

Mapping the Company's Legacy Data Model to SDTM

Nicolas Dupuis

Anja Feuerbacher

Bruce Rogers

Contents

- Background
- Overview of Mapping Tool
- Case 1: DSMB combining old and new data models
- Case 2: Final Analysis with Standard Reporting Tool
- Challenges
- QC strategies
- Conclusions

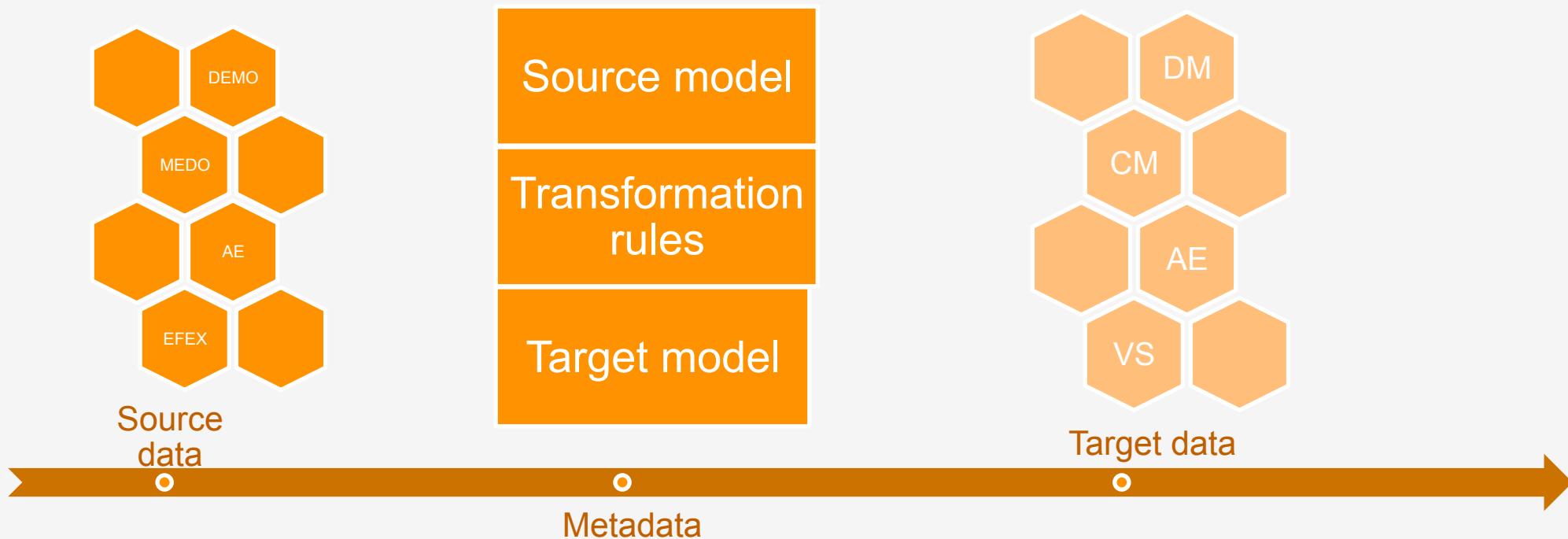
Background

- Historic data
 - "Generic Data Model" in use since ~1996
 - Fixed dataset structures, content not rigidly defined
 - Used for >90% of Roche studies (~10000 study instances)
 - Non-standard data sources (e.g. non-CRF data)
- New standards
 - CDISC SDTM standard
 - Roche implementation - SDTMv
 - Defined in centrally controlled metadata repository
 - Applies extended controlled terminology
 - Includes supplemental qualifiers within core domains

Overview of the Mapping Tool

The Vision:

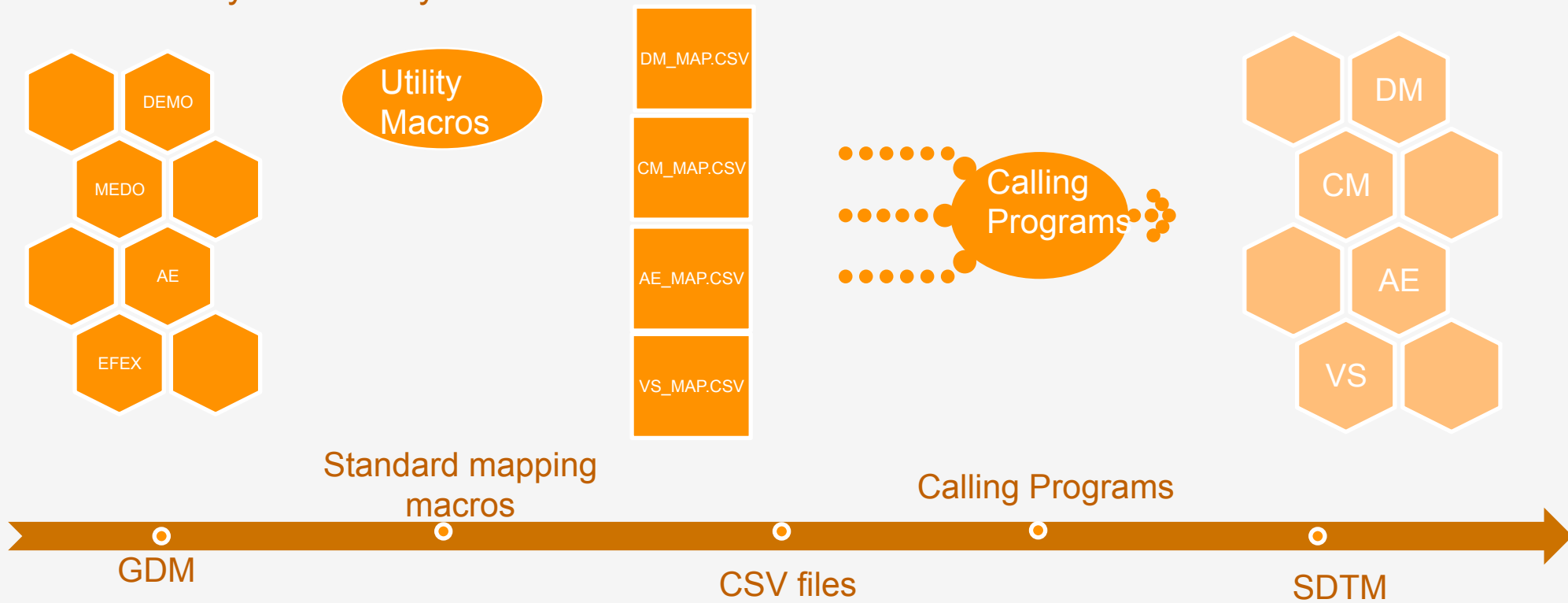
- Rule-based mapping
- Source and target data models, and transformation rules, in central metadata repository
- Complete traceability – one rule per target item
- Re-use of mapping rules, but allowing study-level variations



Overview of the Mapping Tool

Current state:

- Target data model and transformation rules in CSV files, one for each domain
- Some rules use standard utility macros
- Single rule per target domain data element
- One-way traceability



Overview of the Mapping Tool (Maptrans)

Transformation Rule Features

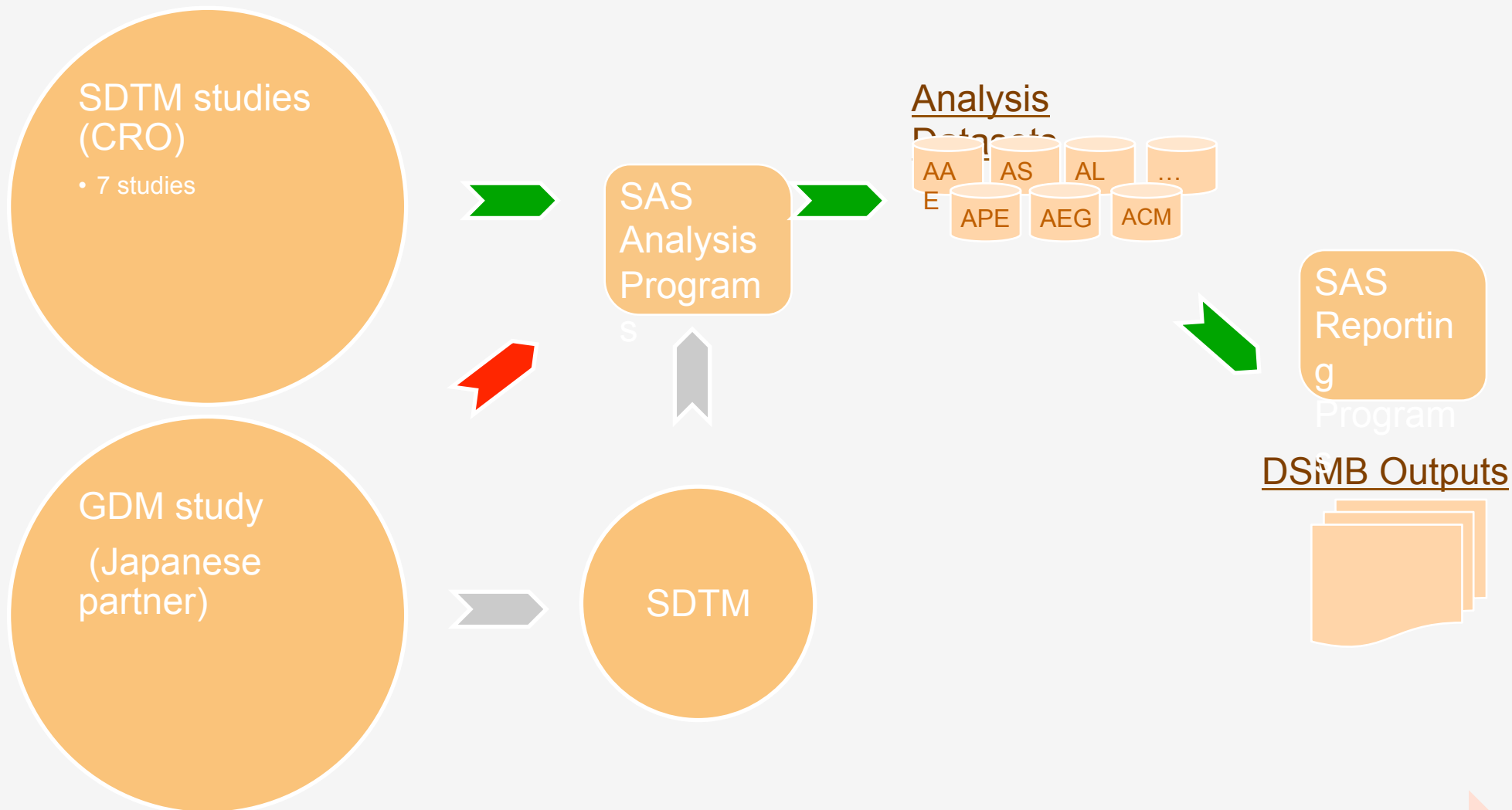
- Rules parsed to create executable SAS code
- Primary source domain per target domain
- Secondary domains will generate SQL joining
- Implied sequence of derivations, allows for temporary variables to be created & reused.
- Repeating variable groups to create multiple output records

Overview of Mapping Tool, cont.

➤ An example of a CSV file:

TMODEL	TDOMAIN	Variable Name	Label	Type	Length	FORMAT	SMODEL	SDOMAIN	SVARIABLE	Formula/Keys
SDTM	DM	STUDYID	Study Identifier	Char	200		GDM	DEMO	PROTO	
SDTM	DM	DOMAIN	Domain Abbreviation	Char	2					"DM"
SDTM	DM	USUBJID	Unique Subject Identifier	Char	50		GDM	DEMO	proto crtn PT	%GDM2SDTM_SUBJID
SDTM	DM	SUBJID	Subject Identifier for the Study	Char	50		GDM	DEMO	PT	
SDTM	DM	RFSTDTC	Subject Reference Start Date/ Time	Char	19		GDM	DEMO	TRT1DC TRT1TC	%dc2iso(TRT1DC) 'T' trim(TRT1TC)
SDTM	DM	RFENDTC	Subject Reference End Date/ Time	Char	19					""
SDTM	DM	SITEID	Study Site Identifier	Char	200		GDM	CENT	CTCNUM	proto crtn
SDTM	DM	BIRTHDT	Date/Time of Birth	Char	19		GDM	DEMO	BIRTHDT	
SDTM	DM	AGE	Age	Num	8		GDM	DEMO	AGE	
SDTM	DM	AGEU	Age Units	Char	200					"YEARS"
SDTM	DM	SEX	Sex	Char	1		GDM	DEMO	SEX	
SDTM	DM	RACE	Race	Char	200	\$race	GDM	DEMO	RACE	
SDTM	DM	ETHNIC	Ethnicity	Char	200					"NOT REPORTED"
SDTM	DM	ARMCD	Planned Arm Code	Char	20		GDM	DEMO	RNDGRP	
SDTM	DM	ARM	Description of Planned Arm	Char	200		GDM	DEMO	RND	
SDTM	DM	COUNTRY	Country	Char	3		GDM	CENT	CTCNTRY	proto crtn
SDTM	DM	CRTN	Country	Num	8		GDM	DEMO	CRTN	

Case 1: DSMB combining GDM and SDTM data

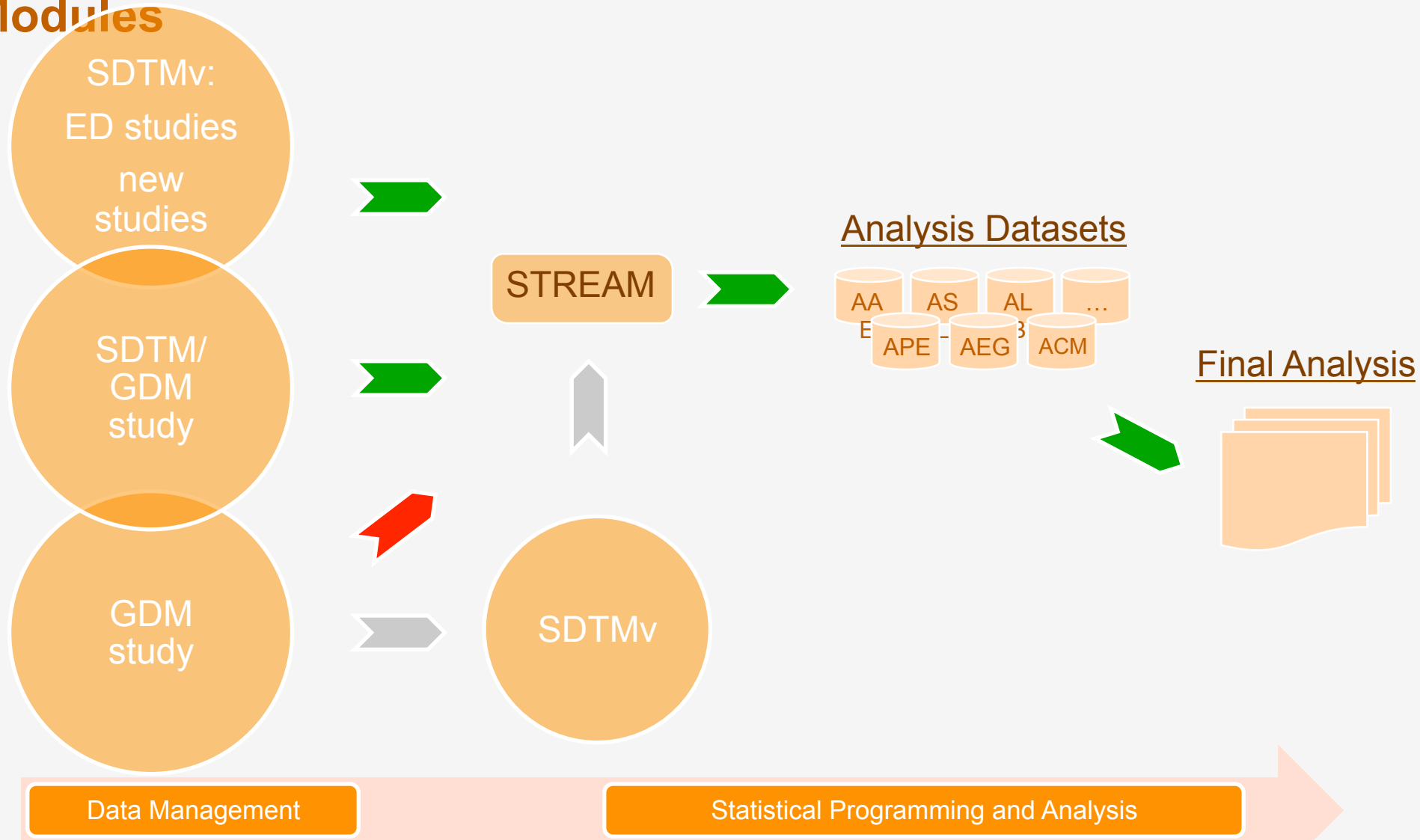


Data Management

Statistical Programming and Analysis

CRO

Case 2: Final Analysis Combining GDM and SDTMv data and using new Standard Reporting and Analysis Modules



Challenge: Implementing the Data Standards

- Many domains do not map directly, or entirely, from a single source to a single target
 - One to many, many to one, and many to many relationships.
 - Example: Adverse Events of special interest, stored in three GDM domains, used in three target (SDTMv) domains, but with elements from each source used in each target.
- Controlled Terminology.
 - GDM usually contain the source values from the CRF.
 - Recoding required when these deviated from the respective SDTM code lists.
 - In many cases, the recoding could be applied simply in the rule, e.g. values “YES” and “NO” would become “Y” and “N”.
 - In some cases, however, a pragmatic decision had to be taken in conjunction with statisticians and clinical scientists.

Challenge: Fitting a banana skin on an apple, using a Japanese manual.

TMODEL	TDOMAIN	Variable Name	Label	Type	Length	FORMAT	SMODEL	SDOMAIN	SVARIABLE	Formula
SDTM	QS	QSTESTCD	Question Short Name	Char		8\$QSTESTCD	GDM	EFEX	EFPARM	

How does it look in :

Japan
GDM.EFEX.EFPARM

WISHDEAD

GDM.EFEX.EFPARM

"SUICIDAL
IDEATION"



CRO
SDTM.QS.QSTESTCD

CSI01

SDTM.QS.QSTEST

"Wish to be Dead"

What does the aCRF say :

CRFV1.10_ECV0.01: Form_PDF_Annotated

プロジェクト名: JN25535

フォーム: C-SSRS-ベースライン評価

Instance Code%

評価日

自殺念慮

1. 死んでしまいたいという願望

EFEX.EFPARM='WISHDEAD'

はい ①
いいえ ②
EFEX.EFVAL



Translate

From: Japanese



To: English

Translate

English Spanish Japanese

English Spanish Arabic

死んでしまいたいという願望

Wish that I want to be dead

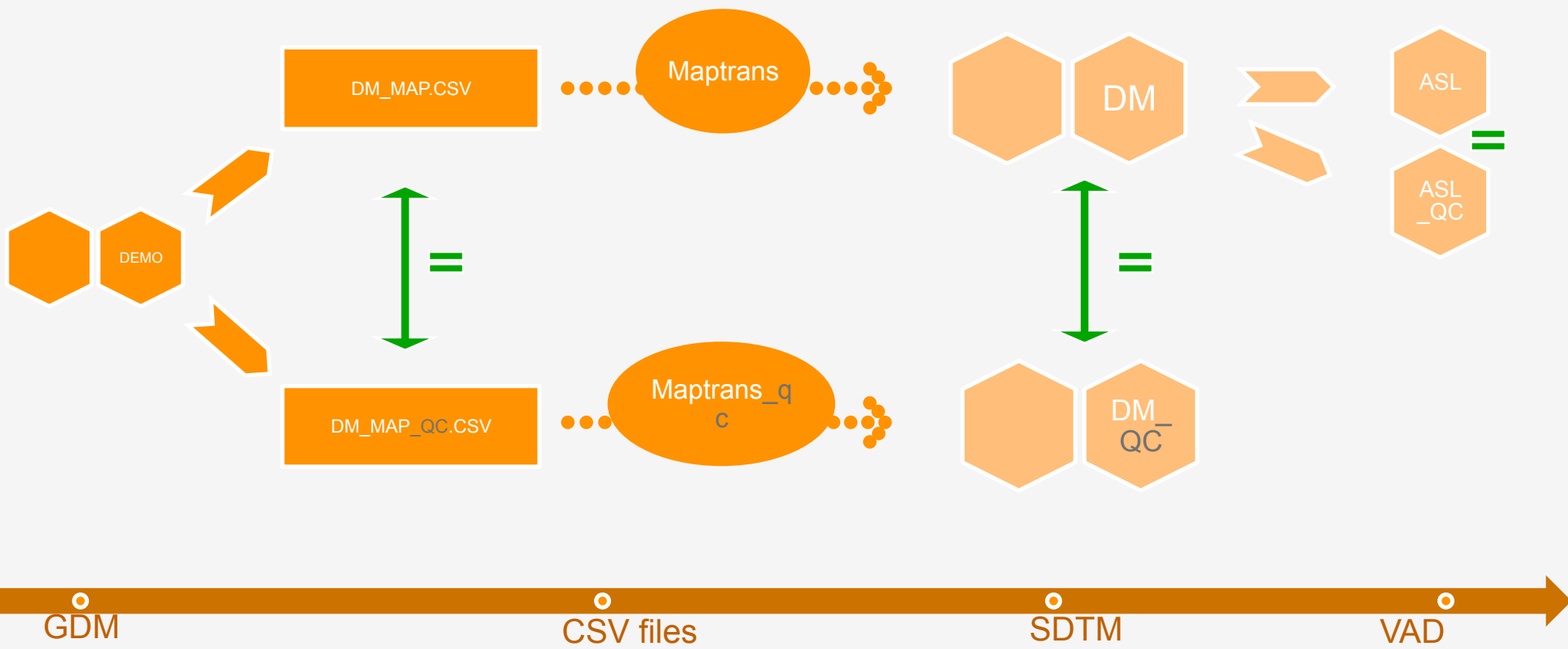
Maptrans uses SAS Formats to transform the text

Challenge: Using the Mapping Tool

- Constraints of Tool
 - Secondary domain mappings not supported for repeating variable blocks
 - Unique keys required for secondary domains tool does not check that supplied keys guarantee uniqueness.
 - Maximum of nine repeat blocks allowed.
- Due to these constraints and time pressures, manually-created code was used for some domains in Case 2.
- Future modifications to toolset will address these issues

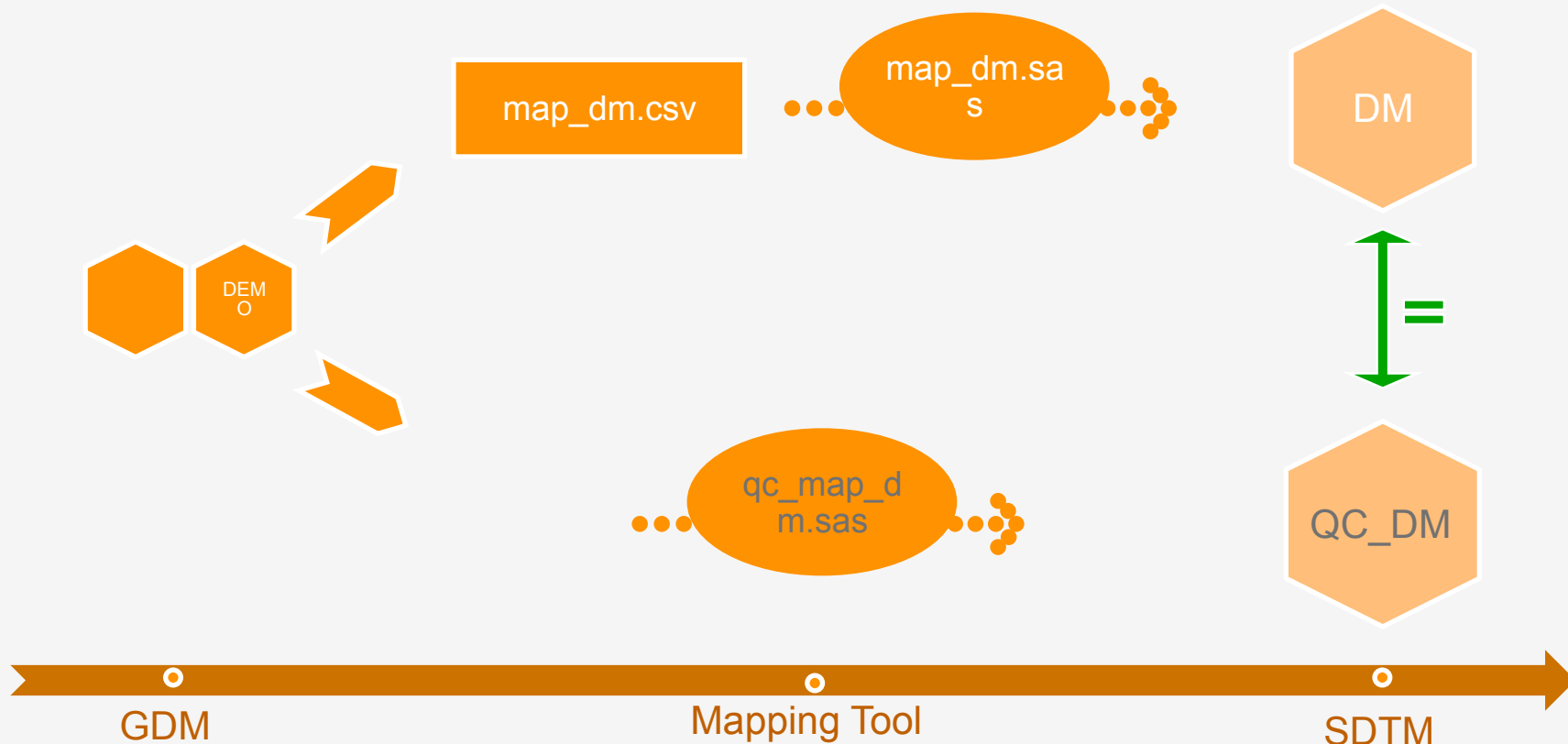
QC Strategies

- Case 1 (DSMB): CSV files and calling programs are created independently by two people and compared. The analysis datasets programs are double programmed from the primary SDTMs.
- Provided useful feedback to ensure sufficient scope of the mapping rules.



QC Strategies

- Case 2 (STREAM): Review of mapping specifications, double programming and proc compare.
- Tool new and not fully validated at outset, so more detailed QC was felt to be justified



Conclusions

- No standard transformations were available, in both cases virtually all mappings needed to be tackled from scratch
 - For future studies, we now have a good base set of mappings for most SDTM(v) domains
 - Developing these for a given new study should then involve minimal effort
- Real advantages already gained for
 - Review, change and validation
 - Transparency and ease of re-use
 - Defined framework for mapping effort
 - Document reduction

The Future is Bright

- Central metadata repository (GDSR)
 - Complete SDTMv data model defined
 - Governance model well established
 - Tools to surface GDSR metadata for program use under construction
- Toolset now needs to be further developed, including
 - Create 'Standard' transformation rules
 - Store rules in central repository
 - Allow study-specific rules to override base set
 - Better support for complicated transformations
 - Improved user interface



We Innovate Healthcare