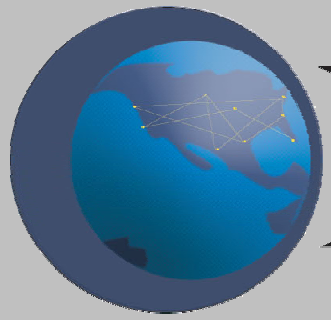


caBIG™ cancer Biomedical Informatics Grid™

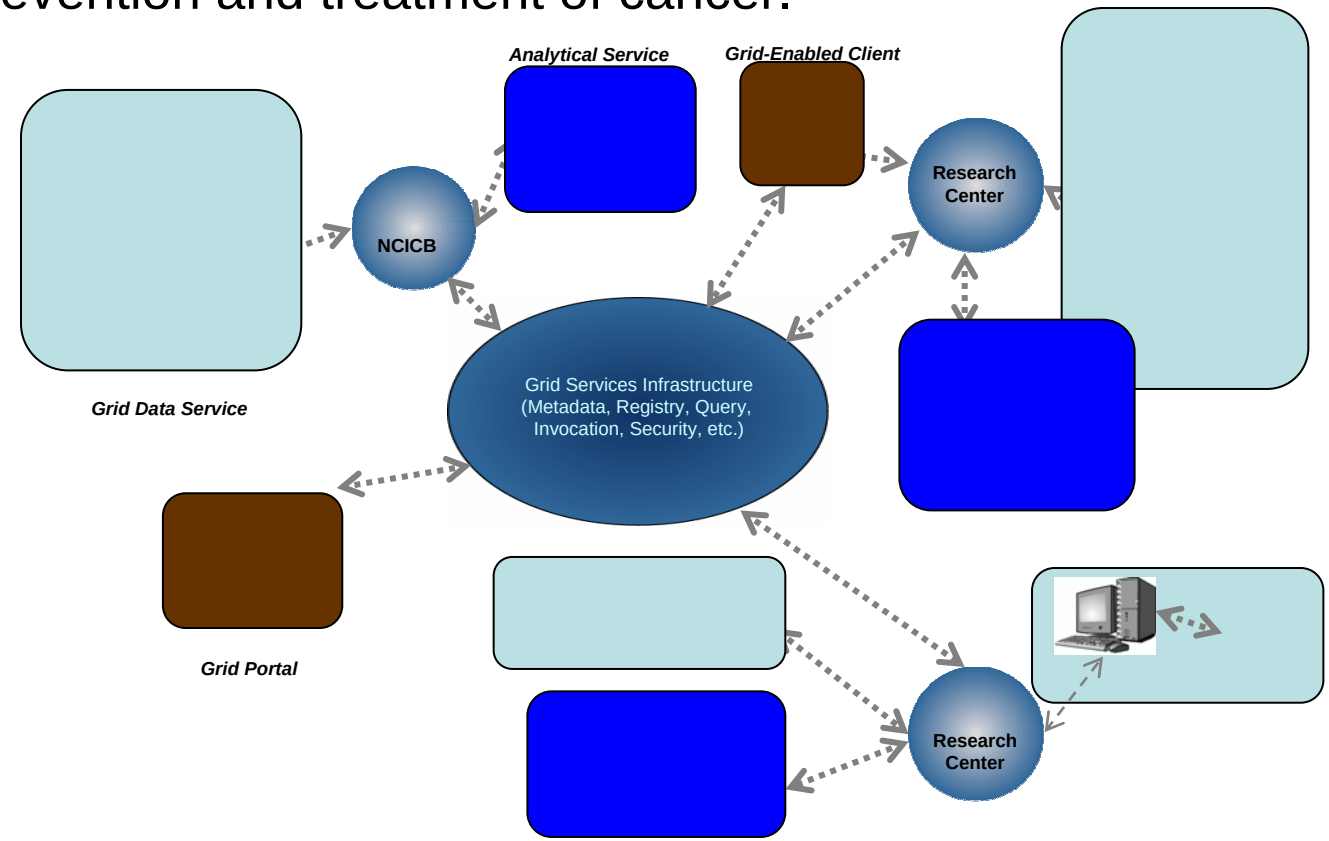
an initiative of the National Cancer Institute

The BRIDG Project: Creating a model of the semantics of clinical trials research

Douglas B. Fridsma, MD PhD
Center for Biomedical Informatics
University of Pittsburgh School of Medicine
Pittsburgh PA
fridsma@cbmi.pitt.edu

caBIG™, What is it?

- The cancer Biomedical Informatics Grid™, or caBIG™, is a virtual network connecting individuals and organizations to enable the sharing of data and tools, creating a World Wide Web of cancer research.
- The goal is to speed the delivery of innovative approaches for the prevention and treatment of cancer.

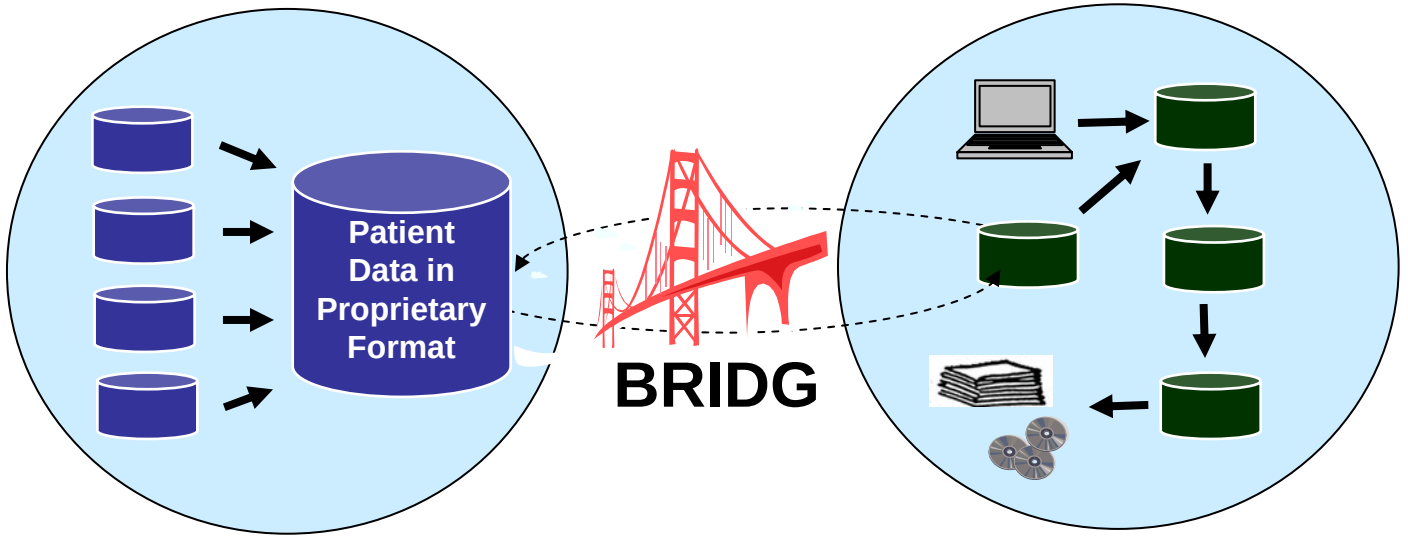


caBIG™ Objectives

- Common, widely distributed infrastructure that permits the cancer research community to focus on innovation
- *Shared*, *harmonized* set of terminology, data elements, and data models that facilitate information exchange
- Collection of *interoperable applications* developed to common standards
- Raw published cancer research data is available for mining and integration

Standards – Why?

Cancer Research: Two Different Worlds



Patient Care World

- Multiple data sources and types
- HL7 is a pervasive standard
- Data are organized around the patient

Clinical Research World

- Protocol defines define elements
- Linear data flow
- CDISC is the emerging standard
- Data are organized around a trial

Acknowledgements:
Landen Bain, CDISC

So why do we need a model of the semantics of clinical trials for caBIG?

“The Map is not the Territory” (Bertrand Russell)

- Domain Experts have a “mental map” of the problems that they hope technology can solve
- In gathering requirements, this map may have flaws or distortions

- Databases schemas are not the territory
  - Implicit semantics in the structure or value-attribute pairs

Acknowledgment: Charlie Mead

The Map is not the Territory

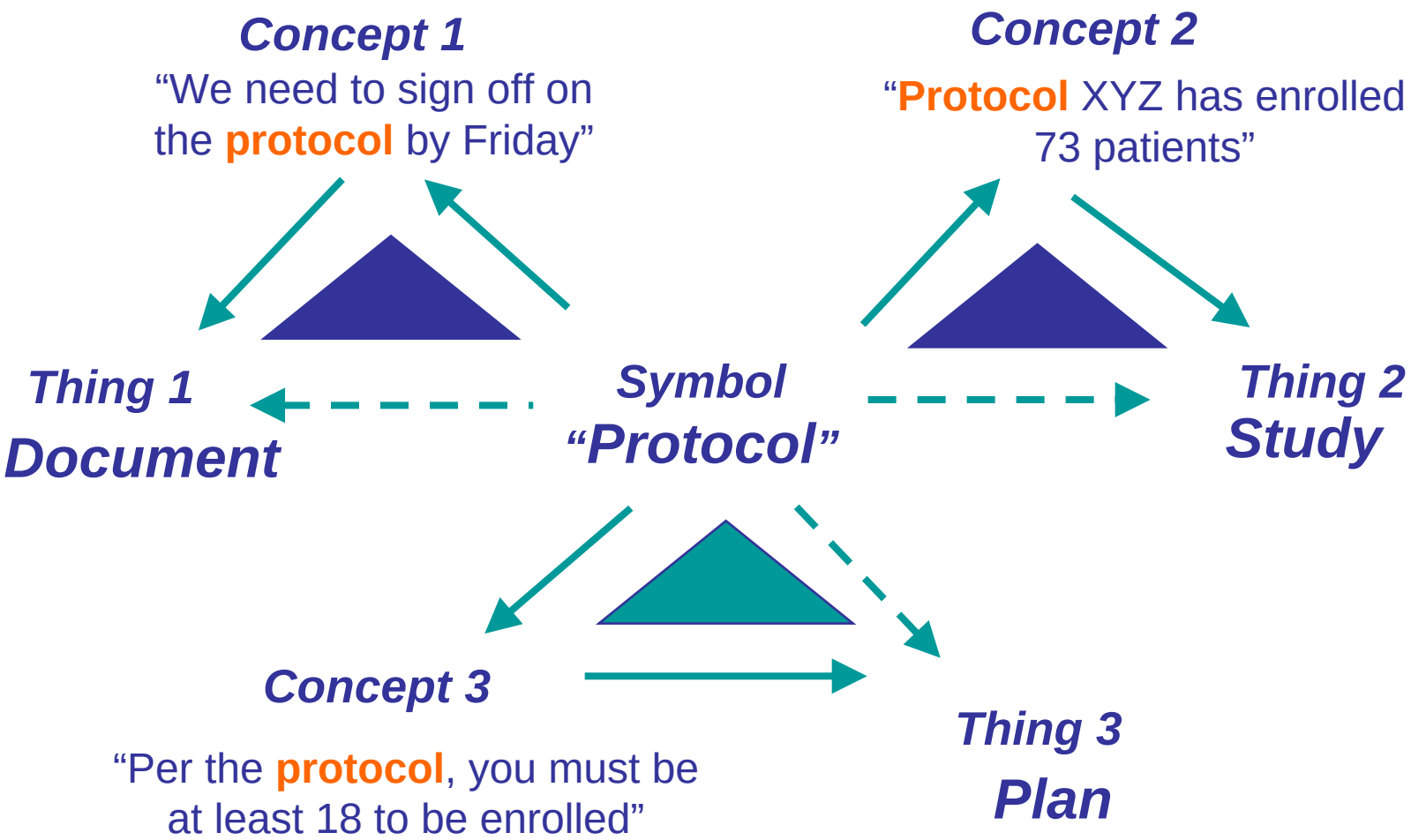
- Deletion (filtered/missing details)– "They use the system to find information about clinical trials."
  - Challenge: "Who uses the system?"
  - Response: "Clinical research coordinators, patients, and investigators."

- Distortion (incorrect or modified details) – "You can't enter a clinical trial protocol until you have an protocol identification number."
  - Challenge: "Are there any circumstances where you can enter a protocol without an identification number?"
  - Response: "Yes, two circumstances…."

- Generalization (abstractions via rules, beliefs, principles)– "Everyone must have a log-on ID to access the information in the system."
  - Challenge: "Are there any system users that can access the information without an log-on ID?"
  - Response: "Organizations and cooperative groups may use the API to access the information directly."

Acknowledgment: Charlie Mead

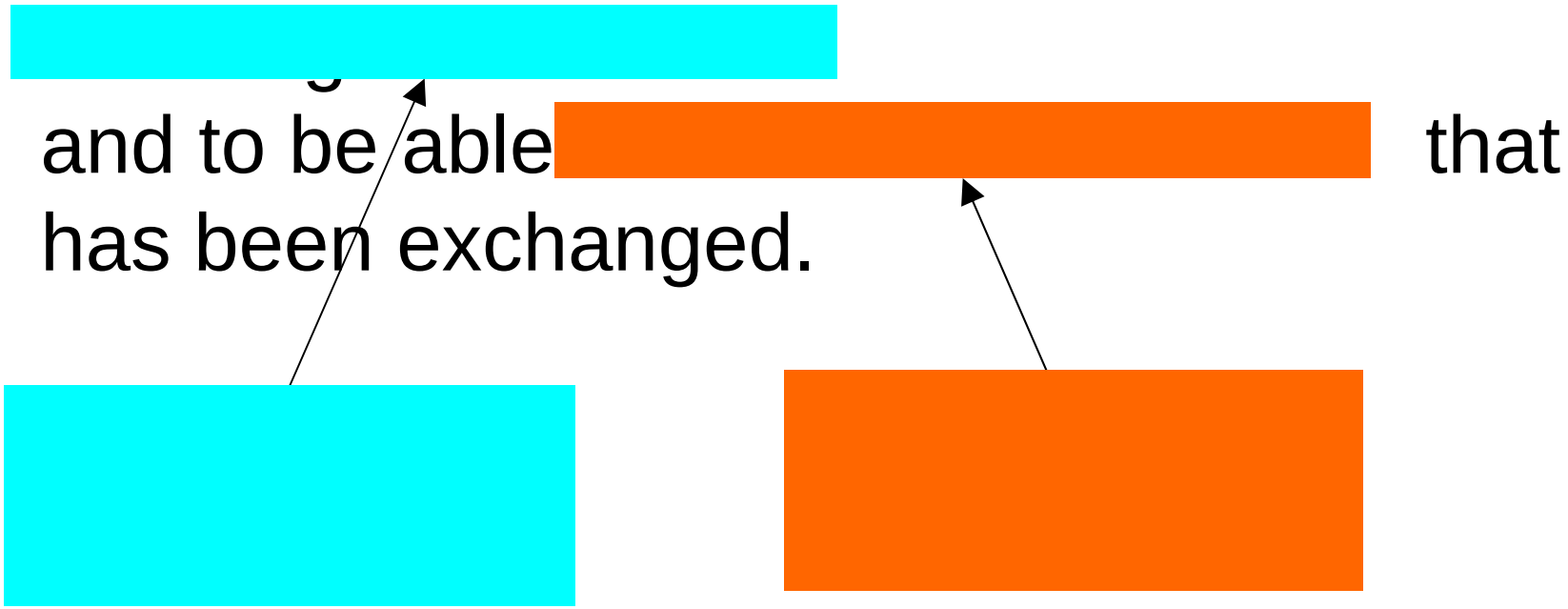
“Protocol” and the Semiotic Triangle



Source: John Speakman/Charlie Mead

Interoperability

- The ability of multiple systems to



and to be able that has been exchanged.

The Pillars of (Semantic) Interoperability

Necessary but not Sufficient

- **Common model across all domains-of-interest**
  - **The representation of clinical trials in BRIDG**
- **Model grounded on robust data type specification**
  - **Common data elements (ISO 11179) in the cancer Data Standards Repository (caDSR)**
- **Methodology for binding terms from concept-based terminologies**
  - **UML loader, semantic connector, Enterprise Vocabulary Server**
- **A formally defined process for defining specific structures to be exchanged between machines, i.e. a "messaging standard"**
  - **HL7 and implementation specifications**
  - **caBIG unified process/model driven architecture**

What is BRIDG?

- A formal model of the *shared semantics* of regulated clinical trials research
- A communication *bridge* between
  - clinical trial domain experts and technical experts
  - different models of clinical trials information
- An open *community of stakeholders* interested in developing standards for exchanging information about clinical trials
  - HL7 Domain analysis model in Regulated Clinical Research (RCRIM) technical committee
  - caBIG analysis model for model-driven development
  - CDISC integrating model for current standards
- The semantic *foundation for application and message development* in HL7, caBIG, and CDISC

So how did we get started?

- Desiderata
  - We did not want to create “yet another protocol representation”
    - “the good thing about standards is that there are so many to chose from…”
  - We wanted the work to be
    - Open
    - Collaborative
    - Standards-based

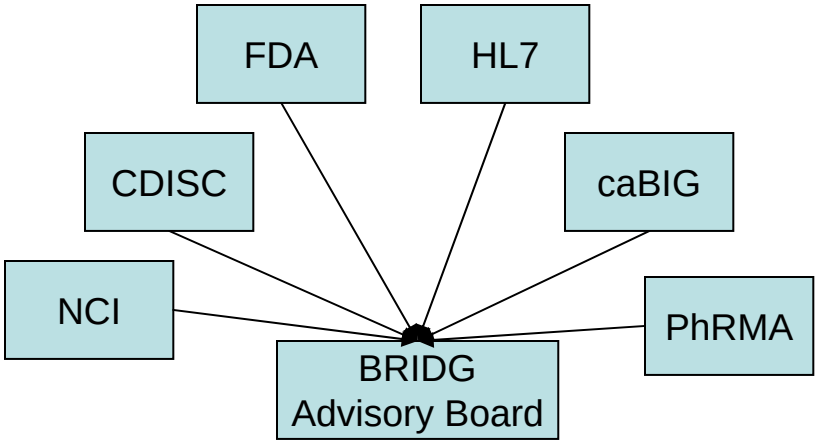
caBIG and the Development of Structured Protocol Representation

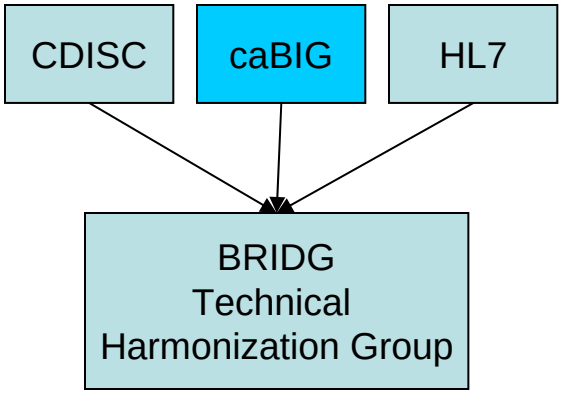
- **Spring 2004** – kick-off of the caBIG project
- University of Pittsburgh award the contract to develop a structured protocol representation to support clinical trials

Merging the caBIG and CDISC projects

- **Fall 2004** – caBIG identified "best of breed" models in the CDISC standards and HL7 messages
  - CDISC started domain modeling in 2003 to integrate their own modeling efforts and to link CDISC to HL7
- **November 2004** – First joint CDISC/HL7/caBIG modeling session
- **Between November 2004 and March 2005** – multiple modeling sessions to develop the "scaffolding" of the domain analysis model (SPR)
- **March 2005 to now**
  - Development of the dynamic aspects of the model
  - Develop scalable processes to support collaboration and expansion of the model, based on software best practices
  - Initiation of 8 subdomain projects within BRIDG

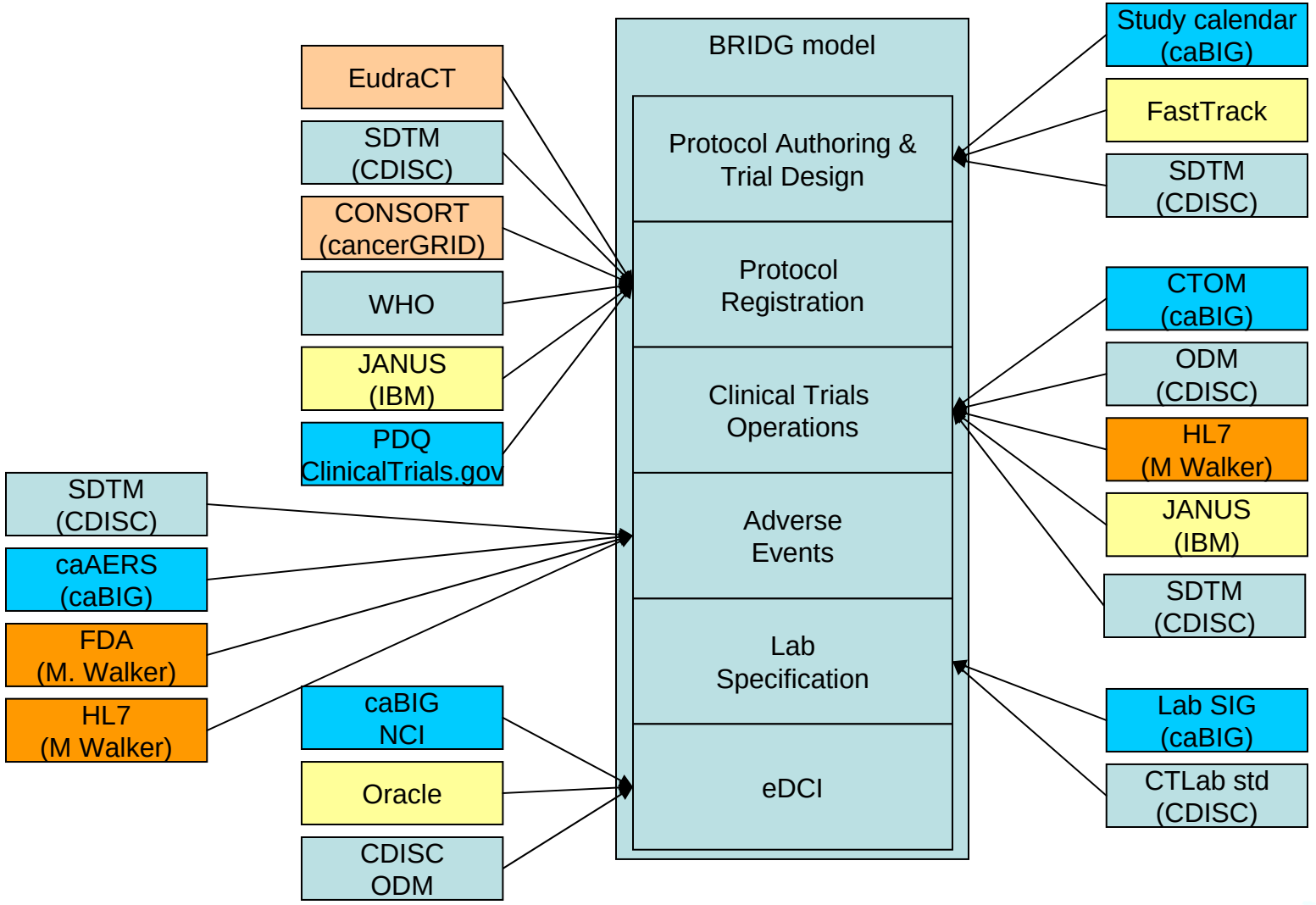
Current Organization of the BRIDG project





- BRIDG Advisory Board
  - Representation from the current stakeholders
  - Help to identify priorities and allocate resources
  - Assist with vetting the model in the various constituents
- Technical Harmonization Group
  - Responsible for ongoing model maintenance
  - Developing shared harmonization processes
- Multiple subdomain projects
  - Representation from pharmaceutical companies, technology companies, government agencies, and cancer centers

BRIDG projects and contributors

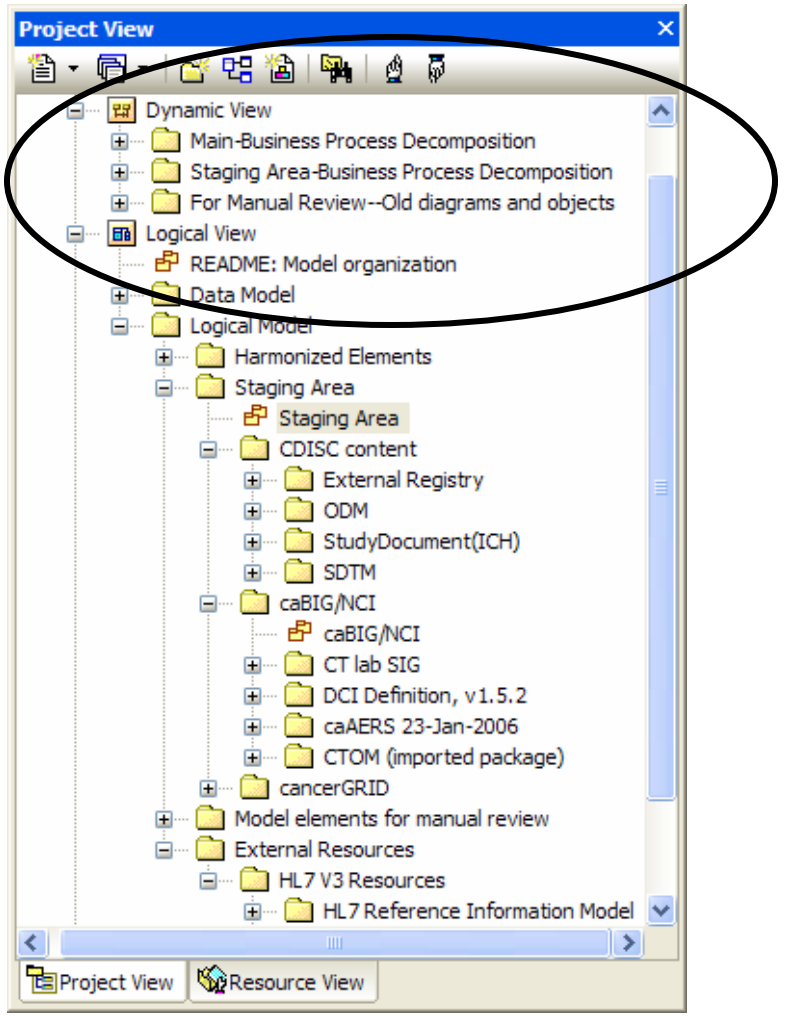


Principles for model organization

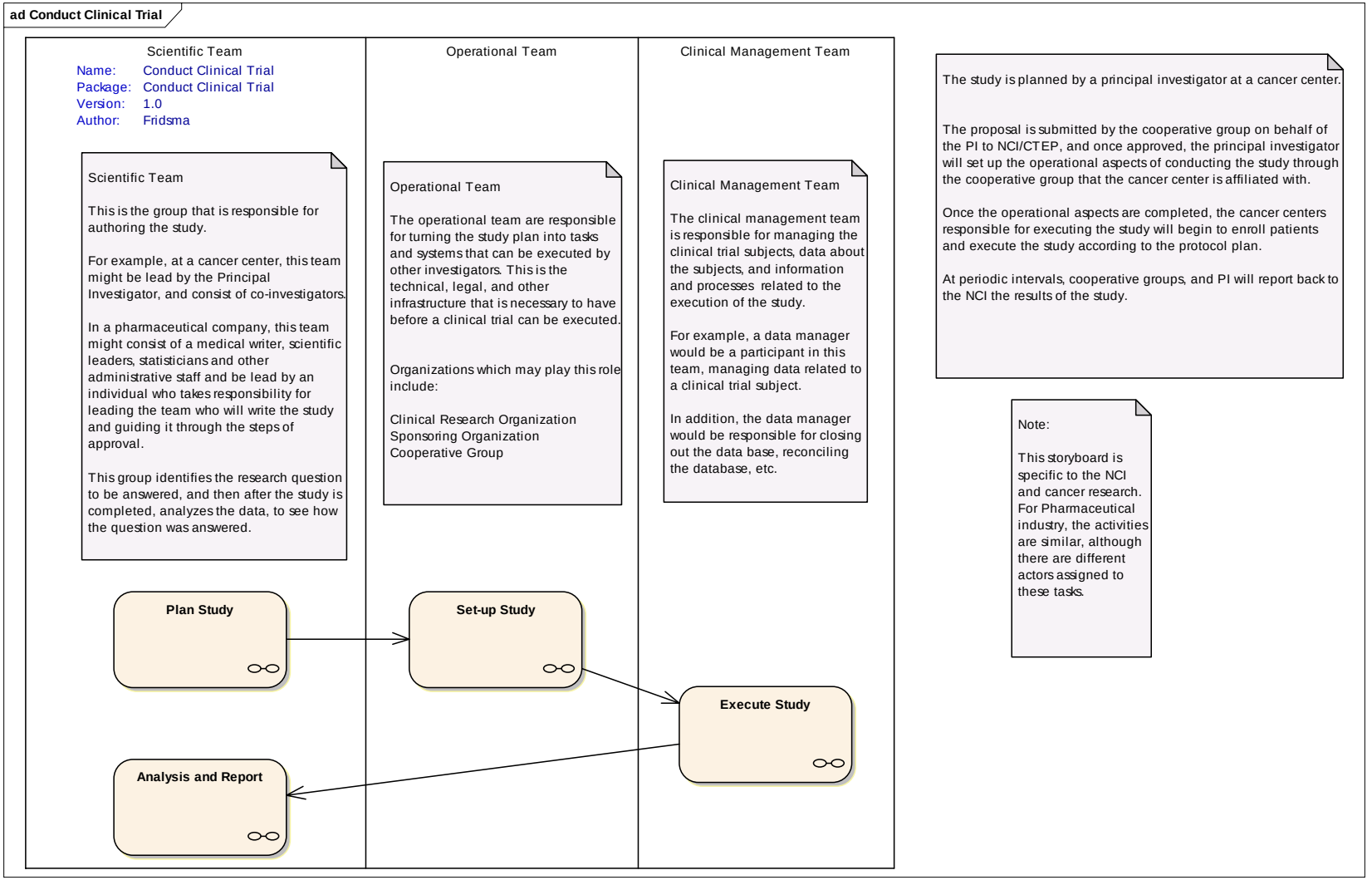
- Make the work process explicit
  - Recognizes that concepts and models are in different stages of development and harmonization
- Provide a mechanism to scale the development work
  - Parallelize the development
  - Prevent collaborators from “colliding” with each other
- Allows us to modeling in the open

Model organization

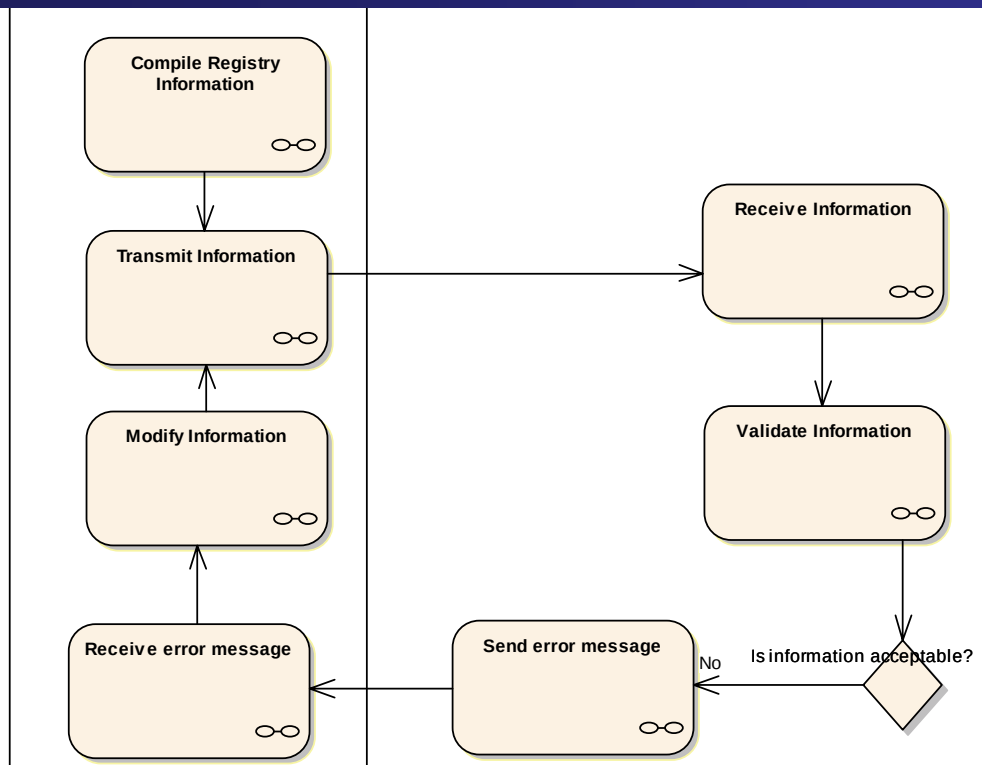
- Dynamic View
  - Captures the business process decomposition of the lifecycle of clinical trials research

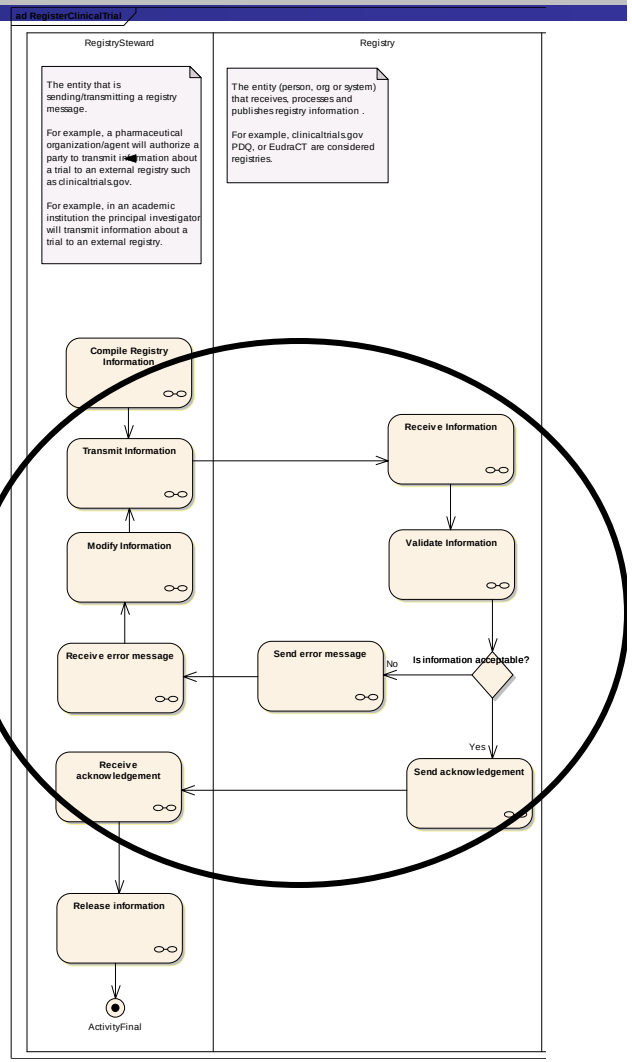


Behavioral Aspects of BRIDG



Behavioral Aspects of BRIDG

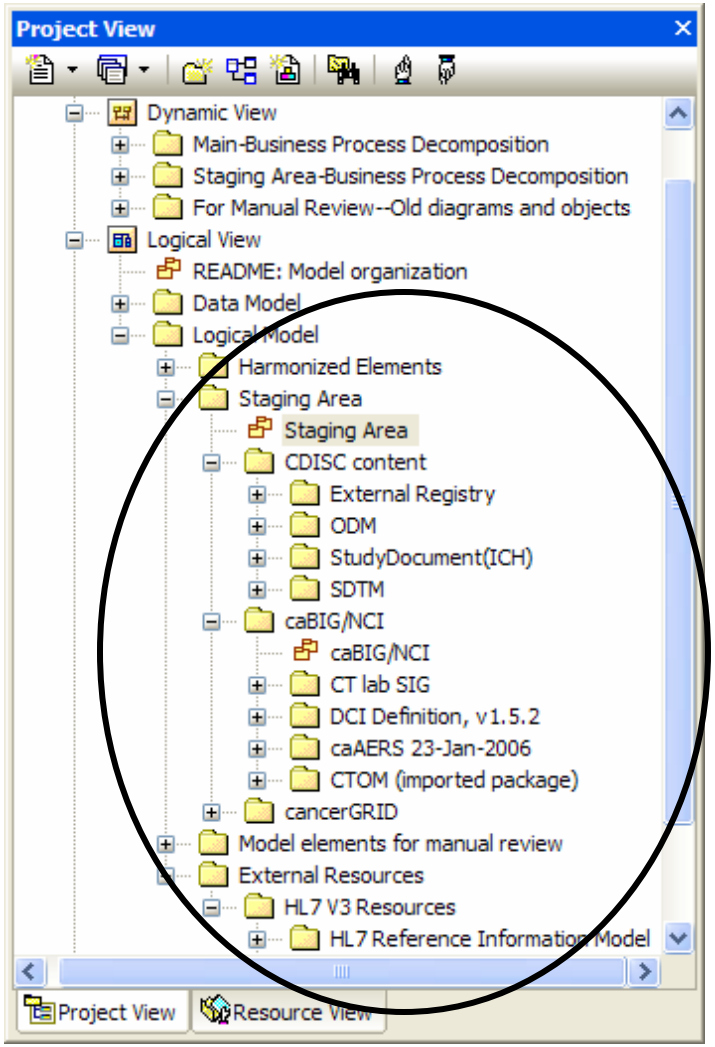




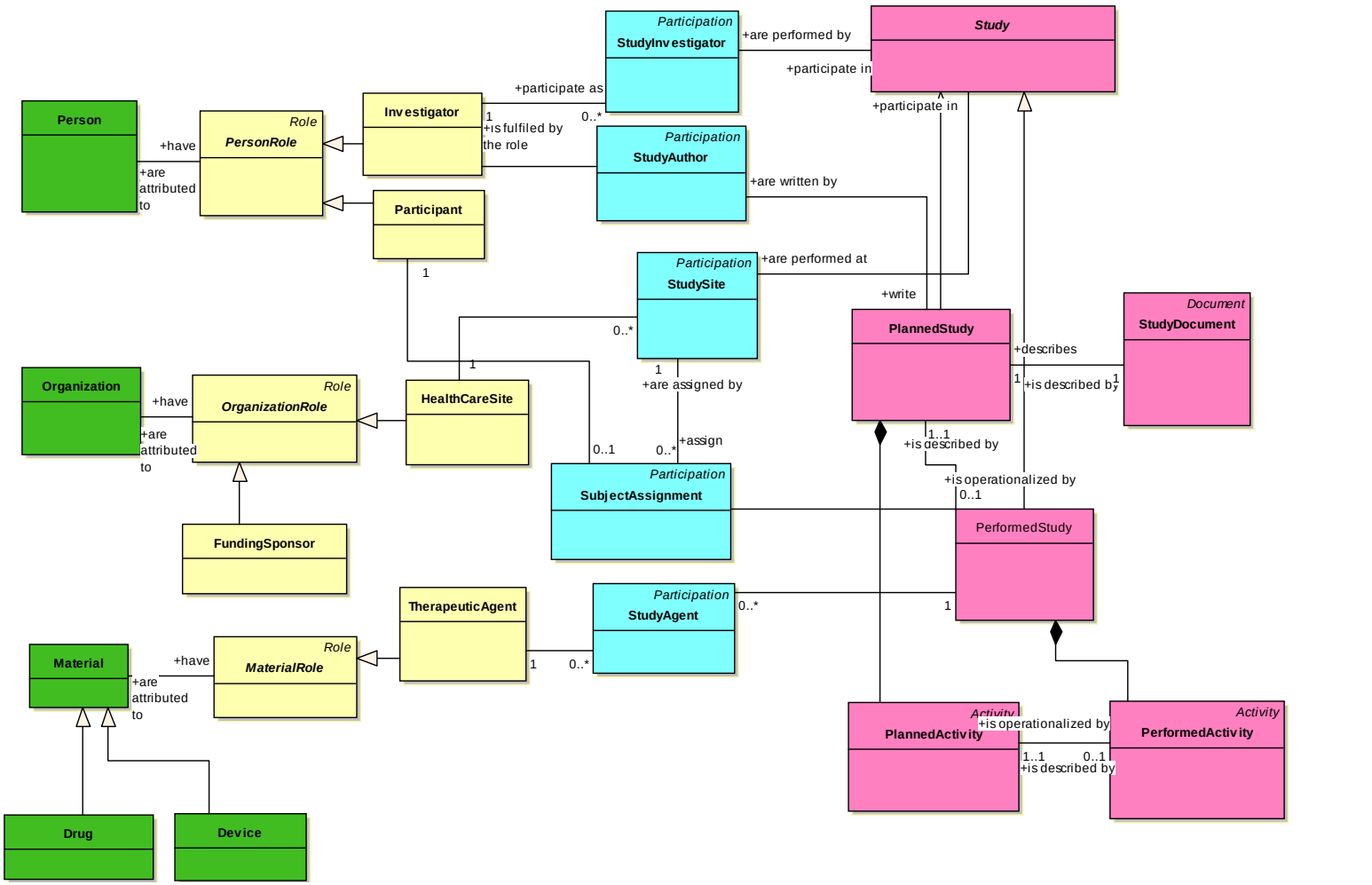
The activities are described in activity diagrams that can be drilled down to provide additional detail. These are linked to the static (logical) portions of the model

Model organization

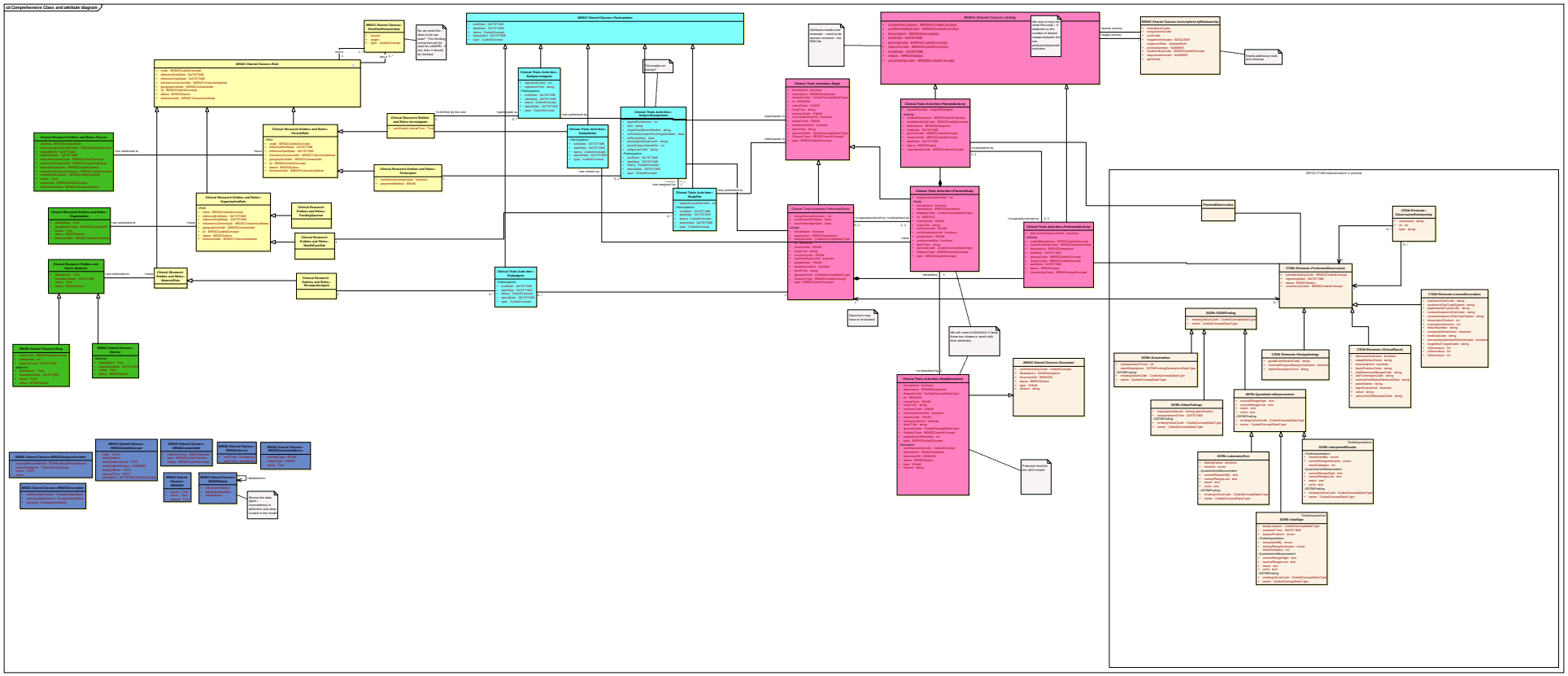
- Logical View
  - Contains three core packages
    - Harmonized elements
    - Staging Area
    - Manual review area
  - Addition resources
    - HL7 V3 RIM
  - Contains the semantics for the static objects (data) that is used in clinical trials research
  - Currently have 9 subdomain models in the process of harmonization



Current Classes in Core Elements

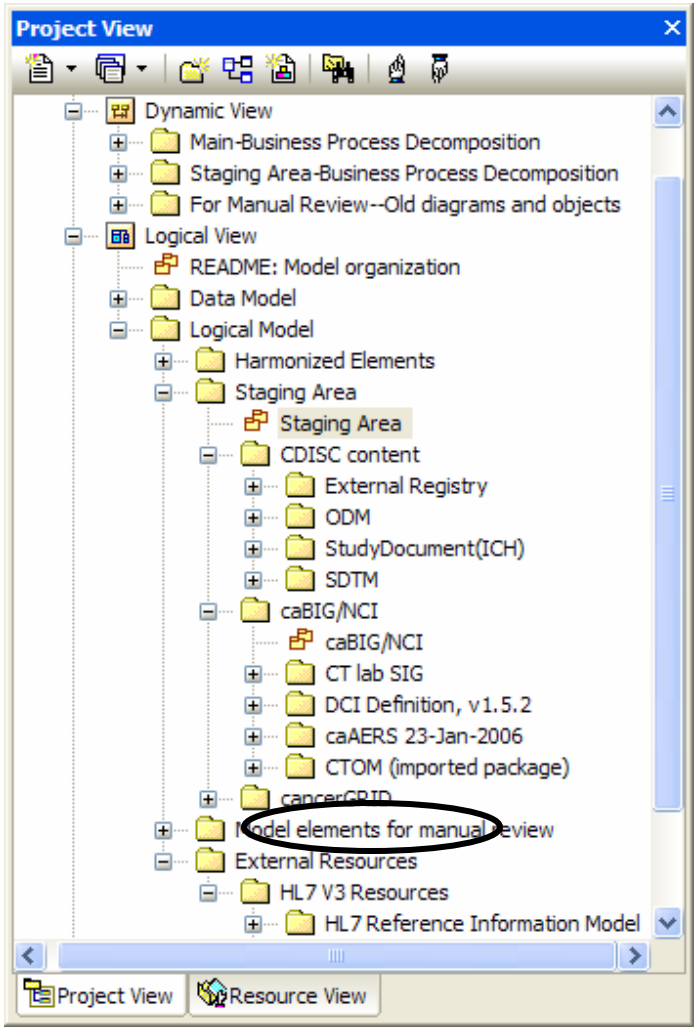


Harmonized BRIDG elements



BRIDG Sub-Projects

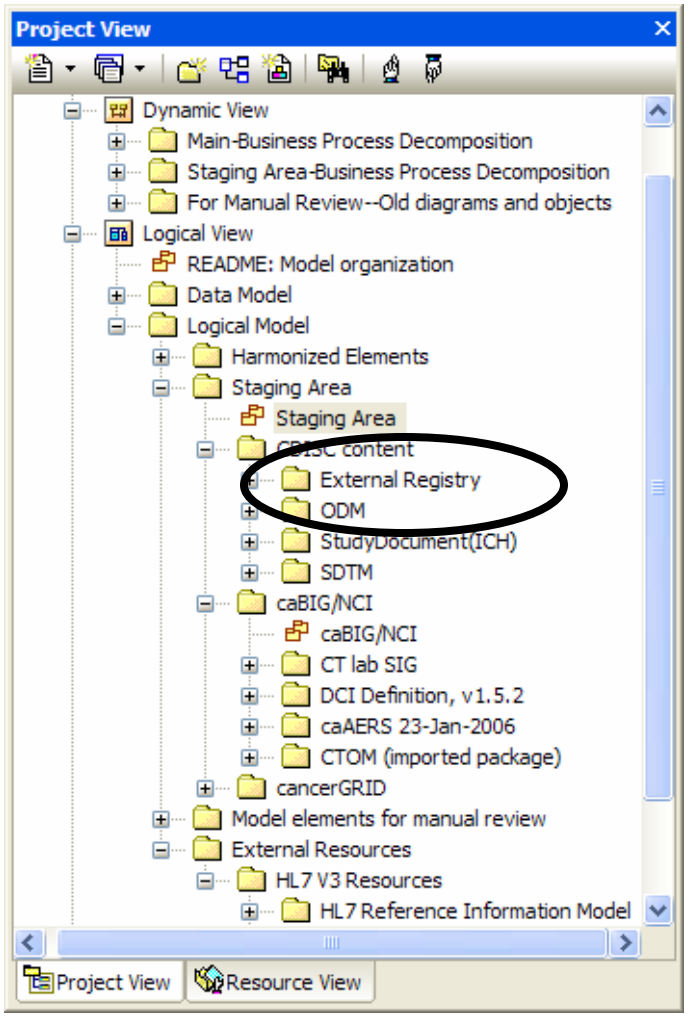
- Trial Design Model
  - Based on CDISC and FDA/Janus standard
  - Developing common concepts and understanding for arms, treatment groups, visits, cycles, courses, etc.
  - At present, input from Pharmaceutical companies thru CDISC and FDA
  - Current Status –
    - Recently worked with CDISC SDTM team to model SDTM requirements
    - Plans to harmonize with BRIDG



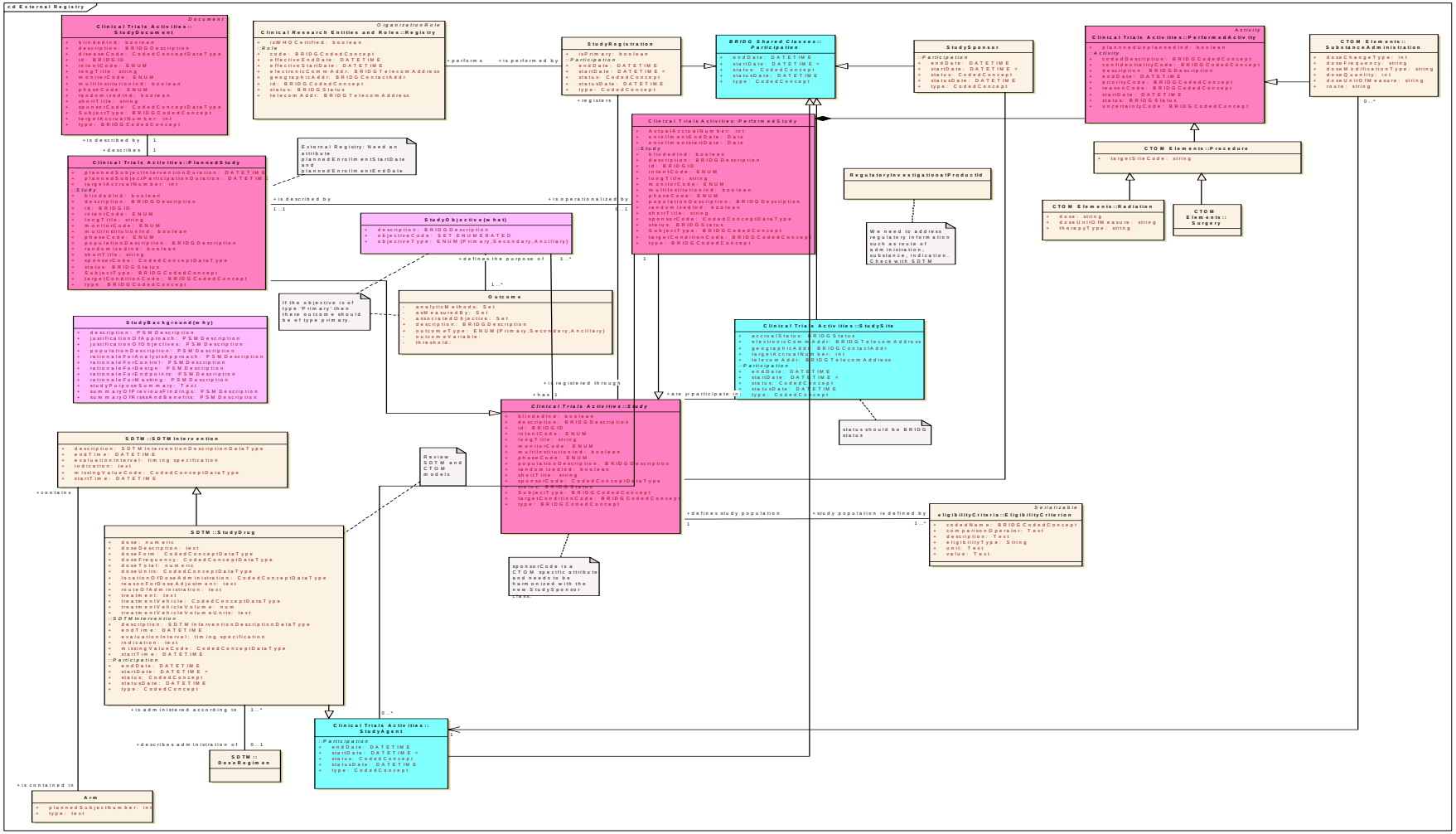
BRIDG Sub-Projects (cont'd)

- Clinical Trial Registry
- Objective: To define requirements for registering a clinical trial in a clinical trial repository
- Working with NCI, CDISC, PDQ, clinicaltrials.gov and European EUDRACT
- Have recently established collaboration with the WHO activity of clinical trials registry
  - Becky Kush (CDISC president) on the advisory board
  - Working with cancerGRID to incorporate and make explicit the CONSORT model

- Current Status –
  - Group has defined a list of 70 elements
  - Modeled in BRIDG April 2006
  - Planning on developing a HL7 v3 message
  - POC – Lakshmi Grama, NCI

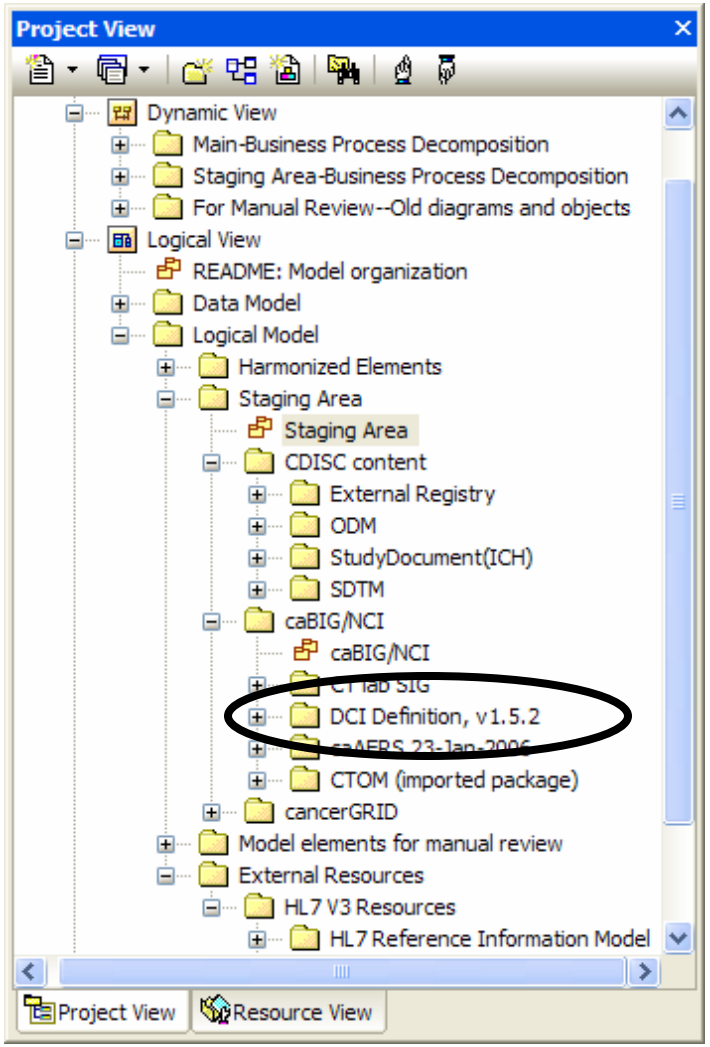


CT Registration message

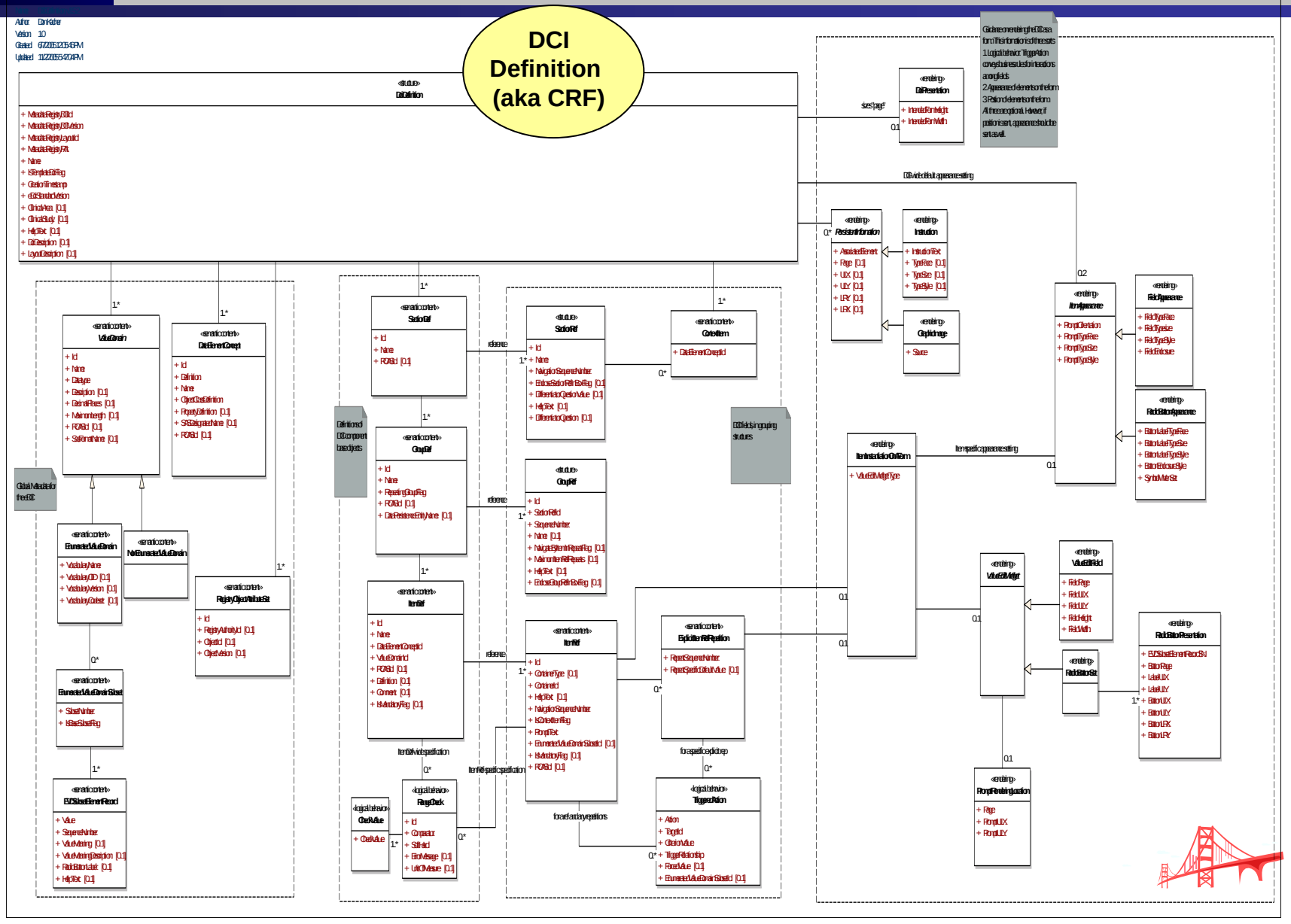


BRIDG Sub-Projects (cont'd)

- eDCI message (electronic Data Capture Instrument)
  - A DCI is a set of related questions for which values are to be collected during a clinical trial visit.
  - This model will be used as an HL7 message definition (or a set of definitions) that can be used to transmit a DCI Definition between databases managed by clinical data management systems (CDMS).
  - Participation from NCI, CDISC, Oracle
  - UML model on bridgproject site -- https://www.bridgproject.org/edci/
  - Current Status –
    - Requirements have been defined for 1st iteration
    - UML class diagram is completed
    - Working on building the message specification (RMIM)
    - POC – Don Kacher, Oracle

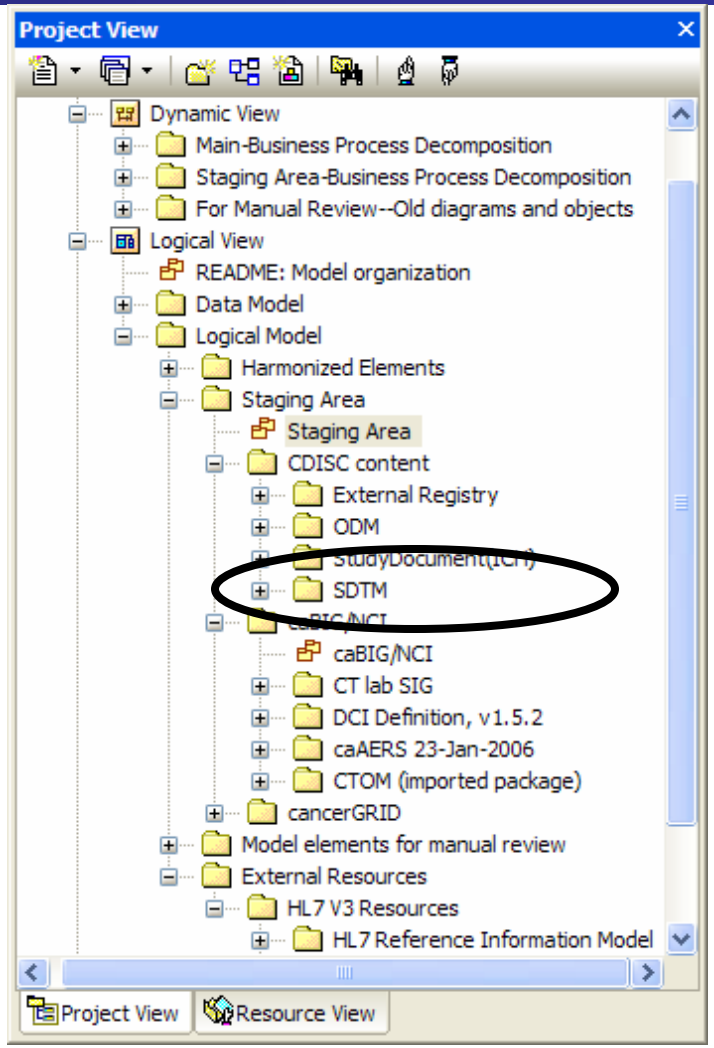


DCI Definition (aka CRF)

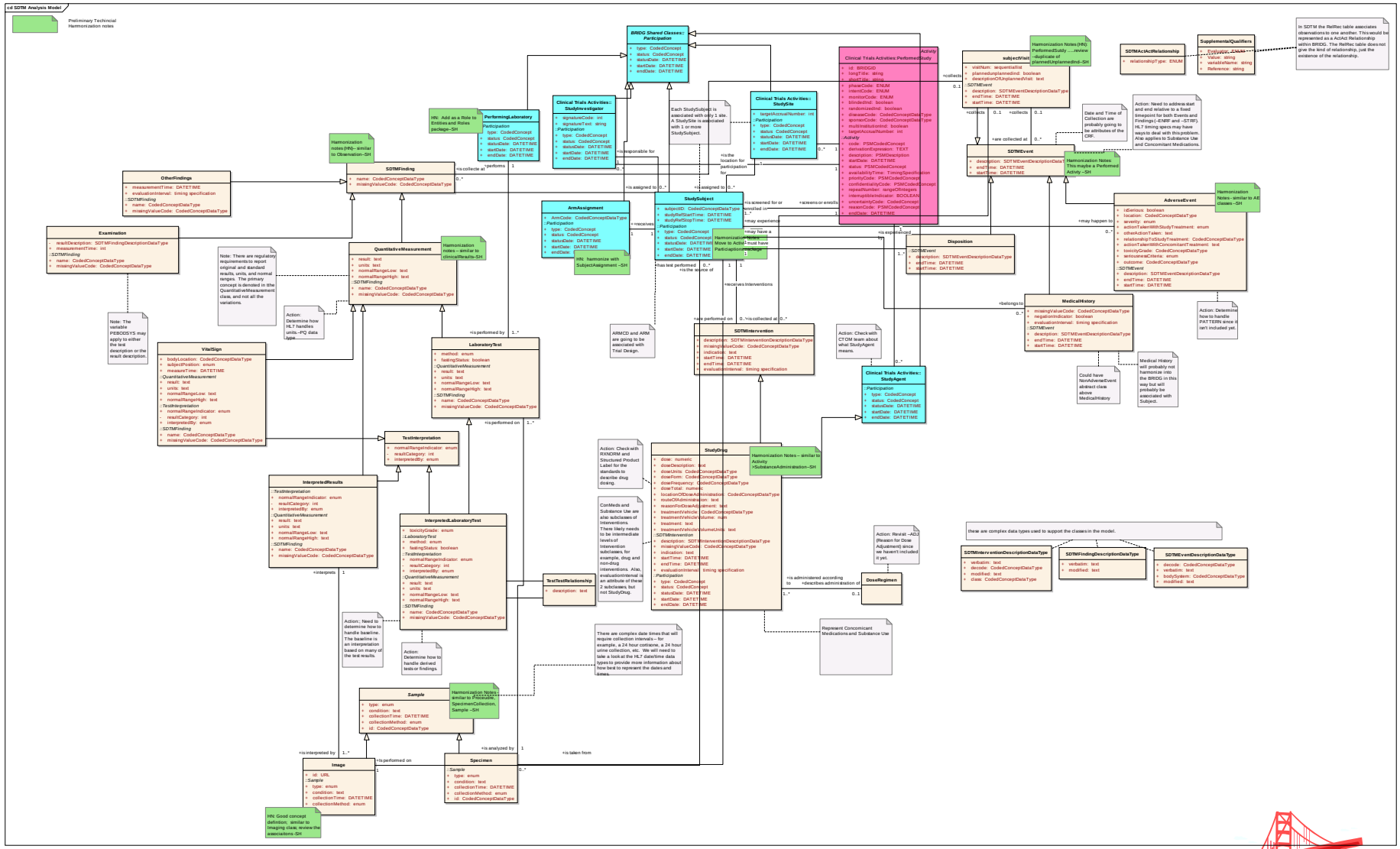


SDTM

- SDTM model
  - Being harmonized with adverse event reporting, CTOM (NCI clinical trial object model) and HL7



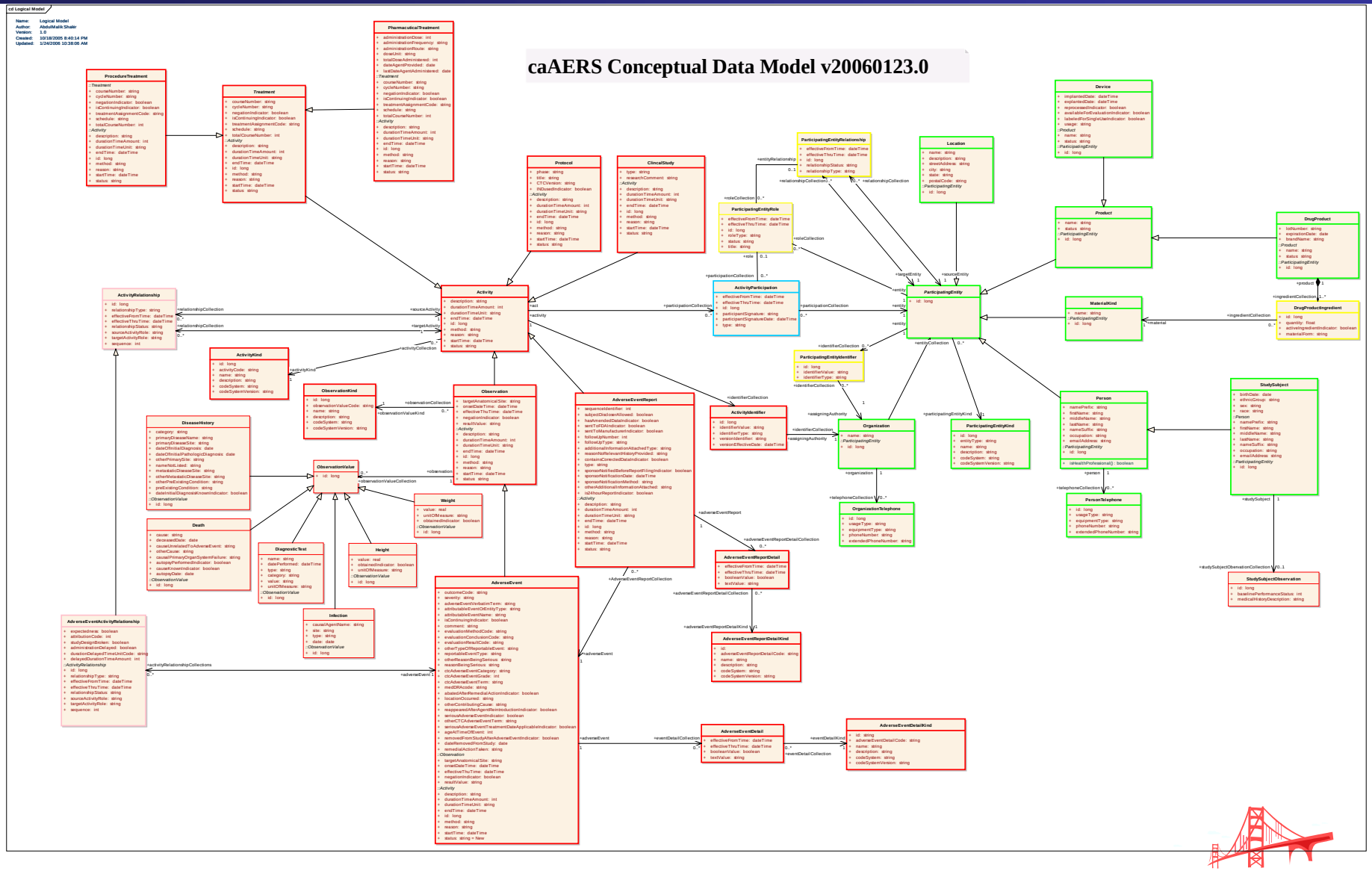
SDTM Class Diagram



Subprojects

- caAERS
  - Project lead: Joyce Niland
  - Developing an HL7 message and application(s) to support adverse event reporting
  - Other AE models –
    - CDC – incidence reporting
    - HL7 – patient safety and public health reporting
    - caBIG (caAERS)
    - FDA and SDTM (CDISC)
  - Harmonization meeting in May with all stakeholders to identify commonalities and differences between these models, and harmonize them into BRIDG

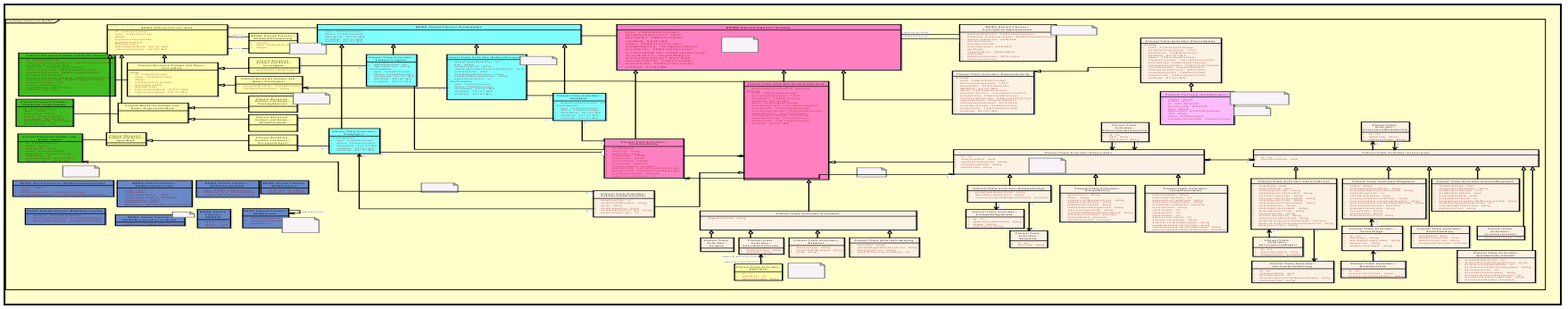
caAERS


cd Logical Model
Name: Logical Model
Author: AbdulMalik Shaikh
Version: 1.0
Created: 10/18/2005 8:40:14 PM
Updated: 1/24/2006 10:38:06 AM
caAERS Conceptual Data Model v20060123.0
ProcedureTreatment
Treatment
PharmaceuticalTreatment
Protocol
ClinicalStudy
ParticipatingEntityRelationship
Location
Device
Product
DrugProduct
DrugProductIngredient
MaterialKind
ParticipatingEntityRole
ParticipatingEntityIdentifier
ActivityParticipation
ParticipatingEntity
Activity
ActivityRelationship
ActivityKind
ObservationKind
Observation
ObservationValue
ActivityIdentifier
AdverseEventReport
Organization
ParticipatingEntityKind
Person
StudySubject
DiseaseHistory
Death
DiagnosticTest
Weight
Height
Infection
AdverseEvent
AdverseEventReportDetail
AdverseEventReportDetailKind
AdverseEventDetail
AdverseEventDetailKind
OrganizationTelephone
PersonTelephone
StudySubjectObservation
AdverseEventActivityRelationship

What does it mean to “adopt BRIDG” or “harmonize with BRIDG?”

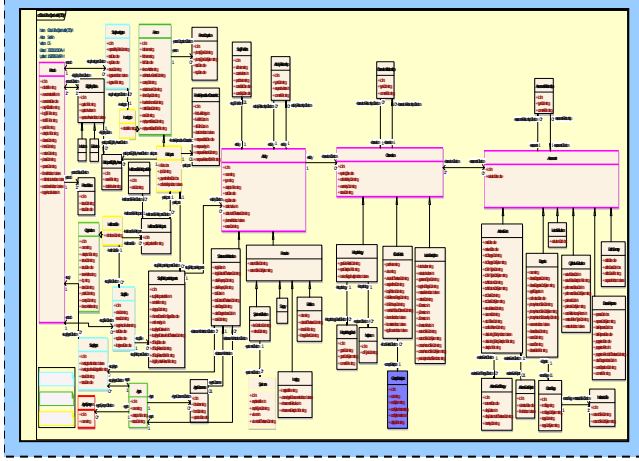
- Adopting and harmonizing with BRIDG is a two-way street
  - The model is not complete, and harmonization and adoption requires participation and contribution to BRIDG from others
  - The model is new and is changing, so harmonization and adoption requires flexibility and change
- Early adopters will have a more significant impact on the direction and development of BRIDG
- Adopting and harmonization with BRIDG is less about a commitment to a specific model, but the realization that
  - A common standard is a shared good that all can benefit from
  - It will require contribution and collaboration as we collectively determine the best approaches
  - It will require compromise and collective action

BRIDG - Implementation Independent Domain Analysis Model

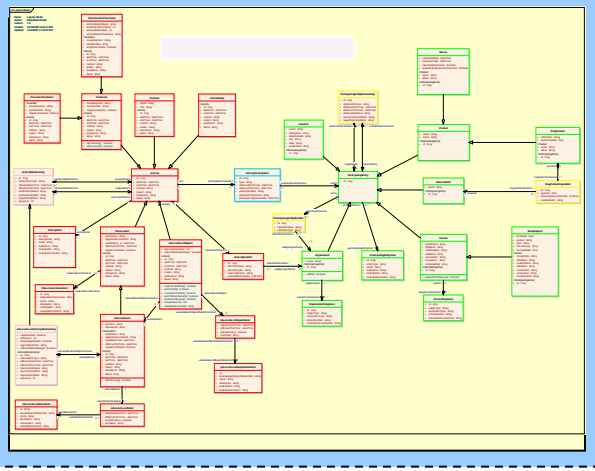


Implementation Specific Models

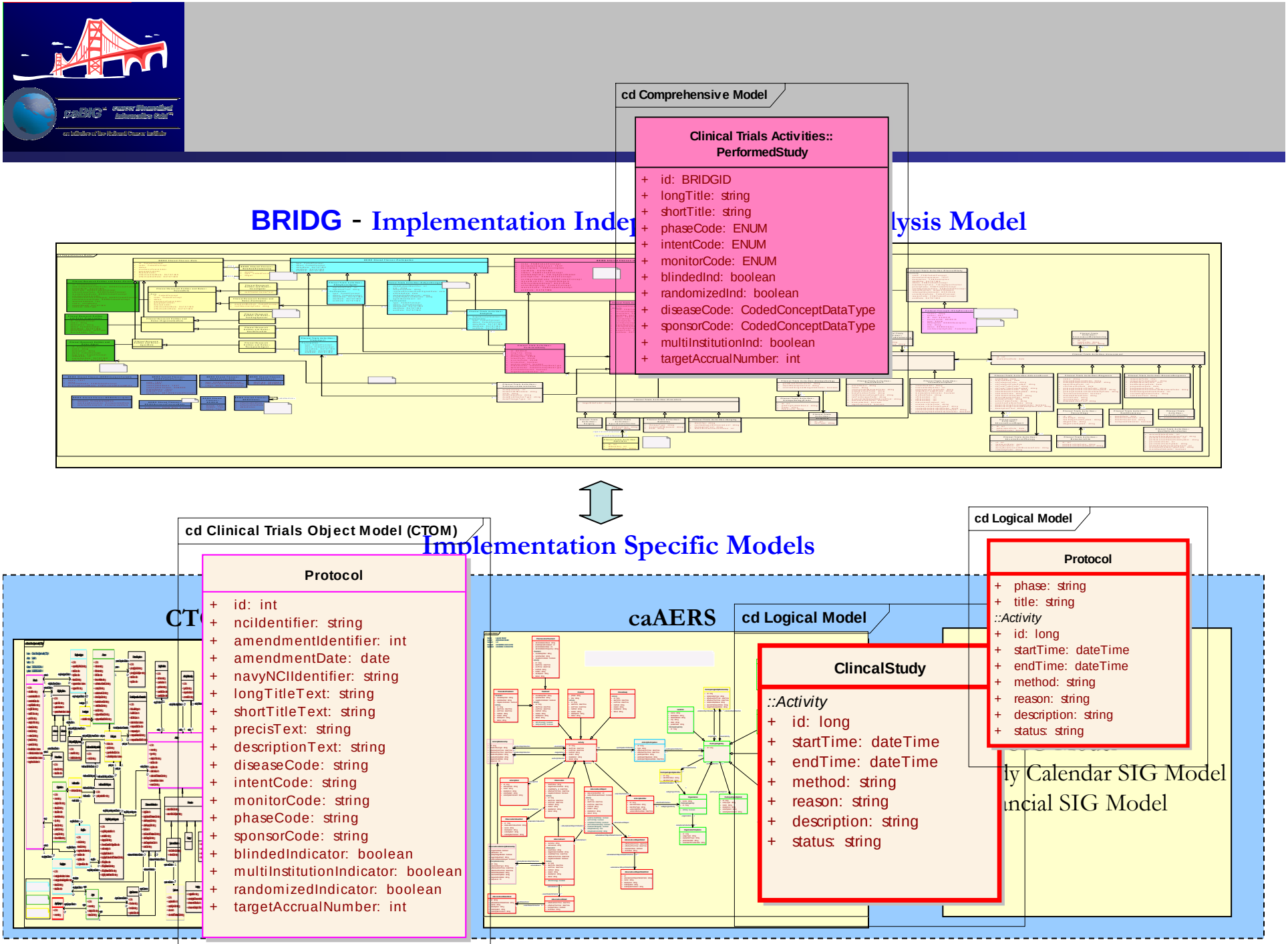
CTOM



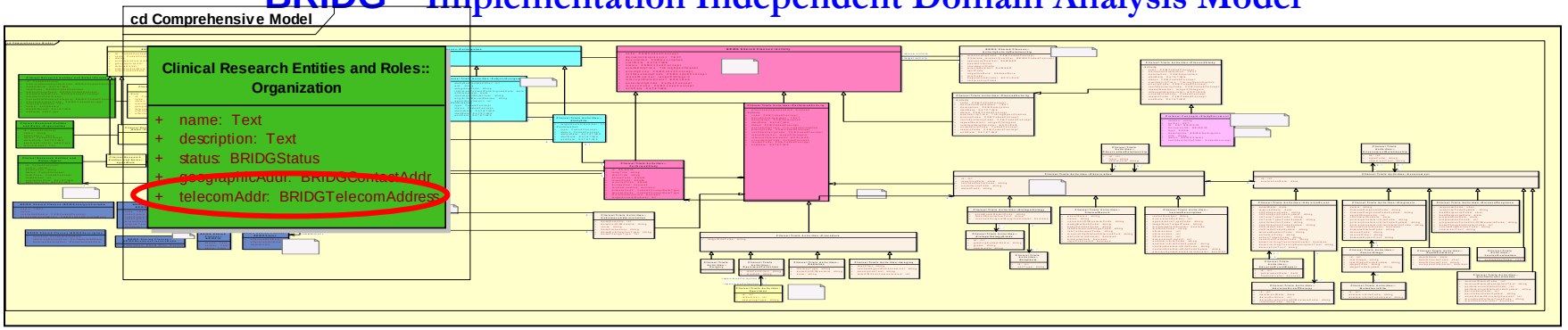
caAERS



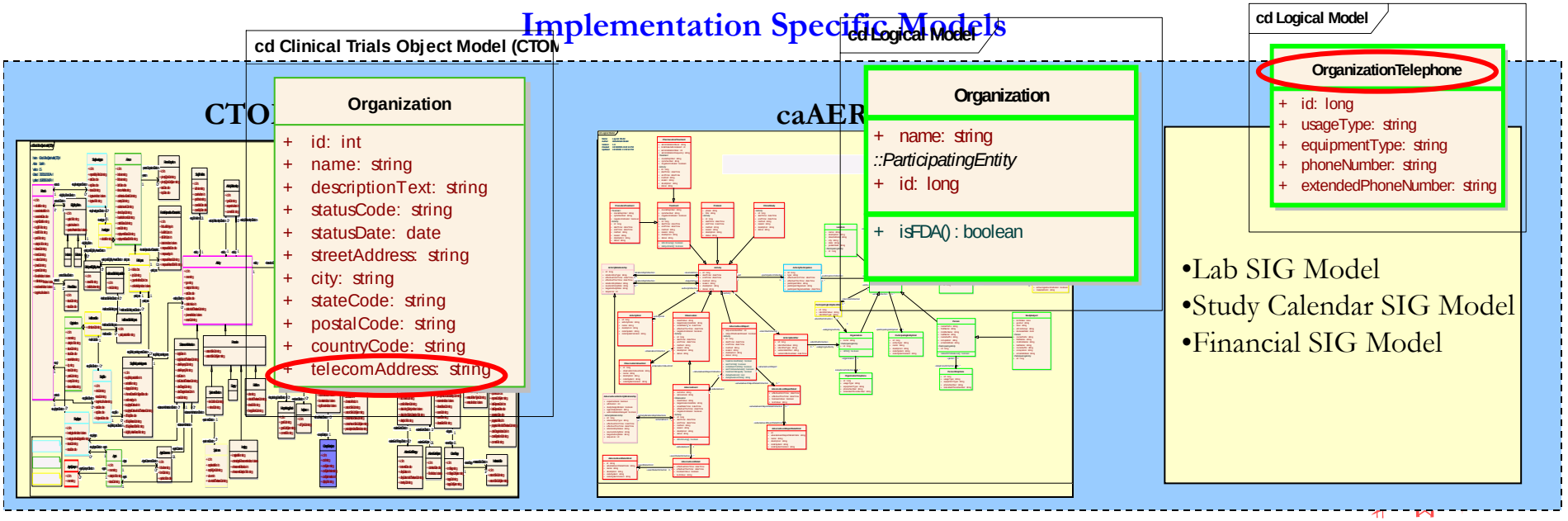
- Lab SIG Model
- Study Calendar SIG Model
- Financial SIG Model


cd Comprehensive Model
Clinical Trials Activities::
PerformedStudy
+ id: BRIDGID
+ longTitle: string
+ shortTitle: string
+ phaseCode: ENUM
+ intentCode: ENUM
+ monitorCode: ENUM
+ blindedInd: boolean
+ randomizedInd: boolean
+ diseaseCode: CodedConceptDataType
+ sponsorCode: CodedConceptDataType
+ multiInstitutionInd: boolean
+ targetAccrualNumber: int
BRIDG - Implementation Indep lysis Model
Implementation Specific Models
cd Clinical Trials Object Model (CTOM)
Protocol
+ id: int
+ nciIdentifier: string
+ amendmentIdentifier: int
+ amendmentDate: date
+ navyNCIIdentifier: string
+ longTitleText: string
+ shortTitleText: string
+ precisText: string
+ descriptionText: string
+ diseaseCode: string
+ intentCode: string
+ monitorCode: string
+ phaseCode: string
+ sponsorCode: string
+ blindedIndicator: boolean
+ multiInstitutionIndicator: boolean
+ randomizedIndicator: boolean
+ targetAccrualNumber: int
CT
caAERS
cd Logical Model
ClincalStudy
::Activity
+ id: long
+ startTime: dateTime
+ endTime: dateTime
+ method: string
+ reason: string
+ description: string
+ status: string
cd Logical Model
Protocol
+ phase: string
+ title: string
::Activity
+ id: long
+ startTime: dateTime
+ endTime: dateTime
+ method: string
+ reason: string
+ description: string
+ status: string
ly Calendar SIG Model
ncial SIG Model

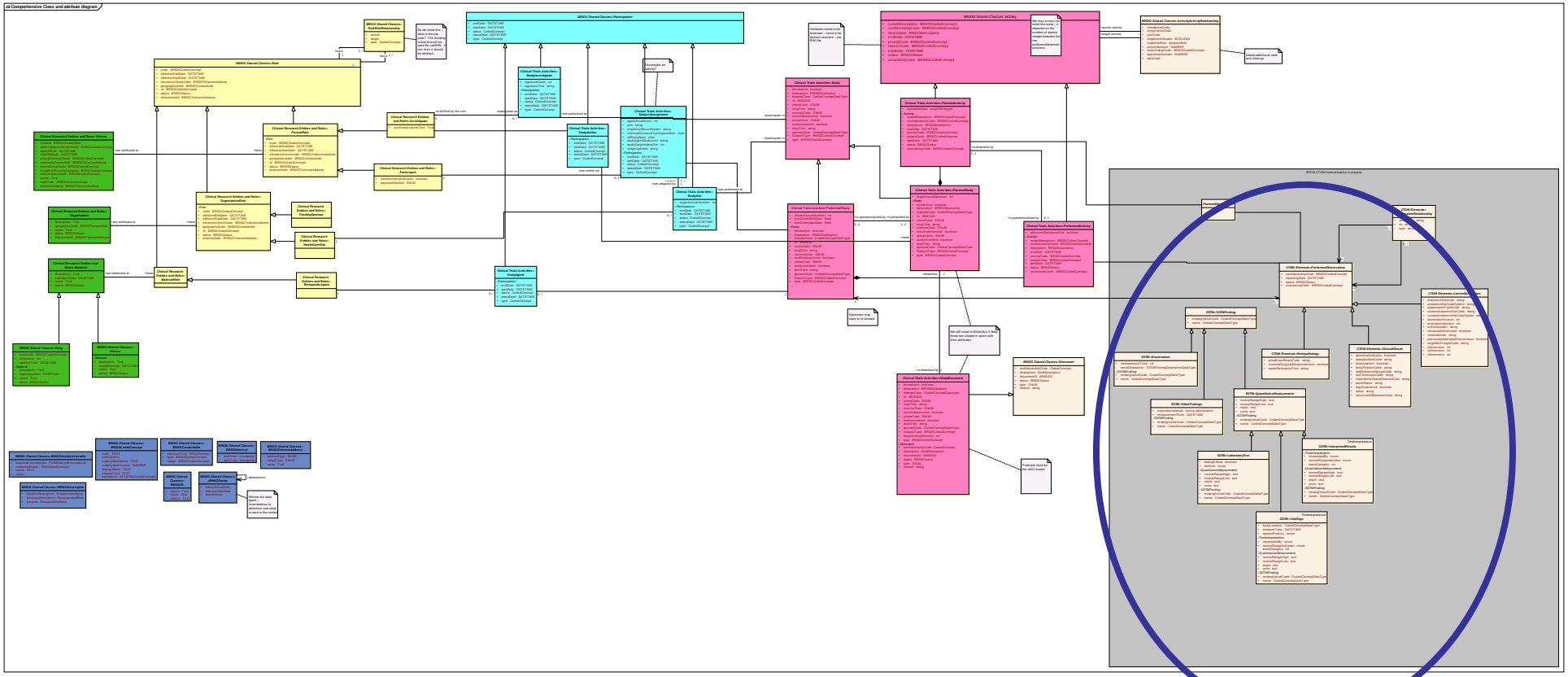
BRIDG - Implementation Independent Domain Analysis Model



Implementation Specific Models

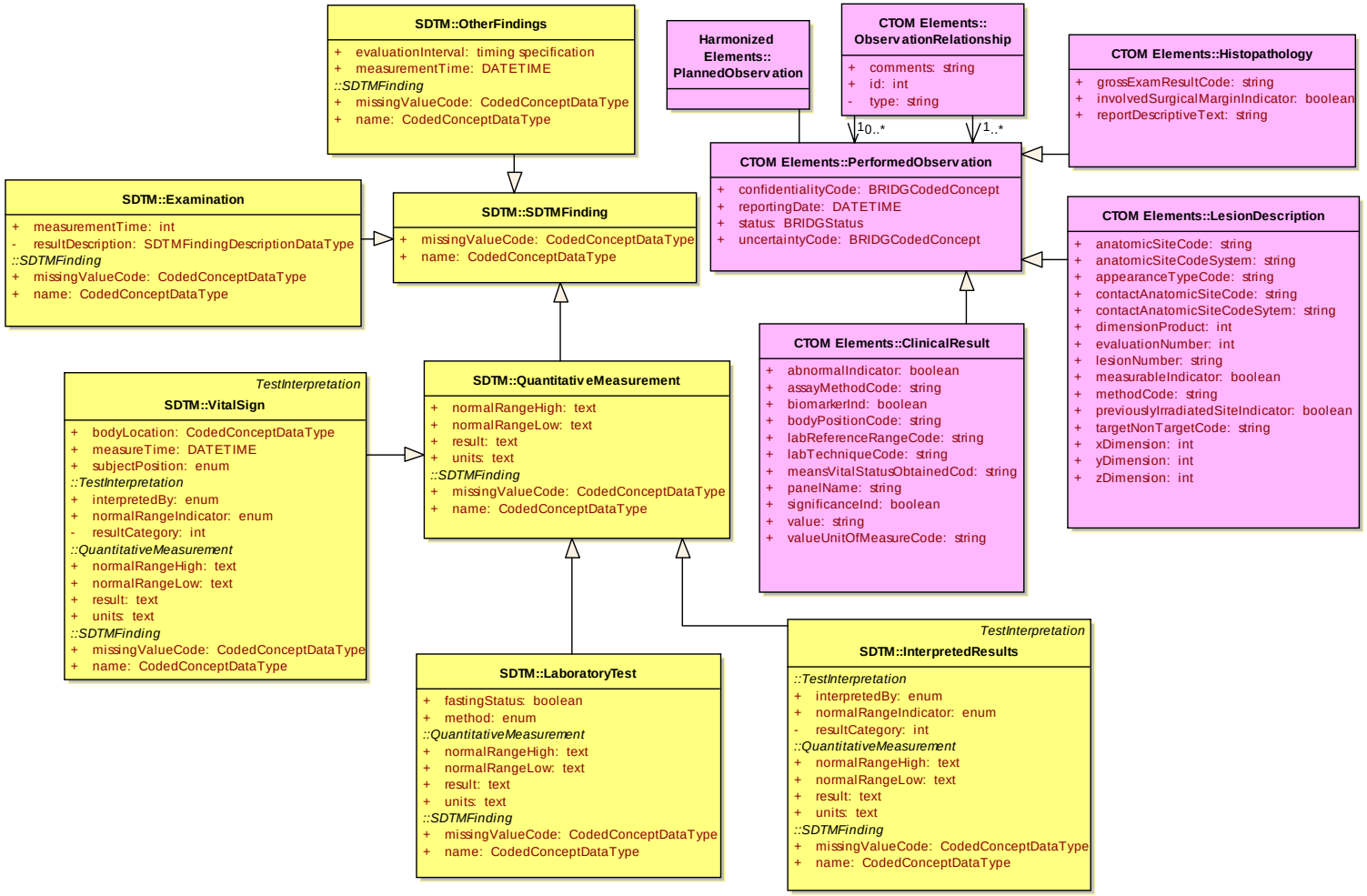


Harmonized BRIDG elements

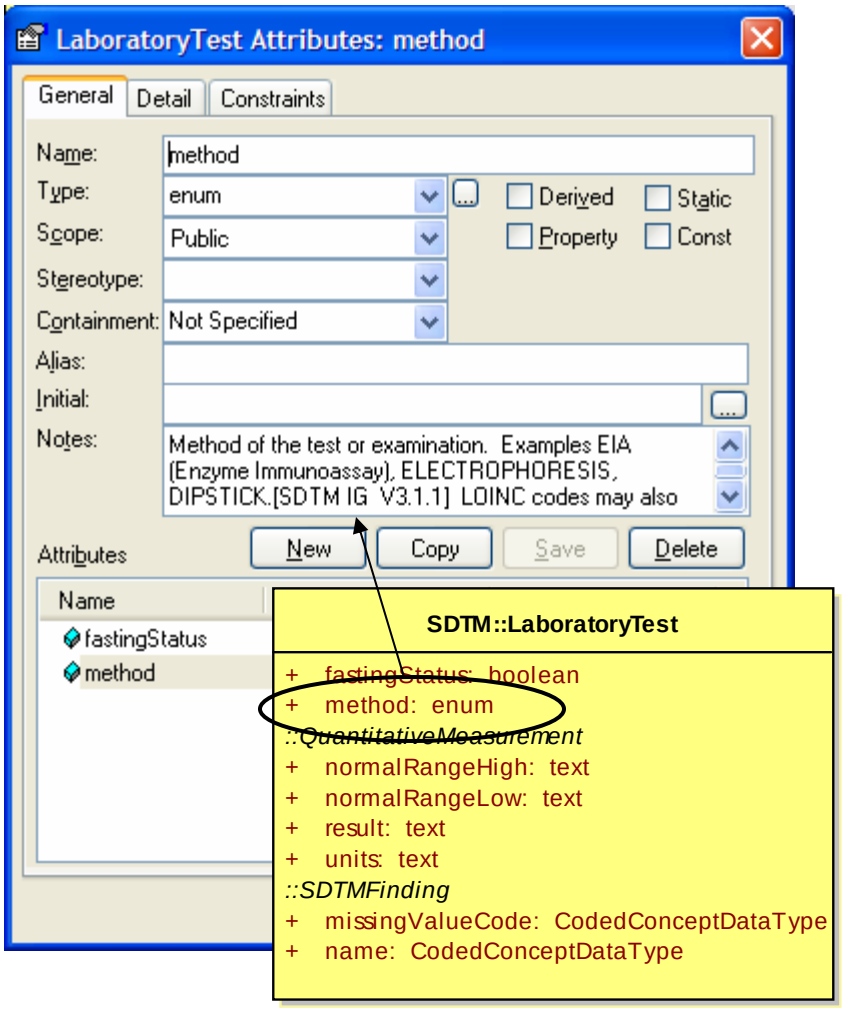


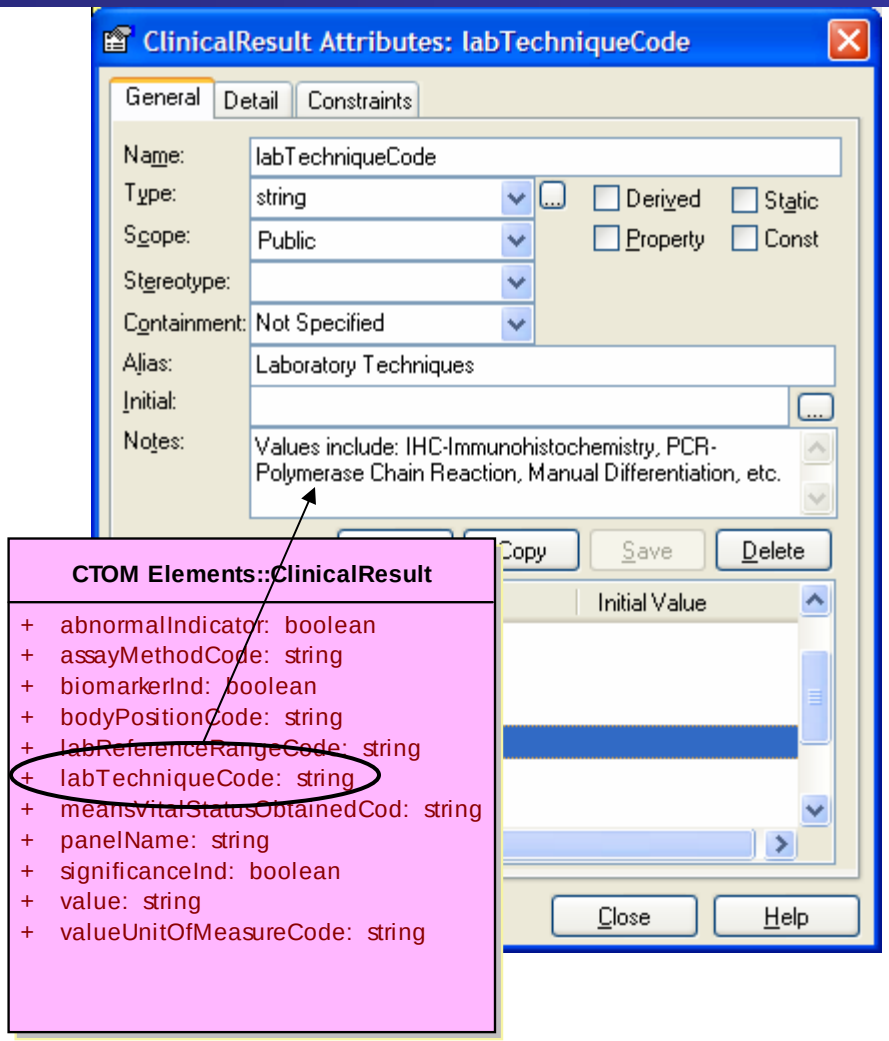
Observation Classes from CTOM and SDTM

CTOM and SDTM harmonization (work in progress)

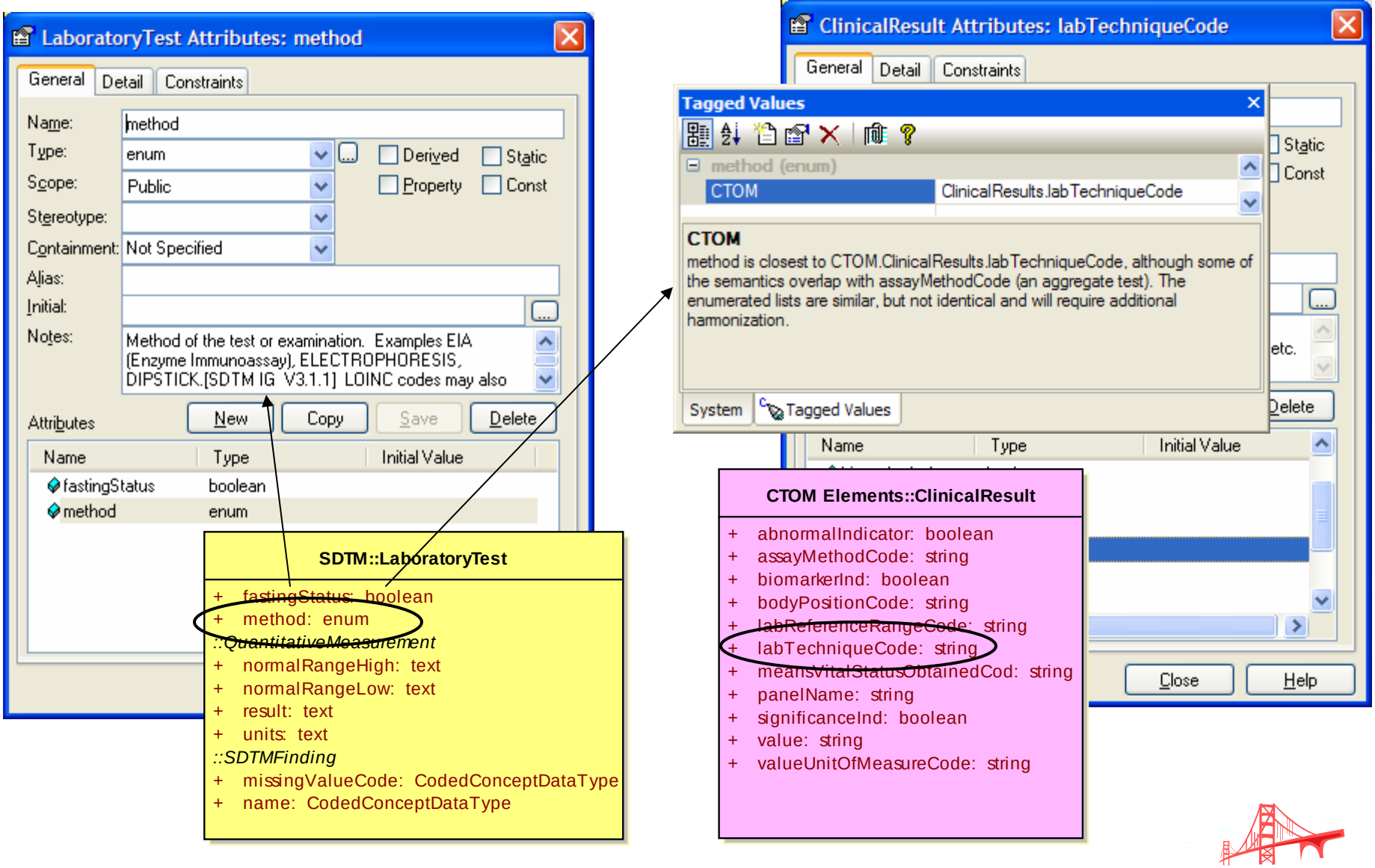


Harmonizing attributes

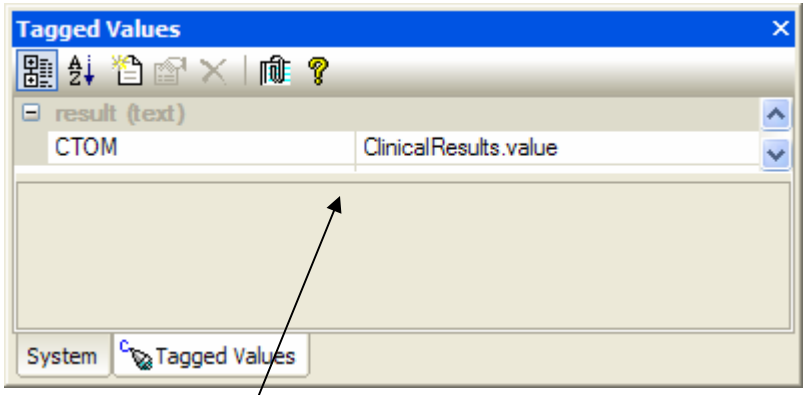


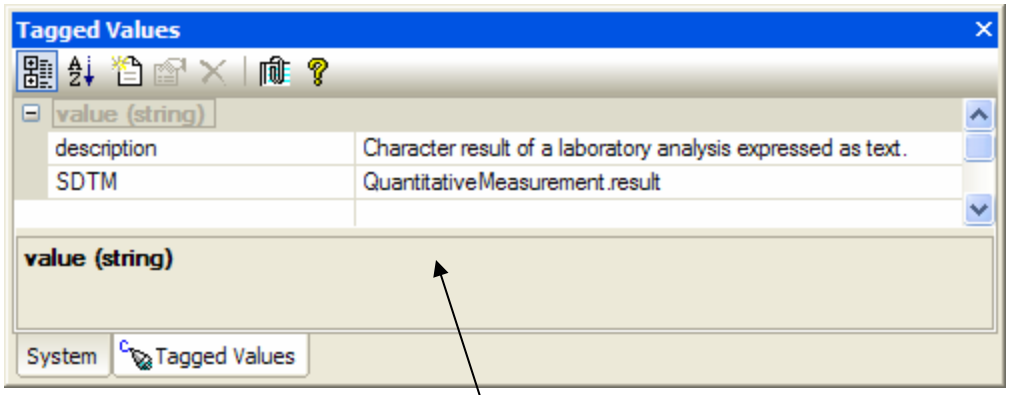


Adding tags to provide semantic traceability (and notes)



Simple semantic can be tracked in tagged values





SDTM::QuantitativeMeasurement

+ normalRangeHigh: text
+ normalRangeLow: text
+ result: text
+ units: text

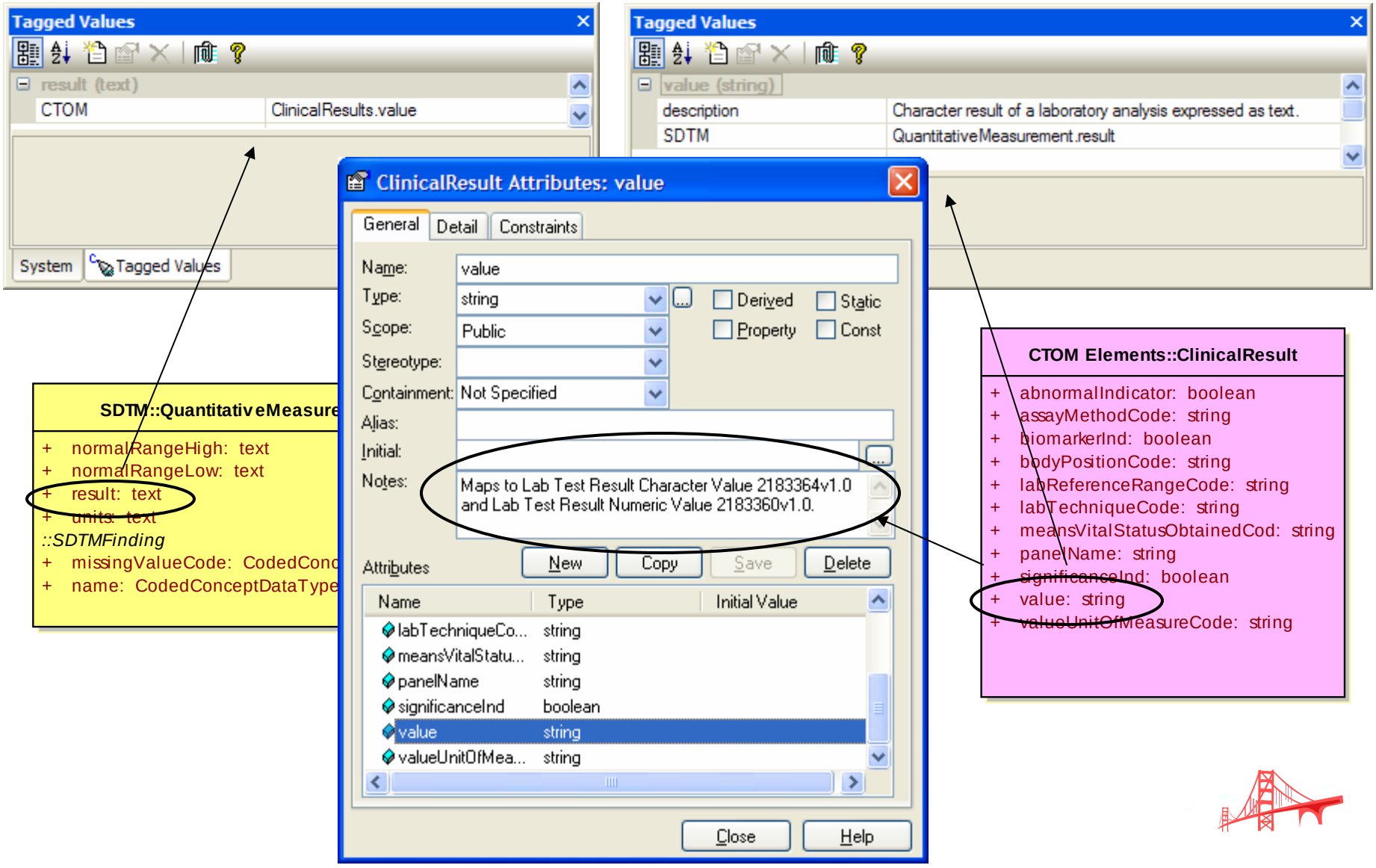
::SDTMFinding

+ missingValueCode: CodedConceptDataType
+ name: CodedConceptDataType

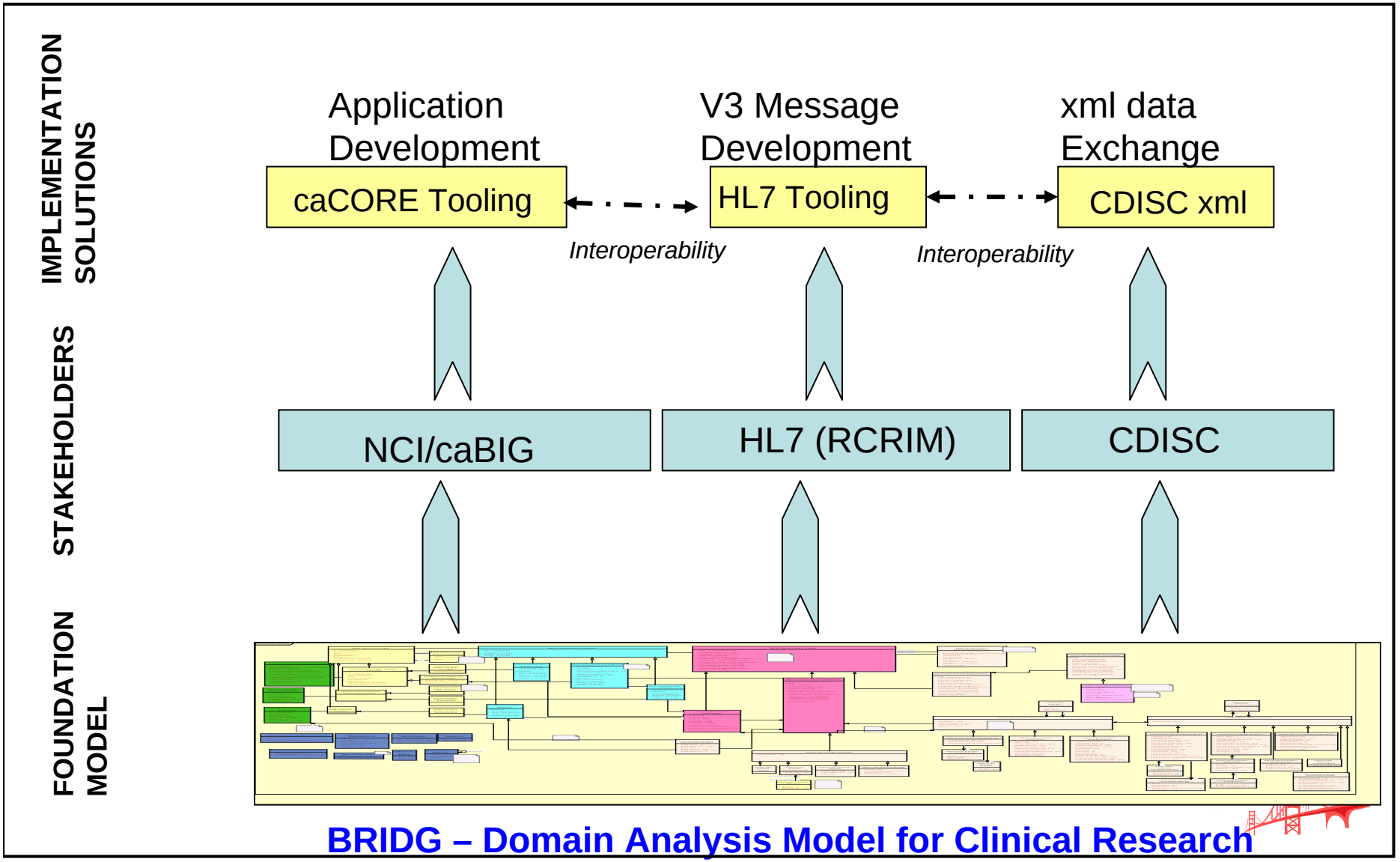
CTOM Elements::ClinicalResult

+ abnormalIndicator: boolean
+ assayMethodCode: string
+ biomarkerInd: boolean
+ bodyPositionCode: string
+ labReferenceRangeCode: string
+ labTechniqueCode: string
+ meansVitalStatusObtainedCod: string
+ panelName: string
+ significanceInd: boolean
+ value: string
+ valueUnitOfMeasureCode: string

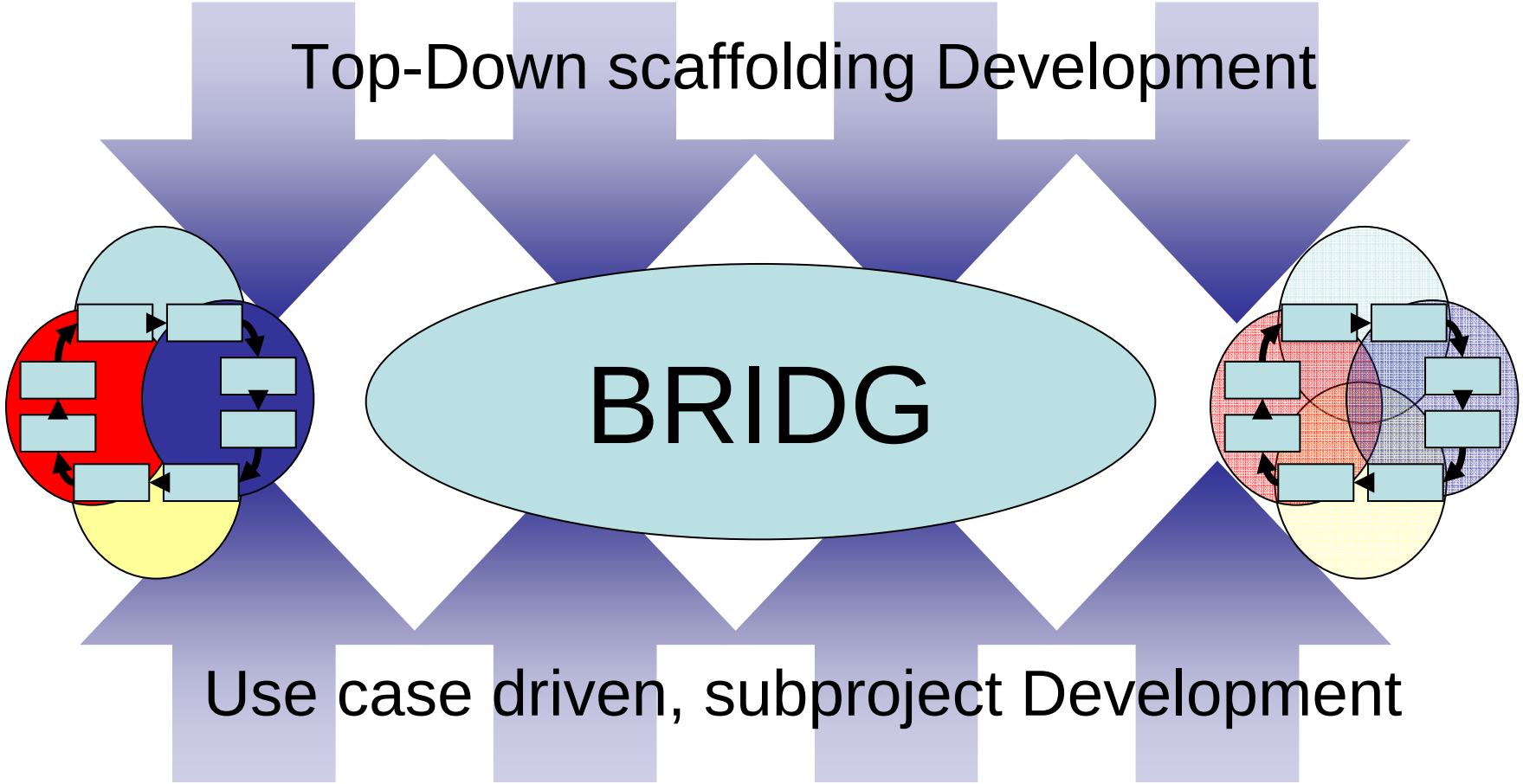
This linking can be extended down to the CDE level



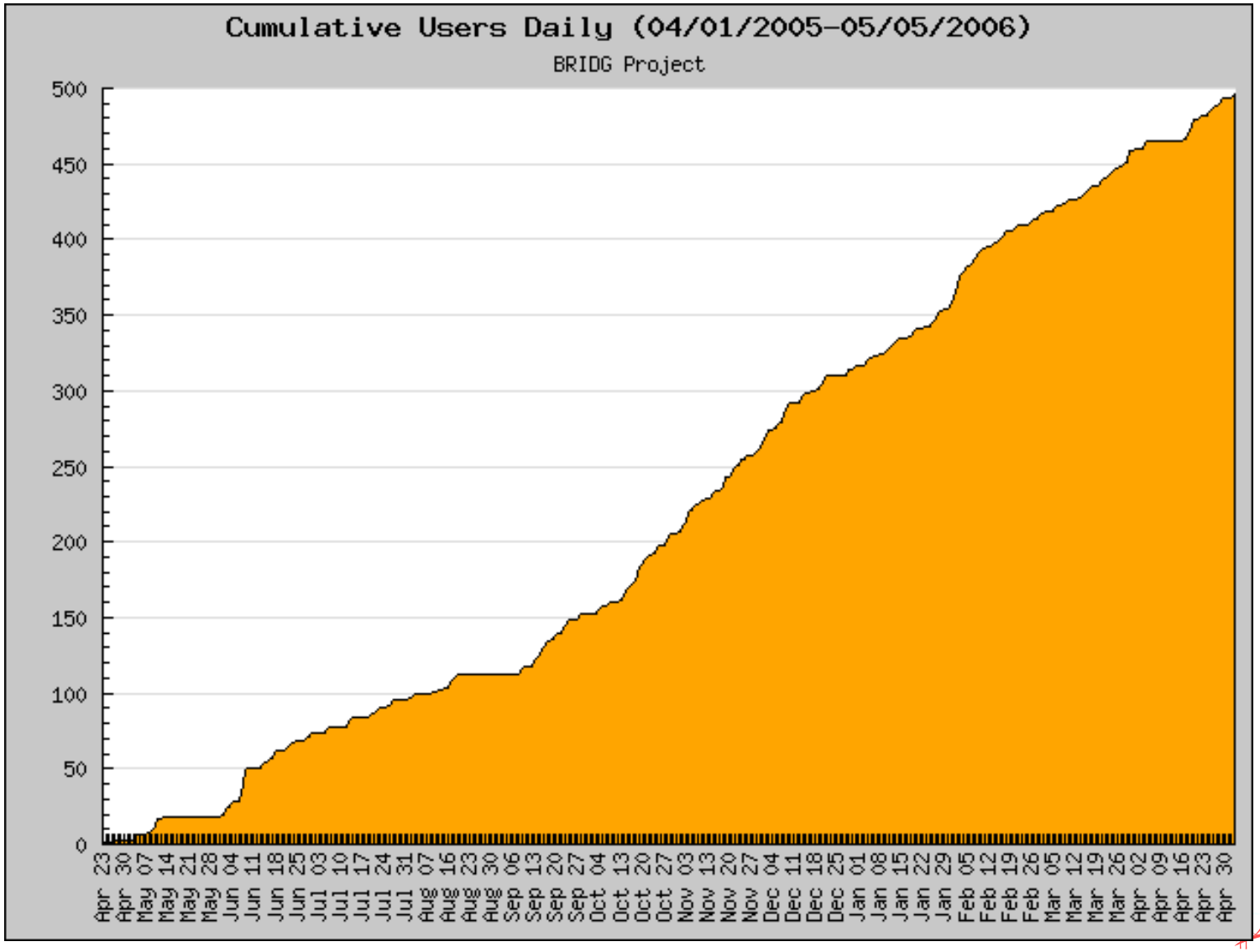
Achieving interoperability



BRIDG development



Cumulative Registered Users



What have we accomplished?

- BRIDG
  - Established excellent collaboration with CDISC, HL7 and other caBIG modelers
  - Constructed the initial pieces of a comprehensive model – still much to do
  - Have developed processes and organization of the model that will support more scaleable collaborative development
  - Demonstrated semantic reuse for subproject development

- We hope that this model will serve as a resource for application and message development within a unified framework, and will define the shared semantics of clinical trials research
  - caBIG for application development
  - HL7 for V3 RIM message development
  - “Semantic traceability” to link analysis model to design and implementation-specific models

Final thoughts: our approach to modeling

- Scope – keep it clear and focused (ie, solve a problem that exists) and standardize to the extend needed
  - Refine through experience, and not endless discussions. This keeps the modeling effort clear and focused
  - BRIDG is not complete – but the scaffolding is there to help organize the analysis and model development in subprojects
- Keep it generic, faithful, free of implementation specific formalisms, and supporting the requirements
- If the tools and models don't work with reality – it is probably the tools and the models that need to change
- If it's broke, fix it
  - The model is in evolution with known problems – the problems should be an opportunities for improvement and a call to arms, not barriers to use
- Model in the open
- Collaborate until it hurts

With thanks to Dipak Kalra for discussion

Acknowledgements –leadership

- Leadership and collaboration
  - CDISC board members
  - Becky Kush, CDISC
  - Sue Dubman, Christo Andonyadis, Peter Covitz NCI
  - Charlie Mead, HL7, BAH
  - Barbara Tardiff, Linda Quade, Randy Levine (FDA), RCRIM technical committee of HL7

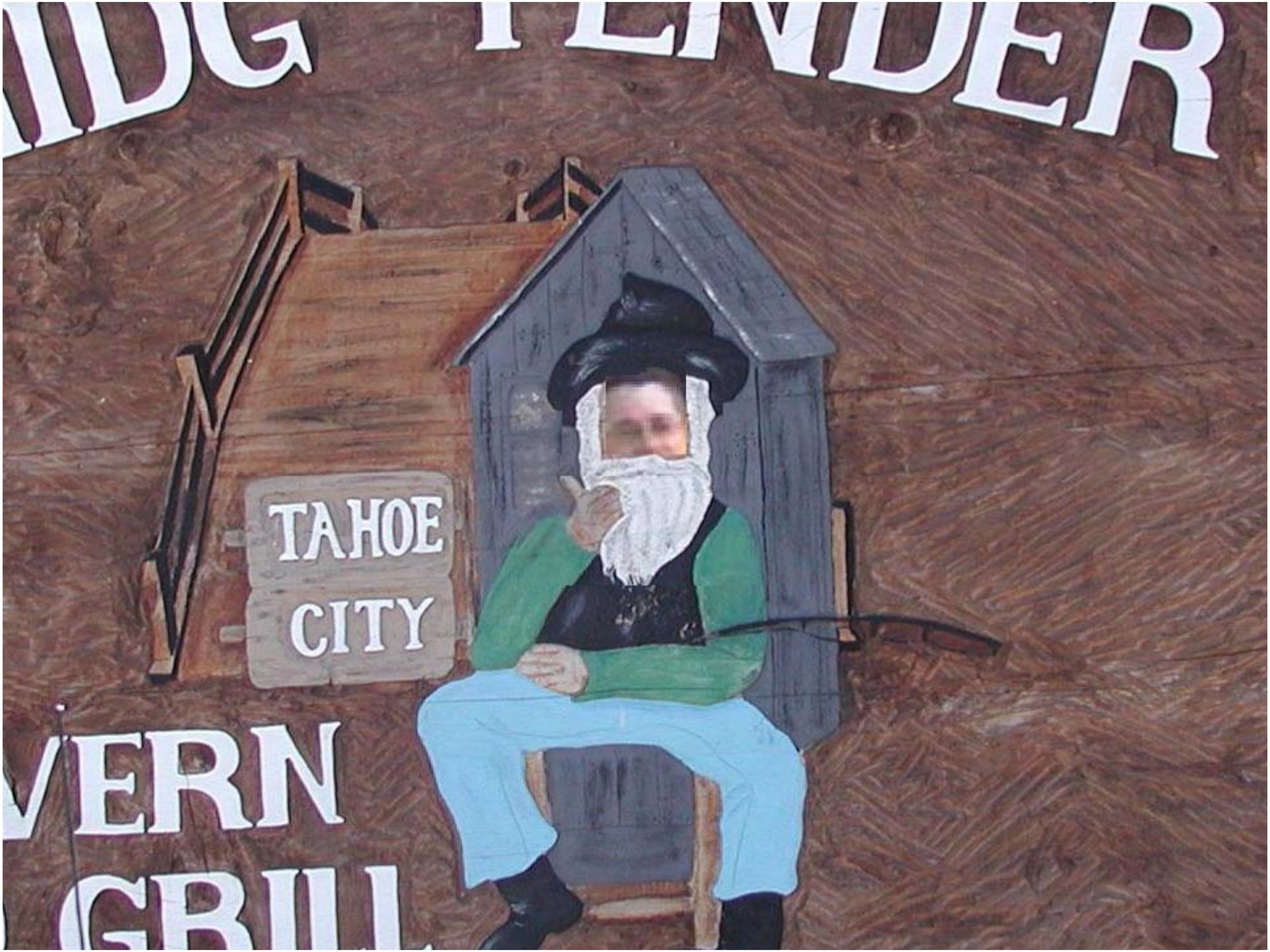
Acknowledgements –modelers

- Christo Andonyadis
- Greg Anglin
- Lisa Chatterjee
- Julie Evans
- Douglas B Fridsma
- Smita Hastak
- Joyce Niland
- John Speakman
- Cara Willoughby
- Diane Wold
- Peter Ambramovich
- Ed Helton
- Pierre-Yves Lastic
- Tony Friebel
- Don Kacher
- Barbara Tardiff
- Chuck Jaffe
- Frank Newby
- Sally Cassels
- Nitin Gupta
- Landen Bain
- Laura Altshuler
- Steve Ruberg
- Elaine Job
- Becky Kush
- Sylvia Collins
- Udo Siegman
- David Hardison
- David Iberson-Hurst

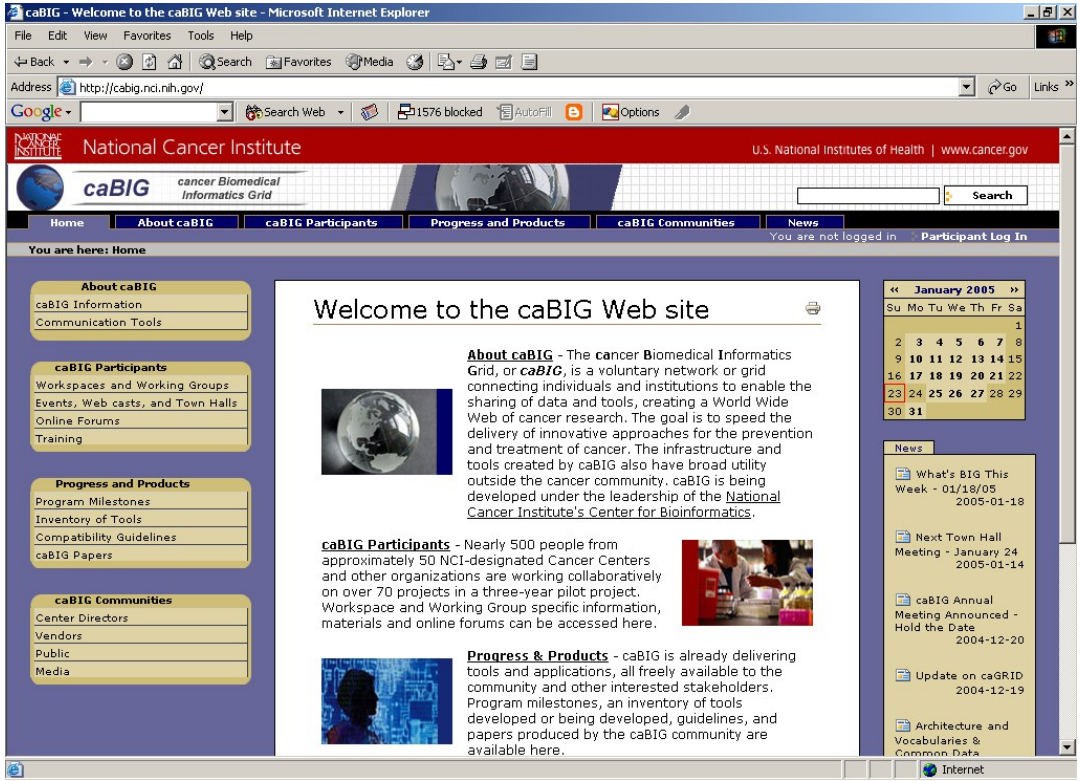
Acknowledgements—Organizations

- Pharmaceutical companies
  - AstraZeneca
  - Boehringer-Ingelheim
  - Eli Lilly
  - GlaxoSmithKline
  - Merck
  - Novartis
  - Pfizer
  - Sanofi-Aventis
- Technology companies
  - ScenPro
  - IBM
  - SAS
  - Fast track
  - SAIC
  - BAH
  - Oracle


BRIDG TENDER
TAHOE CITY
TAVERN AND GRILL
ADDITIONAL BRIDG TENDER PARKING ACROSS THE STREET


TAHOE
CITY

Further Information



- ncicb.nci.nih.gov
- caBIG.nci.nih.gov
- www.BRIDGproject.org
- **fridsma@cbmi.pitt.edu**