CDISC 360 Use Cases: Industry perspectives

Workstream 5 - Build
(Use Case 2) Configure study specification & create artifacts

Tianna Umann, Cross Domain Solution Architect
Clinical Innovation Research Hub, Microsoft
16 October 2019

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

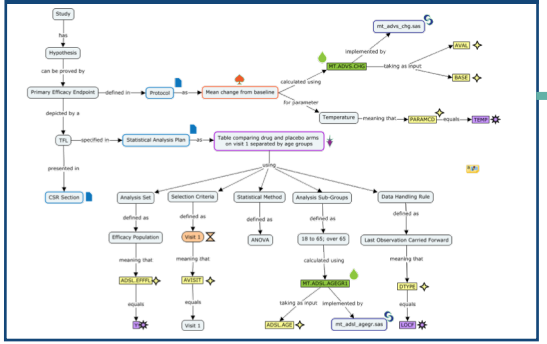
- *{Please disclose any financial relationship or conflict of interest relevant to this presentation here OR}*
- *The author(s) have no real or apparent conflicts of interest to report.*

Workstream 5 (Use Case 2)

Bringing it all together …

CDISC 360 Workstreams

Workstream 1 - *ENHANCE STANDARDS*



Create concepts in knowledge graphs

Workstream 2 - *PUBLISH STANDARDS*

Load into library

Biomedical Concepts
Analysis Concepts
Foundational Standards

cdisc LIBRARY
formerly SHARE

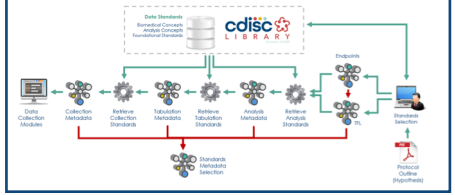
API

Extend API's

Transform concepts in machine readable form

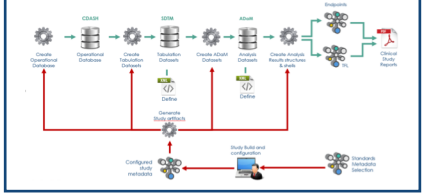
Study Library

Workstream 4 - *DEFINE*



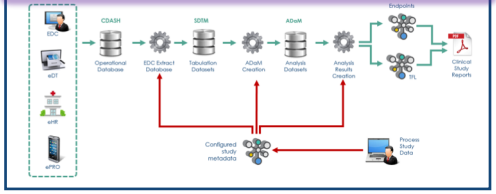
Identify and select standards specification (Use Case 1)

Workstream 5 - *BUILD*



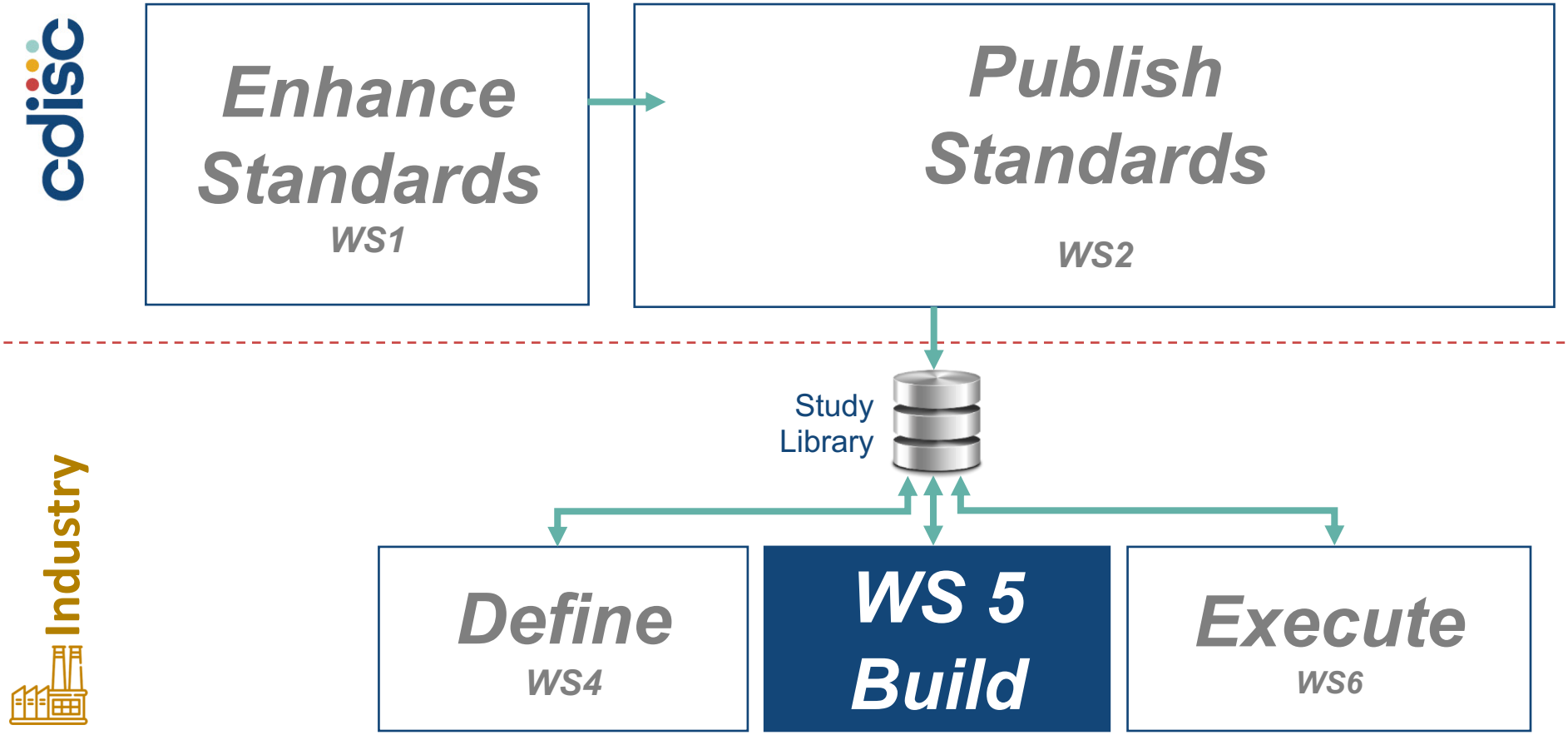
Configure study specification and create artifacts (Use Case 2)

Workstream 6 - *EXECUTE*



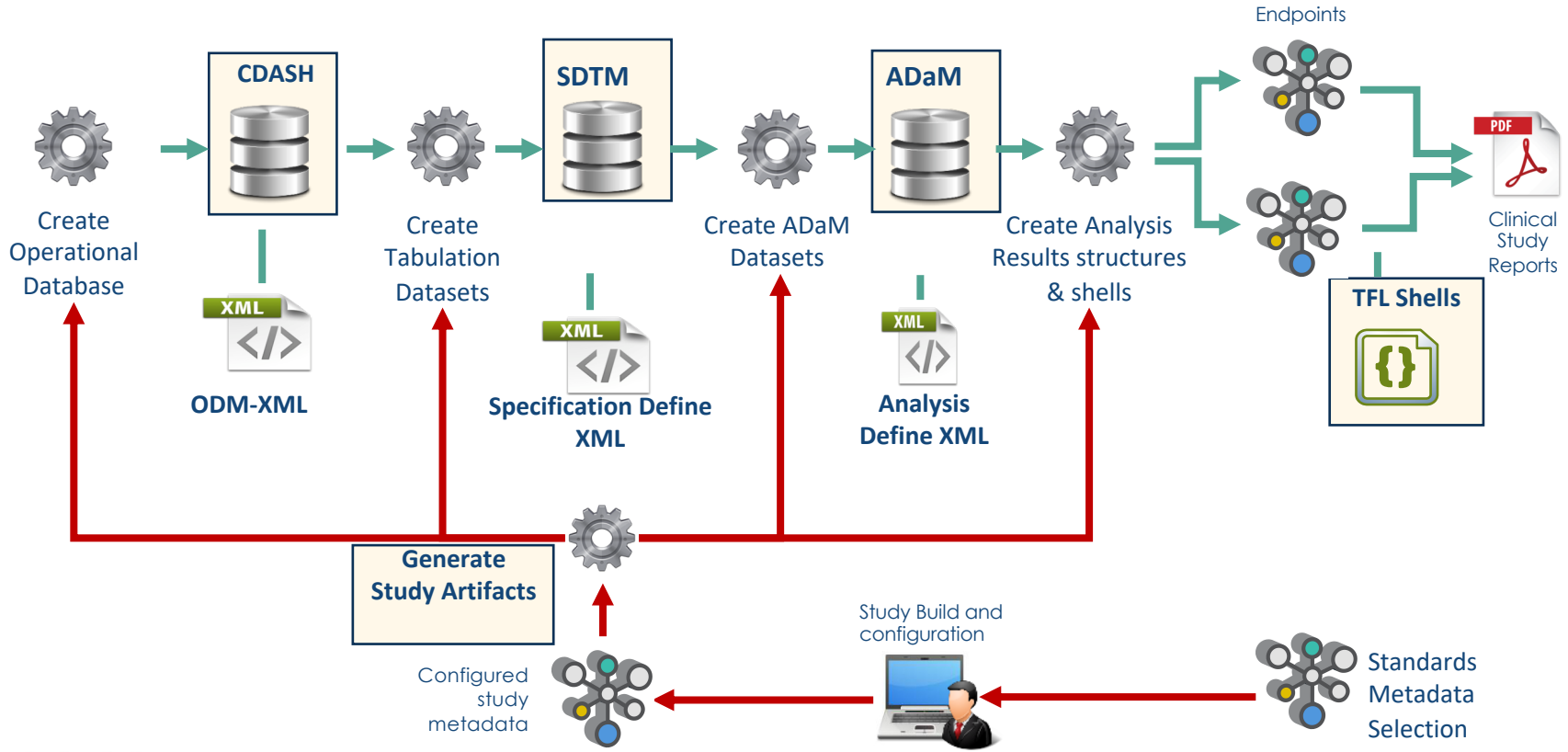
Automatically process and transform data (Use Case 3)

CDISC 360 End to End metadata flow

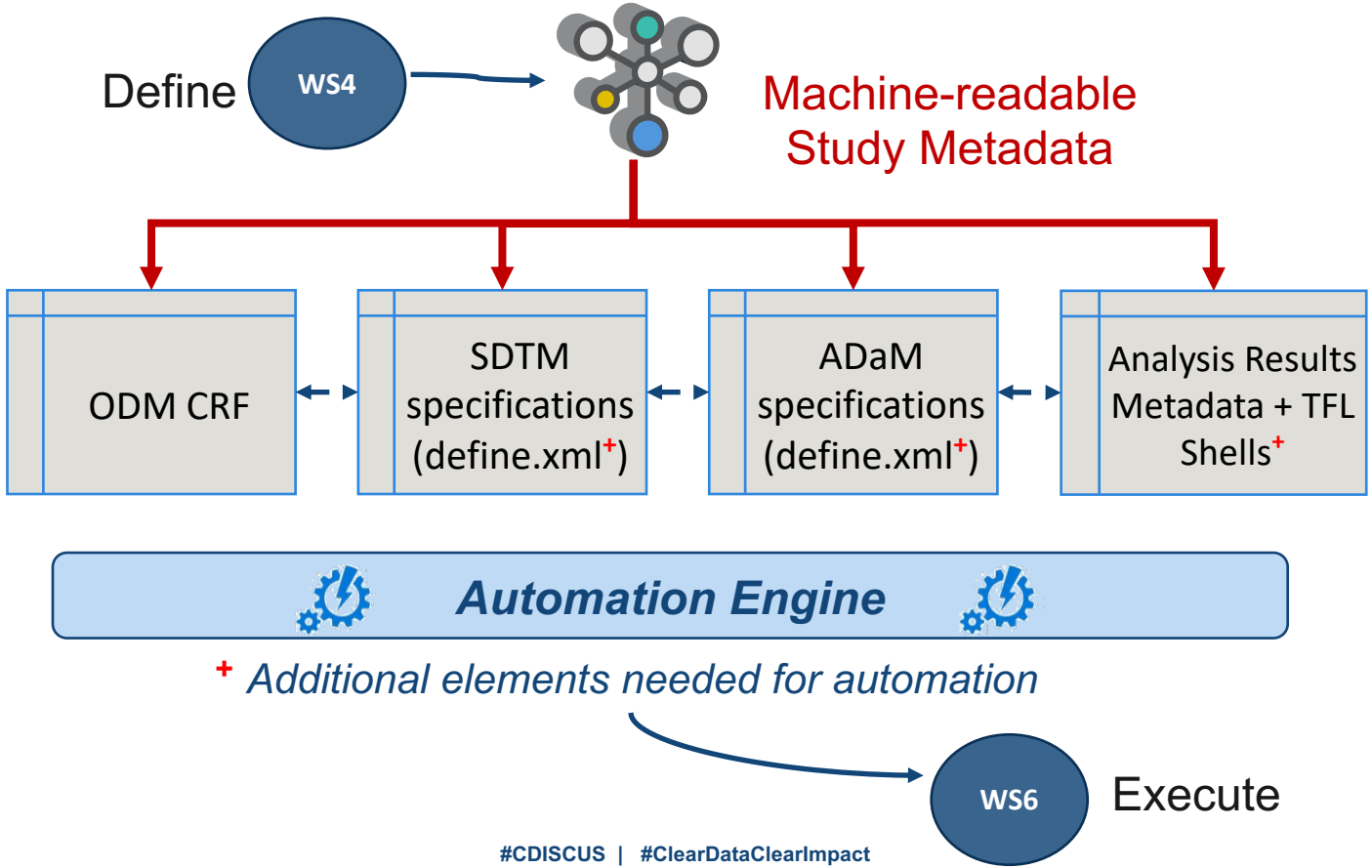


WS 5 - Build

Configure study specification & create metadata artifacts



Build (Workstream 5): Generate Study Metadata Artifacts



Agile Teams Development

Meeting challenges through a Agile and Growth Mindset

CDISC 360 Workstream Deliverable Alignment

WS 1

Create concepts in knowledge graphs

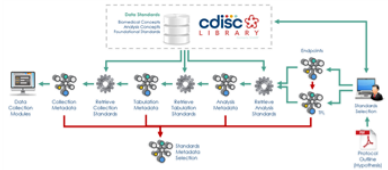
WS 2

Transform concepts in machine readable form

Biomedical Concepts
Analysis Concepts
Foundational Standards

WS 4

Identify and select standards specification

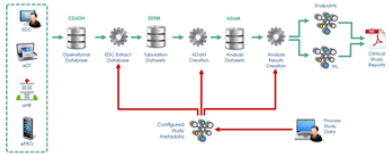


WS 5

Configure study specification and create artifacts

WS 6

Automatically process and transform data



FDA

MACE+/AKI POC Endpoints

Cross Workstream User Stories, Data Flow & Deliverables

As a < type of user >, I want < some goal > so that I can< some reason >

Story Title	User, Goal & Reason	User Action (technology)
Search standard definitions	**As a** <Study Builder> **I want to** <have flexible searching capabilities in the CDISC Library including the new biomedical & analyses concept maps. Search should be non-programmatic & user friendly> **So that I can** <find data standards and concept maps that are to be selected for my study.>	Study Builder can search in the tool to review models /elements directly taken from the CDISC Library.

cdisc

CDISC 360 Cross Workstream Task Teams

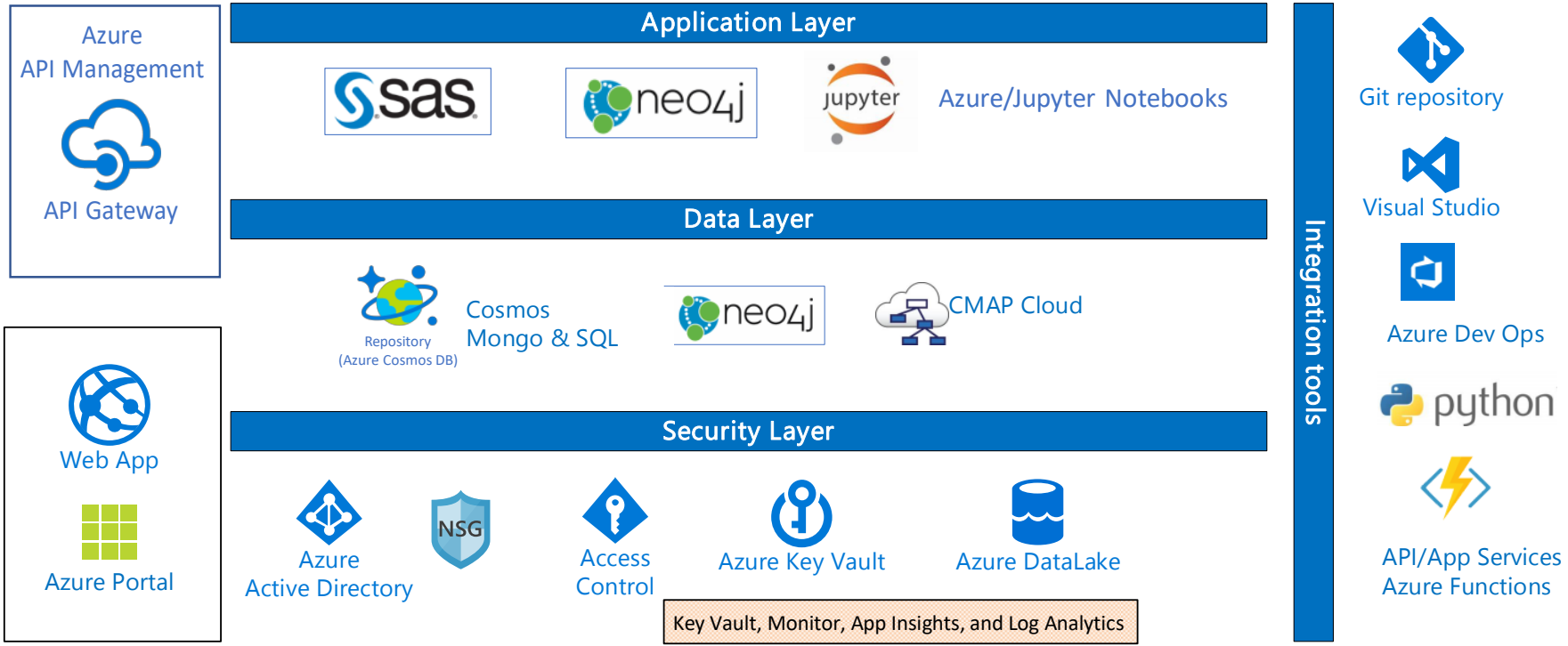
WS1
WS2
WS4
WS5
WS6
FDA

	MACE+/AKI Endpoints Task Team	Specification Define XML+ Task Team	List Study Metadata Task Team	Generate Test Data Task Team
Objective	Define MACE+/AKI Study Endpoints in context of WS6 deliverable requirements	SDTM & ADaM Specification Define XML+ with Value Level meta data, Source Mappings and Derivations	Align WS4 libraries meta data with WS5 workbook, populate query templates and generate CSV files	Coordinate, Develop scripts & Review of generated test data to support the CDISC 360 POC
Sprint 4				
Sprint 5				
Sprint 6				
Sprint 7				

Architecture & Technology approach

Building a flexible and open cloud platform

CDISC 360 POC Azure Cloud & Collaboration Environment



Collaboration Tools

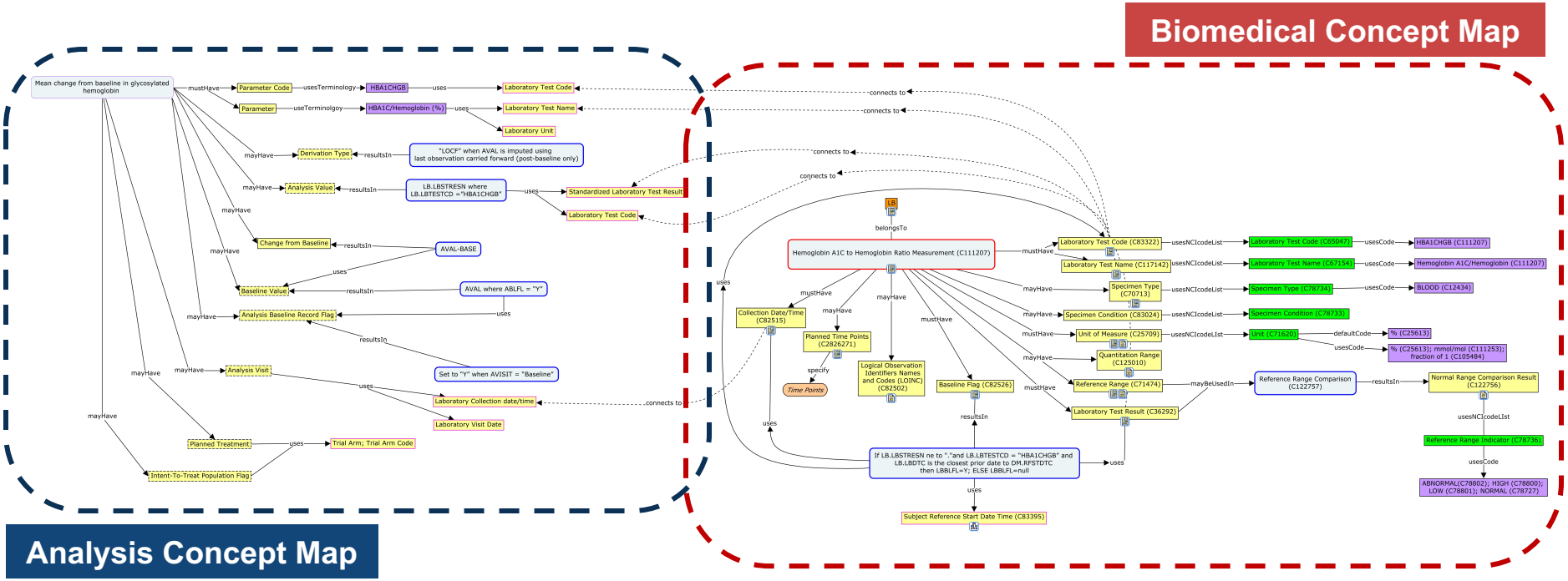
Atlassian Confluence slack JIRA

cdisc

Lessons Learned toward an Automation Model

Team collaboration and Alignment

The Power of One Model



→ The Biomedical Concept and Analysis Concept are ***ONE MODEL***

The Power of a Conceptuel Model for Data Standards

- Linking controlled terminology to the variable – standardize value level metadata
- Machine readable definition of validation rules
- Linking derivations and algorithms to variable(s)
  - Include process metadata (ETL instructions)
- Possibility to standardize Analysis outputs and Collection instruments
  - Combining layout, variables, process information together
- Link Analysis Concepts to Biomedical Concepts
  - Choose an analysis and automatically obtain all related end-to-end metadata

→ ***All of the above:*** *enables automation, increase confidence in results, true analysis traceability*

cdisc

What & How to Evolve – Lessons learned

Agile & Growth Mindset:

- Align through cross workstream teams across sprints to accelerate knowledge sharing and development
- Create and store standards as concepts which create meaning between data
- Standardize the meaning of the information
- Define the data processing (data flow)
- Provide machine-executable data flow definitions

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

Work in Progress

cdisc

Thank You!

Tianna Umann

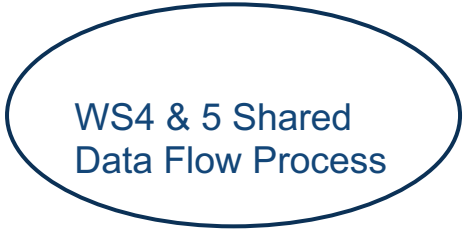
tianna.umann@microsoft.com

Cross Workstream User Stories, Data Flow & Deliverables

As a < type of user >, I want < some goal > so that I can< some reason >

WS5 User Stories

WS4 & 5 Shared
Data Flow Process



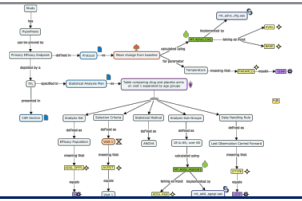
Deliverables

- Metadata Listings
- Protocol Metadata Document
- Study Design specification
- ADaM define.xml specification
- SDTM define.xml specification
- ODM.xml specification
- TFL Specification
- JSON

CDISC 360 Workstream Deliverable Alignment Phase I

WS 1

Create concepts in knowledge graphs



WS 2

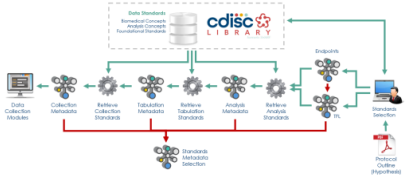
Transform concepts in machine readable form

Biomedical Concepts
Analysis Concepts
Foundational Standards

cdisc LIBRARY

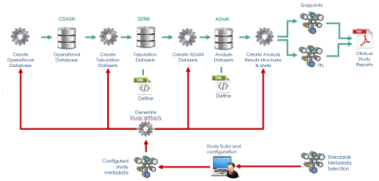
WS 4

Identify and select standards specification



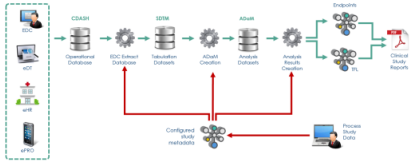
WS 5

Configure study specification and create artifacts



WS 6

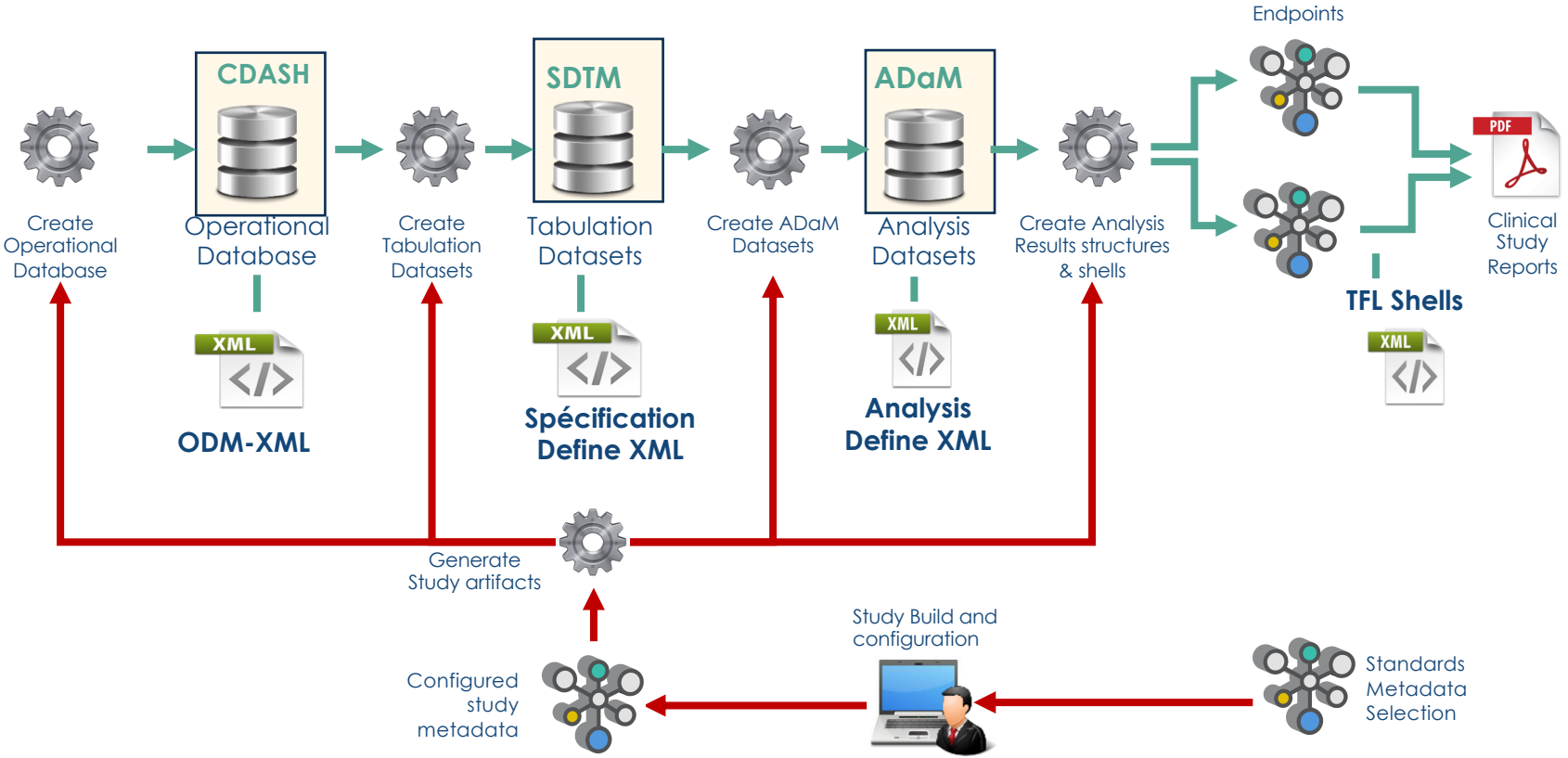
Automatically process and transform data



cdisc

WS 5 - Build

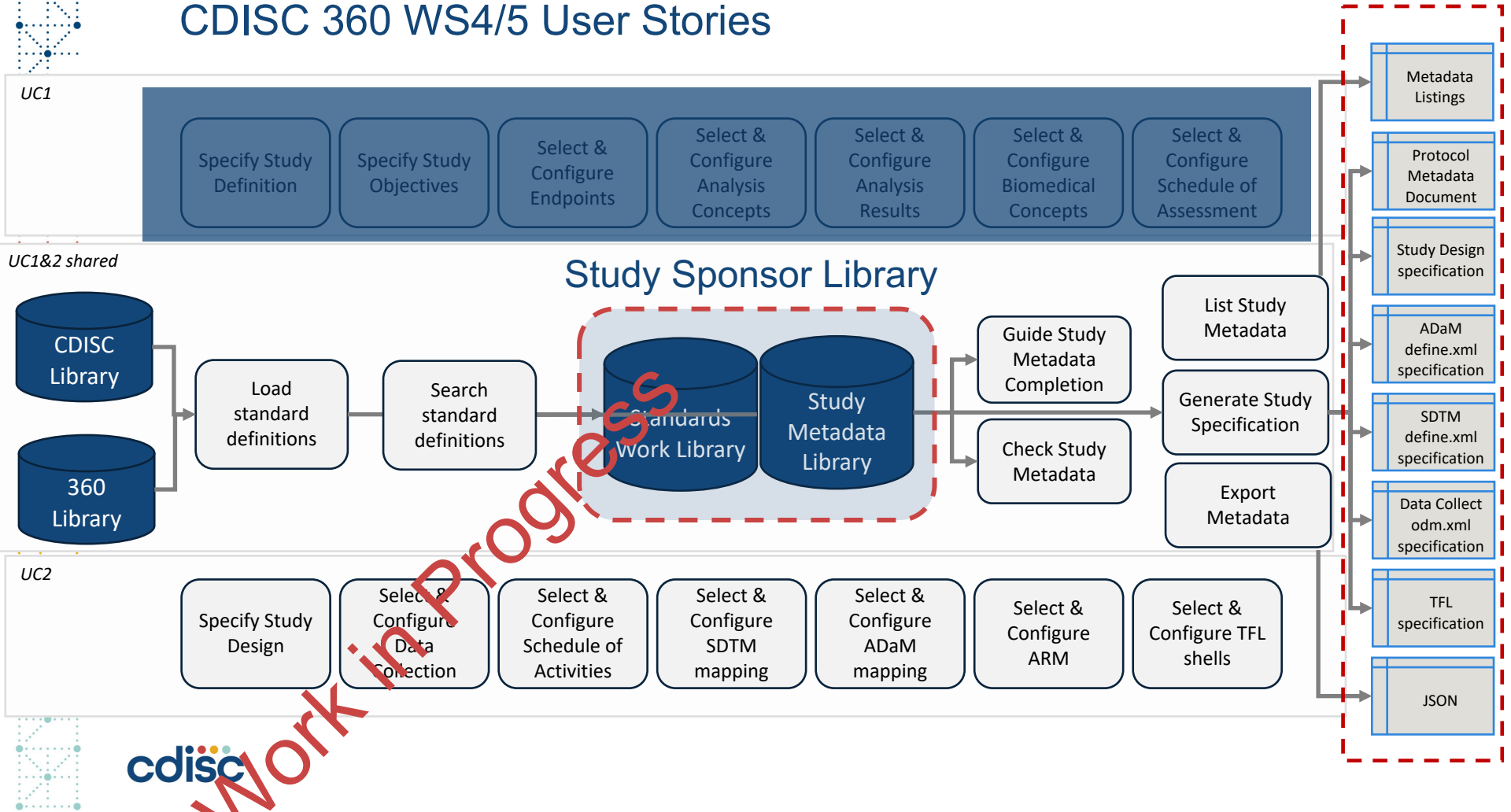
Study Specification Metadata Generation



POC Automation – evolution

- Defining Automation
  - Current state workflow with automation of manual processes
  - Re-engineering the workflow and process
  - The importance of Biomedical Concepts
    - How we do Biomedical Concept mapping now
    - How we might do it in the future leveraging Machine Learning & AI
    - Vision and Goal (sample)
    - Goal of Automation is End to End

CDISC 360 WS4/5 User Stories

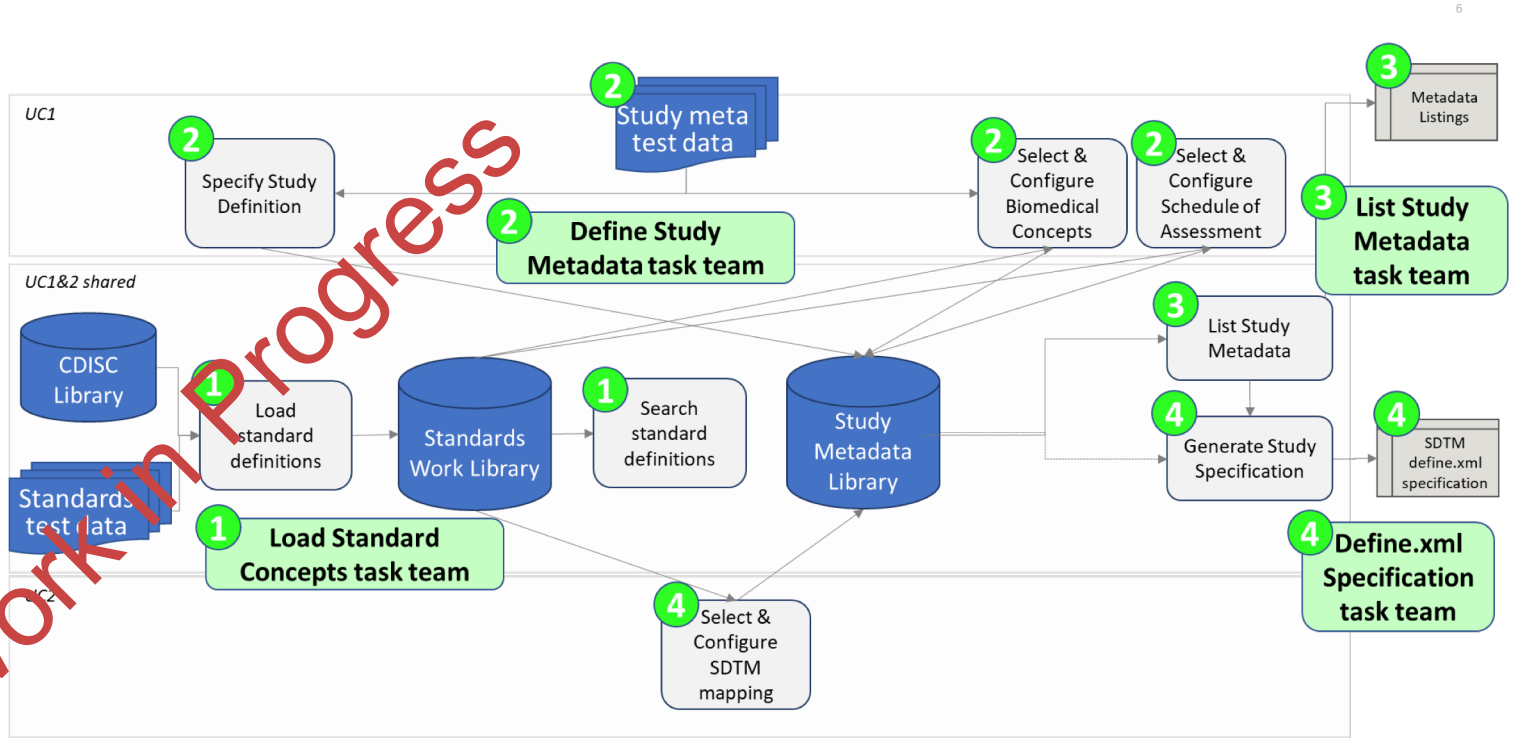


CDISC 360 Cross Workstream Task Teams

	Objective	WS1	WS2	WS4	WS5	WS6	FDA
MACE+/AKI Endpoints Task Team	Define Study Endpoints in context of WS6 deliverable requirements						
Specification Define XML+ Task Team	SDTM & ADaM Specification Define XML+ with Value Level meta data, Source Mappings and Derivations						
List Study Metadata Task Team	Align WS4 libraries meta data with WS5 workbook, populate query templates and generate CSV files						
Generate Test Data Task Team	Coordinate, Develop scripts & Review of generated test data to support the CDISC 360 POC						

cdisc

Ws4-Ws6 User Stories Overview – task teams



What do we need to do to evolve? Where we are

- Standardize the meaning of the information
- Define the data processing (data flow)
- Provide machine-executable data flow definitions
- Standardize missing parts:
  - Protocol content
  - Collection instruments
  - Analysis / endpoint definitions and outputs
- Make standards less complex for the end users
- Publish standards from one trusted source

How do we evolve? Where we are going

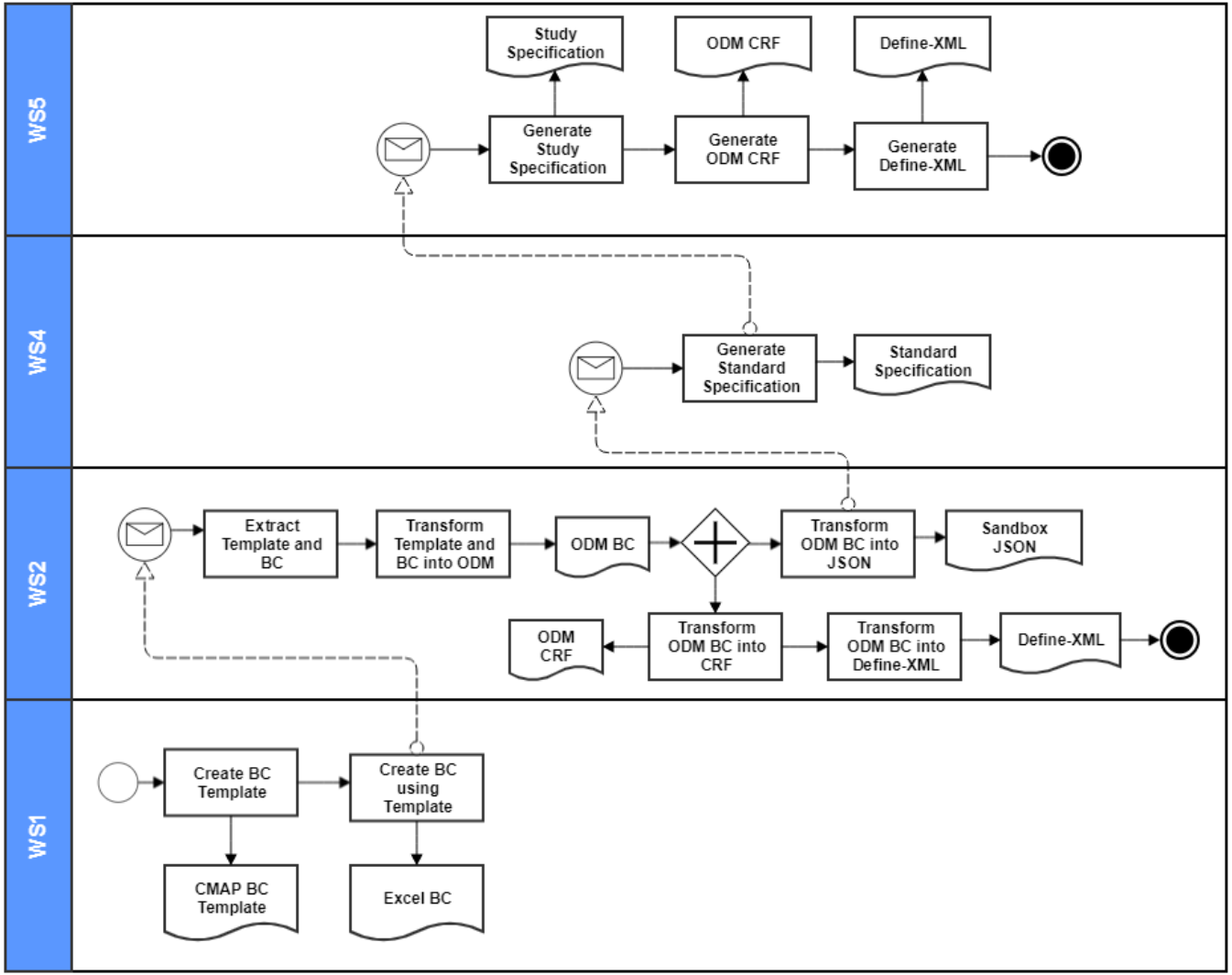
The CDISC 360 Project: Adding a conceptual layer to standards

- Create and store standards as concepts which create meaning between data
- Electronically publish data standards as linked metadata
- Add computer readable process metadata which enables end to end automation
- CDISC 360 will 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

CDISC 360 Interchange Demo

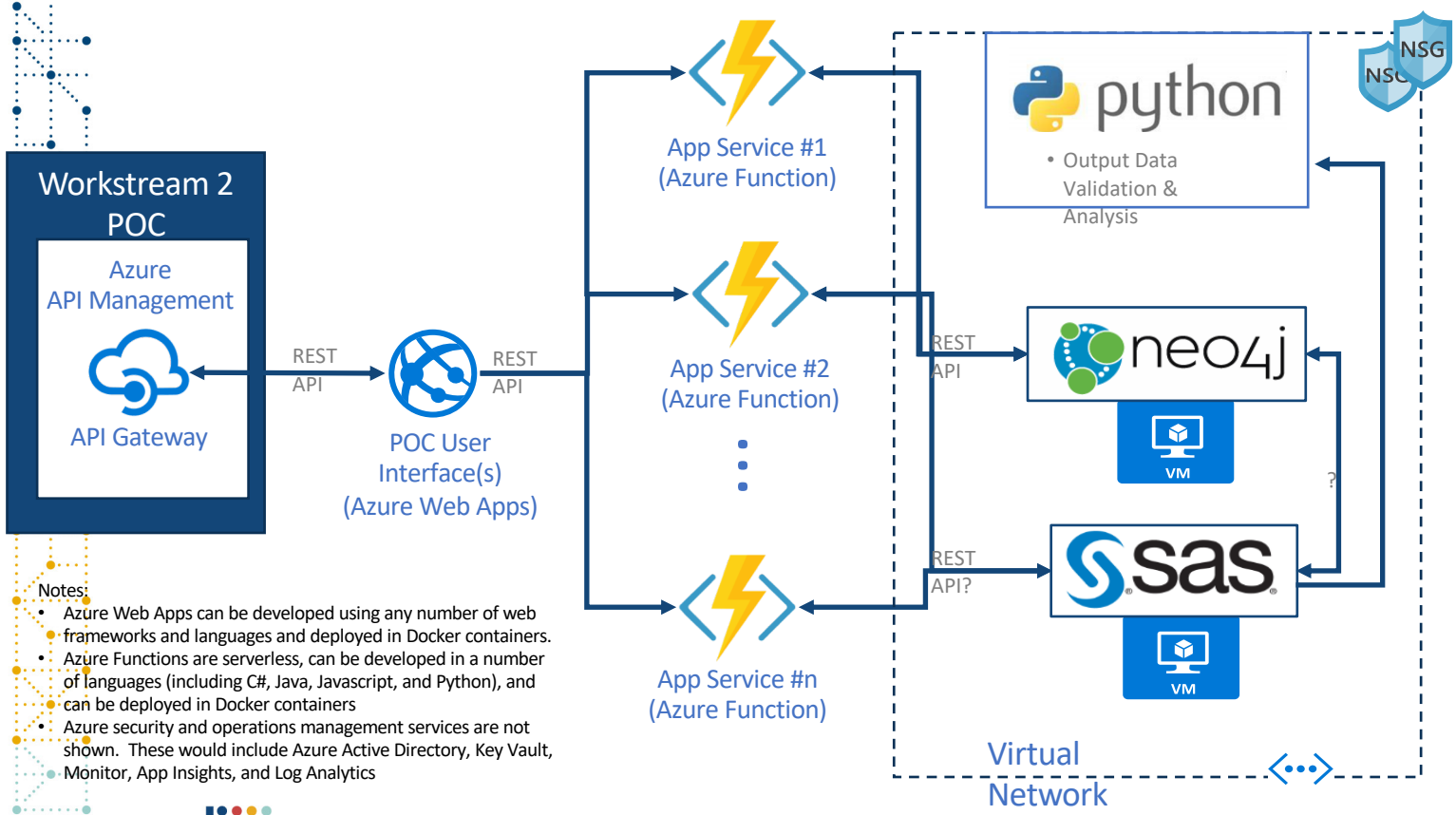


Meeting Purpose

- A bit closer look at each workstream; set foundation for first sprint*
- Provide project members perspective on:
  - Their workstream and all other workstreams
  - Dependencies between workstreams
- Topics covered per workstream:
  - Big picture: objectives, value proposition, dependencies, challenges
  - Inputs
  - Activities
  - Deliverables
- Next steps

**15 min/workstream avg., thus limited dialog is possible today; project leads will further meet with workstream leads/teams as needed*

Workstreams 4-6 POC Target Architecture - Draft



Notes:

- Azure Web Apps can be developed using any number of web frameworks and languages and deployed in Docker containers.
- Azure Functions are serverless, can be developed in a number of languages (including C#, Java, Javascript, and Python), and can be deployed in Docker containers
- Azure security and operations management services are not shown. These would include Azure Active Directory, Key Vault, Monitor, App Insights, and Log Analytics

Open Issues:

- Will we develop one integrated web app for the POC user interface across workstreams 4-6?
- Will Neo4j serve as the primary
- integrated database across workstreams 4-6? Or will some workstreams prefer to use files or other databases?
- Can all the SAS functions/scripts be exposed via REST APIs?
- Can/will SAS connect directly with Neo4J to read inputs and write outputs or will we need to develop custom integration?
- Do we want to consider leveraging Azure Logic Apps for orchestrating long running services and increasing development productivity via config vs. code.... even though Logic Apps can't be containerized?
- Should we run the Azure web app and functions within a vNET (ASE) to ease integration with Neo4J, and SAS?
- Will we need Azure event grid to queue requests between services or is that overkill for this POC?

cdisc

Code Development Management

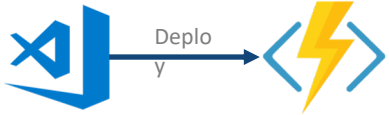
Azure DevOps

- Git Repository
- CI/CD Pipeline
- Test Management

Code Development & Execution

Azure/Jupyter Notebooks

- Code Development
- Code Execution



Visual Studio Code

- Code Development

Azure Functions

- Code Execution

- Code Development
- Code Execution
- Interim Data Storage

- Output Data Validation & Analysis
- Code Development
- Code Execution

Data Storage

Azure Data Lake Store Gen 2

Azure Files

- File Storage

- Graph Database

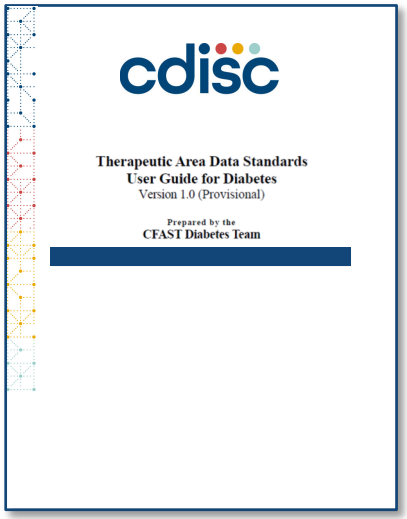
cdisc

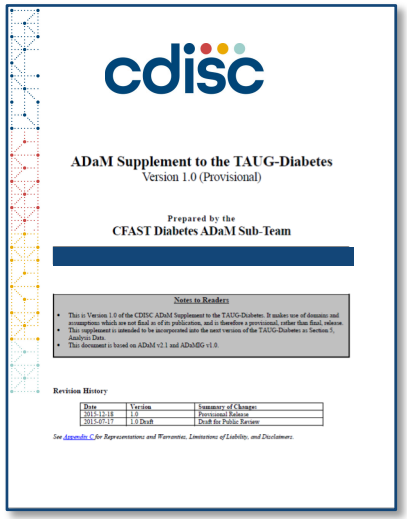
Standards Scope

From the Diabetes TAUG

Project Standards Scope

Diabetes TAUG



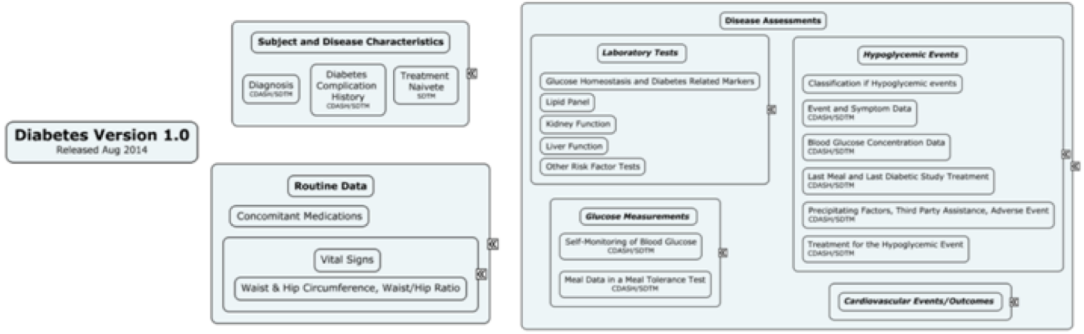


- 1 or 2 statistical endpoints
- 3 to 4 ADaM datasets
- 7 to 8 SDTM datasets
- 15 Data Collection Modules

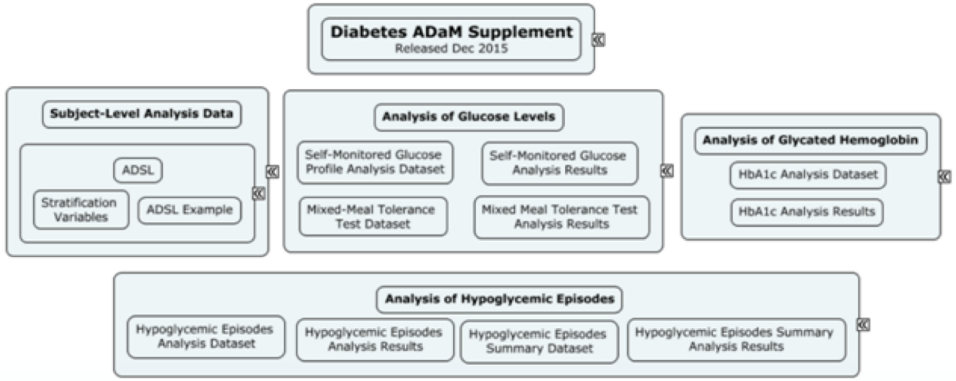
→ ***Reason for this scope:*** *the Diabetes TAUG provides standardized artifacts from analysis outputs to data collection. This allows the project team to focus on innovation and not on establishing a new data standard.*

Diabetes TAUG

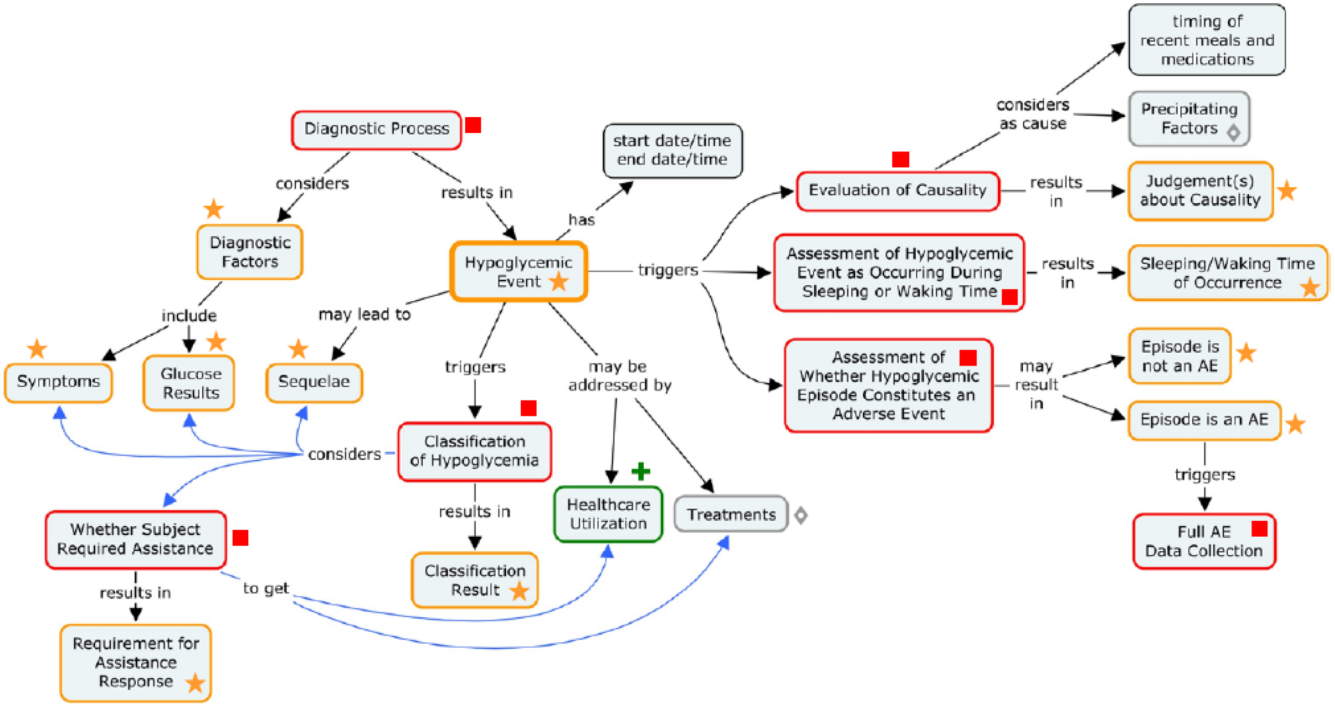
Diabetes V1



Diabetes ADaM V1



Biomedical Concept Map



Analysis Results Shells

Table 3.2.1: Summary of Post-Meal Hypoglycemic Episodes by Severity – Table Shell

Hypoglycemic episodes within 2 hours since last meal by severity
Summary – Safety Analysis Set

	Drug A			Drug B		
	N	(%)	E	N	(%)	E
Number of subjects	xxx			xx		
Diurnal	xxx	(xx.x)	xxx	xx	(xx.x)	xxx
Documented Symptomatic	xx	(xx.x)	xx	xx	(xx.x)	xx
Pseudo Symptomatic	xx	(xx.x)	xx	xx	(xx.x)	xx
Probable Symptomatic	x	(xx.x)	xx	x	(x.x)	x
Nocturnal	x	(x.x)	x	x	(x.x)	x
Documented Symptomatic	x	(x.x)	x	x		
Probable Symptomatic	x			xx	(x.x)	x

N: Number of subjects; %: Percentage of subjects; E: Number of events

Analysis Dataset Metadata

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	35	DZA	Drug B
2	XYZ	000008	ASSYMPC	Asymptomatic Hypoglycemia (cumulative frequency count)	Week 1	3	72	F	35	DZA	Drug B
3	XYZ	000008	ASSYMP	Asymptomatic Hypoglycemia (frequency count)	Week 2	1	72	F	35	DZA	Drug B
4	XYZ	000008	ASSYMPC	Asymptomatic Hypoglycemia (cumulative frequency count)	Week 2	4	72	F	35	DZA	Drug B
5	XYZ	000008	ASSYMP	Asymptomatic Hypoglycemia (frequency count)	Week 3	0	72	F	35	DZA	Drug B
6	XYZ	000008	ASSYMPC	Asymptomatic Hypoglycemia (cumulative frequency count)	Week 3	4	72	F	35	DZA	Drug B
7	XYZ	000008	ASSYMP	Asymptomatic Hypoglycemia (frequency count)	Week 4	1	72	F	35	DZA	Drug B
8	XYZ	000008	ASSYMPC	Asymptomatic Hypoglycemia (cumulative frequency count)	Week 4	5	72	F	35	DZA	Drug B
10	XYZ	000008	ASSYMPC	Asymptomatic Hypoglycemia (cumulative frequency count)	End of Treatment	7	72	F	35	DZA	Drug B
...	...	...	...	...	...	...	...	...	...	...	...
20	XYZ	000008	DOCSEVC	Documented Symptomatic or Severe Hypoglycemia (cumulative frequency count)	End of Treatment	17	72	F	35	DZA	Drug B

Table 3.3.2: ADHYSUM Dataset Metadata

Dataset Name	Dataset Description	Dataset Location	Dataset Structure	Keys	Class	Documentation
ADHYSUM	Hypoglycemic Episodes Summary Data	ADHYSUM.xpt	One record per subject per analysis visit per parameter	STUDYID, USUBJID, AVISIT, PARAMCD	BDS	ADHYSUM.SAS/SAP

Table 3.3.3: ADHYSUM Variable Metadata

Variable Name	Variable Label	Type	Length/Display Format	Codelist/Controlled Terms	Source/Derivation/Comment
STUDYID	Study Identifier	text	$12		ADSL.STUDYID
USUBJID	Unique Subject Identifier	text	$20		ADSL.USUBJID
PARAMCD	Parameter Code	text	$8		See parameter value metadata. *Note that the tables below do not present all possible values for PARAMCD but only those that correspond to the data display.*
PARAM	Parameter	text	$80		See parameter value metadata. *Note that the tables below do not present all possible values for PARAM but only those that correspond to the data display.*
AVISIT	Analysis Visit	text	$13	Week -1; Week 0; Week 1; Week N; End of Treatment	Refer to Section X.X of the SAP for windowing and imputation algorithms based on ADHYPO.ADY. End-of-treatment is defined as the last week during which the subject is on treatment.
AVAL	Analysis Value	integer	8		See parameter value metadata.
TRTDURD	Total Treatment Duration (Days)	integer	8		ADSL.TRTDURD
SEX	Sex	text	$1		ADSL.SEX
AGE	Age	integer	8		ADSL.AGE
COUNTRY	Country	text	$3		ADSL.COUNTRY
TRTA	Actual Treatment	text	$32		ADSL.TRT01A

cdisc

Tabulation Data

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	

suppce.xpt

Row	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL
1	XYZ	CE	XYZ-001-001	CESEQ	1	WHEOCC	When did the hypoglycemic event occur?	BETWEEN BEDTIME AND WAKING
2	XYZ	CE	XYZ-001-001	CESEQ	8	WHENOCC	When did the hypoglycemic event occur?	BETWEEN BEDTIME AND WAKING

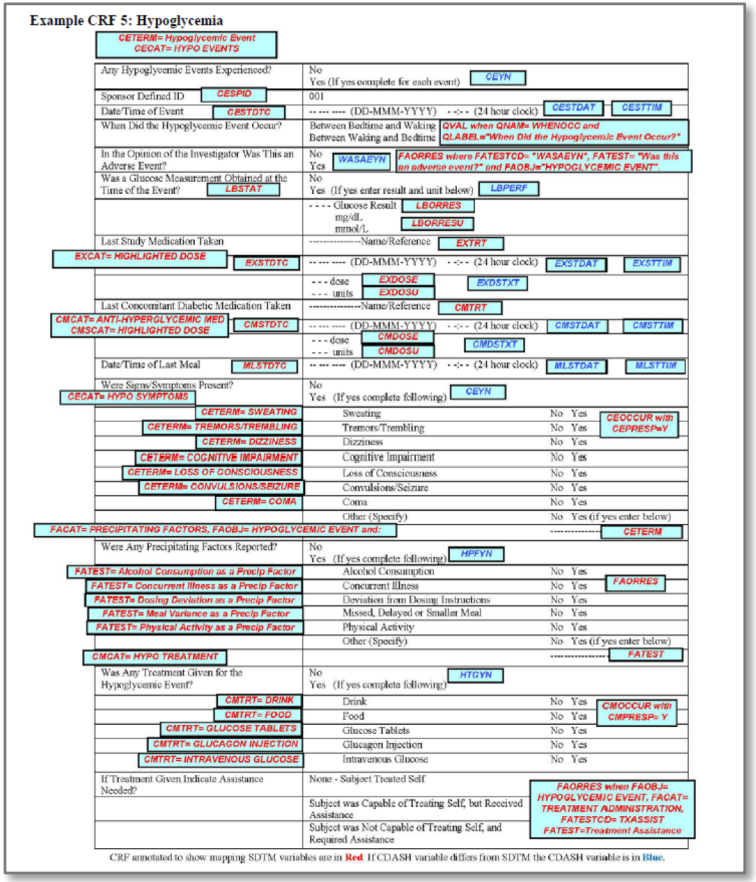
lb.xpt

Row	STUDYID	DOMAIN	USUBJID	SPDEVID	LBSEQ	LBTESTCD	LBTEST	LBORRES	LBORRESU	LBSTRESC	LBSTRESN	LBSTRESU
1	XYZ	LB	XYZ-001-001	GLUCOSE METER	1	GLUC	GLUCOSE	60	mg/dL	3.3	3.3	mmol/l
2	XYZ	LB	XYZ-001-001	GLUCOSE METER	2	GLUC	GLUCOSE	65	mg/dL	3.6	3.6	mmol/l

ml.xpt

Row	STUDYID	DOMAIN	USUBJID	MLSEQ	MLTRT	MLSTDTC	RELMIDS	MIDS	MIDSDTC
1	XYZ	ML	XYZ-001-001	1	MEAL	2013-08-31T20:00	LAST MEAL PRIOR TO	HYPO 1	2013-09-01T11:00
2	XYZ	ML	XYZ-001-001	2	MEAL	2013-09-23T22:30	LAST MEAL PRIOR TO	HYPO 2	2013-09-24T08:48

Collection Metadata



Example CRF 5: Hypoglycemia

CETERM= Hypoglycemic Event
CECAT= HYPO EVENTS

Any Hypoglycemic Events Experienced?	No Yes (If yes complete for each event) CEYN	
Sponsor Defined ID CESPID	001	
Date/Time of Event CESTDTC	-- --- ---- (DD-MMM-YYYY) - -:- - (24 hour clock)	CESTDAT CESTTIM
When Did the Hypoglycemic Event Occur?	Between Bedtime and Waking Between Waking and Bedtime	QVAL when QNAM= WHENOCC and QLABEL="When Did the Hypoglycemic Event Occur?"
In the Opinion of the Investigator Was This an Adverse Event?	No Yes WASAEYN	FAORRES where FATESTCD= "WASAEYN", FATEST= "Was this an adverse event?" and FAOBJ="HYPOGLYCEMIC EVENT"
Was a Glucose Measurement Obtained at the Time of the Event? LBSTAT	No Yes (If yes enter result and unit below) LBPERF	
	- - - - Glucose Result LBORRES mg/dL mmol/L LBORRESU	
Last Study Medication Taken	----------Name/Reference EXTRT	
EXCAT= HIGHLIGHTED DOSE EXSTDTC	-- --- ---- (DD-MMM-YYYY) - -:- - (24 hour clock)	EXSTDAT EXSTTIM
	- - - dose EXDOSE - - - units EXDOSU	EXDSTXT
Last Concomitant Diabetic Medication Taken	----------Name/Reference CMTRT	
CMCAT= ANTI-HYPERGLYCEMIC MED CMSCAT= HIGHLIGHTED DOSE CMSTDTC	-- --- ---- (DD-MMM-YYYY) - -:- - (24 hour clock)	CMSTDAT CMSTTIM
	- - - dose CMDOSE - - - units CMDOSU	CMDSTXT
Date/Time of Last Meal MLSTDTC	-- --- ---- (DD-MMM-YYYY) - -:- - (24 hour clock)	MLSTDAT MLSTTIM
Were Signs/Symptoms Present? CECAT= HYPO SYMPTOMS	No Yes (If yes complete following) CEYN	
CETERM= SWEATING	Sweating	No Yes
CETERM= TREMORS/TREMBLING	Tremors/Trembling	No Yes
CETERM= DIZZINESS	Dizziness	No Yes
CETERM= COGNITIVE IMPAIRMENT	Cognitive Impairment	No Yes
CETERM= LOSS OF CONSCIOUSNESS	Loss of Consciousness	No Yes
CETERM= CONVULSIONS/SEIZURE	Convulsions/Seizure	No Yes
CETERM= COMA	Coma	No Yes
	Other (Specify)	No Yes (if yes enter below)
FACAT= PRECIPITATING FACTORS, FAOBJ= HYPOGLYCEMIC EVENT and:	---------- CETERM	
Were Any Precipitating Factors Reported?	No Yes (If yes complete following) HPFYN	
FATEST= Alcohol Consumption as a Precip Factor	Alcohol Consumption	No Yes
FATEST= Concurrent Illness as a Precip Factor	Concurrent Illness	No Yes
FATEST= Dosing Deviation as a Precip Factor	Deviation from Dosing Instructions	No Yes
FATEST= Meal Variance as a Precip Factor	Missed, Delayed or Smaller Meal	No Yes
FATEST= Physical Activity as a Precip Factor	Physical Activity	No Yes
	Other (Specify)	No Yes (if yes enter below)
CMCAT= HYPO TREATMENT	---------- FATEST	
Was Any Treatment Given for the Hypoglycemic Event?	No Yes (If yes complete following) HTGYN	
CMTRT= DRINK	Drink	No Yes
CMTRT= FOOD	Food	No Yes
CMTRT= GLUCOSE TABLETS	Glucose Tablets	No Yes
CMTRT= GLUCAGON INJECTION	Glucagon Injection	No Yes
CMTRT= INTRAVENOUS GLUCOSE	Intravenous Glucose	No Yes
If Treatment Given Indicate Assistance Needed?	None - Subject Treated Self	
	Subject was Capable of Treating Self, but Received Assistance	
	Subject was Not Capable of Treating Self, and Required Assistance	

CMOCCUR with CMPRESP= Y

FAORRES when FAOBJ= HYPOGLYCEMIC EVENT, FACAT= TREATMENT ADMINISTRATION, FATESTCD= TXASSIST FATEST=Treatment Assistance

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

Standards Selection
(for the “360 Test Study”)

For the “360 Test Study” we will, for these standards:
- Develop standard concepts
- Store concepts in prototype CDISC Library
- Pick & select standards from Library (use case 1)
- Configure study spec & create artifacts (use case 2)
- Populate study artifacts with data (use case 3)

- 1 or 2 statistical end points
  - Analysis of Glycated Hemoglobin
  - Summary of Hypoglycemic episodes
- ~3-4 ADaM datasets
  - ADSL(**Subject-Level Analysis Data (ADSL)**)
  - Hemoglobin A1C Analysis Dataset (**HbA1c Analysis Dataset**)
  - Hypoglycemic Episodes Analysis Dataset (**Hypoglycemic Episodes Analysis Dataset**)
  - Hypoglycemic Episodes Summary Dataset (**Hypoglycemic Episodes Summary Dataset**)
- ~7-8 SDTM datasets
  - DM (Demographics, to support standard variables in ADSL)
  - VS (Vital Signs, for height and weight in ADSL)
  - CM (Concomitant Meds, to support stratification by background treatment, and for treatments of hypoglycemic events)
  - LB (for Hemoglobin A1C data)
  - CE and FACE (for data on hypoglycemic events)
  - EX, ML (for data about meals and study treatments relative to hypoglycemic events)
  - Trial Design datasets (for arms, visit schedule, definition of hypoglycemic events as disease milestones)
- ~15 CDASH CRFs
  - CDASH CRFs needed to support SDTM datasets above. One CRF will support collection of data about hypoglycemic events that will be mapped to multiple SDTM domains.

Workstream 1 and 3

Concept Modeling

(Workstream 3 is a sub-stream of Workstream 1)

WS1: Concept Modeling

Objectives

- Produce biomedical and analysis concepts for 360 test study
- Ensure sufficient linked metadata is available to enable automation of UC1, UC2, UC3
- Develop a robust concept-based standards model
- Develop process, requirements, and effort estimate for scaled-up standard concepts development process

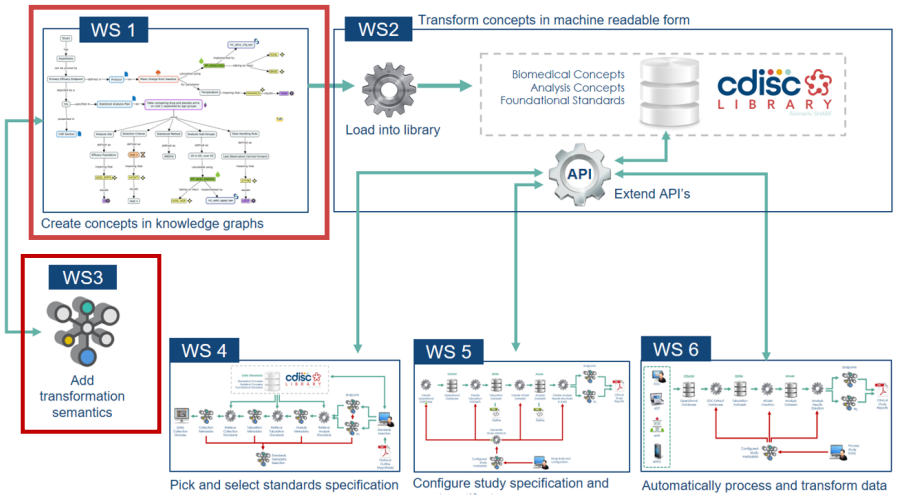
Value Proposition

- A standards metadata model that:
  - Expresses the full meaning of clinical data
  - Provides a single, unified implementation of the CDISC Foundational Standards
  - Enables metadata-driven automation of processes across the clinical research data lifecycle
- Consistent operational processes for concept-based standards development

Dependencies

- WS2 Concepts in CDISC Library [downstream]
- WS4 Standards Specification (UC1) [downstream]
- WS5 Study Specification (UC2) [downstream]
- WS6 Transform Data (UC3) [downstream]

Workstream Snapshot

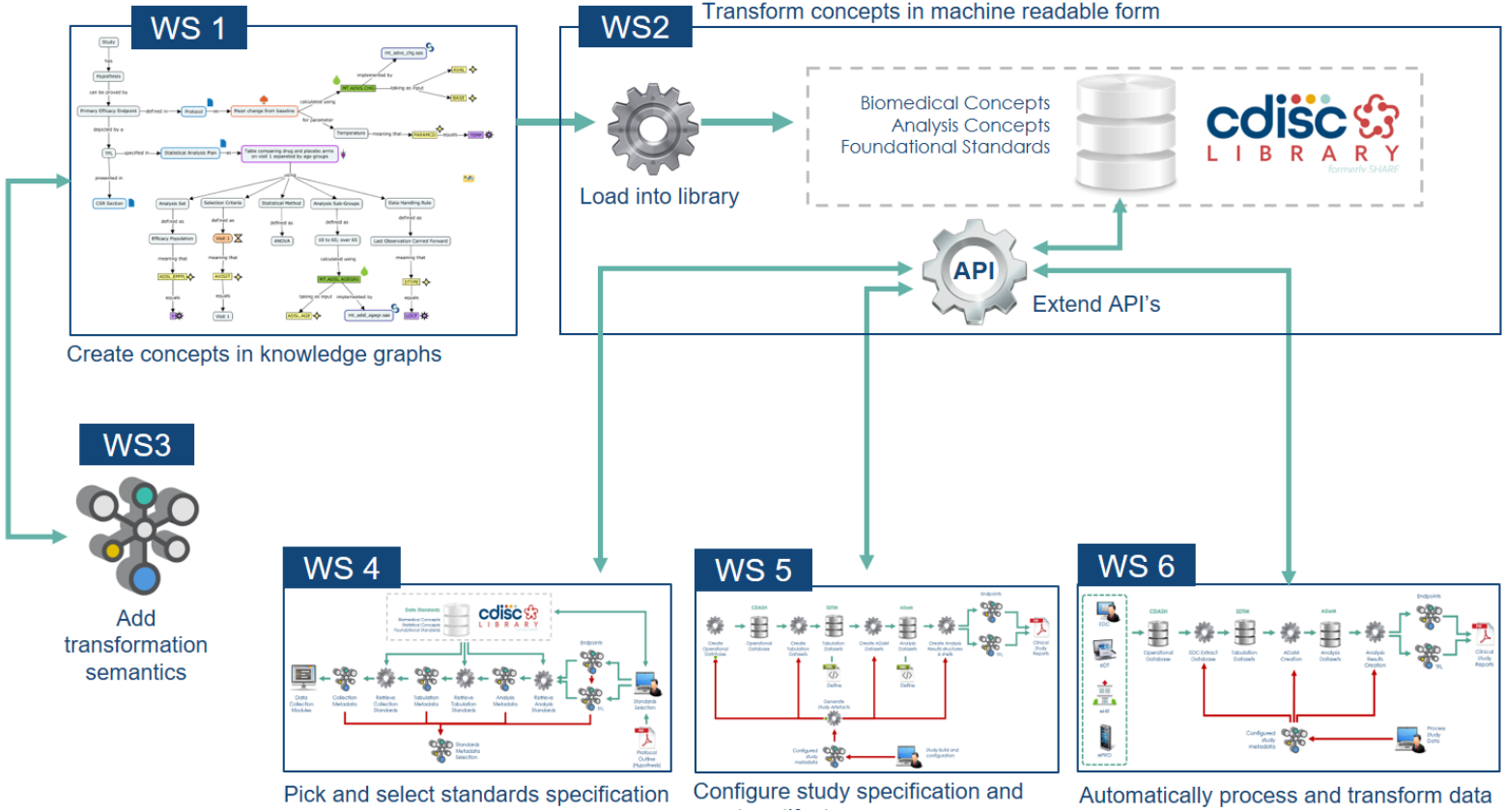


Challenges

- Tool limitation to develop concepts (i.e. CMAP)
- Manual and prone to human error and inconsistencies
- Learning curve to develop biomedical and analysis concepts
- Developing desired level of requirements to achieve ETL automation
- Transformation into machine readable content

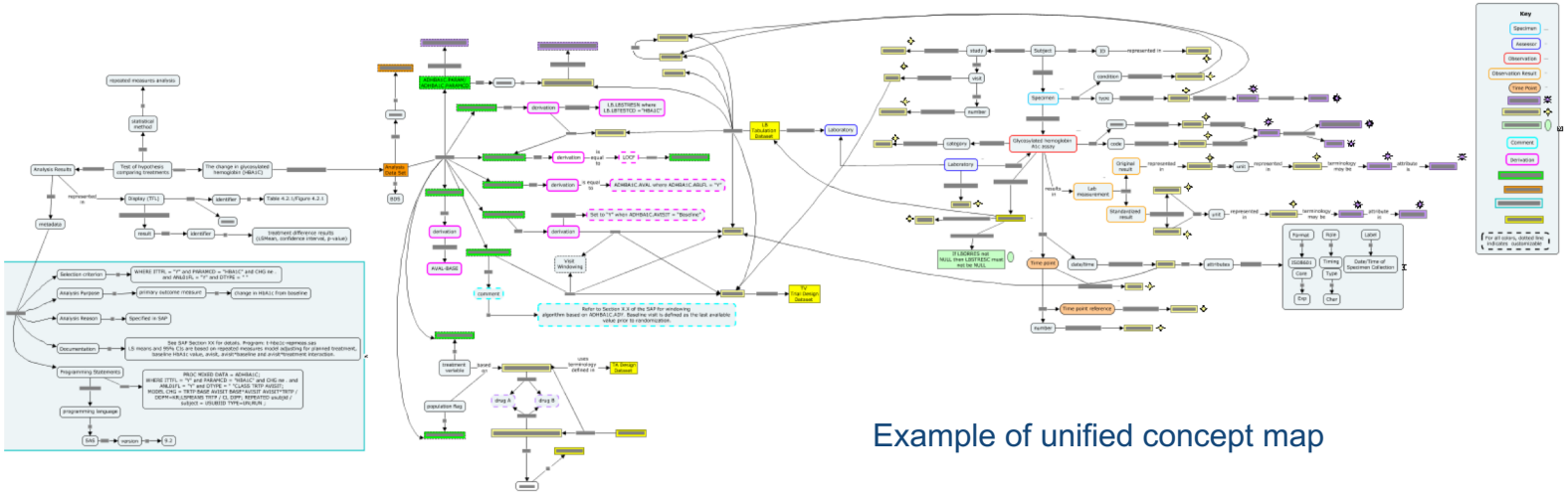
CDISC 360 WS1 details can be found here

Workstreams Overview



WS1 Inputs

- 360 Cmap cloud – has initial mapping of one Diabetes TAUG endpoint:
  - Unified concept map (analysis and biomedical concept combo)
  - Split concept map
    - Analysis results map
    - Analysis parameter map
    - Biomedical concept map to SDTM
    - Biomedical concept map to Data Collection



Example of unified concept map

WS1 Inputs

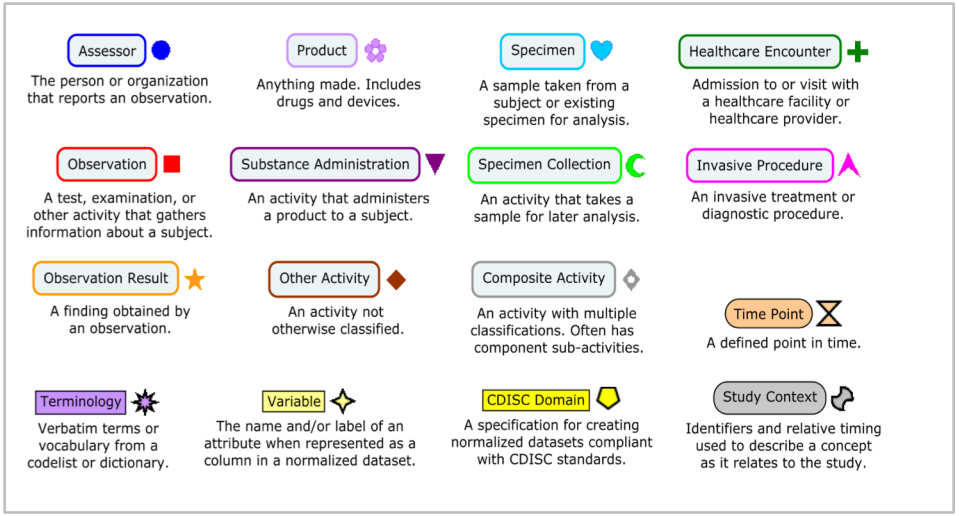
- Cmap map legend
  - Unique map symbols and colors
- Concept map predicates
  - Standard language for predicates that link elements on a map
- Concept Backlog
  - The (many) issues to be addressed to complete the mapping

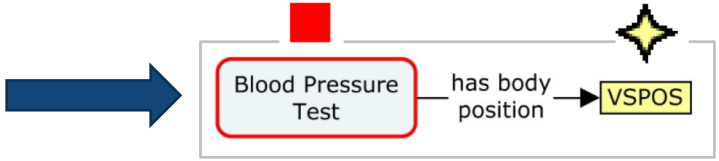
WS1 Inputs - Concept Maps in Progress

- 1 or 2 statistical end points
  - Analysis of Glycated Hemoglobin – in progress
  - Summary of Hypoglycemic episodes
- ~3-4 ADaM datasets
  - ADSL
  - Hemoglobin A1C Analysis Dataset – in progress
  - Hypoglycemic Episodes Analysis Dataset
  - Hypoglycemic Episodes Summary Dataset
- ~7-8 SDTM datasets
  - DM (Demographics, so support standard variables in ADSL)
  - VS (Vital Signs, for height and weight in ADSL)
  - CM (Concomitant Meds, to support stratification by background treatment, and for treatments of hypoglycemic events)
  - LB (for Hemoglobin A1C data) – in progress
  - CE and FACE (for data on hypoglycemic events)
  - EX, ML (for data about meals and study treatments relative to hypoglycemic events)
  - Trial Design datasets (for arms, visit schedule, definition of hypoglycemic events as disease milestones)
- ~15 CDASH CRFs
  - CDASH CRFs needed to support SDTM datasets above. One CRF will support collection of data about hypoglycemic events that will be mapped to multiple SDTM domains.

WS1 Activities

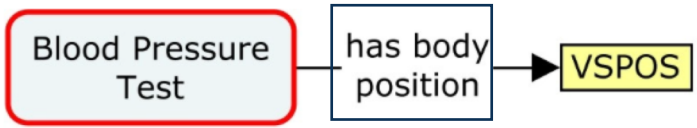
- Create legend definition and apply it to the concept map

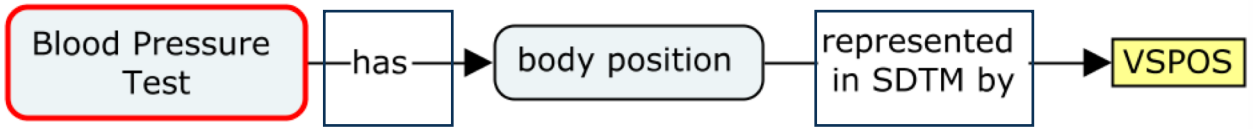




WS1 Activities

- Create predicates definition – still exploring options





WS1 Activities

- Create element definition

USUBJID: A sequence of characters used to uniquely identify a subject across all studies for all applications or submissions involving the product.

Same meaning across CDASH, SDTM, ADaM

WS1 Activities

- Identify concept compliance checks (before loading into Library)
  - Checks to be provided by WS2
  - Checks
    - Legend applied correctly
    - Correct predicate usage
    - Orphan links
    - Map-to-map links
- Co-create user stories of UC1-UC3 to inform concept content
- Add semantic transformation metadata layer to enable UC1-UC3 automation
- Create concept maps for full data standards scope
- Apply feedback from downstream workstreams

WS1 Deliverables

- Concept maps for full data standards scope
- Robust concept-based standards model
- Concept templates
- Approach for scaled-up development & curation of concept-based standards
- Requirements for scaled-up standards development tool set
- Effort estimate for development of ***all*** scaled-up concept-based standards
- Prioritization for standards development
- Links to LOINC, UCUM, and HL7 FHIR

Workstream 2

Concepts in CDISC Library

WS 2: Concepts in CDISC Library

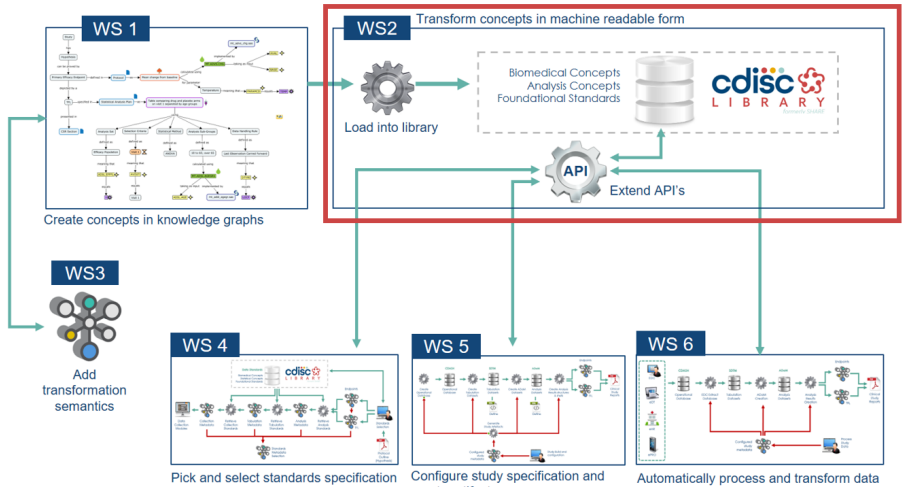
Objectives

- Transformation of biomedical and analysis concepts into machine readable metadata
- Extended API's to extract multifunctional metadata
- Supporting standards development activities

Value Proposition

- Machine readable linked data biomedical and analysis concepts
- Tools to support automated application of biomedical and analysis concepts

Workstream Snapshot



Dependencies

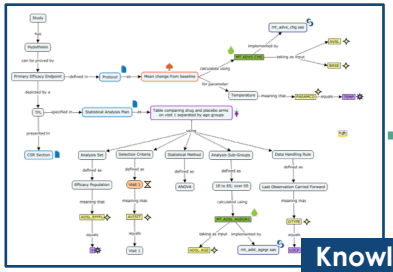
- WS1 Concept Modeling [upstream]
- WS4 Standards Specification (UC1) [downstream]
- WS5 Study Specification (UC2) [downstream]
- WS6 Transform Data (UC3) [downstream]

Challenges

- Automating the transformation of concepts into CDISC Library
- Providing metadata to drive WS 4-6

CDISC 360 WS2 details can be found here

CDISC Library API extension



Knowledge graph

LOAD

Biomedical Concepts
Analysis Concepts
Foundational Standards

cdisc LIBRARY
formerly SHARE

Call

API

Extend Current Set of APIs

Return

API metadata Return

PAA	BBB	X-3	Y+H	DEF	1+2	…

- Dataset metadata
- Concept metadata
- Concept relations
- Controlled terminology
- Process metadata
- Configuration Defaults and options

cdisc

WS 2 Inputs

- Biomedical and Analysis Concepts (BACs) in a machine-readable form (WS-1)
- CMAP conformance rules specification (WS-1)
- Existing 360 user stories
- Current CDISC Library API specification
- Relevant aspects of the current CDISC Library model

WS 2 Activities

- Contribute to authoring the CMAP / BAC user stories
- Support WS-1 to develop BACs that support downstream activities
- Develop a process for working with WS-1 to ingest completed CMAPs
- Develop a process for delivering WS-2 work items to downstream work streams
- Design and implement a sandbox API specification with a supporting data store
- Implement CMAP conformance rule application
- Develop a sandbox API to provide inputs to downstream work streams
- Design the CDISC Library API extensions and document them using OpenAPI
- Design the CDISC Library model extensions and represent them in RDF/OWL
- Develop automated tests for automation deliverables

WS 2 Deliverables

- Application(s) to transform WS-1 biomedical and analysis concepts (BACs) into load files to be ingested by the Sandbox and CDISC Library API data stores
- Conformance rules and supporting automation to ensure the WS-1 BACs conform to the published model and rules
- Sandbox API that provides access to BAC metadata linked to the foundational standards and controlled terminology in the CDISC Library
- Model for representing BACs in the CDISC Library
- CDISC Library API specification
- Test suite for automation deliverables

Workstream 4

Standards Specification – Use Case 1

WS 4: Standards Specification (UC1)

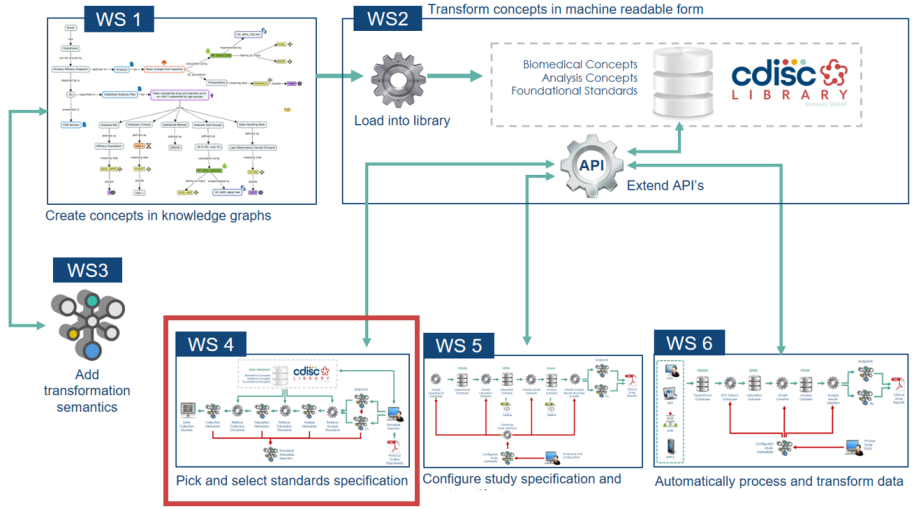
Objectives

- Develop specification for APIs to retrieve specification metadata
- Demonstrate pick and select capability
- Ensure API output is complete
- Combine all metadata in specification pool suitable for WS5

Value Proposition

- Specification for APIs needed in CDISC Library to surface concept-based standards metadata to industry
- Demonstration scripts, for industry to retrieve metadata via API
- Demonstration assembled standards metadata, for industry to develop automated process to configure an E2E study specification and create study artifacts

Workstream Snapshot



Dependencies

- WS1 Concept Modeling [upstream]
- WS2 Concepts in CDISC Library [upstream]
- WS5 Study Specification (UC2) [downstream]
- WS6 Transform Data (UC3) [downstream]

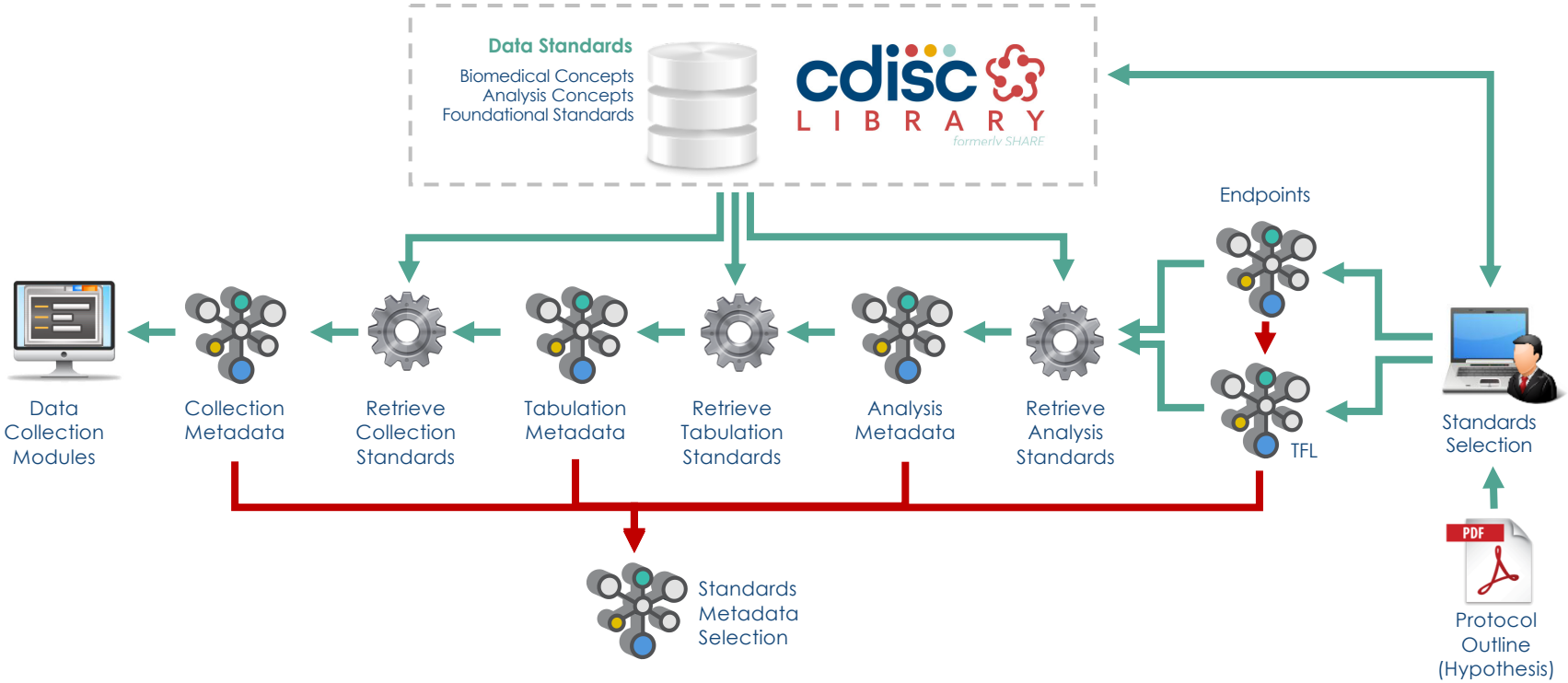
Challenges

- Navigating content to determine what to assemble
- Process to assemble associated selected standards metadata
- Technological means to extract via program script
- Storage and management of assembled selected standards metadata

CDISC 360 WS4 details can be found here

Use Case 1 : End to Start specification

Selecting standards concepts and linked metadata needed for a study



cdisc

WS4 Inputs

Today

- CDISC Library content
  - Foundational Standards
- CDISC Library API overview (as-is)
- Diabetes TAUG standards (outside CDISC Library)

To Come

- 360 prototype CDISC Library
  - Content: standard concepts developed by WS1, for Diabetes TAUG standards in scope
- 360 APIs for prototype CDISC Library

WS4 Activities

- Collaborate with WS5 and WS6 to:
  - Develop user stories for workstreams 4, 5, and 6
  - Create CDISC Library API extension specifications
    - Based upon user stories
    - Based upon needs of WS5 and WS6
- Develop scripts to select standard concepts for study endpoints and associated analysis outputs, via Library APIs
- Develop scripts to navigate standard concepts to pick and retrieve metadata for all involved standard artifacts, via Library APIs
- Develop structure and storage of a metadata specification pool suitable for use by WS5 to execute Use Case 2
  - e.g., How & where is this content stored shared with WS5 and WS6? In a “Study MDR”
- Determine technology for scripts; for scripts execution; for stored metadata specification pool (in collaboration with WS5 and WS6)
  - e.g., do we use Excel to store the configured study specifications and share among WS4, Ws5, WS6

WS4 Deliverables

- Demonstration of automated pick and select capability from CDISC Library
  - For standard endpoint and analysis output concepts

- Demonstration of automated navigation of standard endpoint and analysis output concepts
  - To pick and retrieve metadata for all involved standard artifacts, from CDISC Library

- Demonstration of automated assembly of metadata specification pool suitable for use by WS5 to execute Use Case 2

Workstream 5

Study Specification and Artifact Creation – Use Case 2

WS5: Study Specification (UC2)

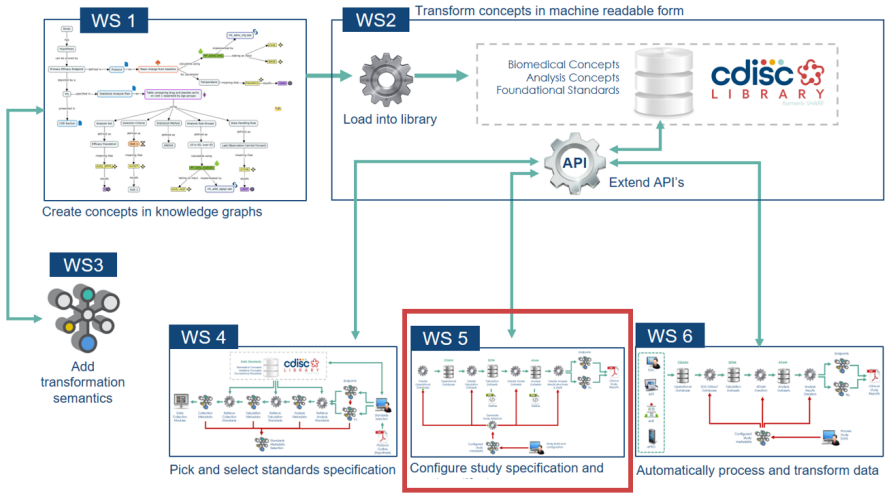
Objectives

- Study-specific configuration of standards metadata
- Instantiate metadata on a study level (datasets, Define-XML, analysis shells)
- Demonstrate study build process (includes trial design)

Value Proposition

- Complete metadata spectrum from collection-to-TFL, to enable industry to auto process study data
- Demonstration scripts, for industry to auto prepare the complete metadata spectrum from selected standard concepts
- Quality and efficient driven process for study setup

Workstream Snapshot



Dependencies

- WS1 Concept Modeling [upstream]
- WS2 Concepts in CDISC Library [upstream]
- WS4 Standards Specification (UC1) [upstream]
- WS6 Transform Data (UC3) [downstream]

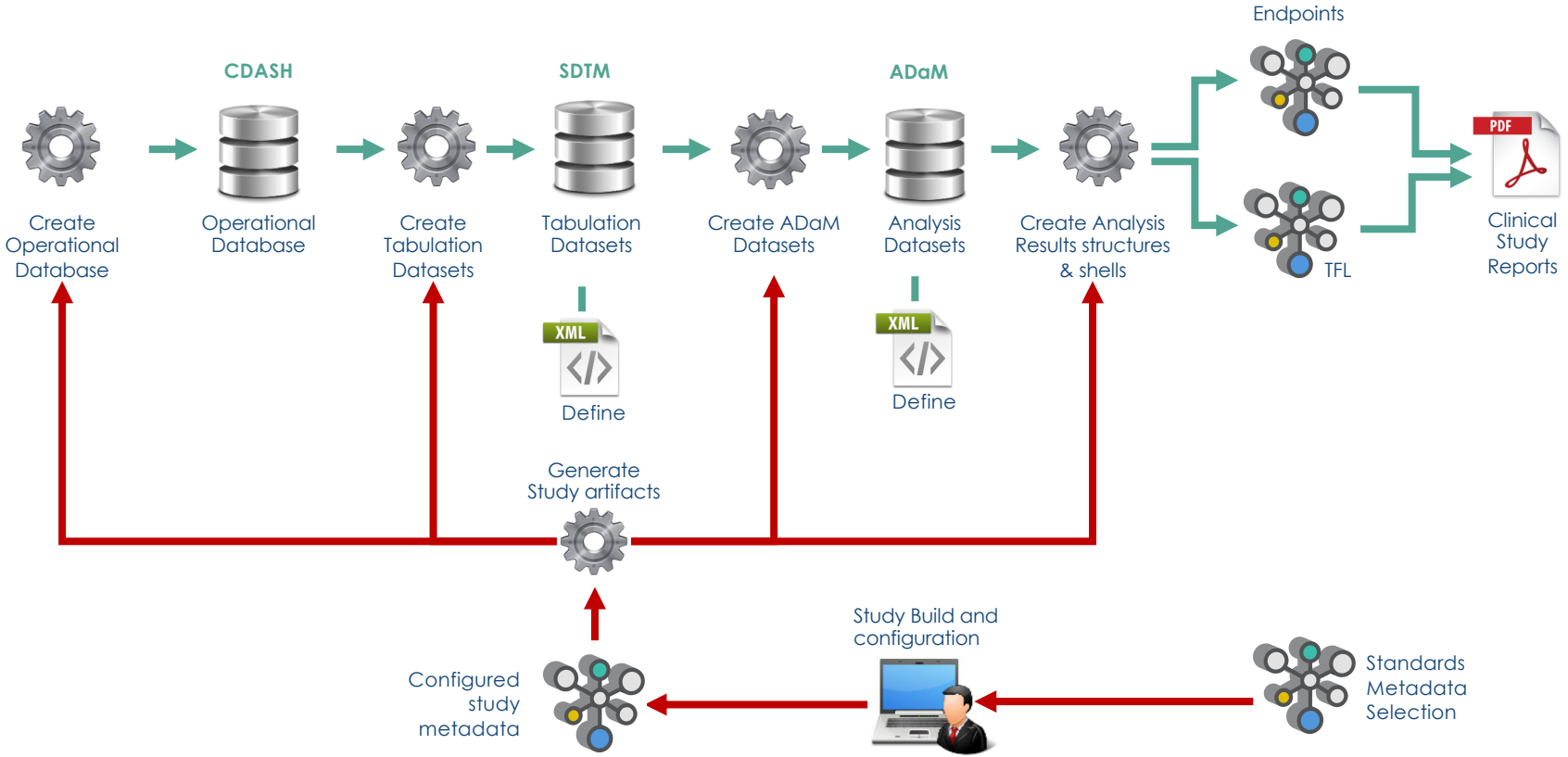
Challenges

- Configuring a study specification from concept-based standards metadata
- Technological means to build a study specification
- Human error when input is required
- Managing a built study specification that incurs changes

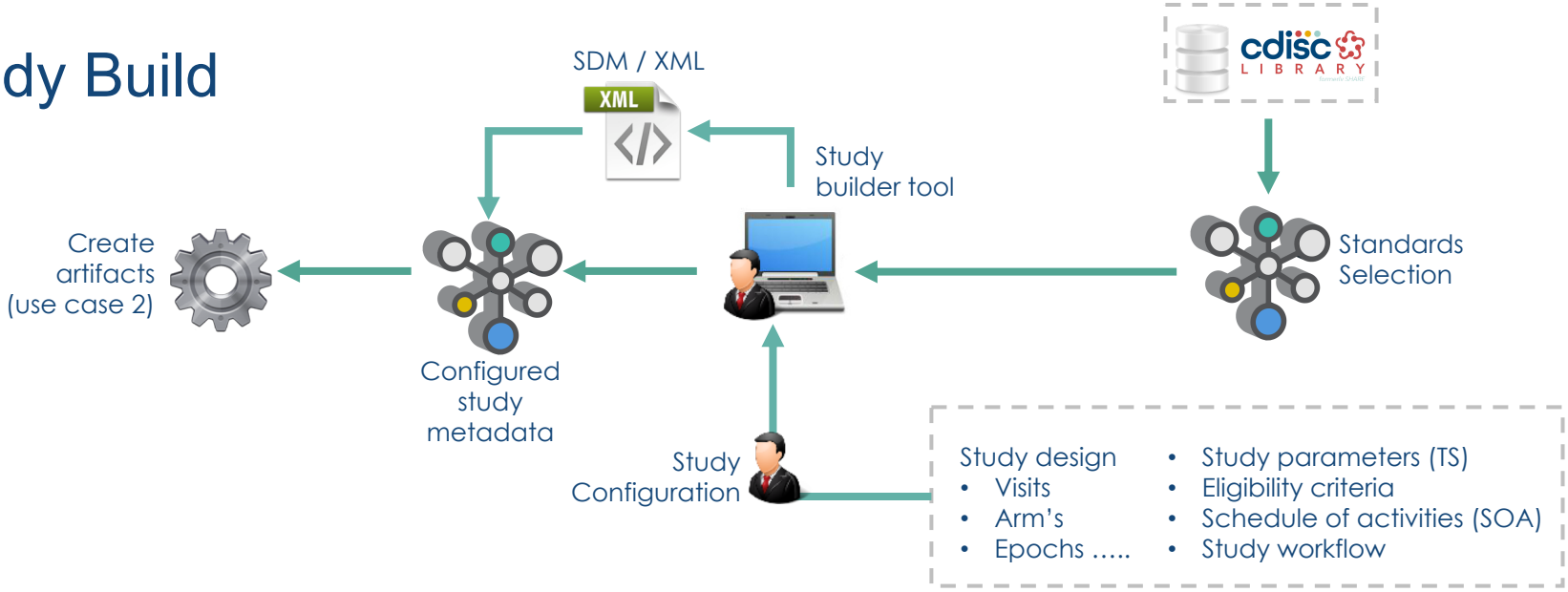
Currently Out of Scope: (1) Creation of study protocol, SAP, CSR and (2) Automated business rules (validation).

Use Case 2 : Start to End Study Metadata

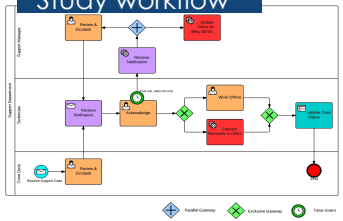
Adding study design, concept configuration & generate artifacts



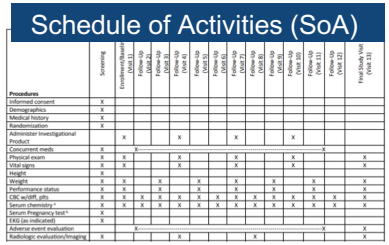
Study Build



Study workflow

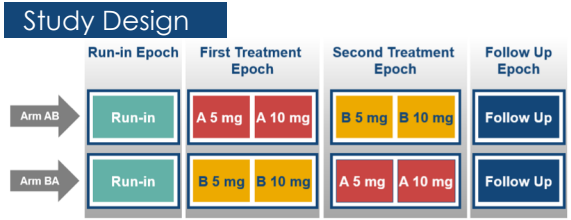


Schedule of Activities (SoA)



Procedures	Screening	Enrollment/Baseline (Visit 1)	Follow-Up (Week 2)	Follow-Up (Week 4)	Follow-Up (Week 6)	Follow-Up (Week 8)	Follow-Up (Week 10)	Follow-Up (Week 12)	Follow-Up (Week 14)	Follow-Up (Week 16)	Follow-Up (Week 18)	Follow-Up (Week 20)	Follow-Up (Week 22)	Follow-Up (Week 24)	Final Study Visit (Week 26)
Informed consent	X														
Demographics	X														
Medical history	X														
Randomization	X														
Administer Investigational Product		X				X				X				X	
Concurrent meds	X		X												X
Physical exam	X	X				X				X				X	
Vital signs	X	X				X				X				X	
Height	X														X
Weight	X	X				X				X				X	
Performance status	X	X				X				X				X	
CBC w/diff, plts	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Serum chemistry*	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Serum Pregnancy test*	X														
EKG (as indicated)	X														
Adverse event evaluation		X													X
Radiologic evaluation/Imaging	X					X				X					X

Study Design



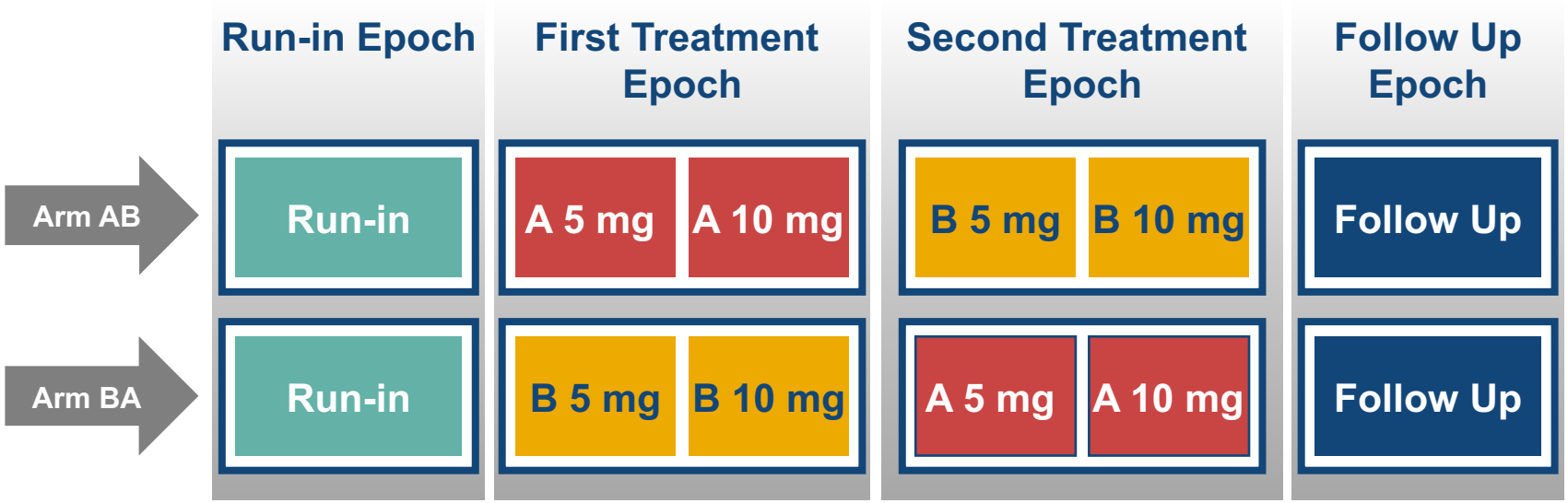
Study Parameters (TS)

XYZ	TS	1		ALSOIN	Existing Treatments	Y		C49488	CDISC	2011-06-10
XYZ	TS	1		AGEMAX	Planned Maximum Age of Subjects	P70Y			ISO 8601	
XYZ	TS	1		AGEMIN	Planned Minimum Age of Subjects	P18M			ISO 8601	
XYZ	TS	1		LENGTH	Planned Trial Length	P3M			ISO 8601	
XYZ	TS	1		PLANSUB	Planned Number of Subjects	300				
XYZ	TS	1		RANDOM	Trial is Randomized	Y		C49488	CDISC	2011-06-10
XYZ	TS	1		SEXPOP	Sex of Participants	BOTH		C49636	CDISC	2011-06-10
XYZ	TS	1		STOPRULE	Study Stop Rules	INTERIM ANALYSES FOR FUTILITY				
XYZ	TS	1		TBLIND	Trial Blinding Schema	DOUBLE BLIND		C15228	CDISC	2011-06-10
XYZ	TS	1		TCNTRL	Control Type	PLACEBO		C49648	CDISC	2011-06-10
XYZ	TS	1		TDIGRP	Diagnosis Group	Neurofibromatosis Syndrome (Disorder)		19133005	SNOMED	
XYZ	TS	1		TINDTP	Trial Indication Type	TREATMENT		C49656	CDISC	2011-06-10

Study Parameters (TS)

STUDYID	DOMAIN	TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVAL	TSVALNF	TSVALCD	TSVCDREF	TSVCDVER
XYZ	TS	1		ADDON	Added on to Existing Treatments	Y		C49488	CDISC	2011-06-10
XYZ	TS	1		AGEMAX	Planned Maximum Age of Subjects	P70Y			ISO 8601	
XYZ	TS	1		AGEMIN	Planned Minimum Age of Subjects	P18M			ISO 8601	
XYZ	TS	1		LENGTH	Planned Trial Length	P3M			ISO 8601	
XYZ	TS	1		PLANSUB	Planned Number of Subjects	300				
XYZ	TS	1		RANDOM	Trial is Randomized	Y		C49488	CDISC	2011-06-10
XYZ	TS	1		SEXPOP	Sex of Participants	BOTH		C49636	CDISC	2011-06-10
XYZ	TS	1		STOPRULE	Study Stop Rules	INTERIM ANALYSIS FOR FUTILITY				
XYZ	TS	1		TBLIND	Trial Blinding Schema	DOUBLE BLIND		C15228	CDISC	2011-06-10
XYZ	TS	1		TCNTRL	Control Type	PLACEBO		C49648	CDISC	2011-06-10
XYZ	TS	1		TDIGRP	Diagnosis Group	Neurofibromatosis Syndrome (Disorder)		19133005	SNOMED	
XYZ	TS	1		TINDTP	Trial Indication Type	TREATMENT		C49656	CDISC	2011-06-10

Study Design



Schedule of Activities (SoA)

Procedures	Screening	Enrollment/Baseline (Visit 1)	Follow-Up (Visit 2)	Follow-Up (Visit 3)	Follow-Up (Visit 4)	Follow-Up (Visit 5)	Follow-Up (Visit 6)	Follow-Up (Visit 7)	Follow-Up (Visit 8)	Follow-Up (Visit 9)	Follow-Up (Visit 10)	Follow-Up (Visit 11)	Follow-Up (Visit 12)	Final Study Visit (Visit 13)
Informed consent	X													
Demographics	X													
Medical history	X													
Randomization	X													
Administer Investigational Product		X			X			X			X			
Concurrent meds	X	X----------											-X	
Physical exam	X	X			X			X			X			X
Vital signs	X	X			X			X			X			X
Height	X													
Weight	X	X		X		X		X		X		X		X
Performance status	X	X		X		X		X		X		X		X
CBC w/diff, plts	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Serum chemistry [a]	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Serum Pregnancy test [b]	X													
EKG (as indicated)	X													
Adverse event evaluation		X----------											-X	X
Radiologic evaluation/Imaging	X				X				X					X

cdisc

WS5 Inputs

Today

- Diabetes TAUG

To Come

- Study configuration data (e.g., Excel sheet to be created by WS5):
  - Study design (arms, epochs, visits)
  - Study parameters (TS)
  - Eligibility criteria
  - Schedule of Activities (SOA)
  - Study workflow
- Standards metadata selection from WS4
  - Metadata should identify study-specific implementation choices
    - E.g., Population selection: (1) 18-65; (2) 18-40, 41-65
    - Metadata should identify default implementation choice

WS5 Activities (1)

- Determine and prepare study configuration for “360 Test Study” (e.g., create Excel workbook with the following):
  - Study design (arms, epochs, visits)
  - Study parameters (TS)
  - Eligibility criteria
  - Schedule of Activities (SOA)
  - Study workflow
- Collaborate with WS6 to acquire and/or develop test data

WS5 Activities (2)

Activities for “validation” instance of “360 Test Study”

Based upon determined study configuration

This is the “double programming”, the “validation” instance of the target artifacts

Any manual/programming means is acceptable (this is not the new automated 360 tooling)

- Pre-create all target artifacts (i.e., datasets, TLFs)
  - Operational database definition (ODM)
  - Tabulation database shells (SDTM datasets)
  - Tabulation metadata (Define.xml)
  - Analysis database shells (ADaM datasets)
  - Analysis metadata (Define.xml for ADaM + Analysis Results Metadata)
  - TLF output shells
- Ensure compliance of target artifacts with Pinnacle 21

WS5 Activities (3)

Activities for “production” instance of “360 Test Study”

These activities are accomplished using the new automated 360 tooling developed during WS5, with the Standards Metadata Selection from WS4 as input

- Store study configuration in SDM/XML
- Enrich standards metadata selection (WS4) to complete study-specific metadata instantiation
- Study instantiation; automate creation of:
  - Operational database (ODM)
  - Tabulation database (SDTM datasets)
  - Tabulation metadata (Define.xml)
  - Analysis database (ADaM datasets)
  - Analysis metadata (Define.xml for ADaM + Analysis Results Metadata)
  - TLF output shells
- Ensure compliance of target artifacts with Pinnacle 21
- Determine technology for scripts; for scripts execution; for stored study specification
  - e.g., do we use Excel to store the configured study specifications and share among WS4, Ws5, WS6

WS5 Deliverables

- Store study design parameters in SDM/XML
  - Verify SDM/XML for completeness
- Complete set of configured study metadata, ready to drive instantiation
- End-to-end artifact creation (push of a button)
- Consider creation of SAP
  - What is needed beyond what we now have to produce the SAP?

Workstream 6

Start to End Data Processing – Use Case 3

WS 6: Transform Data (UC3)

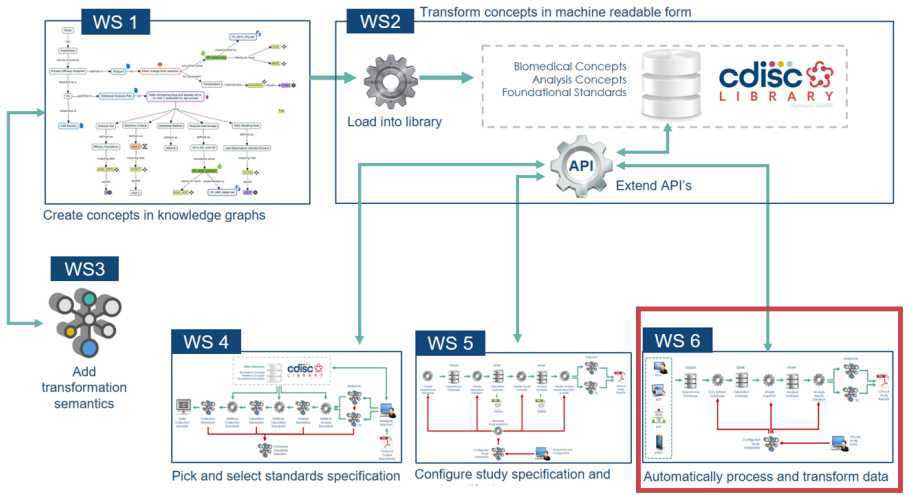
Objectives

- Auto-generate and execute programs to process study data from collection to analysis
- Ensure compliance with CDISC rules

Value Proposition

- Confirmation that concept-based standards metadata holds all the information needed to auto-generate programs to process study data from collection to analysis
- Demonstration scripts, for industry to auto-generate programs to process study data from collection to analysis
- Consistent and repeatable mechanism to produce data and analysis
- Traceability of data transformation

Workstream Snapshot



Dependencies

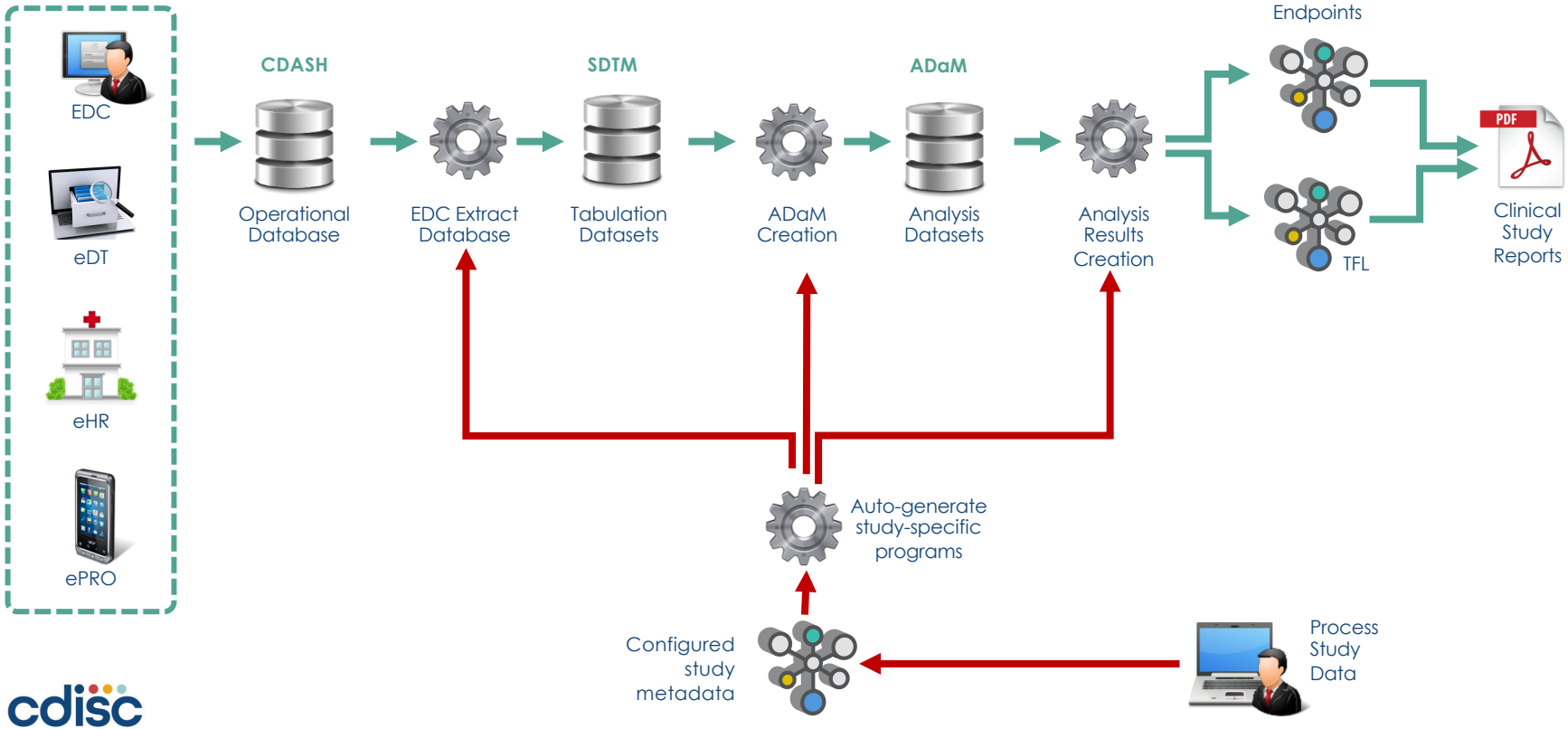
- WS1 Concept Modeling [upstream]
- WS2 Concepts in CDISC Library [upstream]
- WS4 Standards Specification (UC1) [upstream]
- WS5 Study Specification (UC2) [upstream]

Challenges

- Obtain and prepare the test study data
- Mechanism to process consistently and repeatedly
- Managing more robust or complex transformations or statistical analysis
- System agnostic technology assessment
- Programmatic means to produce auto-generated data
- Standardizing a visual representation of TFLs

Use Case 3 : Start to End Data Processing

Automatic population of data into artifacts



cdisc

WS6 Inputs

Earlier

- Study data for “360 Test Study”
  - Determined in collaboration with WS5

Later

- Configured study metadata from WS5
  - Including the semantic transformation metadata layer developed by WS1 sub-stream
- Compliance rules use case from CDISC Library

WS6 Activities (1)

- Determine study data for “360 Test Study”
  - Decision required: use mock data based on TAUG, or actual de-identified sponsor data?
  - Prefer multiple sponsor data, and ensure compliance with TAUG
  - Test data needs to conform with defined standards
    - Do not change standards to accommodate existing data

- Prepare the data collection sources
  - EDC extract
  - Electronic data transfer
  - HL7 FHIR transfer
  - ePRO data

- Determine technology for scripts; for scripts execution; for stored study data
  - For preparation of study data collection sources
  - For “validation” and “production” instances of study data

WS6 Activities (2)

Activities for “validation” instance of “360 Test Study”

Based upon determined study configuration and obtained and prepared test study data

This is the “double programming”, the “validation” instance of the target data

Any manual/programming means is acceptable (this is not the new automated 360 tooling)

- Process raw collected study data to populate all target dataset and TLFs
  - From data sources to operational database (ODM)
  - From operational database to SDTM datasets
  - From SDTM dataset ADaM datasets
  - From ADaM datasets to TLF outputs
- Ensure compliance of transformed data per Pinnacle 21

WS6 Activities (3)

Activities for “production” instance of “360 Test Study”

These activities are accomplished using the new automated 360 tooling developed during WS6, with the Configured Study Specification Metadata from WS5 as input

- Collaborate with WS1 sub-stream on the content of the semantic transformation metadata layer
- Develop scripts to ingest the configured study metadata and auto-generate the needed programs to transform the data
- Execute scripts to populate all the datasets and analysis outputs
- Ensure compliance of transformed data per Pinnacle 21

WS6 Activities (4)

Activities for “production” instance of “360 Test Study” – (continued)

- Compliance rules use case from CDISC Library
- Validate transformed data using the “validation” instance of the target data
- Consider use cases for link to LOINC and UCUM

WS6 Deliverables

- Scripts that ingest the configured study metadata and auto-generate the programs that transform the “360 Test Study” data
  - Creating validated data and analysis outputs

- Consider creation of CSR
  - What is needed beyond what we now have to produce the CSR?

Next Steps

Next Steps

- All hands - make final workstream selection by Tuesday April 16
  - If no change from when you registered, do nothing now
  - If making a change:
    - Send revised selection to Chris Gemma (cgemma@cdisc.org)
    - Identify your active workstream(s)
    - If desired, identify one or more workstreams on which to be informed
      - Invited to meetings; receive all emails; join the WS-specific Slack group

- Workstream 5 leader TBD; volunteers?
- WS Leader - schedule team meetings
- Team – identify scope for sprint 1
- Team – execute scope of sprint 1
- Team – onboard with 360 collaboration tools

Onboard with 360 Collaboration Tools

- CDISC WIKI/JIRA account
  - Sign up if needed
- CDISC 360 Wiki space
  - Access provided with WIKI/JIRA account
- Slack instant messaging
  - Receive email invite to create an account and access the CDISC 360 Slack space
  - Target is to provide this by the end of this week
- Cmap Cloud account
  - Receive email invite to create a Cmap Cloud account and access 360 project folders
  - Target date to provide this is TBD
- JIRA
  - WIKI/JIRA account provides access to JIRA and the 360 space within it
  - Training for JIRA use is TBD