Presenter will speak in Japanese

Validation of Programs in Automated SDTM Datasets Creation

Kunihito Ebi
Product Manger, Fujitsu Limited

Meet the Speaker

Kunihito Ebi

Title: Product Manager, Life Science Solutions

Organization: Fujitsu Limited

- CDISC authorized instructor of XML Technologies since 2015
- Product Manager of CDISC-based metadata management system with SDTM automation capability at Fujitsu since 2015

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author and do not necessarily reflect the official policy or position of the organization where the author belongs.*

What will be discussed in this presentation

- This presentation will introduce validation processes for development of programs at **automated SDTM creation** and discuss effectiveness of the validation processes by comparing with the conventional **double programming** approach.

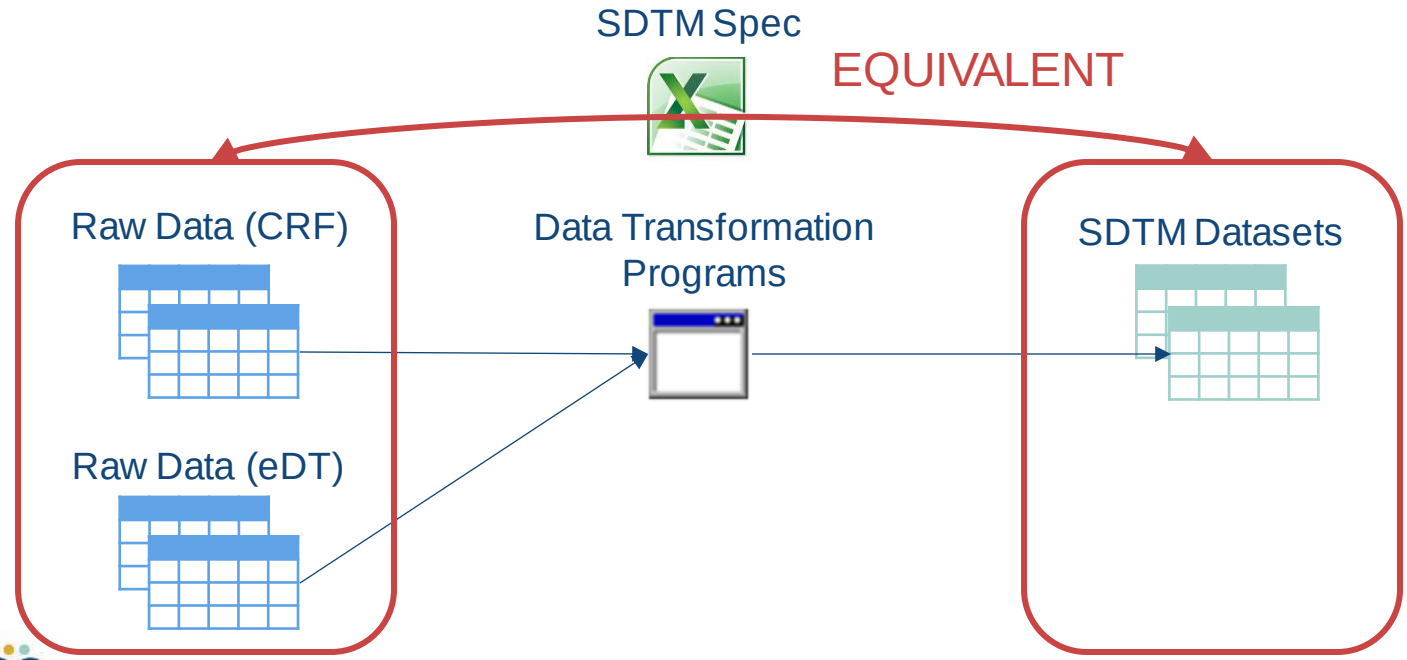
Reference

- The following reference influenced the author's understanding about quality control by the double programming approach.

Takashi Masuda, Quality Control of Statistical Analysis by Program Validation in Drug Development, 2011 SAS User Forum

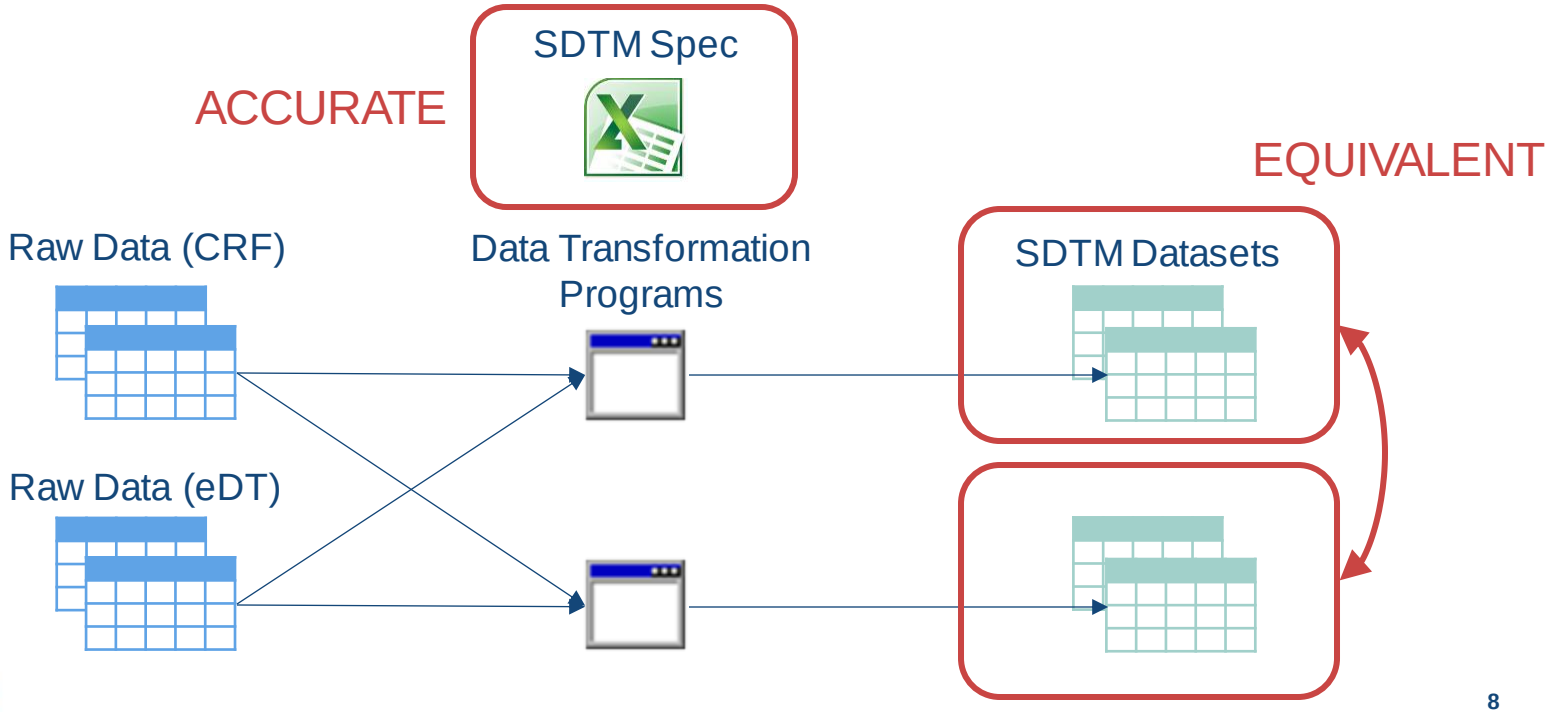
Quality of Transformation from Raw to SDTM

- The transformation process should ensure that Raw Data and SDTM Datasets are equivalent.



Quality of Transformation from Raw to SDTM - Double Programming Approach -

- This approach ensure quality of transformation by accurate SDTM Spec and transformation programs created by independent individuals.

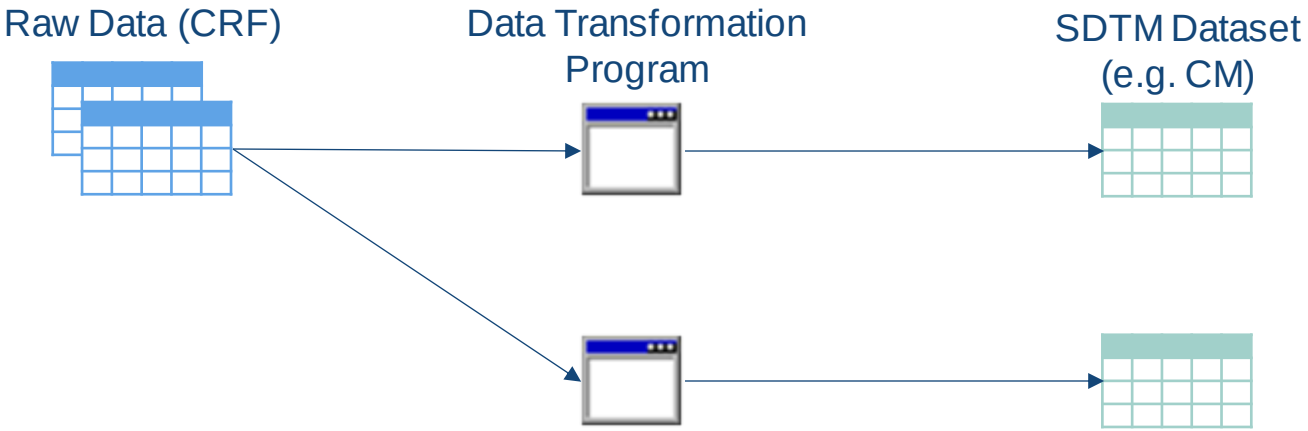


Closer Look at Double Programming Approach

- The SDTM Datasets are considered transformed accurately as per SDTM Spec when the two independent outputs are equal.
- This approach does not validate the Data Transformation Programs but ensures the quality of output.
- The outputs must be compared each time the input data changes.
- SDTM Spec must be accurate to ensure that the Raw Data and SDTM Datasets are equivalent. SDTM Datasets Review is performed to support accuracy of SDTM Spec.

Structure of Data Transformation Programs - Double Programming Approach -

- A program is usually created per target SDTM Dataset.



How SDTM Spec looks like
- Double Programming Approach -

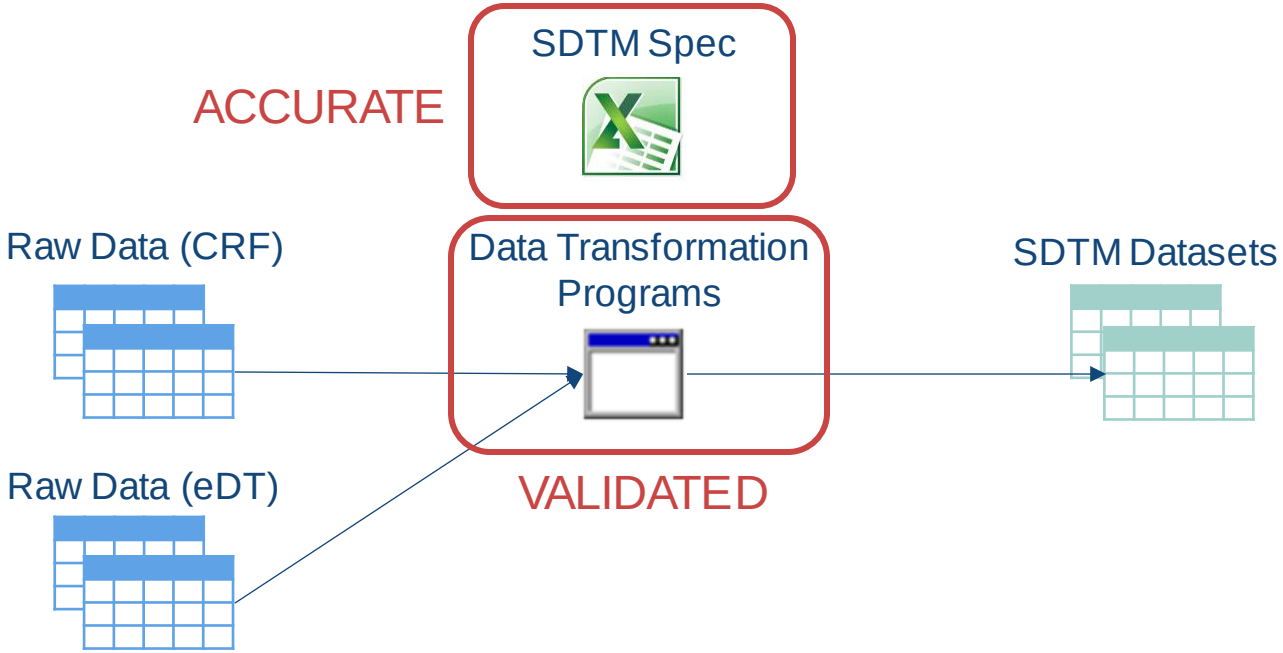
- SDTM Spec often describes mapping rules from Raw to SDTM in a detailed manner in order to explain them accurately to programmers.

- Example:

SDTM Variable	Mapping Rule
CMCAT	If Raw.CONMED1.CMTRT is not missing then 'PREVIOUS MEDICATION', else if Raw.CONMED2.CMONGO = 'Y' then 'ONGOING CONCOMITANT MEDICATION', else 'COMPLETED CONCOMITANT MEDICATION'

Quality of Transformation from Raw to SDTM - Automated Approach -

- This approach qualifies the transformation by accurate SDTM Spec and validated data transformation programs.

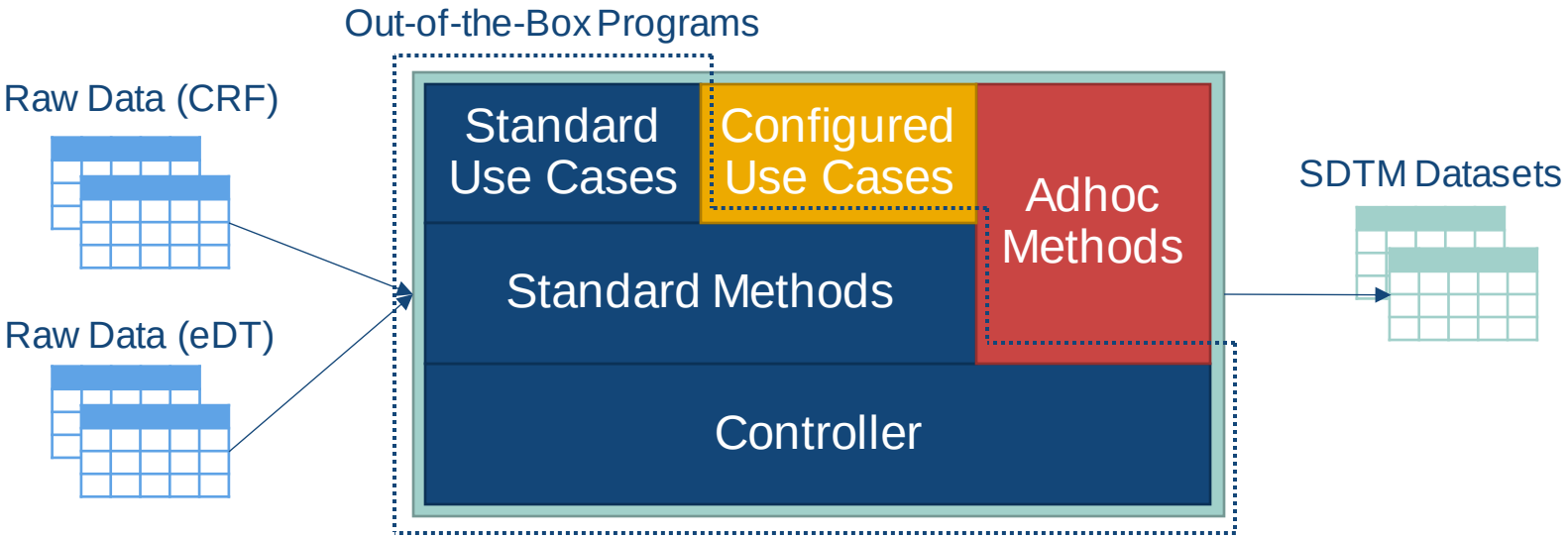


Closer Look at Automated Approach

- This approach validates the Data Transformation Programs by testing.
- The SDTM Datasets are considered transformed accurately as per SDTM Spec when validated programs are used.
- SDTM Spec must be accurate to ensure that the Raw Data and SDTM Datasets are equivalent. SDTM Datasets Review is performed to support accuracy of SDTM Spec.

Structure of Data Transformation Programs - Automated Approach -

- A set of programs is prepared for all SDTM Datasets.



Note: The program structure is dependent on an automation product.

How SDTM Spec looks like
- Automated Approach -

- SDTM Spec looks like that of Double Programming Approach but is more structured and computer-executable.
- SDTM Spec contains "Programs" (i.e. Use Case)
- Example:

Mapping Group	Source Variable	SDTM Variable	Use Case
Previous Medication	Raw.CONMED1.CMTRT	CMTRT	Copy if not missing
		CMCAT	Assign 'PREVIOUS MEDICATION'
Concomitant Medication	Raw.CONMED2.CMTRT	CMTRT	Copy if not missing
		CMCAT	`if Raw.CONMED2.CMONGO='Y' then 'ONGOING CONCOMITANT MEDICATION' else 'COMPLETED CONCOMITANT MEDICATION'`

Standard Use Case

Configured Use Case

Adhoc Method

Validation Process at Automated Approach (1)

- **Out-of-the-box programs** should be validated as per Software Development Life Cycle (SDLC) of the product.

Controller:	Mechanism to load Raw Data, control processing order, and export SDTM Datasets
Standard Method:	Standard program to create a variable in a SDTM dataset `e.g. Copy, Assign`
Standard Use Case:	Standard program with a set of parameters `e.g. Copy if not missing`

Validation Process at Automated Approach (2)

- **Configured Use Cases** should be validated via spec review.

Example:

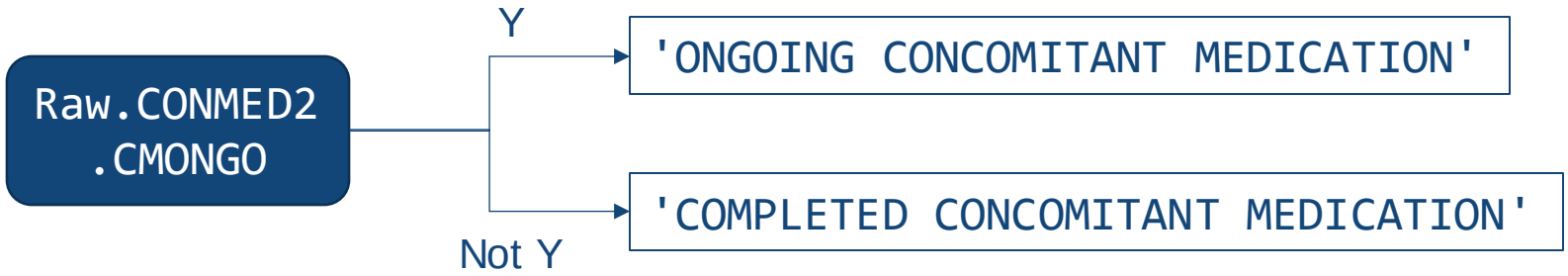
- `Assign 'PREVIOUS MEDICATION'`
- `Max EX.EXENDTC`

Validation Process at Automated Approach (3)

- **Adhoc Methods** should be validated via spec review and testing.

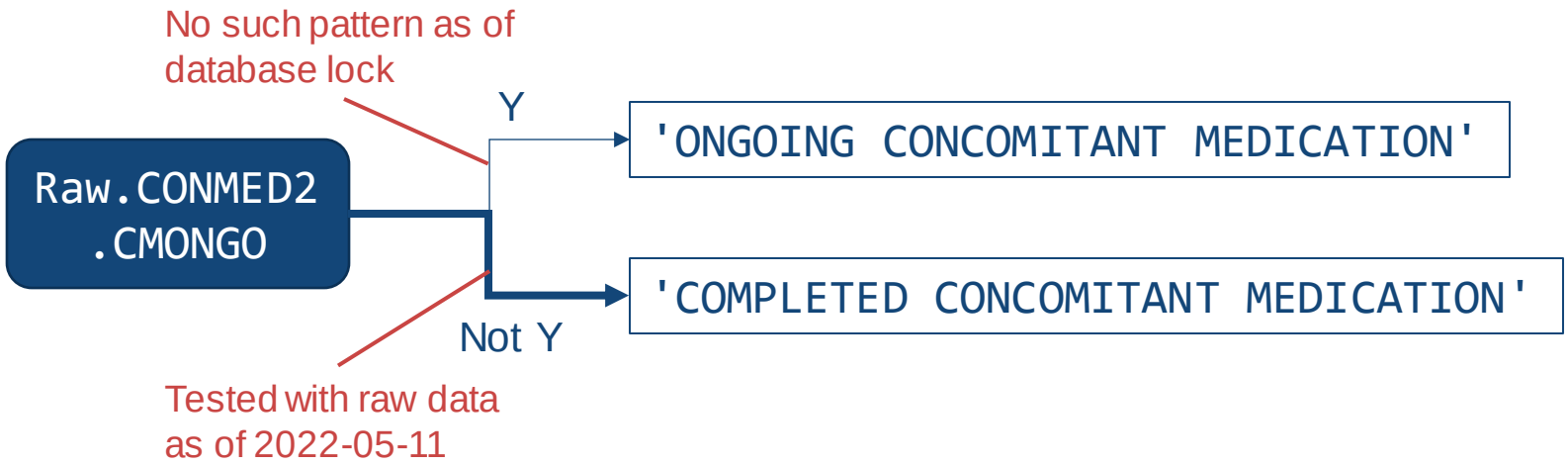
Example:

```
if Raw.CONMED2.CMONGO='Y' then 'ONGOING
CONCOMITANT MEDICATION' else 'COMPLETED
CONCOMITANT MEDICATION'
```



Adhoc Method – Test Data Challenge

- It may be challenging to prepare raw data (test data) that covers all patterns.
- In such case, it may be a realistic approach to track which patterns have been tested and which remain untested.



More Considerations about Automated Approach

- To ensure accuracy of SDTM Spec, authors/reviewers should be trained for:
  - Structure of SDTM Spec
  - Specification of Use Case
- The same applies to Double Programming Approach though.

Create records in CM dataset per Mapping Group

Create one record in CM dataset per record in the source if the Source Variable is not missing

Mapping Group	Source Variable	SDTM Variable	Use Case
Previous Medication	Raw.CONMED1.CMTRT	CMTRT	Copy if not missing
		CMCAT	Assign 'PREVIOUS MEDICATION'
Concomitant Medication	Raw.CONMED2.CMTRT	CMTRT	Copy if not missing
		CMCAT	. . .

Summary and Conclusion

- Equivalence between Raw Data and SDTM Datasets are ensured via:
  - Accurate SDTM Spec and Double Programming

    OR
  - Accurate SDTM Spec and Validated Programs

- Automated Approach validates programs via:
  - Vendor SDLC (for Out-of-the-box Programs)
  - Spec Review
  - Testing

Thank You!

For any questions, please contact the author at:
ebi.kunihito@fujitsu.com