

CT18

Automated Comparison of EDC Transfers: Identifying Data Changes Across Versions

Malavika Biju

Catalyst Flex

malavika.biju@catalystcr.com

Vignesh V Gopal

Catalyst Flex

vignesh.gopal@catalystcr.com

Rahulraj Raju

Catalyst Flex

rahul.raju@catalystcr.com

PHUSE APAC CONNECT

February 2026



Agenda

Background

Problem Statement

Need for a Solution & Solution Overview

Macro Design and Functionality

Dynamic Macro Logic and Processing

Dataset-Level Comparison

Variable-Level Metadata Comparison

Data-Level Comparison

Output and Reporting

Benefits and Impact

Questions?



Background

Why EDC Data Refreshes Matter

- Frequent EDC refreshes are common in modern clinical trials
- Driven by:
 - Ongoing data entry and query resolution
 - Protocol amendments
 - Database updates
- Each refresh may introduce unexpected changes
- These changes often impact downstream programming
- Changes are often not fully documented or communicated



Problem Statement

Challenges Faced by Programmers

- Metadata changes are not always documented
- Programs fail unexpectedly after refresh
- Root cause analysis is manual and time-consuming
- Particularly challenging in:
 - Oncology studies
 - Integrated analysis
 - Studies with frequent transfers

Common Issues Encountered

- Variable attribute changes
- Added or removed variables/datasets
- Structural changes in repeating forms
- Unexpected changes in record counts or missingness



Need for a Solution & Solution Overview

Why Manual Comparison Is Not Enough

- Large number of datasets and variables
- Difficult to identify what changed and where
- High dependency on programmer experience
- Lack of standardized comparison approach
- This leads to need for an automated, standardized, programmer-friendly solution

Automated SAS Macro for EDC Comparison

- Compares two EDC transfers
- Identifies metadata-level changes, dataset-level differences, and data-level discrepancies
- Produces structured, review friendly Excel output
- Easy to reuse across studies and therapeutic areas

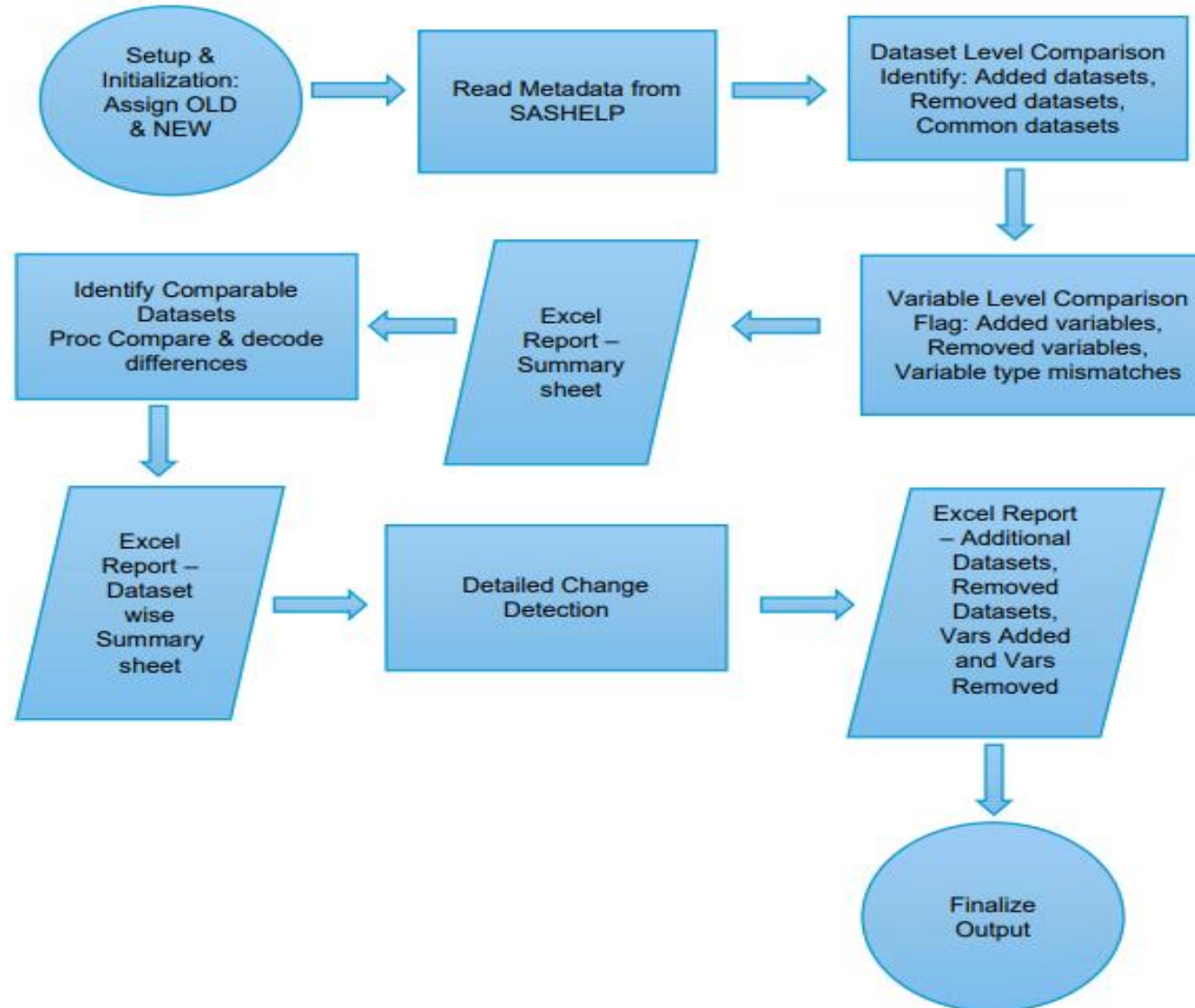


Macro Design and Functionality

Macro Architecture

- Input:
 - Old EDC library
 - New EDC library
- Uses:
 - SASHELP metadata tables
 - PROC COMPARE
 - Dynamic macro logic
- Output:
 - Multi-sheet Excel comparison report

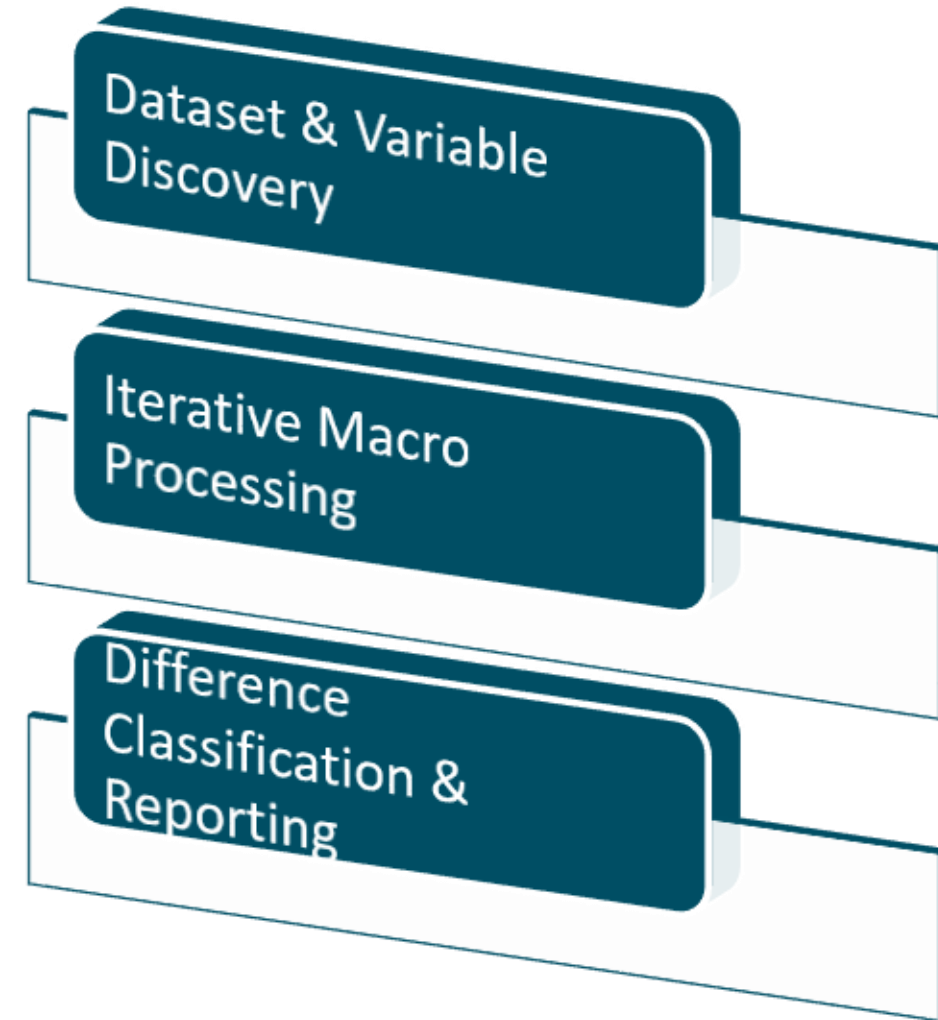
Macro Design and Functionality (continued)





Dynamic Macro Logic and Processing

- Adapts automatically to EDC refresh changes
- Metadata-driven detection (SASHELP.VTABLE & VCOLUMN)
- Iterative dataset processing via macro loops
- PROC COMPARE SYSINFO decoded for differences
- Clear summaries for early impact assessment (SDTM, ADaM, TFL)

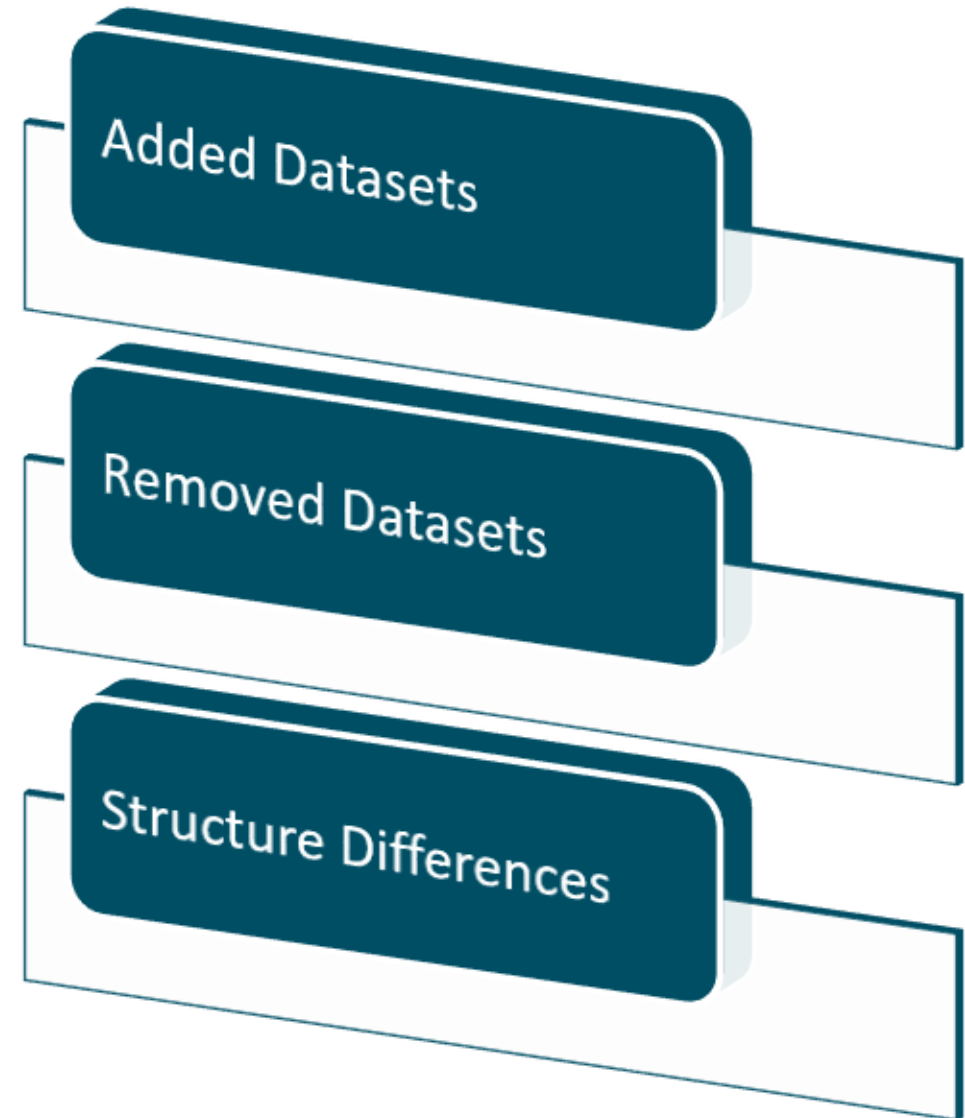




Dataset-level Comparison

What Is Compared?

- Datasets present in OLD vs NEW EDC libraries
- Identifies:
 - Newly added datasets
 - Removed datasets
 - Observation count differences
 - Variable count differences

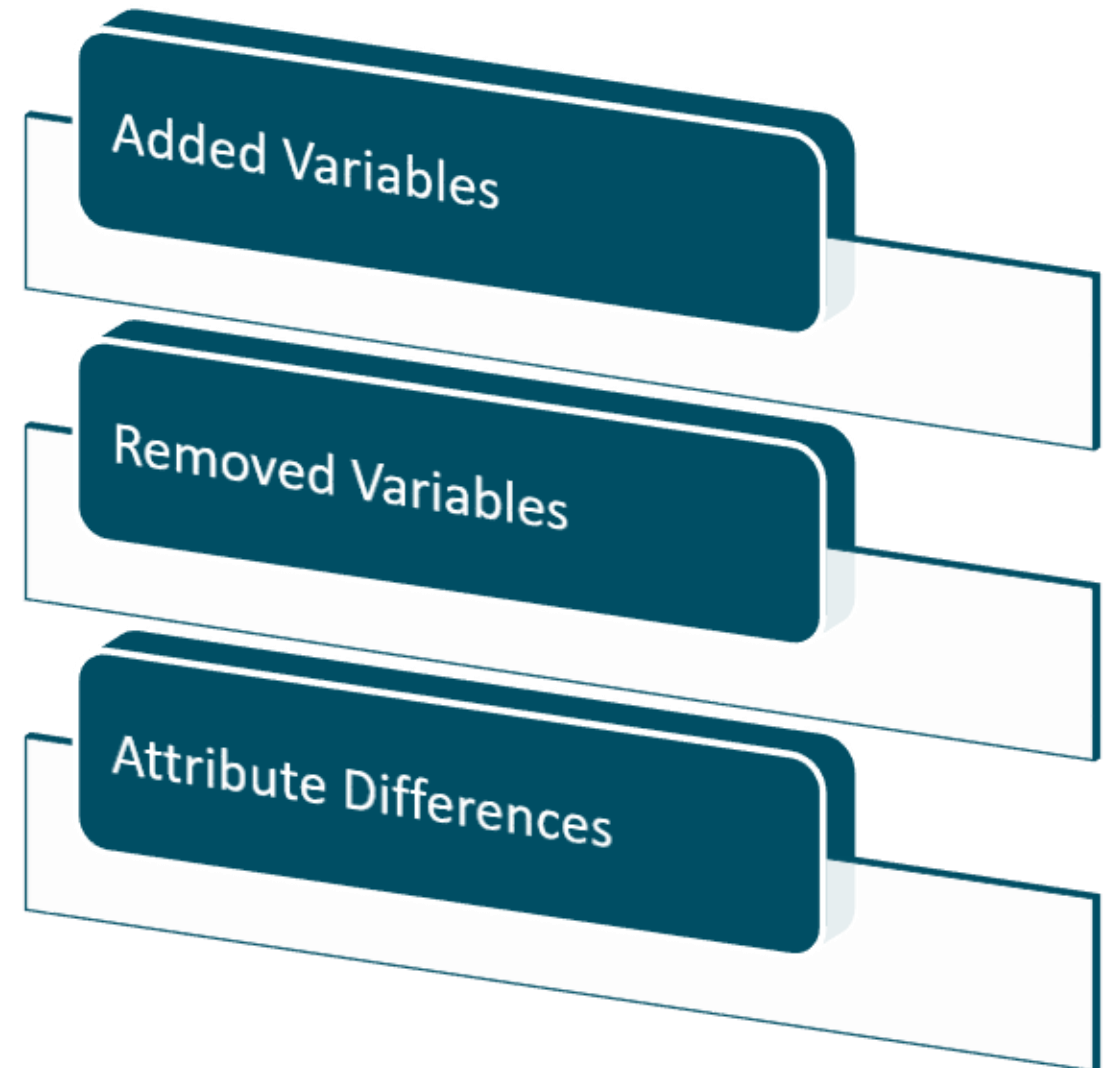




Variable-level Metadata Comparison

Metadata Validation using SASHELP.VCOLUMN

- Variable presence (added/removed)
- Variable type differences
- Length and attribute inconsistencies
- Changes impacting SDTM mapping logic





Data-level Comparison

PROC COMPARE – Based Validation

- Dataset-wise PROC COMPARE execution
- Uses key identifiers:
 - Subject
 - Record ID
 - Record Position
- Non-essential tracking variables are excluded
- SYSINFO decoded to classify differences:
 - Observation changes
 - Attribute changes
 - Value-level differences



Output and Reporting

Excel-based Comparison Report

- Overall dataset summary
- Added/Removed datasets
- Dataset wise comparison results
- Variable level metadata differences
- Detailed record level changes
- Comparison results are organized into separate Excel Sheets for easy review

Output and Reporting (continued)

	A	B	C	D	E	F	G	H
1	Summary Parameters	Count						
2	No. of datasets in old transfer	89						
3	No. of datasets in new transfer	89						
4	No. of datasets in new but not in old transfer	5						
5	No. of datasets in old but not in new transfer	5						
6	No. of datasets showing observation number mismatch	0						
7	No. of datasets showing variable number mismatch	2						
8	No. of datasets with variable type mismatch	0						
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								

<
>
Overall Summary
Additional Datasets
Removed Datasets
Dataset wise Summary
Var ...
+

Fig 1: Overall Summary

	A
1	New Datasets
2	AE_DAT
3	EG
4	PK_APS
5	PK_EVENAMIDE_BPAA
6	RAND_BLIND
7	
8	
9	

Fig 2: Additional Datasets

	A
1	Removed Datasets
2	BLOODSAM
3	CM
4	DV
5	EEG
6	PP
7	
8	
9	

Fig 3: Removed Datasets

Output and Reporting (continued)

	A	B	C	D	E	F	G	H	I
1	Dataset Name	Diff. in Dataset Label	Diff. in Observation Number	Diff. in Variable Length	Diff. in Variables	Diff. in Values	Diff. in Variable Type		
2	ABT								
3	AE			X	X				
4	CAM								
5	CDSS								
6	CGIC								
7	CGIS								
8	CLZBM								
9	CP								
10	CSSRS								
11	CSSRS1								
12	CYP								
13	DA								
14	DA1								
15	DA3								
16	DM				X				
17	ECG								
18	ECG1								
19	ECG2								

<
>
...
Removed Datasets
Dataset wise Summary
Variables Added
Variables Removed

Fig 4: Dataset wise Summary

	A	B
1	Dataset name	Variable name
2	AE	aerel
3	AE	subject
4	DM	rac8
5		
6		
7		

Fig 5: Variables Added

	A	B
1	Dataset name	Variable name
2	AE	AEEVRE
3	AE	AEREL
4	AE	SUBJECT
5	DM	DMMS
6	DM	RAC7
7		
8		
9		

Fig 6: Variables Removed



Benefits and Impact

Benefits

- Early detection of changes
- Reduced manual debugging time
- Improved transparency across refreshes
- Standardized comparison process

Key Takeaways

- EDC refreshes are unavoidable
- Automated EDC comparison improves efficiency and reliability
- This SAS macro enables proactive change management
- Can be easily integrated into standard workflows

THANK YOU

QUESTIONS?