurus

Code Mapping

NCIT

Loinc

UMLS CUI

American Indian or Alaska Native	C41259	LA10608-0	C0282204
Asian or Asian American	C41260	LA6156-9	C0003988
Black or African American	C16352	LA10610-6	C0085756
Hispanic, Latino, or Spanish	C17459	LA6214-6	C0086409
Native Hawaiian or Other Pacific Islander	C41219	LA10611-4	C1513907
Middle Eastern or North African	C43866	Mena no loinc	C1553353
White	C41261	LA4457-3	C0043157

Matched CDE

Income (Project 1)

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

Reported this way

Income (Project 2)

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

Collected this way



Mappable CDE

Income (Project 1)

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

Reported this way

Income (Project 2)

Less than \$10,000 |
\$10,000-\$19,999 |
\$20,000-\$29,999 |
\$30,000-\$39,999 |
\$40,000-\$49,999 |
\$50,000-\$59,999 |
\$60,000-\$69,999 |
\$70,000-\$79,999 |
\$80,000-\$89,999 |
\$90,000-\$99,999 |
.....
\$200,000 or more

Collected this way

Mapped using algorithms

