CDISC / Microsoft Workshop

Peter Van Reusel, Sam Hume, David R. Bobbitt

Workshop Agenda

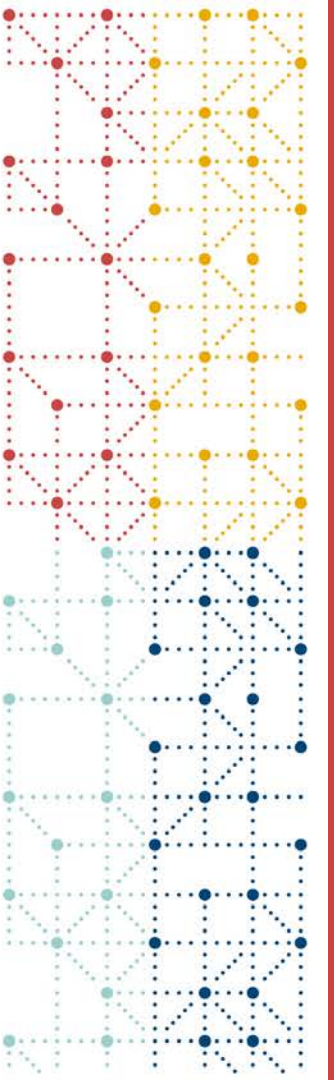
10:00 AM	Welcome	David Bobbitt, Paul Slater
10:30 AM	Current state of the industry (urgency for change)	Ulo Palm
11:00 AM	Technical Briefing • Common challenges for implementation • Missing standards • What follows CDISC 360?	Peter Van Reusel, Sam Hume, Tianna Umann
Lunch break		
12:30 PM	Technical Briefing • Next Generation CDISC Library • Community content curation • Freemium model • New CDISC Titanium membership category	Peter Van Reusel, Sam Hume, Tianna Umann
1:30 PM	Dialog and Q&A	Dave Evans, Moderator
2:30 PM	Wrap up and Concrete next steps	David Bobbitt and Paul Slater
3:00 PM	Adjourn	

Better Standards

Peter Van Reusel, Chief Standards Officer, CDISC

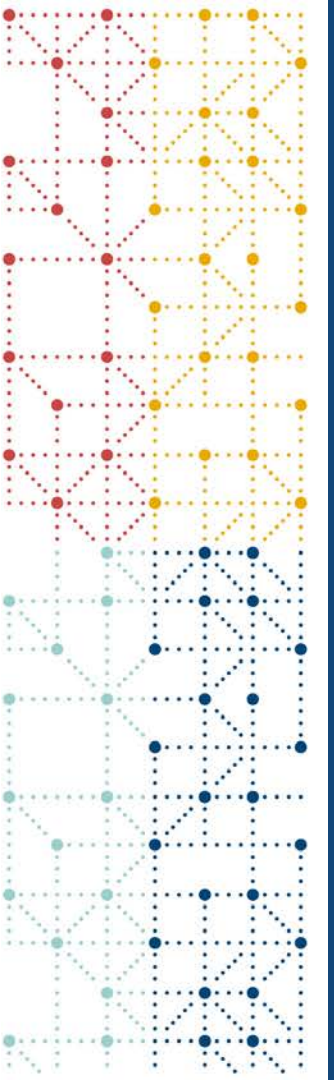
pvanreusel@cdisc.org

10 Sept 2020



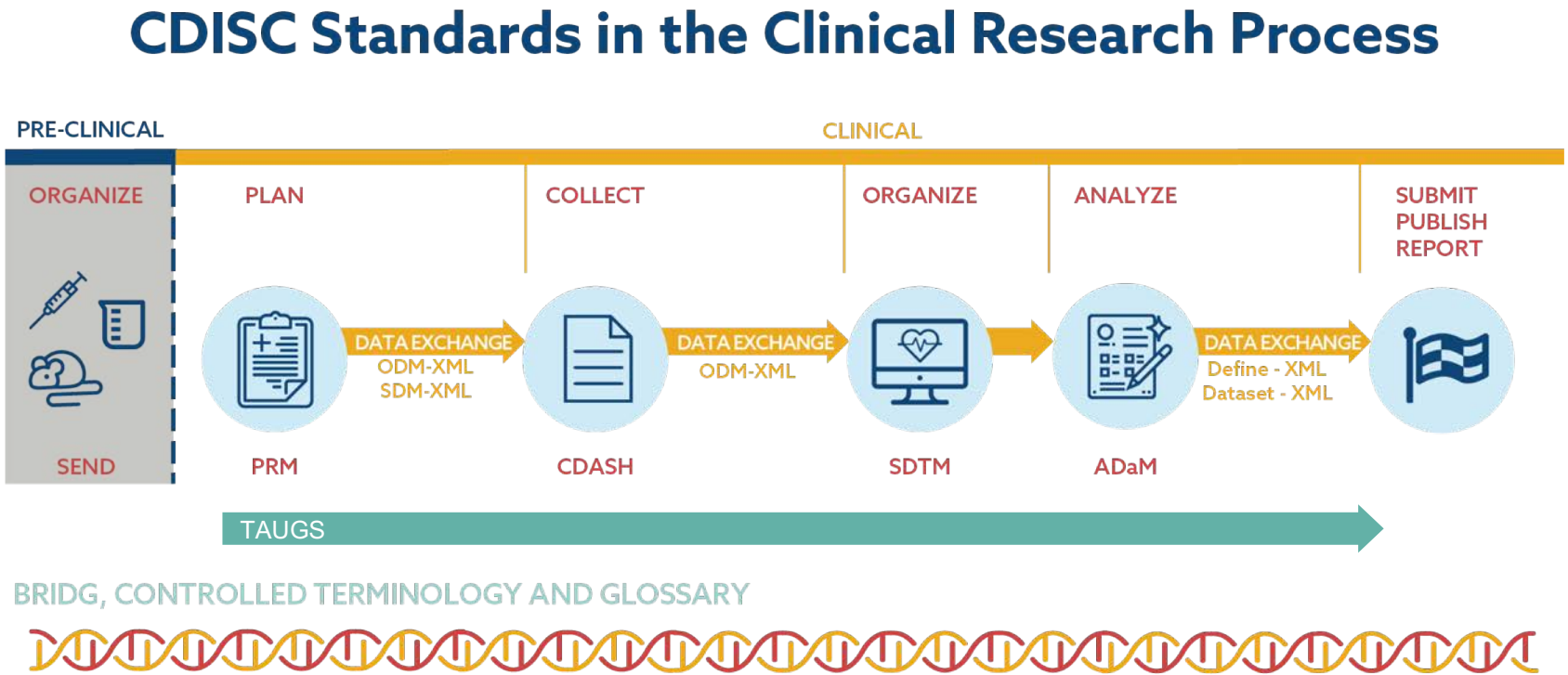
Better Standards

1. Common Implementation Challenges
2. CDISC 360 Status and Lessons Learned
3. What Follows CDISC 360?



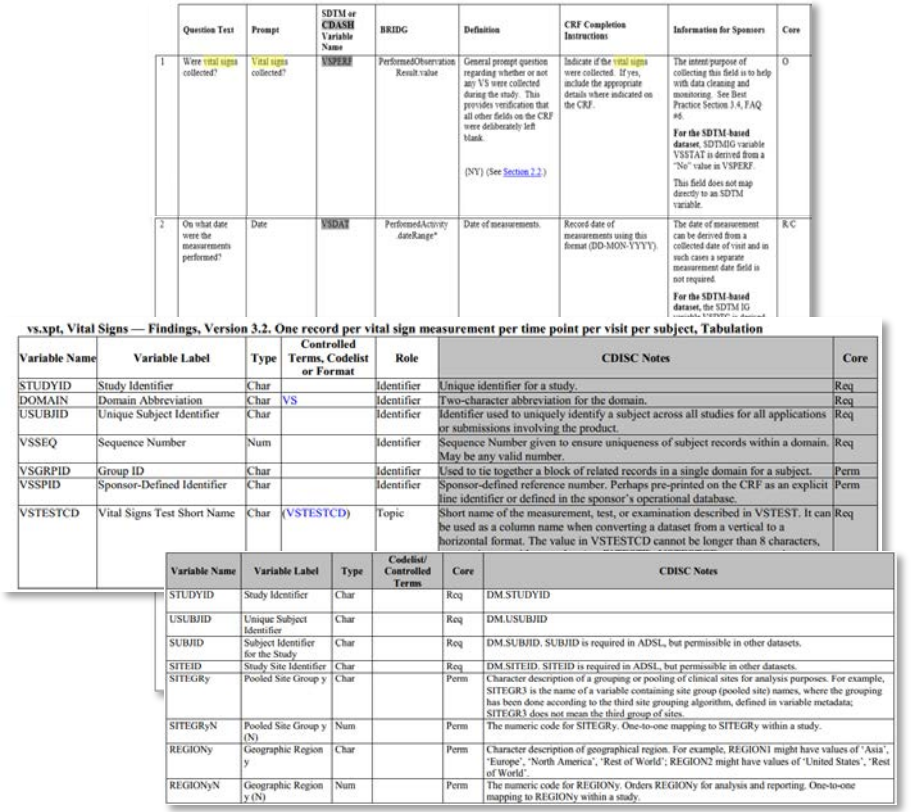
1. Common Implementation Challenges

Today we are here



Benefits Today

- CDISC Foundational models provide much needed structure
  - Normative Content
  - 2 dimensional (tables, columns)
  - Standard to represent data
- The Information itself is not defined
  - We do not need new structures
  - We need to define
    - Entities
    - Semantics (meaning)
    - Relationships between information
    - Rules in the data lifecycle



	Question Text	Prompt	SDTM or CDASH Variable Name	BRIDG	Definition	CRF Completion Instructions	Information for Sponsors	Core
1	Were vital signs collected?	Vital signs collected?	VSPERF	PerformedObservation Result.value	General prompt question regarding whether or not any VS were collected during the study. This provides verification that all other fields on the CRF were deliberately left blank. (NY) (See Section 2.2.)	Indicate if the vital signs were collected. If yes, include the appropriate details where indicated on the CRF.	The intent/purpose of collecting this field is to help with data cleaning and monitoring. See Best Practice Section 3.4, FAQ #6. For the SDTM-based dataset, SDTMIG variable VSTAT is derived from a "No" value in VSPERF. This field does not map directly to an SDTM variable.	O
2	On what date were the measurements performed?	Date	VSDAT	PerformedActivity dateRange*	Date of measurements.	Record date of measurements using this format (DD-MON-YYYY).	The date of measurement can be derived from a collected date of visit and in such cases a separate measurement date field is not required. For the SDTM-based dataset, the SDTM IG	R/C

vs.xpt, Vital Signs — Findings, Version 3.2. One record per vital sign measurement per time point per visit per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
DOMAIN	Domain Abbreviation	Char	VS	Identifier	Two-character abbreviation for the domain.	Req
USUBJID	Unique Subject Identifier	Char		Identifier	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.	Req
VSSEQ	Sequence Number	Num		Identifier	Sequence Number given to ensure uniqueness of subject records within a domain. May be any valid number.	Req
VSGRPID	Group ID	Char		Identifier	Used to tie together a block of related records in a single domain for a subject.	Perm
VSSPID	Sponsor-Defined Identifier	Char		Identifier	Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database.	Perm
VSTESTCD	Vital Signs Test Short Name	Char	(VSTESTCD)	Topic	Short name of the measurement, test, or examination described in VSTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in VSTESTCD cannot be longer than 8 characters,	Req

Variable Name	Variable Label	Type	Codelist/ Controlled Terms	Core	CDISC Notes
STUDYID	Study Identifier	Char		Req	DM.STUDYID
USUBJID	Unique Subject Identifier	Char		Req	DM.USUBJID
SUBJID	Subject Identifier for the Study	Char		Req	DM.SUBJID. SUBJID is required in ADSL, but permissible in other datasets.
SITEID	Study Site Identifier	Char		Req	DM.SITEID. SITEID is required in ADSL, but permissible in other datasets.
SITEGRy	Pooled Site Group y	Char		Perm	Character description of a grouping or pooling of clinical sites for analysis purposes. For example, SITEGR3 is the name of a variable containing site group (pooled site) names, where the grouping has been done according to the third site grouping algorithm, defined in variable metadata; SITEGR3 does not mean the third group of sites.
SITEGRyN	Pooled Site Group y (N)	Num		Perm	The numeric code for SITEGRy. One-to-one mapping to SITEGRy within a study.
REGIONy	Geographic Region y	Char		Perm	Character description of geographical region. For example, REGION1 might have values of 'Asia', 'Europe', 'North America', 'Rest of World'; REGION2 might have values of 'United States', 'Rest of World'.
REGIONyN	Geographic Region y (N)	Num		Perm	The numeric code for REGIONy. Orders REGIONy for analysis and reporting. One-to-one mapping to REGIONy within a study.

What is not in the standards?

- Lack comprehensive data meaning and relationships
- Do not describe the transformations and derivations
- Have flexibility that allows for inconsistencies, making scaling automation difficult, allowing unnecessary variability
- Are published as text instead of machine-readable content with machine executable transformation and derivation algorithms
- Therapeutic Area User Guides provide end-to-end knowledge standardization
  - From data collection to analysis
  - Analog documents, published as text

cdisc

Therapeutic Area User Guide Overview

Therapeutic Area (TA) Standards extend the Foundational Standards to represent data that pertains to specific disease areas. TA Standards include disease-specific metadata, examples and guidance on implementing CDISC standards for a variety of uses, including global regulatory submission.

Autoimmune
- Psoriasis
- Rheumatoid Arthritis

Cardiovascular
- Cardiovascular
- Heart Failure
- QT Studies
- Traditional Chinese Medicine - Coronary Artery Disease-Angina

Endocrine
- Acute Kidney Injury
- Diabetes
- Diabetes - Type 1
- Diabetic Kidney Disease
- Dyslipidemia
- Kidney Transplant
- Polycystic Kidney Disease

Gastrointestinal
- CDAD
- Crohn's Disease

Infectious
- COVID-19
- Ebola
- Hepatitis C
- HIV
- Influenza
- Malaria
- Tuberculosis
- Virology

Mental Health
- Major Depressive Disorder
- Post Traumatic Stress Disorder
- Schizophrenia

Neurology
- Alzheimer's
- Huntington's Disease
- Multiple Sclerosis
- Parkinson's Disease
- Traumatic Brain Injury

Oncology
- Breast Cancer
- Colorectal Cancer
- Lung Cancer
- Pancreatic Cancer
- Prostate Cancer

Other
- Nutrition
- Traditional Chinese Medicine - Acupuncture

Rare Diseases
- Duchenne Muscular Dystrophy

Respiratory
- Asthma
- COPD
- COVID-19

Treatments
- Pain
- Vaccines

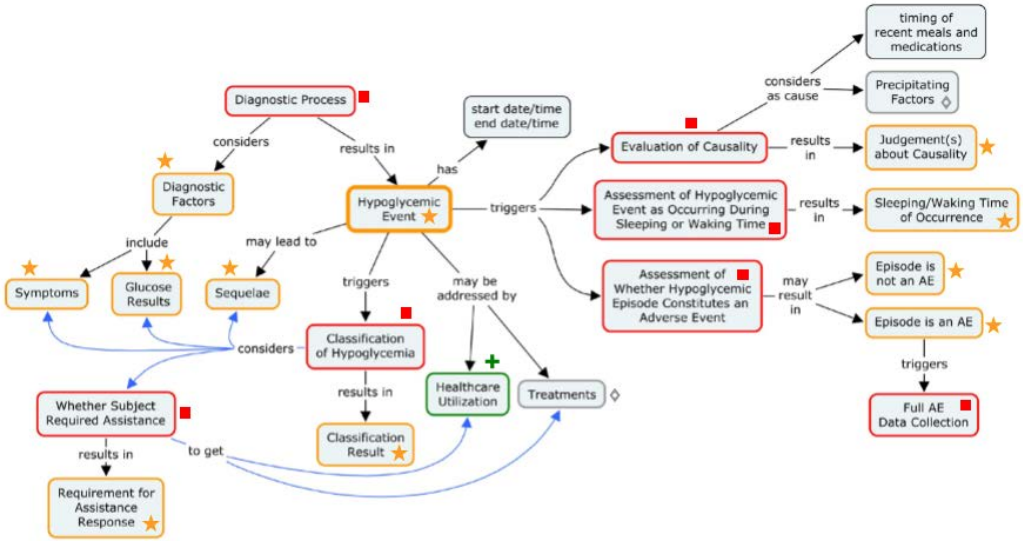
➔ *44 Therapeutic Area User Guides in 8 years*

https://www.cdisc.org/standards/therapeutic-areas/disease-area

Therapeutic Area Concept Maps

- Provides scope and extent of TA User Guide
- Facilitates communication between scientists and data standards experts

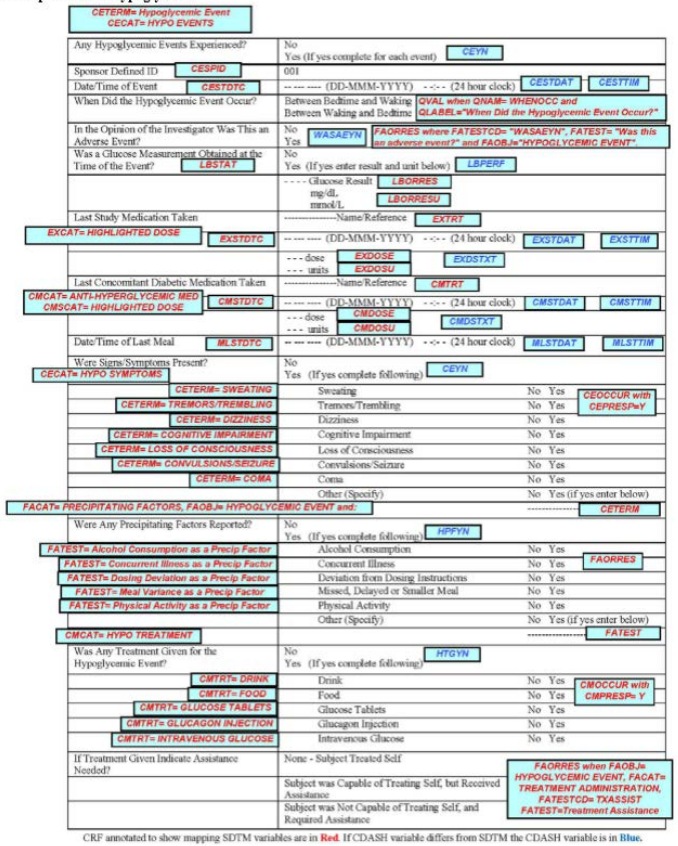
➔ *"This is a picture for humans, this knowledge does not exist for machines"*



Concept Map 7: Hypoglycemic Event

TAUG CRFs and Datasets

Example CRF 5: Hypoglycemia



CRF annotated to show mapping SDTM variables are in **Red**. If CDASH variable differs from SDTM the CDASH variable is in **Blue**.

Row	STUDYID	DOMAIN	USUBJID	CESEQ	CECAT	CETERM	CEDECOD	CEPRESP	CEOCCUR	CESTDTC	CESTDY
2	XYZ	CE	XYZ-001-001	2	HYPO SYMPTOMS	SWEATING	Hyperhidrosis	Y	N		
3	XYZ	CE	XYZ-001-001	3	HYPO SYMPTOMS	TREMORS/TREMBLING	Tremor	Y	N		
4	XYZ	CE	XYZ-001-001	4	HYPO SYMPTOMS	DIZZINESS	Dizziness	Y	N		
5	XYZ	CE	XYZ-001-001	5	HYPO SYMPTOMS	COGNITIVE IMPAIRMENT	Cognitive Disorder	Y	Y		
6	XYZ	CE	XYZ-001-001	6	HYPO SYMPTOMS	LOSS OF CONSCIOUSNESS	Loss of Consciousness	Y	Y		
7	XYZ	CE	XYZ-001-001	7	HYPO SYMPTOMS	CONVULSIONS/SEIZURES	Convulsion	Y	N		
8	XYZ	CE	XYZ-001-001	8	HYPO SYMPTOMS	COMA	Coma	Y	N		
9	XYZ	CE	XYZ-001-001	9	HYPO EVENTS	HYPOGLYCEMIC EVENT	Hypoglycaemia			2013-09-24T08:48	50

Row	RELMIDS	MIDS	MIDSDTC
1 (cont)		HYPO 1	
2 (cont)	DURING	HYPO 1	2013-09-01T11:00
3 (cont)	DURING	HYPO 1	2013-09-01T11:00
4 (cont)	DURING	HYPO 1	2013-09-01T11:00
5 (cont)	DURING	HYPO 1	2013-09-01T11:00
6 (cont)	DURING	HYPO 1	2013-09-01T11:00
7 (cont)	DURING	HYPO 1	2013-09-01T11:00
8 (cont)	DURING	HYPO 1	2013-09-01T11:00
9 (cont)		HYPO 2	

➔ *Human readable only*

TAUG Analysis Datasets and Results

3.3 Hypoglycemic Episodes Summary Dataset

The analysis dataset ADHYSUM is built from an ADHYPO data set and supports both the statistical analysis of the hypoglycemic events and the tabular summary of frequencies of hypoglycemic episodes (see Table 3.3.1). The dataset includes one observation per combination of subject, analysis parameter, time window and indicator (e.g., treatment emergent flag). Each record is a summary of the type of hypoglycemic episode described by the parameter, per subject. For each combination of parameter and the timing variable, AVISIT, records are created even if no hypoglycemic episodes occurred. The statistical model presented below is based on the actual treatment received (TRTA) and adjusted for subject-level values of country and sex. Therefore, these variables are included in ADHYSUM from ADSL to support analysis readiness. The duration of exposure (TRTDURD) is added to the dataset in order to facilitate exposure adjusted incidence rates. For overall summaries the records which have "cumulative frequency count" within the text of PARAM and AVISIT = "End of treatment" can be selected. In this example, parameters for each of the five ADA classification values are defined, along with a derived parameter that represents a grouping of two of the classification values (documented symptomatic or severe hypoglycemia). Mock data for this summary dataset is provided below in Table 3.3.1, yet this mock data shows only a subset of the possible values of analysis parameters. The examples below do not attempt to show all the data needed fully visualize the traceability between ADHYPO and ADHYSUM for a given subject since the volume of required mock data would be large. In practice, however, the counts derived in ADHYSUM for a given subject would be completely traceable to the counts of individual rows for that subject found in the source ADHYPO dataset.

Table 3.3.1: ADHYSUM Analysis Dataset

Row	STUDYID	USUBJID	PARAMCD	PARAM	AVISIT	AVAL	TRTDURD	SEX	AGE	COUNTRY	TRTA
1	XYZ	000008	ASSYMP	Asymptomatic Hypoglycemia (frequency count)	Week 1	3	72	F	55	DZA	Drug B
2	XYZ	000008	ASSYMPC	Asymptomatic Hypoglycemia (cumulative frequency count)	Week 1	3	72	F	55	DZA	Drug B
3	XYZ	000008	ASSYMP	Asymptomatic Hypoglycemia (frequency count)	Week 2	1	72	F	55	DZA	Drug B
4	XYZ	000008	ASSYMPC	Asymptomatic Hypoglycemia (cumulative frequency count)	Week 2	4	72	F	55	DZA	Drug B
5	XYZ	000008	ASSYMP	Asymptomatic Hypoglycemia (frequency count)	Week 3	0	72	F	55	DZA	Drug B
6	XYZ	000008	ASSYMPC	Asymptomatic Hypoglycemia (cumulative frequency count)	Week 3	4	72	F	55	DZA	Drug B
7	XYZ	000008	ASSYMP	Asymptomatic Hypoglycemia (frequency count)	Week 4	1	72	F	55	DZA	Drug B
8	XYZ	000008	ASSYMPC	Asymptomatic Hypoglycemia (cumulative frequency count)	Week 4	5	72	F	55	DZA	Drug B
10	XYZ	000008	ASSYMPC	Asymptomatic Hypoglycemia (cumulative frequency count)	End of Treatment	7	72	F	55	DZA	Drug B

3.4 Hypoglycemic Episodes Summary Analysis Results

The summary statistics in Table 3.4.1 are presented for all hypoglycemic episodes as well as by ADA classification group. The statistics presented in the current example are number of subjects experiencing an event, the number of events, and the raw event rate. To estimate and present the event-rate information, exposure time is needed. Table 3.4.1 is based on the ADHYSUM dataset.

Table 3.4.1: Summary of Hypoglycemic Episodes by Classification – Table Shell

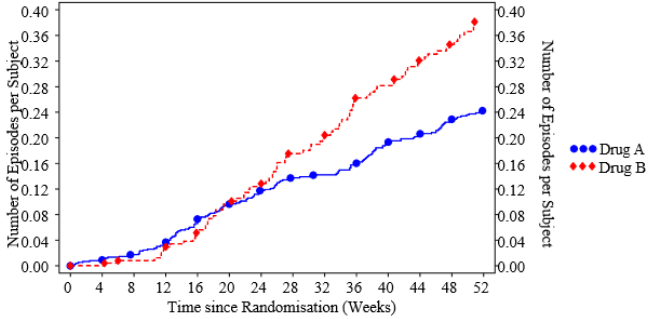
Hypoglycemic Episodes by Classification – Treatment Emergent – Summary – Safety Analysis Set

	Drug A				Drug B				Total			
	N	(%)	E	R	N	(%)	E	R	N	(%)	E	R
Number of subjects	xxx				xxx				xxx			
Total events	xx	(xx.x)	xx	xxx.x	xx	(xx.x)	xx	xxx.x	xx	(xx.x)	xxx	xxx.x
ADA												
Severe hypoglycemia	x	(x.x)	x	xx.x	x	(x.x)	x	x.x	x	(x.x)	x	x.x
Documented symptomatic hypoglycemia	xx	(xx.x)	xx	xxx.x	xx	(xx.x)	xx	xxx.x	xx	(xx.x)	xxx	xxx.x
Asymptomatic hypoglycemia	x	(x.x)	xx	xx.x	x	(x.x)	x	xx.x	xx	(x.x)	xx	xx.x
Probable symptomatic hypoglycemia	x	(x.x)	x	x.x	x	(x.x)	x	x.x	x	(x.x)	x	x.x
Pseudo-hypoglycemia	x				x				x			

N: Number of subjects; %: Percentage of subjects; E: Number of events; R: Event rate per 100 exposure years;
Severe: Subject unable to treat himself/herself and/or have a recorded PG < 3.1 mmol/L (56 mg/dL)
Treatment emergent episodes occur after trial product administration after randomization and no later than 1 day after last trial product administration.

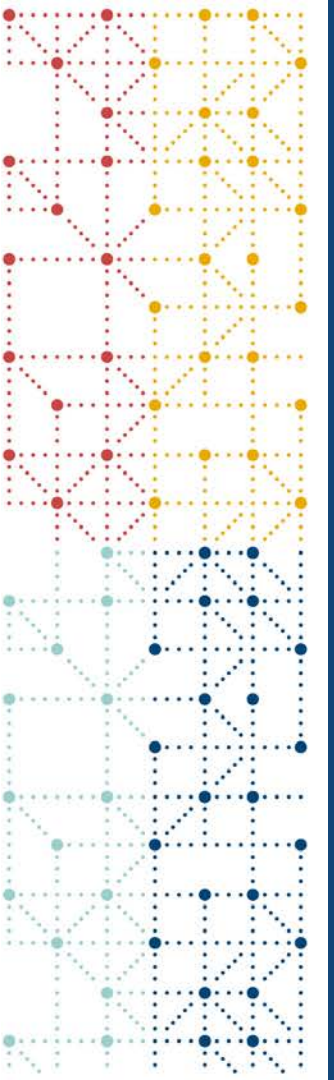
Figure 3.4.1: Mean Cumulative Function Plot of Documented and Severe Symptomatic Hypoglycemic Episodes

Documented and Severe Symptomatic Hypoglycemic Episodes – Treatment Emergent - Mean Cumulative Function - Safety Analysis Set



➔ *Only exists in PDF*

cdisc



2. CDISC 360: Lessons Learned

Piloting, prototyping, testing what works (and what doesn’t)

Biomedical Concepts

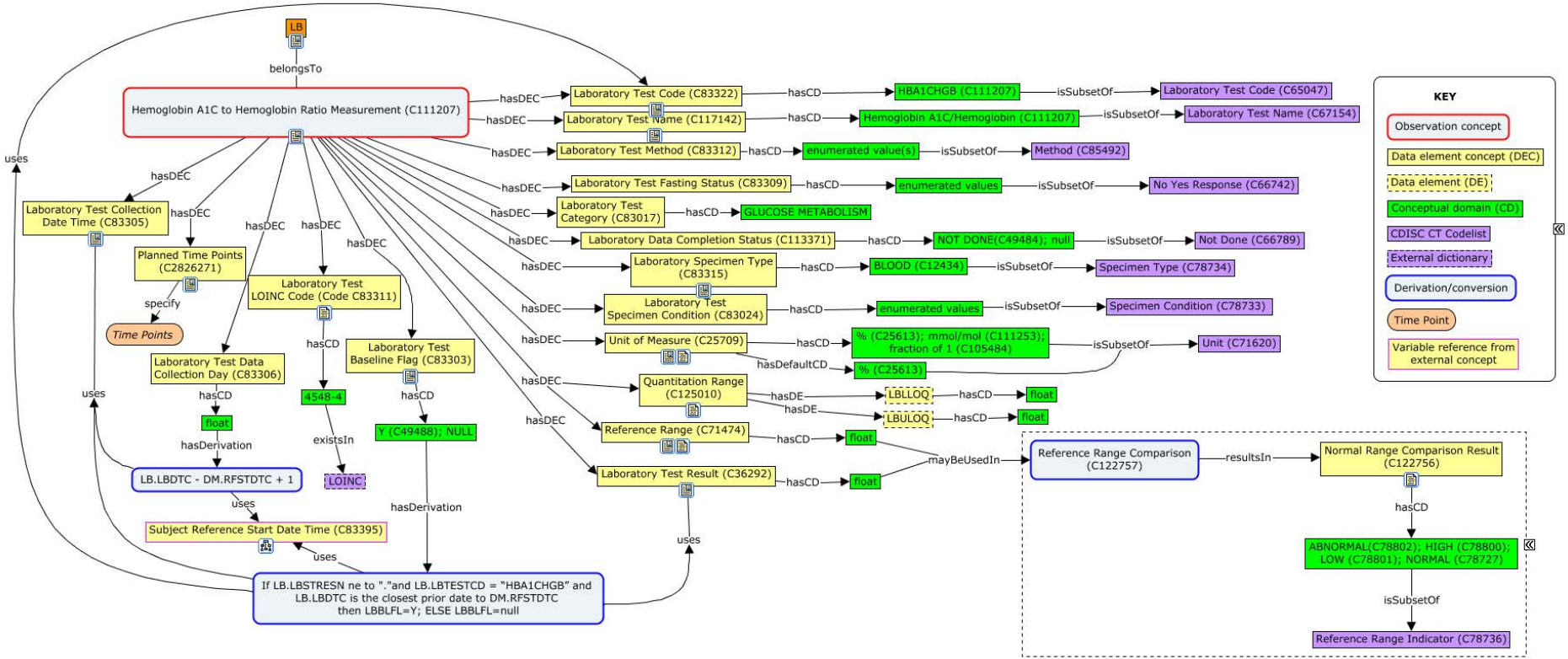
The CDISC 360 Project: Adding a conceptual layer to standards

- Evolve from normative to informative standards
- Create and store standards as concepts which create meaning
- Electronically publish data standards as linked metadata
- Add computer executable process metadata which enables end to end automation
- Develop concept-based standard definitions, and test and demonstrate end-to-end automation of study specification, data processing, and analysis

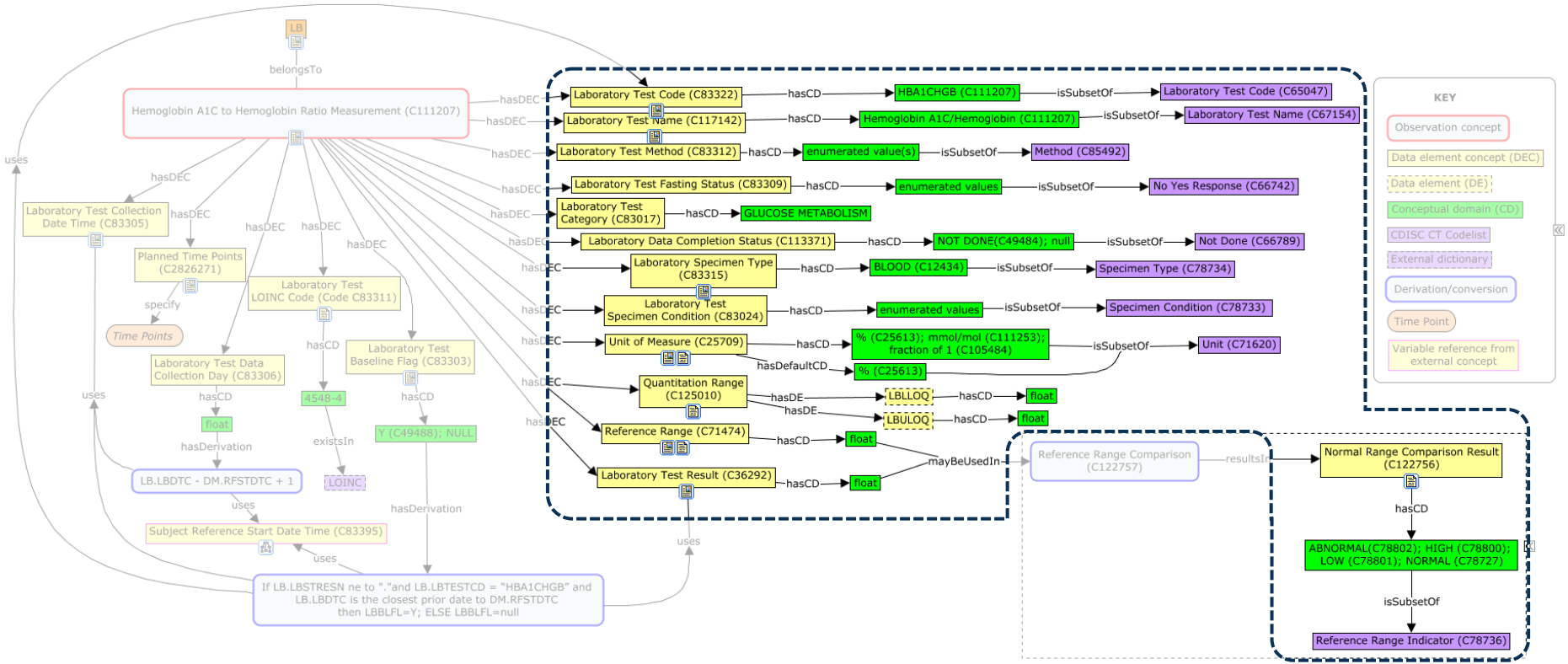
➔ *Test and demonstrate, but* ***not building software***

cdisc

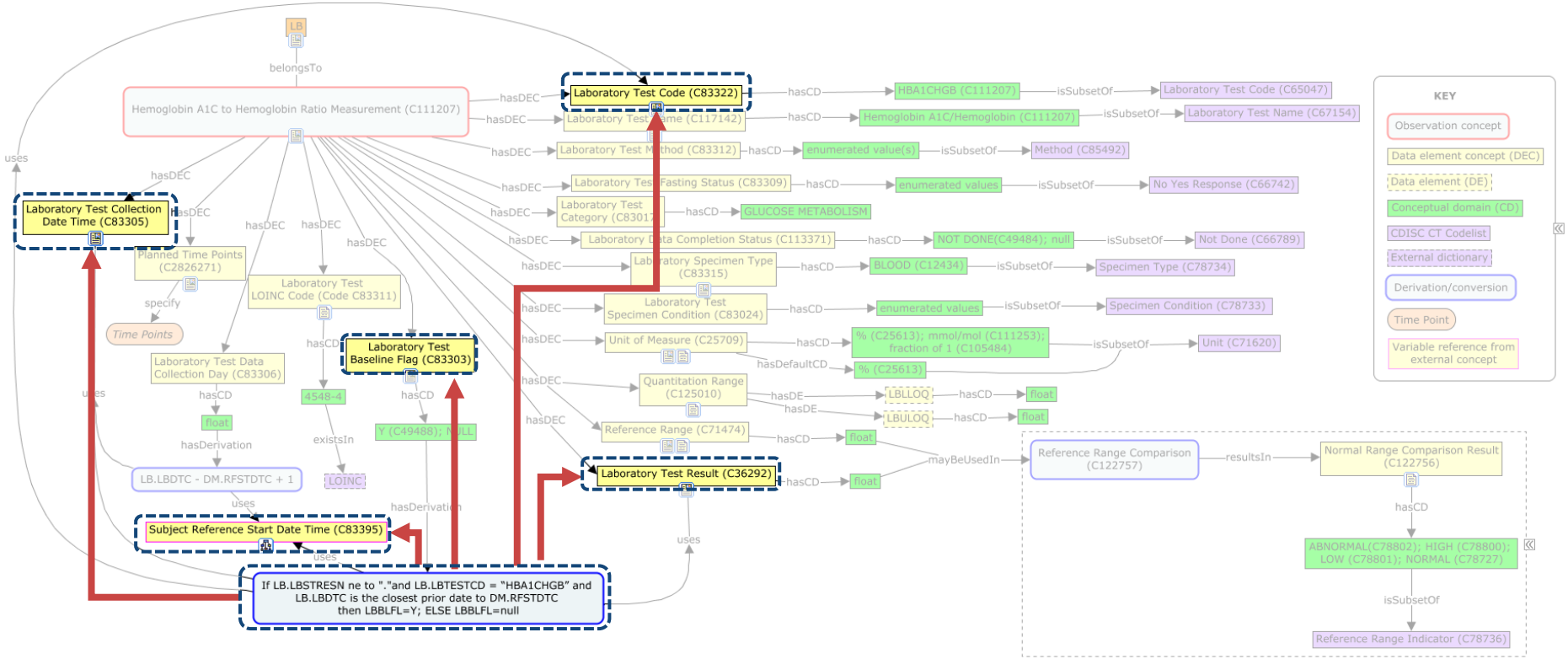
Biomedical Concept

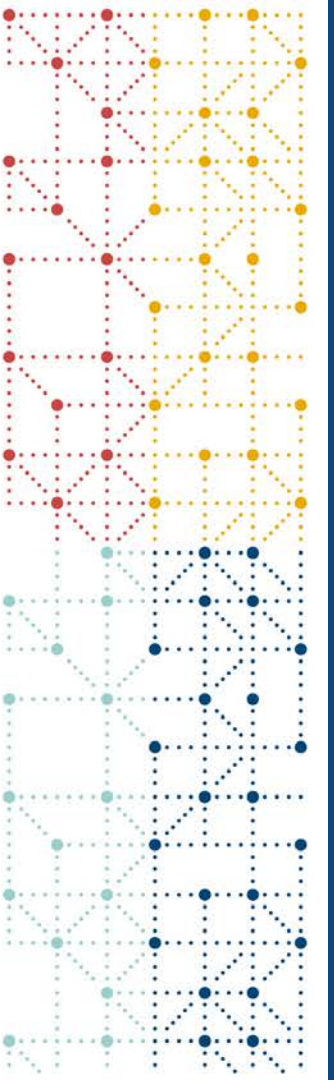


Standardize implementation



Linked derivations & transformations





The Art of the Possible

CDISC 360 – Art of the Possible

- What will follow is a User Experience presentation

- Purpose:
  - Illustrate how the CDISC 360 concept model will enable automation
  - **For illustration only**: CDISC 360 will not deliver software to the industry

- UX presentation link:
  - https://xd.adobe.com/view/93e3e8f6-5b33-405f-4e76-e17af5f29990-e5d2/


abbvie
Allergan
AMGEN
AstraZeneca
Bayer HealthCare
BeiGene
Biogen
BIOMARIN
C&R RESEARCH
Celgene
Cytel
Deloitte.
dMed 缔脉
FDA
formedix
frontier science foundation
GILEAD
Inductive Quotient
Johnson & Johnson
Lilly
Mel Consulting
MERCK
Microsoft
novo nordisk
ORACLE
Otsuka
Pfizer
PharmaStat
PINNACLE 21
POINT CROSS
S-CUBED
SANOFI
sas
Syneos Health
TALENTMINE
VITA DATA SCIENCES a division of SOFTWORLD
xclinical
Clinical Solutions Group

WS1 Highlights and Lessons Learned

- Project Highlights
  - ✓ Developed working definition of biomedical concepts in the context of CDISC 360
  - ✓ Developed biomedical concept template maps based on ISO 11179 for events, findings, and interventions
  - ✓ Developed process for instantiating value-level meta data for biomedical concepts
  - ✓ Developed mappings, derivations, and transformations from CDASH to SDTM

- Lessons Learned
  - Start with an existing metadata model vs. creating one de novo
  - Existing ISO 11179 model was appropriate but not fully sufficient
  - Not an existing road map for this work and thus it took multiple attempts and iterations
  - Better communication needed around project goals scope to decrease confusion about the delineation between CDISC 360 and other similar industry initiatives (e.g., TransCelerate Digital Data Flow).
  - Define components of a biomedical concept better upfront. Struggled with understanding whether a concept should be 360 POC specific vs applicable to all diabetes trials.
  - Need better coordination between work streams with more upfront planning in order to harmonize and streamline deliverables
  - Better unified technology needed within workstreams and across workstreams
    - Use of similar tools across workstreams (CMAP vs. NEO4J)
    - Less variation in the tools needed to create biomedical concepts (CMAPs vs. Excel)
  - Need well defined standards before the metadata can be linked

cdisc

The good

The lessons learned

The not yet achieved

- Standards
- Technology
- Project & Team
- Implementation

Standards

The good

- Developed working definition of Biomedical Concepts in the context of CDISC 360
- Developed template maps based on ISO 11179

The lessons learned

- Need well defined standards before the metadata can be linked

The not yet achieved

- Analysis Concepts are hard to develop, because current analyses are not (sufficiently) standardized

Technology

The good

- Technology plays a critical role
- Modern IT architecture can bring agile, scalable transformation

The lessons learned

- Technical experts dedicated to the project are needed

The not yet achieved

- We did not identify a common (technical) language to automate standard transformations, derivations and validations
- CDISC does not have a ‘sandbox’ Library environment to pilot and test

Project

The good

- Lots of smart, motivated people
- Growth mindset and diversity across team

The lessons learned

- Strong, dedicated agile project manager
- Dedicated staff are needed
- Use case definition and scope are critical

The not yet achieved

- Volunteers provide critical input, but difficult to hang on to the end

Implementation

The good

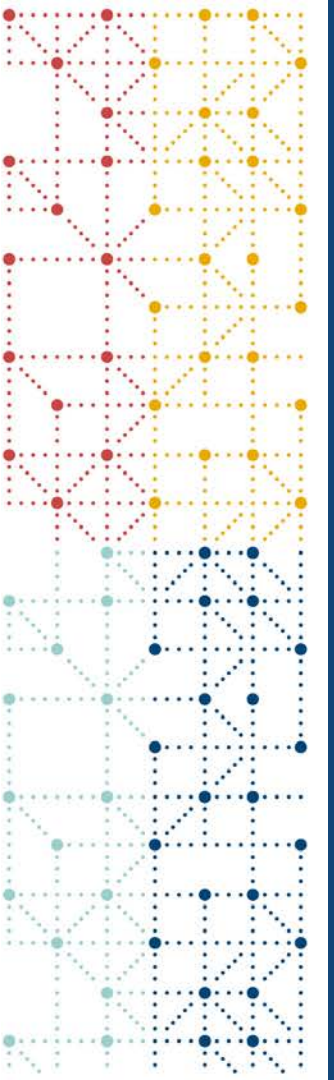
- Implemented basic Study Definition and Design in Study Repository linked to Schedule of Assessment
- Developed & tested metadata elements required for automation

The lessons learned

- Tool agnostic metadata elements work for diverse technologies and tools
- Teams sometimes exceeded expectations, brought new ideas

The not yet achieved

- Where and when can we buy this?



3. What Follows CDISC 360?

What do we need to add or change to our standards?

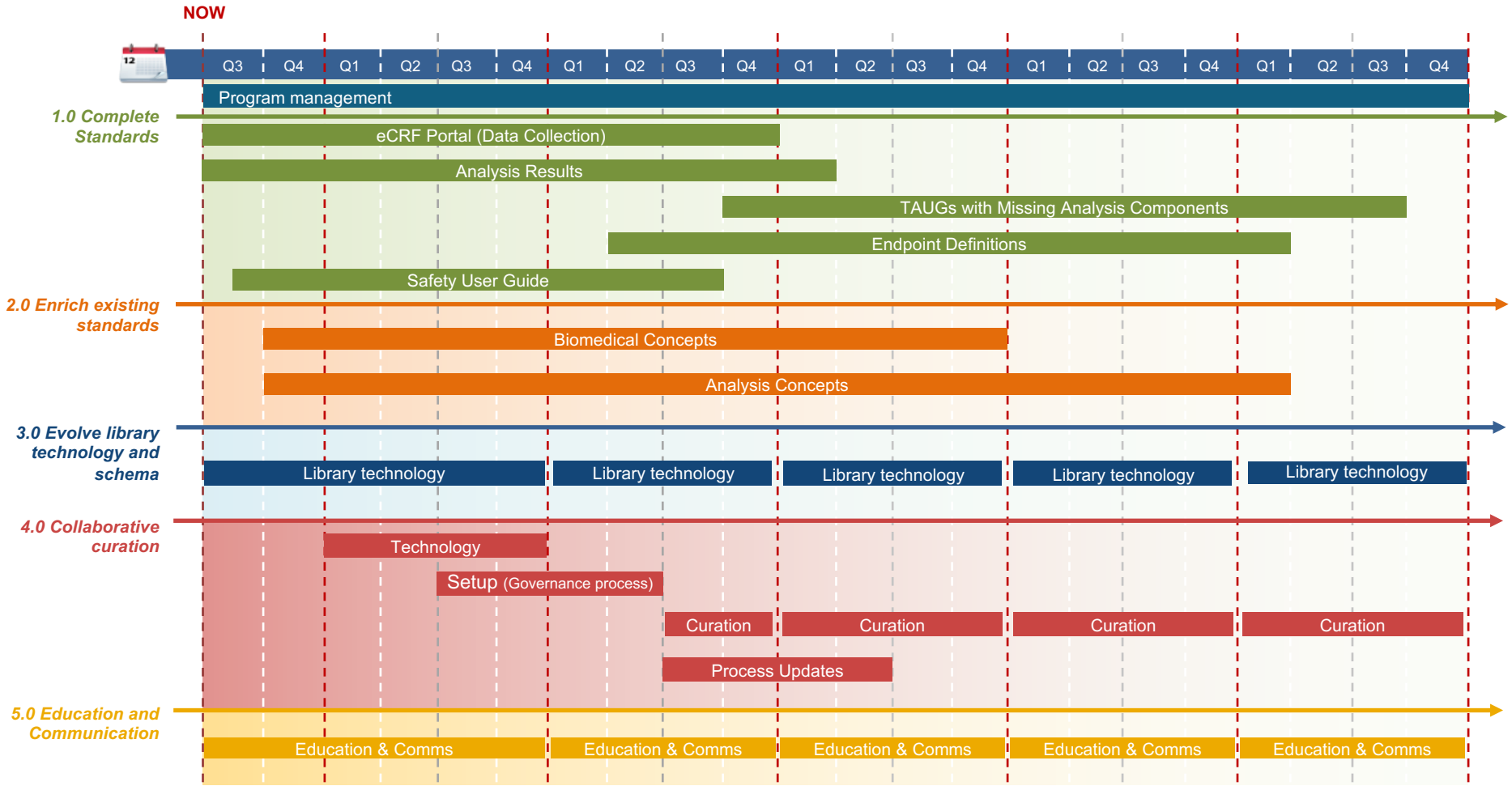
Standards Development

- **Complete end to end** standards
  - Data Collection instruments
  - Analysis Results
  - Endpoint definitions
  - Safety User Guide
    - Collection ➔ Tabulation ➔ Analysis
- **Enrich** existing standards
  - What
    - Standardize implementation (analysis results to collection)
    - Digitizing Therapeutic Areas User Guides
    - Remove unnecessary implementation variability
  - How
    - Stabilize Analysis concept templates (analysis results to biomedical concepts)
    - Stabilize Biomedical concept templates (tabulation to collection)
    - Add transformations and derivations content

Standards Delivery

- Evolve **library** technology and schema
  - Refine and test the CDISC 360 models
  - Refine and deploy CDISC 360 software tools
  - Integrate the CDISC 360 models into the CDISC Library model
  - Update the API to add new CDISC 360 model endpoints
  - Update the CDISC Library Data Standards Browser to include CDISC 360 content
  - Update the CDISC Library standards load software

- Evolve toward collaborative **curation**
  - Develop and rollout governance process
  - Create CDISC Library standards development and curation tools
  - Develop standards curation training
  - Enhance CDISC Library to load community standards implementations


NOW
Q3
Q4
Q1
Q2
Q3
Q4
Q1
Q2
Q3
Q4
Q1
Q2
Q3
Q4
Q1
Q2
Q3
Q4
Q1
Q2
Q3
Q4
Program management
1.0 Complete Standards
eCRF Portal (Data Collection)
Analysis Results
TAUGs with Missing Analysis Components
Endpoint Definitions
Safety User Guide
2.0 Enrich existing standards
Biomedical Concepts
Analysis Concepts
3.0 Evolve library technology and schema
Library technology
Library technology
Library technology
Library technology
Library technology
4.0 Collaborative curation
Technology
Setup (Governance process)
Curation
Curation
Curation
Curation
Process Updates
5.0 Education and Communication
Education & Comms
Education & Comms
Education & Comms
Education & Comms
Education & Comms