



PhUSE US Connect 2019

Paper SI01

End to End SDTM Automation: A Metadata Centric Approach

Roman Radelicki, SGS, Mechelen, Belgium

Swapna Pothula, SGS, Mechelen, Belgium

INTRODUCTION

A metadata driven approach and the seamless integration of processes and people are key to creating clinical databases that are compliant and consistent with CDISC SDTM standards. The implementation of a metadata repository (MDR) is the first step toward a metadata driven approach; we at SGS went a few steps ahead of this and built in-house tools that are linked to the MDR to achieve end to end automation of the SDTM workflow.

METADATA CENTRIC APPROACH

At SGS, we have chosen a metadata centric approach to automating the SDTM workflow, with the end goal of creating databases that are compliant and consistent with SDTM, while at the same time enabling time and cost optimization.

METADATA REPOSITORY

The process began with the creation of an MDR – metadata repository. This global library of metadata is the master source of metadata to be used for study setup.

The structure of the libraries within the MDR is based on the Define-XML model, which is implemented by using extensions to the CDISC Operational Data Model (ODM) XML schema. The MDR libraries contain all the available domains, variables, valuelists, codelists, comments and computational methods that are applicable for a specific implementation of SDTM, SEND or ADaM.

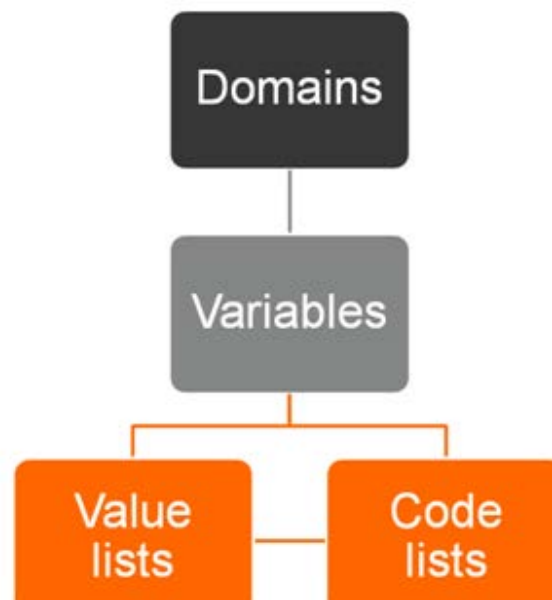


Figure 1 A metadata repository library. Comments and computational methods are not represented in this figure.

Figure 1 illustrates the standard structure of an MDR library. At the top level are all available domains for a specific CDISC implementation e.g. SDTM Implementation Guide v3.2. Further down are all available variables per domain. At the bottom are the value level metadata and the codelists.

Our metadata repository contains multiple libraries. Our SGS implementation guide, SGS IG v5.0, is based on version 3.2 of the SDTM implementation guide. This version of our guide is constantly maintained and kept up to date with new releases of controlled terminology or updates from regulators. On the other hand, the SGS metadata repository is also developed to support metadata libraries from other parties or clients. The most efficient process is to obtain at the start of the collaboration and then periodically the client's files containing their implementation of a specific CDISC standard. Afterwards their client specific CDISC standard is implemented in our metadata repository. Maintenance of the standard occurs at the client side. We upload every new version of the standard.

MDR

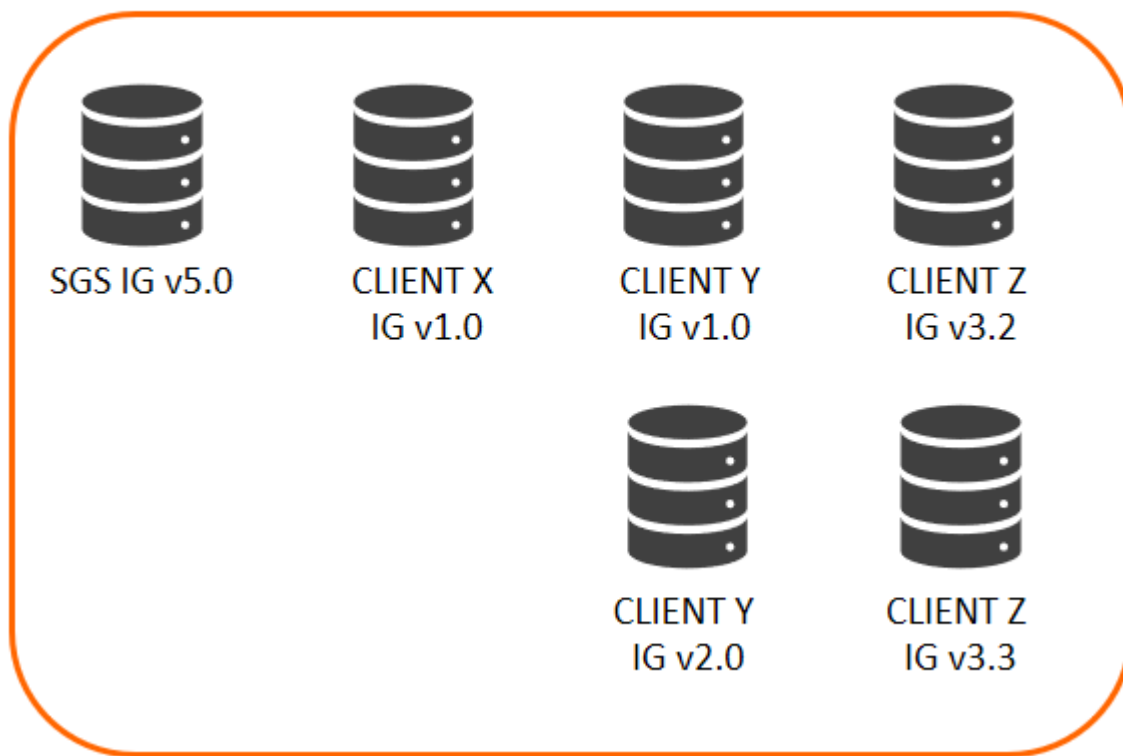


Figure 2 Multiple metadata repository libraries

Another possibility is to create a new library for the client. We usually start with our own implementation and perform updates at the client's request. This way of working allows a lot of flexibility with regards to their implementation of CDISC standards. Depending on the CDISC experience and long-term desires of the client, this MDR concept offers different choices to implement a new CDISC standard either using the SGS or client specifications. For example, the SGS implementation guide can be used by a client as a starting point to organize their trial data in a way which is fully compliant with the latest CDISC and regulatory requirements. Client specific requirements deviating from the SGS IG standards can be updated and implemented in their trial and automatically contributing to the build of a client specific implementation guide. After several trials, the client specific metadata library will take shape, and every new trial will be able to use this library and benefit from it. Maintenance of the metadata can be performed at either side, depending on the client's preference. For clients with their own implementation guide and metadata in place, their metadata library can be implemented in the metadata repository and can be used for their trials, in full accordance with their specifications.

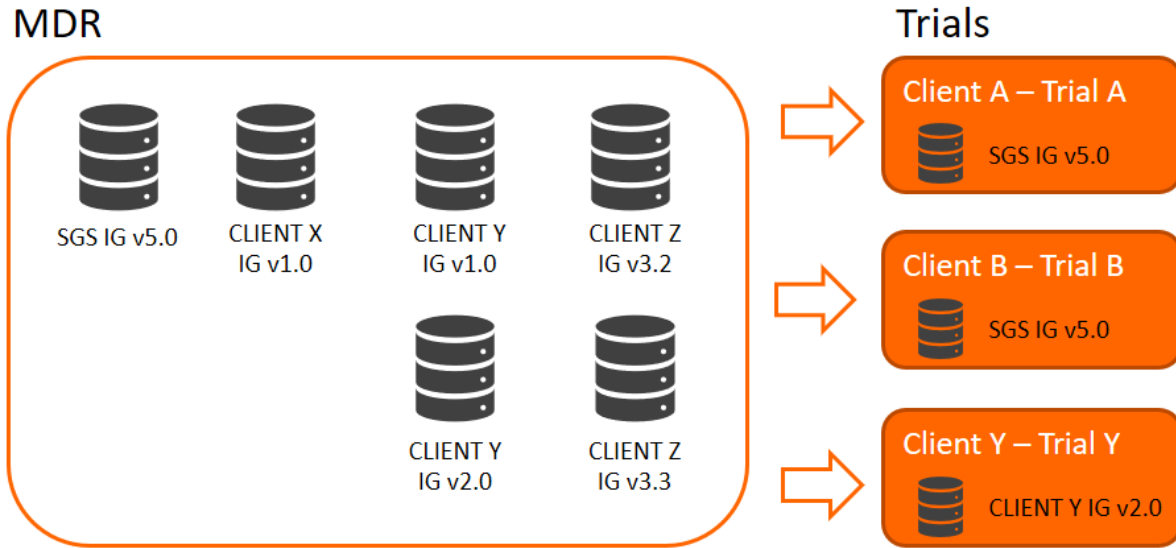


Figure 3 Study specific MDR library implementation

The creation of the MDR was our first step toward a metadata driven approach. As mentioned in the introduction, we at SGS went a few steps beyond this and developed in-house tools linked to the MDR to achieve end to end automation of the SDTM workflow. Linking the in-house tools to MDR paved the way for a metadata centric approach. The tools that have played a major role in the end to end automation are the **AnnotTool** – a tool for automating the CRF annotations, the **MappingTool** – a tool for automating mapping of the EDC extracts to SDTM, the **MetadataCenter** – a tool for automating set up of trial specific metadata and **YADLi** – a tool for automating the upload and mapping of external data from providers. All these tools together enable us achieve end to end automation of the SDTM workflow, including creation of the **Define.xml**.

MDR

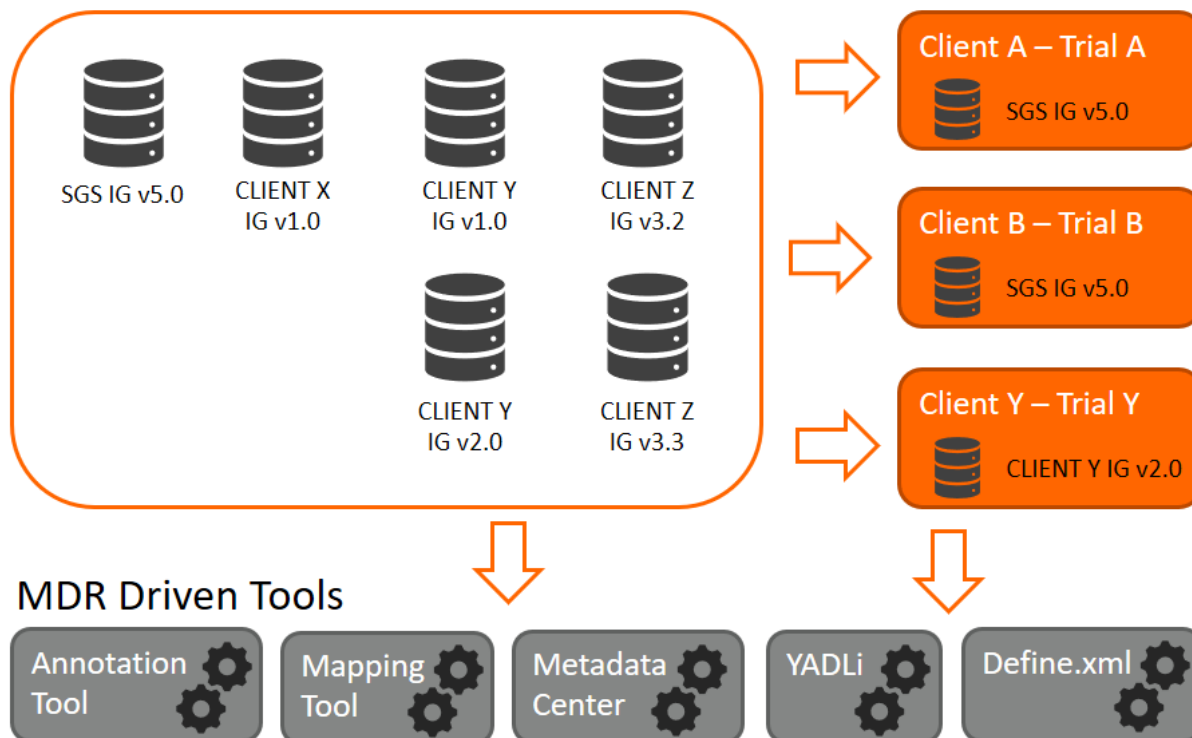


Figure 4 MDR Driven Tools

MDR DRIVEN APPLICATIONS DEVELOPED IN-HOUSE

ANNOTATION TOOL

This tool guides the clinical data manager or programmer in annotating a CRF based on the applicable standard. For the Annotation Tool, we have created different annotation types, see **Figure 5**. These types are all the possible annotations that an end user will be able to use to annotate a CRF. An annotation type can be compared with a template.

E.g. annotation type TESTCD will enable the user to annotate a result of a specific test on the CRF.

#DOMAIN#ORRES when #DOMAIN#TESTCD = #FINDING#

Suppose the user wants to annotate the result of the height of a subject, the end user can select annotation type TESTCD and provide the DOMAIN and FINDING, respectively 'VS' and 'HEIGHT' in this example. The Annotation Tool will replace the template variables #DOMAIN# and #FINDING# with the values provided by the user and will place the newly created annotation on the CRF.

Height [Height] **VSORRES when VSTESTCD = HEIGHT**

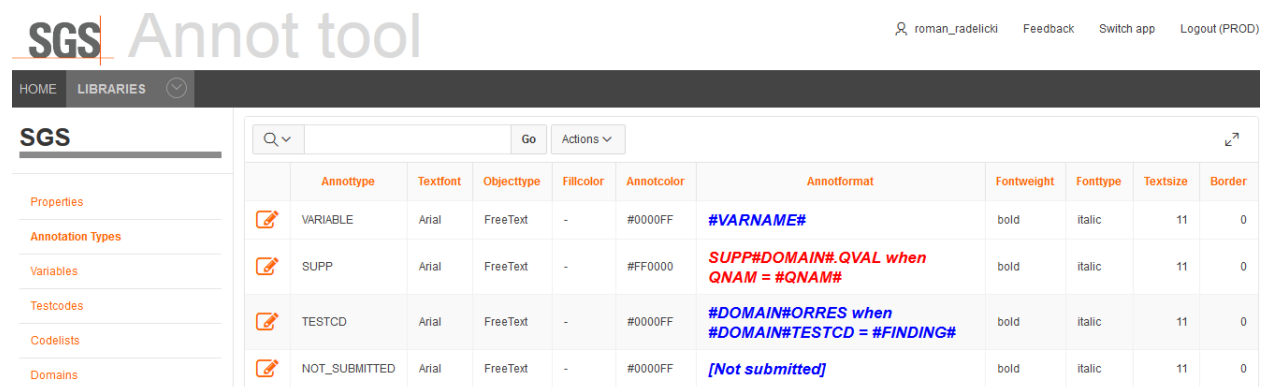
In the Annotation Tool, we have incorporated other annotation types, such as SUPP – for annotating supplemental qualifiers, VARIABLE – for annotating specific variables, RELREC – for RELREC related annotations and COVAL – for the annotation of comments.

In the past, before we had the Annotation Tool, we worked in Adobe Acrobat and were confronted with many obstacles. The look and feel of the annotations: size, font type and color, were not always consistent. The properties of the annotations, such as the author and subject fields were also not populated consistently. When a new page was added to the middle of a CRF, for example, in the case of CRF adaptations due to protocol amendments, all the references to pages suddenly became incorrect. Having different colors per domain on a CRF page, as per the

guideline, was prone to error. As some clients have other specifications about the lay-out of annotations, the most practical and efficient way of annotating a new CRF was to copy from previous trials. We needed to manually select the applicable annotations from the PDF and copy/paste them on the page of the CRF to be annotated. Mistakes made in the previous annotations were copied from trial to trial. Updated guidelines from the client, which impacted the aCRF's from all trials, needed to be addressed separately for each trial.

All these obstacles were overcome with the introduction of the Annotation Tool. As most clients' guidelines in regard to annotations differ from each other, we decided to link their annotation settings to their metadata repositories. This meant that we could define the lay-out of the annotation once per MDR library.

This way of working allowed us to be flexible in regard to client requests, be consistent towards future annotations and keep the lay-out settings for different clients separate from each other.



The screenshot shows the 'SGS Annot tool' interface. On the left is a sidebar with a search bar and a list of categories: Properties, Annotation Types, Variables, Testcodes, Codelists, and Domains. The main area displays a table of annotation types. The table has columns for Annottype, Textfont, Objecttype, Fillcolor, Annotcolor, Annotformat, Fontweight, Fonttype, Textsize, and Border. The data rows are as follows:

Annottype	Textfont	Objecttype	Fillcolor	Annotcolor	Annotformat	Fontweight	Fonttype	Textsize	Border
VARIABLE	Arial	FreeText	-	#0000FF	#VARNAME#	bold	italic	11	0
SUPP	Arial	FreeText	-	#FF0000	SUPP#DOMAIN#.QVAL when QNAM = #QNAM#	bold	italic	11	0
TESTCD	Arial	FreeText	-	#0000FF	#DOMAIN#ORRES when #DOMAIN#TESTCD = #FINDING#	bold	italic	11	0
NOT_SUBMITTED	Arial	FreeText	-	#0000FF	[Not submitted]	bold	italic	11	0

Figure 5 Annotation types and layout settings

The user will be able to copy annotations from the annotation library or will be guided to add a new annotation. Worrying about the layout of annotations, including font, font size and color is a thing of the past. The tool will apply the correct layout for the annotations based on the client or SGS standard. This figure shows all the available annotation types and their layout settings for a specific standard.

Adding a page in the middle of a CRF will have no impact on the referenced pages in the CRF, as the tool will automatically change the page numbers where needed. The domains on top of the annotated page are automatically added by the tool, and the colors per domain are also handled by the system.

After loading a PDF version of the CRF into the tool, the user can start annotating the document and will be guided by the tool. E.g. if a form was already annotated for another trial, the user will be able to copy the annotation by selecting the appropriate page from the library of annotated CRF's and modify if it needed. The more annotations are performed in the Annotation Tool, the smarter it gets, and the user's ability to copy and reuse annotations improves. Should a new annotation be added on a page, a wizard will guide the user in creating the correct annotation.

ADD ANNOTATION

Annotation type: TESTCD.....#DOMAIN#ORRES when #DOMAIN#TESTCD = #FINDING#

Domain: VS.....Vital Signs

Finding:

- MUARMCIR.....Mid-Upper Arm Circumference
- OXYSAT.....Oxygen Saturation
- PULSE.....Pulse Rate
- PULSE.....Pulse Rate
- PULSEPR.....Pulse Pressure
- RESP.....Respiratory Rate
- SAD.....Sagittal Abdominal Diameter
- SSSKNF.....Subscapular Skinfold Thickness
- SYSBP.....Systolic Blood Pressure
- TBW.....Total Body Water
- TEMP.....Temperature
- TRSKNF.....Triceps Skinfold Thickness
- VSALL.....All Vital Signs (not approved)
- WEIGHT.....Weight**
- WSTCIR.....Waist Circumference

Figure 6 Add an annotation

Figure 6 shows the wizard that allows the user to add a finding to the Vital Signs form. First, the user selects the annotation type, in this case, a finding. Then a domain is selected from all finding domains available for the applicable standard. This information is collected from the applicable client library in the MDR. Finally, the user selects the correct test code for the annotation from a list of all available test codes for domain VS for that specific client standard. The Annotation Tool not only ensures the correct look and feel of the annotations in accordance with the client's guidelines, CDISC and regulatory specifications, it also ensures compatibility with client specific metadata via the link with our MDR.

SGS Annot tool roman_radelicki

HOME CRF REPORTS LIBRARIES (1142)

044. ----- Demographics 1.....43|.....16|.....

[DM - Demographics]

Demographics 1

Date of birth	DM1.DOB	BRTHDTC
Sex at birth	DM1.SEX	☐ Male SEX ☐ Female
Race	DM1.RACE	☐ Not allowed to ask as per local regulations ☐ Please specify: RACE
Specify race:	DM1.RACES	RACE
Ethnicity	DM1.ETHN	☐ Not allowed to ask as per local regulations ☐ Hispanic or Latino ☐ Not Hispanic or Latino ETHNIC

ETHNIC = NOT ALLOWED TO ASK AS PER LOCAL REGULATIONS

Figure 7 Annotated Demographics form in the Annotation Tool



Figure 7 illustrates how an annotated Demographics page will look in the Annotation Tool

Another benefit of the tool is a tracking capability during the (CRF) review process. For each annotation on the CRF, a status can be applied which allows the reviewers to indicate that the applicable annotation has been reviewed or approved. **Figure 8** shows all available status settings in the Annotation Tool. When all annotations are reviewed and approved, the user will be able to export the annotations to a PDF and share them with involved parties.

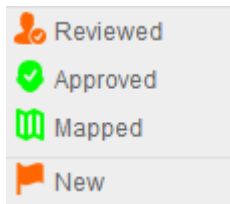


Figure 8 Status in the Annotation Tool

All these features will result in a more accurate and better quality annotated CRF. In the background, all annotations are captured electronically in the database and can therefore be used in other processes, such as the creation of the Define.xml or in the mapping of the CRF to SDTM, further down the line. Knowing the domains, variables, valuelists and codelists used in the annotation enables us to make the metadata for a certain trial more specific. We can automatically define the origin to “CRF” of the variables from the Annotation Tool and can even populate the applicable CRF page. The Annotation Tool also enables requests for amending SGS maintained libraries. For example, if a new TEST/TESTCD doesn't exist in the library, the programmer can add a request in the tool, which, once approved, will automatically be added to the applicable library.

One of the biggest advantages of the metadata centric approach is the ability to apply the plug and play method. If we receive a new client implementation guide, we need to be able to see their metadata library in our MDR. With the Annotation Tool we can copy all settings (look and feel of the annotation) from another implementation guide and simply adapt them to the needs of the client. We can then go ahead, with no additional programming needed. This approach can be applied to most in-house developed applications.

CHALLENGES

As the Annotation Tool is a web application, one of the biggest challenges was being able to ensure an accurate representation of the annotations on the screen and, at the end, in the exported PDF file. We needed to be sure the location and the size of the annotations as presented on screen on different computers were exported to the PDF exactly as they should be.

OPPORTUNITIES

There are additional opportunities for the future. The information gathered by the Annotation Tool should allow us to further automatize the creation of annotations without relying on human input. If a form in a CRF is a close match to an existing annotated form in the library, the tool could automatically annotate the form and immediately present it to the reviewer.

SDTM MAPPING TOOL

Based on the annotated CRF, the clinical data programmer will transform the EDC source data to SDTM.

To facilitate this process, we created the SDTM mapping tool, which will guide the programmer in performing the mapping, providing high flexibility while maintaining high quality CDISC SDTM datasets.

The process begins with the raw data. Raw clinical data used to be captured on paper; now, we receive this data electronically in many different formats and structures. We can distinguish between two major Electronic Data Capture (EDC) systems used to capture clinical data electronically: eCRF and eSource system. Each of these systems has different vendors, for example, Medidata Rave®, Oracle Inform® and Foundry Health ClinSpark®. Each of those vendors provides a structure and file formats for delivering clinical data to our SDTM programmer.

To enable the programmer to map all these different kinds of formats and structures of (e)source data, we have come up with a process called pre-conversion. Pre-conversion will transform any kind of source data into one general structure which is then used by the SDTM mapping tool. The general structure can be compared to a pivot table, where for each form, all input fields from the CRF are placed in rows. This transformation needs to be programmed once per vendor/system and can then be used for all future trials. We can apply the plug and play approach here. For

every new source system, we must program a pre-conversion for the general structure, after which we can plug it into the SDTM mapping tool to ensure it can be used for all future trials in that specific source system.

To learn more about our pre-conversion, check out the PhUSE US Connect 2019 paper DH03 “From Source Data to SDTM: Enabling a Semi-automated Conversion” from our SGS colleague Pieter Blomme.

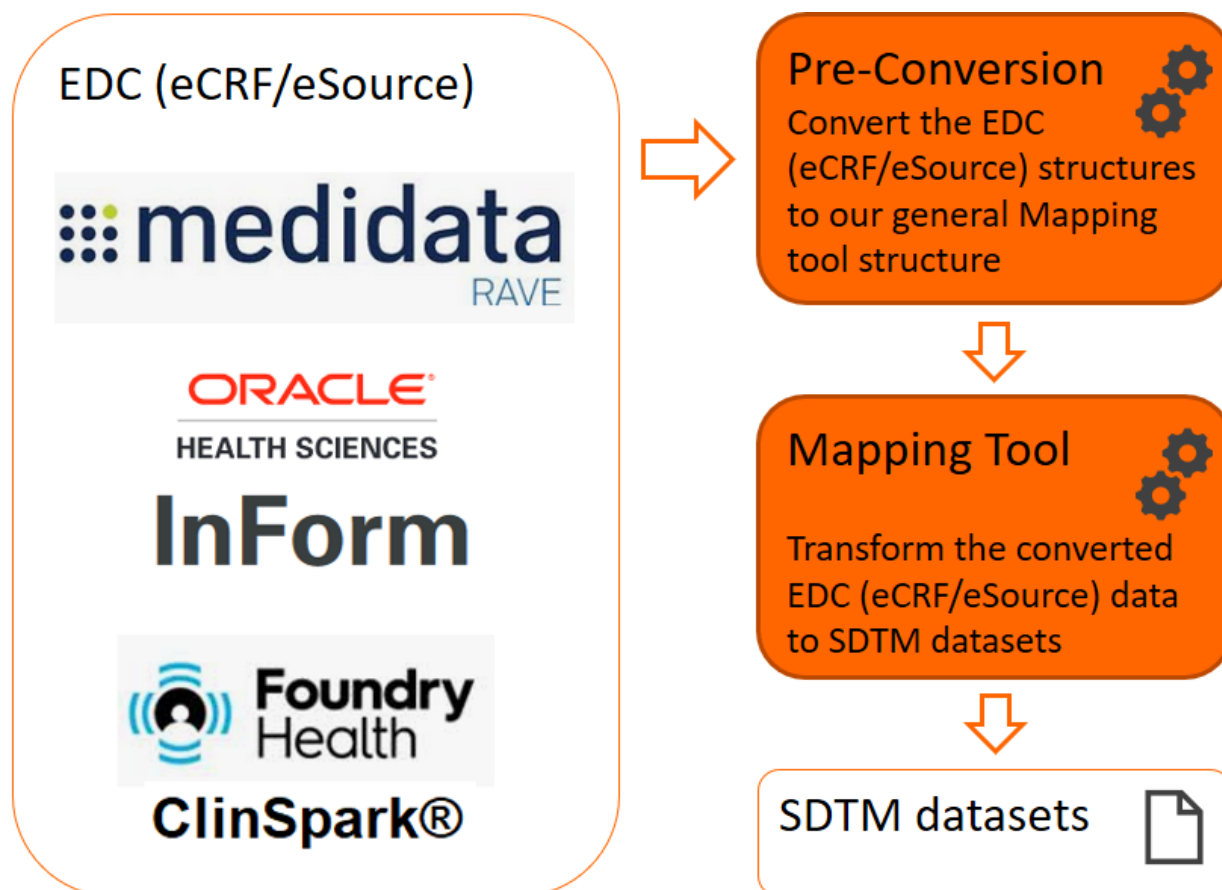
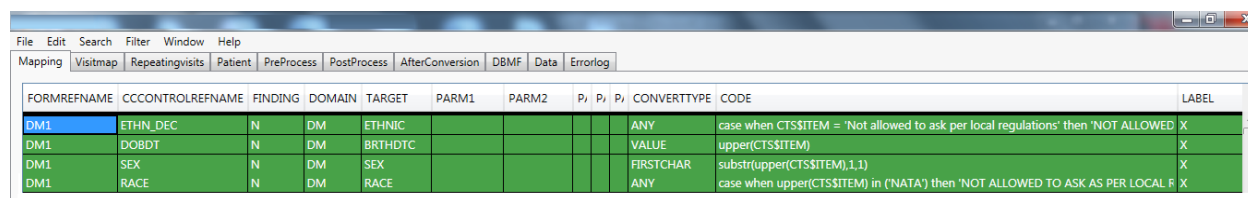


Figure 9 SDTM conversion flow



The screenshot shows the SDTM Mapping tool interface. It has a menu bar (File, Edit, Search, Filter, Window, Help) and a tab bar (Mapping, Visitmap, Repeatingvisits, Patient, PreProcess, PostProcess, AfterConversion, DBMF, Data, Errorlog). The main area displays a table with columns: FORMREFNAME, CCCONTROLREFNAME, FINDING, DOMAIN, TARGET, PARM1, PARM2, P1, P2, CONVERTTYPE, CODE, and LABEL. The table contains four rows of data for form DM1.

FORMREFNAME	CCCONTROLREFNAME	FINDING	DOMAIN	TARGET	PARM1	PARM2	P1	P2	CONVERTTYPE	CODE	LABEL
DM1	ETHN_DEC	N	DM	ETHNIC					ANY	case when CTSSITEM = 'Not allowed to ask per local regulations' then 'NOT ALLOWED	X
DM1	DOBDT	N	DM	BRTHDTC					VALUE	upper(CTSSITEM)	X
DM1	SEX	N	DM	SEX					FIRSTCHAR	substr(upper(CTSSITEM),1,1)	X
DM1	RACE	N	DM	RACE					ANY	case when upper(CTSSITEM) in ('NATA') then 'NOT ALLOWED TO ASK AS PER LOCAL R	X

Figure 10 SDTM Mapping tool

Figure 10 illustrates the use of the mapping tool. All input fields, radio buttons and checkboxes from the CRF are captured and presented to the programmer, in a format resembling a pivot table, in the mapping tool. For each form, the programmer can map each column to a specific SDTM variable. The first line in **Figure 10** DM1.ETHN_DEC corresponds with the question ‘Ethnicity’ on the CRF in **Figure 7**.

The SDTM mapping tool relies on the MDR libraries to guide the user in selecting the correct SDTM domains and variables for mapping. When selecting a domain or variable, the programmer will be prompted with all possible domains and variables that are applicable to that specific metadata library. The mapping tool keeps a history of all forms previously mapped, enabling the programmer to check if a form was mapped before, and then simply copy the mapping and apply updates if needed.

METADATA CENTER

Metadata Center is a tool to automate the set up of trial specific metadata. At trial setup, a copy of the applicable MDR library is performed on trial level. The intention is to produce a define.xml file based on the trial specific metadata. To make this happen, the programmer will have to make the copy of the MDR library on trial level specific to the trial.

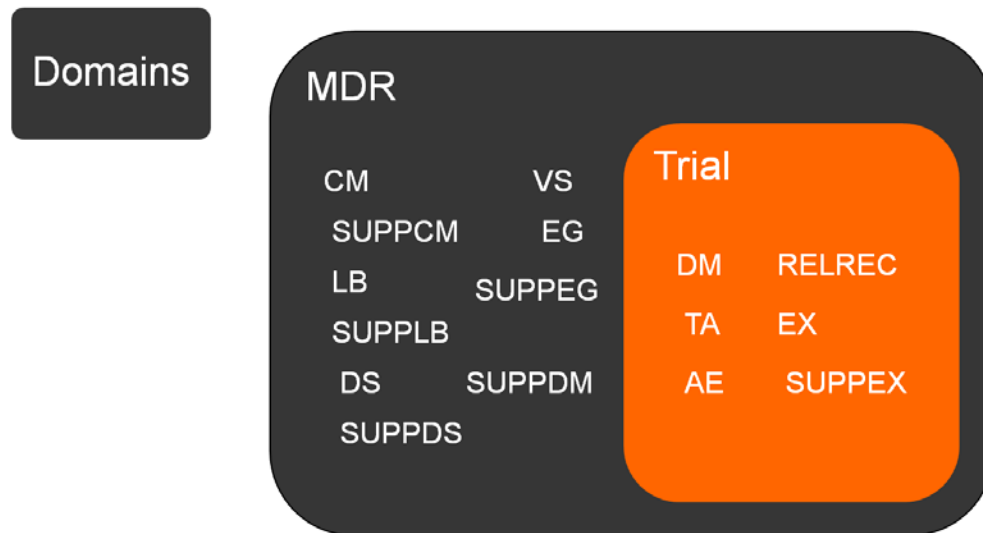


Figure 11 In orange, a selection of domains applicable for a specific trial

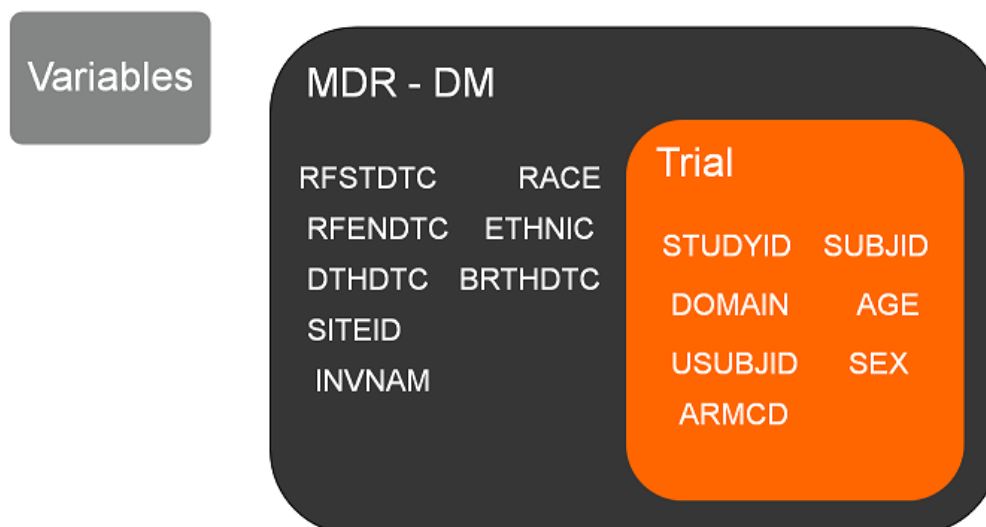


Figure 12 In orange, a selection of variables for domain DM for a specific trial



Figure 11 and **Figure 12** illustrate how we select, from the copy of the complete MDR library, only the domains and variables that are applicable for a specific trial.

DATASET	DESCRIPTION	CLASS	STRUCTURE	PURPOSE	KEYS	DSORDER	SGSSTATE
TA	Trial Arms	Trial Design	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD	001	IN
DM	Demographics	Special-Purpose	One record per subject	Tabulation	STUDYID, USUBJID	009	IN
CM	Concomitant/Prior Medications	Interventions	One record per recorded intervention occurrence or const...	Tabulation	STUDYID, USUBJID, CMTRT, CMSTDTC, CMSCAT	013	EX
EX	Exposure	Interventions	One record per protocol-specified study treatment, const...	Tabulation	STUDYID, USUBJID, EXTRT, EXSTDTC, EXPTPTNUM, EXREFID	015	IN
AE	Adverse Events	Events	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AEDECOD, AESTDTC	020	IN
DS	Disposition	Events	One record per disposition status or protocol milestone p...	Tabulation	STUDYID, USUBJID, DSDECOD, DSSTDTC, DSSCAT	023	EX
EG	ECG Test Results	Findings	One record per ECG observation per time point per visit p...	Tabulation	STUDYID, USUBJID, EGTESTCD, VISITNUM, EGTPTRF, EGTPPTNUM	031	EX
LB	Laboratory Test Results	Findings	One record per lab test per time point per visit per subject	Tabulation	STUDYID, USUBJID, LBTESTCD, LBSPEC, VISITNUM, LBPTREF, LBPTPTNUM	036	EX
VS	Vital Signs	Findings	One record per vital sign measurement per time point pe...	Tabulation	STUDYID, USUBJID, VSTESTCD, VISITNUM, VSTPTREF, VSTPTPTNUM	053	EX
RELREC	Related Records	Relationship	One record per related record, group of records or dataset	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, RELID	060	IN
SUPPEX	Supplemental Qualifiers for EX	Relationship	One record per IDVAR, IDVARVAL and QNAM per subject	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM	074	IN
SUPPLB	Supplemental Qualifiers for LB	Relationship	One record per IDVAR, IDVARVAL and QNAM per subject	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM	080	EX
SUPPVS	Supplemental Qualifiers for VS	Relationship	One record per IDVAR, IDVARVAL and QNAM per subject	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM	103	EX

Figure 13 Database view of the selected domains

Figure 13 shows the modified metadata on trial level for domains in Metadata Center. Indicating a domain is applicable to your trial is done by using the SGSSTATE flag (last column in **Figure 13**). SGSSTATE 'IN' will include the domain in the Define.xml and SGSSTATE 'EX' will exclude the domain. This method of in-or excluding domains is used for variables, valuelists, codelists, comments and computational methods.

SDTM-IG 3.2

Annotated Case Report Form

- ▶ Tabulation Datasets
- ▶ Value Level Metadata
- ▶ Controlled Terminology
- ▶ Computational Algorithms
- ▶ Comments

Tabulation Datasets for Study SGSTRIAL (SDTM 3.2) (SDTM-IG 3.2)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
TA	Trial Arms	Trial Design	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD	TA.xpt	
DM	Demographics	Special-Purpose	One record per subject	Tabulation	STUDYID, USUBJID	DM.xpt	
EX	Exposure	Interventions	One record per protocol-specified study treatment, constant-dosing interval, per subject	Tabulation	STUDYID, USUBJID, EXTRT, EXSTDTC	EX.xpt	
AE	Adverse Events	Events	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AETERM, AESTDTC	AE.xpt	
RELREC	Related Records	Relationship	One record per related record, group of records or dataset	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, RELID	RELREC.xpt	
SUPPEX	Supplemental Qualifiers for EX	Relationship	One record per IDVAR, IDVARVAL and QNAM per subject	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM	SUPPEX.xpt	

Go to the [top](#) of the define.xml

Go to the [top](#) of the define.xml

Figure 14 provides an example of the trial specific Define.xml

From the trial level metadata database tables SGS has an application that will create a Define.xml. **Figure 14** illustrates the domain part of the Define.xml based on the tables from **Figure 13**.

After using the SGSSTATE flag, the programmer can further manipulate the metadata where needed. Depending on when the metadata needs to be set up, some metadata manipulation can be performed automatically by Metadata Center. We use information collected via other in-house developed applications to automatically make the metadata trial specific. If the aCRF is already available in the Annotation Tool, Metadata Center can indicate automatically which domains, variables, valuelists and codelists from the CRF should be included in the trial specific metadata. The origin can then be populated with 'CRF' and the CRF page can be provided without manual action. When the mapping of the EDC data to SDTM has been performed with the SDTM mapping tool, Metadata Center can calculate the maximum length of the variables in the SDTM datasets according to the FDA technical conformance guide, which states that the allotted length for each column containing character (text) data should be set to the maximum length of the variable used across all datasets in the study, except for suppqual datasets.

Another feature of the Metadata Center is metadata checks. These checks are an addition to the Pinnacle21 validation of the define.xml and will help the programmer to set up the metadata before the define.xml is produced. These are rather straightforward checks that validate the integrity of the metadata and compliance with the specific metadata repository library.

The final feature of Metadata Center is the audit trail. Every manipulation to the metadata on trial level will leave an audit record. The audit records contain the old record before the manipulation, the person who manipulated the record, the manipulation performed (update, delete or insert) and the date when the manipulation was performed. The audit trail will allow us to identify all the changes that were performed from and to specific dates. This is helpful when a client requires a report with metadata changes performed since the previous transfer of metadata. It also enables us to retrieve the metadata as it was on a specific date.



CHALLENGES AND OPPORTUNITIES

In the past, we mainly used Metadata Center very late in the process of creating SDTM datasets. When the annotation and conversion of data to SDTM took place, metadata was partly automatically created based on the annotated CRF and the SDTM datasets. Because metadata was created only at the end of the process, it didn't receive the attention it required and because of time constraints, it was difficult to deliver the high quality metadata needed for submissions.

Now, we see opportunities for creating the metadata closer to the trial setup. Some aspects still need to be maintained during trial conduct, such as the length of the variables, which can change with every transfer of new data. Nevertheless the creation of qualitative metadata at the beginning of the process will pave the path for future task such as the conversion to SDTM datasets. Due to the increased awareness and interest in the define.xml at client side, we will also benefit from discussing the metadata upfront and working together towards a stress-free database lock.

YADLI

YADLi is short for Yet Another Data Loader Improved. This is a tool which enables the clinical programmer to load any kind of source data into the database and perform data manipulations to load the data in the SDTM database. With the creation of the DTA – Data Transfer Agreement, the format and structure of the data is agreed upon by the provider and client. Afterwards a test transfer is received from the provider, which will be checked against the DTA using YADLi.

Our programmer stores the transfer file on our filesystem, after which YADLi loads the file onto the database. YADLi is able to load multiple formats including any delimited text file in TXT or CSV format, Excel files and SAS® files: sas7bdat or XPT. YADLi can automatically determine the delimiter used in the text file and uses the Oracle external tables technology to present the data as a database table.

Now the work begins. The programmer will have to map the original data to SDTM. This can be achieved by accessing the YADLi web application. However, the data will first need to be checked against what was agreed upon in the DTA. The programmer can add some checks and constraints to YADLi, which will run with every new transfer being loaded in the tool. This way, DTA compliance is guaranteed from the first test transfer until the final data is received from the provider. Some of the checks which a programmer can implement are: check if the column is required, check the data against a predefined codelist, check the case, length and type of column. At this point YADLi also makes use of our MDR libraries. The tool contains SDTM checks which will check the mapped data against the applicable metadata library used.

When we have already received a transfer for the same type of data from the same provider, the programmer will be able to perform a copy of the mapping and modify it if needed. The more transfers we receive and load, the smarter YADLi becomes.

When the mapping is performed, the tool generates the mapped dataset based on the original loaded data transfer. This dataset can then be loaded into the SDTM database or be used for any other purpose, for example for sample identifier reconciliation.

Some of the best features of YADLi are logging and traceability. In the tool, we have a complete audit history of all modifications performed on the mapping, when they were performed and by whom. We also know which transfer was loaded, when and by whom. This is not only useful from an auditor's point of view, but also provides the clinical data manager and programmer with a clear overview of which data is present in the SDTM database.

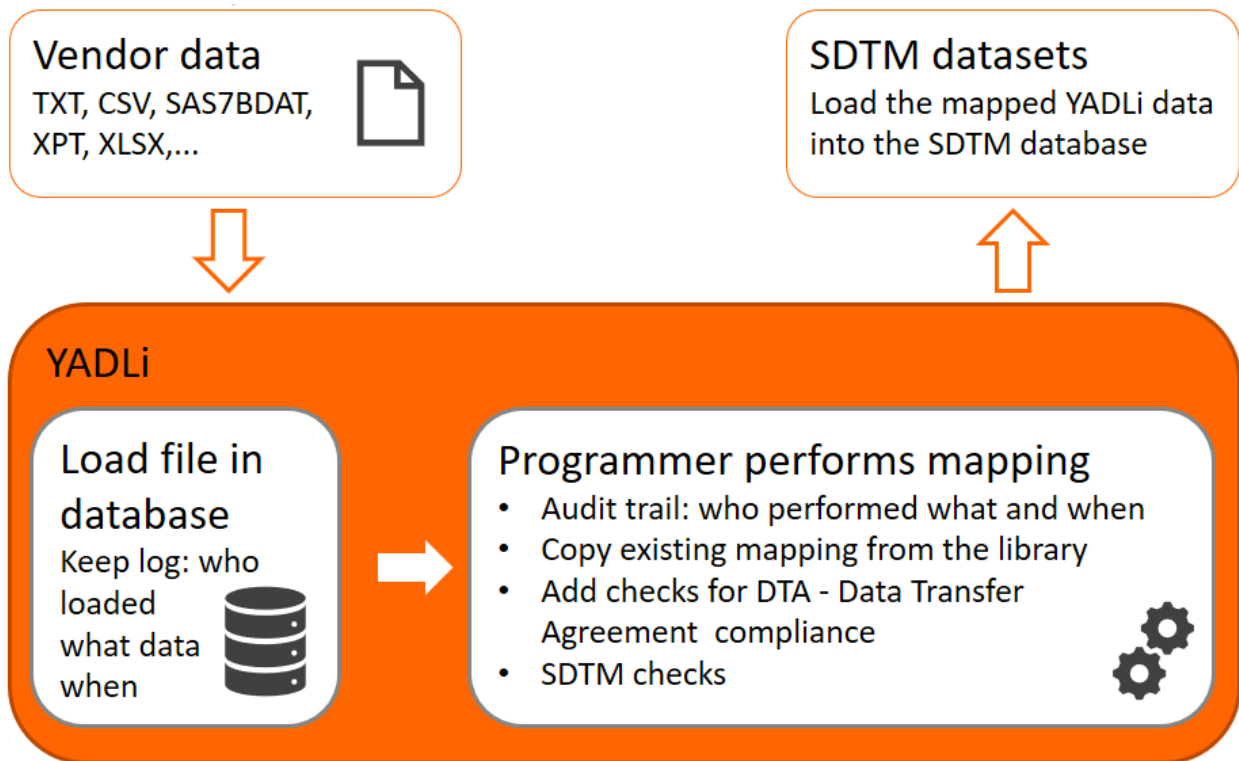


Figure 15 YADLi flow

CHALLENGES

The biggest challenge we have had with YADLi was being able to load any kind of data transfer in any kind of format. The ideal format would be JSON or XML, but based on our experience with providers, this does not seem to be achievable in the near future. In the best cases, we receive ASCII delimited text files or SAS files. Unfortunately, we often have to use Excel files. Yes, dear reader, I know what you are thinking, it's 2019 and there are still people out there who prefer using excel files to transfer data.

OPPORTUNITIES

One opportunity might be to integrate YADLi into the SDTM Mapping Tool. They are basically designed to perform the same task, map external data to SDTM.

CONCLUSION

Linking the tools to MDR to automate SDTM workflows has paid off big time at SGS. While the road to the metadata centric approach hasn't been an easy one, we have taken it and are reaping the benefits in more ways than we have foreseen. At the same time, this is a work in progress with constant improvements being made in accordance with the changing needs of the client, industry and regulatory agencies.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

clinicalresearch@sgs.com
 EUROPE: +32 15 27 32 45
 AMERICAS: + 877 677 2667
 WWW.SGS.COM/CRO

Brand and product names are trademarks of their respective companies.