



From SDTM to displays, through ADaM & Analyses Results Metadata, a flight on board METADATA Airlines

Omar SEFIANI - Stéphane BOUGET, Boehringer
Ingelheim

DH13, PhUSE Barcelona 2016, October, 12th



Boehringer
Ingelheim

- Background
- Metadata Driven Programming
- Metadata Repository
- SDTM mapping & derivations
- ADaM implementation
- Analysis Results Metadata (ARM)
- Data exportation / documentation
- Summary



- Many changes can occur during the conduct of a project:
 - ❖ standard updates leading to structural changes
 - ❖ new scientific approaches
 - ❖ new regulatory requests
 - Find a solution to:
 - ❖ Reduce maintenance effort
 - ❖ Increase readability
 - ❖ Ensure consistency between similar implementations
- => One single program per area dealing with multiple trials simultaneously (centric approach)



- Metadata used for:
 - ❖ Generic SAS programs development
 - ❖ SAS Code generation
 - ❖ Multiple small macros implementation (elementary tasks)
 - ❖ Modular programming using KEYWORDS
- The concept presented is currently used in a multi-trial respiratory project
 - ❖ BI legacy data conversion to SDTM
 - ❖ ADaM transformation
 - ❖ Reporting program generation

Definition of
algorithms

Identification in the code
via meaningful
KEYWORDS describing
the functionality

- Flexibility
- Reduction of maintenance efforts
- Centralization and automatization

Name	ADAM	<pre> * All The Rules for Numerical and Character variables only; if RULECB = 'FIRST' then do; if rowid=1 then &__varname.=r_&__varname.; end; else if RULECB = 'FIRST-NON-MISSING' then do; if missing(&__varname.) then &__varname.=r_&__varname.; end; else if RULECB = 'LAST' then do; if rc2=0 and has_nxtchk=0 then &__varname.=r_&__varname.; end; else if RULECB = 'LAST-NON-MISSING' then do; if not missing(r_&__varname.) then &__varname.=r_&__varname.; end; else if RULECB = 'NONE' then do; &__varname.=r_&__varname.; end; </pre>	RULECB03
Label	Target ADaM		Rule to apply for COMB03FL
Type	text		text
Format	\$8.		\$40.
Case	UPPER		UPPER
Context (Where condition)			
	ADSL		FIRST
	ADSL		FIRST
	ADSL		FIRST
	ADSL		FIRST
	ADSL		FIRST
	ADSL		LAST

MS Excel or any MDR system

SDTM / ADAM

- One sheet per SDTM domain / ADAM dataset (structure)
- Differentiation of variable types (mapped or derived / predecessor or derived)

~~Controlled Terminology embedded~~ **SDTM only**

- Link derived variables to derivation methods keywords
- One sheet for all methods (link keyword to SAS macro)

ARM

TOC (Table of contents Generator) :

- Unique display template
- Output (TLF)
- Statistics for main analyses

Domain Name	Variable	Label	Type	Length	Format	Class	Conversion Rule
DM	STUDYID	Study Identifier	Text	20		Source	
DM	DOMAIN	Domain Abbreviation	Text	8		Derived Level 0	RULE_DOMAIN
DM	USUBJID	Unique Subject Identifier	Text	20		Source	
DM	SUBJID	Subject Identifier for the Study	Text	20		Source	
DM	RFSTDTC	Subject Reference Start Date/Time	Datetime	20	ISO 8601	Derived Level 1	RULE_FIRST_DRUG_INTAKE
DM	RFENDTC	Subject Reference End Date/Time	Datetime	200	ISO 8601	Derived Level 1	RULE_TRIAL_COMPLETION
DM	RFXSTDTC	Date/Time of First Study Treatment	Datetime	20	ISO 8601	Derived Level 1	RULE_FIRST_DRUG_INTAKE
DM	RFXENDTC	Date/Time of Last Study Treatment	Datetime	20	ISO 8601	Derived Level 2	RULE_LAST_DRUG_INTAKE

Name	ADSVAR	TSPARMCD	TSGRPID	VALTYPE
Label	Variable of ADTRIALS	Parameter Code of TS	Group ID	Conversion
Type	text	text	text	text
Format	\$8.	\$8.	\$40.	\$20.
Case	UPPER	UPPER	UPPER	UPPER
Context (Where condition)				
	DBLDT	DCUTDTC	DBLOCK	DATE
	SNAPMED	MEDVER		CHAR4
	SNAPWHO	WHOVER		CHAR6
	TPHASELG	TPHASE		CHAR50
	CNTRIES	FCNTRY		CHAR200
	TITLE	TITLE		CHAR200
	INTMODEL	INTMODEL		CHAR200
	TBLIND	TBLIND		CHAR200
	TCNTRL	TCNTRL		CHAR200
	RANDOM	RANDOM		CHAR10
	SEXPOP	SEXPOP		CHAR10
	AGEMIN	AGEMIN		YEARS
	AGEMAX	AGEMAX		YEARS
	PLANENR	PLANSUB		INTEGER
	FVFPATDT	SSTDTC		DATE
	TSTOPDT	CTARDTC		DATE
	EXTFL	EXTTRL		CHAR1
	PROJID	PROJNUM		CHAR10
	WOTIME	REPDUR		DAYS

Appendix	Number	Section	Type	Title 1	Program	Outputno	Display Template
15							
15 1		TRIAL PATIENTS					
15 1.1			Table	Disposition of patients	dispo.sas	1	EOT-DISP1
15 2		EFFICACY EVALUATION					
15 3		SAFETY EVALUATION					
15 3.1		Adverse events					
15 3.1.1			Table	Frequency [N (%)] of patients with adverse events - TS	ae.sas	1	XAE-1
15 3.2		Vital signs - physical findings and other relevant observations related to safety					
15 3.2.1			Table	Descriptive statistics and change from baseline over time - TS	vs.sas	1	EOT-DESC1

Optionnal						
Multiple exposure flag	Highest population flag	Treatment Analysis Number	ADSL replacement Class name	ADSL variables used	BDS endpoints	BDS where restriction
01	TRTFL	1000		NONE	ADVS.SYSBP[AVAL] / ADVS.SYSBP[CHG]	AVISIT = 52

SDTM transformation engine :

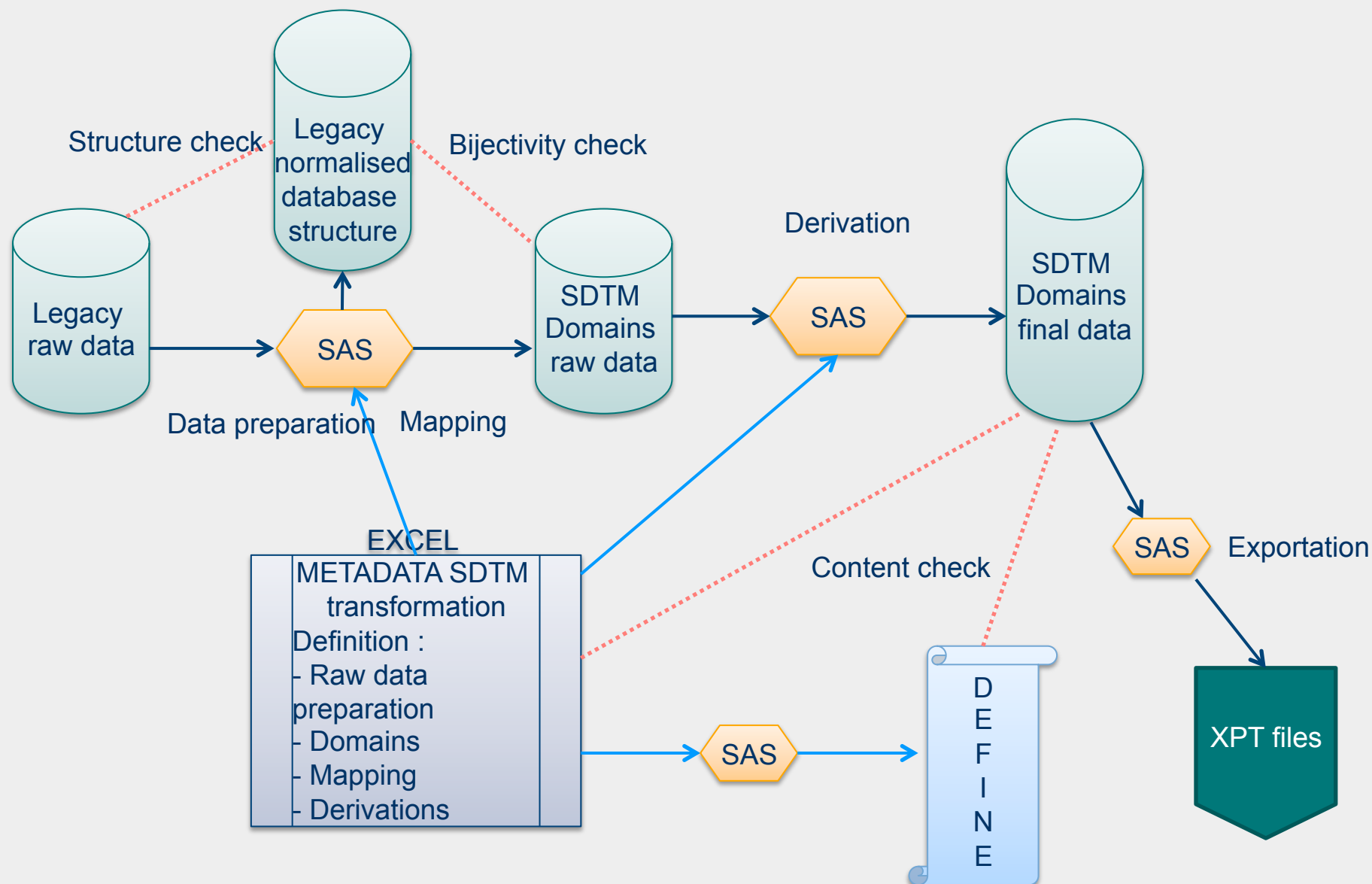
- All legacy data converted to SDTM on an ongoing basis
- Includes additional non-required information by SDTM (e.g. flags)

Step 1 : SDTM mapping

- one-to-one relationship between one legacy variable and one SDTM variable
- obtain a bijection between one raw value and one SDTM value

Step 2 : SDTM derivations

- after all domains are created
- facilitating the interaction between different domains
- execution order needed
example : EPOCH
variable derived first in SE
and then populated in
other domains



Metadata ADaMs

- Global datasets that contains attributes (study or patient level)
- Can be used by all subsequent ADaMs
- Automatically replicated to the main sheet as 'Assigned' variables to be extracted from the define.xml during the creation of the ADS.

Manual entry

	<i>Context (Where condition)</i>	<i>GARMCD</i>
1199_0214	ARMCD EQ "E"	B300T
1199_0214	ARMCD EQ "F"	P300T
1199_0214	ARMCD EQ "DBLUE"	DBLUE
1199_0214	ARMCD EQ "DRED"	DRED
1199_0214	ARMCD EQ "SCRNFALL"	SCRNFALL
1199_0214	ARMCD EQ "SCRNONG"	SCRNONG
1199_0214	ARMCD EQ "NOTASSGN"	NOTASSGN

ADaM dataset ADTARM in ADS plan

<i>Dataset Name</i>	<i>Context (Where condition)</i>	<i>Variable Name</i>	<i>Variable Label</i>	<i>Origin</i>	<i>Source / Derivation</i>
ADTARM	*ALL*	STUDYID	Study Identifier	Predecessor	TS.STUDYID
ADTARM	*ALL*	SNAPSHOT	Snapshot Name	Predecessor	TS.TSVAL [TSPARMCD = 'SNAPSHOT']
ADTARM	*ALL*	SNAPDT	Snapshot Date	Predecessor	TS.TSVAL [TSPARMCD = 'SNAPDT']
ADTARM	*ALL*	ARMCD	Planned Arm Code	Predecessor	TA.ARMCD
ADTARM	*ALL*	ARM	Description of Planned Arm	Predecessor	TA.ARM
ADTARM	*ALL*	GARMCD	Global Arm Code	Assigned	
ADTARM	ARMCD EQ "E"	GARMCD	Global Arm Code	Assigned	B300T
ADTARM	ARMCD EQ "F"	GARMCD	Global Arm Code	Assigned	P300T
ADTARM	ARMCD EQ "DBLUE"	GARMCD	Global Arm Code	Assigned	DBLUE
ADTARM	ARMCD EQ "DRED"	GARMCD	Global Arm Code	Assigned	DRED
ADTARM	ARMCD EQ "SCRNFALL"	GARMCD	Global Arm Code	Assigned	SCRNFALL
ADTARM	ARMCD EQ "SCRNONG"	GARMCD	Global Arm Code	Assigned	SCRNONG
ADTARM	ARMCD EQ "NOTASSGN"	GARMCD	Global Arm Code	Assigned	NOTASSGN

Setup metadata datasets

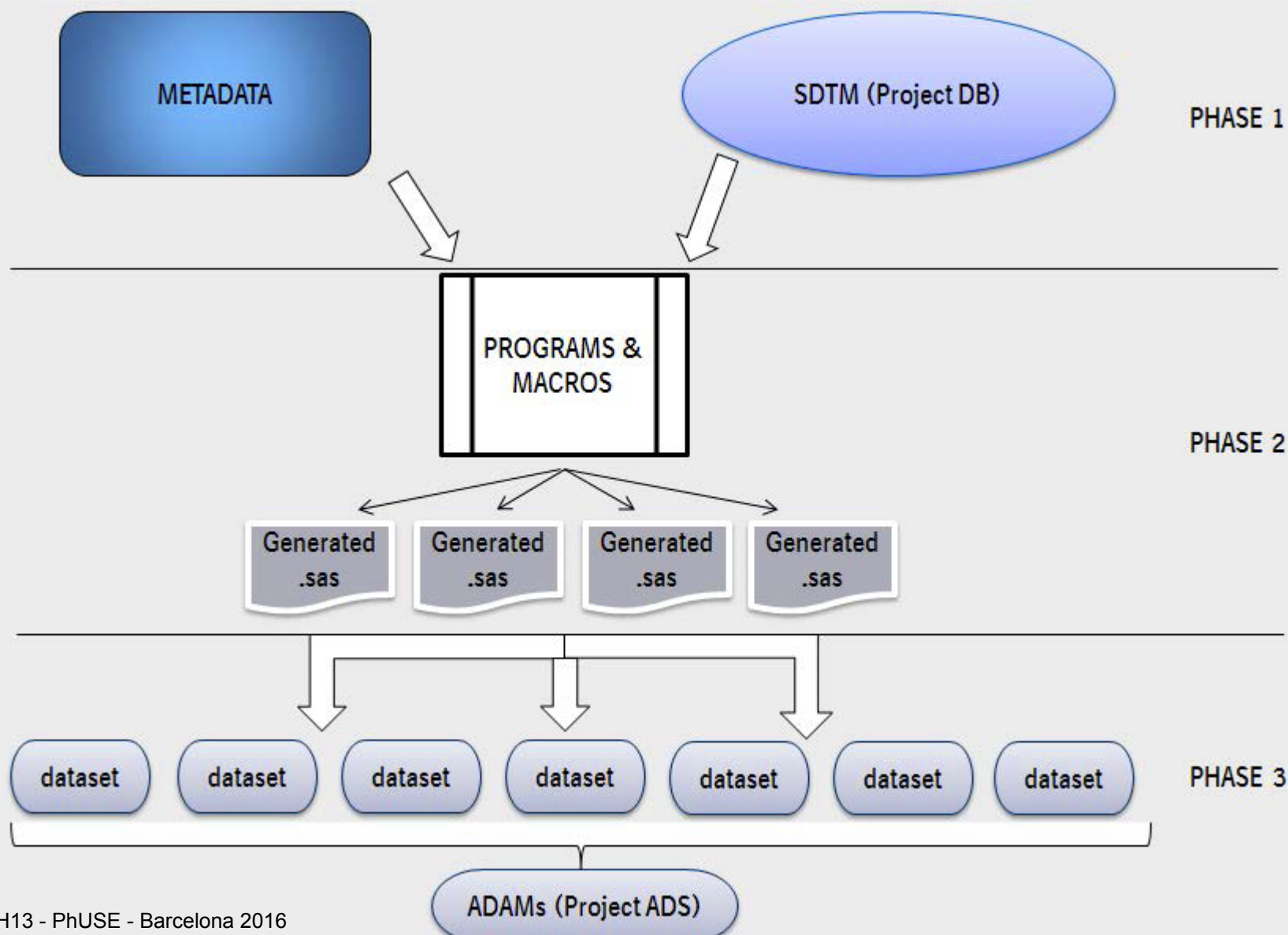
- Contains algorithms identification
- Read via a generic macro
- Used to generate a SAS program creating the corresponding metadata dataset
- Used as setup dataset during the creation of an ADaM

Name	ADSVAR	TSPARMCD	TSGRPID	VALTYPE
Label	Variable of ADTRIALS	Parameter Code of TS	Group ID	Conversion
Type	text	text	text	text
Format	\$8.	\$8.	\$40.	\$20.
Case	UPPER	UPPER	UPPER	UPPER
Context (Where condition)				
	DBLDT	DCUTDTC	DBLOCK	DATE

```

data SETUP_SPTRIALS;
  attrib      ADSVAR      label="Variable of ADTRIALS"  format=$8.
              TSPARMCD    label="Parameter Code of TS"   format=$8.
              TSGRPID     label="Group ID"               format=$40.
              VALTYPE      label="Conversion"            format=$20.
;
  call missing(ADSVAR,TSPARMCD,TSGRPID,VALTYPE);
  delete;
run;

proc sql;
  insert into SETUP_SPTRIALS
  values (
    'DBLDT',
    'DCUTDTC',
    'DBLOCK',
    'DATE'
  )
    
```



ARMs used to:

- Automatically create the ARM section in the define.xml V2 for ADaMs
- Dynamically generate a part of SAS macro calls for displays

An output program generator is currently under development to automatically generate the different macro calls for:

- Data building (ADaM preparation)
- Analysis & reporting (calculation+output)

Table Of Content (metadata repository)

TO

- Output description
- Statistical part of ARM

GENERAT

- Data selection
- Macros parametrization

SAS MACROS

SAS Program.sas
SAS Program.sas
SAS Program.sas
SAS Program.sas

define.xml

ARMs

Echantillon	D _{ext} (yr)	D _{ext} (β)		D _{ext} (1+β)		D(1)		Age ESR (ks)		US-ESR	
		EU	LU	EU	LU	EU	LU	EU	LU	Age (ks)	β ²
cas11459e	2171 ± 27	3,134		3,311	1,610	6,127	236 ± 14	354 ± 19	163 (+15-13)	327 (+34-31)	-0.6
cas11459e			2,746	1,343					174 ± 7		-0.6
cas405e	2291 ± 31	1,516		5,007	2,941	10,064	6,106	214 ± 13	372 ± 22	171 (+11-10)	268 (+43-36)
cas405e		(β=0)	1,633	0,636						227 (+19-16)	-0.9
cas405e			1,928	0,975						219 (+13-11)	-0.8
cas3439e	2101 ± 37	1,516		4,205	2,131	10,182	5,903	206 ± 14	356 ± 22	217 (+18-15)	271 (+37-34)
cas3439e		(β=0)	2,100	1,062						171 (+9-7)	-0.7
cas3439e			2,351	1,194						233 (+14-12)	-0.9
cas1515e	2151 ± 42	1,516		5,208	2,643	10,063	6,104	202 ± 13	350 ± 21	178 (+10-9)	260 (+37-33)
cas1515e		(β=0)	1,006	0,844						196 (+12-11)	-0.9
cas1515e			2,273	1,151						223 (+15-13)	-0.9
cas7599e	2617 ± 41	1,516		12,177	6,363	21,210	11,773	123 ± 8	222 ± 14	232 (+16-14)	nd
cas7599e		(β=0)	3,283	1,701						248 (+14-13)	nd
cas7599e			4,254	2,163						271 (+21-18)	nd

D_{ext} = dose équivalente (± erreur statistique) ; D_{ext} = dose externe déterminée à partir des teneurs en U, Th, K des sédiments collectés sur le site ; D_{ext} (dérivée) = D_{ext} (jeune) ; D_{ext} (jeune) sont déterminées à partir des rapports d'activité U-238/U-235 et Th-232/U-235 déterminés par spectrométrie alpha. La dose annuelle est calculée d'après Adams et Allison (1966). EU = early uptake, LU = linear uptake (Key, 1993) ; Les teneurs en eau sont (% poids) : addiment = 20 ± 5, dérivée = 15 ± 5, émail = 2 ± 1 ; k₁ = 0,13 ± 0,02 (Gron et Katzenberger-Apel, 1994). D(1) = D_{ext} + D_{ext} ; β² = paramètre de diffusion de l'uranium d'après la méthode US-ESR (Gron et al., 1995) ; nd = non déterminé.

SDTM

- Define.xml generated automatically at each execution
- .XPTs generated at the same time
- Pinnacle 21 checks performed
- Bijectivity check
- Possible data restriction for the export

ADAM

- Define.xml main source of metadata for the ADaMs creation
- .XPTs generated at the same time
- Pinnacle 21 checks performed
- Analysis Data Reviewer's Guide contains partially derived information based on the ADS plan

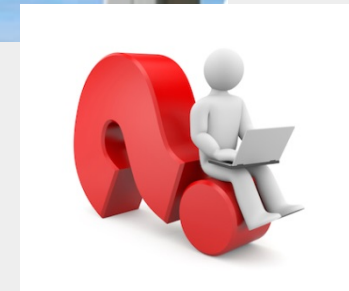
ARM

- Present in the define.xml
- Synchronized with the displays
- Standardized results datasets exported for validation

- Powerful approach to minimize risks of inconsistencies across different packages and studies - fits with a project centric approach (multiple trials)
- Development of programs in a generic manner:
 - ❖ high level of control needed (user ERRORS and WARNINGS)
 - ❖ custom code or hardcoding should be avoided
 - ❖ good level of algorithmic and use of complex technical solutions (eg. hashcode, extended attributes, arrays , doSubl, macros, ...)
- Synergy between standardization and flexibility
 - ❖ more flexible with data diversity
 - ❖ reduction of implementation time and maintenance facility
 - ❖ improvement in productivity for the creation of similar outputs
 - ❖ facilitates delegation and simplify oversight



Thank you



Omar SEFIANI
Stéphane BOUGET
omar.sefiani@boehringer-ingelheim.com
stephane.bouget@boehringer-ingelheim.com