

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
- Conclusions

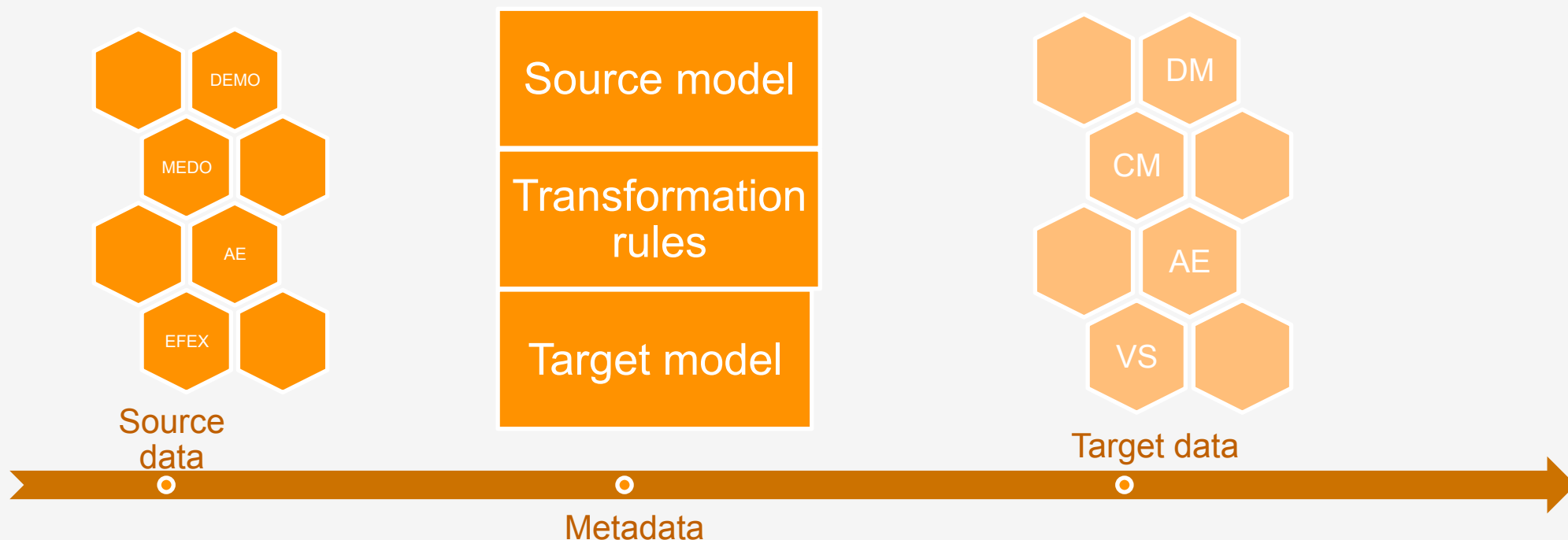
Background

- Historic data
 - "Generic Data Model" in use since ~1996
 - Fixed dataset structures, content not rigidly defined
 - Used for >90% of Roche studies until recently
 - 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 supp. quals. within core domains

Overview of the Mapping Tool

The Vision:

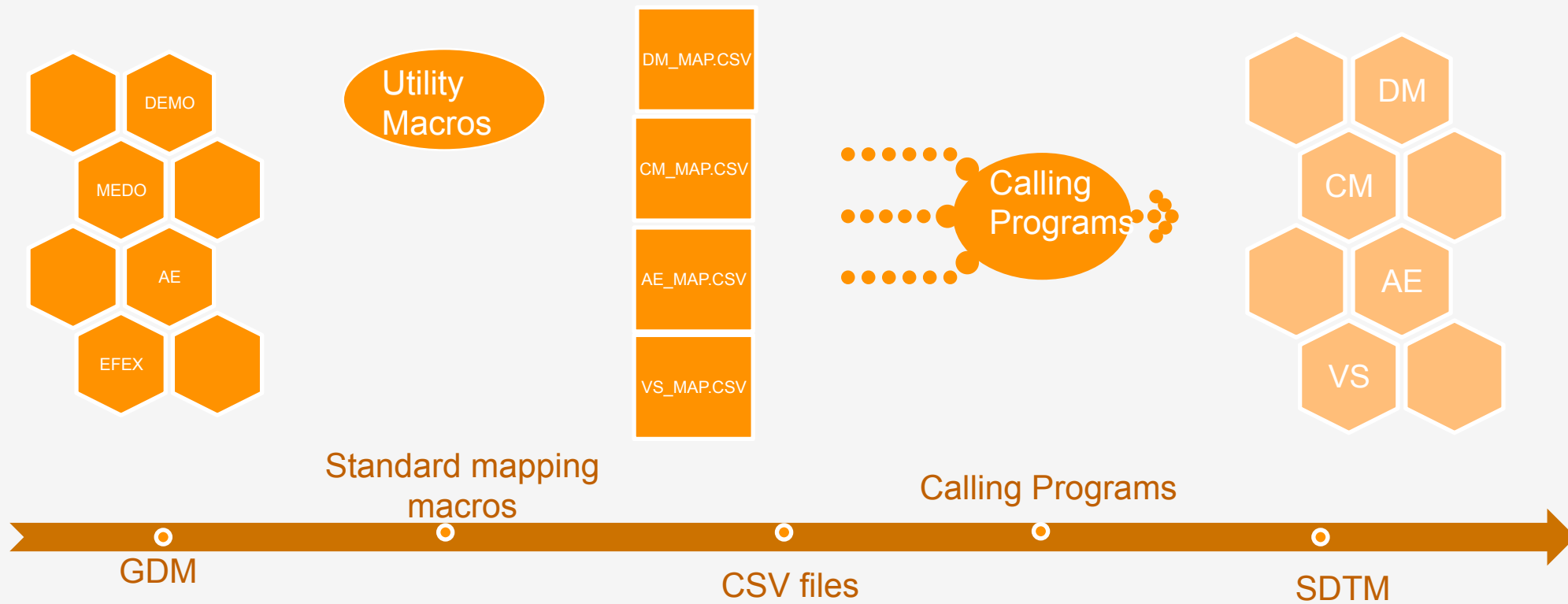
- Rule-based mapping
- Source and target data models, and transformation rules, in central metadata repository
- Complete traceability – one rule per target item
- Re-use of mapping rules, but allowing study-level variations



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

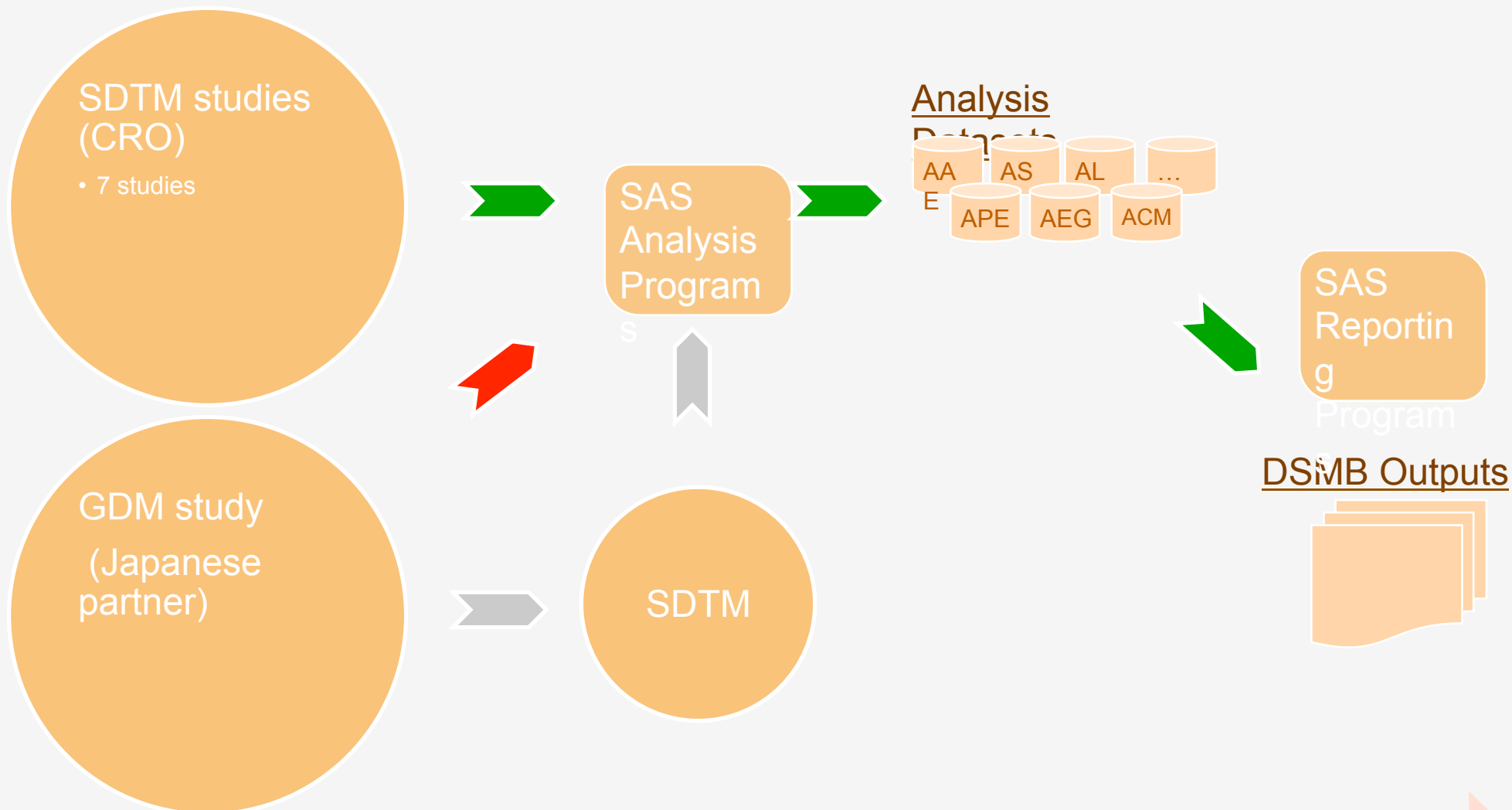
- Rules parsed to create executable SAS code
- Primary source domain per target domain
- Secondary domains will generate SQL joining
- Rules may include
 - simple (variable name) assignments,
 - SAS formulae,
 - SQL statements
 - Literal values
 - SAS formats,
 - conditional assignments ('case')
 - Repeating groups (e.g. for multiple parameter/values)

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 crt n
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 crt n
SDTM	DM	CRTN	Country	Num	8		GDM	DEMO	CRTN	

Case 1: DSMB combining GDM and SDTM data



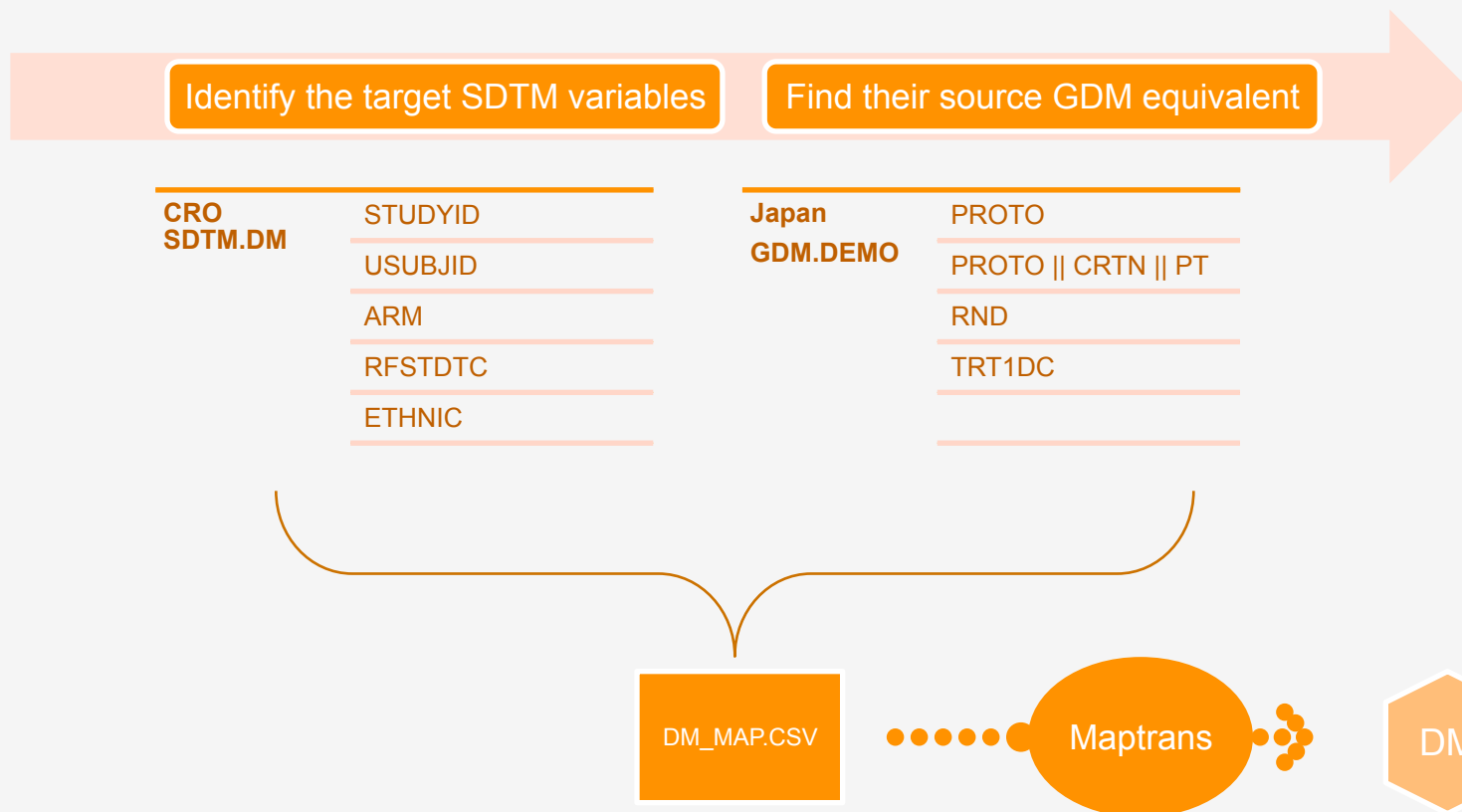
Data Management

Statistical Programming and Analysis

CRO

Case 1, cont.

- The idea was to create ‘Japan’ SDTMs as close as possible to the SDTMs from the CRO.
- One CSV file for each target domain. One observation for each target SDTM variable.



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'

EFEX.EFVAL

はい

いいえ

1

2

4

How did we get a confirmation :



Translate

From: Japanese



To: English

Translate

English

Spanish

Japanese

English

Spanish

Arabic

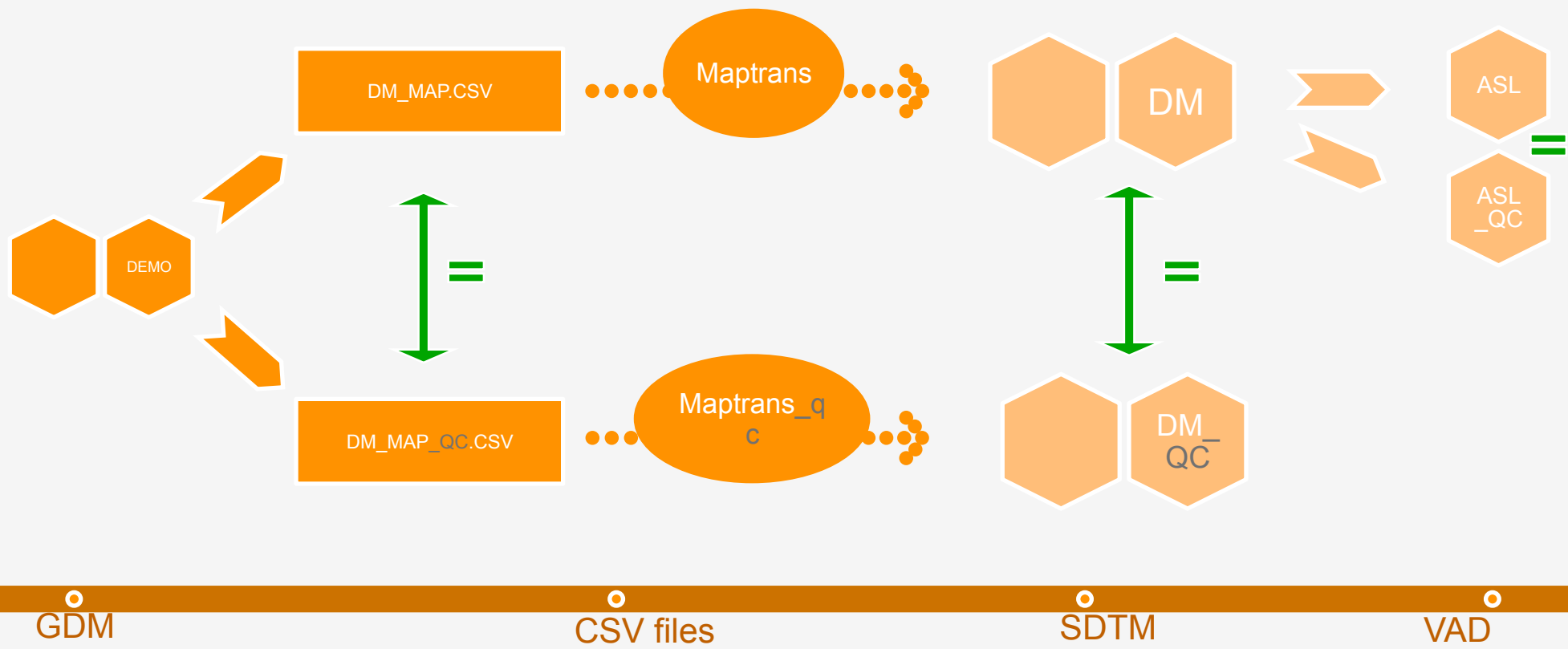
死んでしまいたいという願望

Wish that I want to be dead

Maptrans uses SAS Formats to transform the text

QC Strategy

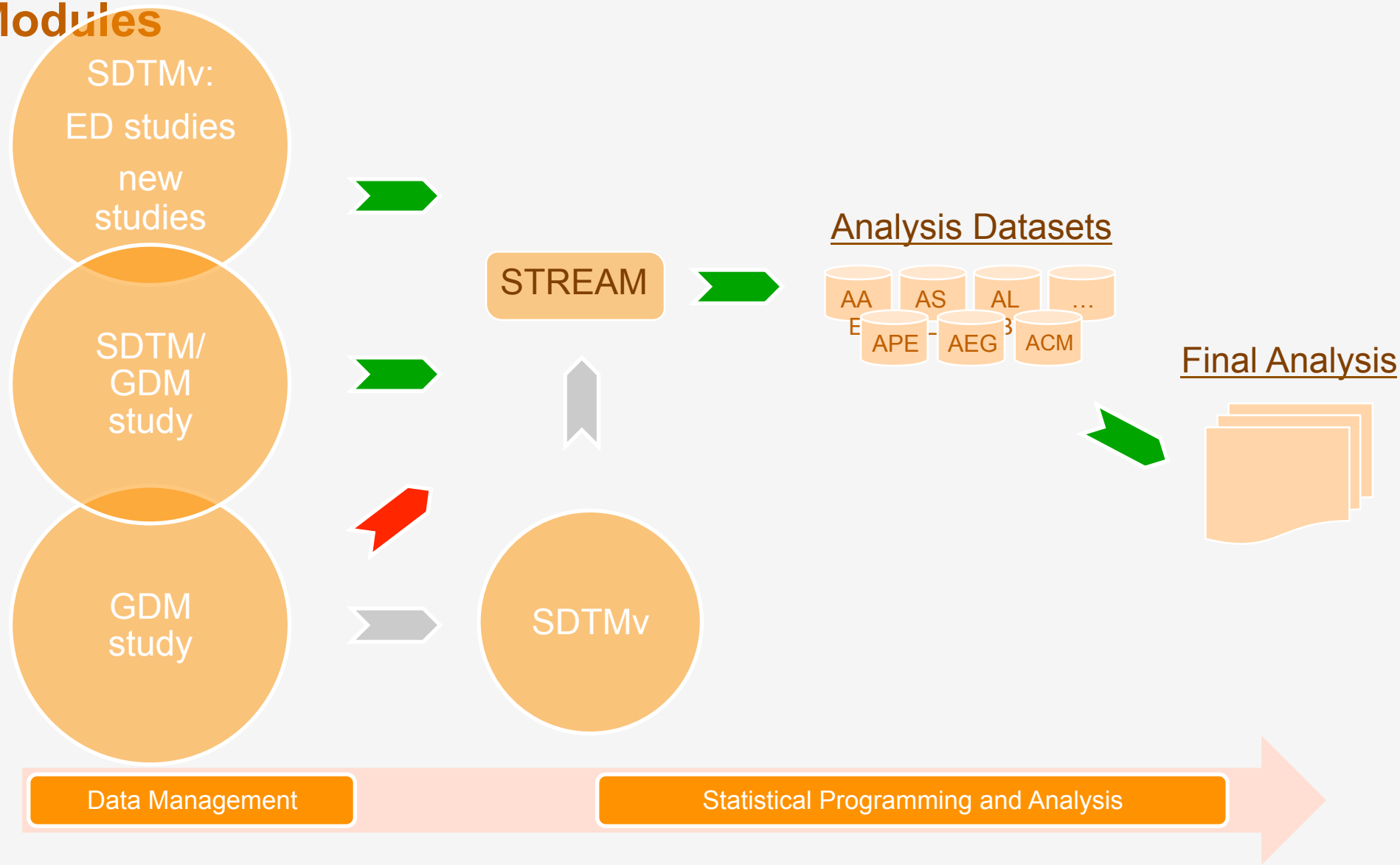
- CSV files are created independently by two people and compared.
- Mapping programs called separately for each CSV: results from each then compared
- The analysis datasets programs are double programmed from the primary SDTMs. Provided feedback to ensure sufficient scope of the mapping rules



Effort Estimate

- Because the goal was to create analysis datasets for DSMB, not all the domains were needed. We mapped 8 SDTM domains
- Overall it took around 6 weeks, 2 programmers 50% part-time
- Initial learning curve much steeper
- Significant savings on validation

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

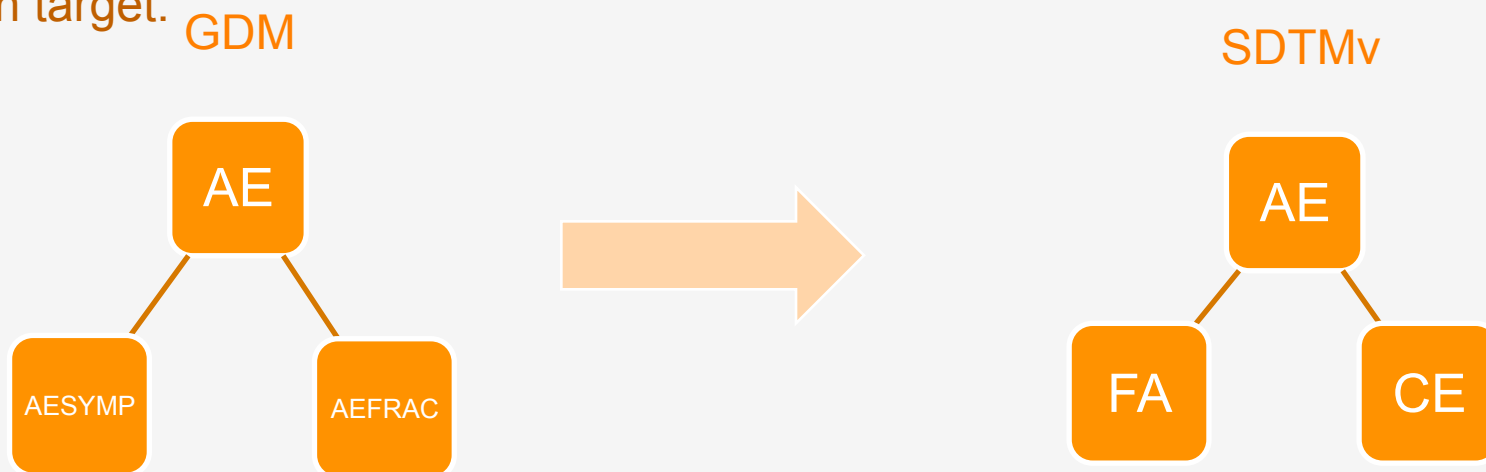


Challenges: Using the Mapping Tool

- Tool sometimes not straightforward
 - Problems using multiple inputs for repeating variable blocks (i.e. parameters/values)
 - Needs effort to correctly design merging of primary and secondary input datasets
- Due to constraints of the current version of the tool and time pressures, pre-existing manually-created code was used for some domains
- Example: SDTMv domain DS
 - Input: 5 GDM datasets , 1 external csv file
 - 12 parameters per patient plus information for screen failures

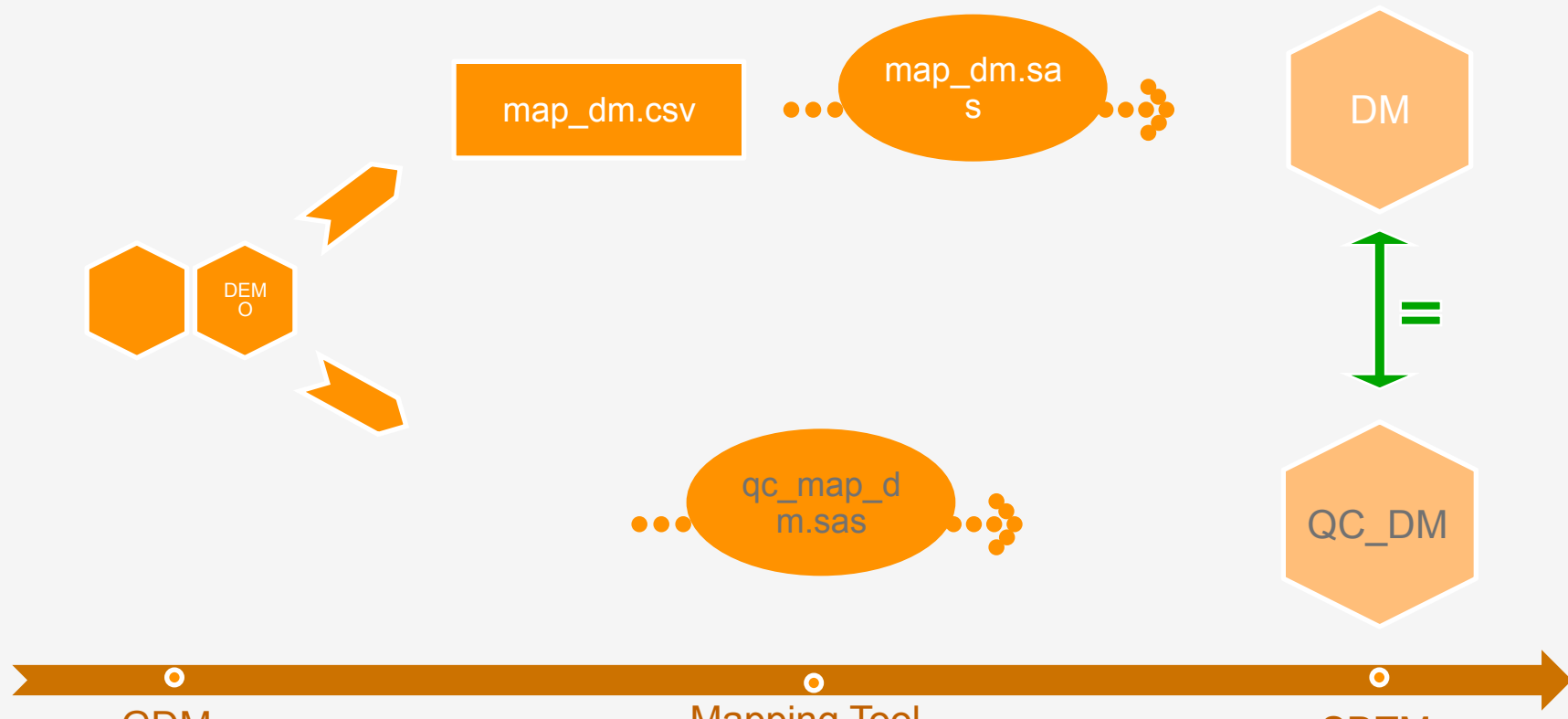
Challenge: Implementing the Data Standards

- Mapping sometimes not straightforward
- Requirement of reporting tool that content is standard (Controlled Terminology)
 - Different terminologies used in source and target
 - Values therefore need to be re-mapped.
- Many domains do not map directly, or entirely, from a single source to a single target
 - 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.



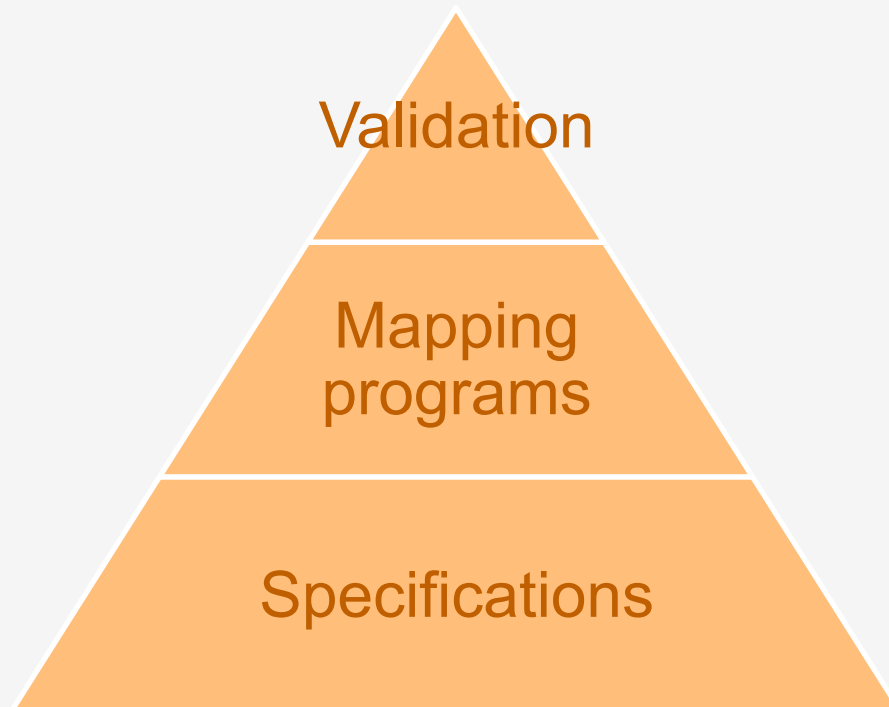
QC Strategy

- Validation of SDTMv domains:
 - Review of mapping specifications with respect to GDSR and SDTM IG
 - Double programming and comparison with Proc Compare
 - Tool new in this context and not fully validated at outset
 - Some 'QC' programs already existed, so a good test of mapping rules



Effort Estimate

- Started mapping of GDM to SDTMv in November 2011
- Study team: 1 full-time new-starter, 2 part time experienced programmers
- Other tasks had to be done
- Mapping and validation completed in June 2012



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
 - Specification entered directly into CSV and SAS formats
 - Self-documenting - no other program code used.
 - Human-readable CSV file is both specification and 'mapping' documentation

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