



Improving CDISC SDTM Data Quality & Compliance Right from the Beginning

Bharat Chaudhary, Cytel
Padamsimh Balekundri, Cytel

Session CD08 – PhUSE 2015, Vienna

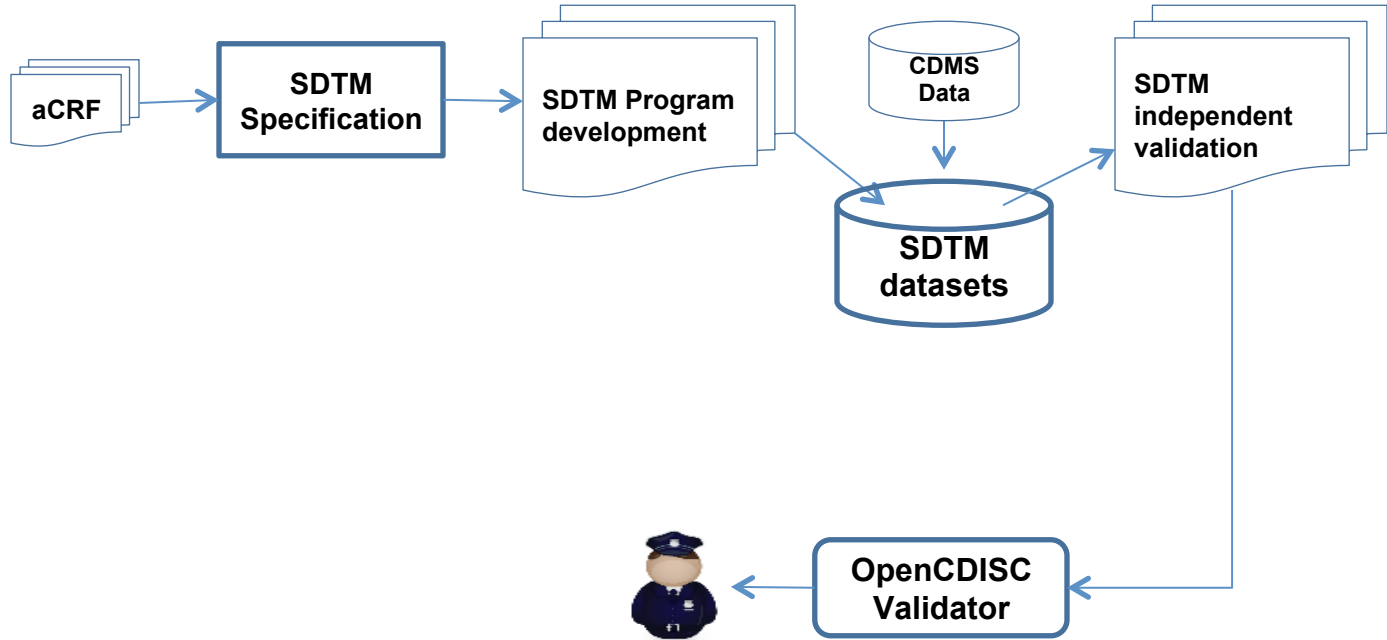
Agenda

- Background
- Overview: SDTM Development
- The Problem: In SDTM Development
- Our Solution: Set of Utilities
- Summary

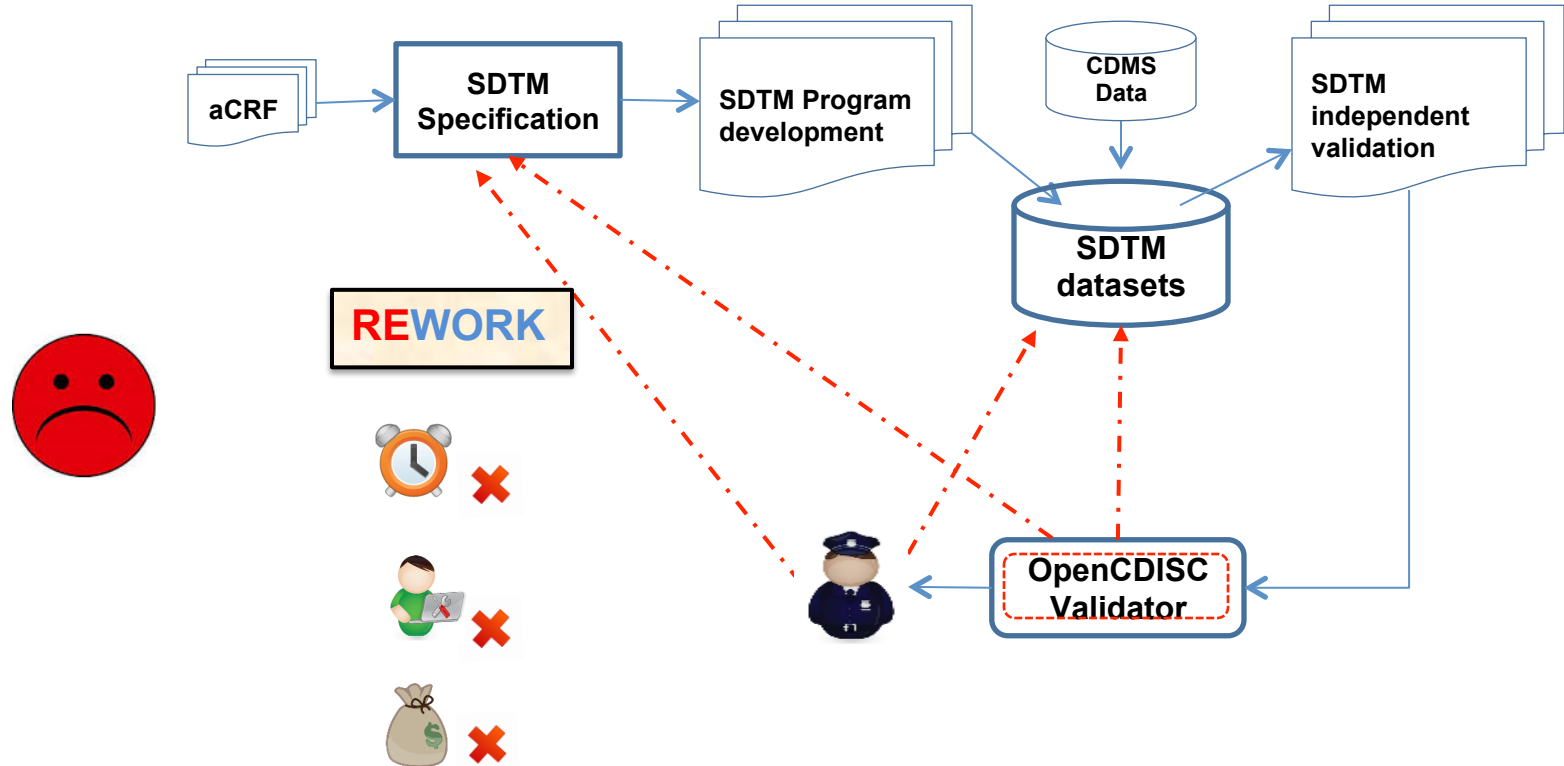
Background

- SDTM (Study Data Tabulation Model)
- Beneficial for clinical data reviewers & analysts as it brings standardization in presenting the data
- Complexity of standards - Risk of compliance issue – Rework
- Validation tools & Techniques - Limitations
- Scope for Improvement – Ideas

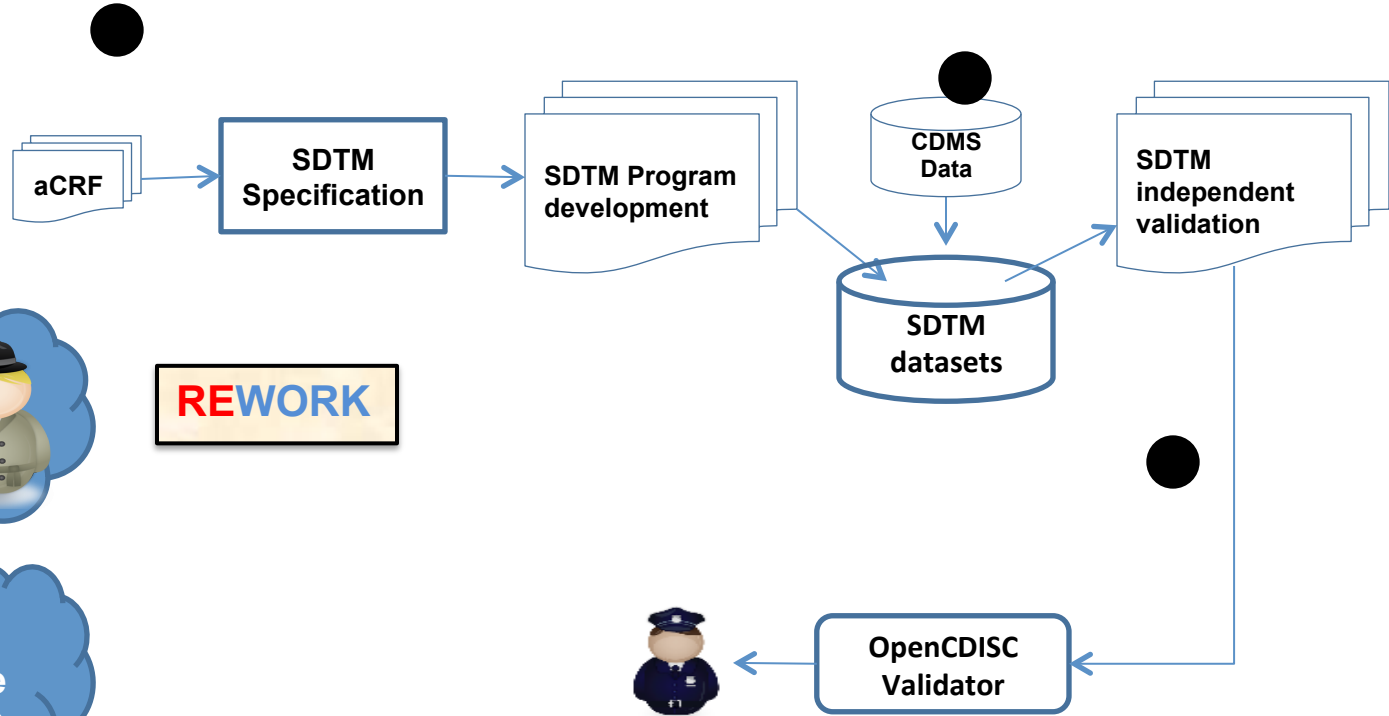
Overview: SDTM Development



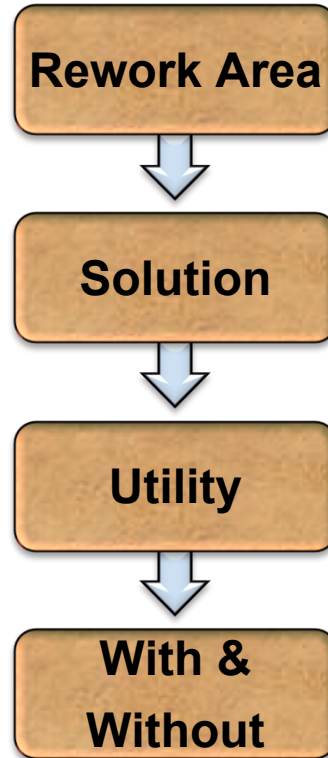
The Problem: In SDTM development



Solution

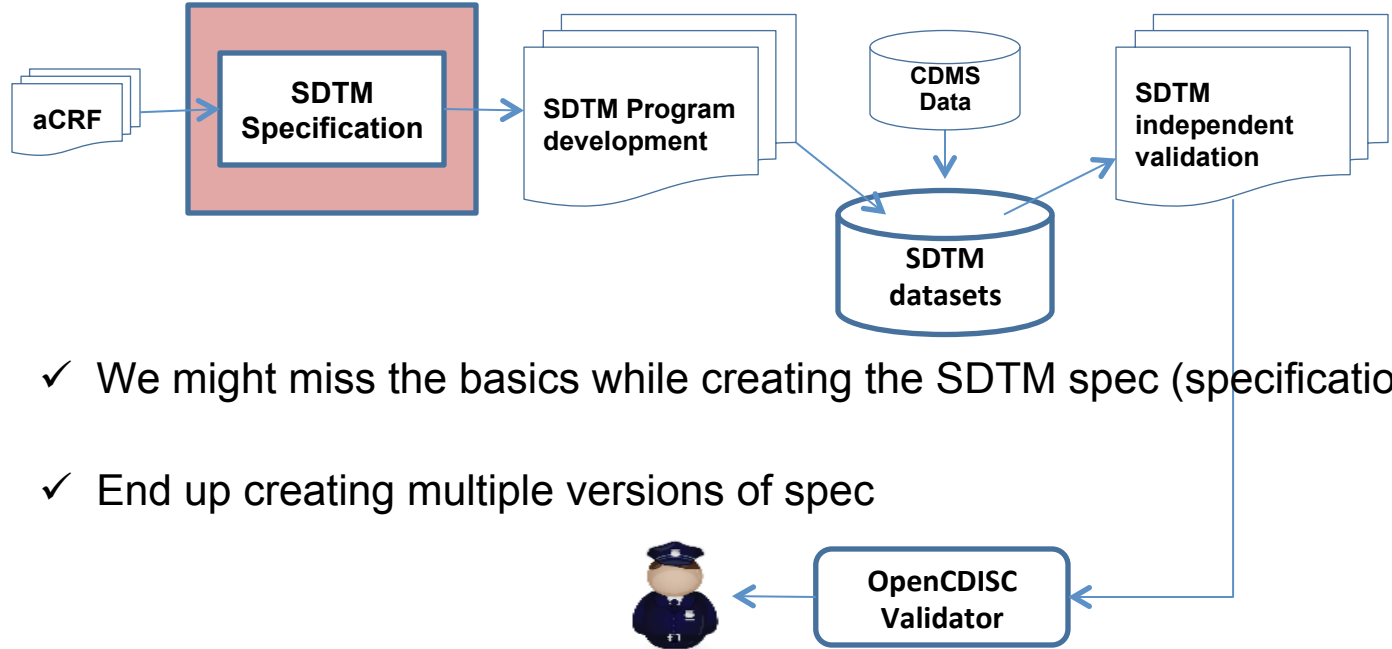
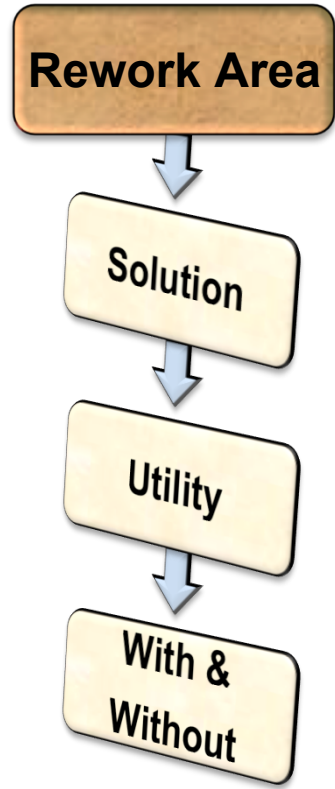


Steps to go through



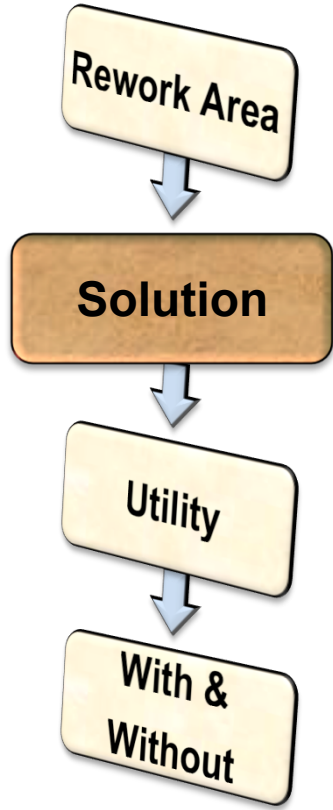
Solution 1

Solution 1



- ✓ We might miss the basics while creating the SDTM spec (specification)
- ✓ End up creating multiple versions of spec

Solution 1



SDTM Spec Checker (SSC)



- ✓ Identify compliance issues at spec level, even before programming
- ✓ Minimize the number of spec versions, which saves time & efforts

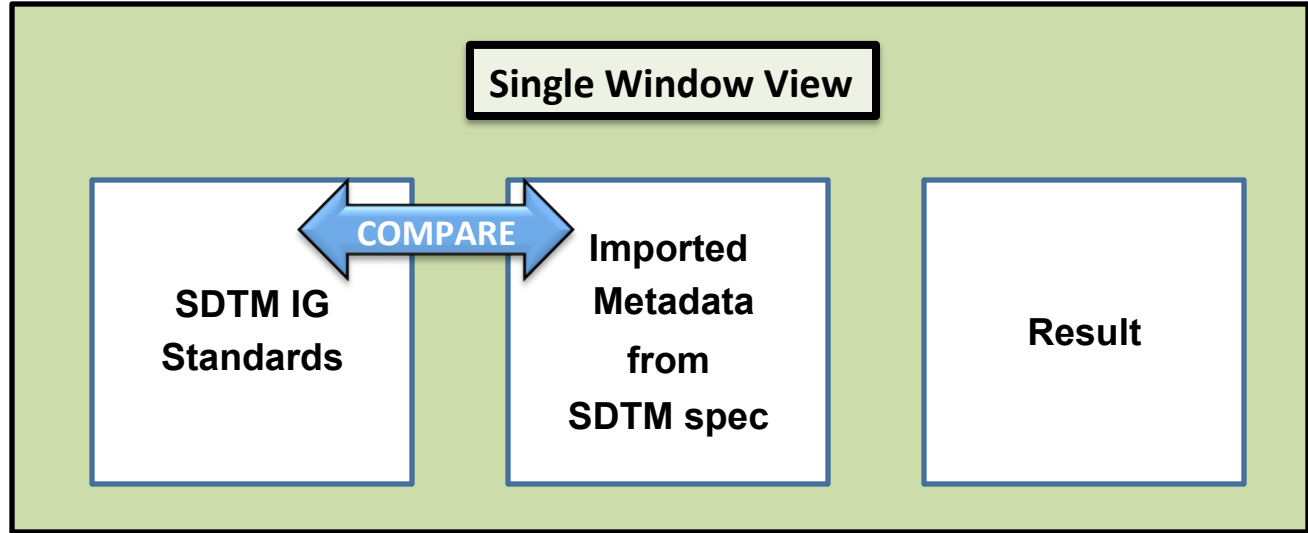
Solution 1



(SSC)

Framework

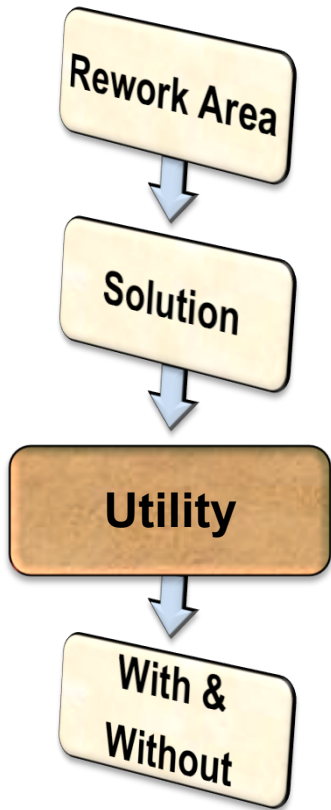
Technology:	MS Excel
Inputs:	SDTM Spec
Execution Time:	< 1 Min





(SSC)

Solution 1



Execution Step

1. SDTM Mapping specification open/active

Variable Name	Variable Label	Type	Length	Controlled Terms or Format
STUDYID	Study Identifier	Char	20	
DOMIN	Domain Abbreviation	Char	2	
USUBJID	Unique Identifier	Char	30	
SUBJID	Subject Identifier for the Study	Char	8	

2. SDTM Spec Checker

The SDTM Spec Checker				Enter the SDTM Spec Name --> Test_spec.xlsx					
CDISC SDTM Version 3.1.2				Data From Spec			Results		
Domain	Variable Name	Label	Type	Label	Type	Length	Variable Check	Label Check	Type Check

Solution 1



(SSC)

Rework Area



Solution



Utility



With &
Without

The SDTM Spec Checker				Enter the SDTM Spec Name -->	Test_spec.xlsx		(SSC)		
CDISC SDTM Version 3.1.2				Data From Spec			Results		
Domain	Variable Name	Label	Type	Label	Type	Length	Variable Check	Label Check	Type Check
DM	STUDYID	Study Identifier	Char	Study Identifier	Char	20	Mapped	Label Matched	Type Matched
DM	DOMIN	Domain Abbreviation	Char	Domain Abbreviation	Char	2	Mapped	Label Matched	Type Matched
DM	USUBJID	Unique Subject Identifier	Char	Unique Identifier	Char	30	Mapped	Label doesn't match	Type Matched
DM	SUBJID	Subject Identifier for the Study	Char	Subject Identifier for the Study	Char	8	Mapped	Label Matched	Type Matched
DM	RFSTDTC	Subject Reference Start Date/Time	Char	Subject Reference Start Date/Time	Char	20	Mapped	Label Matched	Type Matched
DM	RFENDTC	Subject Reference End Date/Time	Char	Subject Reference End Date/Time	Char	20	Mapped	Label Matched	Type Matched
DM	SITEID	Study Site Identifier	Char	Study Site Identifier	Num	4	Mapped	Label Matched	Type doesn't match
DM	INVID	Investigator Identifier	Char	Investigator Identifier	Char	8	Mapped	Label Matched	Type Matched

Solution 1



(SSC)

Rework Area



Solution



Utility



With & Without

The SDTM Spec Checker				Enter the SDTM Spec Name -->	Test_spec.xlsx		(SSC)		
CDISC SDTM Version 3.1.2				Data From Spec			Results		
Domain	Variable Name	Label	Type	Label	Type	Length	Variable Check	Label Check	Type Check
DM	STUDYID	Study Identifier	Char	Study Identifier	Char	20	Mapped	Label Matched	Type Matched
DM	DOMIN	Domain Abbreviation	Char	Domain Abbreviation	Char	2	Mapped	Label Matched	Type Matched
DM	USUBJID	Unique Subject Identifier	Char	Unique Identifier	Char	30	Mapped	Label doesn't match	Type Matched
DM	SUBJID	Subject Identifier for the Study	Char	Subject Identifier for the Study	Char	8	Mapped	Label Matched	Type Matched
DM	RFSTDTC	Subject Reference Start Date/Time	Char	Subject Reference Start Date/Time	Char	20	Mapped	Label Matched	Type Matched
DM	RFENDTC	Subject Reference End Date/Time	Char	Subject Reference End Date/Time	Char	20	Mapped	Label Matched	Type Matched
DM	SITEID	Study Site Identifier	Char	Study Site Identifier	Num	4	Mapped	Label Matched	Type doesn't match
DM	INVID	Investigator Identifier	Char	Investigator Identifier	Char	8	Mapped	Label Matched	Type Matched

Solution 1



(SSC)

Rework Area



Solution



Utility



With & Without

The SDTM Spec Checker				Enter the SDTM Spec Name -->	Test_spec.xlsx		(SSC)		
CDISC SDTM Version 3.1.2				Data From Spec			Results		
Domain	Variable Name	Label	Type	Label	Type	Length	Variable Check	Label Check	Type Check
DM	STUDYID	Study Identifier	Char	Study Identifier	Char	20	Mapped	Label Matched	Type Matched
DM	DOMIN	Domain Abbreviation	Char	Domain Abbreviation	Char	2	Mapped	Label Matched	Type Matched
DM	USUBJID	Unique Subject Identifier	Char	Unique Identifier	Char	30	Mapped	Label doesn't match	Type Matched
DM	SUBJID	Subject Identifier for the Study	Char	Subject Identifier for the Study	Char	8	Mapped	Label Matched	Type Matched
DM	RFSTDTC	Subject Reference Start Date/Time	Char	Subject Reference Start Date/Time	Char	20	Mapped	Label Matched	Type Matched
DM	RFENDTC	Subject Reference End Date/Time	Char	Subject Reference End Date/Time	Char	20	Mapped	Label Matched	Type Matched
DM	SITEID	Study Site Identifier	Char	Study Site Identifier	Num	4	Mapped	Label Matched	Type doesn't match
DM	INVID	Investigator Identifier	Char	Investigator Identifier	Char	8	Mapped	Label Matched	Type Matched

Solution 1



(SSC)

Rework Area



Solution



Utility



With &
Without

Final Report

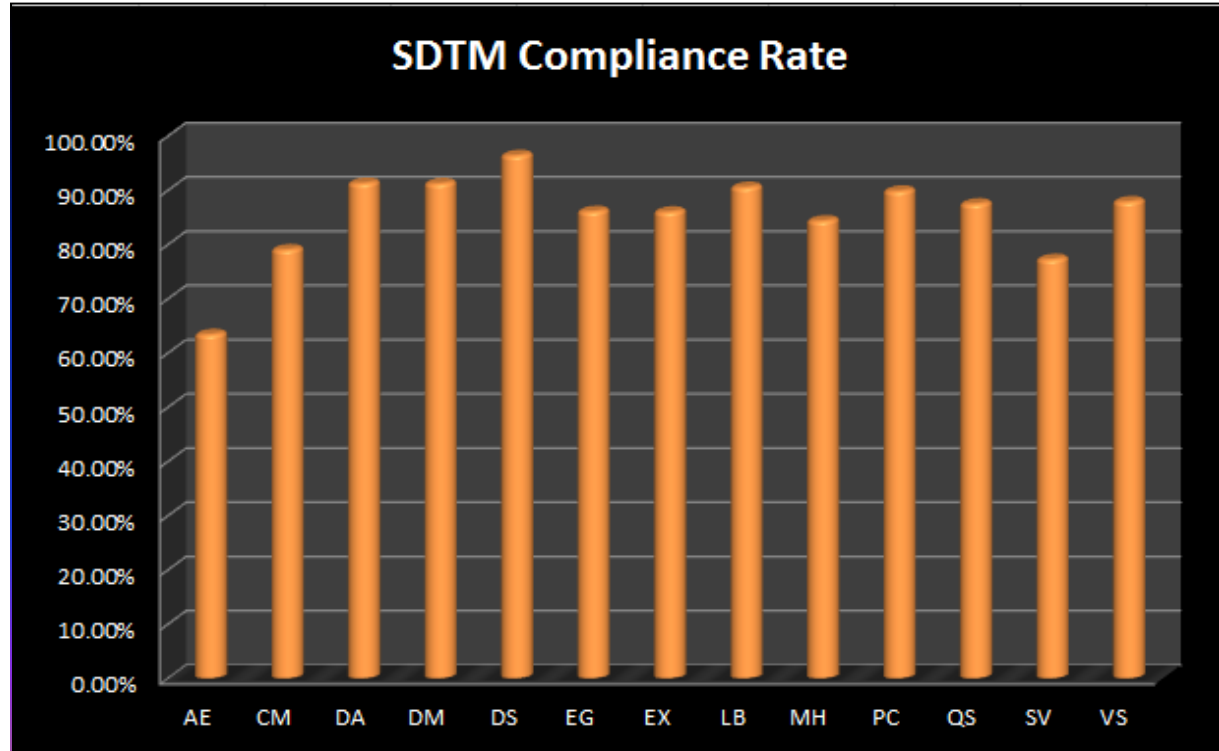
The SDTM Compliance Report									
Sr#	Domain	Total variables	Mapped Variables	Label Match	Type Match	Length Match	Total Match Value	Expected Match Value	SDTM Compliance Rate
1	AE	41	34	34	19	17	104	164	63.41%
3	CM	37	31	30	31	25	117	148	79.05%
5	DA	23	22	22	22	18	84	92	91.30%
6	DM	20	20	17	19	17	73	80	91.25%
7	DS	14	14	13	14	13	54	56	96.43%
9	EG	36	33	31	33	27	124	144	86.11%
10	EX	34	31	31	31	24	117	136	86.03%
12	LB	45	43	41	43	36	163	180	90.56%
14	MH	24	21	20	21	19	81	96	84.38%
16	PC	37	35	34	35	29	133	148	89.86%
19	QS	30	28	26	28	23	105	120	87.50%
23	SV	11	9	9	9	7	34	44	77.27%
24	VS	31	29	27	29	24	109	124	87.90%
							1298	1532	84.73%
The Overall Quality of spec									
		84.73%							

Solution 1



(SSC)

Final Report



Solution 1



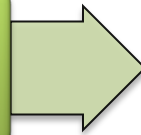
(SSC)

Average time taken to review the SDTM Specification with **10 domains**.

3-4 Hrs



<20 Min



Rework Area



Solution



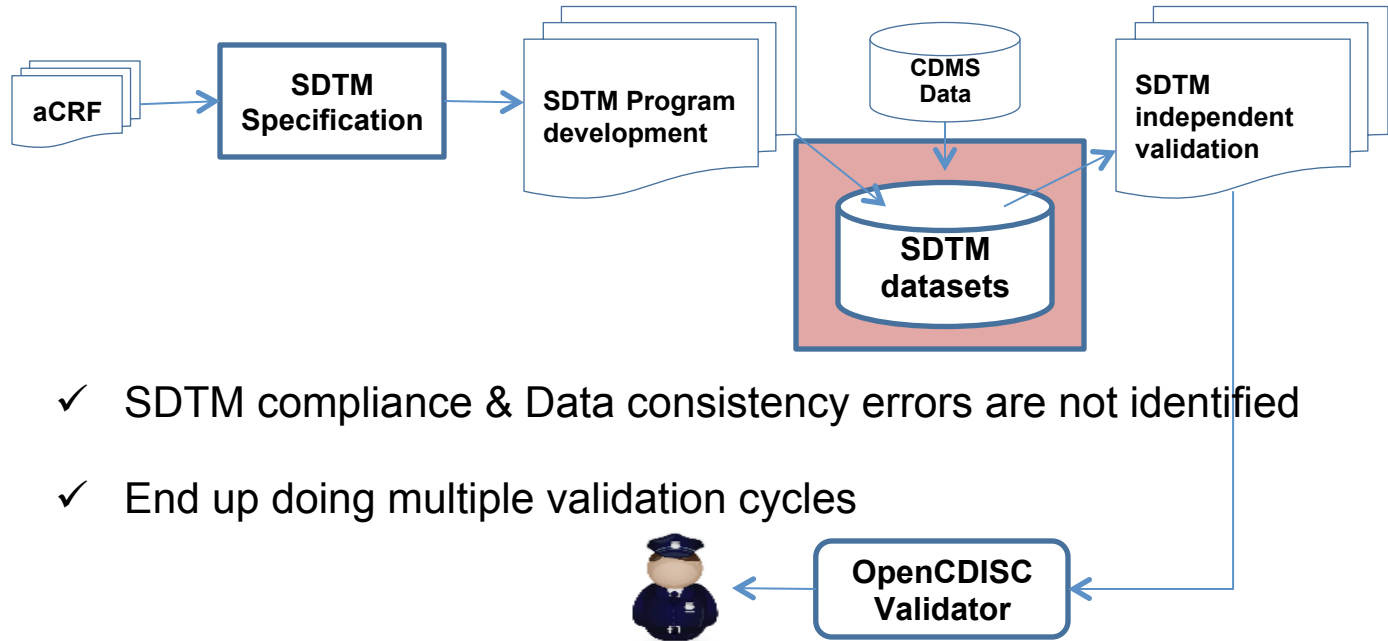
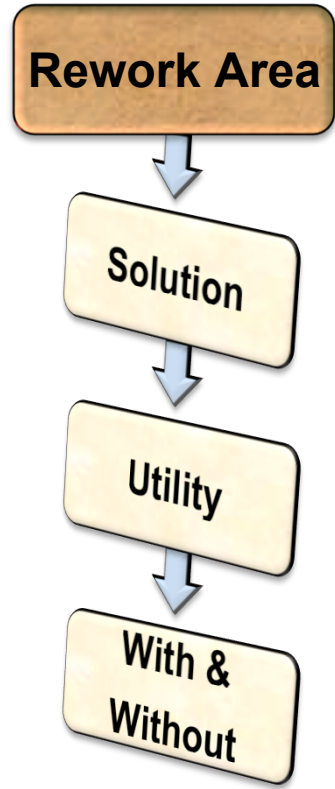
Utility



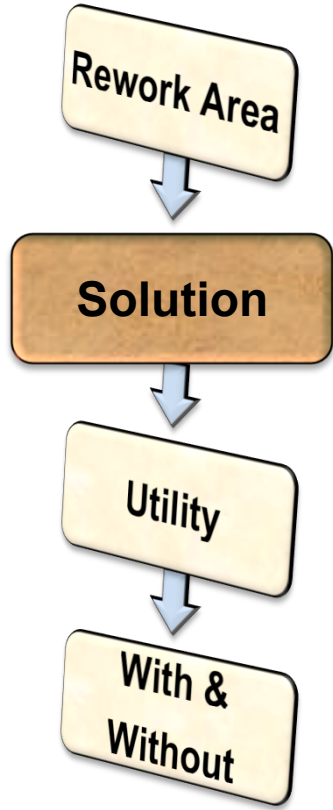
With &
Without

Solution 2

Solution 2



Solution 2



SDTM Dataset Checker (SDC)



- ✓ Identify data quality and SDTM compliance issues at development stage
- ✓ Minimize the number of validation cycle, which saves time & efforts

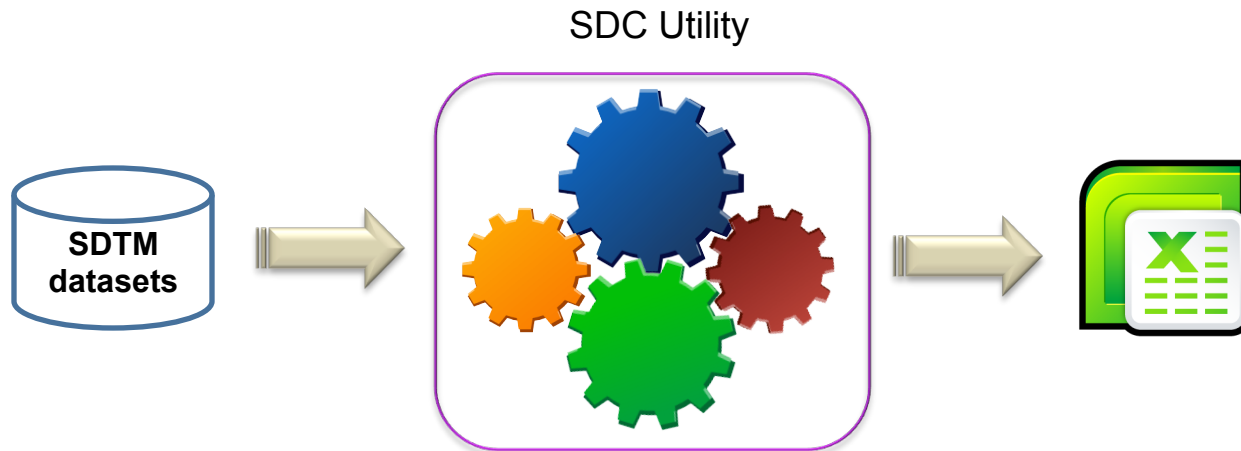
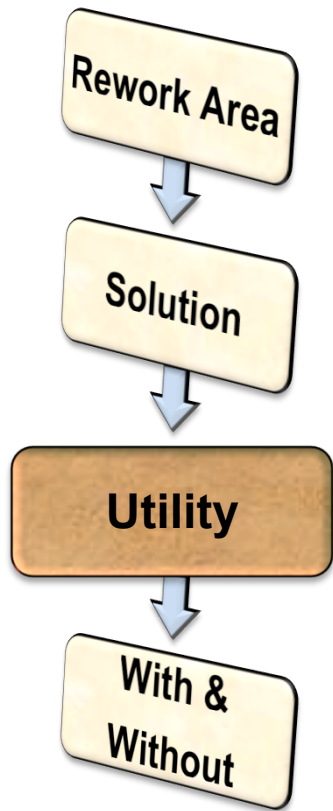
Solution 2



(SDC)

Technology:	SAS Macro
Inputs:	SDTM SAS Dataset
Execution Time:	< 5 Min

Framework





Solution 2



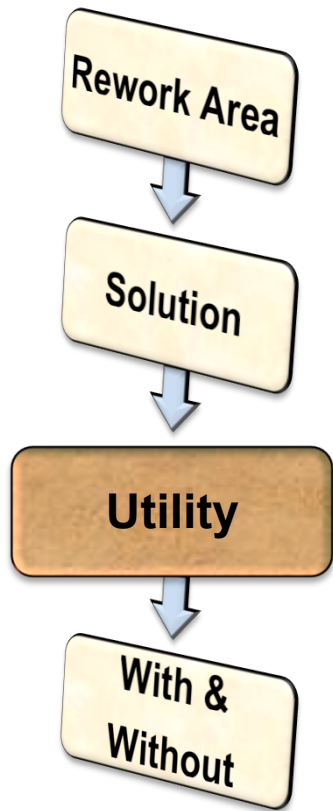
Framework



Master Macro

- ✓ Having inbuilt checks
- ✓ First these checks will apply on the Domain

Solution 2



Framework



CDISC Master file

- ✓ It's a reference file
- ✓ Having metadata of commonly used CDISC SDTM IG versions



(SDC)

Solution 2

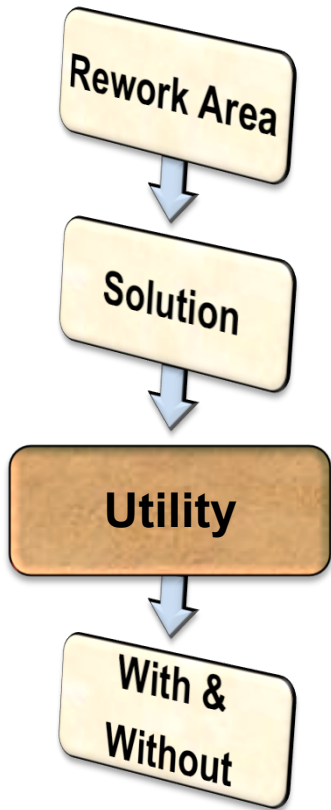
Framework



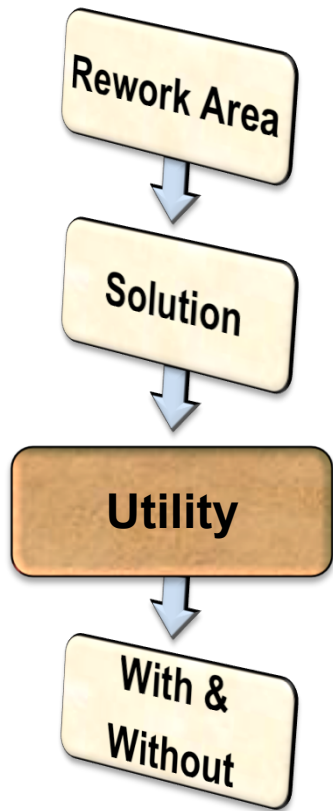
SDTM Checks file

- ✓ Having Checks in this excel file
- ✓ User can add any number of custom checks

ISSUE	CODE	VARIABLE
Birth date is greater then Reference start date	" < scan(rfstdtc,1,'T') < scan(brthdtc,1,'T')	rfstdtc,brthdtc
--BLFL (Baseline Flag) have value other than ('Y','')	--blfl not in ('Y','')	--blfl
--STDTC is after --ENDTC	--STDTC > --ENDTC	--STDTC, --ENDTC
TEMP < 0	vsorresu ne " and vstestcd = 'TEMP' and vsstresn < 0	vstestcd,vsstresn, vsorresu



Solution 2



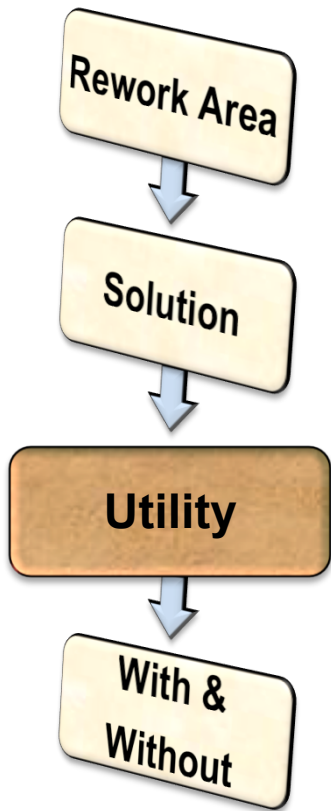
Framework



Study Specific macro

- ✓ User-defined checks
- ✓ If user wants to add any complex algorithm

Solution 2



Execution Step

- To apply on all domains in sdtm library :

```
%sdtm_chk (sdtm_v=3.1.2, lib= sdtm);
```

- To apply on specific domain (i.e. LB) :

```
%sdtm_chk (sdtm_v=3.1.2, lib= sdtm, check=lb );
```



Solution 2



(SDC)



Final Report

REPORT of SDTM Checker: 2015-08-18T01:20

ISSUE	FINDING	DOMAIN	STUDYID	USUBJID	SEQ
Duplicate records	Duplicate records group 001	AE	XX	xx-12	15
Duplicate records	Duplicate records group 001	AE	XX	xx-12	16
Null value in Required variable	Null value in Required variable [AEDECOD]	AE	XX	XX-07	4
Null value in Required variable	Null value in Required variable [COVAL]	CO	XX	xx-09	1
Label of variables are not as per CDISC SDTM IG v 3.1.1	Label of variable AGE "Age" should be "Age in AGEU at RFSTDTC"	DM	XX		
Order of variables are not as per CDISC SDTM IG v 3.1.1	Order of variable 19 (VISITNUM) should be 20	EG	XX		
Order of variables are not as per CDISC SDTM IG v 3.1.1	Order of variable 20 (VISIT) should be 19	EG	XX		
VISIT vs DATE issue	Date of CYCLE 10 DAY 15 [2013-01-06] {seq=1} < Date of CYCLE 10 DAY 8 [2013-12-30] {seq=48}	EX	XX	XX-03	

✓ Easy navigation & filtering can be applied for better understanding of issues



(SDC)

Solution 2

Real Life Example

- Date of early visit > Date of later visit

ISSUE	VSTESTCD	VISITNUM	VISIT	VSDTC	VSDY
VISIT vs. --DTC issue	WEIGHT	25.00	CYCLE 8 DAY 8	2013-04-03	-205
VISIT vs. --DTC issue	WEIGHT	24.00	CYCLE 8 DAY 1	2014-03-27	154

- Duplicate records

FINDING	DOMAIN	USUBJID	EGSEQ	EGTEST	EGORRES	EGMETHOD	VISIT	EGDTC
Duplicate records group 001	EG	ABC-001	46	ECG Interpretation	NORMAL	12 LEAD STANDARD	Week 3	2012-11-18
Duplicate records group 001	EG	ABC-001	47	ECG Interpretation	NORMAL	12 LEAD STANDARD	Week 3	2012-11-18

- Unexpected Results

ISSUS	VSTESTCD	VSORRES	VSSTRESC	VSORRESU	VSSTRESN	VSSTRESU
Unexpected result	RESP	0	0	breaths/min	0	breaths/min
Unexpected result	TEMP	100	100	C	100	C
Unexpected result	PULSE	0	0	beats/min	0	beats/min

Rework Area

Solution

Utility

With & Without

Solution 2



Real Life Example

- Non missing --ORRES & --ORRESU but --STRESC is blank

FINDING	USUBJID	LBTESTCD	LBORRES	LBORRESU	LBSTRESC
Missing value for LBSTRESC, when LBORRES and LBORRESU are provided	ABC-003	XYZ	0.4	g/L	

- Incorrect order of variables, USUBJID before DOMAIN

FINDING	STUDYID	USUBJID	DOMAIN	MHSEQ
Order of variable 2 (USUBJID) should be 3	ABC			.
Order of variable 3 (DOMAIN) should be 2	ABC			.



Solution 2



(SDC)

Average time taken to review the SDTM SAS Dataset with **1 domain**.

2-3 Hrs

<25 Min

Rework Area

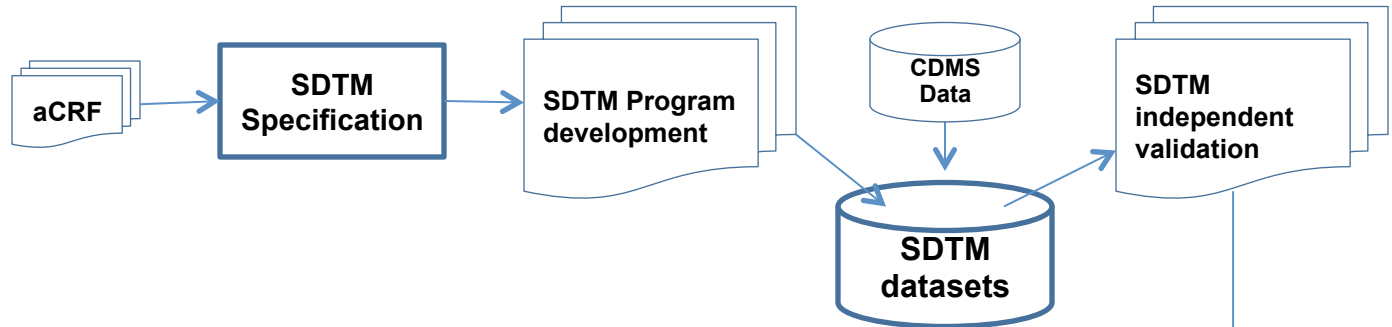
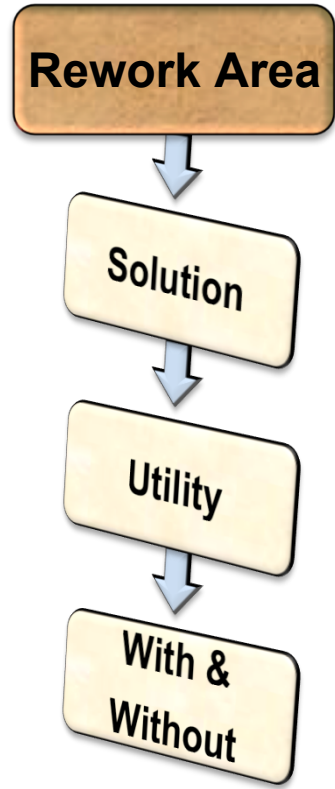
Solution

Utility

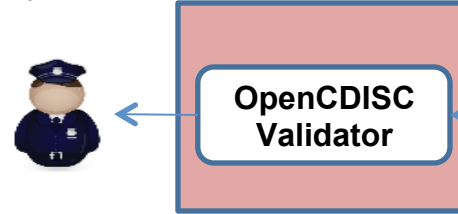
**With &
Without**

Solution 3

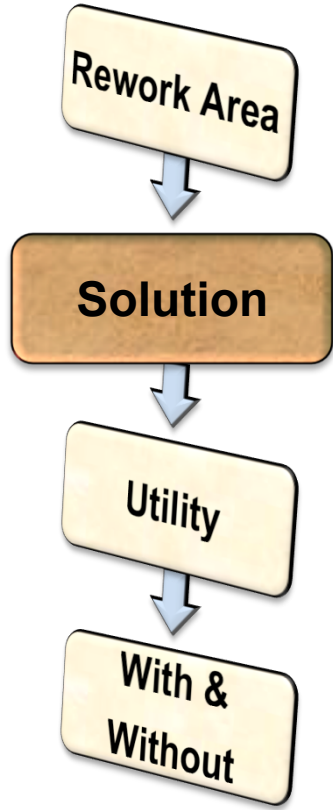
Solution 3



- ✓ Repeat runs of OpenCDISC validator for same study
- ✓ Possibility of missing to identify new issues each time we run the validator



Solution 3



OpenCDISC Report Reviewer (OCR)



- ✓ It saves time on multiple report reviews for the same study, by
 - Importing and aligning comments from previous/first report
 - Identifying new errors/warnings

Solution 3



(OCR)

Framework



Technology:	MS Excel-VBA Macro
Inputs:	Reviewed OpenCDISC Report
Execution Time:	< 2 min

Reviewed Report



OCR Utility



Importing Comments



New Summarized Report





(OCR)

Solution 3

Execution Step

- ✓ First reviewed OpenCDISC Report

Domain	Record	Count	Variables	Values	Rule ID	Message	Category	Severity	Comments
CM		1	CMDOSU	18 cc's	CT0049	Value for CMDOSU	Terminology	Warning	Mapped as per raw data
CM		2	CMDOSU	drop	CT0049	Value for CMDOSU	Terminology	Warning	Mapped as per raw data
CM		18	CMDOSU	tablet	CT0049	Value for CMDOSU	Terminology	Warning	Mapped as per raw data
CM		2	CMDOSU	tbasp	CT0049	Value for CMDOSU	Terminology	Warning	Mapped as per raw data

- ✓ New report

Domain	Record	Count	Variables	Values	Rule ID	Message	Category	Severity
CM		1	CMDOSU	18 cc's	CT0049	Value for CMDOSU	Terminology	Warning
CM		2	CMDOSU	drop	CT0049	Value for CMDOSU	Terminology	Warning
CM		18	CMDOSU	tablet	CT0049	Value for CMDOSU	Terminology	Warning
CM		2	CMDOSU	tbasp	CT0049	Value for CMDOSU	Terminology	Warning
AE			VARIABLE	AELLT	SD0058	Variable appears in Metadata		Error

Rework Area

Solution

Utility

With &
Without

Solution 3



(OCR)



Execution Step

- ✓ Utility will import comments from previous report and align them with new report

Domain	Record	Count	Variables	Values	Rule
CM		1	CMDOSU	18 cc's	CT0
CM		2	CMDOSU	drop	CT0
CM		18	CMDOSU	tablet	CT0
CM		2	CMDOSU	tbasp	CT0
AE			VARIABLE	AELLT	SD0

Dataset Summary Issue Summary De

Text Filters

- ☒ (Select All)
- ☐ Codelist error
- ☐ External data not available
- ☐ Mapped as per raw data
- ☐ Mapped as per spec
- ☒ New Errors
- ☐ Non-issue
- ☐ Other Comments
- ☐ Raw data issue
- ☐ Spec update required
- ☐ Values coming from external data
- ☐ Visit coming from external data

OK Cancel

Comments

- Mapped as per raw data
- Mapped as per raw data
- Mapped as per raw data
- Mapped as per raw data
- New Errors

- ✓ Filter out 'New Errors' and address it, Instead of reviewing entire report



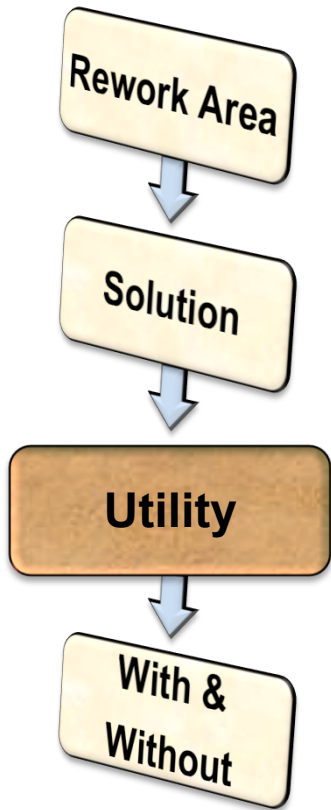
(OCR)

Solution 3

Final Report

- ✓ The OCR utility also generates the summary report

OPEN-CDISC Report Review Summary		
Sr.No	Comments Category	Found
1	Mapped as per spec	48
2	Mapped as per raw data	558
3	Raw data issue	0
4	External data not available	54
5	Codelist error	0
6	Non-issue	0
7	Spec update required	0
8	Values coming from external data	0
9	Visit coming from external data	0
10	New Errors	1
11	Other Comments	429
Total Found		1090

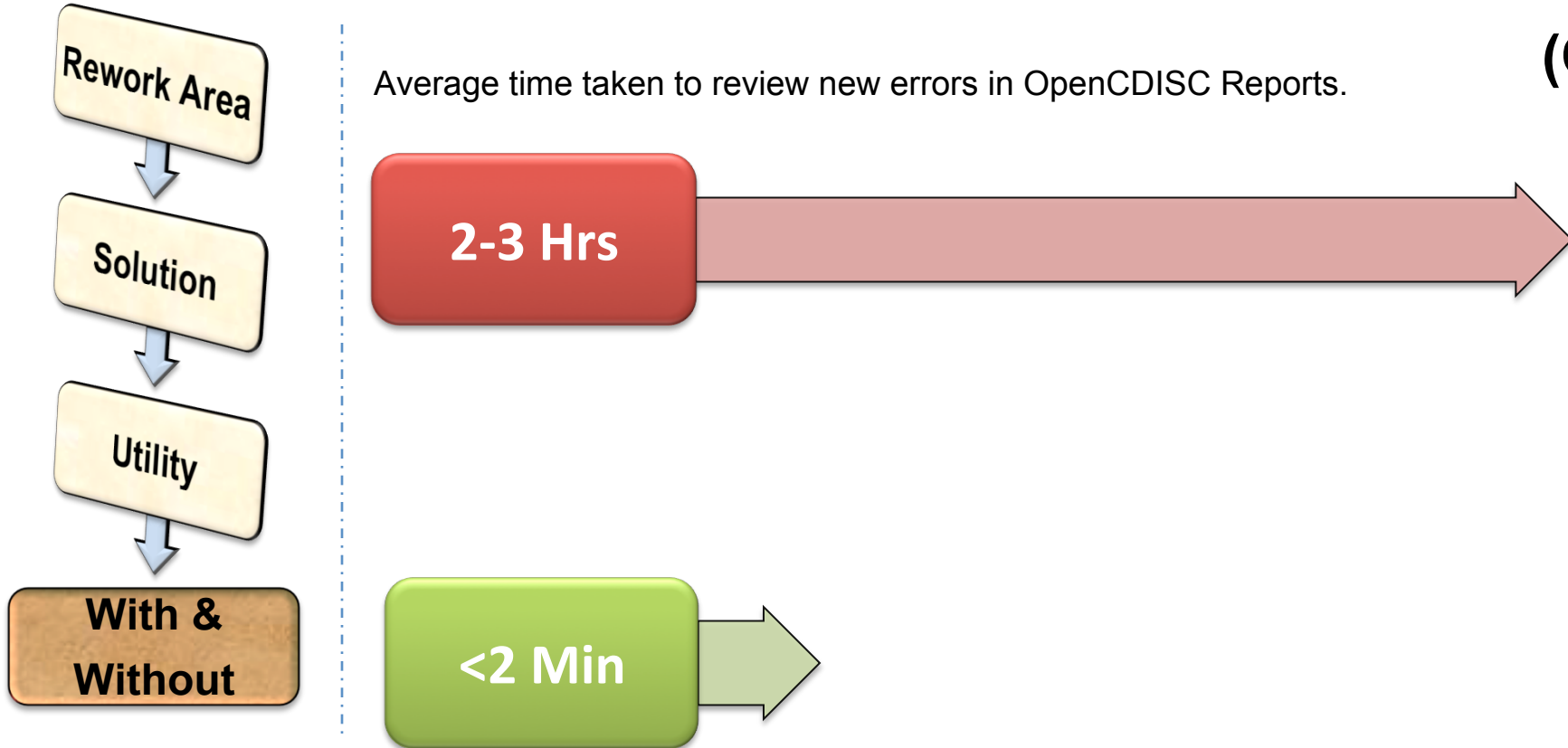


Solution 3

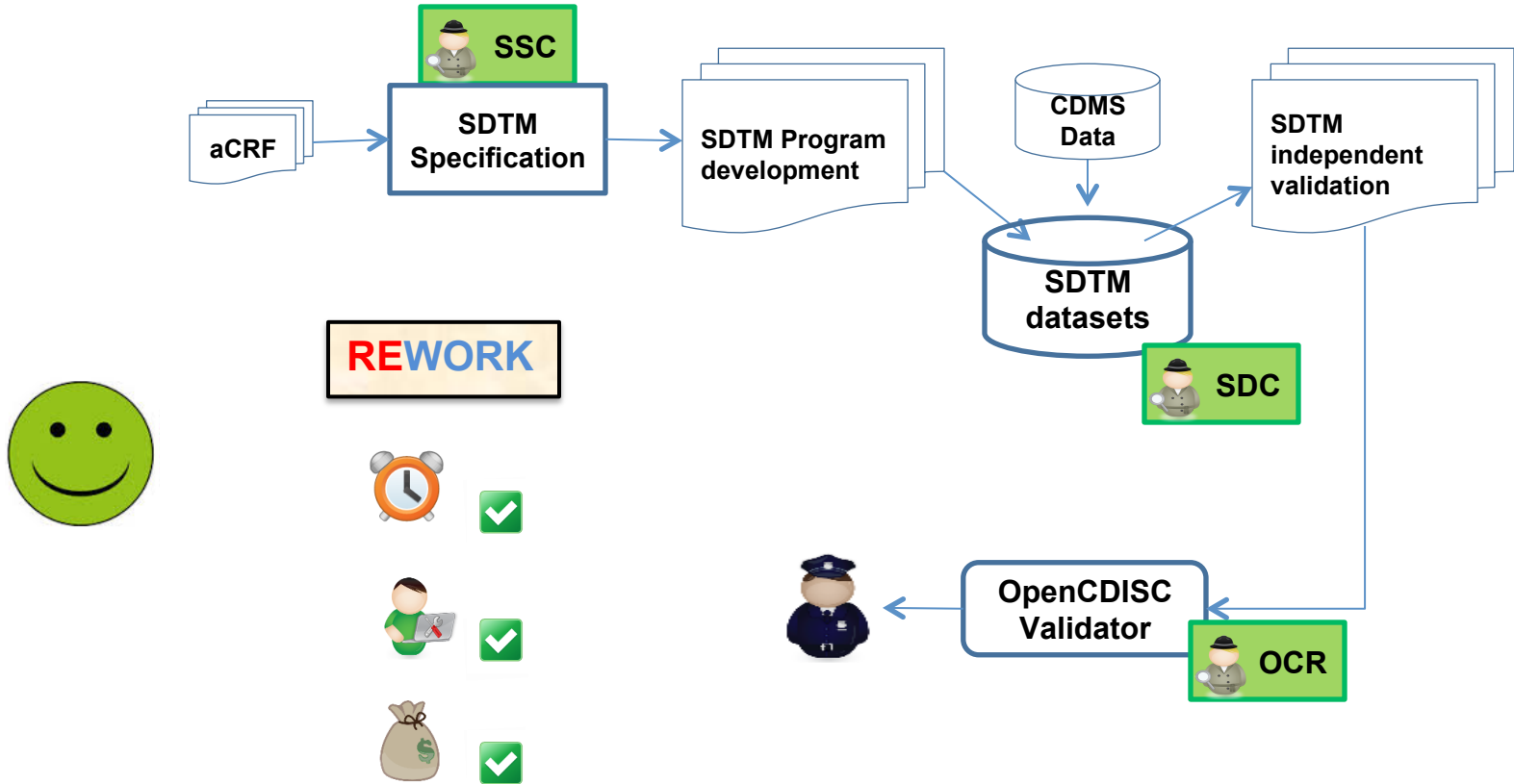


(OCR)

Average time taken to review new errors in OpenCDISC Reports.



Summary



Questions ?

Thank you !!

bharat.chaudhary@cytel.com
padamsimh.balekundri@cytel.com