cdisc

2023
JAPAN
INTERCHANGE
TOKYO | 10-11 JULY

Populating DDF Study Definitions Repository from study protocol using AI

Kunihito Ebi, Senior Consultant/Product Manager, Fujitsu Limited.

Meet the Speaker

Kunihito Ebi

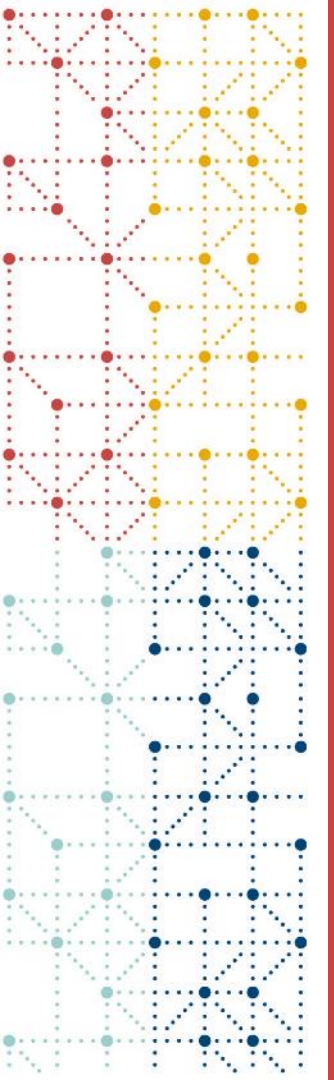
Title: Product Manager, Life Science Solutions

Organization: Fujitsu Limited

- CDISC authorized instructor of XML Technologies since 2015
- Product Manager of CDISC-based metadata management system with SDTM automation capability at Fujitsu since 2015

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 the organization where the author belongs.*



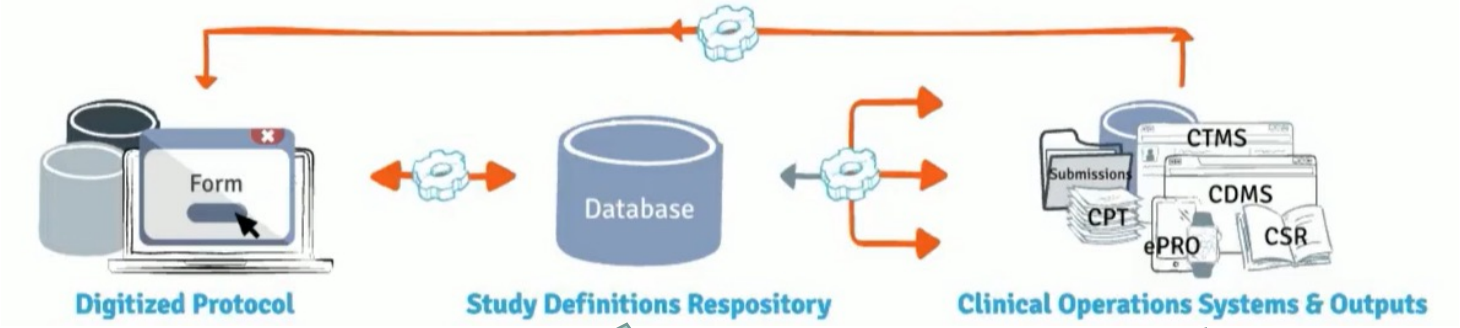
Agenda

1. Digital Data Flow (DDF) Overview and Current State
2. Our Implementation Approach to DDF
3. Deep Dive into Technical Solutions using AI

Digital Data Flow (DDF) Overview and Current State

Digital Data Flow (DDF) Overview

- Digital Data Flow (DDF) is an initiative by TransCelerate that aims to create digitized study protocol and automate creation of study assets.



Study builders or digital protocol authoring tools

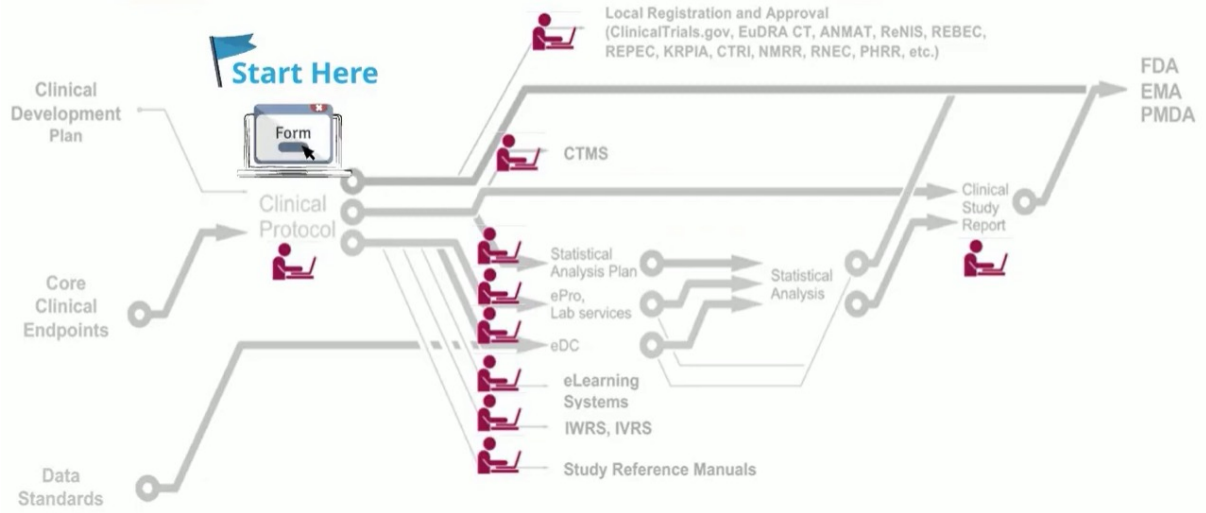
Central repository of digitized study protocol

Automate downstream documents and systems

Note: The content in this slide has been created by summarizing contents on the TransCelerate web site at the author's own discretion.

Benefits of DDF

- Accelerates study startup and study execution
- Increases automation while reducing manual hand-offs/transcription
- Reduce errors ad improve quality

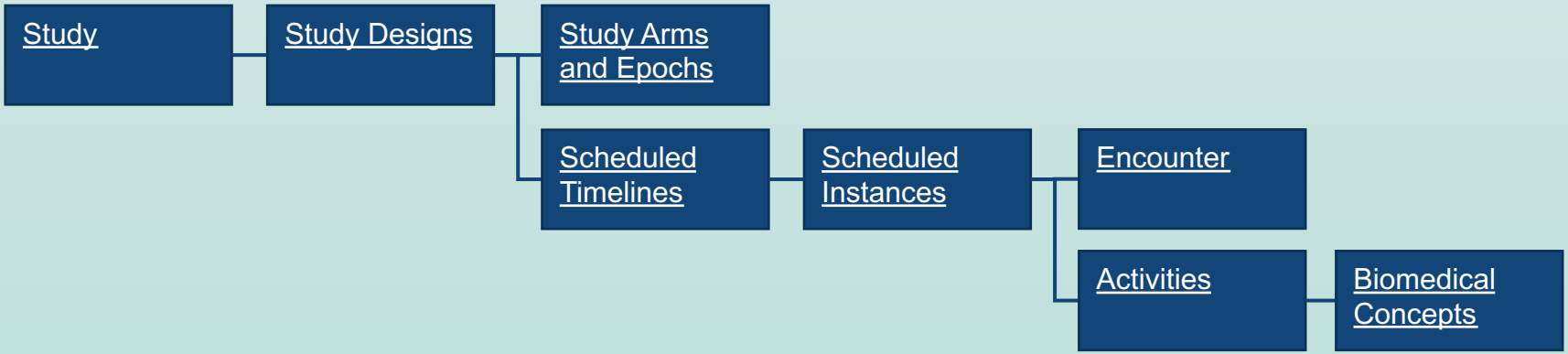


Note: The content in this slide has been created by summarizing contents on the TransCelerate web site at the author's own discretion.

DDF and USDM

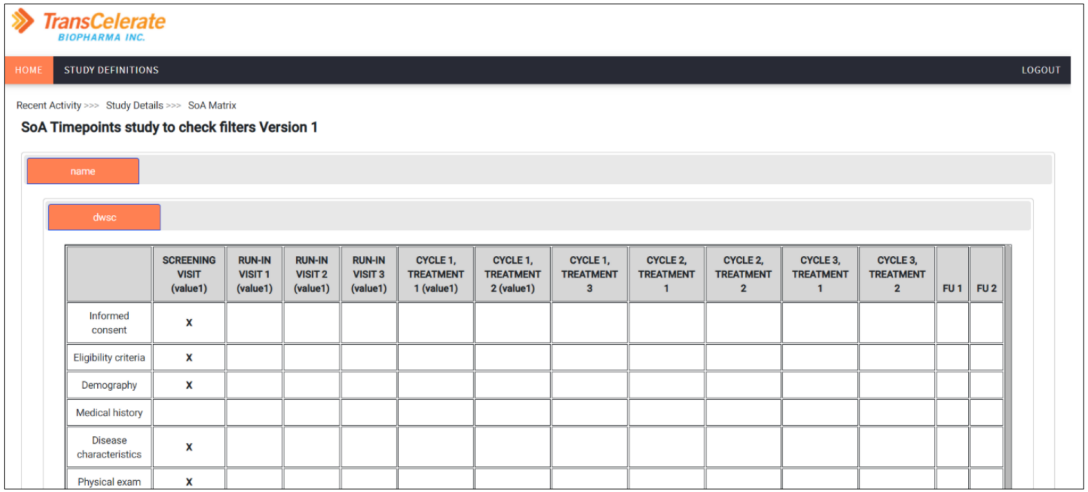
- CDISC is collaborating with TransCelerate to develop standard model for Study Definitions Repository. The model is called **Unified Study Definitions Model** (USDM).

Conceptual diagram of USDM (partial, simplified)



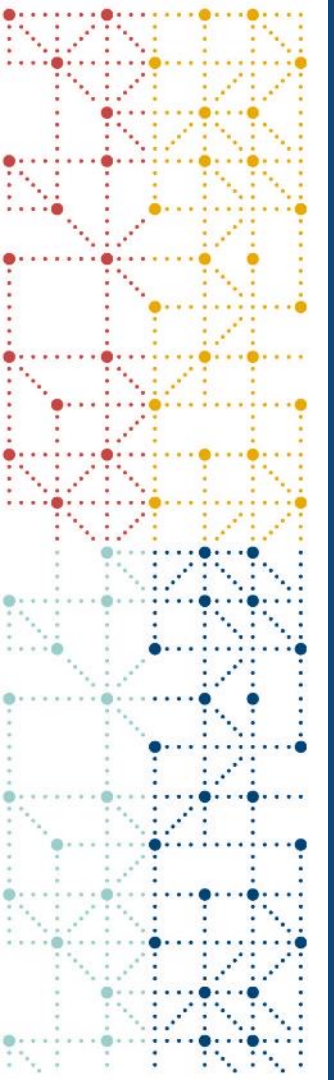
Current State of DDF

- **DDF Open-Source Tools** available on GitHub, including DDF SDR UI v2.0 released in March 2023



	SCREENING VISIT (value1)	RUN-IN VISIT 1 (value1)	RUN-IN VISIT 2 (value1)	RUN-IN VISIT 3 (value1)	CYCLE 1, TREATMENT 1 (value1)	CYCLE 1, TREATMENT 2 (value1)	CYCLE 1, TREATMENT 3	CYCLE 2, TREATMENT 1	CYCLE 2, TREATMENT 2	CYCLE 3, TREATMENT 1	CYCLE 3, TREATMENT 2	FU 1	FU 2
Informed consent	X												
Eligibility criteria	X												
Demography	X												
Medical history													
Disease characteristics	X												
Physical exam	X												

- **USDM-IG v2.0** published in June 2023



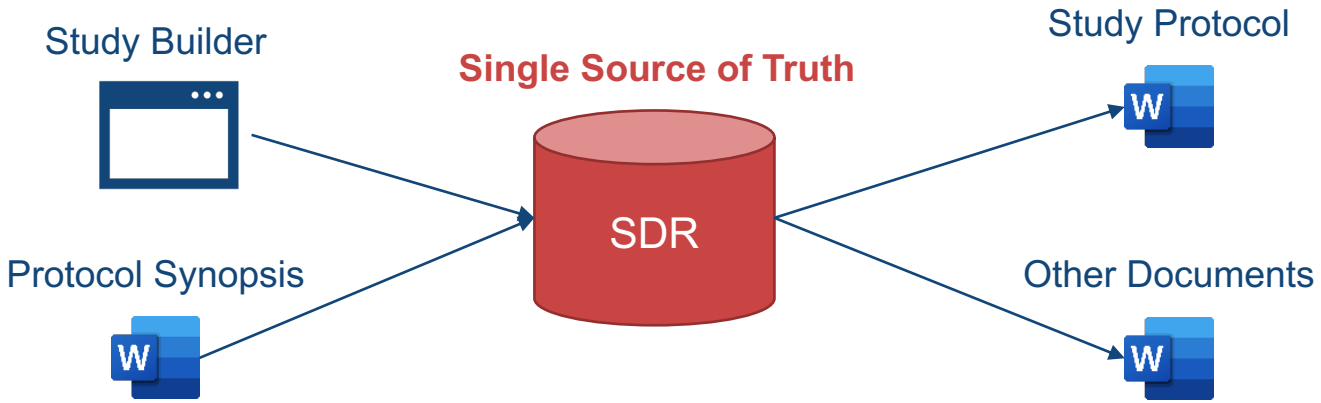
Our Implementation Approach to DDF

Study Builder, Authoring Tool or both?

- **Study Builder** is useful to digitize study protocol contents from the beginning, especially to complicated information to be reused overtime.
  - e.g. Schedule of Activities (SoA)
- **Traditional or digitized authoring tools** (i.e. document templates) have been used as a common practice for authoring study protocols and other documents.
- We assume that both Study Builder and Authoring Tool are used for document authoring in the DDF era.

SDR First, or Document First?

- SDR may be filled out first if every content is entered via Study Builder.
- A document may already exist before Study Protocol (e.g. protocol synopsis, or study protocol in other studies).
- We assume that information is transferred from a document to SDR and vice versa.



Combining USDM and ICH M11 (CeSHarP)

- Mapping Exercise between USDM-IG v2.0 and ICH M11
  - Between M11 Template Titles (up to outline LV3) and USDM Classes/Attributes

CeSHarP Title	Outline Level	USDM Class	USDM Attribute
Protocol Full Title	1	study.protocolVersions	officialTitle
Sponsor Confidentiality Statement	1	(N/A)	(N/A)
Protocol Number	1	study.studyIdentifiers	studyIdentifier
Version	1	study.protocolVersions	protocolVersion
Amendment Number	1	(N/A)	(N/A)
Amendment Scope	1	(N/A)	(N/A)
Compound Name(s)	1	study.studyDesigns	studyInvestigationalInterventions
Trial Phase	1	study	studyPhase
Acronym	1	study	studyAcronym
Short Title	1	study.protocolVersions	briefTitle
Sponsor Name and Address	1	study.studyIdentifiers	organizationName
Sponsor Name and Address	1	study.studyIdentifiers	organizationAddress
Manufacturer Name and Address	1	(N/A)	(N/A)

- 59 among 175 titles have been used to extend our SDR
  - This is essential to enable reuse of document contents

Reuse of Document Contents

- Essential elements of study protocol is available in USDM-IG v2.0.
- Broader context may be available only in document content, so we link document sections to our SDR elements to enable later reuse.

USDM Class and Attributes

Study Populations
Population Description
Planned Maximum Age of Participants
Planned Minimum Age of Participants
Planned Number of Participants
Planned Sex of Participants

CeSHarP Template Sections

5.1 Selection of Trial Population

Describe the population selected (for example, healthy volunteers, adult participants, paediatric participants) and how the enrollment criteria reflect the populations that are likely to use the drug if approved. Specify the population age range (for example, ≤3 months, ≥18 to ≤80 years old) and any key diagnostic criteria for the population (for example, "acute lung injury", or a specific biomarker profile). If applicable, describe similar conditions or diseases and their differential diagnosis.

[Selection of Trial Population]

5.2 Rationale for Trial Population

Provide a rationale for the trial population ensuring that the population selected is well defined and clinically recognisable. Justify whether the trial intervention is to be evaluated in children, in adults unable to consent for themselves, other vulnerable participant populations, or those that may respond to the trial intervention differently (for example, elderly, hepatic or renally impaired, or immunocompromised participants).

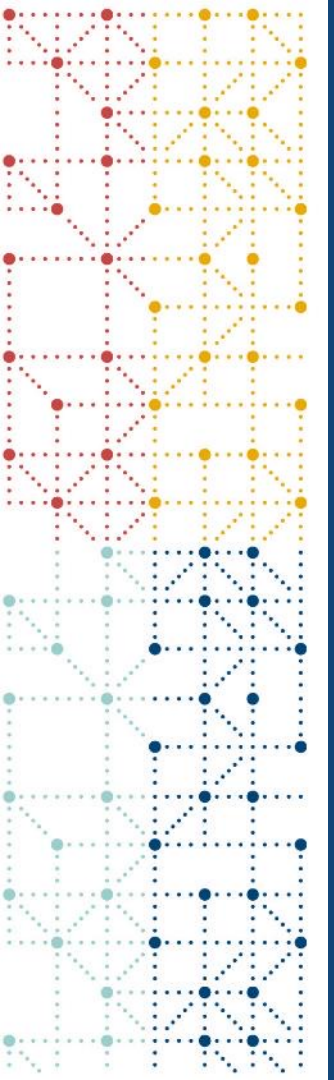
[Rationale for Trial Population]

Automation of EDC Setup and SDTM Mapping

- Draft eCRF Specification can be created from SDR by mapping from USDM to ODM (Operational Data Model developed by CDISC).

USDM Element		ODM Element
Encounter	→	Event
Activity	→	Form
Biomedical Concept	→	Item

- Mapping from eCRF to SDTM becomes easier because the same set of biomedical concepts in the SDR are applied to both eCRF and SDTM spec.

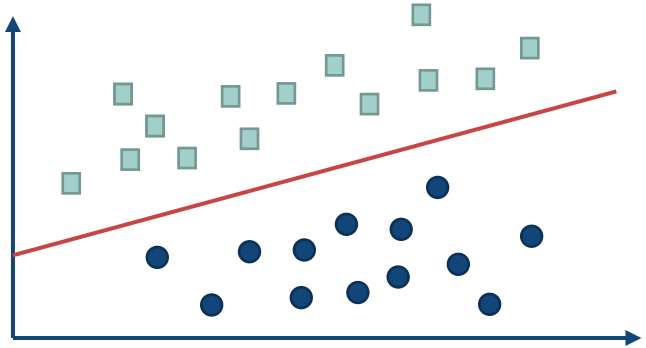


Deep Dive into Technical Solutions using AI

Why can AI be used for DDF?

- Artificial Intelligence (AI) is good at classification problems, which we often encounter at DDF implementation.
- Availability of training data is often a challenge, but they may be available in public domain.

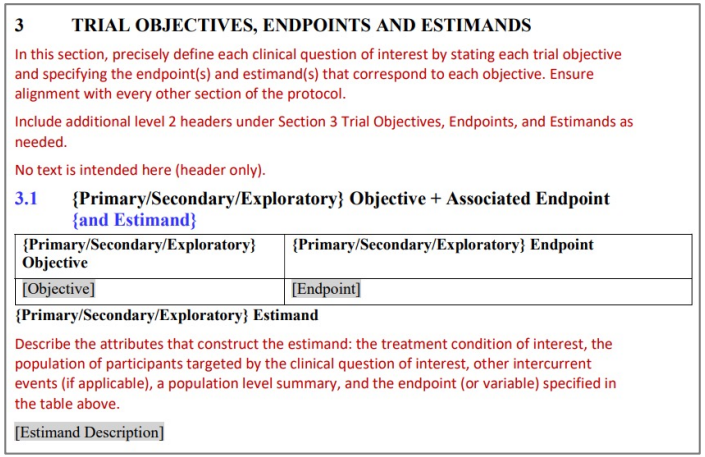
Example: Finding Biomedical Concepts from Study Protocol Document



Problem:
Binary classification of whether a sentence in study protocol represents a biomedical concept

From Study Protocol Document to SDR

- It is **"Tagging"**, another classification problem, to associate document sections, which may or may not be conformant to M11 CeSHarP, with SDR.
- NLP techniques for **Similarity Measurement** work fine when Study Protocol is based on or pre-mapped to CeSHarP.
- Calculating distance of meanings by **Vectorization** is more practical when Study Protocol is NOT pre-mapped to CeSHarP.



3 TRIAL OBJECTIVES, ENDPOINTS AND ESTIMANDS

In this section, precisely define each clinical question of interest by stating each trial objective and specifying the endpoint(s) and estimand(s) that correspond to each objective. Ensure alignment with every other section of the protocol.

Include additional level 2 headers under Section 3 Trial Objectives, Endpoints, and Estimands as needed.

No text is intended here (header only).

3.1 {Primary/Secondary/Exploratory} Objective + Associated Endpoint {and Estimand}

{Primary/Secondary/Exploratory} Objective	{Primary/Secondary/Exploratory} Endpoint
[Objective]	[Endpoint]

{Primary/Secondary/Exploratory} Estimand

Describe the attributes that construct the estimand: the treatment condition of interest, the population of participants targeted by the clinical question of interest, other intercurrent events (if applicable), a population level summary, and the endpoint (or variable) specified in the table above.

[Estimand Description]

Tags from USDM

Estimand | Objective | Endpoint

Tags from CeSHarP

Trial Objectives, Endpoints and Estimands

SoA Table to SDR

- Special technique is sometimes used to digitize document content, such as automatic recognition of Schedule of Activities (SoA):

	Screening	Run-In	Treatment									Follow-Up			Comment
Visit	1	2	3	4	5	7	8	9	10	11	12	13	Early Termination	Retrieval	
Week	-2	-0.3	0	2	4	6	8	12	16	20	24	26			
Informed Consent	X														
Patient Number Assigned	X														
Hachinski <4	X														
MMSE 10-23	X														
Physical Examination	X											X	X		
Medical History	X														
Habits	X														
Chext X-ray	X														
Apo E Genotyping				X											
Patient Randomization			X												
Vital Signs Temperature	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Ambulatory ECG placed		X													
Ambulatory ECG removed			X												
ECG	X			X	X	X	X	X	X	X	X	X	X		
Placebo TTS Test	X														
CT Scan	X														
Concomitant Medications	X		X	X	X	X	X	X	X	X	X	X	X	X	
Laboratory (Chem/Hema)	X			X	X	X	X	X	X	X	X	X	X		
Laboratory (Urin)	X			X				X			X		X		
Plasma Specimen			X	X	X	X		X		X			X		
Hemoglobin A1C	X														
Study Drug Record			X	X	X	X	X	X	X	X	X	X	X		
Adverse Events	←													→	

Encounters

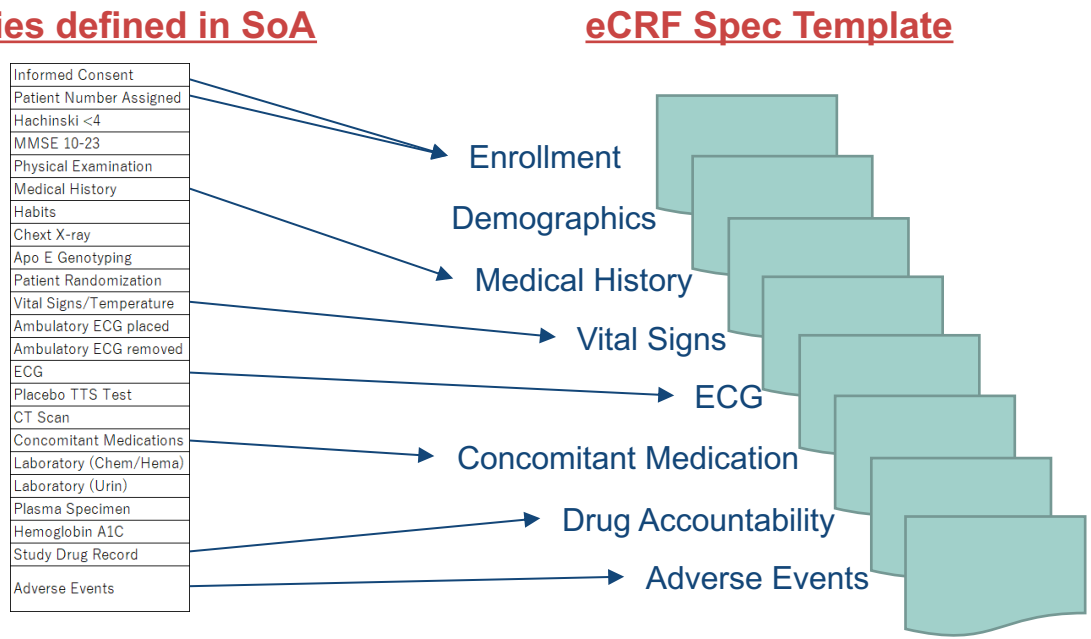
Activities

Encounters x Activities

Note: The SoA table has been created by the author based on the CDISC pilot study available on the CDISC GitHub.

SDR to CRF Spec

- eCRF Template will be used to create eCRF Spec because information available in SDR is limited. Mapping from Activities in SDR to eCRF Forms is another classification problem.

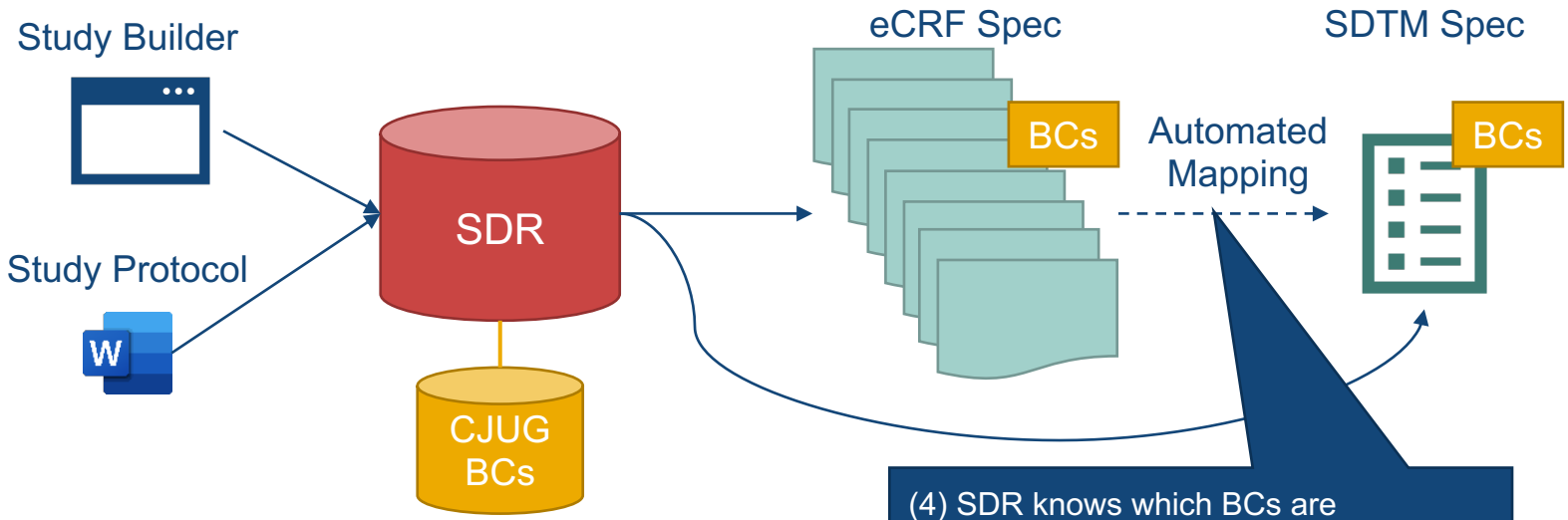


Mapping from CRF to SDTM

(1) Activities/BCs are populated via Study Builder or read from Study Protocol.

(2) Activities with BCs are mapped to forms when a draft eCRF Spec is created from SDR.

(3) Activities with BCs are mapped to SDTM domains when a draft SDTM Spec is created from SDR.



Summary and Future Steps

- DDF aims to digitize study protocol and automate creation of study assets.
- USDM defines standard model for SDR, that serves as the central repository of DDF.
- SDR should be extended with ICH M11 CeSHarP.
- SDR may be populated via Study Builder or Authoring Tool.
- Many classification problems to be solved by AI are found in DDF impl.
- Draft eCRF and SDTM Specs with mappings can be created from SDR.
- The same principles explained in this presentation should be applicable to statistical processes.

i.e. Protocol → SAP → ARM → ADaM Spec

Thank You!

For questions, please contact the speaker at:

ebi.kunihito@fujitsu.com