Implementation of a harmonized, report-friendly SDTM and ADaM Data Flow

General by Marie-Rose Peltier

Experience by Marie Fournier

Groupe Utilisateurs Francophones du CDISC – 08/02/05

Framework of SDTM SDS

- SDTM SDS variables are mostly CRF variables. For report-writing and graphs, derived variables are lacking.
- SDTM SDS variables are mostly character variables. Numeric values are needed for sorting and analysis.
- Date variables are ISO dates. It is necessary to have numeric dates for the calculation of relative days, duration, etc.
- Population flags, drug variables and other common variables are not readily available for use in analysis.
- Some data structures are designed to facilitate standardization, and are not one PROC away.

Framework of ADaM

- CRF variables are further processed. It is hard to trace back to the source variables and to validate.
- Variable attributes can be inconsistent across domains.
- The lack of standardization makes it difficult to develop macros or programs for variables that are common to or similar in root construction for both SDTM and ADaM.

Where implementing SDTM ADaM ?

- **Logically SDTM created in CDM and ADaM in Biostatistics**

  Concerns

  - SDTM contains derived variables (Baseline Flags, Reference dates, Treatment Emergence Flag, Population Flags, ..)
  - SDTM proposed structure not efficient (ECG, Vital Signs)
  - Some SDTM domain contains data from different domains (ex disposition)
  - SDTM defines not all CRF variables (problem Suppqual)
  - Problem for specific TA domains

➔**Decided to create final SDTM and ADS in Biostatistics**

Summary of issues

- Derived variables should be created in ADaM first, and “back-populated” to SDTM SDS!
- Logically, derived variables should be created directly in SDTM SDS first!
- Same Variables in SDTM and ADaM REDUNDANCY !!
- SDTM only for submission : not sure whether study will be submitted.
- Priority Anaysis

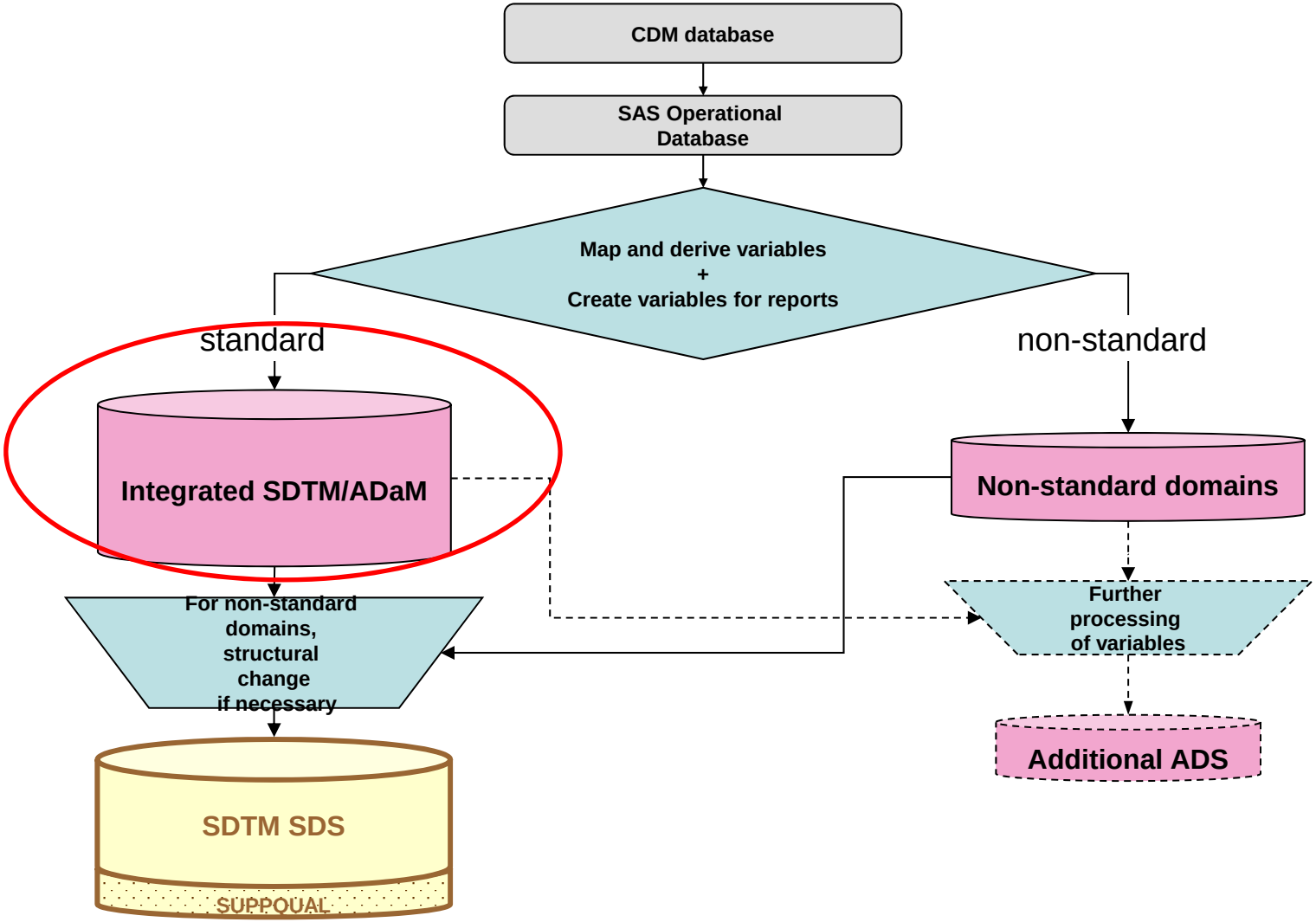
Solution

- **To resolve redundancy and process flow issues, integrate the same and yet separate SDTM and ADaM domains into single domains.**
- **Enhance the integrated domains for SDTM domains to complement the incompleteness of ADaM domains and vice versa:**
  - **SDTM SDS variables become more analysis-friendly and report-friendly**
  - **ADaM ADS variables become easier to validate and more structured**

The use of the integrated domains

- **Each integrated SDTM and ADaM domain is usable for:**
  - producing tables, listings and graphs,
  - creating SDTM SDS domains and adding SUPPQUAL and RELREC at submission,
  - creating additional domains which are dependent on the variables from the integrated domains. The additional domains are also used for producing tables, listings and graphs.

An Optimal SDTM/ADaM Data Flow

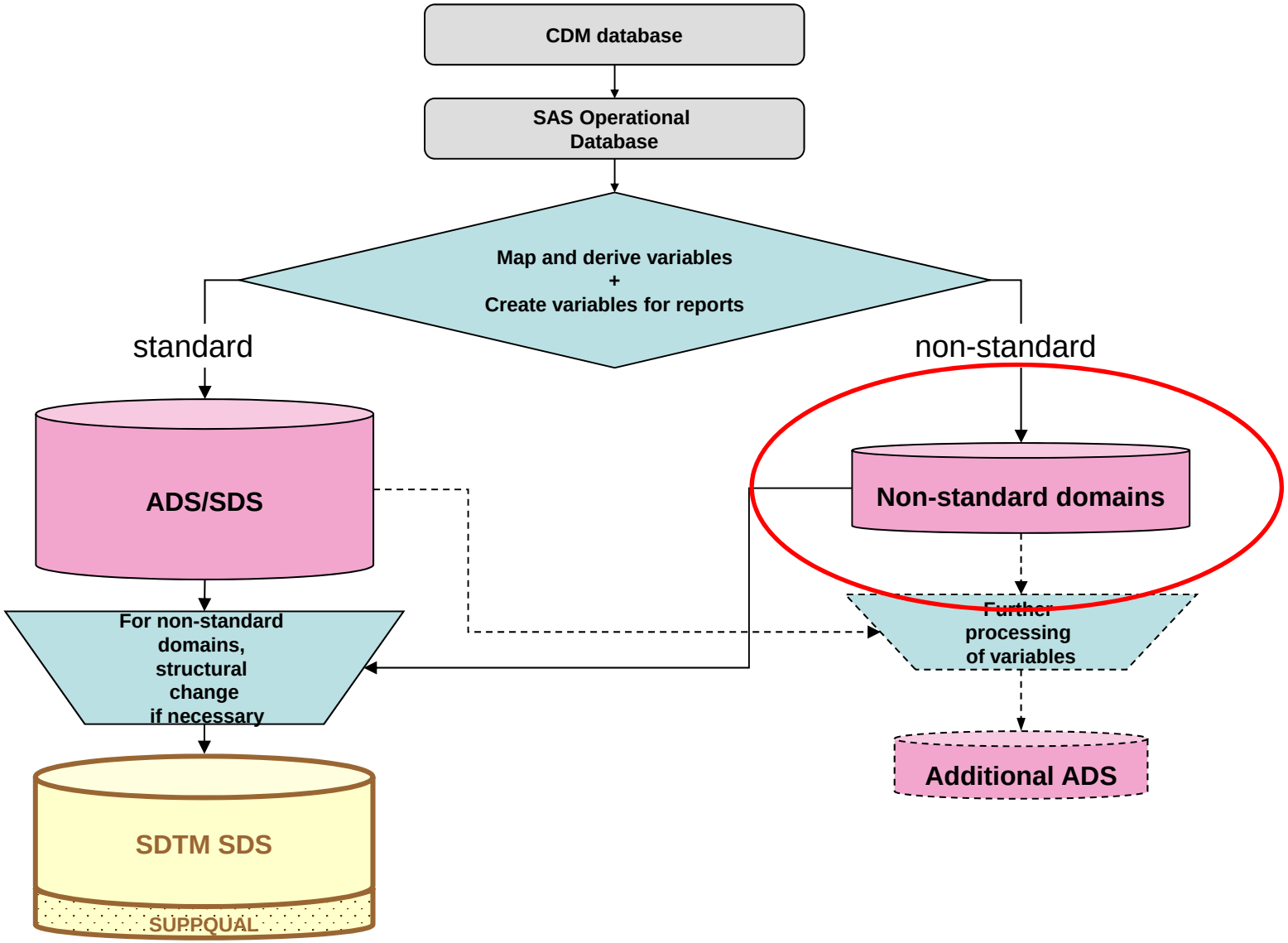


Note: The color pink stands for ADS and are used for reporting.
Both the ADS and the SDTM SDS/SUPPQUAL data are submitted.
The dotted line means when needed.

Data Flow Integrated SDS/ADS Domains

- **Mostly Safety data**
- **Ready for reporting**
- **Following the definition of the SDTM (Study Data Tabulation Model) principles, which is intended to provide a generalized model for submitting all types of data**
  - fit in one of the three SDS classes: interventions, events and findings (fixed structure)
  - Naming conventions
  - ISO8601 format (character)
- **Contains**
  - All CRF variables (SDTM naming conventions and attributes)
  - Common Variables
  - SDS standard derived variables
  - ADS standard derived variables
  - Code and decode (ex. SEX=F SEX**N**=1)
- **Add flag 'Target Dataset' : Variable ADS or SDS or Suppqual**

An Optimal SDTM/ADaM Data Flow

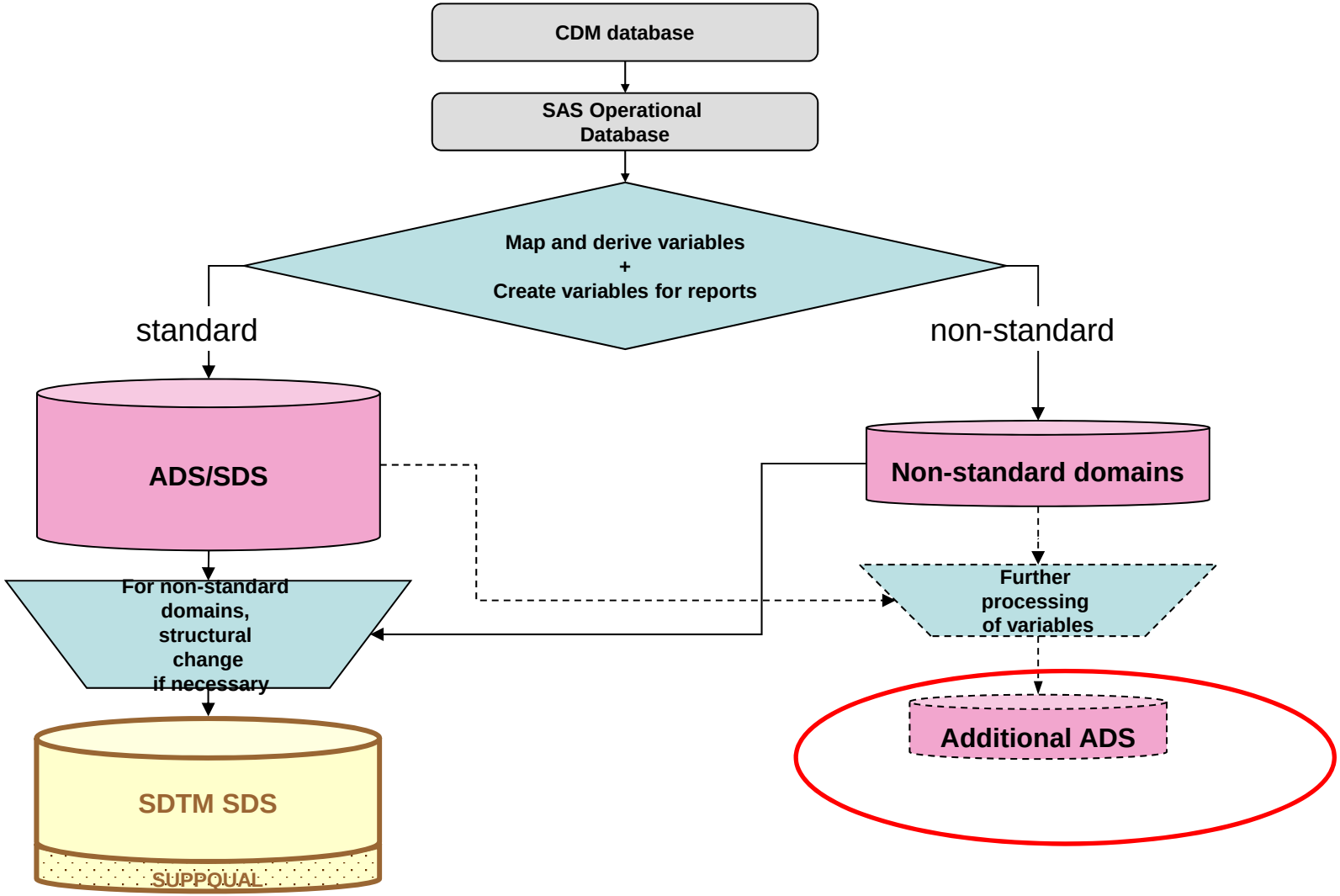


Note: The color pink stands for ADS and are used for reporting.
Both the ADS and the SDTM SDS/SUPPQUAL data are submitted.
The dotted line means when needed

Data Flow : Non Standard Domains

- Datasets that don’t fit in one of the CDISC SDTM structures (Interventions, Events, Findings)
- Ready for reporting
- The structure for analysis (reporting) is equal to the structure of the raw data (horizontal)
- Contains
  - All CRF variables
  - Common Variables
  - All analysis variables
  - ISO8601 format (character)
  - Code and decode (ex. SEX=F SEXN=1)

An Optimal SDTM/ADaM Data Flow



Note: The color pink stands for ADS and are used for reporting.
Both the ADS and the SDTM SDS/SUPPQUAL data are submitted.
The dotted line means when needed.

Data Flow : Additional Analysis Datasets

- **Data which directly support the assessment of primary and secondary study objectives**
  - **“Direct”: one statistical procedure away from statistical results**
  - **Usually highly derived (i.e., many computational steps) from CRT data**
- **Structure depends on Analysis (Ready for reporting)**
- **Contains**
  - **CRF variables needed for analysis**
  - **All analysis variables**

Advantages

- Analysis-ready, report-ready, SDTM-ready and ADaM-ready
  It is more efficient to have one set of data specifications that combines the requirements of SDTM, ADaM and report-writing than to have two separate sets of data specifications (one for SDTM SDS and one for ADaM) that have redundant variables.
- It is easier to create, maintain and update one set of data domains.
- There is no need to program, validate and crosscheck two sets of domains on the same topic. (one set of programs/macros)
- There is no need to break variables in domains apart into SUPPQUAL with RELREC for SDTM and to put them back again for report writing.
- Possibility to give priority to analysis if remapping to SDTM to complicated

Disadvantage

- **Need for extra step before submission**
  - Remap non standard domains to SDTM
  - Need to break variables to define relations (RELREC), and break variables into domain SUPPQUAL