Central Metadata Repository for Automation in SDTM Dataset Generation

Naoko Izumi, Ajay Sinha
24Oct2019

NOVARTIS | Reimagining Medicine

Disclaimer

- All information provided in this slides is provided for information purposes only
- Views expressed in this presentation are those of the speaker and not necessarily of Novartis

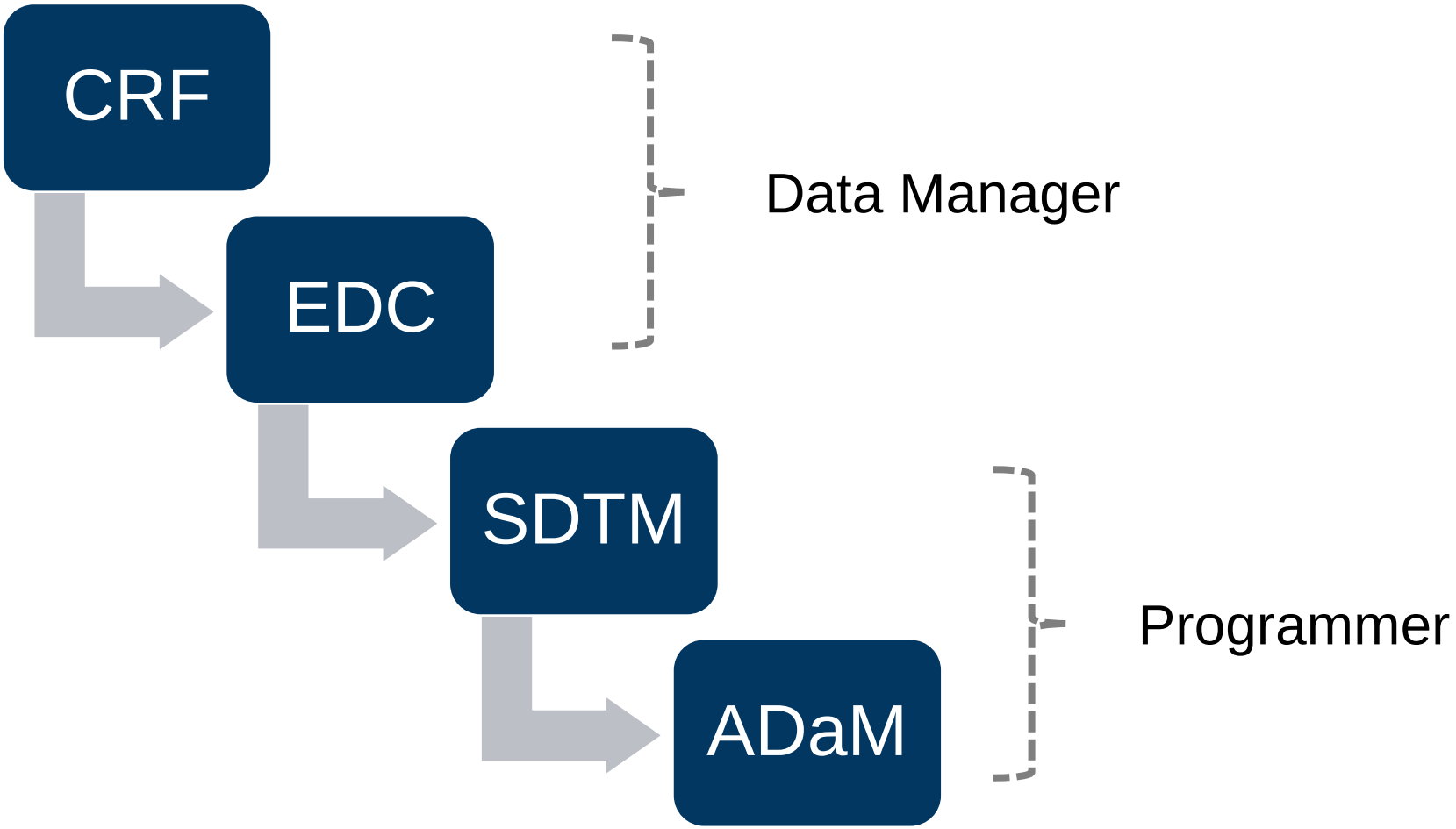
Agenda

- Introduction
- Pre-requisite to start SDTM
- SDTM Process
- SDTM Automation Process
- Key Benefits

Introduction

- In a world of continual improvement in processes need for automated tool has become the next tag line. In context of data submitted to Health Authorities quality and consistency of data is of prime importance.
- In a properly managed setup, standards and metadata can be used to drive automation. This presentation describes **metadata-driven approach** that can be followed for generation of SDTM datasets.

Introduction - Data Flow



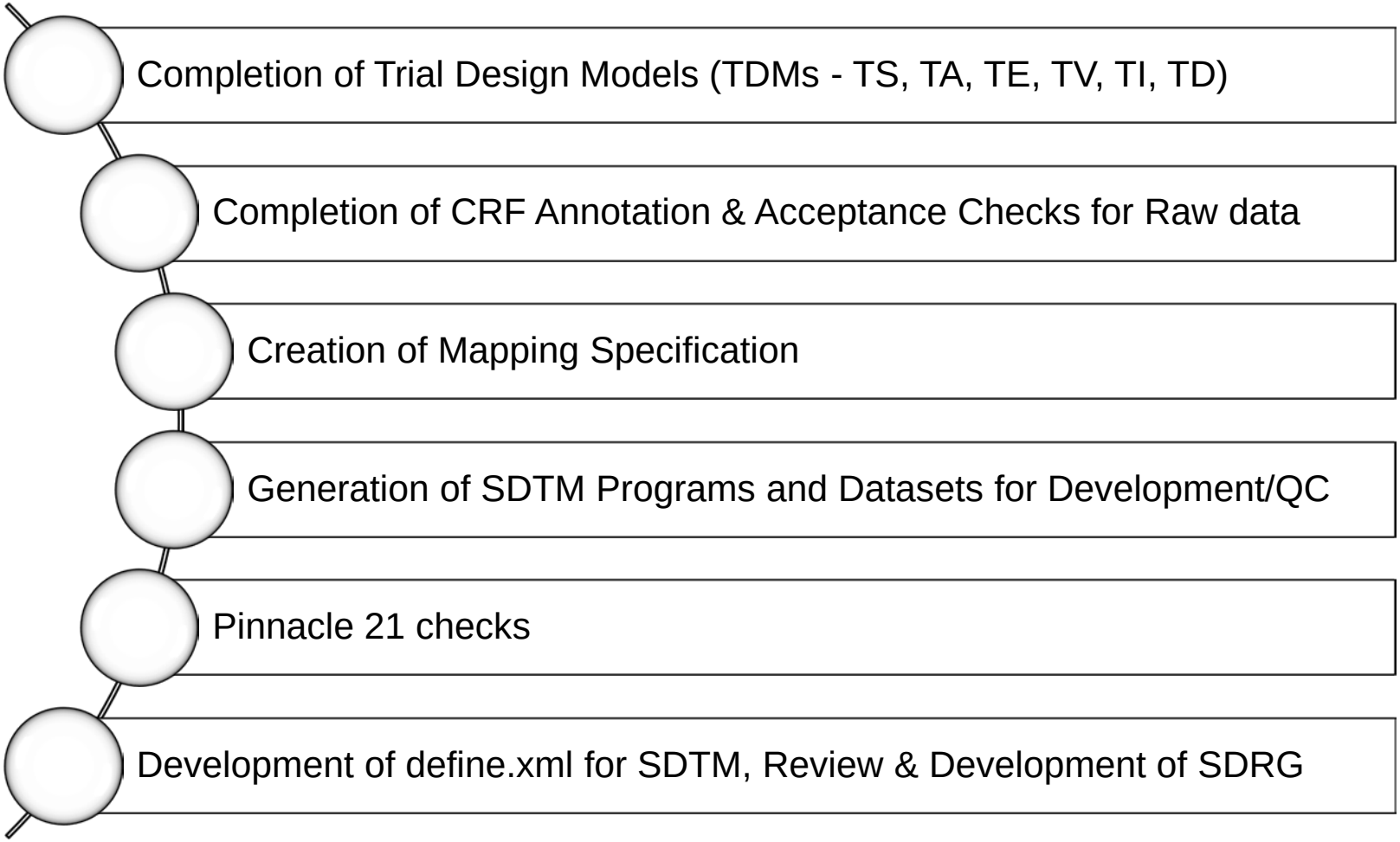
Pre-requisites to Start SDTM

1. SDTM-IG
2. Study Protocol
3. CRF
4. ALS (Architectural Load Sheet)
5. DQP (Data Quality Plan)
6. DTS (Data Transfer Specifications) for Vendor Data
7. Study Data Technical Conformance Guide
8. Data Standards Catalog

Pre-requisites to Start SDTM Contd..

9. Versions
   - MedDRA
   - WHO Drug Dictionary
   - SNOMED-CT (Systematized Nomenclature of Medicine Controlled Terminology)
   - NDF-RT (National Drug File - Reference Terminology)
   - UNII (FDA Unique Ingredient Identifier)
   - Pinnacle 21
   - Controlled Terminology
10. Raw data

SDTM Process



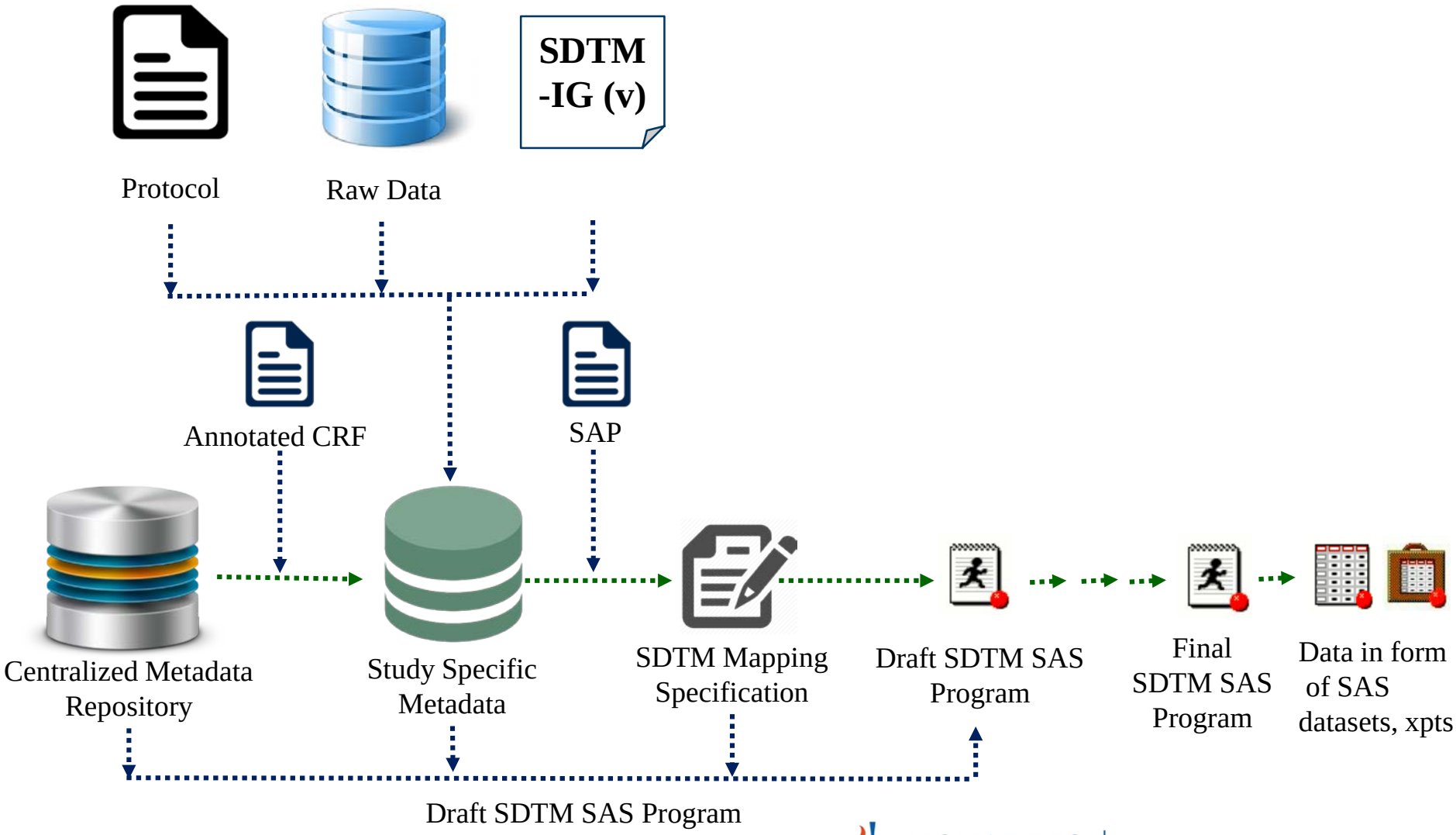
SDTM Automation Process

NOVARTIS | Reimagining Medicine

SDTM Automation Process

- SDTM Structure is pre-defined (SDTM-IG)
- SDTM Metadata can be generated using SDTM-IG
  - Ordering of variables
  - Datatypes of variables
  - Usage of variables
- Controlled Terminology freely available on Website (https://www.cancer.gov/research/resources/terminology/cdisc)

Working Model Steps



Centralized Metadata Repository

1. Repository created using SDTM-IG metadata
2. Availability of Sponsor defined variables
3. Precision values pre-defined
4. Origin of the variables pre-defined
5. Grouped into SDTM defined Class (including custom domains)
6. Ability to create study specific metadata - using Raw data, Controlled Terminology Sheet

Study Specific Controlled Terminology

Raw data values can be easily compared with CT to get SDTM accepted values

Controlled Terminology Sheet

Raw data

Study Specific Controlled Terminology

SAS Raw Data

Controlled Terminology Sheet

	SEX
1	Female
2	Female
3	Male
4	Female
5	Female
6	Female
7	Female
8	Female
9	Female
10	Female
11	Female
12	Female
13	Female
14	Female
15	Female
16	Female

D1985 Anatomical Location

	A Code	B Codelist Code	C Codelist Extensible (Yes/No)	D Codelist Name	E CDISC Submission Value	F CDISC Synonym(s)
24628	C49636	C66732		Sex of Participants Response	BOTH	
	C16576	C66732		Sex of Participants Response	F	Female
24629						
	C20197	C66732		Sex of Participants Response	M	Male
24630						

Study Specific Metadata

Global Metadata Raw data Study Specific Metadata

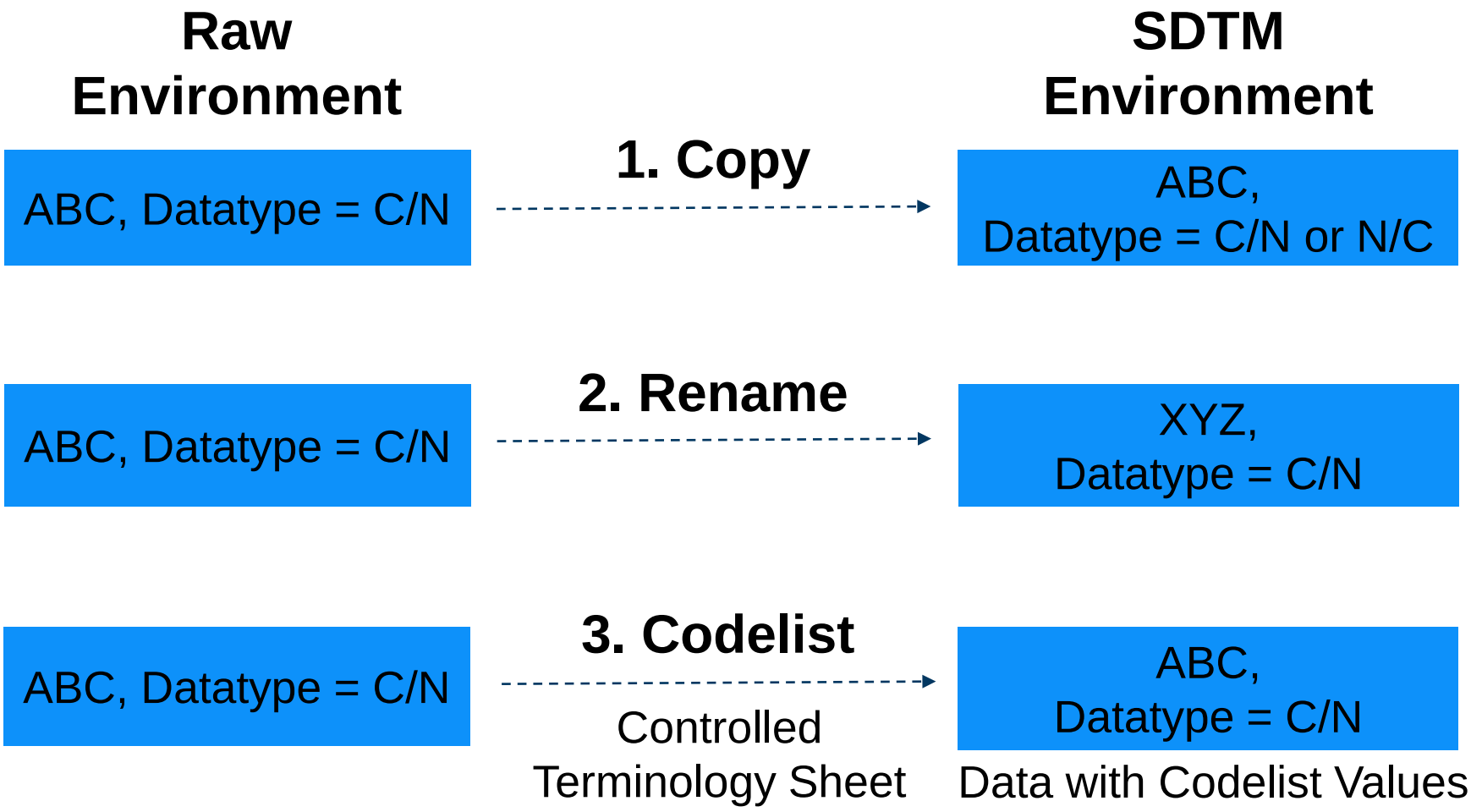
Status	Domain Name	Variable Name	Derivation Name	Derivation logic
EDC	AE	AEACN	N/A	N/A
DRM	AE	AEACN	COPY	EDC.AE.AEACN
SDTM	AE	AEACN	COPY	DRM.AE.AEACN
AD	ADAE	AEACN	COPY	SDTM.AE.AEACN

	MEMLABEL	NAME	TYPE	LENGTH	VARNUM	LABEL
1	Adverse Events	AEACN	2	25	21	Action Taken with Study Treatment
2	Adverse Events	AEBDSYCD	1	8	17	Body System or Organ Class Code
3	Adverse Events	AEBODSYS	2	100	16	Body System or Organ Class
4	Adverse Events	AECAT	2	20	15	Category for Adverse Event
5	Adverse Events	AECONTRT	2	2	30	Concomitant or Additional Trtmnt Given
6	Adverse Events	AEDECOD	2	200	9	Dictionary-Derived Term
7	Adverse Events	AEENDTC	2	20	34	End Date/Time of Adverse Event
8	Adverse Events	AEENDY	1	8	36	Study Day of End of Adverse Event
9	Adverse Events	AEHLGT	2	200	13	High Level Group Term
10	Adverse Events	AEHLGTCD	1	8	14	High Level Group Term Code

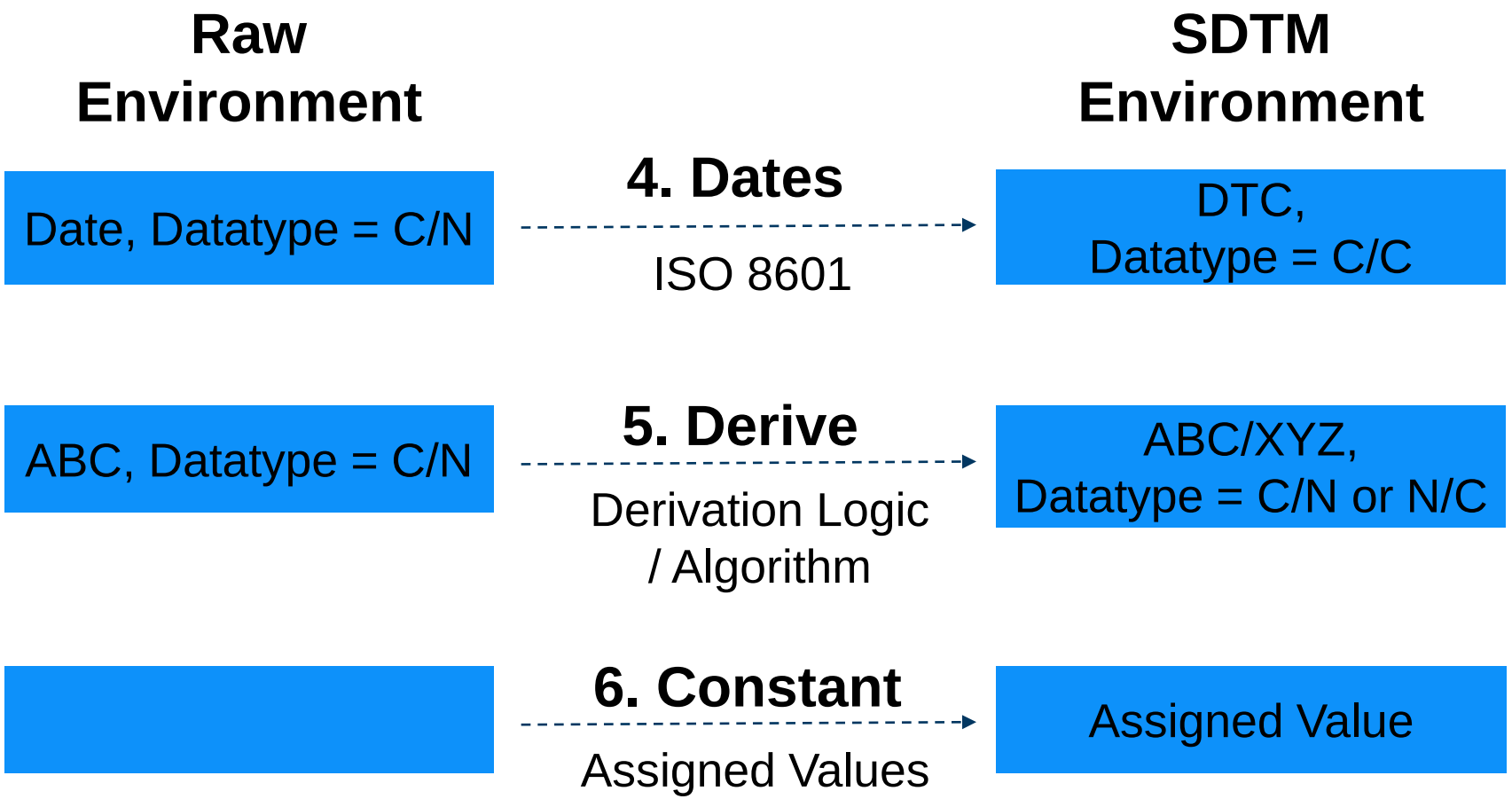
Categorization of SDTM Variables

1. **Copy** - from Source data (may be change in datatype)
2. **Rename** - Just rename the raw to SDTM variable
3. **Codelist** - Apply study codelist to raw data to populate SDTM Variables with accepted values
4. **Dates** - Direct copy with ISO 8601 format
5. **Derive** - Per SDTMIG provide generic derivation algorithm to get SDTM Variables from raw
6. **Constant** - Assign a value for a variable

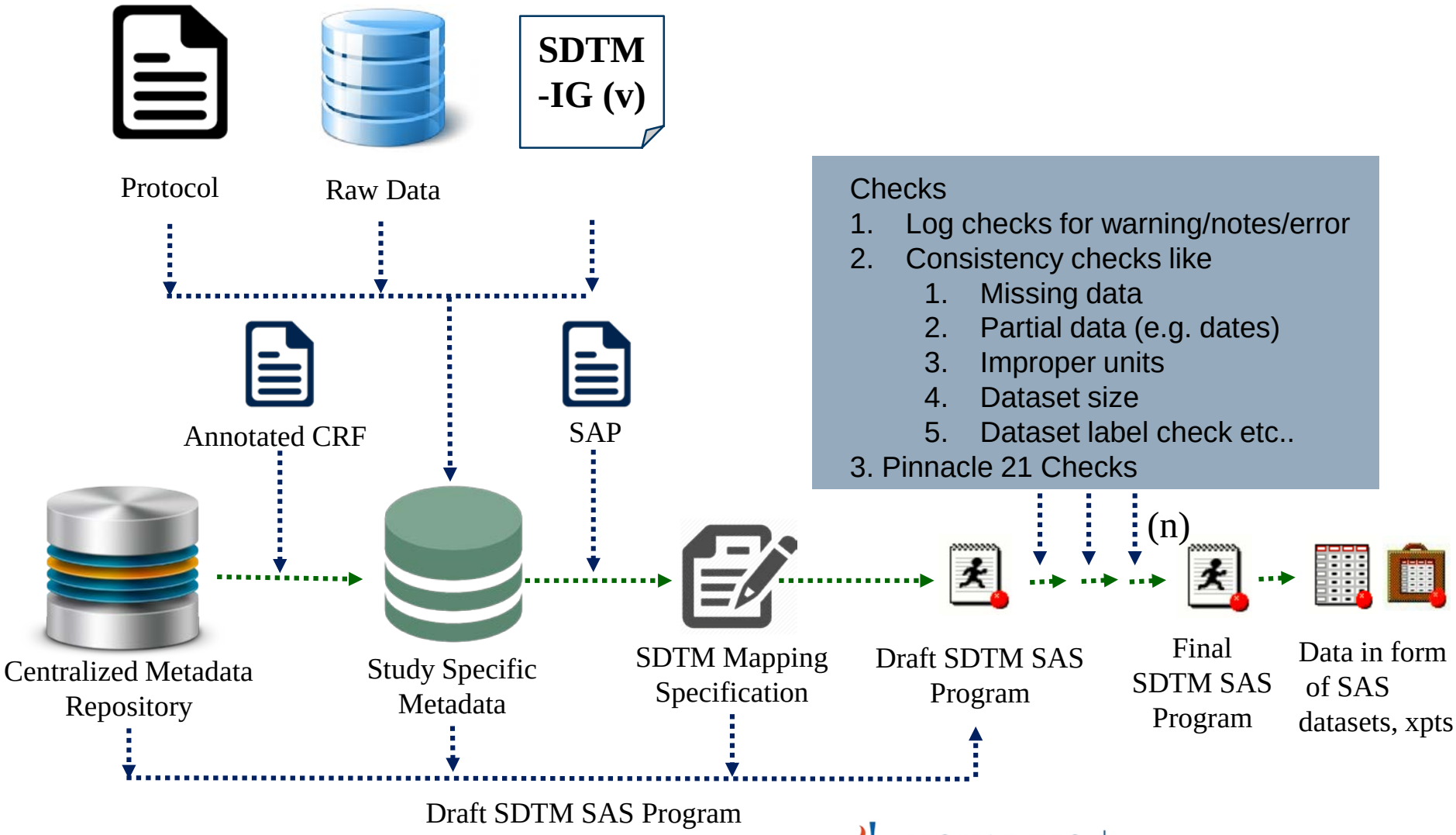
Categorization of SDTM Variables Contd..



Categorization of SDTM Variables Contd..



Working Model Steps



SDTM Final Deliverables

- Define.xml
- Study Data Reviewers Guide (Pinnacle 21 Reports justified)
- Annotated CRF
- Pure SDTM xpt files

Key Benefits

- ✓ Greater consistency between CRF and SDTM
  - Traceability of changes
- ✓ Easy to construct automation process
  - Create dataset specification and programs from metadata automatically
- ✓ Maintenance of metadata by global standard team
  - Centralize all changes made to metadata

Thank you

NOVARTIS | Reimagining Medicine