Introduction to Semantic Interoperability and the Role of Concepts

Denise Warzel, BBA, MSc

ISO SC 32/WG 2 Metadata Standards

Convenor

US National Institutes of Health – National Cancer Institute

Semantic Infrastructure, Metadata and Models Program Manager

Agenda

- Semantic Interoperability

- Interoperability across Healthcare, Healthcare Research, and Biomedical Research

- Lessons learned/challenges

- Role of Concept Systems

Semantic Interoperability

- **Defined**: Ability of computer systems to exchange data, with clear unambiguous meaning

- **Principles**:
  - Standards for data exchange
  - Shared semantic alignment and mapping
  - Rich metadata
  - Governance among stakeholders

Why Semantic Interoperability in Healthcare, Healthcare Research and Biomedical Research?

Scientists need data to conduct research

New cures and treatments

Lack of interoperability inhibits data integration

More Data = Increased confidence in new findings

Interoperability Inhibitors

Different Data Models

Different Coding Standards

Different Terminology

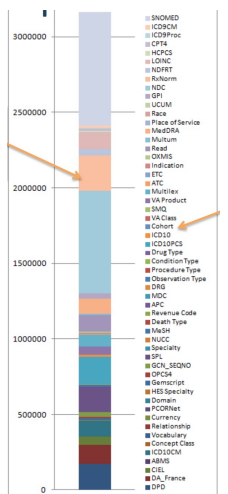# Approach → Challenges and Lessons Learned

Ontologies

- Highly specialized
- Too many "Standard Ontologies"
    - Biomedical Research: 1,051*
        *Stanford University: https://bioportal.bioontology.org/
- Mappings expensive and error prone

Common Data Models

- Healthcare and Healthcare Research
    - Healthcare – e.g. 8 Major Electronic Medical Record (EMR) Vendors – 8 Proprietary Data Models
    - Healthcare Research – e.g 4 "Common Data Models
- Biomedical Research
    - Domain Specific: Genomics, Proteomics, Cancer Specific (Pediatric, Lung, Ovarian), Imaging, etc
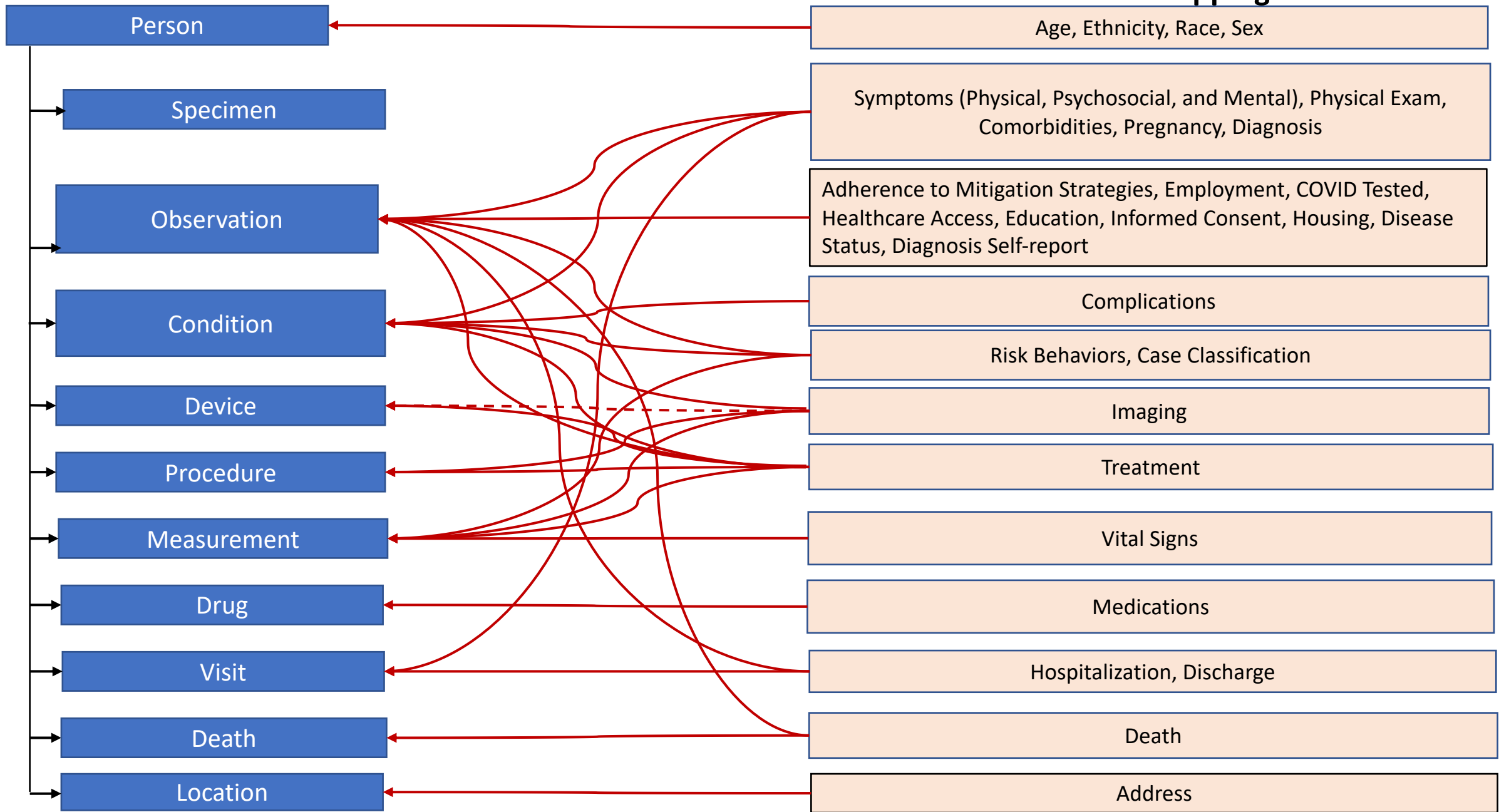
Standard Terminologies

- Healthcare and Healthcare Research
    - Hundreds of terminologies, not fully aligned
    - Multiple mappings by different groups
- Biomedical Research
    - Same problem

Healthcare Research OMOP Data Model

- Person
- Specimen
- Observation
- Condition
- Device
- Procedure
- Measurement
- Drug
- Visit
- Death
- Location

Biomedical Research Partial mapping of NIH COVID CDEs

- Age, Ethnicity, Race, Sex
- Symptoms (Physical, Psychosocial, and Mental), Physical Exam, Comorbidities, Pregnancy, Diagnosis
- Adherence to Mitigation Strategies, Employment, COVID Tested, Healthcare Access, Education, Informed Consent, Housing, Disease Status, Diagnosis Self-report
- Complications
- Risk Behaviors, Case Classification
- Imaging
- Treatment
- Vital Signs
- Medications
- Hospitalization, Discharge
- Death
- Address

What to do?

Leverage ISO/IEC 11179 Metadata Registry Standard

Structured Metadata Supports:

- Accurate, reliable information about data
    - Unique Identifiers
- Data interchange referencing data elements
    - Data Sets, Models
- **Semantics based on Concepts**
- Rich Metadata
    - Multiple Names, Definitions, Languages
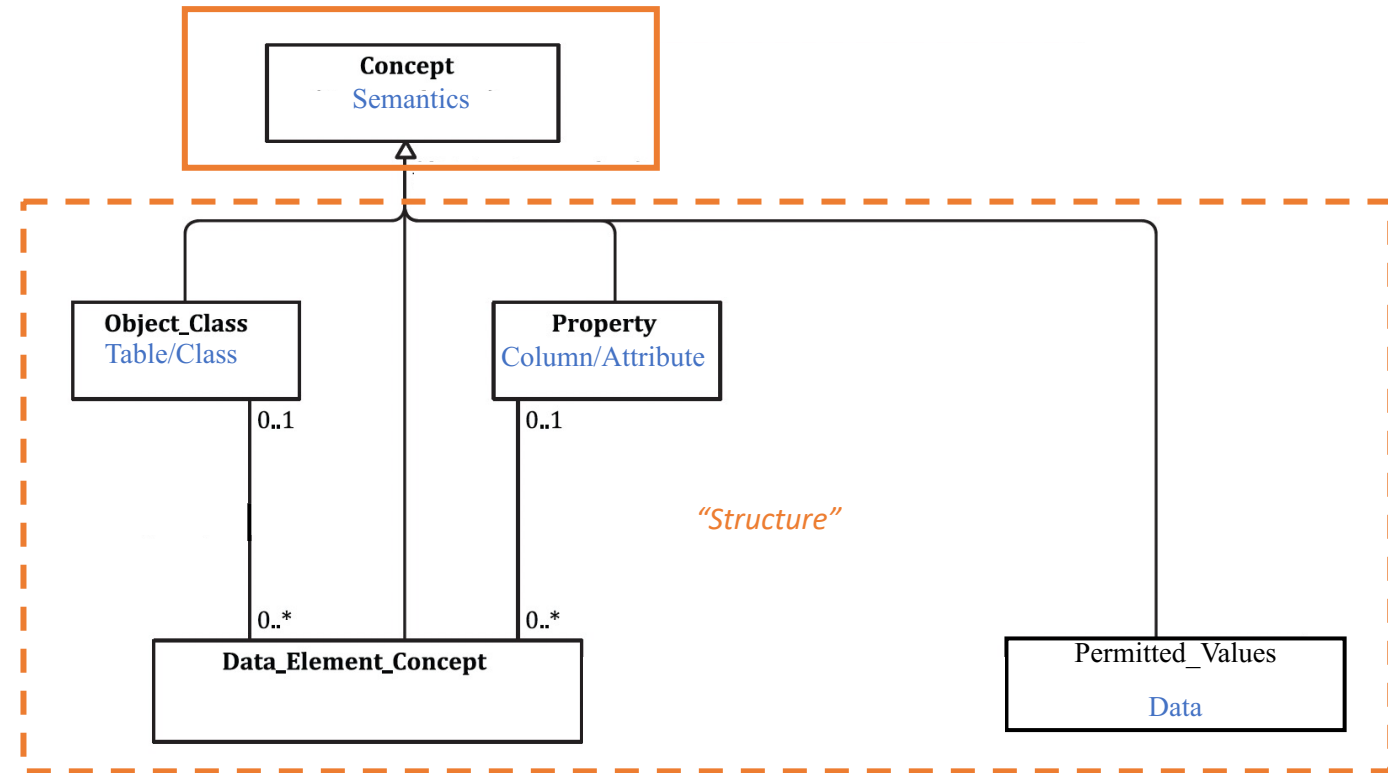
ISO/IEC 11179-31:2023 Data Specification, Based on Figure 4, Data Element Concept metamodel region

Concepts and Semantic Interoperability

Concepts
- Ideas, units of knowledge

Concept Systems
- Composed of terms and definitions
- Domain specific
- Human and computer readable
- Provide semantic clarity
- Enhanced understanding
  - Synonyms, roles, relationships

Semantic Spectrum
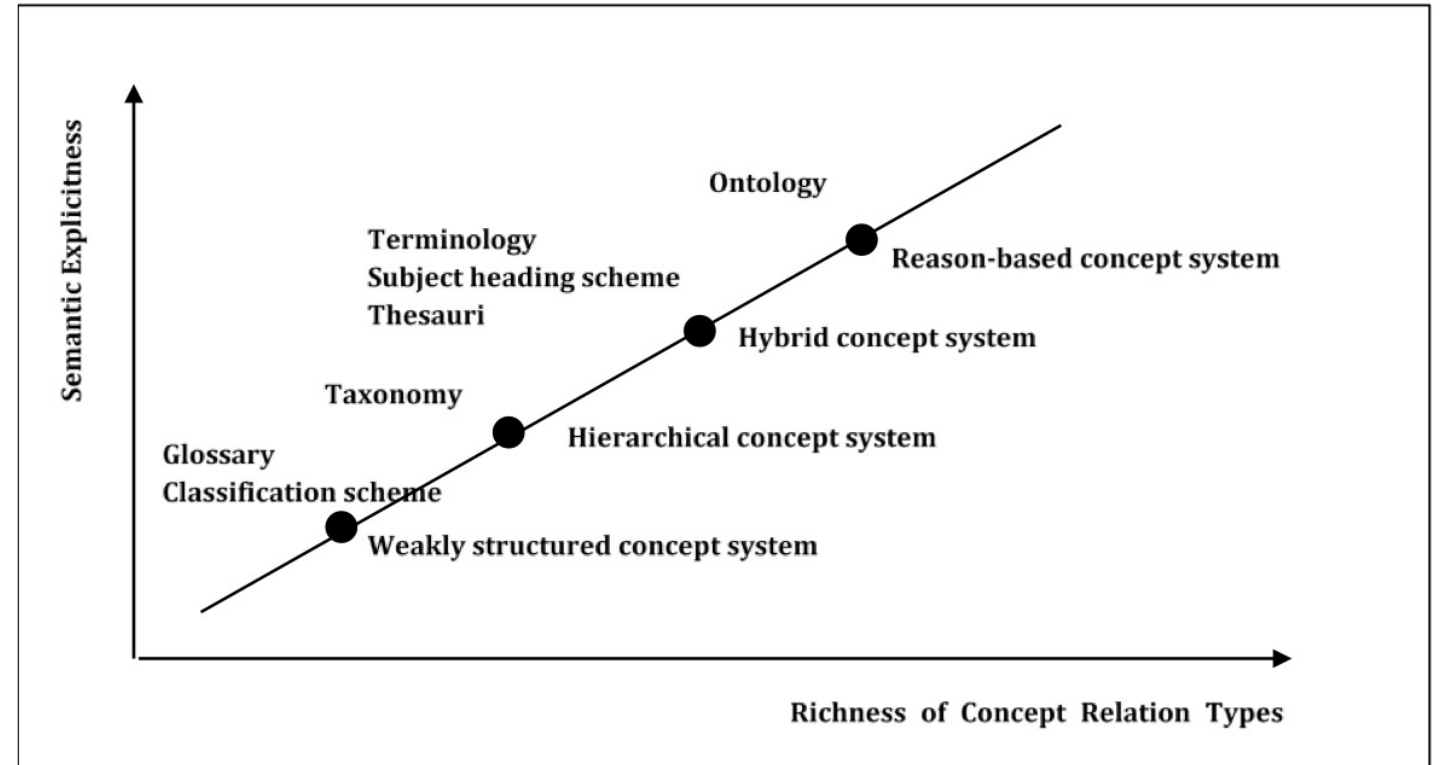
Figure 2 — Level of structure of concept systems

ISO/IEC 5394 Criteria for concept systems

Example: ISO/IEC 11179 Concept Annotations
4 Ethnicity Data Elements, 4 Data Models, 3 Domains

Permitted Values Aligned by Concepts		Healthcare Research Domain		Healthcare Domain	Biomedical Research Domain
		Data Element Semantics Aligned by Concepts C25190 Person C16564 Ethnic Group			
Concept Code	**Concept Name**	**PCORnet** DEMOGRAPHIC HISPANIC	**OMOP** PERSON ethnicity_concept_id	**HL7 US Core** Person Ethnicity	**CDISC** DEMOGRAPHIC Ethnicity
C17998	Unknown	UN		UNK	UNKNOWN
C53269	No Information Available	NI		NI	
C17649	Other	OT		OTH	
C51024	Response Declined	R			
C17459	Hispanic or Latino	Y	38003563	2135-2	HISPANIC OR LATINO
C41222	Not Hispanic or Latino	N	38003564	2186-5	NOT HISPANIC OR LATINO

1ˢᵗ Concepts define the Semantics of the Data Elements across the Data Models

2ⁿᵈ Concepts define Semantics of the data values permitted in each Data Model

3ʳᵈ Concepts enable semantic interoperability

Role of Concepts in Semantic Interoperability

- Shared meaning across disparate domains
  - Healthcare, Healthcare Research, and Biomedical Research
- Enables knowledge discovery
  - Concept Systems → Synonyms, relationships, roles
- Enables data transformation
  - Discover the multiple ways a data element such as **Person Ethnicity** was captured/stored
- Organize data for comparison
  - Use in AI/ML algorithms
  - Representation in graph databases

NCI's Approach - Critical Components

- Common Semantic Model supports Harmonization and Mappings
  - Declared by the US Office of the National Coordinator based on Health Level 7 Standards (HL7)
    - United States Core for Data Interoperability (USDCI)
      - Set of Classes and Data Elements

- Shared Concept System for Semantic Alignment
  - Domain Specific Concept System → NCI Thesaurus (more than a Thesaurus)
  - Used to annotate Common Semantic Model and align Data Elements

- Central Governance/Coordination for Data Registration in ISO/IEC 11179
  - Standard, Structured Metadata based on Concepts
  - Consistent application of the Concepts in the Semantic Model

Common Semantic Model

- US Office of the National Coordinator (ONC)
  - United States Core for Data Interoperability (USDCI)
- Register USDCI in ISO/IEC 11179 Registry
- Annotated with Concepts from NCI Thesaurus
- Map disparate Data Models to Common Semantic Model

USCDI v3 Summary of Data Classes and Data Elements

Shared Concept System

NCI Thesaurus - Formal declarations make meaning machine interpretable

Figure 2. The figure shows an example of a logical definition for the concept C9003, "adrenal cortex adenoma," in the NCIt.

Example from the National Cancer Institute Thesaurus

Common Semantic Model vs Common Data Model

- Shared Semantics
  - *Provides clear, unambiguous meaning*
  - Independent of Data Model
  - Enables mapping and transformation
- Annotated and Registered using Shared Concept Annotations
  - ISO/IC 11179, NCI Thesaurus
- Concept Systems enhance Knowledge Discovery

USCDI*
Shared Semantics
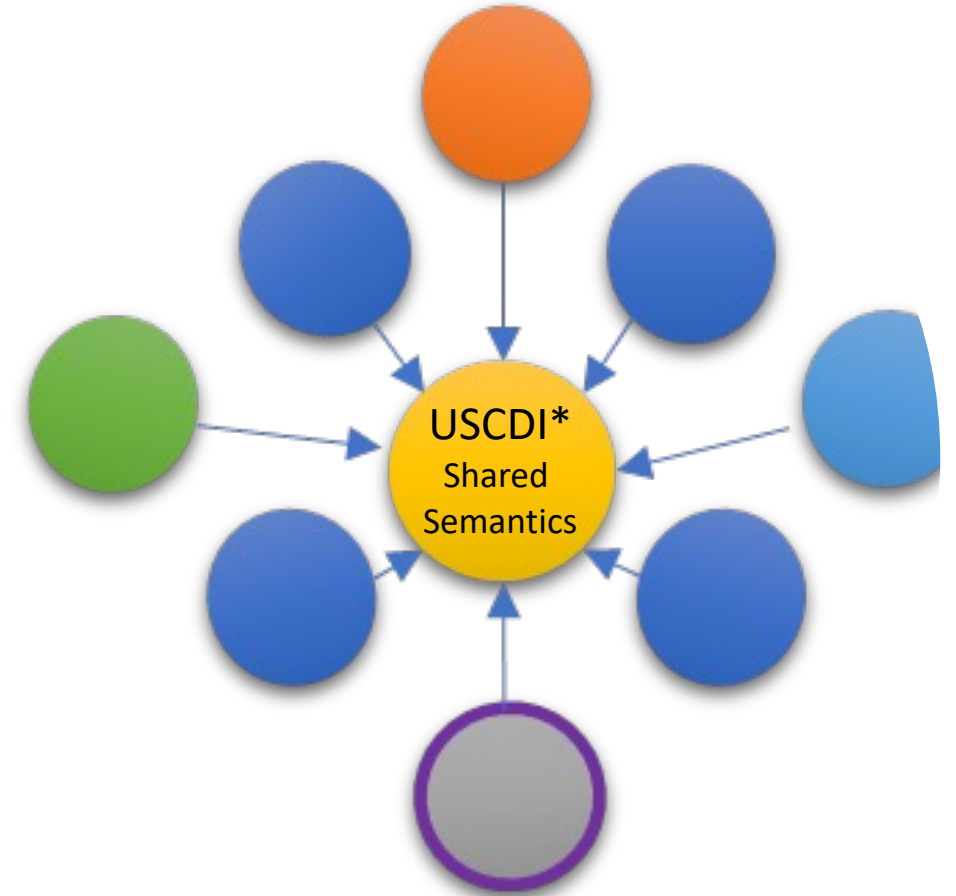
United States Core Data for Interoperability

References

- ISO/IEC 11179 Information technology — Metadata registries (MDR) — Part 3: Metamodel for registry common facilities
- ISO/IEC 11179 Information technology — Metadata registries (MDR) — Part 31: Data Specification
- ISO/IEC 5493 Criteria for concept systems
- United States Core for Data Interoperability (USDCI)
- NCI Semantics Primer - Blog
  - https://datascience.cancer.gov/news-events/blog/semantics-primer