entimo

Automated ETL – from OC to SDTM

Astrid Wolter
(Entimo AG - Berlin/Germany)

Agenda

- From OC to SDTM - automated ETL
  - Introduction & Characteristics
  - Prerequisites
  - Overview / Description of process
- Discussion
  - Metrics
  - Developed best practices
  - Challenges
- Conclusion

From OC to SDTM - automated ETL
Project initiation / introduction

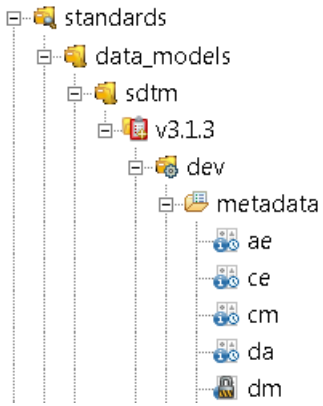
- Project participants:
  - Major large-size pharmaceutical company
- Project goals:
  - End-to-end, standards-driven initiative
  - Develop a data-model and process framework to transform data from Oracle Clinical (OC) to SDTM in an automated way
  - Scheduled overnight for multiple, selectable studies

From OC to SDTM - automated ETL
Project characteristics

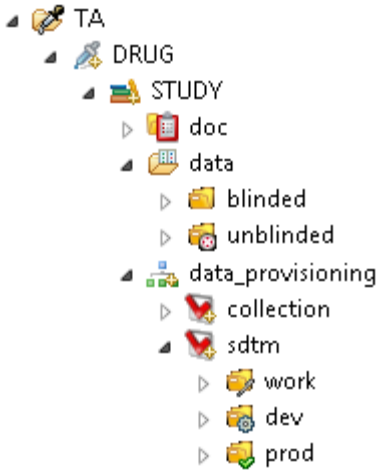
- Standards-driven:
  - Areas (global, study)
  - Study Structure
- Metadata-driven
  - Data Models (OC, SDTM)
- Framework
  - OC load
  - SDTM conversion
- Parameterized

From OC to SDTM - automated ETL Prerequisites (1)

- Global / Standards areas:
  - Data Models / Metadata
  - Programs
  - Utilities / Frameworks

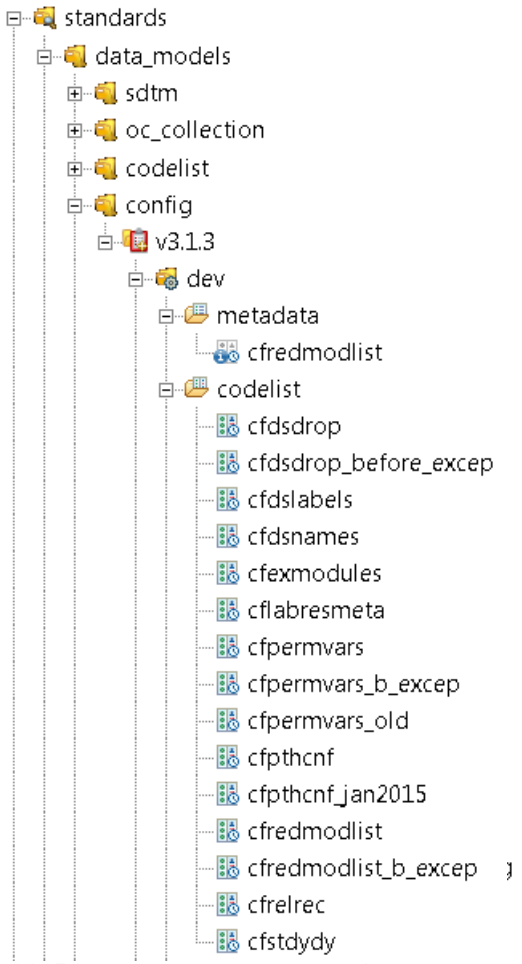


- Study area
  - Standard TA-DRUG-STUDY structure
  - Standard STUDY-DATA structure
  - Development / Production area



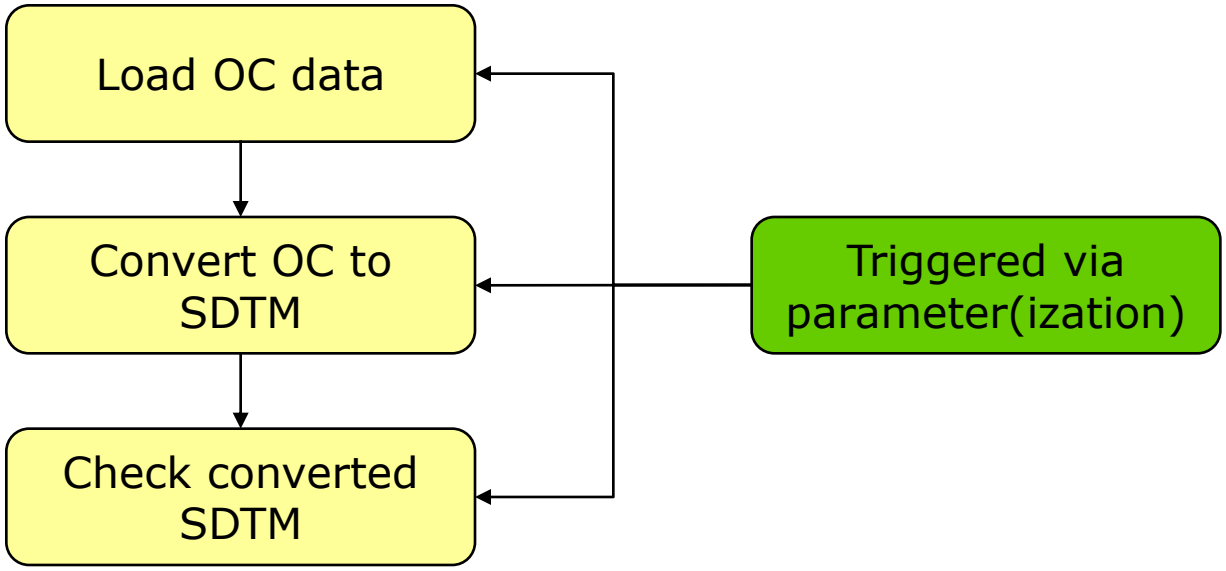
From OC to SDTM - automated ETL Prerequisites (2)

- Data Models:
  - OC Collection - Metadata
  - SDTM – Metadata
  - Codelists
  - Study Definition
  - Configuration
  - Mappings
- Utilities / Programs:
  - OC Load
  - Utility programs for SDTM conversio
- Framework:
  - SDTM conversion

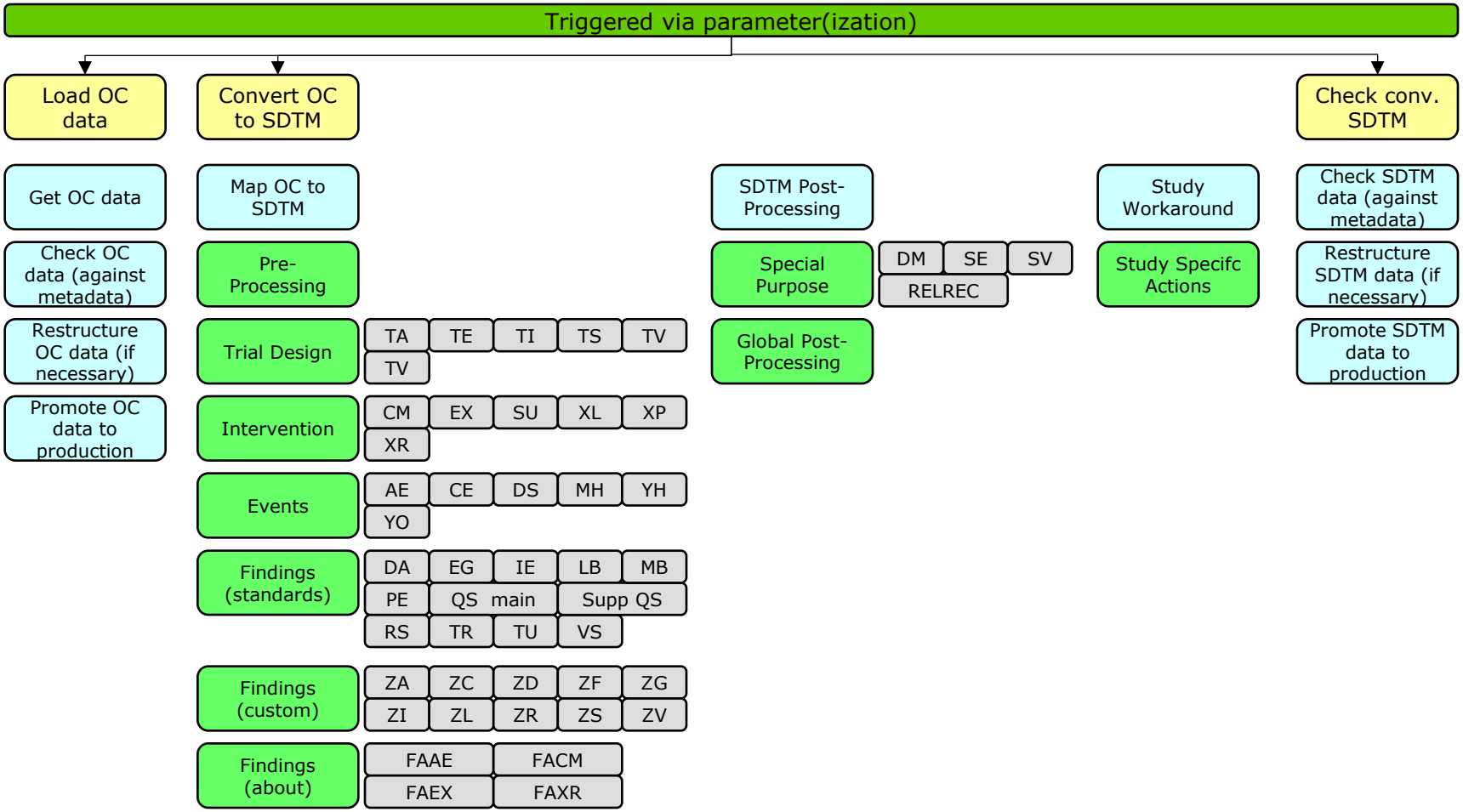


From OC to SDTM - automated ETL Overview of Process – Framework (1)

- High Level Process (Main Framework):

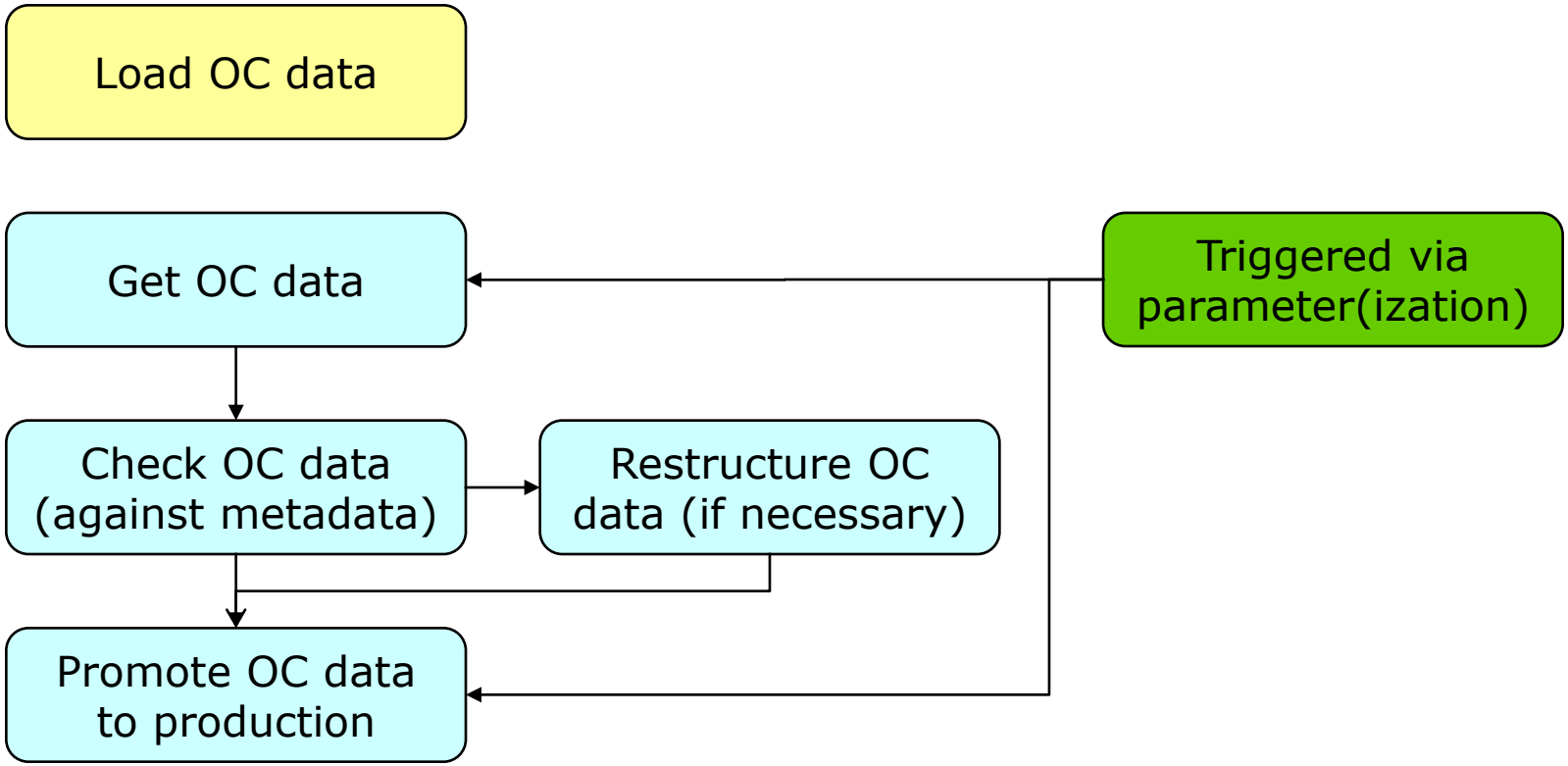


From OC to SDTM - automated ETL
Overview of Process – Framework (2)



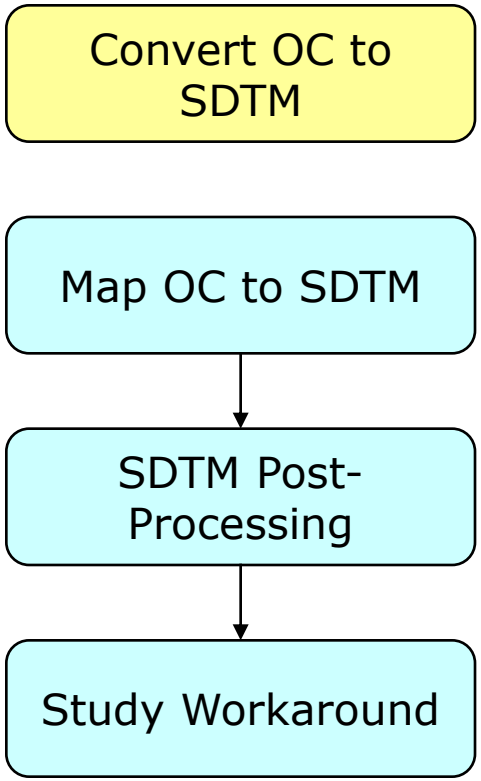
From OC to SDTM - automated ETL
Overview of Process – Framework (3)

- Sub-Framework 'Load OC':



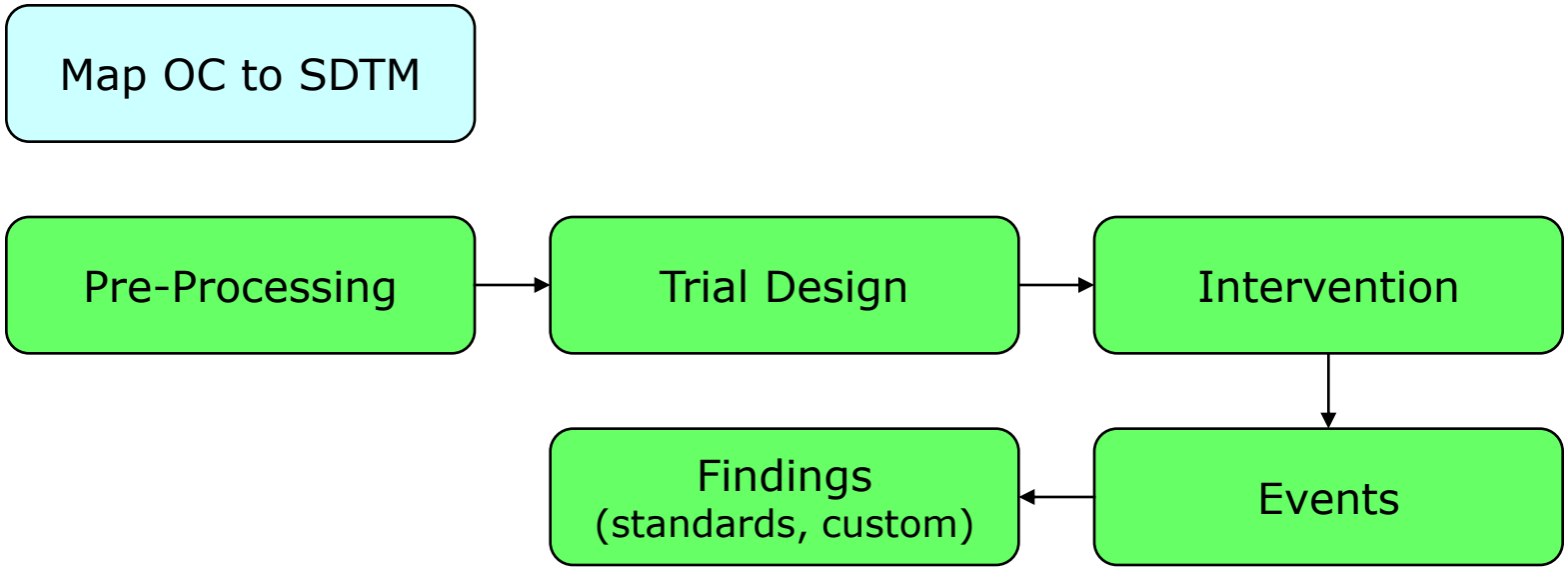
From OC to SDTM - automated ETL Overview of Process – Framework (4)

- Sub-Framework ‘Convert OC to SDTM’:



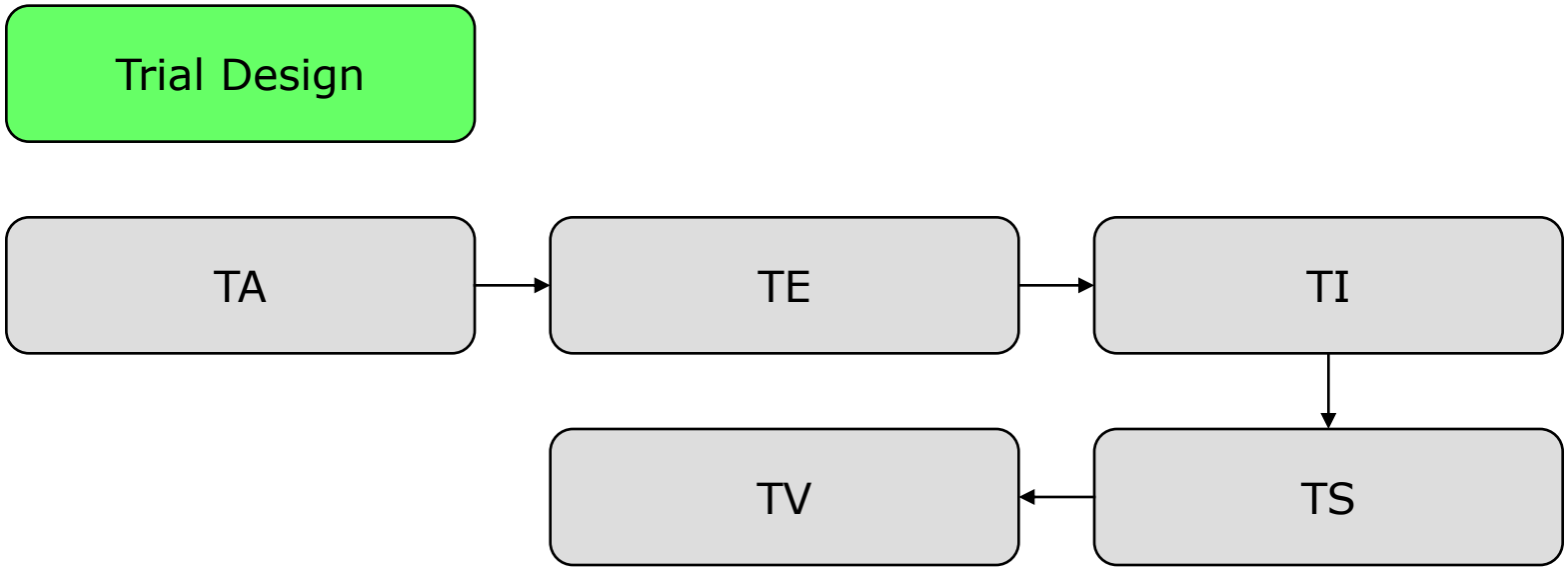
From OC to SDTM - automated ETL
Overview of Process – Framework (5)

- Sub-Framework 'Convert OC to SDTM':



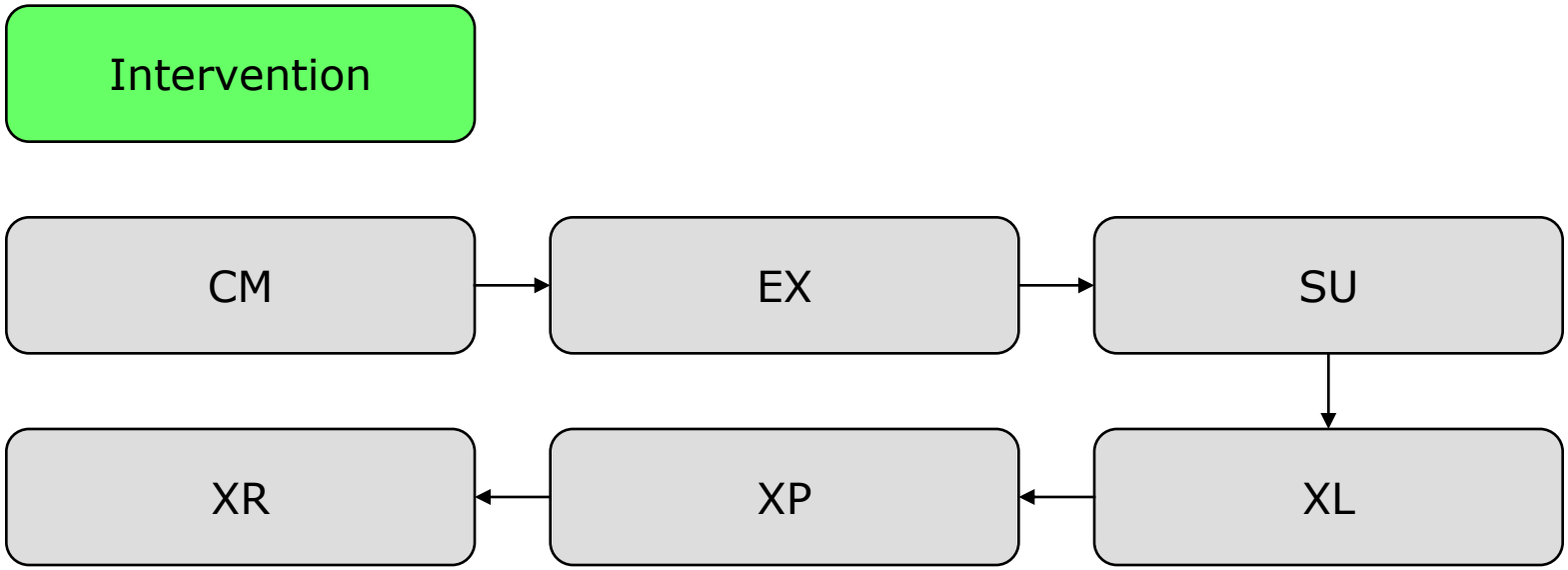
From OC to SDTM - automated ETL Overview of Process – Framework (6)

- Sub-Framework ‘Trial Design’:



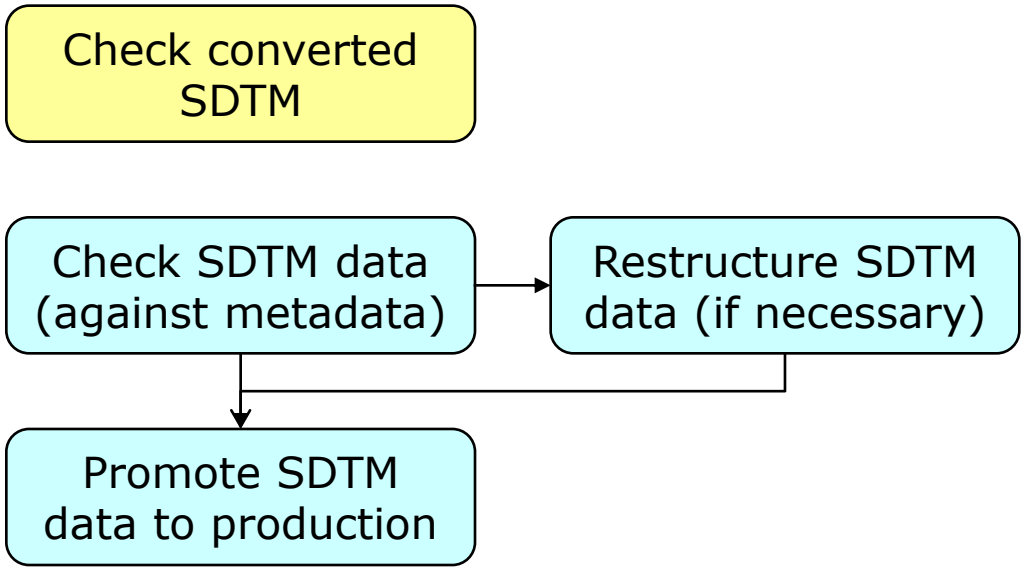
From OC to SDTM - automated ETL
Overview of Process – Framework (7)

- Sub-Framework 'Intervention':



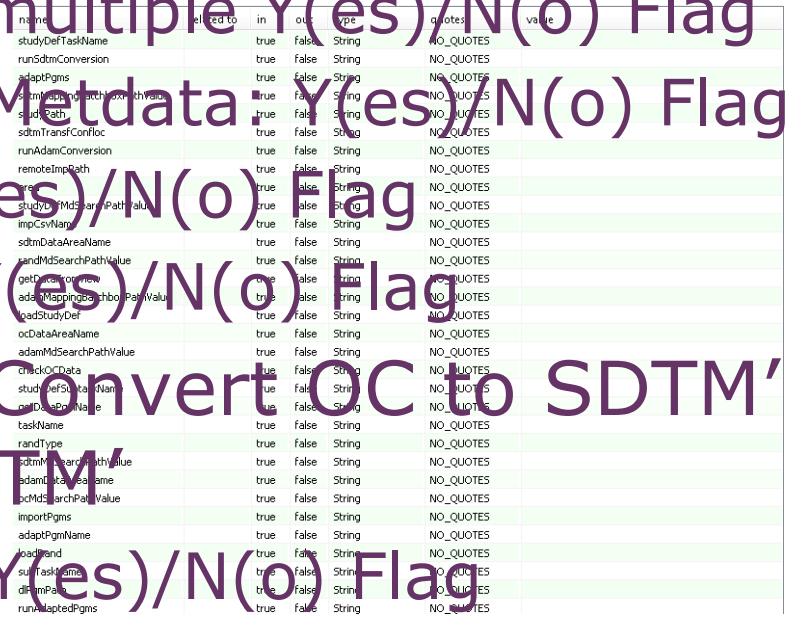
From OC to SDTM - automated ETL Overview of Process – Framework (8)

- Sub-Framework ‘Check converted SDTM’:



From OC to SDTM - automated ETL Parameterization Samples (1)

- General Parameters
  - Study Path (e.g. TA/DRUG/STUDYNO)
- Study-specific trigger 'Load OC data'
  - Refresh Data from OC: multiple Y(es)/N(o) Flag
  - Check OC Data against Metdata: Y(es)/N(o) Flag
  - Load Randomization: Y(es)/N(o) Flag
  - Load Study Definition: Y(es)/N(o) Flag
- Study-specific trigger 'Convert OC to SDTM' & 'Check converted SDTM'
  - Run SDTM Conversion: Y(es)/N(o) Flag



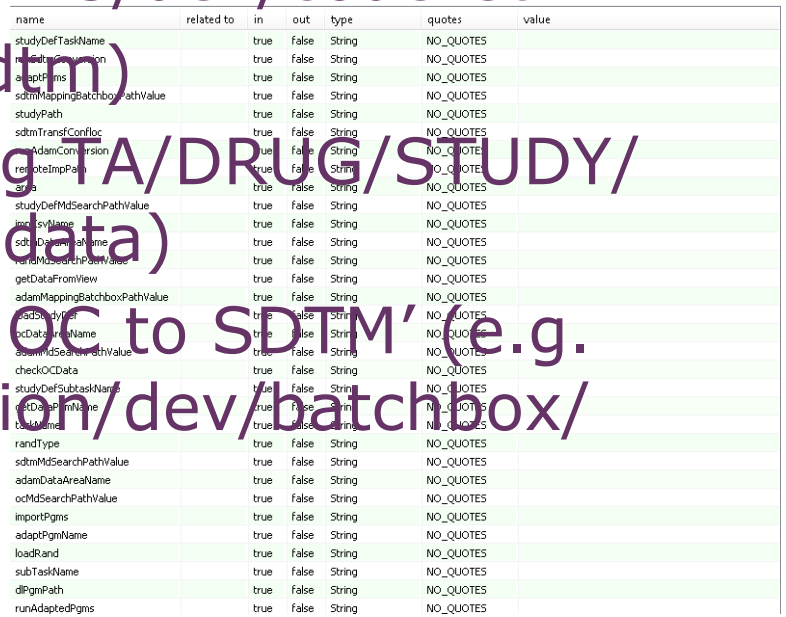
From OC to SDTM - automated ETL Parameterization Samples (2)

- Additional Parameters for 'Load OC data'
  - Remote Import Path
  - OC Data Area (e.g. collection)
  - OC Metadata Path (e.g standards/data_models/oc_collection/v1.0/dev/metadata)
  - Study Definition Metadata Path (e.g standards/data_models/oc_collection/v1.0/dev/metadata)
  - Randomization Metadata Path (e.g. standards/data_models/oc_collection/v1.0/dev/metadata)
  - Randomization Type (b or d)

name	related to	in	out	type	quotes	value
studyDefTaskName		true	false	String	NO_QUOTES	
runSdtmConversion		true	false	String	NO_QUOTES	
adaptPgmName		true	false	String	NO_QUOTES	
sdtmMappingBatchboxPathValue		true	false	String	NO_QUOTES	
studyPath		true	false	String	NO_QUOTES	
sdtmTransfConfloc		true	false	String	NO_QUOTES	
runAdamConversion		true	false	String	NO_QUOTES	
remoteImpPath		true	false	String	NO_QUOTES	
area		true	false	String	NO_QUOTES	
studyDefMdSearchPathValue		true	false	String	NO_QUOTES	
impCsvName		true	false	String	NO_QUOTES	
sdtmDataAreaName		true	false	String	NO_QUOTES	
randMdSearchPathValue		true	false	String	NO_QUOTES	
getDataFromView		true	false	String	NO_QUOTES	
adamMappingBatchboxPathValue		true	false	String	NO_QUOTES	
loadStudyDef		true	false	String	NO_QUOTES	
ocDataAreaName		true	false	String	NO_QUOTES	
adamMdSearchPathValue		true	false	String	NO_QUOTES	
checkOCData		true	false	String	NO_QUOTES	
studyDefSubtaskName		true	false	String	NO_QUOTES	
getDataPgmName		true	false	String	NO_QUOTES	
taskName		true	false	String	NO_QUOTES	
randType		true	false	String	NO_QUOTES	
sdtmMdSearchPathValue		true	false	String	NO_QUOTES	
adamDataAreaName		true	false	String	NO_QUOTES	
ocMdSearchPathValue		true	false	String	NO_QUOTES	
importPgms		true	false	String	NO_QUOTES	
adaptPgmName		true	false	String	NO_QUOTES	
loadRand		true	false	String	NO_QUOTES	
subTaskName		true	false	String	NO_QUOTES	
dlPgmPath		true	false	String	NO_QUOTES	
runAdaptedPgms		true	false	String	NO_QUOTES	

From OC to SDTM - automated ETL Parameterization Samples (3)

- Additional Parameters for 'Convert OC to SDTM'
  - Path to SDTM config location (e.g standards /data_models/config/v3.1.3/dev/codelist
  - SDTM Data Area (e.g. sdtm)
  - SDTM Metadata Path (e.g TA/DRUG/STUDY/ data/blinded/sdtm/metadata)
  - Path to framework 'Map OC to SDTM' (e.g. TA/DRUG/STUDY/collection/dev/batchbox/ bb_study_sdtm_call)



name	related to	in	out	type	quotes	value
studyDefTaskName		true	false	String	NO_QUOTES	
[illegible]		true	false	String	NO_QUOTES	
adaptPgms		true	false	String	NO_QUOTES	
sdtmMappingBatchboxPathValue		true	false	String	NO_QUOTES	
studyPath		true	false	String	NO_QUOTES	
sdtmTransfConfloc		true	false	String	NO_QUOTES	
[illegible]		true	false	String	NO_QUOTES	
remoteImpPath		true	false	String	NO_QUOTES	
area		true	false	String	NO_QUOTES	
studyDefMdSearchPathValue		true	false	String	NO_QUOTES	
[illegible]		true	false	String	NO_QUOTES	
sdtmDataAreaName		true	false	String	NO_QUOTES	
[illegible]		true	false	String	NO_QUOTES	
getDataFromView		true	false	String	NO_QUOTES	
adamMappingBatchboxPathValue		true	false	String	NO_QUOTES	
[illegible]		true	false	String	NO_QUOTES	
ocDataAreaName		true	false	String	NO_QUOTES	
[illegible]		true	false	String	NO_QUOTES	
checkOCData		true	false	String	NO_QUOTES	
studyDefSubtaskName		true	false	String	NO_QUOTES	
[illegible]		true	false	String	NO_QUOTES	
[illegible]		true	false	String	NO_QUOTES	
randType		true	false	String	NO_QUOTES	
sdtmMdSearchPathValue		true	false	String	NO_QUOTES	
adamDataAreaName		true	false	String	NO_QUOTES	
ocMdSearchPathValue		true	false	String	NO_QUOTES	
importPgms		true	false	String	NO_QUOTES	
adaptPgmName		true	false	String	NO_QUOTES	
loadRand		true	false	String	NO_QUOTES	
subTaskName		true	false	String	NO_QUOTES	
dlPgmPath		true	false	String	NO_QUOTES	
runAdaptedPgms		true	false	String	NO_QUOTES	

Discussion
Metrics (1) - Load OC

- OC datasets found
- Metadata check
  - successful | failed | irrelevant
- Structuring of datasets
  - successful | failed
- Promote to production (Checkin)
  - successful | failed | irrelevant
  - no metadata linked

Discussion
Metrics (2) – SDTM conversion

- Per Map
  - Existence of source datasets (check)
  - Existence of source variable in source datasets (check)
  - Structural correctness of input datasets (check)
  - Number of obervations in source datasets
  - Actual existing variables in source
  - Missing optional variables in source

Discussion
Developed best practices

- Data Model
  - Supports broad range of study flavors
- Framework
  - Enables study specific definition
- Superset / Subset of Metadata
  - Superset on standards level
  - Subset on study level
- Parameterization
  - Ensures flexibility

Discussion
Challenges

- Maintainance
  - No. of programs / framework
- Complexity
  - No. of parameter
- Handling of (study) exceptions

Conclusion

- Full Automation from OC to SDTM
  - achieved
- Flexibility
  - achieved
- Areas for improvement / optimization
  - ongoing