Team building through legacy data conversion

Lieke Gijsbers, OCS Life Sciences, the Netherlands
Jasmine Kestemont, Innovion, Belgium

PhUSE Single Day Event, Beerse – 11 September 2019

Outline

- Project Scope
- Project Approach
  - To conduct conversion in an efficient and consistent way
  - To keep the team engaged
- Using team diversity to improve problem resolution

OCS Life Sciences Innovion

Project scope

- Prepare study portfolio for submission to Food and Drug Administration (FDA)

- Conversion clinical trials to SDTM and ADaM standards
  - First study from 2006, last study currently in set-up

1

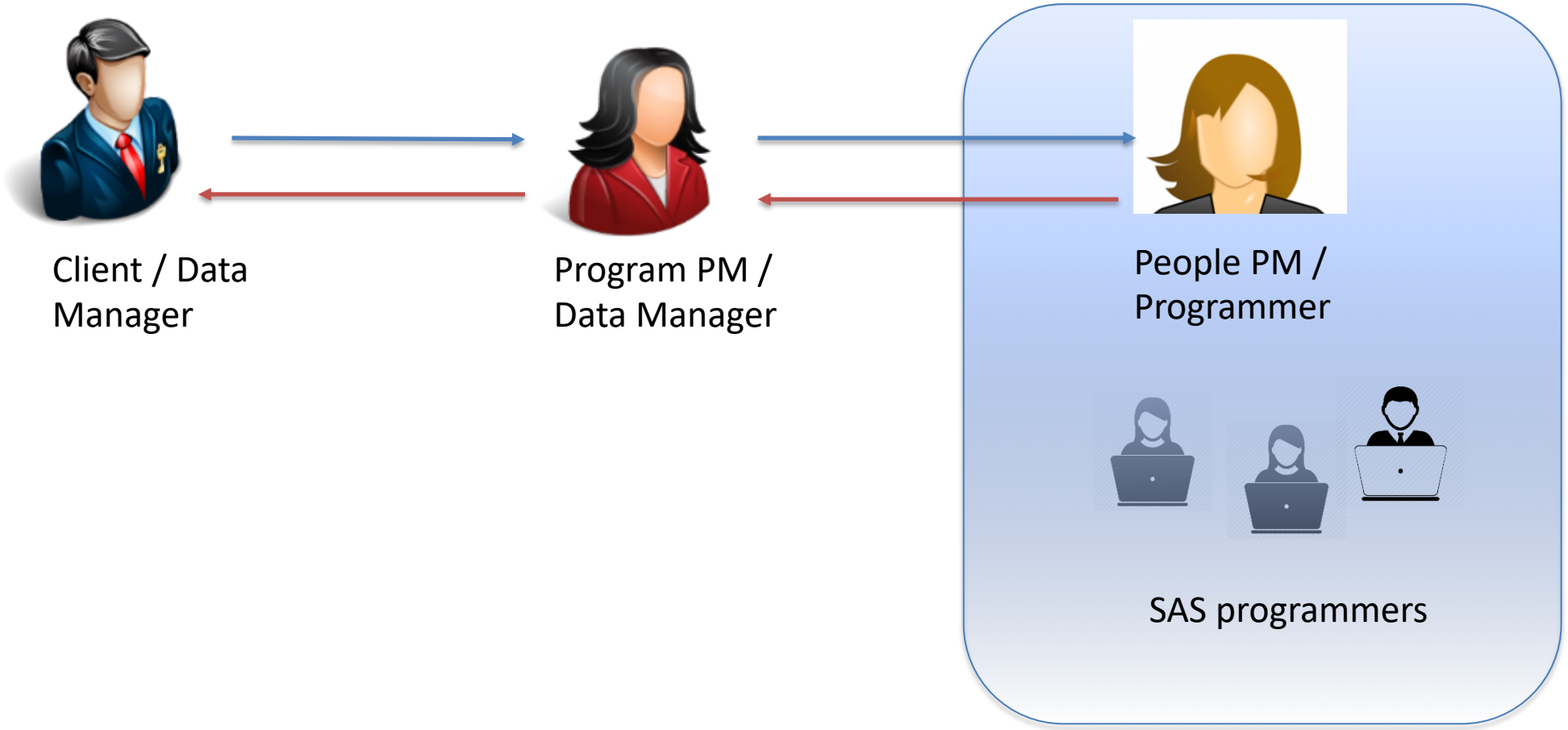
- Pilot project
- Conversion of a single phase II trial to SDTM and ADaM, and reproduction of tables

2

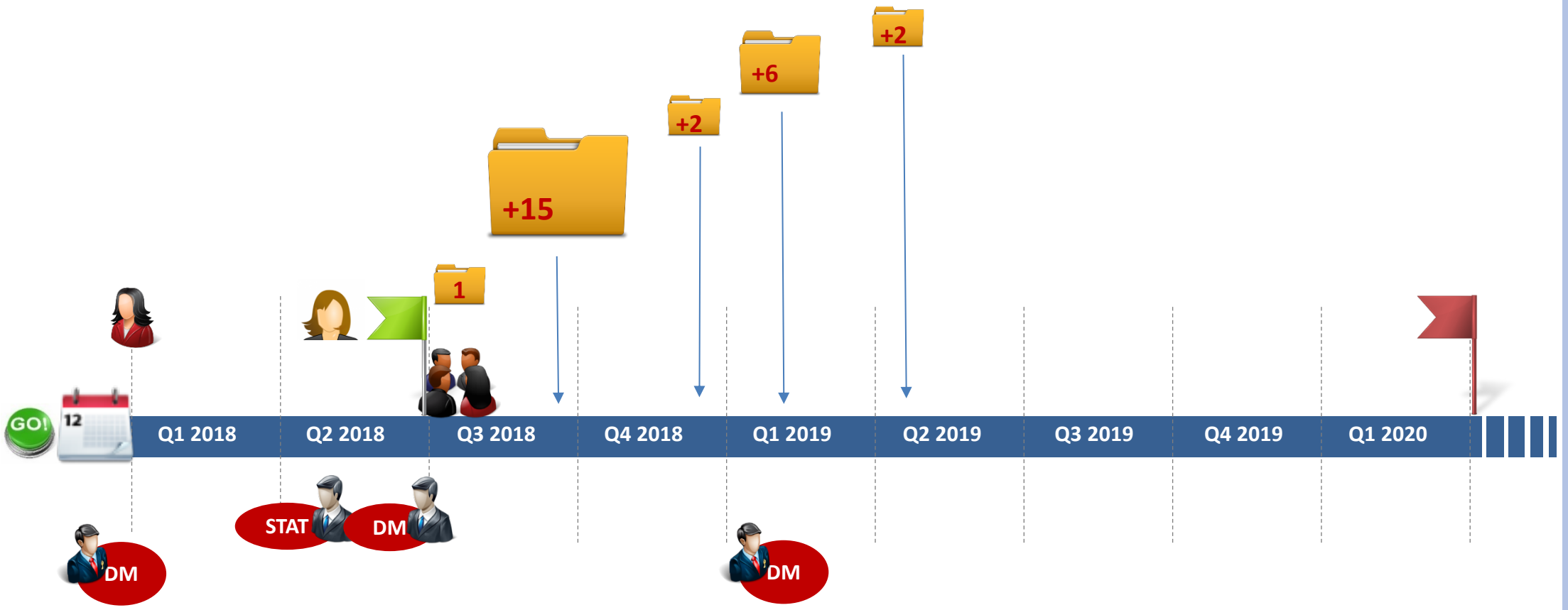
- Conversion of 11 phase I trials to SDTM
- Conversion of 4 phase II trials to SDTM and ADaM, and reproduction of tables

OCS Life Sciences Innovion

Project start – team composition



Project start – timelines and scope



Project goals and challenges

- Project goals
  - Deliver on time, on budget, with quality!
  - Ensure consistency across all trials
  - Scale up team
  - Onboard junior staff with limited experience
  - Keep the team motivated

- Project challenges
  - New team
  - New client
  - Complex Data Structures – dataset containing **10306** variables
  - Multiple CROs
    - But, data from same CRO had similar data structure

OCS Life Sciences Innovion

Project start – approach

1

- Pilot: conversion of a phase II trial to SDTM and ADaM, and reproduction of tables
- Time frame: 5 months
- Set standards for all subsequent trials

- Team
  - 1 programmer >15 year experience
  - 2 programmers with little to no experience

- Approach
  - Team member responsible for certain domains

OCS Life Sciences Innovion

Project continuation – approach

2

- Conversion of 20 phase I trials to SDTM
- Conversion of 5 phase II trials to SDTM and ADaM and reproduction of tables
- Time frame: 1 year

- Team
  - Expanded with OCS Life Sciences' ex-trainees
    - Learning by doing

- Adjusted approach
  - Team member responsible for a complete study rather than certain domains

OCS Life Sciences Innovion

Efficient and consistent conversion

- Tools available, but also developed and enhanced by team members
- Reduce repetitive tasks: focus on the data

Pre-analysis

- Determine similarities between studies in terms of design, aCRF and source data

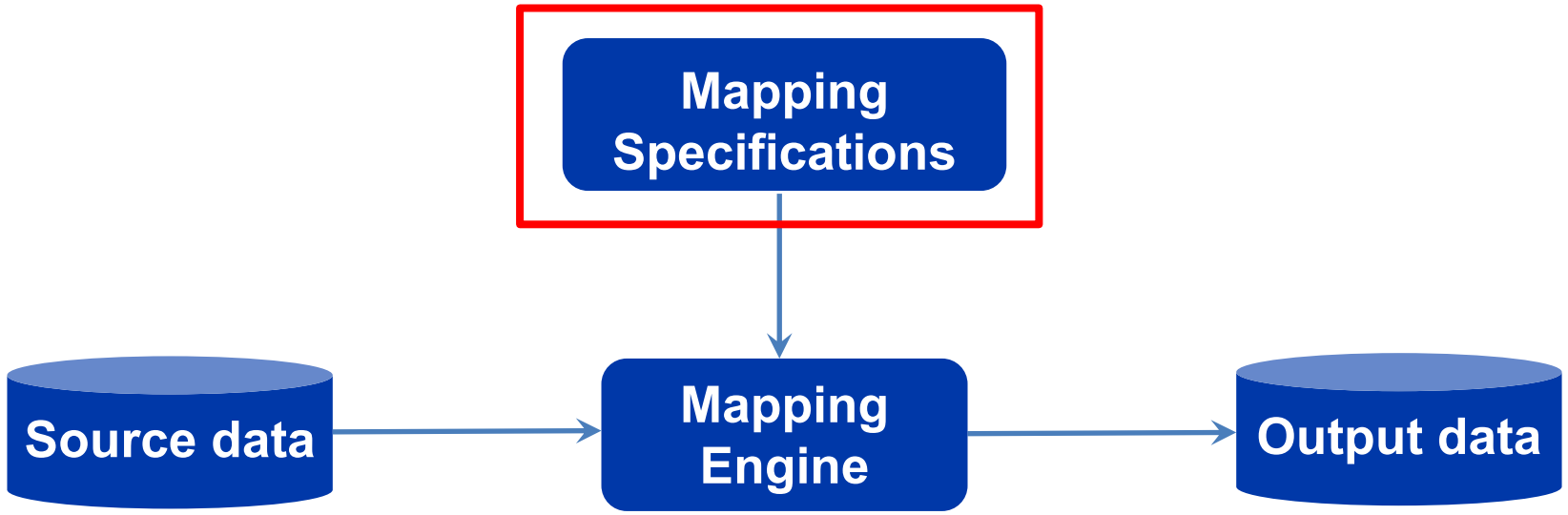
Dataset	Variable	Label	Data Type	Format	STUDY 1	STUDY 2	STUDY 3	STUDY 4	Total
AE	SUBJID	Subject's No	text		Y	Y	Y	Y	4
AE	PGNO	Page No	integer		Y	Y	Y	Y	4
AE	AESEQ	Adverse event No	text		Y	Y	Y	Y	4
AE	AESTATUS	Adverse event	integer	YESNO.	Y		Y	Y	3
AE	AETERM	Description	text		Y	Y	Y	Y	4
AE	SOCAE	SOC MedDRA code	text		Y	Y	Y		3
AE	SOCAE_	SOC MedDRA term	text		Y	Y	Y	Y	4
AE	HLGTAE	HLGT MedDRA code	text		Y	Y	Y		3
AE	HLGTAE_	HLGT MedDRA term	text		Y	Y	Y	Y	4
AE	HLTAE	HLT MedDRA code	text		Y	Y	Y		3
AE	HLTAE_	HLT MedDRA term	text		Y	Y	Y	Y	4
AE	PTAE	PT MedDRA code	text		Y	Y	Y		3
AE	PTAE_	PT MedDRA term	text		Y	Y	Y	Y	4
AE	LLTAE	LLT MedDRA code	text		Y	Y	Y		3
AE	LLTAE_	LLT MedDRA term	text		Y	Y	Y	Y	4
AE	AEEXP	Expected/unexpected	integer	AEEXP.			Y	Y	2

  - Resource allocation
  - Efficient conversion

	STUDY 1	STUDY 2	STUDY 3	STUDY 4
STUDY 1	100			
STUDY 2	97	100		
STUDY 3	90	92	100	
STUDY 4	82	85	92	100

Mapping engine

- SDTM datasets created using mapping engine developed by OCS Life Sciences



Mapping engine

- One Excel spreadsheet containing all:
  - Source and target variables
  - Mapping specifications
  - Translation of the specifications into SAS (pseudo-)code needed to generate the target variables

SOURCE_DS	SOURCE_VAR	TARGET_DS	TARGET_VAR	SPECIFICATION	FUNCTION
SOURCE.AE	AE	AE		Keep only records where an adverse event occurred	WHERE[AE=1]
SOURCE.AE		AE	STUDYID	Assign "GFT505-111-7-1"	ASSIGN[GFT505-111-7-1]
SOURCE.AE		AE	DOMAIN	Assign "AE"	ASSIGN[AE]
SOURCE.AE	IDSTUD			Not mapped	NOT MAPPED
SOURCE.AE	IDSUBNUM	AE	USUBJID	Concatenate STUDYID and IDSUBNUM, separated by "-"	CONCAT[STUDYID || "-" || IDSUBNUM]
SOURCE.AE	IDAE	AE	AESPID	Copy from source variable	COPY
SOURCE.AE	VERBA	AE	AETERM	Copy from source variable	COPY
SOURCE.AE	AELLT	AE	AELLT	Copy from source variable	COPY
SOURCE.AE	AELLTC	AE	AELLTCD	Copy from source variable	COPY
SOURCE.AE	DTSAE	AE	AESTDTC	Copy from source variable, converted to ISO8601 format	FUNCTION[%CONVERT_DATETIME(INDATE = DTSAE, INDATETIMEFMT = DDMMMYYYY, OUTDATETIME = AESTDTC);]
SOURCE.AE	INTENS	AE	AESEV	Recode according to AESEV recoding list	RECODE[AESEV]
SOURCE.AE	SAE	AE	AESER	Recode according to NY recoding list	RECODE [NY]

CODELIST	TYPE	FROM	TO
AESEV	N2C	1	MILD
AESEV	N2C	2	MODERATE
AESEV	N2C	3	SEVERE
AESEV	N2C	.	
NY	N2C	0	N
NY	N2C	1	Y
NY	N2C	96	NA
NY	N2C	.	

- Easy to learn and use
- Reduces repetitive code
- Re-use of mapping in studies with same set-up

OCS Life Sciences Innovion

Data steward report

- Overview of SDTM/ADaM datasets, variable names, labels, types and values
- Act as a library

Dataset	Variable	Label	Data Type	Value	STUDY 1	STUDY 2	STUDY 3	STUDY 4	Total	Remarks
LB	STUDYID	Study Identifier	text		Y	Y	Y	Y	4	Values are too diverse to present.
LB	DOMAIN	Domain Abbreviation	text	LB	Y	Y	Y	Y	4	
LB	USUBJID	Unique Subject Identifier	text		Y	Y	Y	Y	4	Values are too diverse to present.
LB	LBSEQ	Sequence Number	integer		Y	Y	Y	Y	4	Values are too diverse to present.
LB	LBTESTCD	Lab Test or Examination Short Name	text	CHOL	Y	Y	Y		3	
LB	LBTESTCD	Lab Test or Examination Short Name	text	COCAINE	Y	Y	Y	Y	4	
LB	LBTESTCD	Lab Test or Examination Short Name	text	CREAT	Y	Y	Y	Y	4	
LB	LBTEST	Lab Test or Examination Name	text	Cholesterol	Y	Y	Y		3	
LB	LBTEST	Lab Test or Examination Name	text	Cocaine	Y	Y	Y	Y	4	
LB	LBTEST	Lab Test or Examination Name	text	Creatinine	Y	Y	Y	Y	9	
LB	LBCAT	Category for Lab Test	text	BIOCHEMISTRY	Y	Y	Y	Y	9	
LB	LBCAT	Category for Lab Test	text	DRUG TEST	Y		Y	Y	3	
LB	LBCAT	Category for Lab Test	text	DRUG TESTING		Y			1	
LB	LBCAT	Category for Lab Test	text	URINALYSIS	Y	Y	Y	Y	4	
LB	LBORRES	Result or Finding in Original Units	text	NEGATIVE	Y	Y	Y	Y	4	Only character values are presented.
LB	LBORRES	Result or Finding in Original Units	text	POSITIVE			Y	Y	2	Only character values are presented.

OCS Life Sciences Innovion

Define-XML

- Start of the project: created using Pinnacle 21 tool
  - Metadata on dataset and variable level

Ord	Datas	Variable	Label	Data Typ	Leng	Significant Digits	Form	Mandato	Codelist	Orig	Pages	Method	Predecess	Role	Comme
1	CM	STUDYID	Study Identifier	text	16			Yes						Identifier	
2	CM	DOMAIN	Domain Abbreviation	text	2			Yes						Identifier	
3	CM	USUBJID	Unique Subject Identifier	text	20			No						Identifier	
4	CM	CMSEQ	Sequence Number	integer	8			Yes						Identifier	
5	CM	CMSPID	Sponsor-Defined Identifier	text	1			No						Identifier	
6	CM	CMTRT	Reported Name of Drug, Med, or Therapy	text	21			Yes						Topic	
7	CM	CMDECOD	Standardized Medication Name	text	11			No						Synonym Qualifier	
8	CM	CMCAT	Category for Medication	text	22			No						Grouping Qualifier	
9	CM	CMINDC	Indication	text	28			No						Record Qualifier	
10	CM	CMCLAS	Medication Class	text	37			No						Variable Qualifier	
11	CM	CMCLASCD	Medication Class Code	text	5			No						Variable Qualifier	
12	CM	CMDOSE	Dose per Administration	integer	8			No						Record Qualifier	
13	CM	CMDOSU	Dose Units	text	11			No						Variable Qualifier	
14	CM	CMDOSFRM	Dose Form	text	19			No						Variable Qualifier	
15	CM	CMDOSFRQ	Dosing Frequency per Interval	text	2			No						Variable Qualifier	
16	CM	CMROUTE	Route of Administration	text	11			No						Variable Qualifier	
17	CM	EPOCH	Epoch	text	9			No						Timing	
18	CM	CMSTDTC	Start Date/Time of Medication	date				No						Timing	
19	CM	CMENDTC	End Date/Time of Medication	date				No						Timing	
20	CM	CMSTDY	Study Day of Start of Medication	integer	8			No						Timing	
21	CM	CMENDY	Study Day of End of Medication	integer	8			No						Timing	

OCS Life Sciences Innovion

Define-XML

- During the project: program to create and validate Define-XML
  - Metadata on dataset, variable and value level including CRF pages, codelists, methods

Ord	Datas	Variable	Label	Data Typ	Leng	Significant Digits	Form	Mandato	Codelist	Orig	Pages	Method	Predecess	Role	Comme
1	CM	STUDYID	Study Identifier	text	16			Yes						Identifier	
2	CM	DOMAIN	Domain Abbreviation	text	2			Yes	CM.DOMAIN					Identifier	
3	CM	USUBJID	Unique Subject Identifier	text	20			Yes		Derived		USUBJID		Identifier	
4	CM	CMSEQ	Sequence Number	integer	8			Yes		Derived		SEQ		Identifier	
5	CM	CMSPID	Sponsor-Defined Identifier	text	1			No		CRF	51			Identifier	
6	CM	CMTRT	Reported Name of Drug, Med, or Therapy	text	21			Yes		CRF	851			Topic	
7	CM	CMDECOD	Standardized Medication Name	text	11			No						Synonym Qualifier	
8	CM	CMCAT	Category for Medication	text	22			No		CRF	851			Grouping Qualifier	
9	CM	CMINDC	Indication	text	28			No		CRF	851			Record Qualifier	
10	CM	CMCLAS	Medication Class	text	37			No						Variable Qualifier	
11	CM	CMCLASCD	Medication Class Code	text	5			No						Variable Qualifier	
12	CM	CMDOSE	Dose per Administration	float	4	1		No		CRF	851			Record Qualifier	
13	CM	CMDOSU	Dose Units	text	11			No	CM.UNIT	CRF	851			Variable Qualifier	
14	CM	CMDOSFRM	Dose Form	text	19			No	CM.FRM	CRF	851			Variable Qualifier	
15	CM	CMDOSFRQ	Dosing Frequency per Interval	text	2			No	CM.FREQ	CRF	851			Variable Qualifier	
16	CM	CMROUTE	Route of Administration	text	11			No	CM.ROUTE	CRF	851			Variable Qualifier	
17	CM	EPOCH	Epoch	text	9			No		Derived		CM.EPOCH		Timing	
18	CM	CMSTDTC	Start Date/Time of Medication	date			ISO8601	No		CRF	851			Timing	
19	CM	CMENDTC	End Date/Time of Medication	date			ISO8601	No		CRF	851			Timing	
20	CM	CMSTDY	Study Day of Start of Medication	integer	8			No		Derived		CMSTDY		Timing	
21	CM	CMENDY	Study Day of End of Medication	integer	8			No		Derived		CMENDY		Timing	

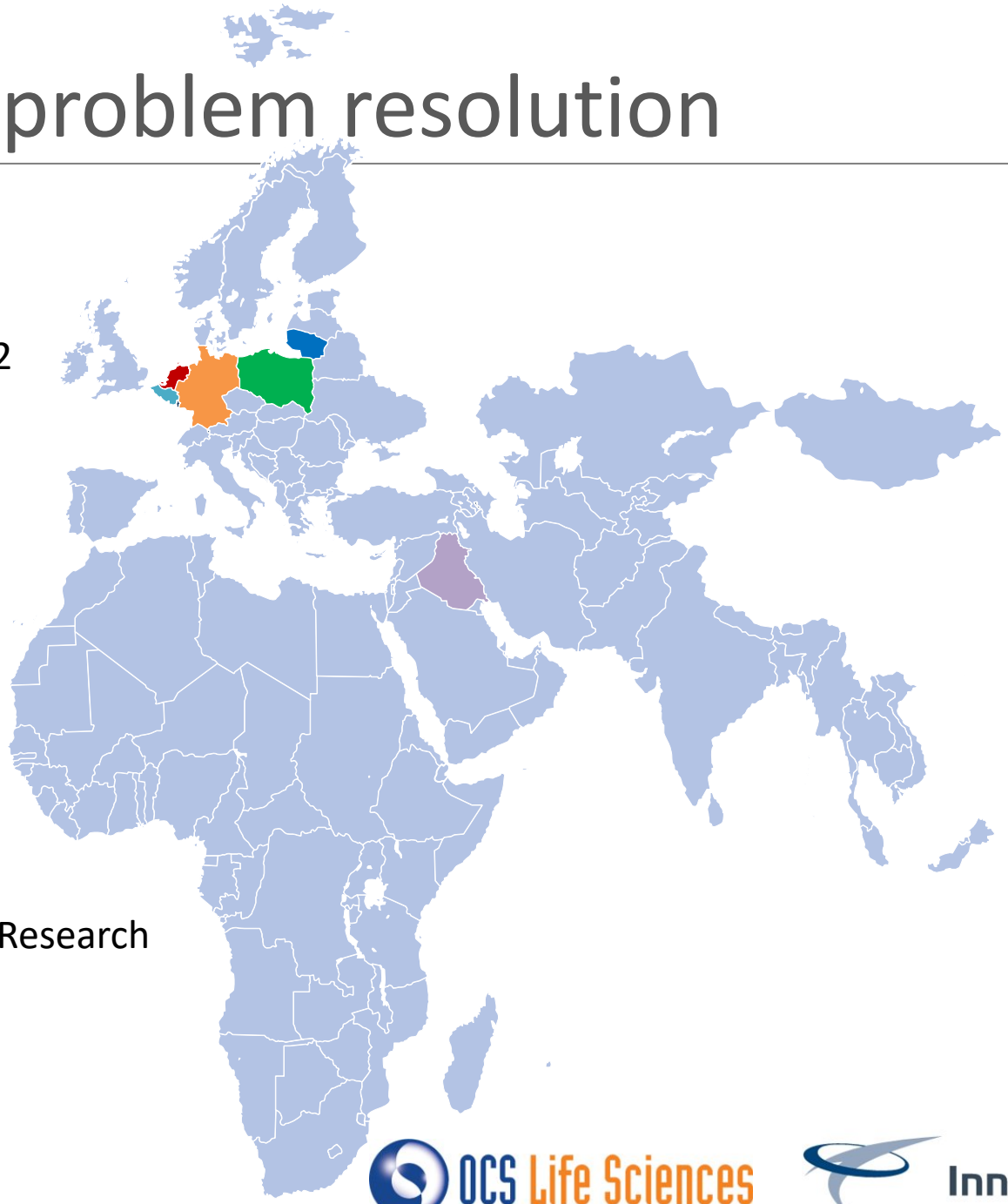
  - Reduces development and validation time
  - Keeps the team motivated

OCS Life Sciences Innovion

Team diversity for problem resolution

Data Manager: 1
Technical/SAS Programmers: 2
Clinical Programmers: 7
Statistician: ad hoc

IT
Business Mathematics
Biochemistry
Chemical Sciences
Nutrition and Health
Medical and Pharmaceutical Research
(Molecular) Neurosciences
Bio-Engineering



OCS Life Sciences Innovion

Data challenges

- Data
  - From multiple CROs
  - In multiple languages
  - With many source issues
  - Poorly described metadata

Example challenge and team resolution

- After generation of TFL

CSR

Change in glycemic parameters				
		BASELINE	**END POINT**	**Δ B - EP**
Plasma glucose	Placebo	5.18 ± 0.63	5.22 ± 0.42	0.04 ± 0.51
(mmol/L)	DRUG A 10 mg	5.13 ± 0.42	5.14 ± 1.06	0.01 ± 0.97
	DRUG A 20 mg	5.06 ± 0.74	5.16 ± 0.78	0.1 ± 0.42
HbA1c (%)	Placebo	xxx	xxx	xxx
	DRUG A 10 mg	xxx	xxx	xxx
	DRUG A 20 mg	xxx	xxx	xxx

Re-calculation after conversion

Change in glycemic parameters					
		BASELINE	**END POINT**	**Δ B - EP**	**….**
Plasma glucose	Placebo	5.18 ± 0.63	5.16 ± 0.45	0.02 ± 0.49	xxx
(mmol/L)	DRUG A 10 mg	5.13 ± 0.42	5.14 ± 1.06	0.01 ± 0.97	xxx
	DRUG A 20 mg	5.06 ± 0.74	5.16 ± 0.78	0.1 ± 0.42	xxx
HbA1c (%)	Placebo	xxx	xxx	xxx	xxx
	DRUG A 10 mg	xxx	xxx	xxx	xxx
	DRUG A 20 mg	xxx	xxx	xxx	xxx

OCS Life Sciences Innovion

What did we do?

Source
data

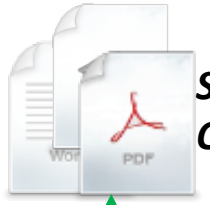
COMPARISON

SDTM

Source
analysis

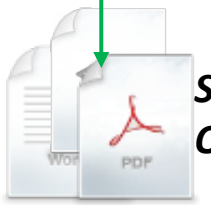
COMPARISON

ADaM



Statistical
Output

COMPARISON



Statistical
Output

What was the issue?

In source analysis

If multiple samples for same date/time

- Only keep one record

In conversion analysis

If multiple samples for same date/time

- Use mean of results

Problem:

Source programming code was available

Rationale for deleting records was not available

(not clear if highest or lowest, first or re-test results were kept)

OCS Life Sciences Innovion

Proposed solution

Upfront agreement:

No hardcoding to reconstruct results

Re-analysis

Results were not significantly different with or without deleted records

OCS Life Sciences Innovion

Current status

- On track
- Team is highly motivated
- High profile project within our companies
- Planning for celebration in March

Conclusions

MOTIVATION IS THE KEY TO SUCCESS

THERE IS NO “ONE APPROACH TO DOING IT RIGHT” – TEAM IDEAS

FOCUS ON WHAT IS IMPORTANT: THE DATA

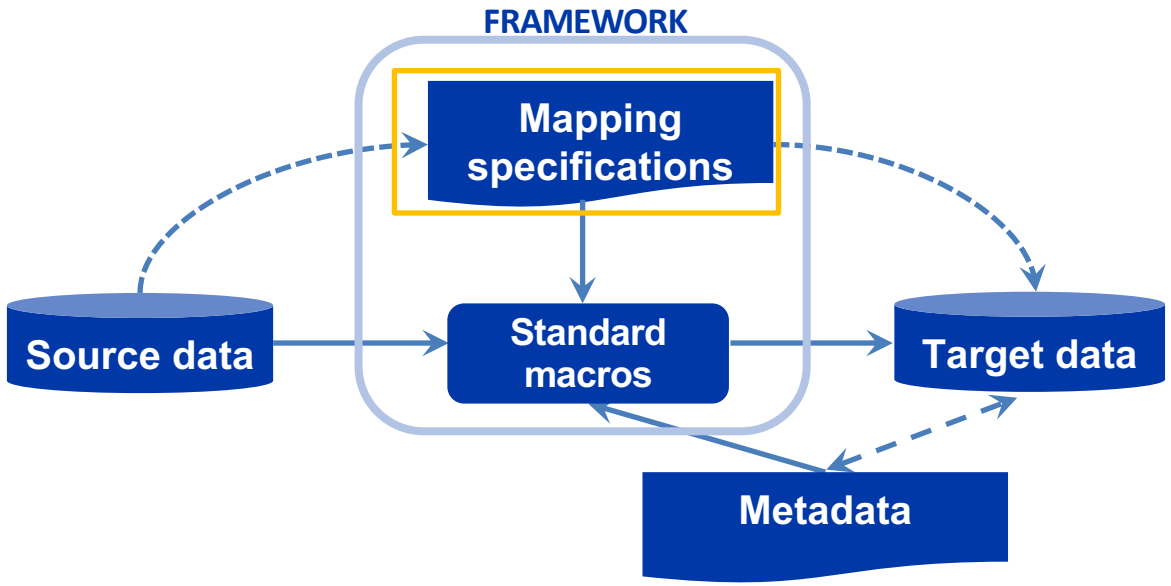
OCS Life Sciences Innovion

Thank you! Questions?

Lieke Gijsbers
OCS Life Sciences
Netherlands
lieke.gijsbers@ocs-consulting.com

Jasmine Kestemont
Innovion
Belgium
jasmine.kestemont@innovion.be

Back-up slides



OCS Life Sciences Innovion

Mapping Engine

- Excel spreadsheet defining:
  - Order of DRM variables
  - The attributes of the DRM variables
  - Key variables for the sorting of records

DOMAIN	NAME	LABEL	TYPE	LENGTH	FORMAT	ORIG	EVAL	SUPPKEY	SORTVAR	UPCASE
AE	STUDYID	Study Identifier	C	200					1	Y
AE	DOMAIN	Domain Abbreviation	C	200						Y
AE	USUBJID	Unique Subject Identifier	C	200					2	Y
AE	AESEQ	Sequence Number	N	8						
AE	AESPID	Sponsor-Defined Identifier	C	200						
AE	AETERM	Reported Term for the Adverse Event	C	200						Y
AE	AELLT	Lowest Level Term	C	200						Y
AE	AELLTCD	Lowest Level Term Code	N	8						
AE	AEDECOD	Dictionary-Derived Term	C	200					3	Y
AE	AEPTCD	Preferred Term Code	N	8						
AE	AEHLT	High Level Term	C	200						Y
AE	AEHLTCD	High Level Term Code	N	8						
AE	AEHLGT	High Level Group Term	C	200						Y
AE	AEHLGTCD	High Level Group Term Code	N	8						
AE	AEBODSYS	Body System or Organ Class	C	200						Y
AE	AEBDSYCD	Body System or Organ Class Code	N	8						
AE	AESOC	Primary System Organ Class	C	200						Y
AE	AESOCCD	Primary System Organ Class Code	N	8						
AE	AESEV	Severity/Intensity	C	200						Y
AE	AESER	Serious Event	C	200						Y
AE	AEACN	Action Taken with Study Treatment	C	200						Y
AE	AEACNOTH	Other Action Taken	C	200						Y
AE	AEREL	Causality	C	200						Y
AE	AEOUT	Outcome of Adverse Event	C	200						Y
AE	AECONTRT	Concomitant or Additional Trtmnt Given	C	200						Y
AE	EPOCH	Epoch	C	200						Y
AE	AEDTC	Date/Time of Collection	C	200						
AE	AESTDTC	Start Date/Time of Adverse Event	C	200					4	
AE	AEENDTC	End Date/Time of Adverse Event	C	200						
AE	AEDY	Study Day of Visit/Collection/Exam	N	8						
AE	AESTDY	Study Day of Start of Adverse Event	N	8						
AE	AEENDY	Study Day of End of Adverse Event	N	8						
SUPPAE	AETRTEM	Treatment Emergent Analysis Flag	C	200		DERIVED		AESEQ		

OCS Life Sciences Innovion

Mapping Engine

SOURCE_DS	SOURCE_VAR	TARGET_DS	TARGET_VAR	SPECIFICATION	FUNCTION
SOURCE.AE	AE	AE		Keep only records where an adverse event occurred	WHERE[AE=1]
SOURCE.AE		AE	STUDYID	Assign "GFT505-111-7-1"	ASSIGN[GFT505-111-7-1]
SOURCE.AE		AE	DOMAIN	Assign "AE"	ASSIGN[AE]
SOURCE.AE	IDSTUD			Not mapped	NOT MAPPED
SOURCE.AE	IDSUBNUM	AE	USUBJID	Concatenate STUDYID and IDSUBNUM, separated by "-"	CONCAT[STUDYID || "-" || IDSUBNUM]
SOURCE.AE	IDAE	AE	AESPID	Copy from source variable	COPY
SOURCE.AE	VERBA	AE	AETERM	Copy from source variable	COPY
SOURCE.AE	AELLT	AE	AELLT	Copy from source variable	COPY
SOURCE.AE	AELLTC	AE	AELLTCD	Copy from source variable	COPY
SOURCE.AE	DTSAE	AE	AESTDTC	Copy from source variable, converted to ISO8601 format	FUNCTION[%CONVERT_DATETIME(INDATE = DTSAE, INDATETIMEFMT = DDMMMYYYY, OUTDATETIME = AESTDTC);]
SOURCE.AE	INTENS	AE	AESEV	Recode according to AESEV recoding list	RECODE[AESEV]
SOURCE.AE	SAE	AE	AESER	Recode according to NY recoding list	RECODE [NY]
		AE		**POSTSTEP1: Add SDTM DM.RFSTDTC to mapped AE data** Obtain USUBJID and RFSTDTC from SDTM.DM by merging the processed mapped AE data with the SDTM.DM domain on USUBJID and keep only records from the mapped AE data.	POSTSTEP1[PROC SORT DATA=work.mapped_source_ae OUT =work.mapped_ae01; BY usubjid; RUN; PROC SORT DATA=outlib.dm(KEEP=usubjid rfstdtc) OUT =work.mapped_dm01; BY usubjid; RUN; DATA work.mapped_ae02; MERGE work.mapped_ae01(IN=inae) work.mapped_dm01; BY usubjid; IF inae; RUN;]

OCS Life Sciences Innovion