



It takes two to tango Pre-checker & Create Mapper

Author: Carla Santillan (Labcorp)

Co-author: Marc-Antoine Prodhomme (Janssen)

PHUSE Connect November 2022
Presentation SI02

Clinical Data Standards

Integrated Data Analytics & Reporting

DISCLAIMER

The presenter is working at Janssen on behalf of Labcorp as part of Data Standard team during DRM development.

The two tools described in this presentation have been created by Janssen.

Agenda

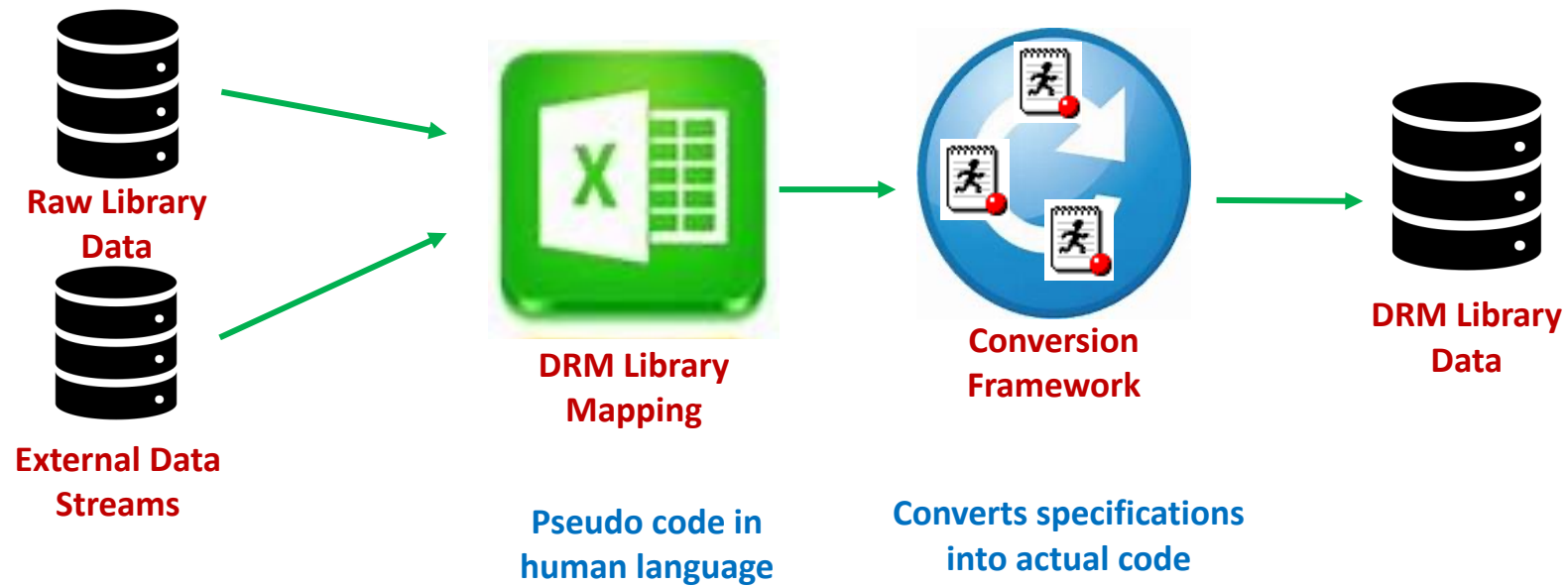
1 What is DRM, Pre-Checker and Create mapper?

2 Why were these tools created?

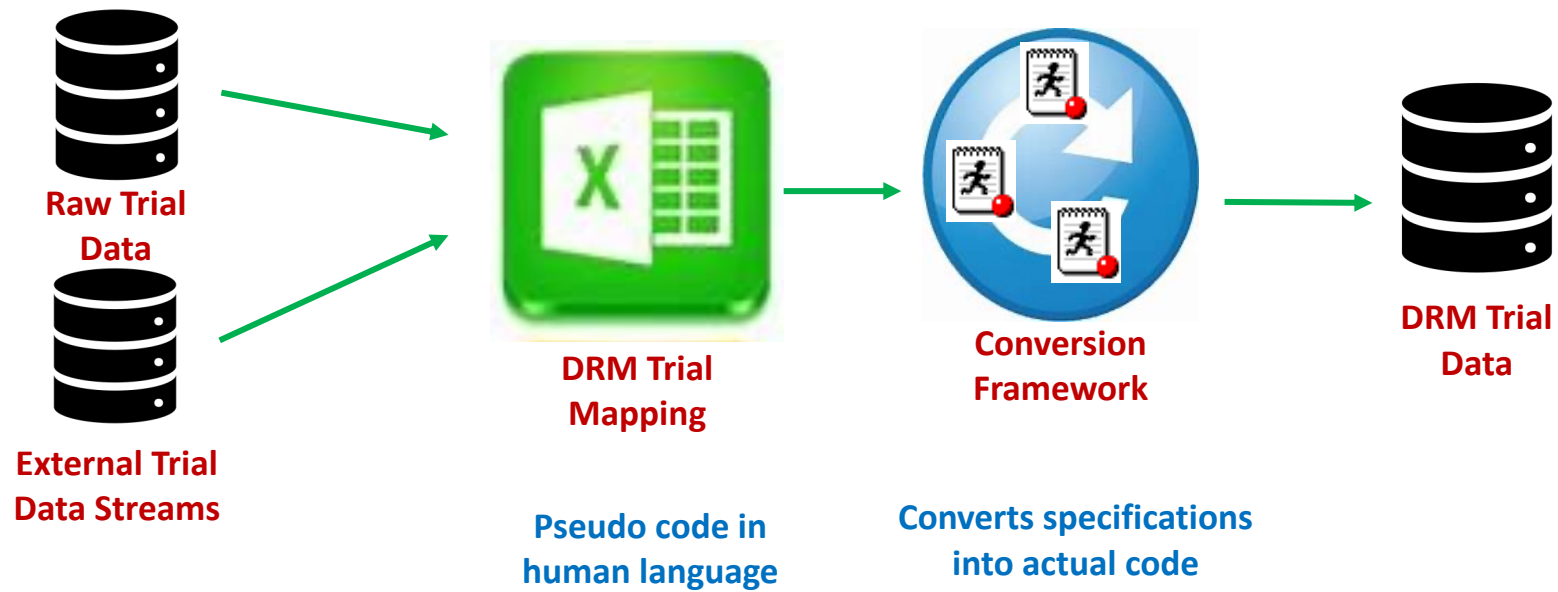
3 How do we use it?

What is Data Review Model (DRM)

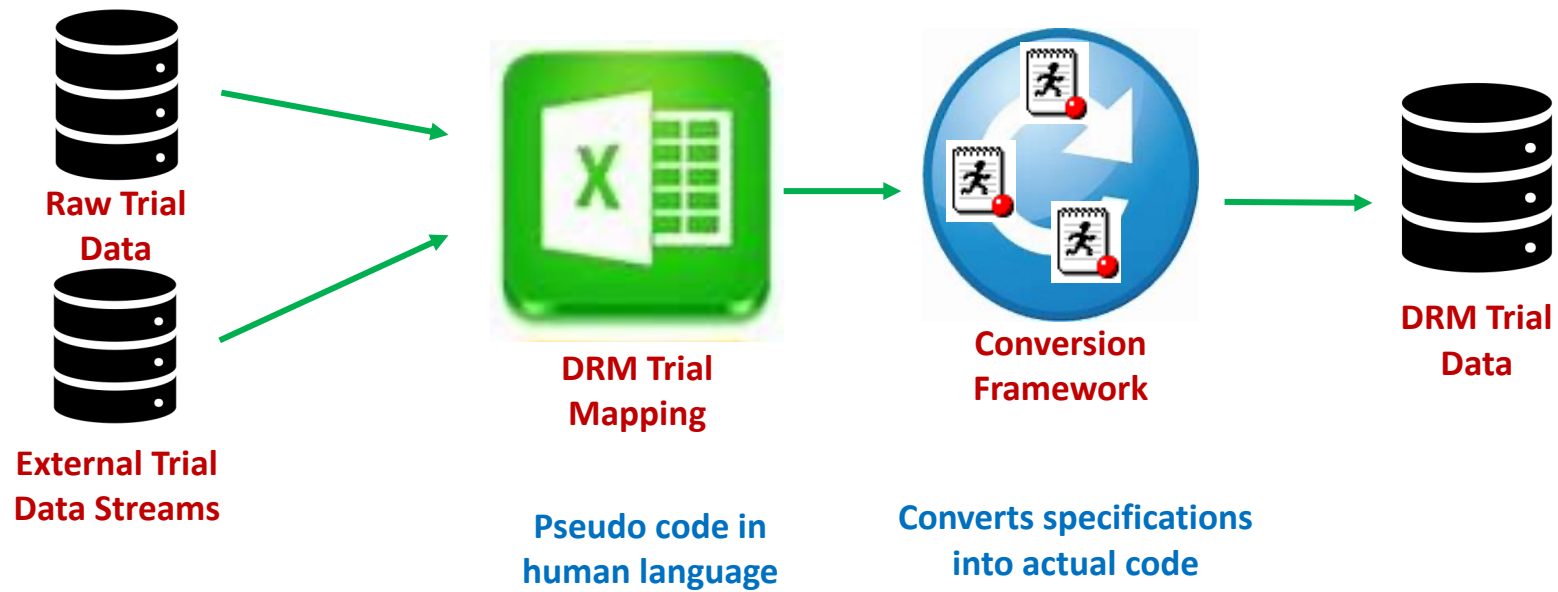
For more details on DRM see
Posters Presentations PP12



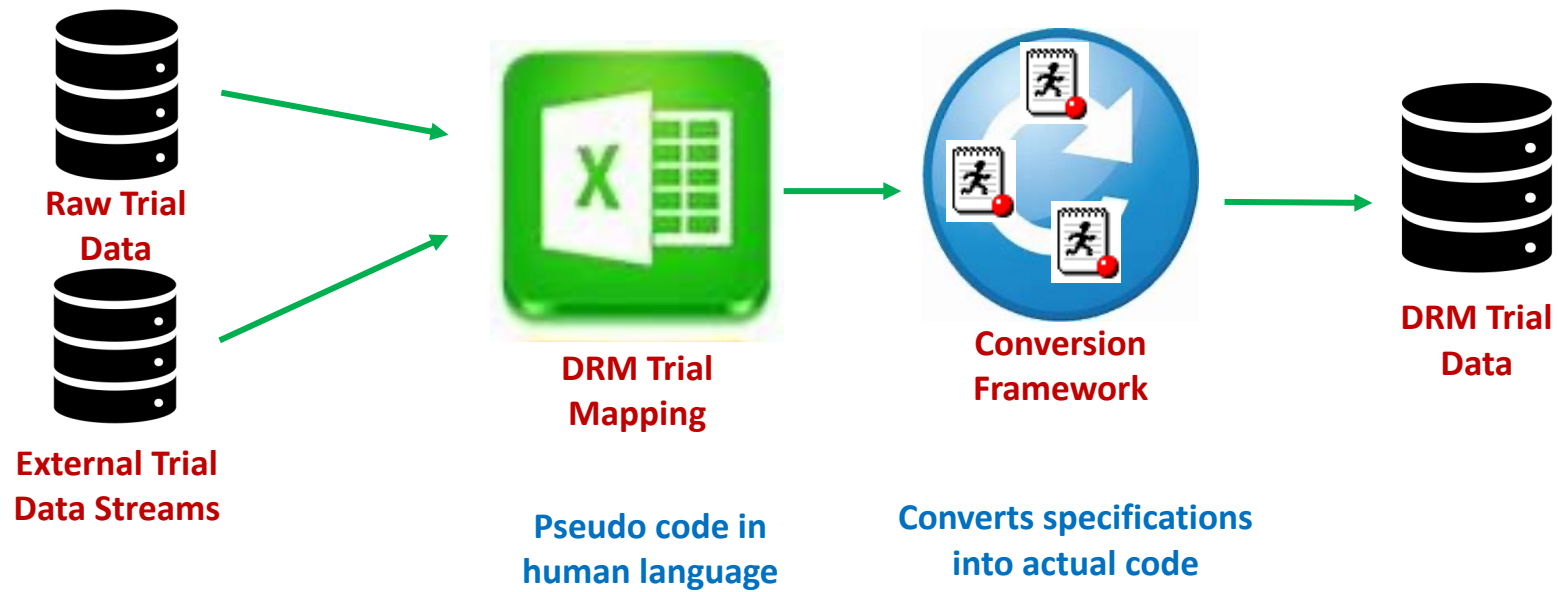
DRM at trial level



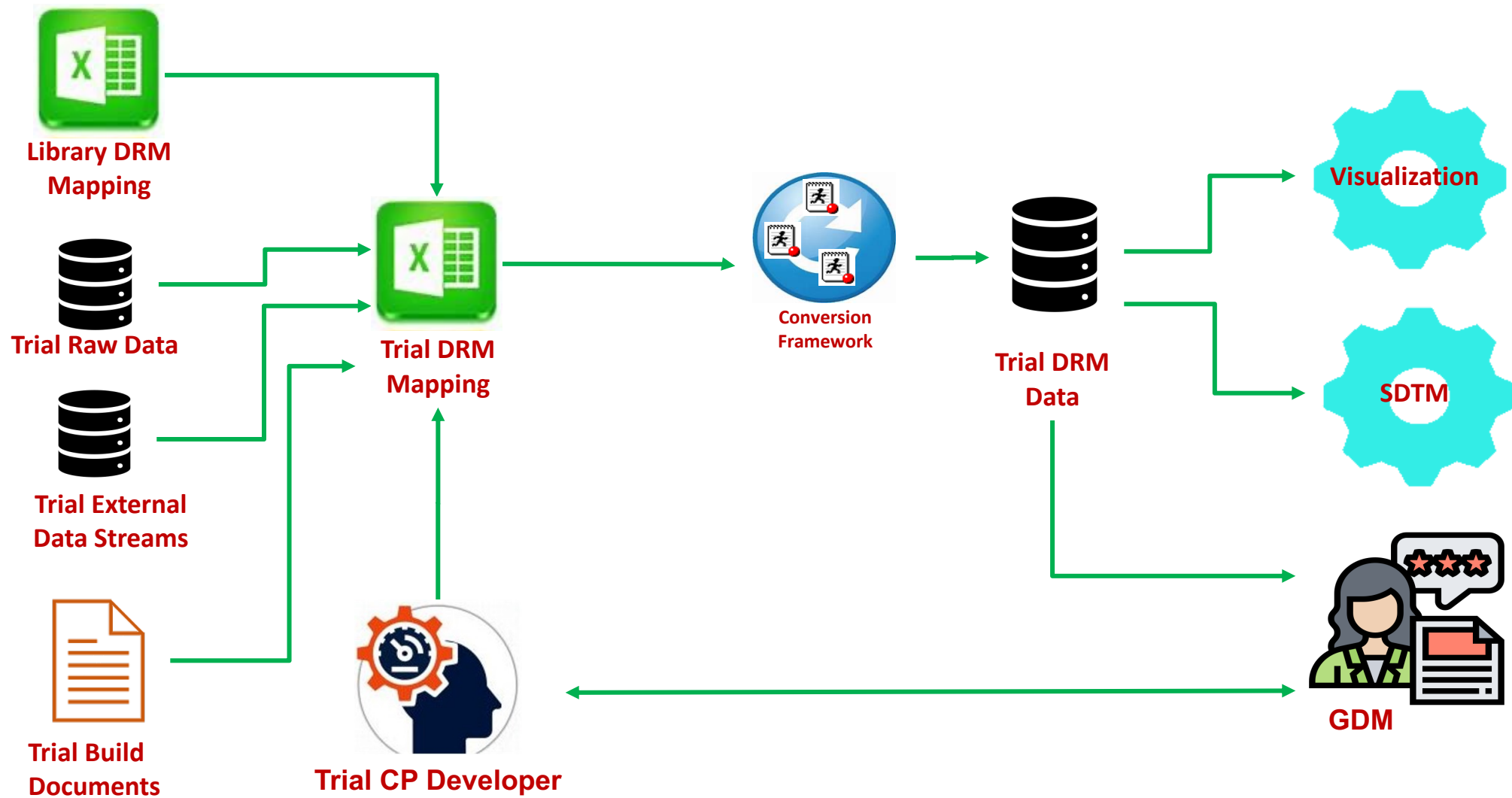
DRM at trial level



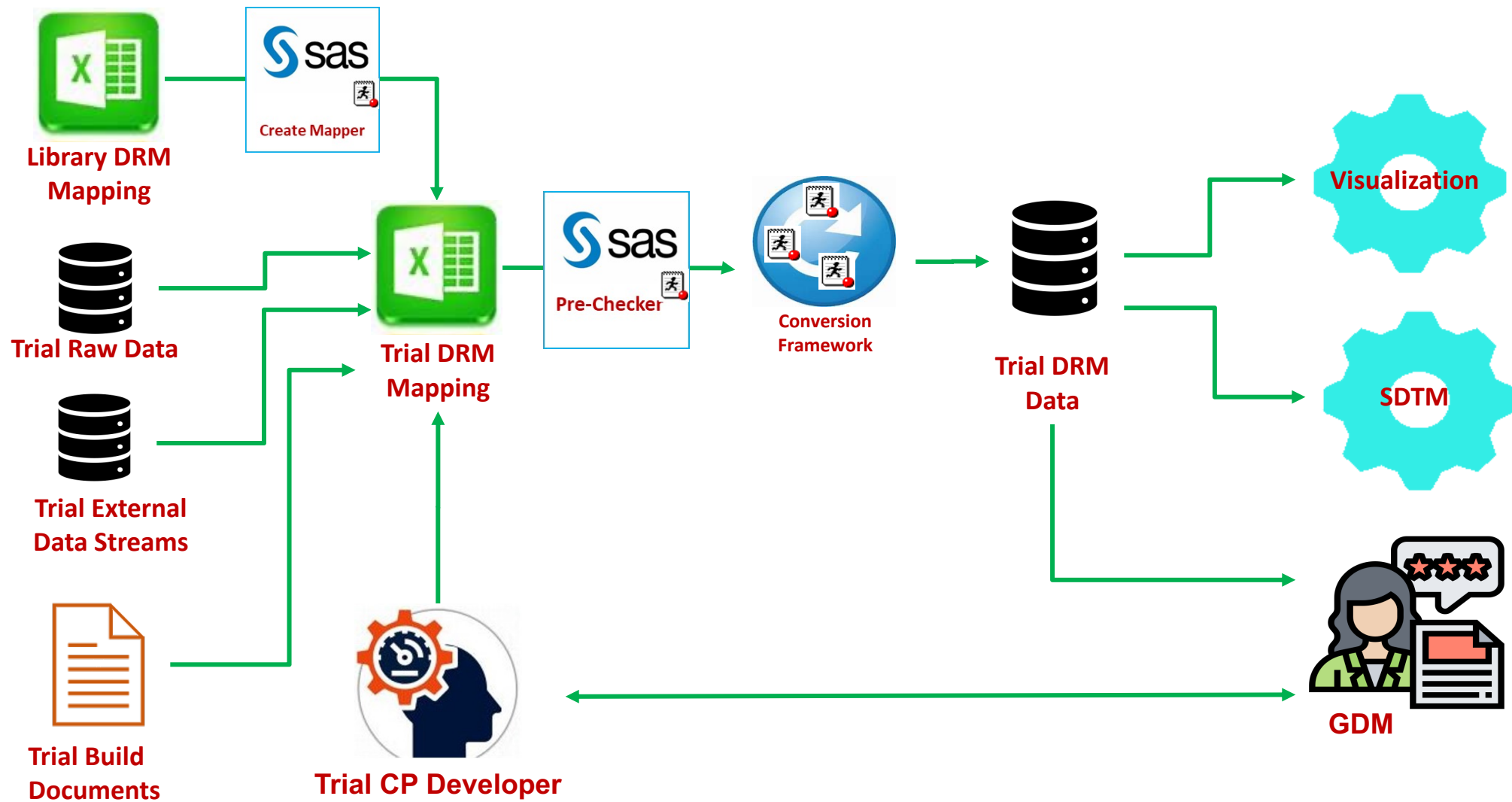
DRM at trial level



DRM – Create Mapper – Pre-checker



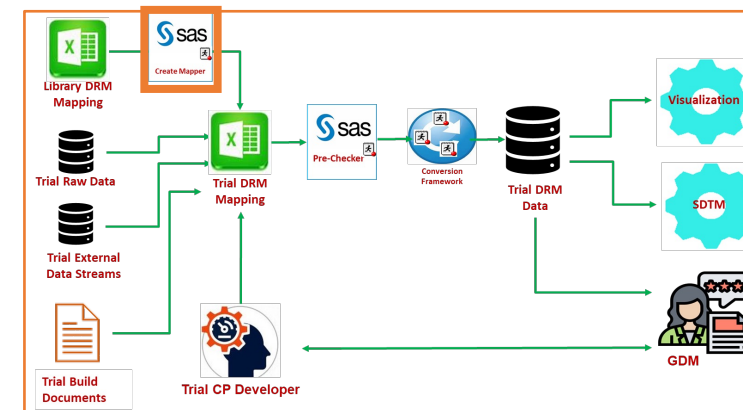
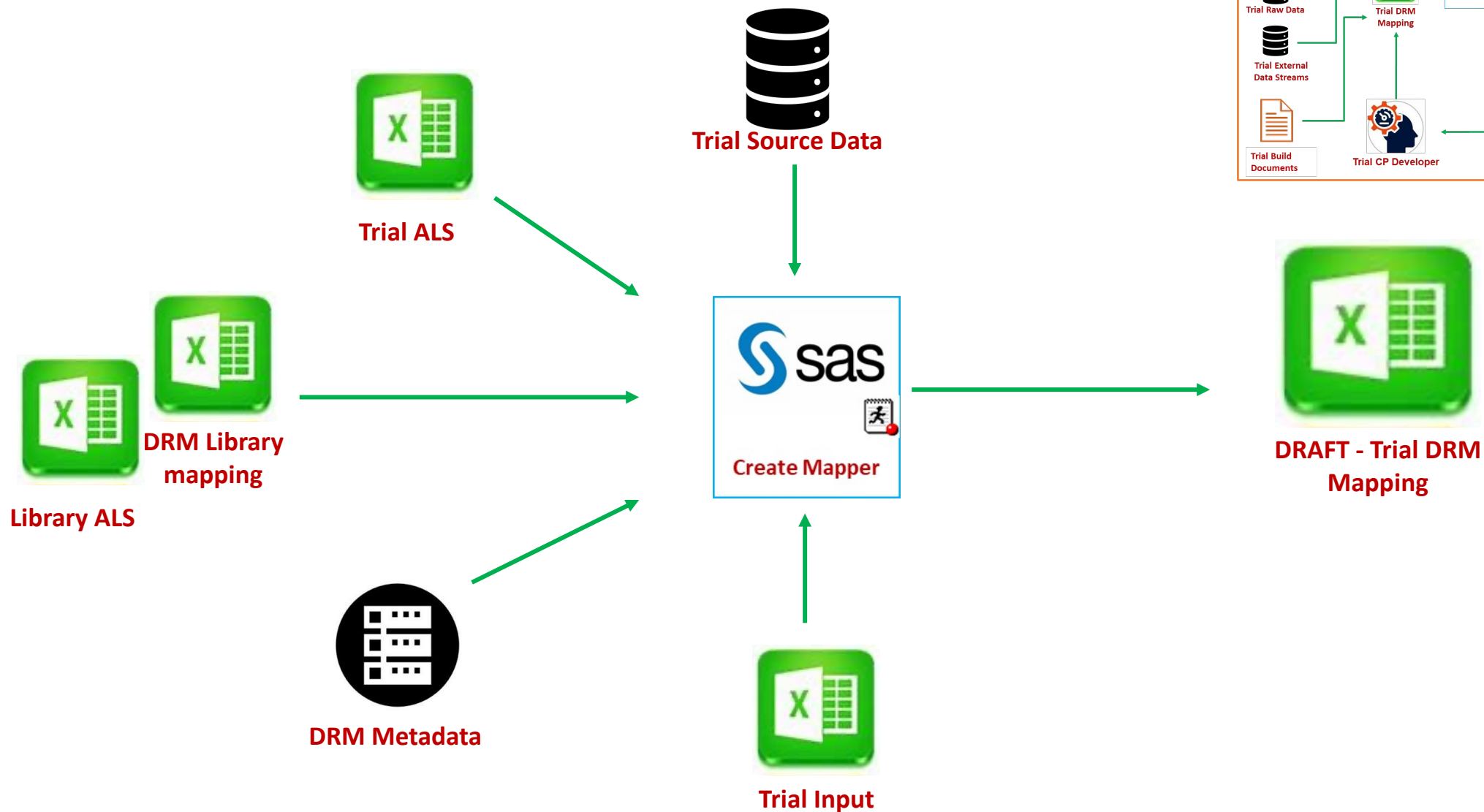
DRM – Create Mapper – Pre-checker



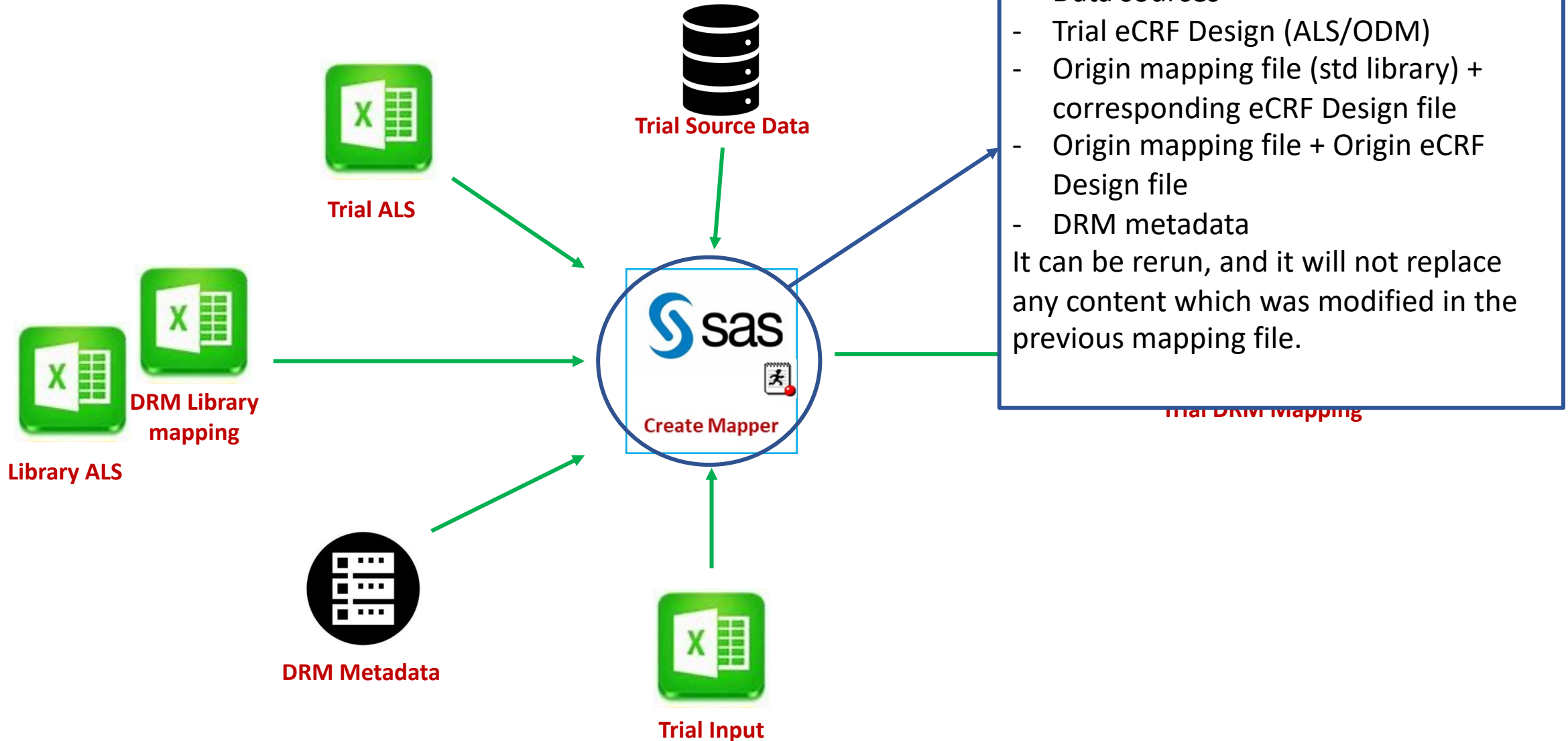
Why were these tools created?

- Reduce time producing DRM
- Avoid rework
- Avoid manipulation
- Metrics of trials following standards

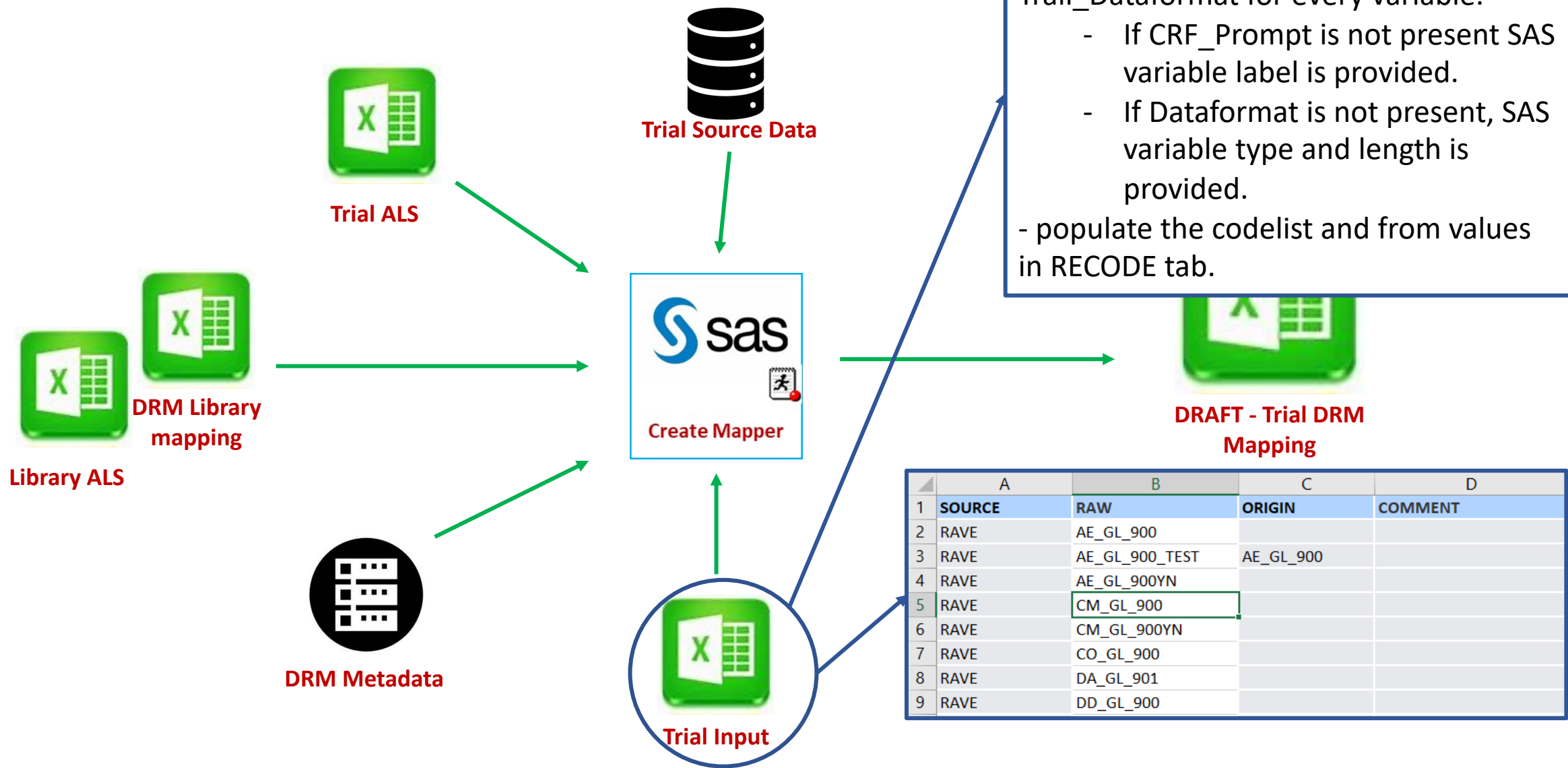
Create Mapper



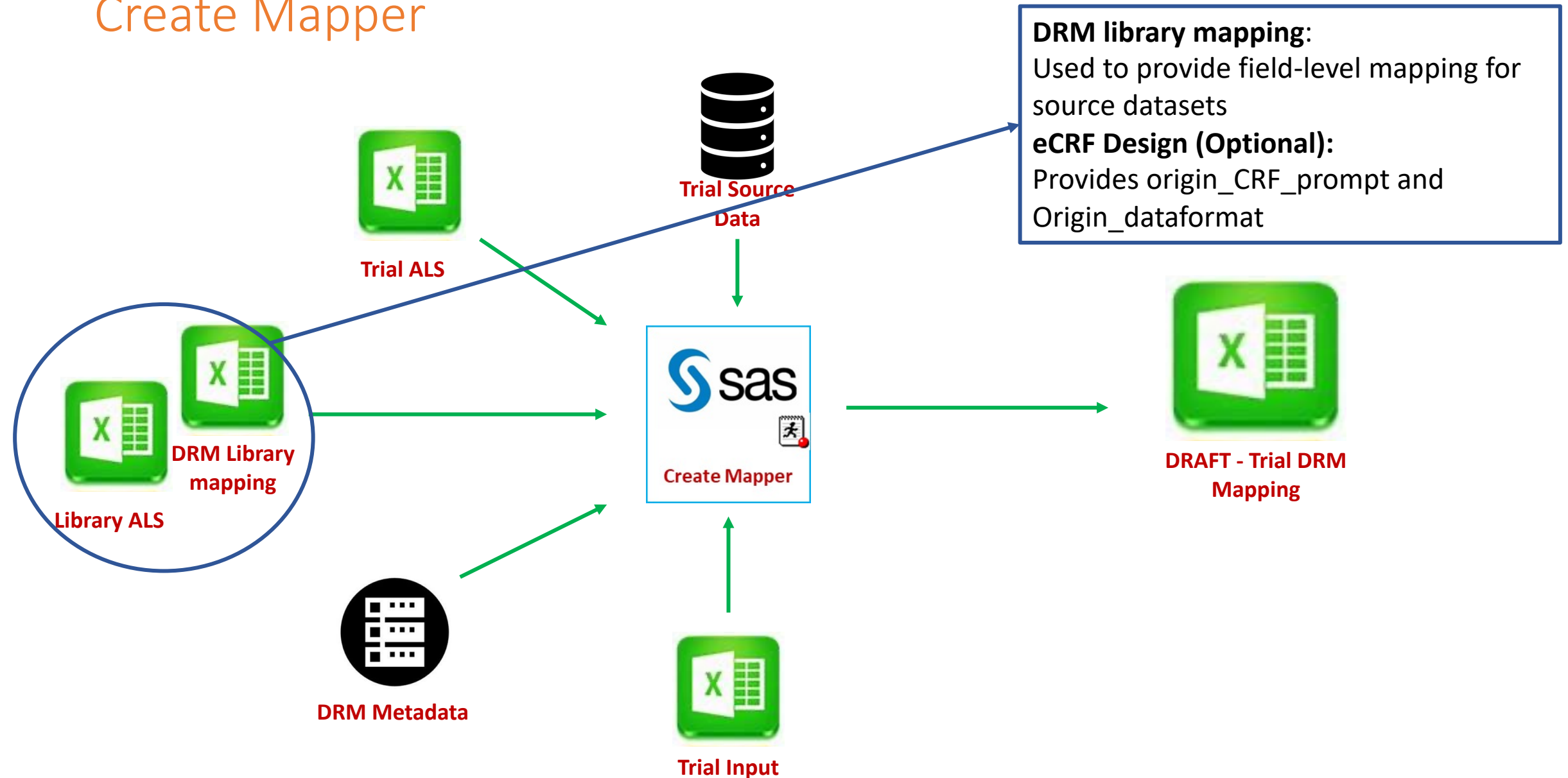
Create Mapper



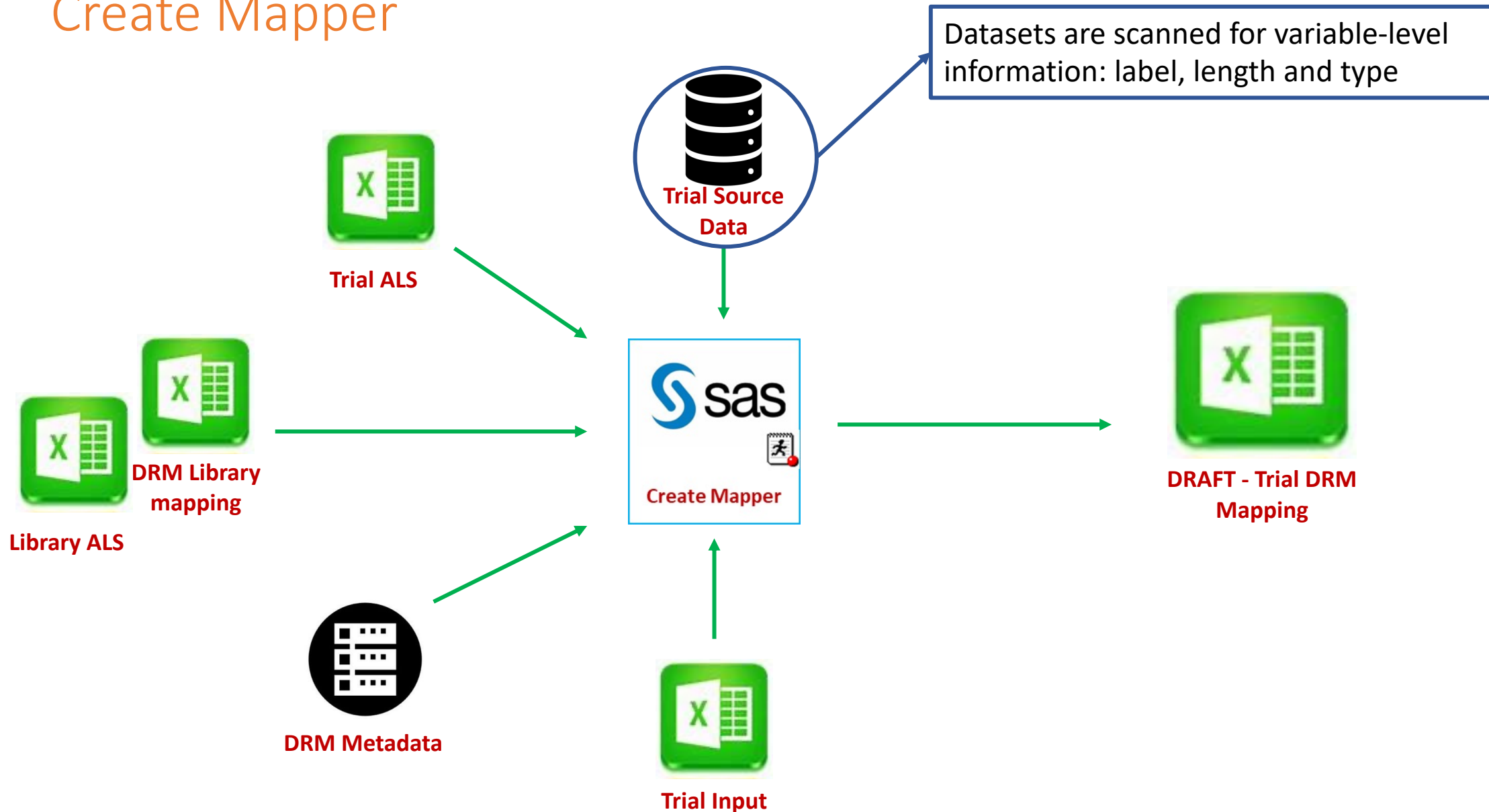
Create Mapper



Create Mapper



Create Mapper



Create mapper output

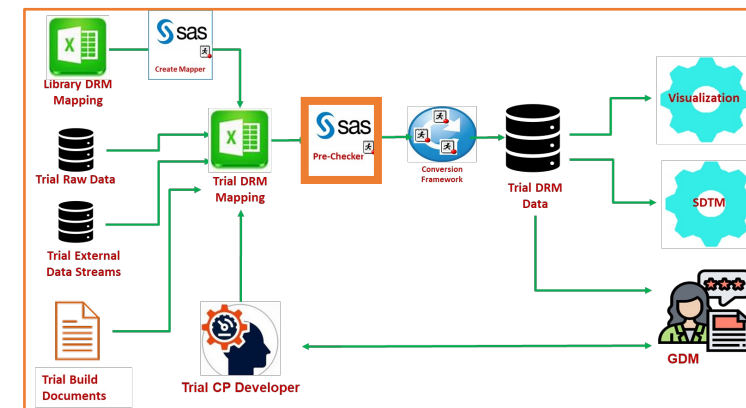
SOURCE_DS	SOURCE_VAR	TRIAL CRF PROMPT	TRIAL DATAFORMAT	ORIGIN CRF PROMPT	ORIGIN DATAFORMAT	TARGET_DS	TARGET VAR	SPECIFICATION	FUNCTION
RAVE_EX_GL_900B	RECORDID	[SAS] Internal id for the record	[SAS] 8 10.			DRM_EC	RRID	Copy from source variable	COPY
RAVE_EX_GL_900B	RECORDPOSITION	[SAS] Record number	[SAS] 8 10.			DRM_EC	RSEQ	Copy from source variable	COPY
RAVE_EX_GL_900B	MAXUPDATED	[SAS] Latest data update time	[SAS] 8 DATETIME22.3			DRM_EC	MAXUPDAT	Copy from source variable	COPY
RAVE_EX_GL_900B	AEDSL2	[ALS] If Adverse Event, choose the corresponding AE log line, start date, and term	[ALS] \$80	[ALS] If Adverse Event, choose the corresponding AE log line, start date, and term	[ALS] \$80	DRM_EC	AEDSL2_DRM	Copy from source variable	COPY
RAVE_EX_GL_900B	EXRSNDO	[ALS] If Other, Specify	[ALS] \$200	[ALS] If Other, Specify	[ALS] \$200	DRM_EC	ECREASOO	Copy from source variable	COPY
RAVE_EX_GL_900B	EXAMONT	[ALS] Total volume infused	[ALS] 4.1	[ALS] Dose	[ALS] 8.3	DRM_EC	ECDOSE	Copy from source variable	COPY
RAVE_EX_GL_900B	DATAPAGENAME	[SAS] eCRF page name	[SAS] \$765 \$765.			DRM_EC	RDATASET	Set to 'EX_GL_900B'	ASSIGN[EX_GL_900B]
RAVE_EX_GL_900B	EXYN_DOS	[ALS] Was IV study agent administered for the subject?	[ALS] \$2	[ALS] Was the dose administered?	[ALS] \$2	DRM_EC	ECPRESP	Assign Y	ASSIGN[Y]

- The following tabs are present in the file:
 - » Tabs containing all mapping information: MAPPING, RECODE, ORDER, TABLES, COLUMNS.
 - » Tabs with unused metadata, if any: UNUSED_TABLES, UNUSED_COLUMNS, UNUSED_COLUMNS_EXISTING_TABLES.
 - » If there is a previous DRM mapping file, the following tabs may be present if differences were found between the two versions: DIFFER_MAPPING, DIFFER_RECODE, DIFFER_COLUMNS.

Create mapper output - Example of the DIFFER_MAPPING tab

ORIGIN		SOURCE_DS	SOURCE_VAR	TARGET_DS	TARGET_VAR	SPECIFICATION	SPECIFICATION	FUNCTION	IS_SDTM_SUPP	ORIGIN	EVAL
TYPE	OBS	SOURCE_DS	SOURCE_VAR	TARGET_DS	TARGET_VAR	SPECIFICATION	FUNCTION	IS_SDTM_SUPP	ORIGIN	EVAL	
TRIAL	68	RAVE.DM_GL_900	RACEA	DRM_DM	RACE	Assigning the value of RACE in case multiple RACES are selected	FUNCTION[length MULTI 8.; MULTI=sum (_RACEA, _RACEAIAN, _RACEBA, _RACENHOP, _RACEW, _RACENR); IF MULTI=1 then do; RACE= COALESCEC(RACEA, RACEAIAN, RACEBA, RACENHOP, RACEW, RACENR_CHAR_DRM); RACEA=""; RACEAIAN=""; RACEBA=""; RACENHOP=""; RACEW=""; end; else if MULTI>1 then do; RACE='MULTIPLE'; end; else if MULTI=0 then do; RACE=""; end; ;]				
ORIGIN	68	RAVE.DM_GL_900	RACEA	DRM_DM	RACE	Assigning the value of RACE in case multiple RACES are selected	FUNCTION[length MULTI 8.; MULTI=sum (_RACEA, _RACEAIAN, _RACEBA, _RACENHOP, _RACEW, _RACENR, _RACEUNK); IF MULTI=1 then do; RACE= COALESCEC(RACEA, RACEAIAN, RACEBA, RACENHOP, RACEW, RACENR_CHAR_DRM, RACEUNK_CHAR_DRM); RACEA=""; RACEAIAN=""; RACEBA=""; RACENHOP=""; RACEW=""; end; else if MULTI>1 then do; RACE='MULTIPLE'; end; else if MULTI=0 then do; RACE=""; end; ;]				
TRIAL	74	RAVE.DM_GL_902	INSTANCENAME	DRM_DM	RVISIT	Copy from source variable	COPY				
TRIAL	75	RAVE.DM_GL_902	MAXUPDATED	DRM_DM	MAXUPDAT	Copy from source variable	COPY				

Pre-checker - trial flow



DRAFT - Trial DRM Mapping



Trial DRM Mapping



Conversion Framework

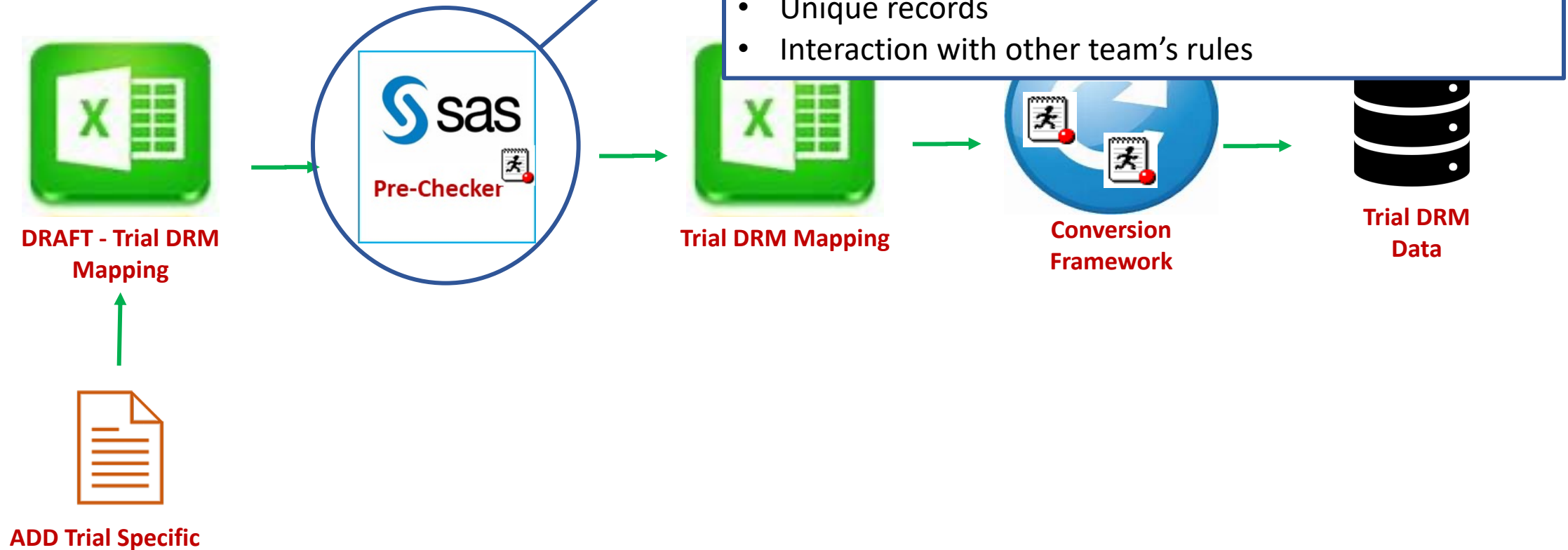


Trial DRM Data



ADD Trial Specific

Pre-checker - trial flow



Pre-checker output – Report

Tabs present in the file:

- CHECK_RESULTS
- CHECK_SUMMARY
- Original Mapping specifications (MAPPING, RECODE, ORDER, TABLES, COLUMNS)

Pre-checker output

CHECKID	CHECKNAME	CHECKTYPE	PARAMETER	DESCRIPTION	MESSAGE	RUN CHECK	EXECUTED CHECK	NUMBER_OF_OBSERVATIONS
CHECK165	check_columns_parameter	COLUMNS	length	Variable level information for each domain must be complete in metadata library - i.e. name/length/type/label must have value in columns sheet.	Columns domain {#COLUMNS_DOMAIN#} variable {#COLUMNS_NAME#} has no defined {#PARAMETER#}	Y	Y	136
CHECK166	check_columns_parameter	COLUMNS	type	Variable level information for each domain must be complete in metadata library - i.e. name/length/type/label must have value in columns sheet.	Columns domain {#COLUMNS_DOMAIN#} variable {#COLUMNS_NAME#} has no defined {#PARAMETER#}	Y	Y	136
CHECK167	check_mapping_poststep	MAPPING		For each domain TARGET_DS must have a poststep, only if the corresponding SOURCE_DS consists of more than one source dataset.	{#MAPPINGS_TARGET_DS#} does not have any POSTSTEP function	Y	Y	0
CHECK168	check_unique_poststep	MAPPING		Each POSTSTEP number must be unique within a target domain.	{#MAPPINGS_TARGET_DS#} POSTSTEP function does not use an unique number	Y	Y	0
CHECK169	check_source_var	SOURCE		The checks ensures that every source variable has been included in the Mapping tab	Source variable {#SOURCE_SOURCE_VAR#} in source dataset {#SOURCE_SOURCE_DS#} was not included in mapping metadata	Y	N	0
CHECK170	check_metadata_var	MAPPING		Variable that is defined in the mapping sheet must be present in the source data set, regardless whether it is 'NOT MAPPED' or not.	Variable {#MAPPINGS_SOURCE_VAR#} in source dataset {#MAPPINGS_SOURCE_DS#} does not exist in the source data	Y	N	0
CHECK171	check_unique_function	MAPPING		Functions in poststeps must be unique within a target domain.	{#MAPPINGS_TARGET_DS#} FUNCTION does not contain an unique function	Y	Y	0
CHECK172	check_linebreak_specification	MAPPING		The text in SPECIFICATION cannot end with a line break.	MAPPING: TARGET_DS {#MAPPINGS_TARGET_DS#} - TARGET_VAR {#MAPPINGS_TARGET_VAR#}: SPECIFICATION has an extra line break at the end that needs to be removed.	Y	Y	0
CHECK173	check_linebreak_function	MAPPING		The text in FUNCTION cannot end with a line break.	MAPPING: TARGET_DS {#MAPPINGS_TARGET_DS#} - TARGET_VAR {#MAPPINGS_TARGET_VAR#}: FUNCTION has an extra line break at the end that needs to be removed.	Y	Y	0
CHECK174	check_unique_order	ORDER		Order tab records must be unique (key: CREATION_ORDER_OF_DOMAINS).	ORDER record for {#ORDER_CREATION_ORDER_OF_DOMAINS#} is not unique.	Y	Y	0
CHECK175	check_unique_tables	TABLES		Tables tab records must be unique (key: DOMAIN).	TABLES record for {#TABLES_DOMAIN#} is not unique.	Y	Y	0
CHECK176	check_unique_columns	COLUMNS		Columns tab records must be unique (key: DOMAIN, NAME).	COLUMNS record for {#COLUMNS_DOMAIN#}-{#COLUMNS_NAME#} is not unique.	Y	Y	0
CHECK177	check_urecid_columns	COLUMNS		URECID must be present in the COLUMNS tab for each DOMAIN	COLUMNS: DOMAIN {#COLUMNS_DOMAIN#} does not contain a record for URECID.	Y	Y	0
CHECK178	check_urecid_mapping	MAPPING		There must be one record per source dataset and target dataset for source variable URECID, where the variable is copied or kept.	MAPPING: SOURCE_DS {#MAPPINGS_SOURCE_DS#} - TARGET_DS {#MAPPINGS_TARGET_DS#}: No record found for URECID where the variable is either copied or kept.	Y	Y	23
CHECK179	check_mergeds_target_var	MAPPING		If a target dataset has more than 1 source, then there must be a step that uses the macro _001_merge_ds_ or the macro _001_concat_ds_.	MAPPING: TARGET_DS {#MAPPINGS_TARGET_DS#} has more than 1 source, but its mapping does not make use of the macro _001_merge_ds_ or the macro _001_concat_ds_.	Y	Y	8

Pre-checker output

CHECKID	CHECKNAME	CHECKTYPE	DESCRIPTION	PARAMETER	MESSAGE	EXCEL_ROW NUMBER	DOCUMENTATION
CHECK057	check_variables_completed_map	MAPPING	For each tab records (mapping, recode, order, tables, columns, librarysearchorder) the variable defined in PARAMETER must not have missing values.	SPECIFICATION	MAPPING tab: Column SPECIFICATION contains unexpected missing values - Number of occurrences: 8	802	
CHECK058	check_variables_completed_map	MAPPING	For each tab records (mapping, recode, order, tables, columns, librarysearchorder) the variable defined in PARAMETER must not have missing values.	FUNCTION	MAPPING tab: Column FUNCTION contains unexpected missing values - Number of occurrences: 9	182	
CHECK129	check_recode_to_type	RECODE	The data type of each value in the TO column must conform to the TYPE column in recode tab (e.g. If TYPE = "N2N", TO must be numeric).		Codelist {VISITNUM}: TO value { } does not match the expected data type: {C2N} - Number of occurrences: 2	239	doc
CHECK138	check_function	MAPPING	Mapping functions have to be part of the expected Conversion Framework functions (COPY, ASSIGN, RECODE, BLANK, NOT MAPPED, FUNCTION, ISODATE,...)		MAPPING: TARGET_DS {MEDS}- TARGET_VAR {MDSTDTC}: FUNCTION { } is not part of the expected Conversion Framework functions - Number of occurrences: 1	192	doc
CHECK138	check_function	MAPPING	Mapping functions have to be part of the expected Conversion Framework functions (COPY, ASSIGN, RECODE, BLANK, NOT MAPPED, FUNCTION, ISODATE,...)		MAPPING: TARGET_DS {MEDS}- TARGET_VAR {MDENDTC}: FUNCTION { } is not part of the expected Conversion Framework functions - Number of occurrences: 1	197	doc
CHECK146	check_mapping_function	MAPPING	IF SOURCE_DS/TARGET_DS/TARGET_VAR are populated then FUNCTION must have a value.		Mapping: FUNCTION was not populated when {CLINSPRK.CM_GL_900} {MEDS} and {MDSTDTC} were defined - Number of occurrences: 1	192	doc
CHECK146	check_mapping_function	MAPPING	IF SOURCE_DS/TARGET_DS/TARGET_VAR are populated then FUNCTION must have a value.		Mapping: FUNCTION was not populated when {CLINSPRK.CM_GL_900} {MEDS} and {MDENDTC} were defined - Number of occurrences: 1	197	doc
CHECK147	check_not_mapped_target_ds	MAPPING	If FUNCTION is equal to "NOT MAPPED", then TARGET_DS must be empty.		{CLINSPRK.MH_GL_900_CS_1}-{MHCAT}: FUNCTION is 'NOT MAPPED' but TARGET_DS {MEDHIST} is not empty - Number of occurrences: 1	441	doc
CHECK165	check_columns_parameter	COLUMNS	Variable level information for each domain must be complete in metadata library - i.e. name/length/type/label must have value in columns sheet.	length	Columns domain {AE} variable {PERIOD} has no defined {length} - Number of occurrences: 25	2	doc
CHECK165	check_columns_parameter	COLUMNS	Variable level information for each domain must be complete in metadata library - i.e. name/length/type/label must have value in columns sheet.	length	Columns domain {AE} variable {PHASE} has no defined {length} - Number of occurrences: 25	3	doc
CHECK165	check_columns_parameter	COLUMNS	Variable level information for each domain must be complete in metadata library - i.e. name/length/type/label must have value in columns sheet.	length	Columns domain {AE} variable {VISIT} has no defined {length} - Number of occurrences: 25	4	doc

Pre-checker output

Prechecks documentation

check_recode_to_type

Check ID: **CHECK129**

Checkname: check_recode_to_type

Parameter:

Checktype: Recode

Correct example:

Recode tab:

CODELIST	TYPE	FROM	TO	COMMENTS	AREA_DESC	AREA_VERSION	RULE FROM	RULE TO	RECID	RECODE ISSUE
AESEV	N2C	.			RAVE.GLOBAL	20-DEC-2019	19-MAR-2020			
AESEV	N2C	1	MILD		RAVE.GLOBAL	20-DEC-2019	19-MAR-2020			
AESEV	N2C	2	MODERATE		RAVE.GLOBAL	20-DEC-2019	19-MAR-2020			
AESEV	N2C	3	SEVERE		RAVE.GLOBAL	20-DEC-2019	19-MAR-2020			
REL	C2C				RAVE.GLOBAL	20-DEC-2019	19-MAR-2020			
REL	C2C	Not Related	NOT RELATED		RAVE.GLOBAL	20-DEC-2019	19-MAR-2020			
REL	C2C	Related	RELATED		RAVE.GLOBAL	20-DEC-2019	19-MAR-2020			
VISITNUM	C2N	Day 1	1		RAVE.GLOBAL	20-DEC-2019	19-MAR-2020			
VISITNUM	C2N	Day 8	8		RAVE.GLOBAL	20-DEC-2019	19-MAR-2020			
VISITNUM	C2N	Day 15	15		RAVE.GLOBAL	20-DEC-2019	19-MAR-2020			
VISITNUM	C2N	Day 22	22		RAVE.GLOBAL	20-DEC-2019	19-MAR-2020			

Explanation:

The example above shows that column TO associated with codelist is equal to VISITNUM has character value "BASELIN
TYPE has value "C2N", values in column TO must be numeric.

The corrected example below shows that in column TO associated with codelist is equal to VISITNUM character value "BASELINE has been replaced by a numeric value "1" (indicated in green).

Conclusion

- These are two important basic tools on the creation of DRM, they reduce the time and increase the efficiency
- There is still a lot of manual interaction
- We will keep working to increase CP and DM's time to focus on quality of the data and standards compliance



With the collaboration of

