

# *Experiences and lessons learned from a first SDTM submission project*

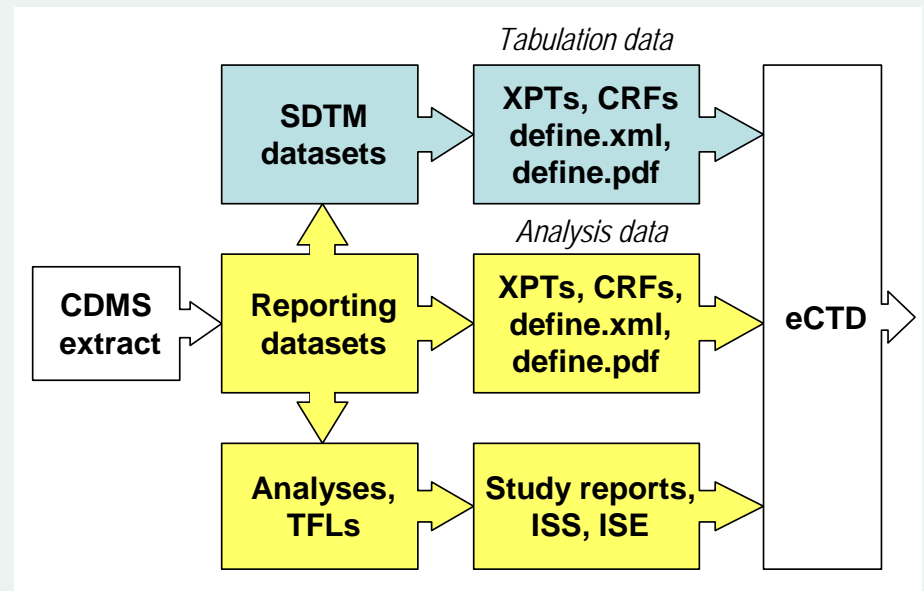
*Paul Vervuren, Schering-Plough  
Bas van Bakel, OCS Consulting*

# Agenda

- Introduction
- Conversion Process
- Sample Submission
- Project Activity Metrics
- Conclusion

# Approach

- Primary submission component: reporting datasets
  - internal data standard/conventions
- SDTM datasets: additional component
- define.xml, define.pdf



# Scope

- 6 studies, ISS datasets
- 25 domains (20 in 1<sup>st</sup> study)
- 7 custom domains (5 in 1<sup>st</sup> study)
- SUPPQUAL, RELREC
- Trial Design datasets

# Starting point

- Toolset
  - Sample conversion programs for 13 standard domains
  - Macro library
  - Metadata repository
  - Mapping sheet
- Training SDTM Theory & Application

# Project strategy

- In-house
- Expert input
- Sample submission
- define.xml, pragmatic
- Separate project

# Define.xml

- Toolset CDISC SDTM/ADaM pilot
- Modifications to MD2ODM macro:
  - ODM1.2 instead of ODM1.3
  - Interface to metadata repository
  - Value-level metadata section
- define.pdf
  - define.xml as input
  - method: SAS XML mapper, SAS ODS (paper Lex Jansen)

The screenshot shows a software interface with a tree view on the left and a table of Value Level Metadata on the right. The tree view includes folders for 'Annotated Case Report Form', 'Datasets', and 'Value Level Metadata'. Under 'Value Level Metadata', there are sub-folders for 'ValueList.CM.CMCAT', 'ValueList.DA.DATESTCD', 'ValueList.IE.IETESTCD', 'ValueList.LB.LBCAT', 'ValueList.LB.LBTTESTCD', 'ValueList.PE.PETESTCD', 'ValueList.QS.QSCAT', 'ValueList.SC.SCTESTCD', 'ValueList.SU.SUTRT', and 'ValueList.SUPPAE.QNAM'. The right pane displays three tables of metadata for 'ValueList.CM.CMCAT', 'ValueList.CM.CMCAT.CONTRACEPTIVES.CMCONTRA', and 'ValueList.DA.DATESTCD'. Each table has columns for Source Variable, Value, Label, Type, Controlled Terminology, and Origin.

Source Variable	Value	Label	Type	Controlled Terminology	Origin
CMCAT	CONTRACEPTIVES	CONTRACEPTIVES	text	CMCONTRA	CRF Page 5
CMCAT	GENERAL MEDICATION	GENERAL MEDICATION	text		CRF Page 39

Source Variable	Value	Label	Type	Controlled Terminology	Origin
CONTRACEPTIVES	CONTRACEPTIVE HISTORY	Contraceptive History	text		Derived
CONTRACEPTIVES	CONTRACEPTIVE STATUS	Contraceptive Status	text		Derived

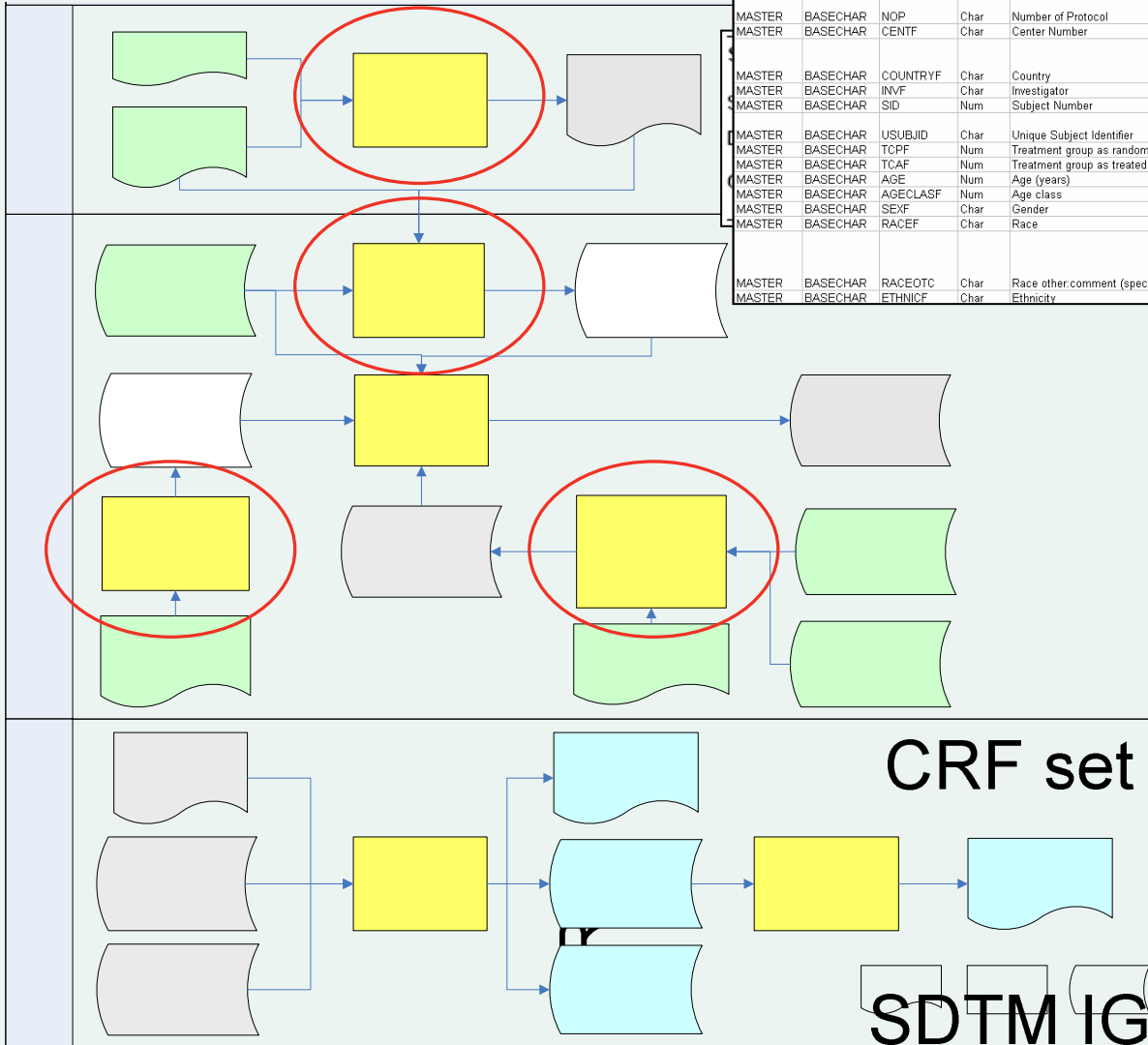
Source Variable	Value	Label	Type	Controlled Terminology	Origin
DATESTCD	DISPAMT	Dispensed Amount	text		CRF Page 31
DATESTCD	RETAMTS	Returned Amount (Strip)	text		CRF Page 31
DATESTCD	RETAMIT	Returned Amount (Tablet)	text		CRF Page 31

The screenshot shows Adobe Acrobat Professional displaying a PDF document titled 'define.pdf'. The document content is a table of Value Level Metadata for 'ValueList.CM.CMCAT'. The table has columns for Source Variable, Value, Label, Type, Controlled Terminology, and Origin. The 'Controlled Terminology' column contains the value 'CMCONTRA' for the first row.

Source Variable	Value	Label	Type	Controlled Terminology	Origin
CMCAT	CONTRACEPTIVES	CONTRACEPTIVES	text	CMCONTRA	CRF Page 5 [27]
CMCAT	GENERAL MEDICATION	GENERAL MEDICATION	text		CRF Page 39 [29]

# Conversion Process

Source Libname	Dataset	Variable	Type	Label	SDTM Domain	SDTM Variable	Mapping type	SDTM Mapping Comments
MASTER	BASECHAR	NOP	Char	Number of Protocol	DM, DS, SC, SU, XT	STUDYID		
MASTER	BASECHAR	CENTF	Char	Center Number	DM	SITEID		
MASTER	BASECHAR	COUNTRYF	Char	Country	DM	COUNTRY	_CODE_ country=country	
MASTER	BASECHAR	INVF	Char	Investigator	DM	INVID		
MASTER	BASECHAR	SID	Num	Subject Number	DM	SUBJID		
MASTER	BASECHAR	USUBJID	Char	Unique Subject Identifier	DM, DS, SC, SU, XT	USUBJID		
MASTER	BASECHAR	TCPF	Num	Treatment group as randomized	DM	ARM, ARMCD		Use format TCPF for ARM. NOT MAPPED
MASTER	BASECHAR	TCAF	Num	Treatment group as treated				See derivation DMD1 on Derivation sheet NOT MAPPED
MASTER	BASECHAR	AGE	Num	Age (years)	DM	AGE		
MASTER	BASECHAR	AGECLASF	Num	Age class				
MASTER	BASECHAR	SEXF	Char	Gender	DM	SEX		
MASTER	BASECHAR	RACEF	Char	Race	DM	RACE		Use format RACE
MASTER	BASECHAR	RACEOTC	Char	Race other comment (specification)	SC	SCORRES, SCTEST, SCTESTCD, SCCAT		Record created only if value present. SCTEST = "Race Other", SCTESTCD = "RACEOTH", SCCAT = "DEMOGRAPHIC DATA" NOT MAPPED
MASTER	BASECHAR	ETHNICF	Char	Ethnicity	DM	ETHNIC		Use format RACE



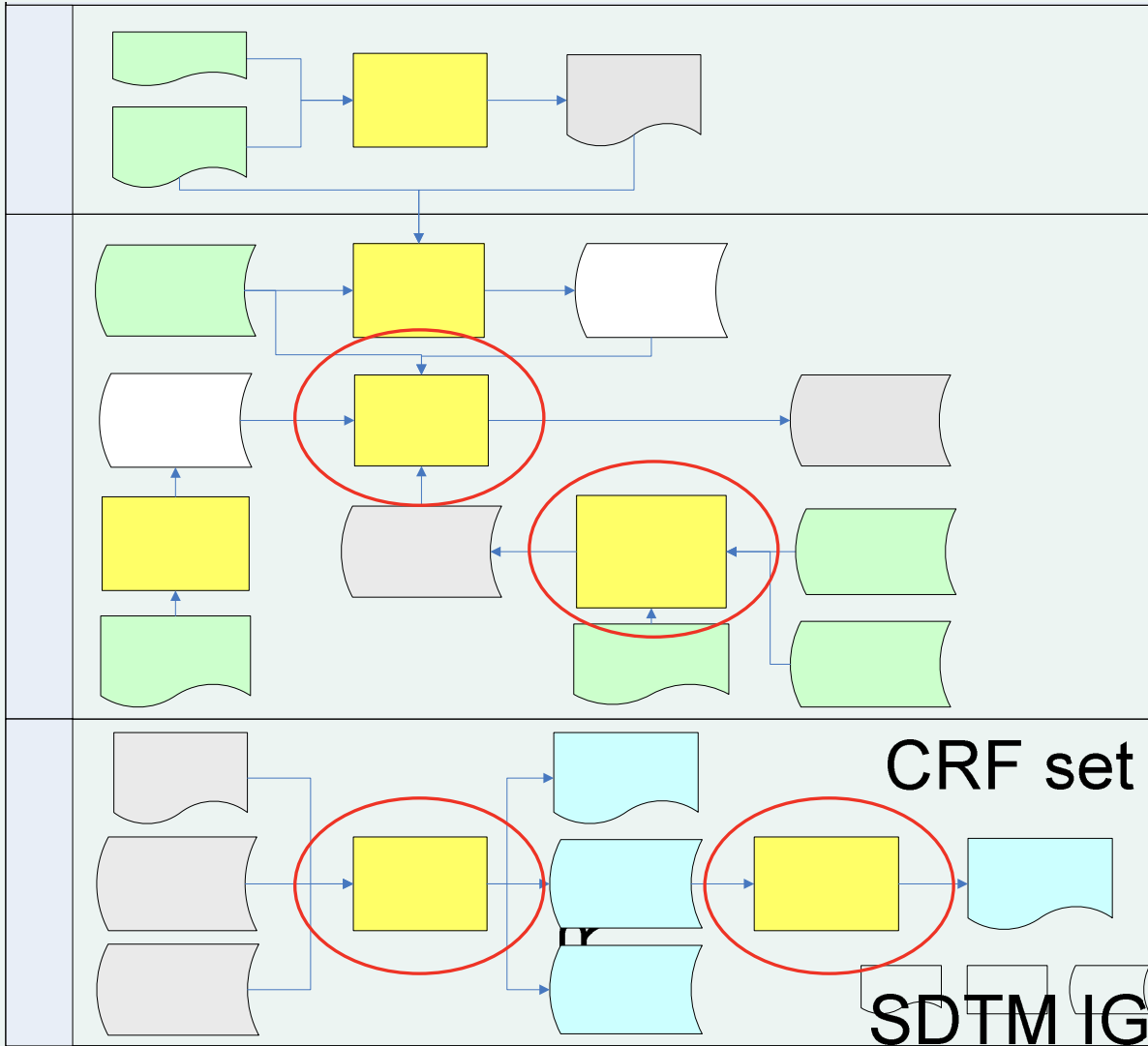
1. SDT  
annota

# Conversion Process

	A	B	C	D	E	F	G	
1	Source	Domain	Description	Location	Structure	Class	Purpose	Keys
2	CDISC	AE	Adverse Events	ae.xpt	One record per adverse event per subject	Events	Tabulation	STUDYID, USUBJID, AETERM, A
3	SPONSOR	BR	Endometrical Biopsy	br.xpt	One record per biopsy per time point per vis	Findings	Tabulation	STUDYID, USUBJID, VISITNUM,
4	CDISC	CM	Concomitant Medications	cm.xpt	One record per medication intervention episode	Interventions	Tabulation	STUDYID, USUBJID, CMCAT, V
5	CDISC	CO	Comments	co.xpt	One record per comment per subject	Special Purpose	Tabulation	STUDYID, USUBJID, COSEQ
6	CDISC	DA	Drug Accountability	da.xpt	One record per accountability observation per	Findings	Tabulation	STUDYID, USUBJID, DACAT, D
7	CDISC	DM	Demographics	dm.xpt	One record per subject	Special Purpose	Tabulation	STUDYID, USUBJID
8	CDISC	DS	Disposition	ds.xpt	One record per disposition status or protocol	Events	Tabulation	STUDYID, USUBJID, DSCAT, D
9	CDISC	DV	Protocol Deviations	dv.xpt	One record per protocol deviation per subject	Events	Tabulation	STUDYID, USUBJID, DVCAT, D
10	CDISC	EG	ECG Test Results	eg.xpt	One record per ECG observation per time po	Findings	Tabulation	STUDYID, USUBJID, VISITNUM,

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	source	domain	seq	variable_name	variable_label	variab	variab	variab	controlled_term	controlled_term	origin	role	core	spons	spon	notes
2	CDISC	DM	1	STUDYID	Study Identifier	Char	20	1			CRF	Identifier	Req		N	
3	CDISC	DM	2	DOMAIN	Domain Abbreviation	Char	2		DOMAIN	SPONSOR	Derived	Identifier	Req		N	
4	CDISC	DM	3	USUBJID	Unique Subject Identifier	Char	40	2			Sponsor Defined	Identifier	Req		N	Concatenation of STUDYID
5	CDISC	DM	4	SUBJID	Subject Identifier for the	Char	40				CRF	Topic	Req		N	
6	CDISC	DM	5	RFSTDTC	Subject Reference Start	Char	19		ISO8601		Sponsor Defined	Timing	Exp		N	Equal to date of first admini
7	CDISC	DM	6	RFENDTC	Subject Reference End	Char	19		ISO8601		Sponsor Defined	Timing	Exp		N	Equal to date of last admini
8	CDISC	DM	7	SITEID	Study Site Identifier	Char	20				CRF or Derived	Record Qualifier	Req		N	of last administration of In
9	CDISC	DM	8	INVID	Investigator Identifier	Char	20				CRF or Derived	Record Qualifier	Perm		N	will be used. Equal to date
10	CDISC	DM	9	INVNAM	Investigator Name	Char	40				CRF or Derived	Synonym Qualifier	Perm		N	
11	CDISC	DM	10	BRTHTDC	Date/Time of Birth	Char	19		ISO8601		CRF or Derived	Result Qualifier	Perm		N	
12	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
23	CDISC	DM	21	ASRY	All-Subjects-Randomized	Char	1		NY	CDISC	Derived	Result Qualifier		Perm	Y	Subject included in the All
24	CDISC	DM	22	ASTY	All-Subjects-Treated Gro	Char	1		NY	CDISC	Derived	Result Qualifier		Perm	Y	Subject included in the All
25	CDISC	DM	23	ITTY	Intent-to-Treat Group	Char	1		NY	CDISC	Derived	Result Qualifier		Perm	Y	Subject included in the Int

# Conversion Process



1. SDT  
annota

# Sample Submission

Consisting of:

- Define.xml;
- Annotated CRFs;
- 36 SDTM domains (as transport files);
- Limited number of subjects from one trial.

FDA feedback:

- Loading of the datasets in WebSDM produced a list of 79 WebSDM validation check messages;
- There was an undefined problem loading the define.xml into WebSDM. To be able to load the SDTM datasets FDA had used standard SDTM 3.1.1 metadata.

# WebSDM check violations

Check ID	Severity	Error message
IR4001	High	Null value in column
IR4135	High	Missing character result when original result provided
IR4252	High	SDTM Required variable not found
R4104	High	Rule failed to execute
IR4000	Medium	No rows in domain table
IR4009	Medium	Either Original Result or Status should be specified, but not both
IR4108	Medium	Invalid BEFOREAFTER code
IR4112	Medium	Derived Flag = 'Y' but (Character) Standard Result is null
IR4253	Medium	SDTM Expected variable not found
IR4501	Medium	Invalid Subject Visit/Visit Number
IR4005	Low	No Baseline result
IR4117	Low	__ENRF expected when __ENDTC is null
IR4118	Low	__STRF expected when __STDTC is null
IR4125	Low	Missing units on value
IR4128	Low	Missing units on value
IR4130	Low	Start date expected when end date provided

# WebSDM check violations

Check ID	Severity	Error message
IR4001	High	Null value in column
<b>IR4135</b>	<b>High</b>	<b>Missing character result when original result provided</b>
IR4252	High	SDTM Required variable not found
<b>R4104</b>	<b>High</b>	<b>Rule failed to execute</b>
IR4000	Medium	No rows in domain table
IR4009	Medium	Either Original Result or Status should be specified, but not both
<b>IR4108</b>	<b>Medium</b>	<b>Invalid BEFOREAFTER code</b>
IR4112	Medium	Derived Flag = 'Y' but (Character) Standard Result is null
IR4253	Medium	SDTM Expected variable not found
IR4501	Medium	Invalid Subject Visit/Visit Number
IR4005	Low	No Baseline result
IR4117	Low	__ENRF expected when __ENDTC is null
IR4118	Low	__STRF expected when __STDTC is null
IR4125	Low	Missing units on value
IR4128	Low	Missing units on value
IR4130	Low	Start date expected when end date provided

# WebSDM check violations

Check ID	Severity	Error message
IR4135	High	Missing standard result when original result provided

- **CAUSE**

Character result in standard format (QSSTRESC) was set to empty when the original result (QSORRES) was 'NOT APPLICABLE'.

- **RESOLUTION**

Fill QSSTRESC with 'NOT APPLICABLE'.

# WebSDM check violations

Check ID	Severity	Error message
IR4108	Medium	Invalid BEFOREAFTER code

- **CAUSE**

In CM one subject had a contraceptive method for which start date was not collected. As a result, CMSTRF could not be determined and was set to 'U' (unknown), in line with SDTM-IG 3.1.1 but WebSDM only accepts 'BEFORE', 'DURING' or 'AFTER' for the –STRF variables.

- **RESOLUTION**

According to specification → No adaptations.  
Will be mentioned in Reviewer's Guide.

# WebSDM check violations

Check ID	Severity	Error message
R4104	High	Rule failed to execute

- **CAUSE**

Description of the check: Identifies records that violate the condition “COUNT(DISTINCT Protocol Deviation Term) = COUNT(DISTINCT Protocol Deviation Coded Term)”.

Check seems to be incorrect as the SDTM model clearly specifies it is possible to code several Protocol Deviation Terms (--TERM) to the same Protocol Deviation Coded Term (--DECOD).

- **RESOLUTION**

According to specification → No adaptations.  
Will be mentioned in Reviewer’s Guide.

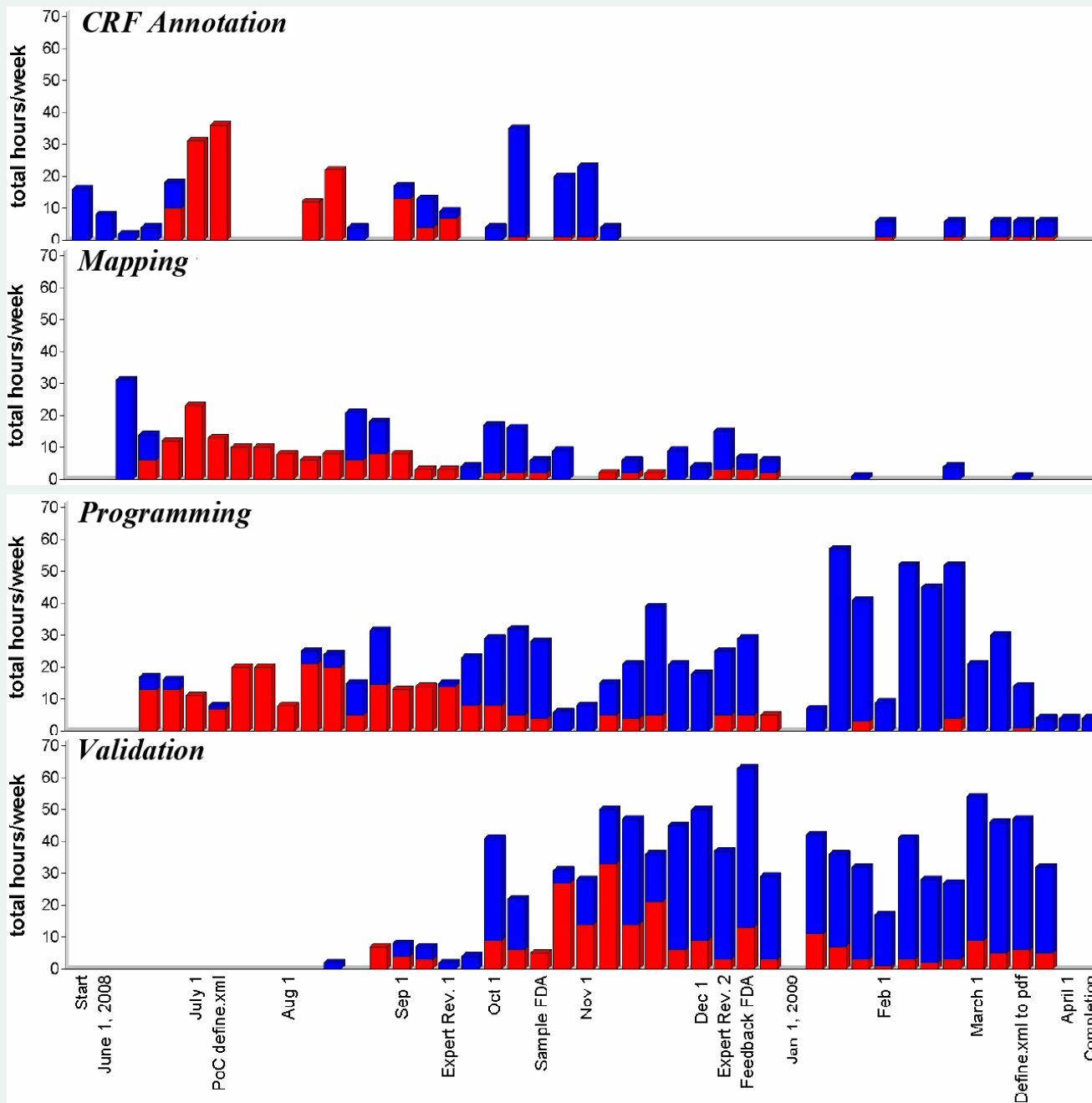
# Loading problem

- Additional information was requested from the FDA.
- FDA response: Misspecification of the 'Mandatory' attribute related to the ItemRef element identified by OID="TS.TSVVAL1" as the possible cause of non-loading in WebSDM.
- Action taken: We implemented the rule described on the CDISC forum ('Mandatory field in Define.xml'):

Perm(issible) variables:	Mandatory = "No"
Exp(ected) variables:	Mandatory = "No"
Req(uired) variables:	Mandatory = "Yes"
- We have not been able to conduct a final test to check whether this change had resolved the WebSDM loading issue.

# Project Activity Metrics

<b>Activity</b>	<b>Total hours</b>	<b>%</b>
CRF annotation	308	8.0
Define.xml	574	14.9
Mapping	309	8.0
Meetings	267	6.9
Programming	795	20.6
Project Management	626	16.2
Validation	985	25.5
<b>Total</b>	<b>3864</b>	<b>100</b>



# Conclusion

- Define.xml
  - Toolset worth considering, no black-box scenario
- Metadata repository
  - Efficiency, quality and process management, tools & approach
- Sample submission
  - Strong milestone, Initiated eCTD creation
- Expert input
  - Domain selection and modeling, confirmation!, ad-hoc support



# Back-up Slides

# Project Activity Metrics

Study information			Metrics per activity							
			CRF Annotation		Mapping		Programming		Validation	
Study number	# CRF pages	# SDTM Domains	Hours	Hours / page	Hours	Hours / domain	Hours	Hours / domain	Hours	Hours / domain
001	45	25	143	3.2	158	6.3	215	8.6	322	12.9
002	46	27	47	1.0	14	0.5	43	1.6	199	7.4
003	25	23	27	1.1	14	0.6	41	1.8	119	5.2
004	20	23	23	1.2	12	0.5	70	3.0	69	3
006	24	25	19	0.8	16	0.6	75	3.0	91	3.6
011	23	23	23	1.0	4	0.2	105	4.6	84	3.7
ISS	n.a.	13	n.a.	n.a.	12	0.9	120	9.2	101	7.8
General	n.a.	n.a.	26	n.a.	79	n.a.	126	n.a.	n.a.	n.a.