Case Study of CDISC Submission at Novartis Oncology

PhUSE Boston – May 7, 2014

Ram Govindaraju

ram.govindaraju@novartis.com

NOVARTIS

Disclaimer

The opinions expressed in this presentation and on the following slides are solely those of the presenter and not necessarily those of Novartis. Novartis does not guarantee the accuracy or reliability of the information provided herein

Introduction

- The following presentation will focus on the process flow for the creation of FDA submission data in CDISC[1] format for a recent submission at Novartis Oncology
- Emphasizes on the Metadata driven systems and their effective use in CDISC compliant submissions
- Covers the data flow from source to SDTM[2] & ADaM[3]
- Explains about Metadata and it's use in the derivation of elements in the data
- Illustrates generic templates for the creation of SDTM+ data

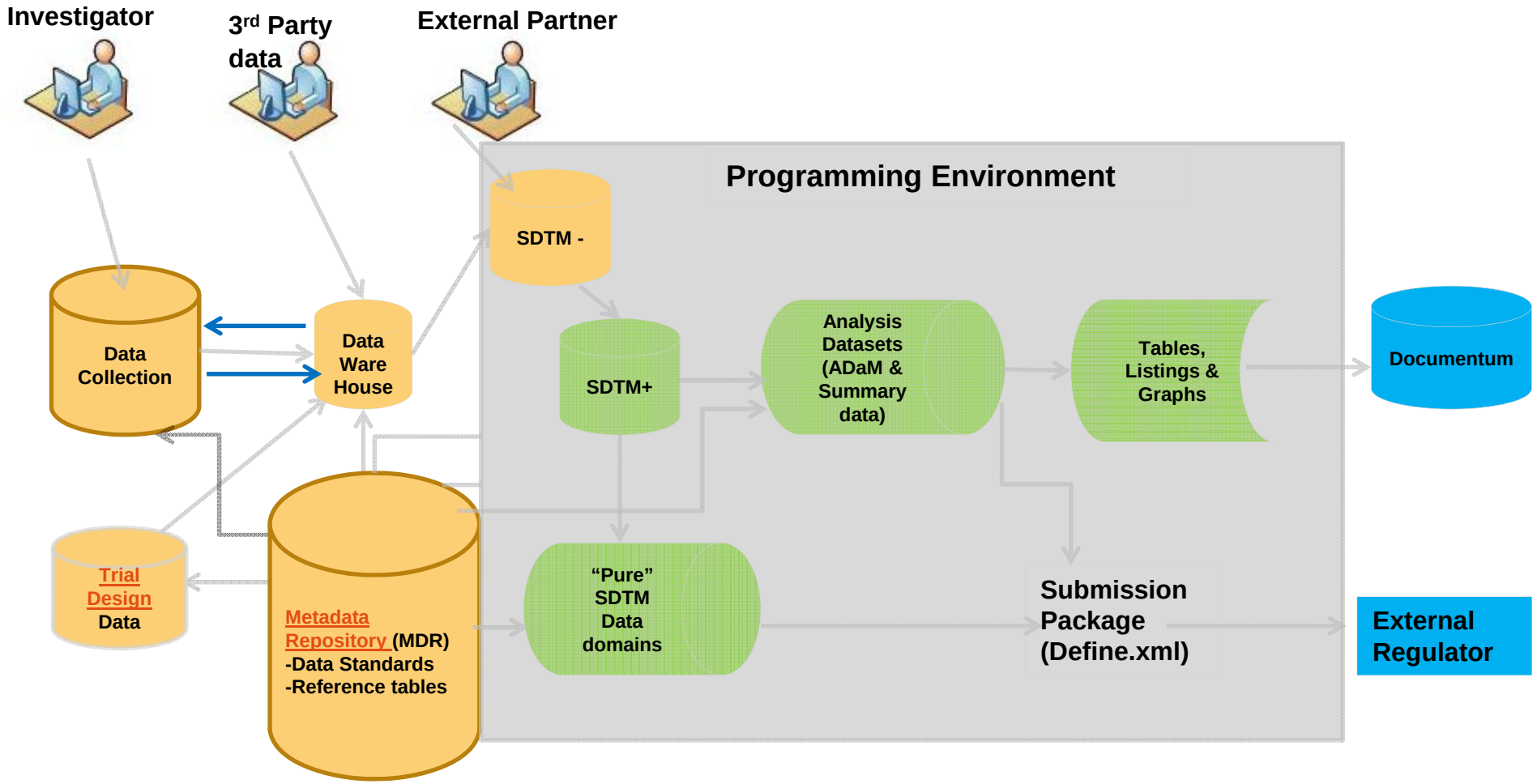
NOVARTIS

1 CDISC = Clinical Data Interchange Standards Consortium

Data Flow

Flow Chart – Global Picture

Data Std. Governance Boards / ADaM Governance Board / Oncology Std Outputs GB



NOVARTIS

Data Flow – Cont'd

- As shown in the flow chart, collected data from Database including data from 3rd party transfer(s) (eg. Lab, Biomarker etc.) will be loaded into the Data Warehouse
- Data then goes through pre-defined derivations and transformations in the Warehouse and subsequently extracted into the programming environment (this data is referred as SDTM-)
- SDTM+ data that contains a given SDTM and its corresponding Supplemental Qualifier variables in one domain, will be derived from SDTM- data
- ADaM datasets and pure SDTM domains (along with Supplemental Qualifiers) will be created from SDTM+ domains

Global Metadata

Key points

- Consistency in the Submission data
- Compliance with CDISC standards
- Standard programming templates were created based on Metadata
- Global Metadata contains data structure details for staging area within GPS for every data domain:
  - Source (RP) – SDTM-
  - Derived (DR) – SDTM+
  - Pure SDTM and Supplemental Qualifier (SD)
  - ADAM (AD)

Key fields in Metadata

Key fields:

- DSTYPE: Dataset type related to staging area. (S = Source (RP), D = Derived (DR), B = Pure SDTM and Supplemental Qualifier (SD), A=ADAM (AD))
- DSNAME: Dataset name (Ex: AE)
- ETNAM: Element name. This is the name of the variable represented in the OC / LSH systems and is different to the SAS variable name (Ex: CPEVENT)
- NAME: SAS variable name as it appears in SDTM / SDTM+ / ADaM (Ex: VISIT)
- FETNAM: Also called FQE for Fully Qualified Element name; this is the unique identifier for an element, within a given domain for a giving staging area. FQE value are composed of Staging area, Data Domain name (dataset name) and Element name separated by “.” Sign (Ex: DR.AE.CPEVENT)
- CODELIST: Codelist associated to the element (Ex: NOYES for corresponding AEYN field)

Key fields in Metadata (cont'd)

- CT_TERM: Controlled Terminology Term (Ex: SDTM, ABBREVIATED etc.)
- CONET: Contributing element: List of FQE used for the derivation (Ex: RP.AE.CPEVENT)
- DRVNAM: Derivation name (Ex: COPY_ELEMENT, DECODE etc.). Detailed description of derivations available in SPS Share point.
- DRVSEQ: Derivation sequence number (setup for a given domain for a given staging area) (Ex: 1)
- DRVDOM: List of domain (logical name) involved in the related derivation (Ex: RP.AE)
- DRV_LOGIC: Derivation logic description (Ex: VISIT = AE.CPEVENT)
- DRV_PARAM: Derivation parameter: Addition information (parameters, list of additional FQE) (Ex: RANDOMIZATION)
- KEY: Key variable number, populated only for key fields and missing otherwise (Ex: 1)

Key fields in Metadata (cont'd)

- LABEL: SAS variable label (Ex: Visit Name)
- LENGTH: SAS variable length (Ex: 20)
- MAND: Mandatory status (Y/N)
- SORT: Sorting variable number. Similar to Key field, populated only for sort fields and missing otherwise (Ex: 1)
- TYPE: SAS variable Type like Character, Numeric etc. (Ex: C)
- SASFMT: SAS variable format (Ex: e8601dt.)

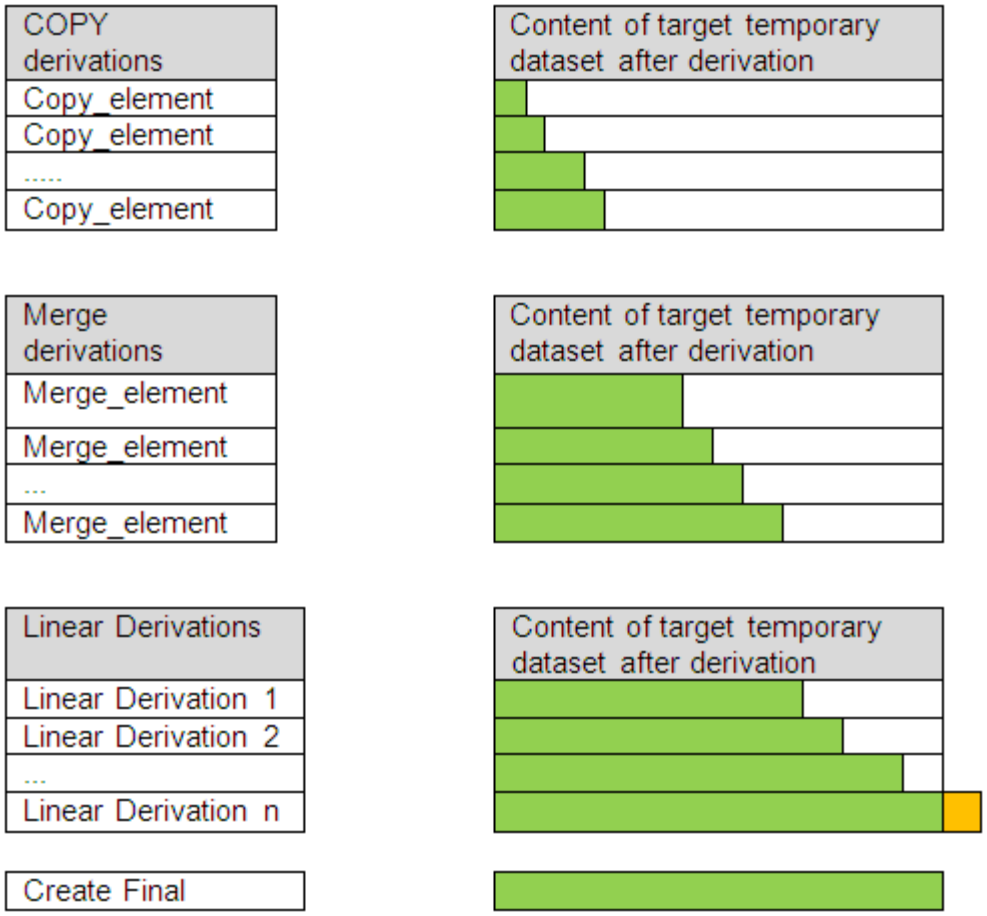
Creating SDTM+ / SDTM / ADaM datasets

- Following are three generic logical steps followed to create the datasets:
  - Get Main input data information (COPY_ELEMENT)
  - Merge information from secondary file(s) (MERGE_ELEMENT)
  - Apply existing derivation and/or add user code
- Standard template programs are created for applying it to a study. Some of the study specific domains like SE and EX have to be created by study teams (generic templates are available)
- For ADaM datasets, the variables SRCSEQ, SRCDOM, SRCVAR were included for the purpose of traceability between SDTM and ADaM
- Efficacy ADaM programs as chosen by Biostatistics were included along with the submission package

SDTM+ Templates

Template Structure

Derivation Part



General logic is to follow 3 main steps:

1 - Get Main input data information

2 – Merge information from secondary file(s)

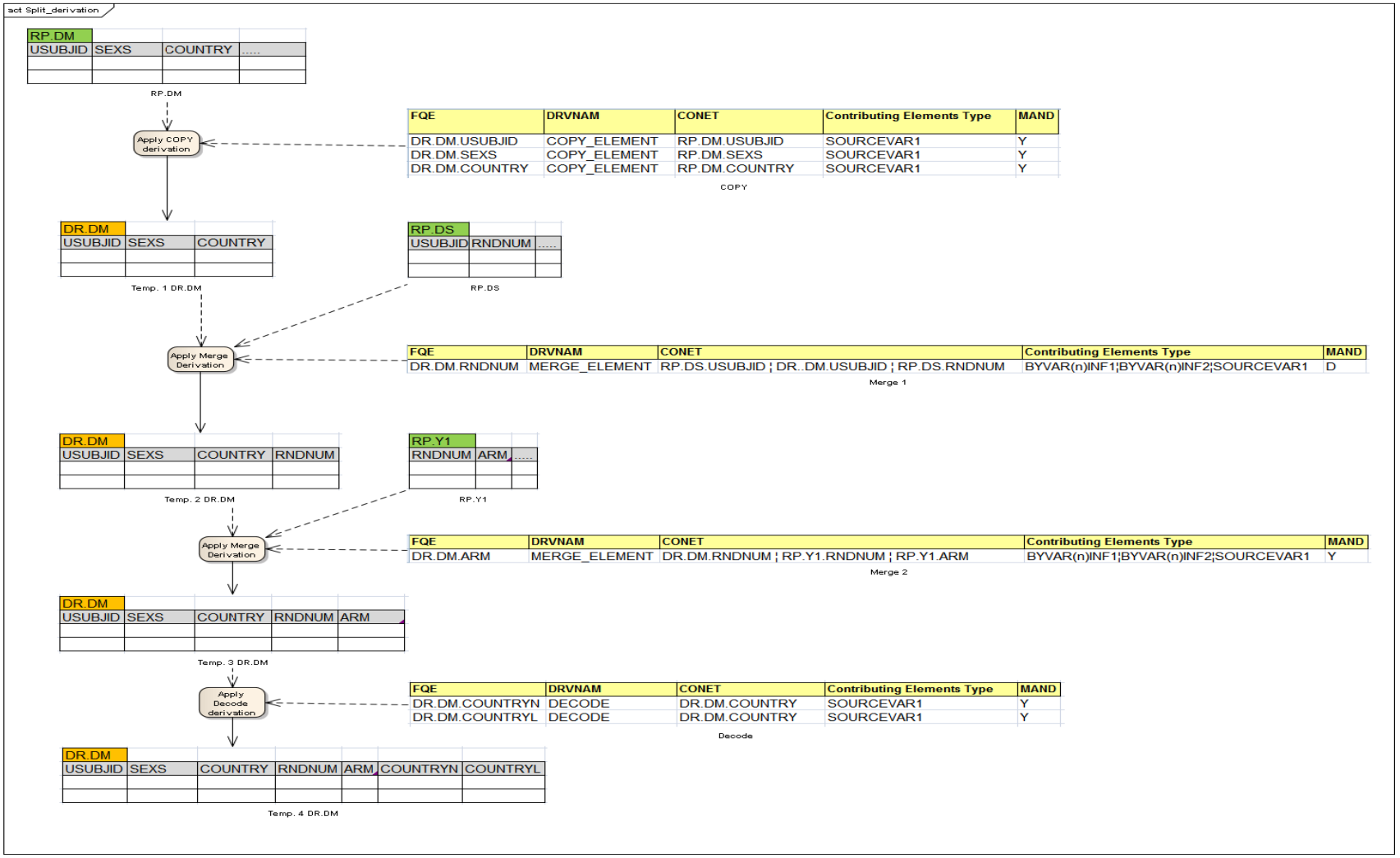
3 – Apply existing derivation and/or add user code

Keeping in mind that for variable dependency data flow reason step 2 and 3 can be repeated e.g. 1-2-3-2-3-2-3

NOVARTIS

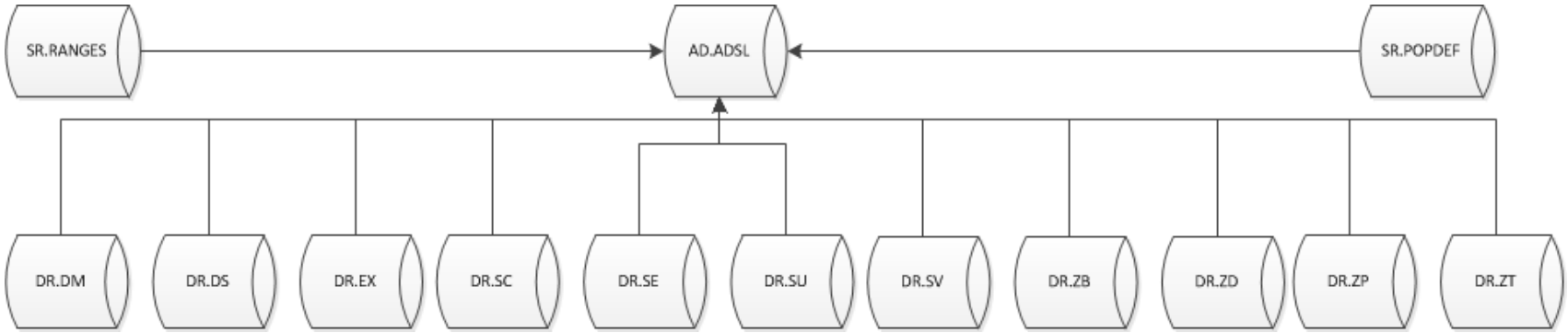
SDTM+ Templates

Derivation Process Overview



NOVARTIS

Which data files are needed for ADSL



- **Means, potentially 13 input data files are need to create ADSL**
- **RANGES and POPDEF are the reference files to define variable grouping and Analysis Set definition**

Controlled terminology re-mapping

- Open CDISC® validator was ran on SDTM+ data primarily to find the discrepancies in the Controlled terminology used (Novartis Controlled terminology).
- A re-mapping file was created and applied to the data for the code list items that could be mapped
- Traceability between SDTM and ADaM was maintained as the above SDTM+ data was the source for both of them
- Other Sponsor defined codelist items that couldn't be re-mapped were explained under Data Conformance Summary in the Study Data Reviewer's guide along with any other discrepancies in the data (reported by Open CDISC® validator)

Conclusion

- Creation of SDTM and ADaM using the standard templates and Metadata was effective since:
  - Time was spent only on the study specific custom derivations
  - For the future studies too, the data produced would be consistent between studies and compounds created in this standard approach
- There were lessons learned in terms of Metadata, Controlled terminology and Standard templates and were reported back to the respective functions
- Challenge would be to maintain Metadata, Controlled Terminology and other reference files with the latest CDISC standards

Questions?