

Metadata Standards and CDISC

Wayne R. Kubick

CDISC

National Children's Study Metadata Repository Workshop

January 2012

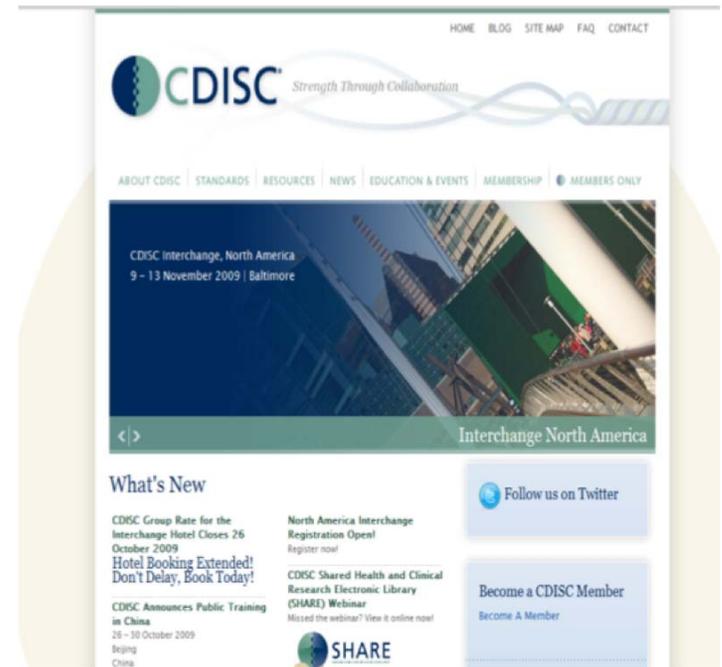


Strength through Collaboration

CDISC Snapshot

- Global, open, multi-disciplinary, vendor-neutral non-profit standards developing organization (SDO)
 - Founded in 1997; incorporated in 2000
 - > 290 organizational members (academia, biopharma, service and technology providers, etc)
 - Liaison A Status with ISO TC 215
 - Charter agreement with HL7 since 2001; Board
 - Member/Leader of Joint Initiative Council (JIC) for Global Harmonization of Standards
 - Member of ANSI-led ISO TAG
 - Seat on U.S. HIT Standards Committee
 - Active Coordinating Committees
 - Europe, Japan, China, Korea
 - Standards downloaded in over 65 countries

Established global industry standards to support the electronic acquisition, exchange, submission and archiving of data to streamline biomedical research (open via www.cdisc.org)



www.cdisc.org

The Clinical Data Interchange Standards Consortium Vision

*Informing patient care and safety
through higher quality medical
research*



What is Metadata?

- Metadata ('data about data') is defined as data providing information about aspects of the data:
 - Descriptive: definition, source, purpose, comments
 - Structural: data type, attributes, format, terminologies, relationships, standards and constraints
 - Administrative: date/time of data creation version, audit trail, provenance
- Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource

»Adapted from Wikipedia

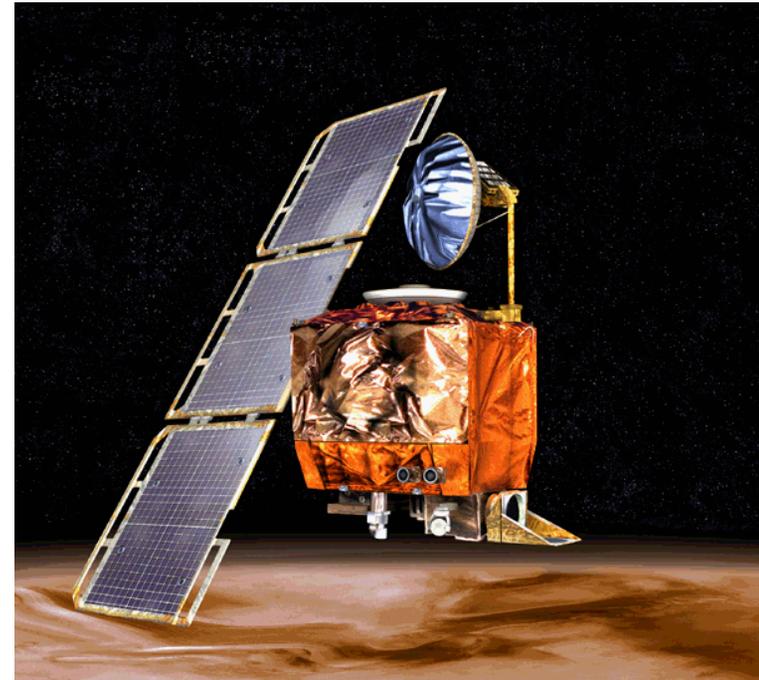
How Valuable is Metadata?

Event	Time	D	S	F
Begin	9/23/99 02:01:00	121,900,000	12,300	143.878
End	9/23/99 02:17:23		9,840	

Event	Time	D	S	F
Start	19990923 05:01:00	196,200,000	5.5	640
Finish	19990923 05:17:23		4.4	

Source – Dave Christiansen, 2000

In this case \$125,000,000: Mars Climate Orbiter



Mars Orbit Insertion Burn	M/D/Y HH:MM:SS PDT (Earth Receive Time, 10 min. 49 sec. Delay)	Distance (miles)	Speed (miles/hr)	Force (Pounds)
Begin	9/23/99 02:01:00	121,900,000	12,300	143.878
End	9/23/99 02:17:23		9,840	
Mars Orbit Insertion Burn	YYYYMMDD EDT (Earth Receive Time, 10 min. 49 sec. Delay)	Distance (km)	Speed (km/sec)	Force (Newtons)
Start	19990923 05:01:00	196,200,000	5.5	640
Finish	19990923 05:17:23		4.4	

Source – Dave Christiansen, 2000

CDISC ADaM Analysis Metadata

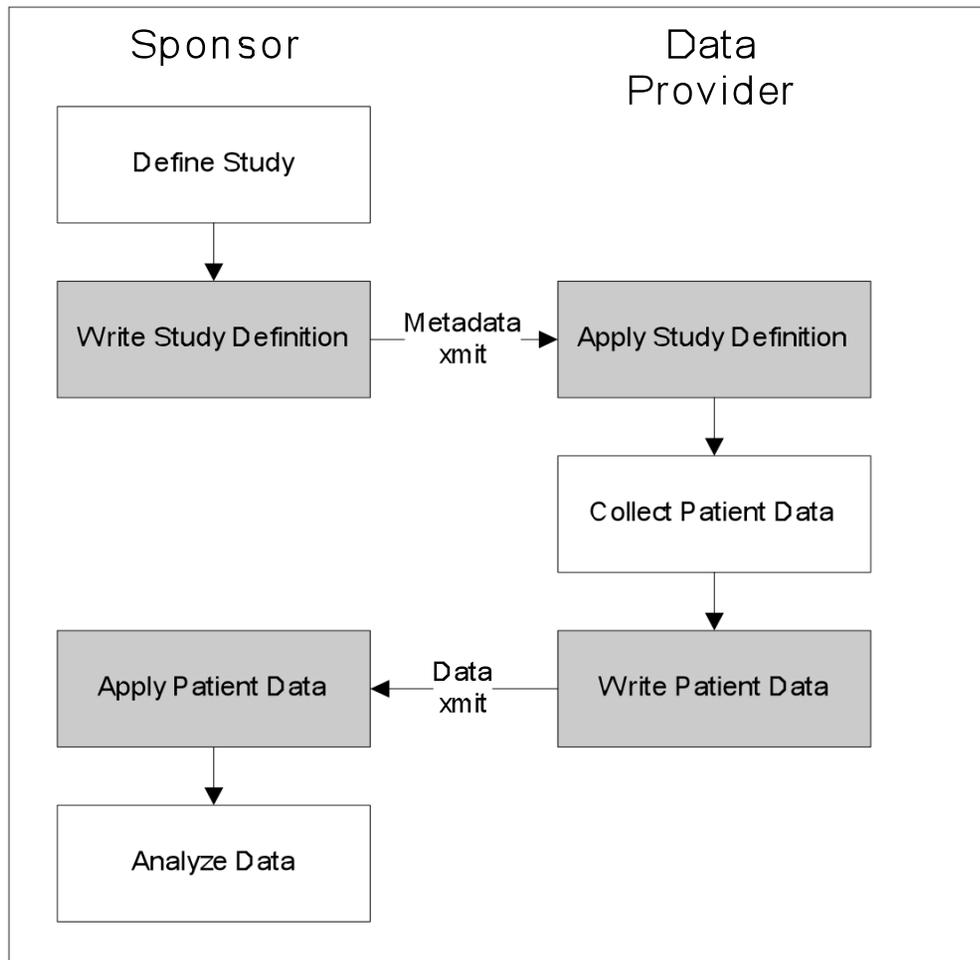
- Analysis dataset metadata
 - Describe each dataset as a whole
- Analysis variable metadata
 - Describe each variable within a dataset
- Analysis results metadata
 - Link analysis result with analysis plans, programs, data
 - Characterizes overall analysis – were objectives met?
- Thus, describe each statistical object in detail:
 - Statistical methods
 - Transformations
 - Assumptions
 - Derivations and imputations performed

Data Provenance

- Provenance, from the French *provenir*, "*to come from*", refers to the chronology of the ownership or location of an historical object. This will typically be accomplished by tracing the whole history of the object up to the present ... *establishing provenance is essentially a matter of documentation.*
- Scientific research is generally held to be of good provenance when it is documented in detail sufficient to allow reproducibility. Scientific workflows assist scientists and programmers with *tracking their data through transformations, analyses and interpretations.*

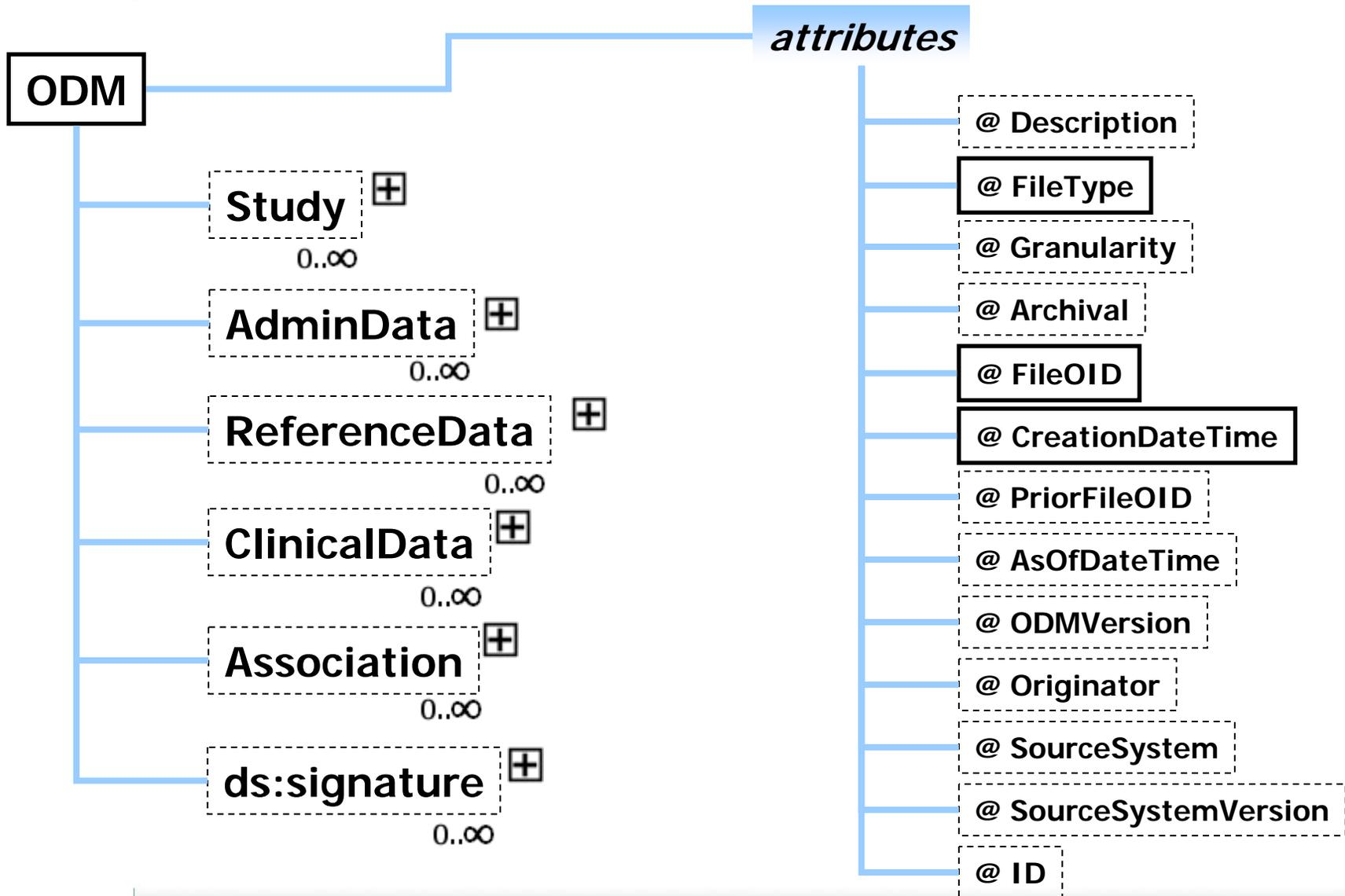
Source: Wikipedia

CDISC Operational Data Model Study Data Interchange Scenario

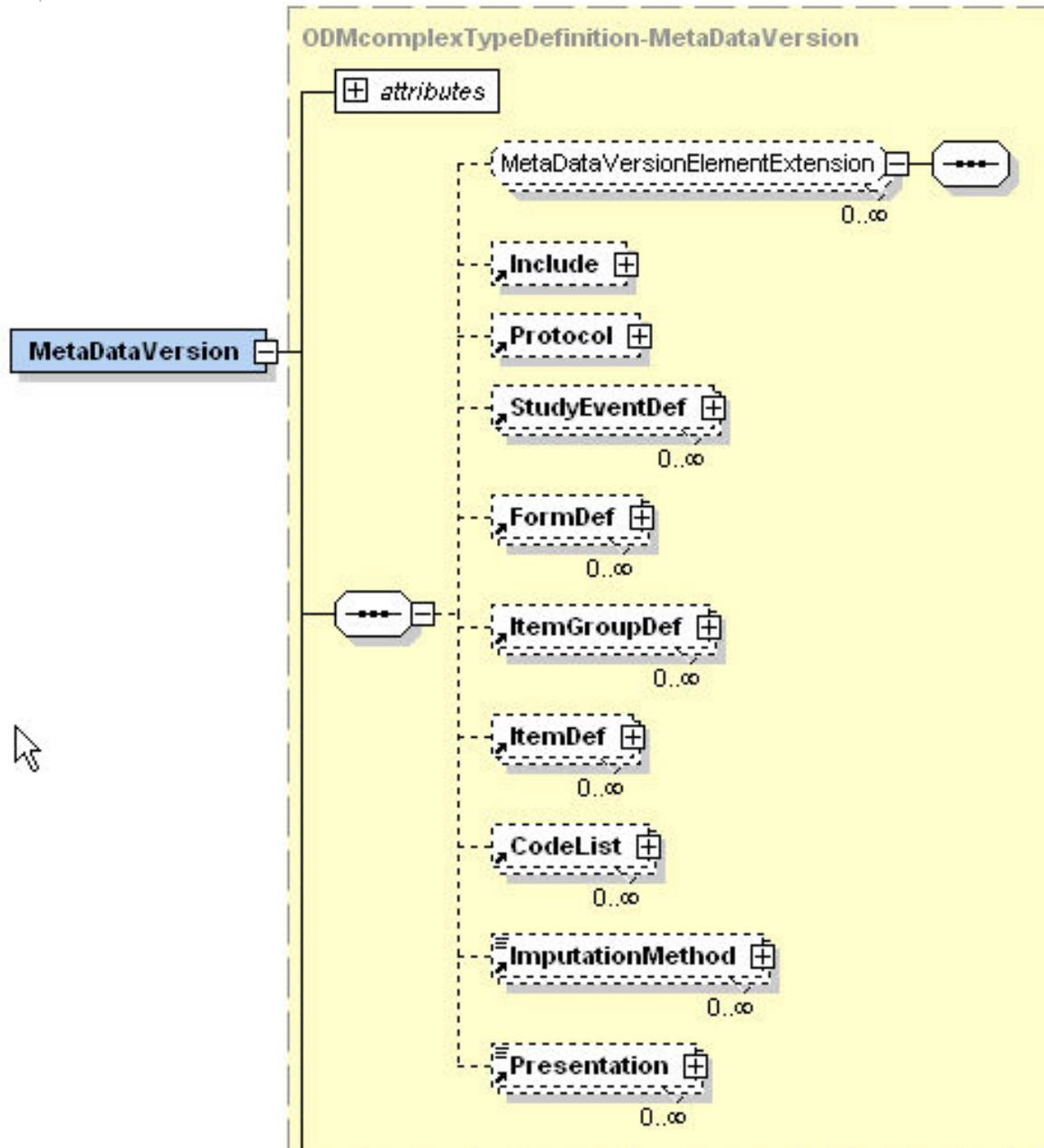


- XML transmission:
 - Metadata
 - Data
 - Audit Trail
 - Traceability and Transformations

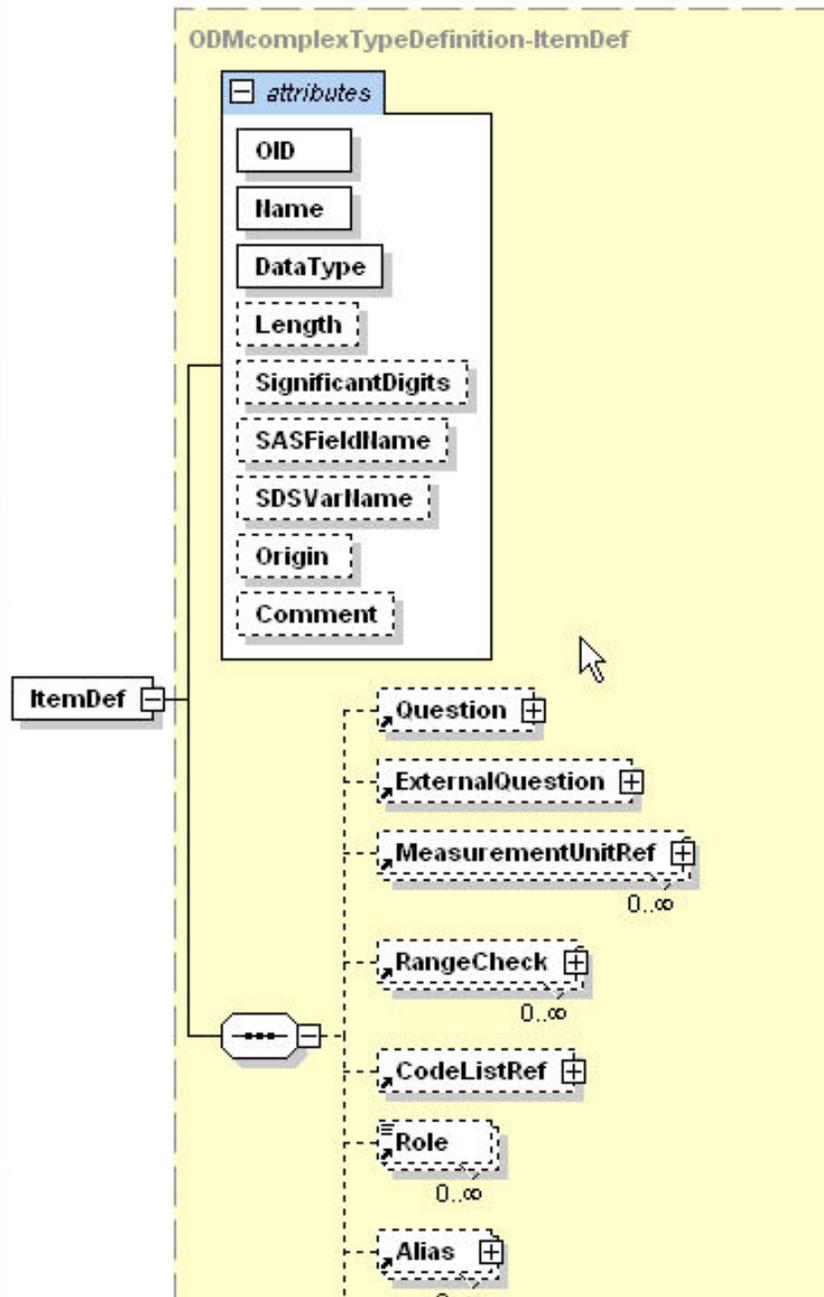
ODM Element



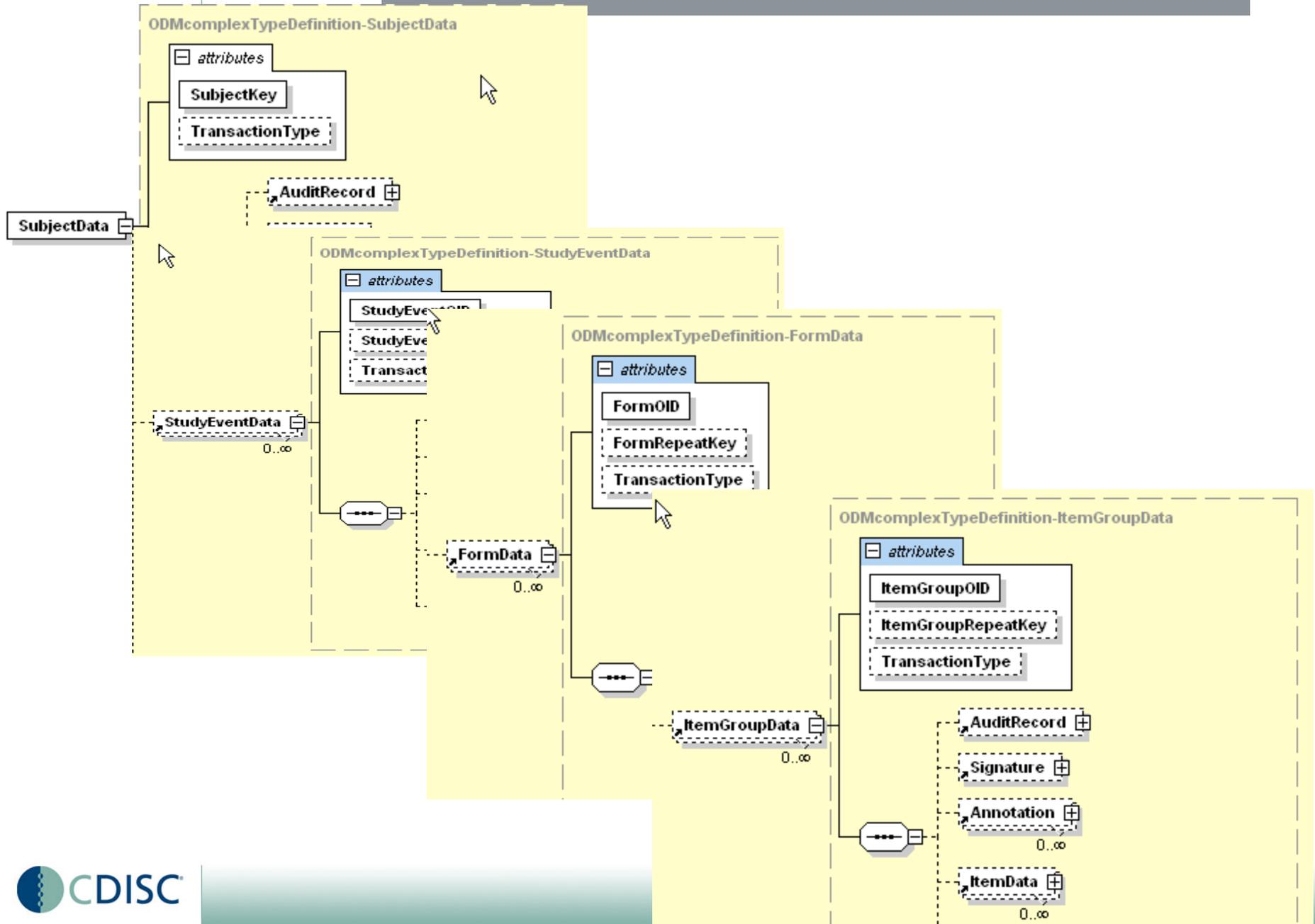
ODM MetaDataVersion Element



ItemDef

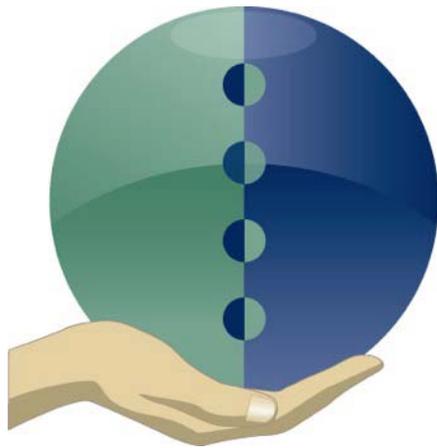


Clinical Data Hierarchy



Need for a Global Metadata Repository

- To enhance Data Quality and Compliance
- To decrease the time needed to aggregate and review results
- Machine readable standards to improve “compliance”
- Illustrate inherent relationships between metadata
- Speed up standards development



CDISC
SHARE
SHARED HEALTH AND CLINICAL RESEARCH ELECTRONIC LIBRARY

CDISC SHARE VISION

A global, accessible electronic metadata library, which through advanced technology, enables precise and standardized data element definitions **and richer metadata** that can be used in applications and studies to improve biomedical research and its link with healthcare.

<http://www.cdisc.org/cdisc-share>

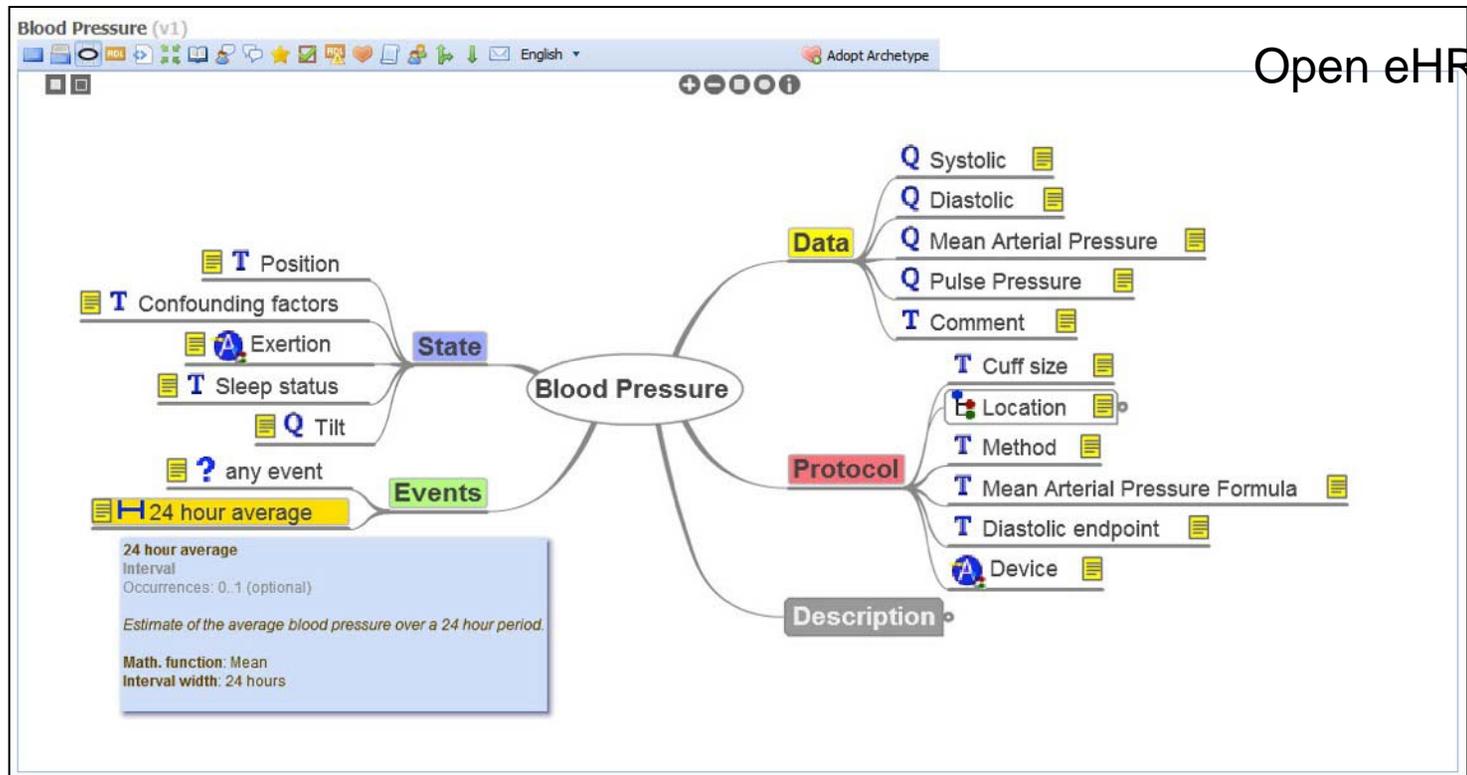
CDISC SHARE Library Contents

- Metadata (SDTM and CDASH)
 - Trial Design Metadata
 - Definitions
 - Datatypes (ISO 21090)
- Links to controlled terminology (CT) dictionaries via the NCIIt (which links to CDISC CT, SNOMED, ICD9, ICD10, UMLS, etc.)
- Implementation instructions
- CDASH CRF metadata and instructions

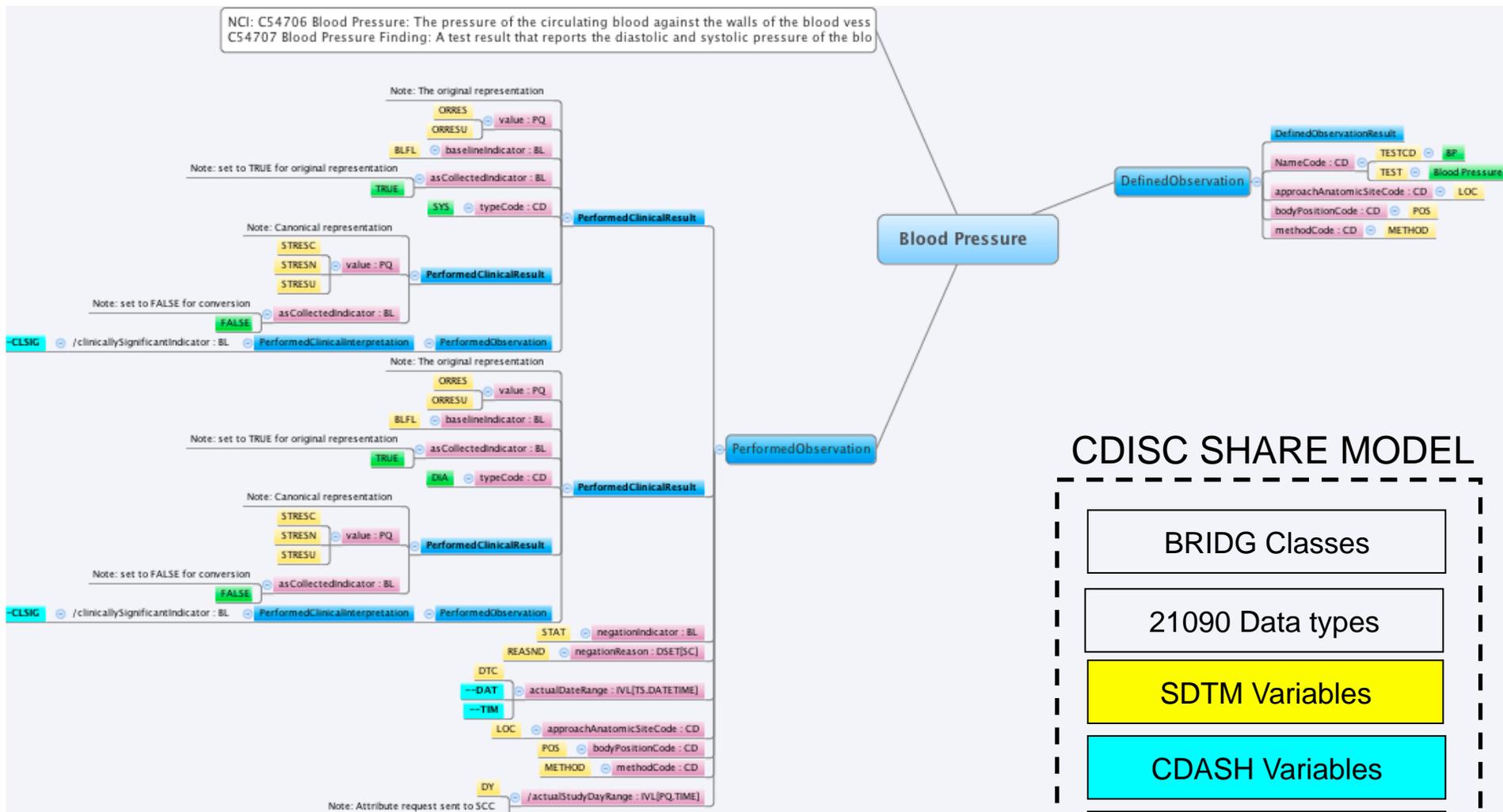
Which Metadata Model?

SDTM

USUBJID	VSSEQ	VSTESTCD	VSTEST	VSPOS	VSLOC	VSORRES	VSORRESU
ABC-001-001	12	SYSBP	Systolic Blood Pressure	SITTING	LEFT ARM	95	mmHg
ABC-001-001	13	DIABP	Diastolic Blood Pressure	SITTING	LEFT ARM	44	mmHg

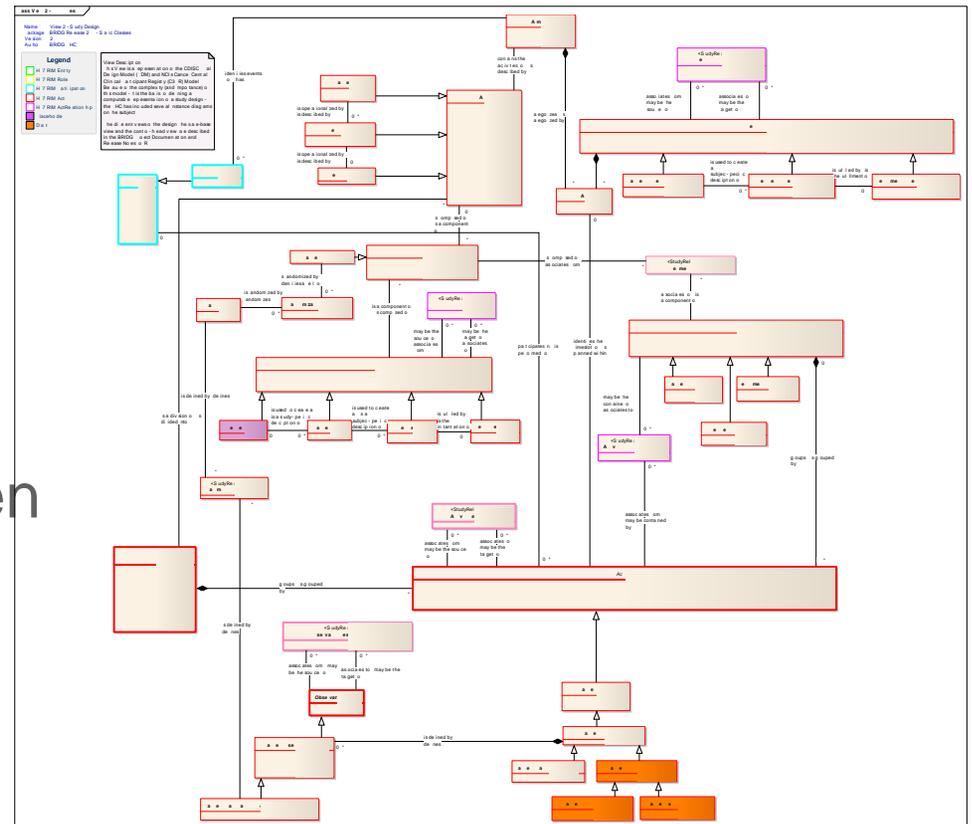


The CDISC SHARE MODEL

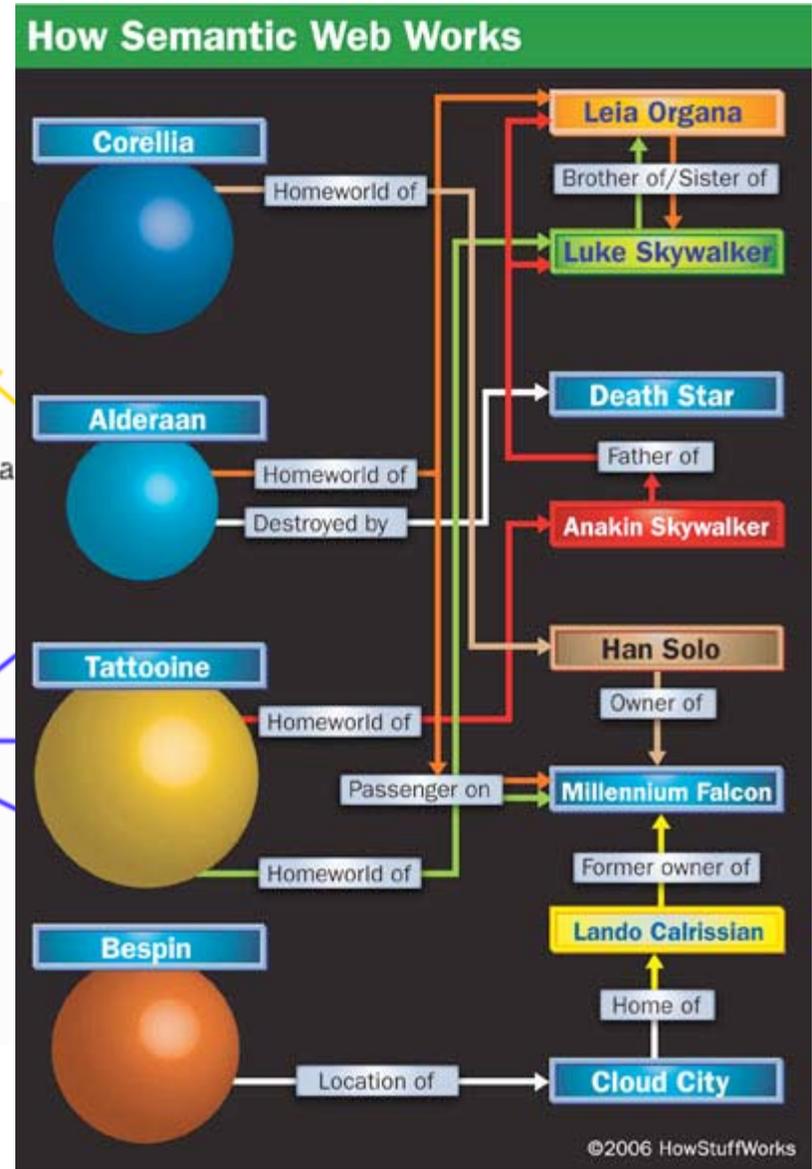
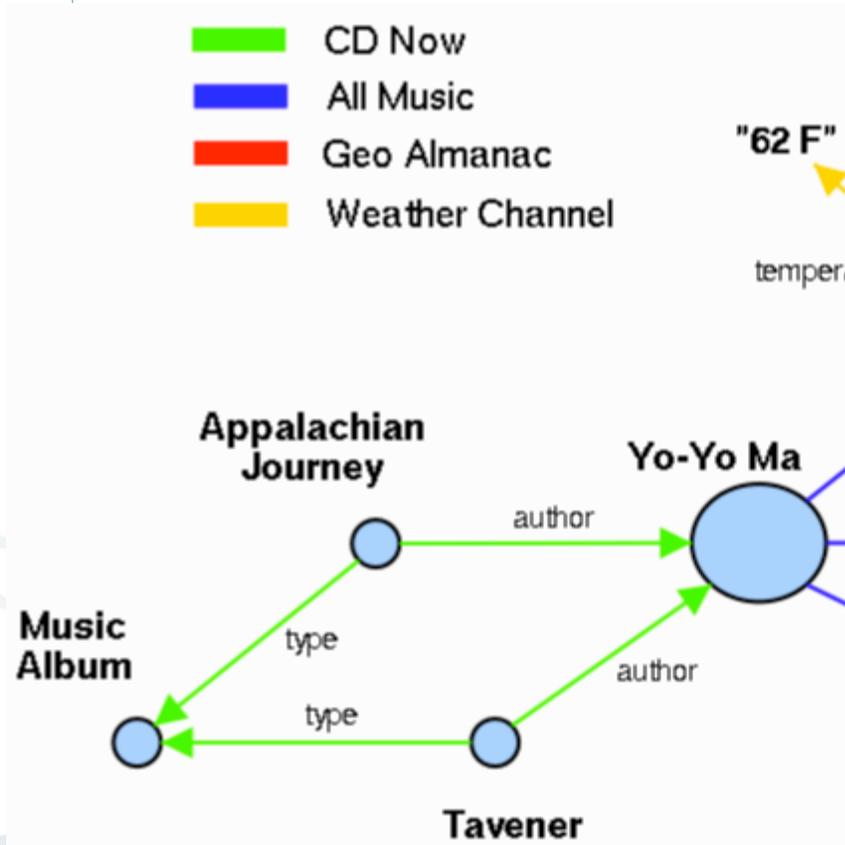


BRIDG | Biomedical Research Integrated Domain Group

- A domain analysis information model representing protocol-driven biomedical/clinical research
 - Provides a **basis for harmonization** among standards within the clinical research domain and between biomedical/clinical research and healthcare.
 - ISO 21090 compliant, HL7 alignment, RIM alignment
- <http://www.cdisc.org/bridg>

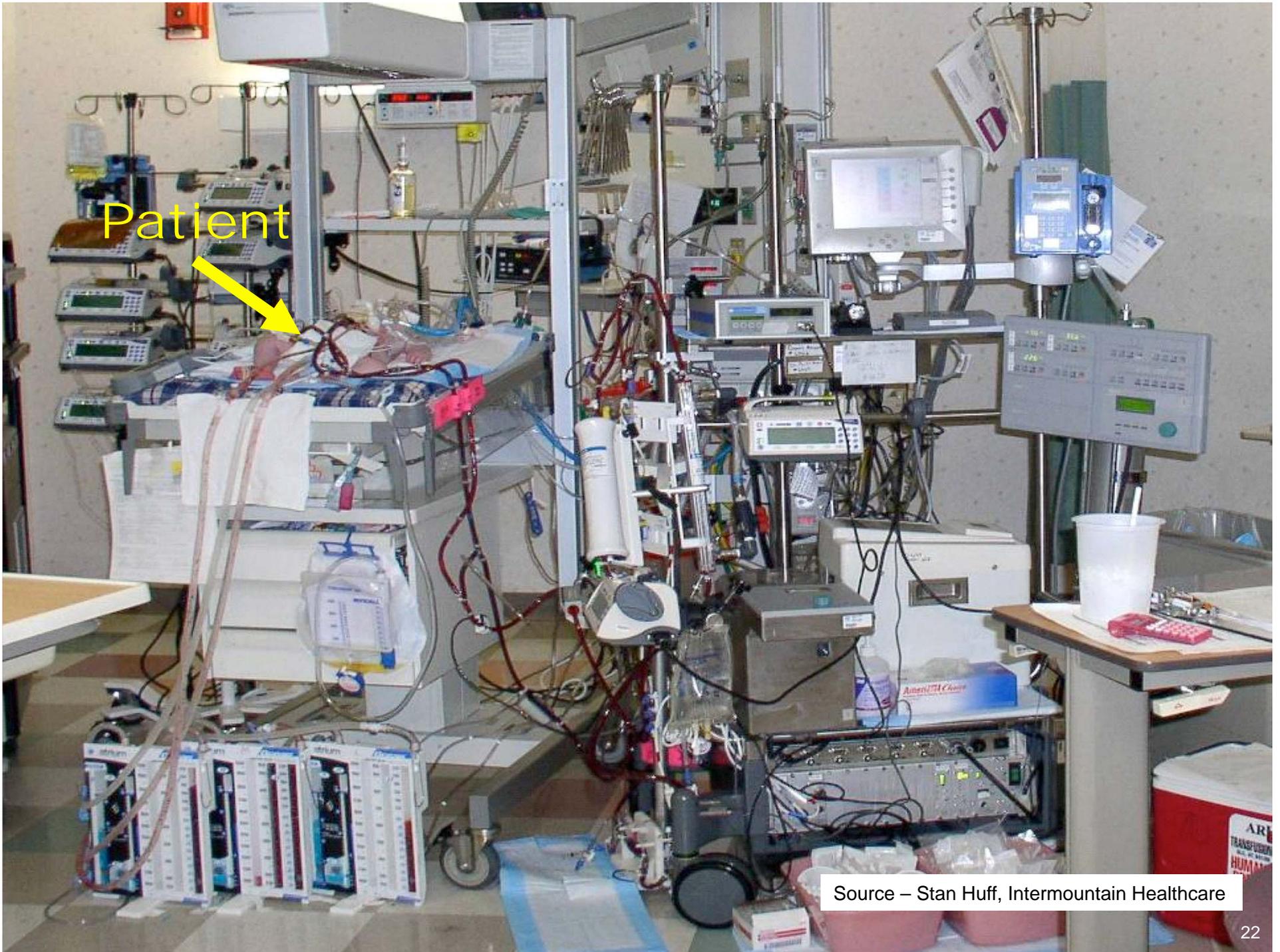


The Semantic Web



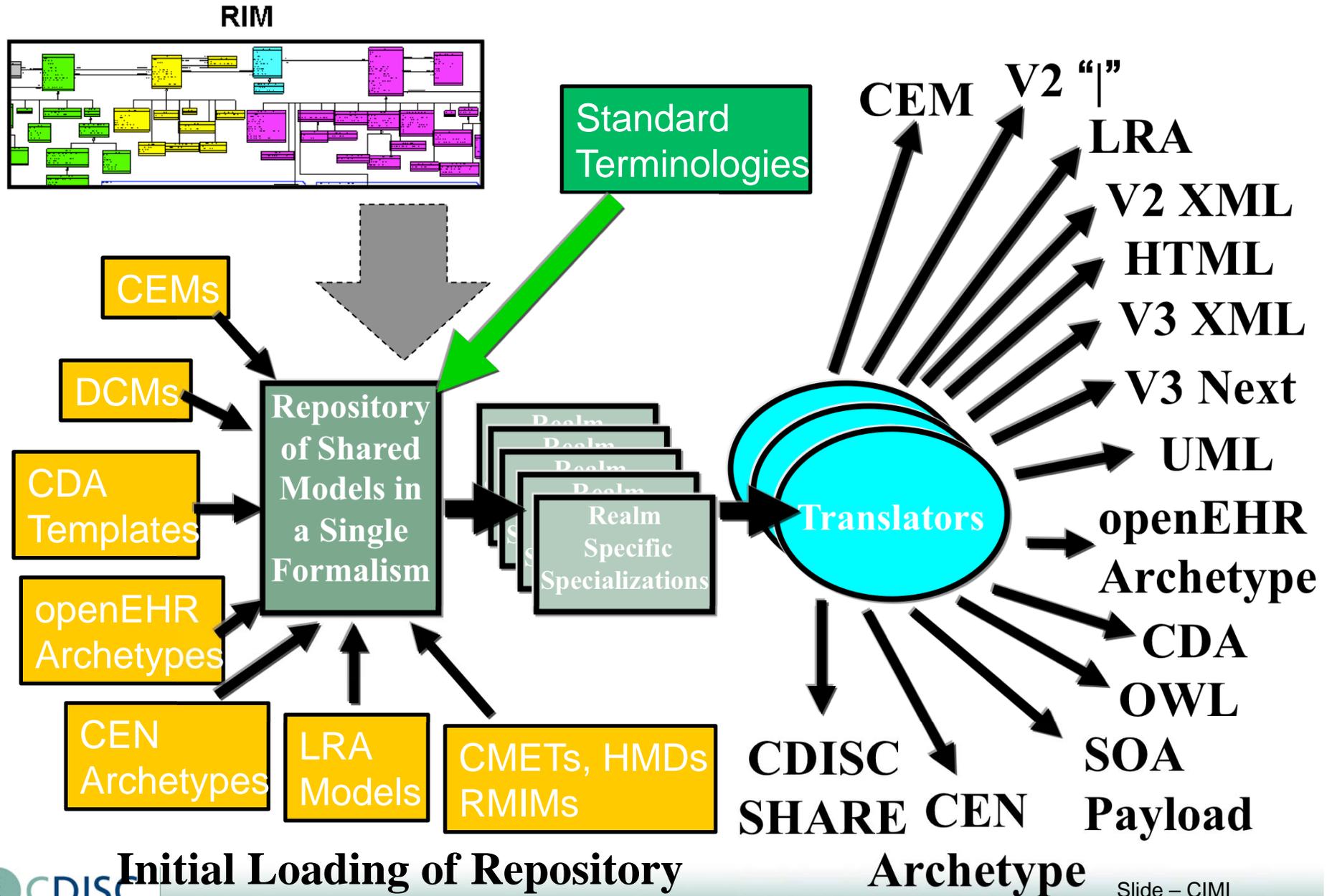
Clinical Information Modeling Initiative

- Improve the interoperability of healthcare systems through shared implementable clinical information models using:
 - **Shared repository** of detailed clinical information models
 - Using a single formalism
 - Based on a common set of base data types
 - With formal bindings of the models to standard coded terminologies
 - **Repository is open and models are free** for use at no cost
- Expressed by Intermountain Healthcare as:
 - Be able to share applications, reports, alerts, protocols, and decision support with anyone in the WORLD



Source – Stan Huff, Intermountain Healthcare

CIMI Information Models



Decomposition Mapping

Precoordinated Model (User Interface Model)

SystolicBPRightArmSittingObs

SystolicBPRightArmSitting

data 138 mmHg

Post coordinated Model (Storage Model)

SystolicBPObs

SystolicBP

data 138 mmHg

quals

BodyLocation

BodyLocation

data Right Arm

PatientPosition

PatientPosition

data Sitting

Source – Stan Huff, Intermountain Healthcare

Data Entry Styles

Hair Color

- Brown
- Blonde
- Red

Hair Color

Brown

**Brown
Blonde
Red**

Evaluation Styles

Finding

Brown hair

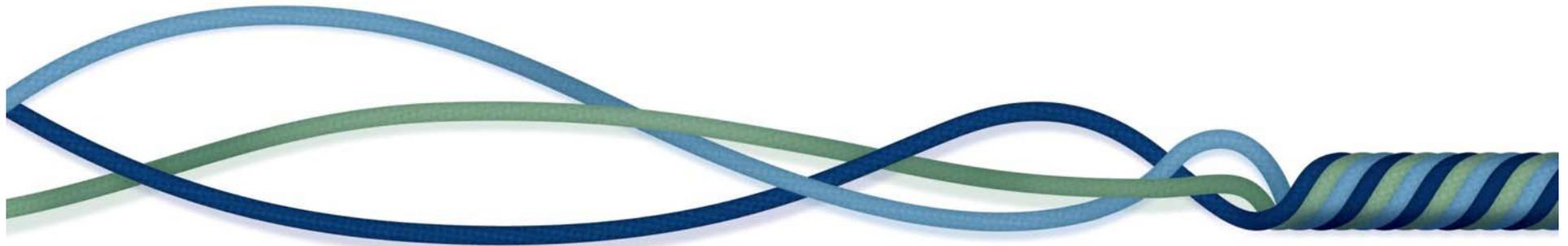
**Brown hair
Blonde hair
Red hair**

Assertion Style

Top Three Metadata Challenges

1. Identifying and maintaining reusable data elements across different child life stages and over time as technologies, devices and protocols change...
2. Representing other data types besides questionnaires in the data collection stage of operations...
3. Providing frameworks for the construction and use of objects like health development trajectories and exposomes...





Strength through collaboration.

Wayne R. Kubick
wkubick@cdisc.org