

USDM Synthetic Test Data Generation



AD08 PHUSE EU Connect 2025

Stuart Malcolm

Head of Standards, Efficiency and Automation

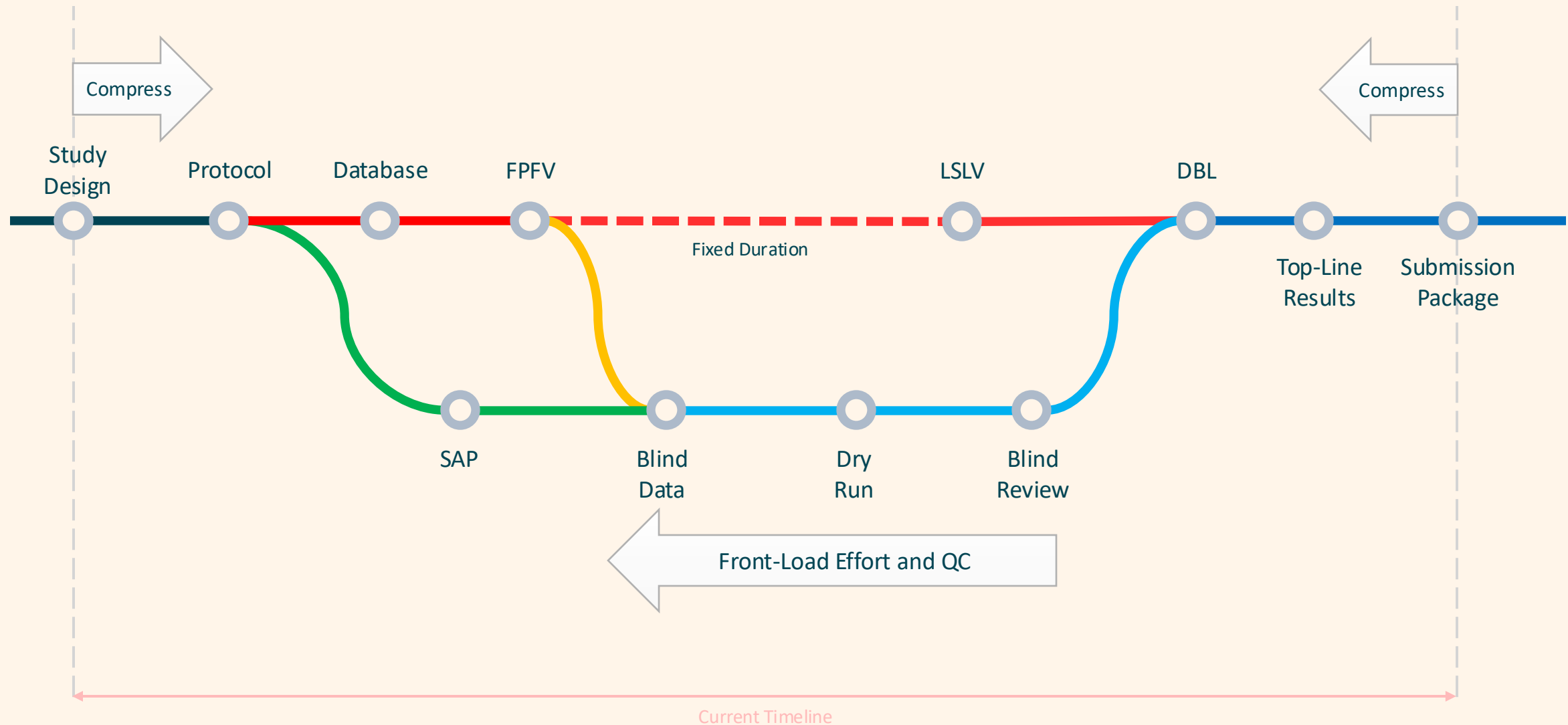
VeraMed



“I cannot start
until I get some
data”

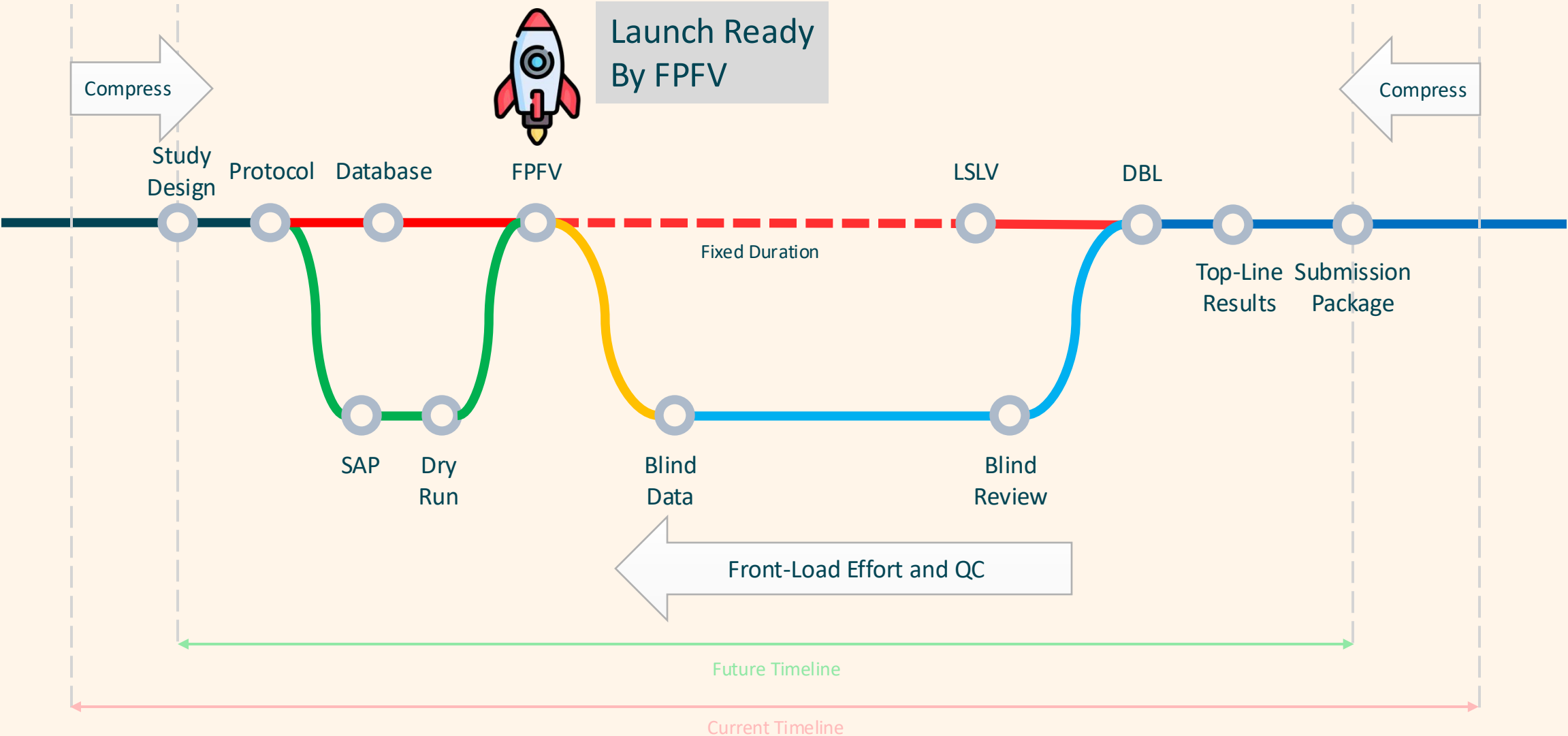
Current Process – Manual and Data-Driven

How do we compress timelines by front-loading QC to reduce amendments and downstream rework.



Future State – Automated and Metadata-driven

Leveraging digital Protocol and End-to-End machine-readable standards to be ‘submission ready’ by FPFV



Moonshot Vision

Leveraging machine-readable standards to be 'launch ready' by FPFV

Change of mindset

- From analysis to modelling

Digital Transformation

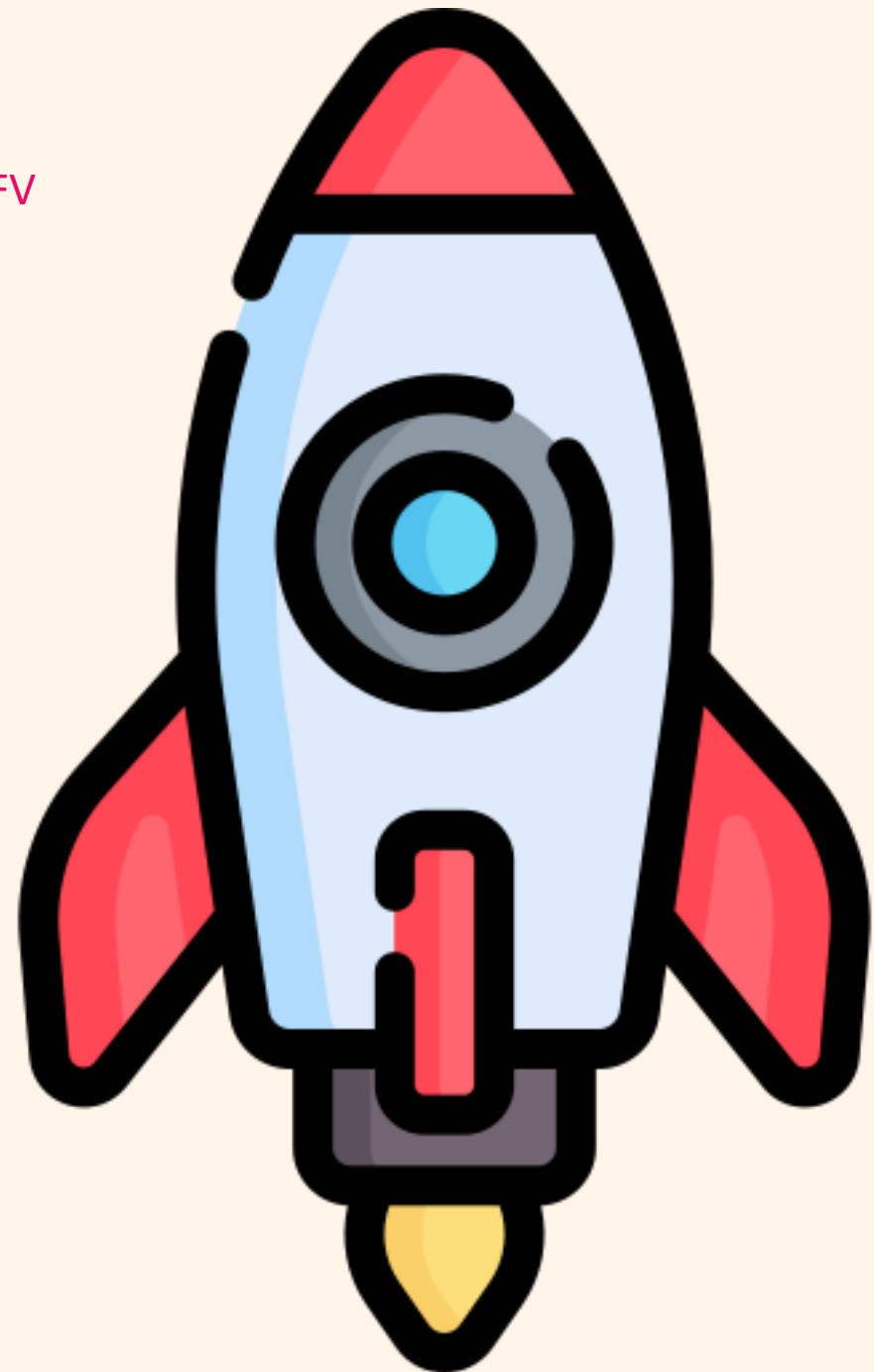
- Digital Protocol, SAP, DM Plan, etc.

Robust interchange standards

- Multi-discipline collaboration

Quality Assurance over Quality Control

- Realistic Synthetic Data





Realistic Synthetic Data

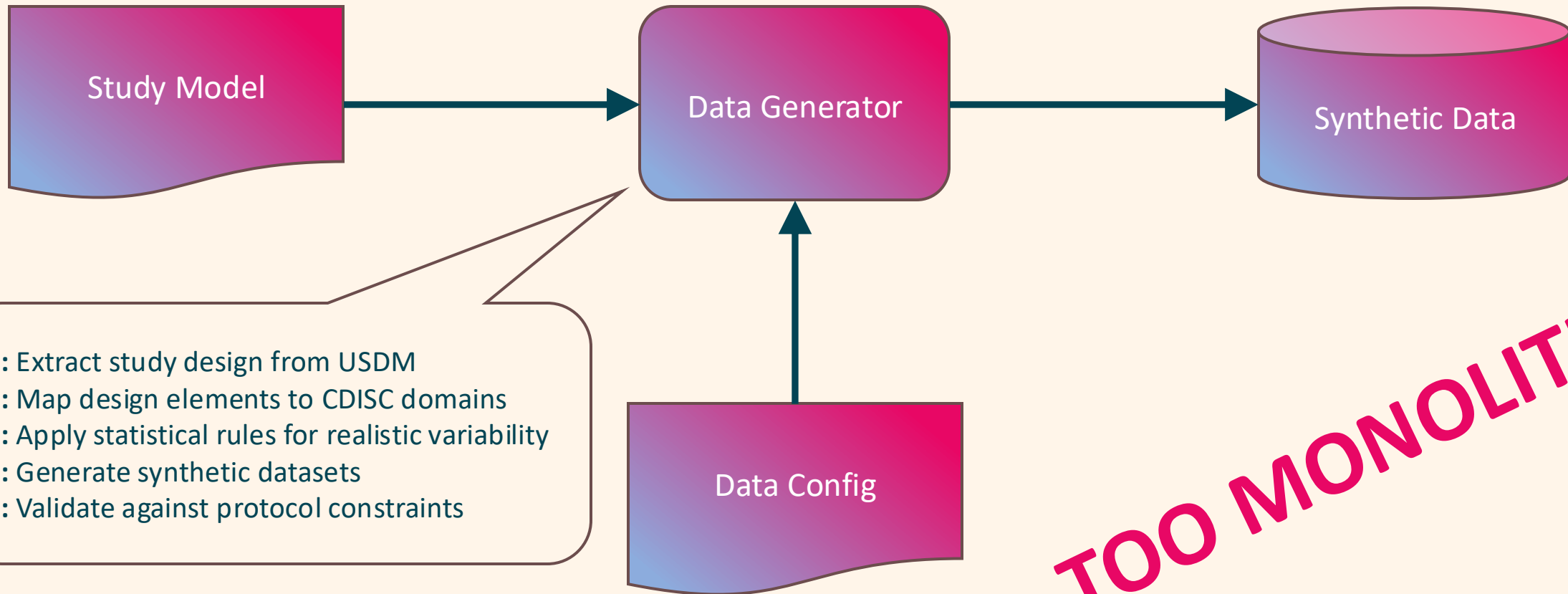


Prior Art: Existing synthetic data solutions

Project / Tool	Standards	Capability	License
cdiscdataset.com	EDC, CRF, SDTM, SEND, ADaM, TLF	number of subjects, TA, Treatment Arms only. No Population control.	Proprietary
pharmaverseraw	EDC-agnostic. Mix of CDASH/Non-CDASH.	based on the CDISC Pilot Study SDTM	Apache 2
pharmaversesdtm	SDTM	Mix of TA-agnostic (DM, VS, EG, RS...) and TA-specific (oe_ophtha, rs_onco, rs_onco_irecist...)	Apache 2
random.cdisc.data	ADaM	Generate wide range of 'realistic' ADaM domains programatically	Apache 2
raw.synthetic.data	n/a	Proposed COSA project. Not released.	n/a
SsdTM	SDTM	Synthetic data generators in Python	MIT
Synthpop	Any	R Synthetic data created from original dataframe using statistical distributions	GPL 3

Design Architecture #1

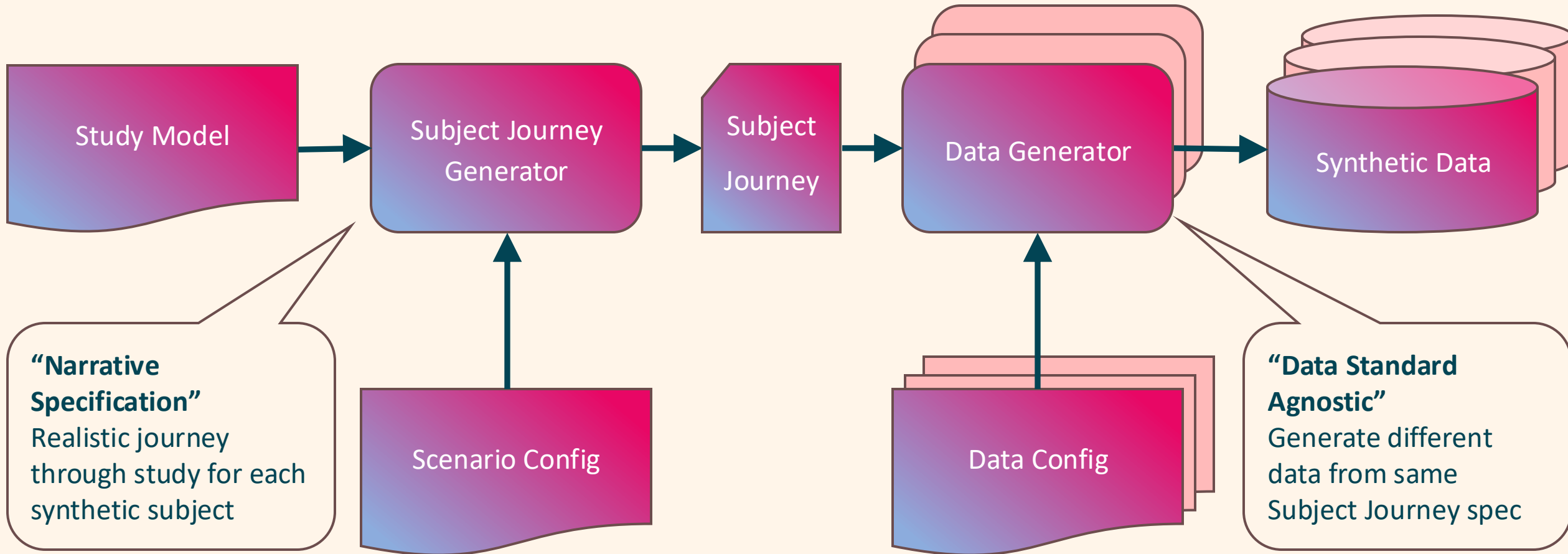
Generate CDISC compliant datasets directly from USDM source.



TOO MONOLITHIC!

Design Architecture #2

Generate intermediate subject journey description which is then used to generate synthetic data.





Unified Study Definitions Model Implementation Guide (USDM-IG)

Version 4.0 (Final)

Prepared by the DDF Team

Notes to Readers

This is the final Version 4.0 of the Unified Study Definitions Model Implementation Guide (USDM-IG v4.0).

Revision History

Date	Version
2025-06-03	4.0 Final
2024-04-16	3.0 Final
2023-06-07	2.0 Final
Not applicable*	1.0

*No USDM-IG was developed for USDM Version 1.0.

See [Appendix E](#) for Representations and Warranties, Limitations of Liability, and Disclaimers.

CONTENTS

1	INTRODUCTION	4
1.1	PURPOSE	4
1.2	ORGANIZATION OF THIS DOCUMENT.....	4
1.3	HOW TO READ THIS DOCUMENT	4
2	FUNDAMENTALS OF THE USDM.....	6
3	RELATIONSHIP TO OTHER STANDARDS AND FORMATS	9
3.1	RELATIONSHIP TO OTHER CDISC STANDARDS	9
3.2	RELATIONSHIP TO OTHER STANDARDS.....	11
4	USDM FEATURES	12
4.1	OVERVIEW	12
4.2	PRINCIPLES	12
4.3	NAMING CONVENTIONS	13
4.4	INTERNAL IDENTIFIERS WITHIN THE MODEL.....	14
4.5	CONTROLLED TERMINOLOGY	14
4.6	STUDY, PROTOCOLS, AND AMENDMENTS	14
4.7	STUDY IDENTIFIERS AND TITLES.....	16
4.8	STUDY DESIGN.....	18
4.9	STUDY ROLES AND ORGANIZATIONS	20
4.10	ARMS AND EPOCHS	23
4.11	ACTIVITIES.....	23
4.12	PROCEDURES.....	25
4.13	BIOMEDICAL CONCEPTS	25
4.14	STUDY TIMING.....	26
4.15	STUDY INTERVENTIONS	32
4.16	STUDY OBJECTIVES AND ENDPOINTS.....	34
4.17	STUDY ESTIMANDS	35
4.18	POPULATIONS, COHORTS, AND ELIGIBILITY CRITERIA	35
4.19	ABBREVIATIONS	38
4.20	UNSTRUCTURED CONTENT	40
4.21	SYNTAX TEMPLATES.....	42
4.22	XHTML ATTRIBUTES.....	44
4.23	ADDRESSING FOOTNOTES.....	44
4.24	COMPLEX STUDY DESIGNS	51
4.25	SCHEDULE OF ACTIVITY VIEWS.....	57
5	USDM DATA DICTIONARY	61
6	USDM API.....	99
6.1	GENERAL	99
6.2	SERIALIZATION	99
6.3	ADDITIONAL ATTRIBUTES AND REQUIRED CONTENT	99
6.4	EXTENSION MECHANISM	100
7	MAPPING TO OTHER STANDARDS AND FORMATS.....	108
7.1	CREATION OF SDTM TRIAL DESIGN DOMAINS	108
7.2	INFORMING CLINICALTRIALS.GOV REGISTRY	108
7.3	INFORMING CTIS REGISTRY	109
7.4	CREATION OF M11 DOCUMENTS	109





3.1	RELATIONSHIP TO OTHER CDISC STANDARDS	9
3.2	RELATIONSHIP TO OTHER STANDARDS.....	11
4	USDM FEATURES	12
4.1	OVERVIEW	12
4.2	PRINCIPLES	12
4.3	NAMING CONVENTIONS.....	13
4.4	INTERNAL IDENTIFIERS WITHIN THE MODEL.....	14
4.5	CONTROLLED TERMINOLOGY	14
4.6	STUDY, PROTOCOLS, AND AMENDMENTS	14
4.7	STUDY IDENTIFIERS AND TITLES.....	16
4.8	STUDY DESIGN.....	18
4.9	STUDY ROLES AND ORGANIZATIONS	20
4.10	ARMS AND EPOCHS.....	23
4.11	ACTIVITIES.....	23
4.12	PROCEDURES.....	25
4.13	BIOMEDICAL CONCEPTS	25
4.14	STUDY TIMING.....	26
4.15	STUDY INTERVENTIONS.....	32
4.16	STUDY OBJECTIVES AND ENDPOINTS.....	34
4.17	STUDY ESTIMANDS.....	35
4.18	POPULATIONS, COHORTS, AND ELIGIBILITY CRITERIA.....	35
4.19	ABBREVIATIONS	38
4.20	UNSTRUCTURED CONTENT	40
4.21	SYNTAX TEMPLATES.....	42
4.22	XHTML ATTRIBUTES.....	44
4.23	ADDRESSING FOOTNOTES.....	44
4.24	COMPLEX STUDY DESIGNS	51
4.25	SCHEDULE OF ACTIVITY VIEWS.....	57
5	USDM DATA DICTIONARY.....	61

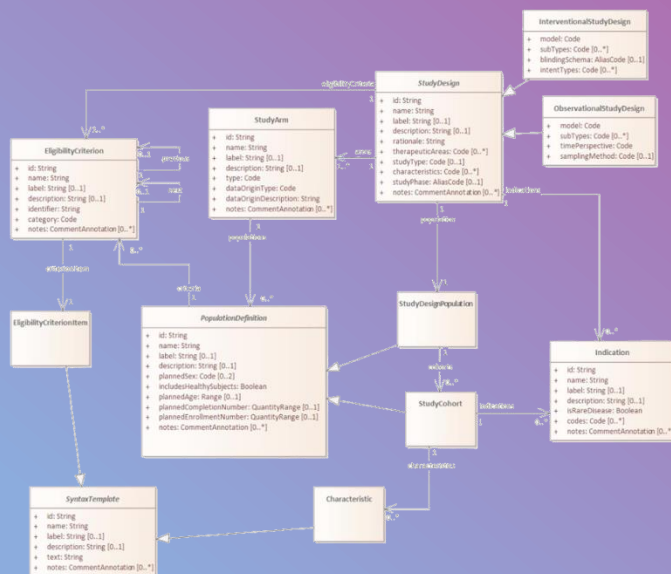
Unified Study Definition Model v4.0



Populations

4.18 Populations, Cohorts and Eligibility

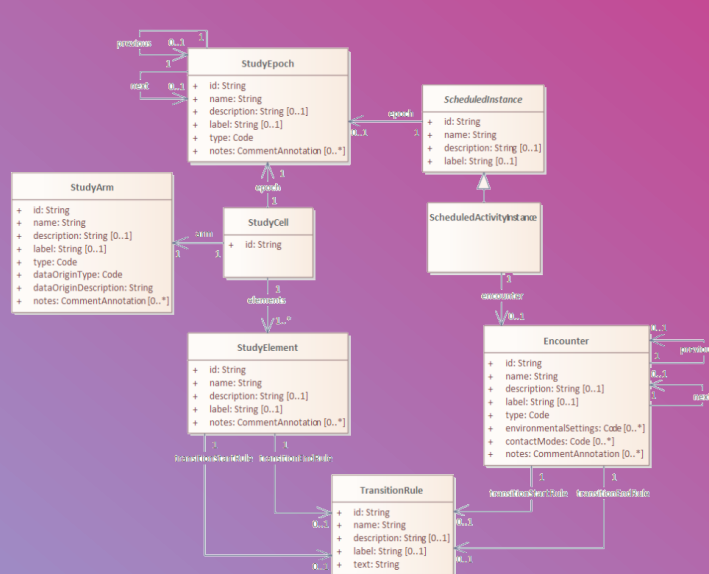
- Sub-population on Subject characteristic
- Population size, enrollment number
- Indication
- plannedAge, plannedSex
- Eligibility criteria (Syntax Template)



Study Design

4.8 Study Design, Arms and Epochs

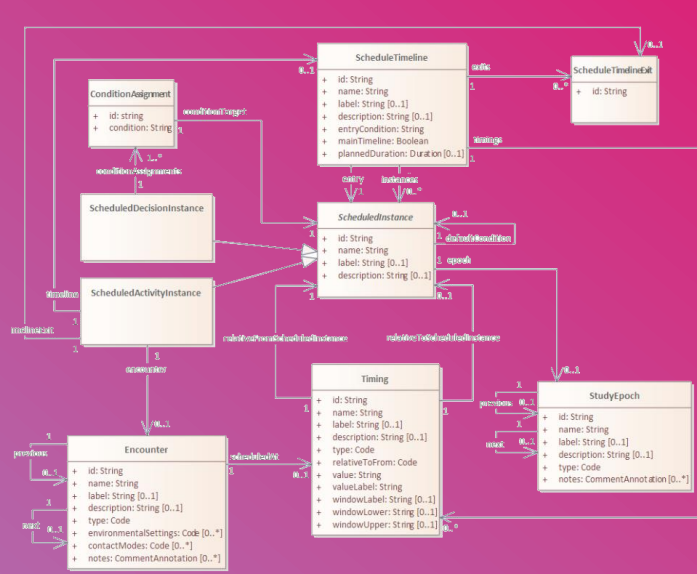
- Interventional or Observational
- Study Phase, Type and Design Parameters
- StudyEpoch, StudyArm , StudyCell
- StudyElement and TransitionRule



Schedule of Activities

4.14 Study Timing

- Timeline of Activities and Encounters
- Intervals and Windowing
- Decisions and Branching
- Profiles and Sub-Timelines
- Unscheduled Visits



4.18 class Population Cohort Eligibility

4.10 class Arms & Epochs UML diagram

4.14 Class Timeline UML diagram



Subject Journey Through a Study

Populations

1. Generate Synthetic Population

Study-Level
Distribution



Subject-level
Distribution



2. Generate Synthetic Subjects

- Demographics
- Medical History
- Associated Persons



Study Design

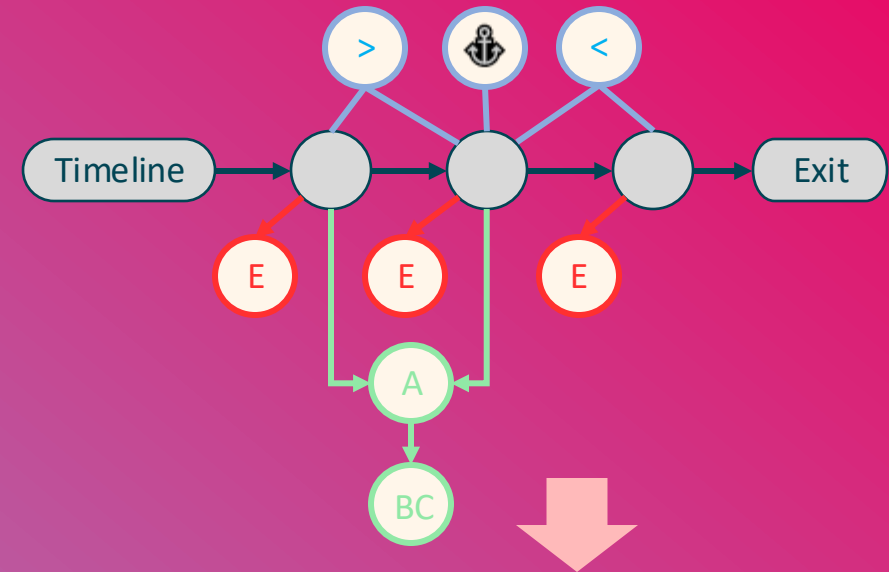
4. Walk subjects through SoA



3. Enroll into Study

- Trial Design
- Subject Elements
- Disposition

Schedule of Activities



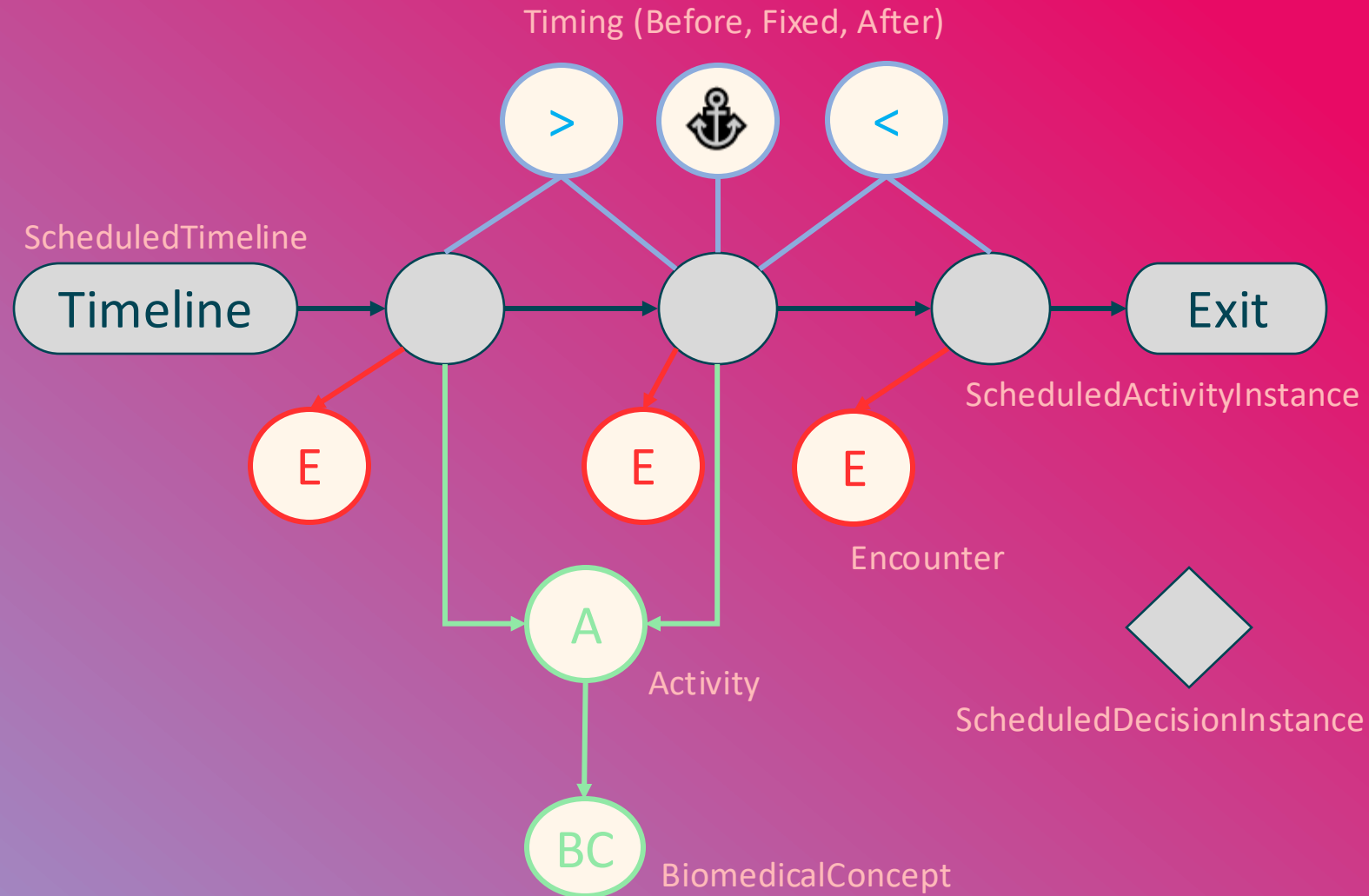
5. Generate Subject Journey Description

Subject, Encounter, Activity, BC
Subject, Encounter, Activity, BC
Subject, Encounter, Activity, BC
Subject, Encounter, Activity, BC
...



	Screening	Day 1	
Subject related Assessments			
Informed consent	X		
In/Exclusion criteria	X	X	
Demography	X		
Medical history	X		
Randomisation		X	
Efficacy			
Lab efficacy assessments		X	X
PRO questionnaire		X	X
Safety	A	E	E
Vital signs			E
ECG	X	X	
Hematology	X	X	
Biochemistry	X	X	
Adverse events	X	X	X
Intervention			
Drug dispensation		X	X
Drug accountability		X	

Encounters





Wrap-up

Summary



Realistic Synthetic Data

Simulated study not random data



Architecture

Split study modelling from data generation



Digital Protocol

Using USDM to create synthetic study



Subject Journey

Intermediate format to specify path through study



Unstructured text?

How to deal with text fields in USDM?



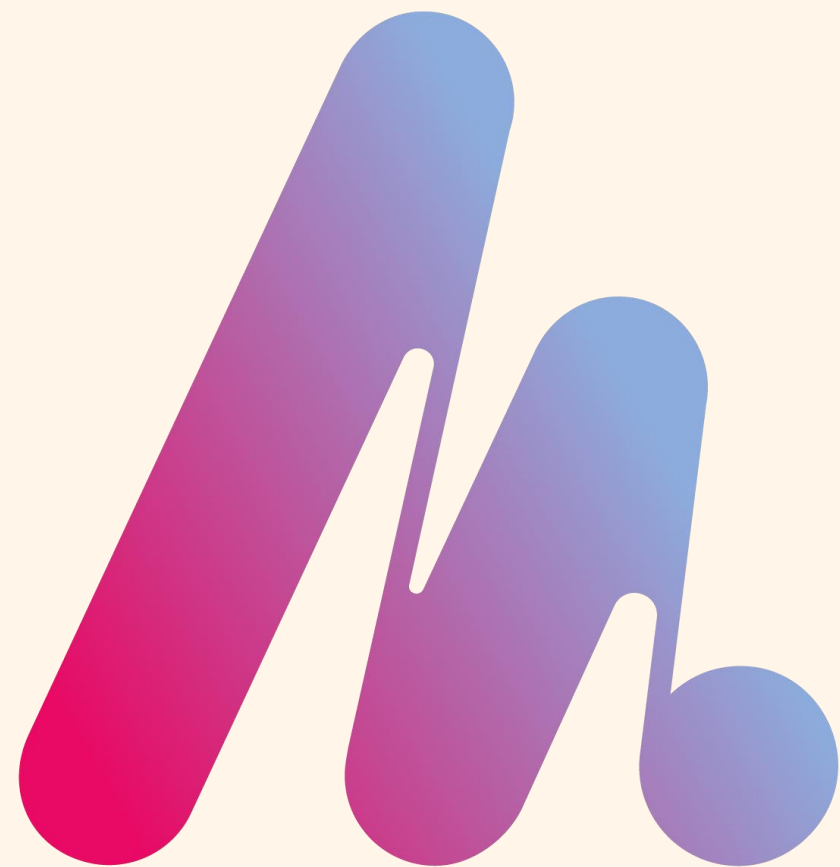
Use of AI/ML?

Alternative approach to traditional simulation techniques



User Experience?

User interfaces and workflow





Thank You!

Questions?

VeraMed

Subject Journey Through a Study



Populations

1. Generate Synthetic Population

Study-Level
Distribution



Subject-level
Distribution



2. Generate Synthetic Subjects

- Demographics
- Medical History
- Associated Persons



Study Design

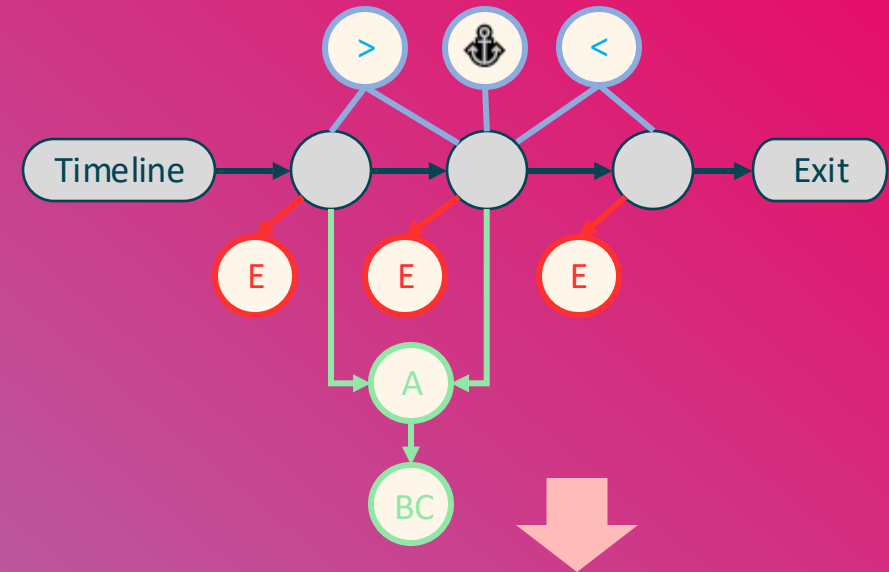
4. Walk subjects through SoA



3. Enroll into Study

- Trial Design
- Subject Elements
- Disposition

Schedule of Activities



5. Generate Subject Journey Description

Subject, Encounter, Activity, BC
Subject, Encounter, Activity, BC
Subject, Encounter, Activity, BC
Subject, Encounter, Activity, BC
...

1. Generate Synthetic Population

Study-Level
Distribution



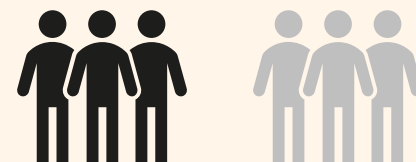
Subject-level
Distribution



2. Generate Synthetic Subjects

- Demographics
- Medical History
- Associated Persons

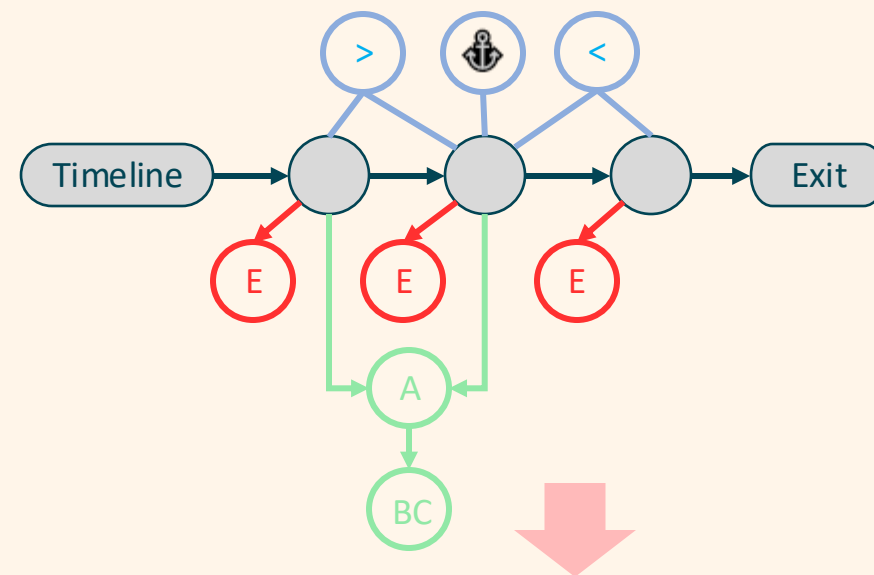
4. Walk subjects through SoA



Study-level
Distribution

3. Enroll into Study

- Trial Design
- Subject Elements
- Disposition



5. Generate Subject Journey Description

Subject, Encounter, Activity, BC
Subject, Encounter, Activity, BC
Subject, Encounter, Activity, BC
Subject, Encounter, Activity, BC
...