

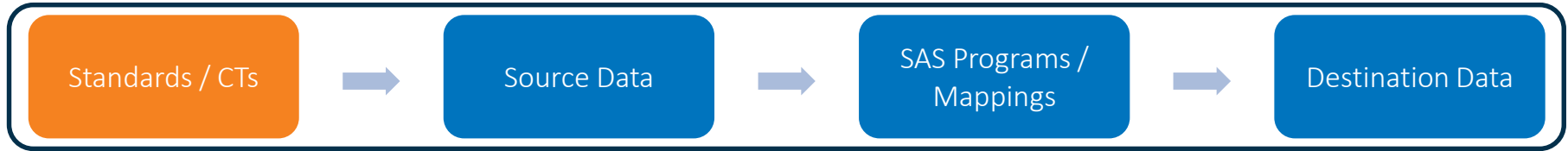


# Data Mapping Using Machine Learning

SAS Institute Japan Ltd.  
Toru Tsunoda (Toru.Tsunoda@sas.com)

# Data Mapping

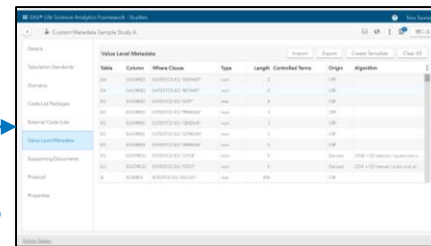
## Problem Definition



- Source Data mapped to Destination Data based on Standards
  - Time-consuming process
- Standards and Implementation Guides leave room for individual interpretation
  - Inconsistent Mappings – Not what a “Standard” should be
- Mapping is generally done in individual SAS programs
  - Lacks collection of central metadata/mappings
  - Only way to re-use previous mappings is copying programs

# History of Clinical Data Mapping tools

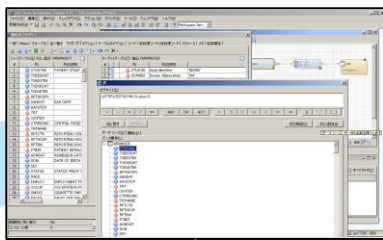
## Various SAS tools



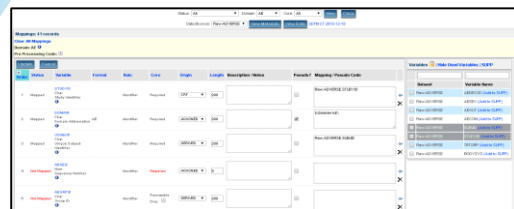
**Life Science Analytics Framework**  
(ex SAS Drug Development)

CDR(Clinical Data Repository) plus  
MDR(Meta Data Repository)

**Clinical Data Integration**



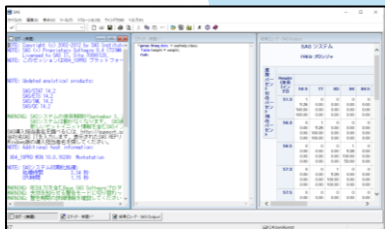
**Analytics Pro (PC-SAS)**



**Data Mapper**  
(ex LSAF Extension)

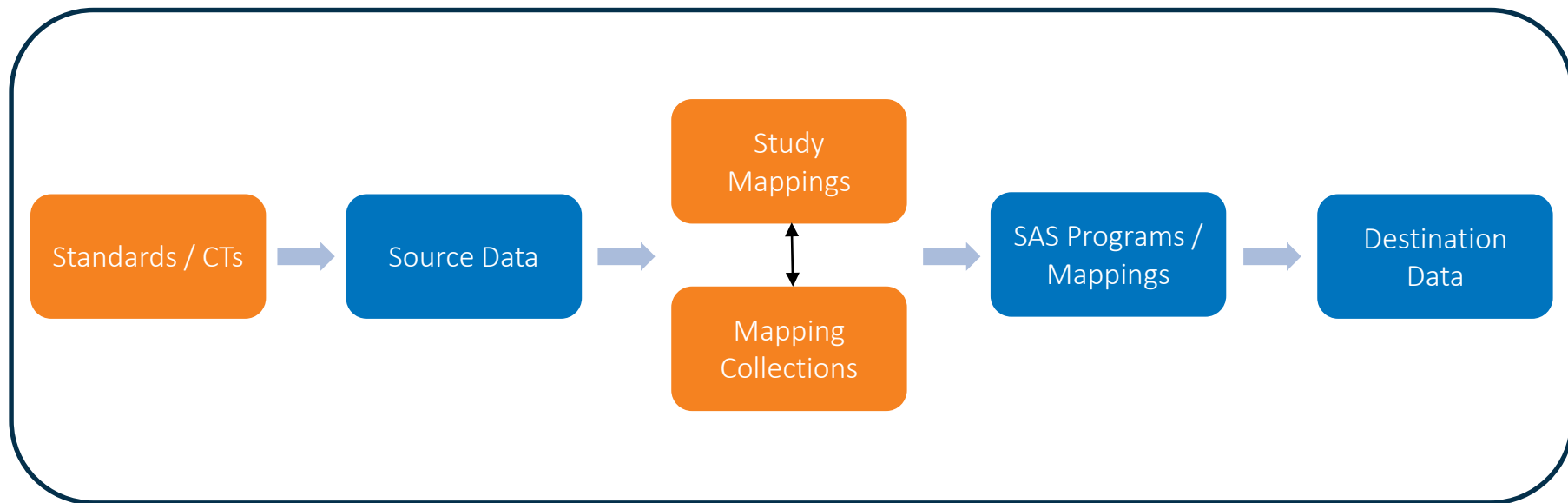
Programing tool  
using GUI

Programing tool  
using Machine Learning



# Data Mapping using “Data Mapper”

Solution

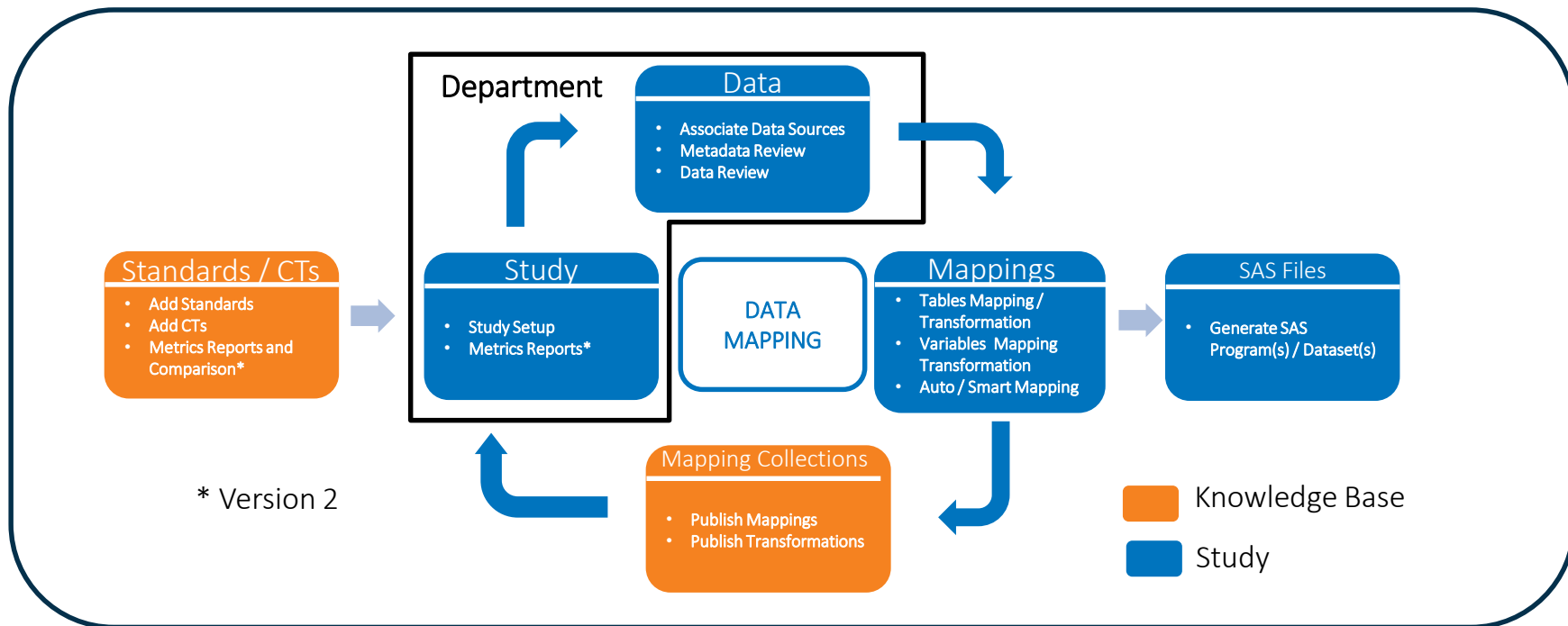


# Data Mapper

## Solution

- User interface to mapping process
- **Libraries** - Collection of mapping rules in central database
- Collect data and standards, and allow user to define rules for mapping
- Ability to re-use mapping rules in future studies
- **Auto Mapping** - One-to-One mapping Rule between Source Data and its variable, to Destination Data and its variable
  - e.g. adverse.patid → AE.USUBJID
  - Next study, adverse.patid is automatically mapped
- **Smart Mapping** - Similar Variables mapped in past provide guidelines to map new variables
  - e.g. adverse.ptid → AE.USUBJID
  - Suggested mapping based on previous mapping, adverse.Patid → AE.USUBJID
- Generate the SAS programs based on defined mapping

# Data Mapper



# Data Mapper

## Administration - Users

- Users – Ability to register user in system
- Authentication Type: Application / Active Directory

SAS® Data Mapper Administration

**Users** Number of Users: 36

**Roles**

**Departments**

**Data Sources**

Username	First Name	Last Name	Email	Company	Role	Authentication Method
admin	System					Application
avharv	Avy					Active Directory
bebocc	Ben					Active Directory
dechoi	Daniel					Active Directory
erbole	Eric					Active Directory
jeelio	Jenni					Active Directory
jeeliocu	Jenni					Active Directory
jimbox	Jim					Active Directory
jmadmin	Jenni					Application
jmelion	Jenni					Application
jmelion2	Jenni				wner	Application
jminactive	Jenni					Application

**Edit User**

Details

Account

All fields are required.

**Role\***

Admin

**Authentication Type\***

Active Directory

Please select an Authentication Type

Active Directory

Application

Active

**Last Login**

never

**Created**

August 1, 2018, 7:01 PM (EDT) by System Administrator (admin)

**Last Modified**

August 1, 2018, 7:01 PM (EDT) by System Administrator (admin)

# Data Mapper

## Administration - Roles

- Roles – Pre-defined set of Roles with different privileges
- Ability to add /remove privileges

SAS® Data Mapper Administration

Users

**Roles**

Departments

Data Sources

Number of Roles: 8

Name	Description
Admin	Predefined administrator role.
Mapper	Predefined mapper role.
Mapping Lead	Predefined mapping lead role.
Standards Admin	Predefined role to modify Standards.
Study Owner	Predefined study owner role.
Study-Add	QA Role-Add Study
Study-Both	QA Role-Add and Edit Study
Study-Edit	QA Role-Edit Study

Edit Role

Details

**Privileges**

Available

Ability to download data source file(s)  
Add study  
Admin: access to Admin Page and perform actions  
Generate SAS programs  
Modify data sources on a study  
Modify mapping libraries  
Modify study metadata

Selected






Modify controlled terminologies  
Modify data standards



# Data Mapper

## Administration – Data Sources

- Data Sources – define data locations

SAS® Data Mapper Administration	
 Users	Number of Data Sources: 14
 Roles	 Click refresh to scan for new data sources.
 Departments	
 Data Sources	
Name	▲ 🔊 Path
custom_formats	/tla/warehouse/oracle_content/data_sources/custom_formats
dataSourceA	/tla/warehouse/oracle_content/data_sources/dataSourceA
dataSourceB	/tla/warehouse/oracle_content/data_sources/dataSourceB
dataSourceC	/tla/warehouse/oracle_content/data_sources/dataSourceC
dupe	/tla/warehouse/oracle_content/data_sources/dupe
empty	/tla/warehouse/oracle_content/data_sources/empty
FSDDataSource1	/tla/warehouse/oracle_content/data_sources/FSDDataSource1
FSDDataSource2	/tla/warehouse/oracle_content/data_sources/FSDDataSource2
NICSAH_DATA	/tla/warehouse/oracle_content/data_sources/NICSAH_DATA

# Data Mapper

## Administration - Departments

- Departments – Group similar studies with associated data sources and members
- Provides better permission and data access control

SAS® Data Mapper Administration

Users

Roles

Departments

Data Sources

Number of Departments: 13

Name	▲ ▼	Users ▼	Studies ▼	Data Sources
CardioVascular		34	9	
CNS		0	0	
Dermatology		0	0	
Endocrinology		0	0	
Gastroenterology		0	0	

# Data Mapper

## Administration - Departments

- Details / Studies / Users / Data Sources

Edit Department

Details

Name\*

CardioVascular

Studies

Users

Data Sources

Description

Enter department description

Created

August 1, 2018, 7:02 PM (EDT) b

Last Modified

August 2, 2018, 10:28 AM (EDT)

### Edit Department

Details

Studies

Users

Data Sources

#### Available

Testing Study2  
Testing Study4  
Testing Study6  
Testing Study8  
Testing Study9

#### Selected

Testing Study  
Testing Study3  
Testing Study5  
Testing Study7

### Edit Department

Details

Studies

Users

Data Sources

#### Available

Jenni Test8 (jmtest8) : Mapper  
Jenni Test9 (jmtest9) : Mapping Lead

#### Selected

Avy Harvey (avharv) : Admin  
Ben Bocchicchio (bebocc) : Admin  
Brian Durham (sasbkd) : Admin  
Daniel Choi (dachoi) : Admin  
Dongliang Gao (scndlg) : Admin  
Eric Bolender (erbole) : Admin  
Jenni Admin (jmadmin) : Admin

### Edit Department

Details

Studies

Users

Data Sources

ⓘ Data sources can be refreshed by clicking the refresh button on the Data Sources page.

#### Available

dupe  
FSDataSource1  
FSDataSource2  
space test  
T4214320

#### Selected

custom\_formats  
dataSourceB  
dataSourceC  
empty  
qatest  
rwe\_1k  
sample

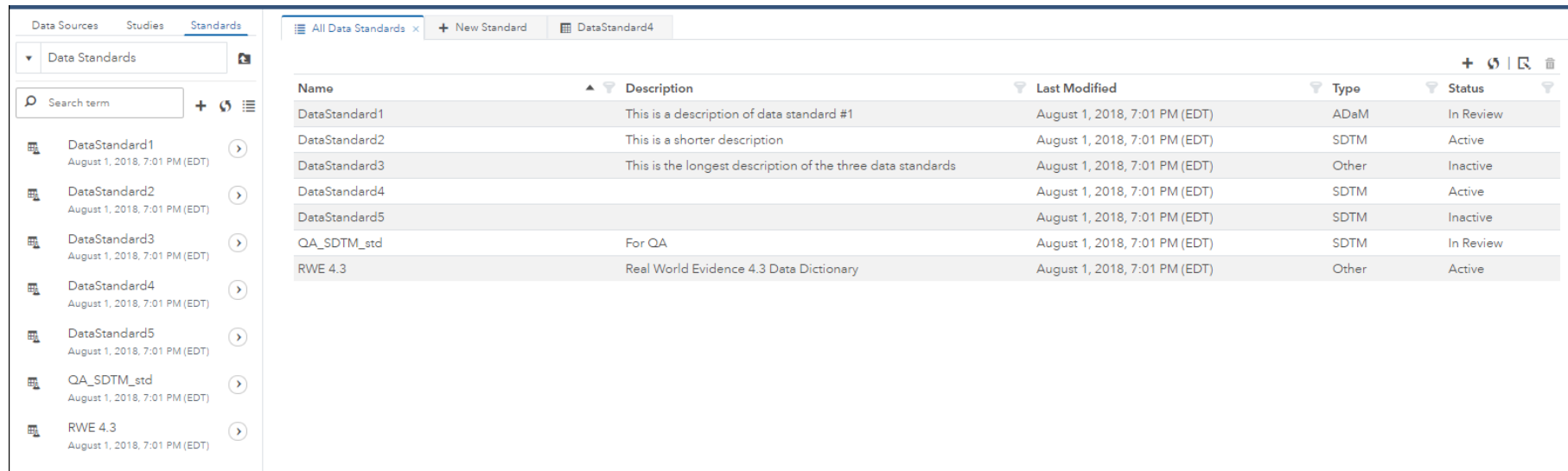
#### In Use ⓘ

- dataSourceA
- NICSAH\_DATA

# Data Mapper

## Knowledge Base - Standards

- Register Standard by Importing
- View list of Registered Standards



The screenshot displays the 'Standards' section of the Data Mapper Knowledge Base. The interface includes a sidebar on the left with a search bar and a list of standards. The main area shows a table of standards with columns for Name, Description, Last Modified, Type, and Status.

Name	Description	Last Modified	Type	Status
DataStandard1	This is a description of data standard #1	August 1, 2018, 7:01 PM (EDT)	ADaM	In Review
DataStandard2	This is a shorter description	August 1, 2018, 7:01 PM (EDT)	SDTM	Active
DataStandard3	This is the longest description of the three data standards	August 1, 2018, 7:01 PM (EDT)	Other	Inactive
DataStandard4		August 1, 2018, 7:01 PM (EDT)	SDTM	Active
DataStandard5		August 1, 2018, 7:01 PM (EDT)	SDTM	Inactive
QA_SDTM_std	For QA	August 1, 2018, 7:01 PM (EDT)	SDTM	In Review
RWE 4.3	Real World Evidence 4.3 Data Dictionary	August 1, 2018, 7:01 PM (EDT)	Other	Active

# Data Mapper

## Knowledge Base - Standards

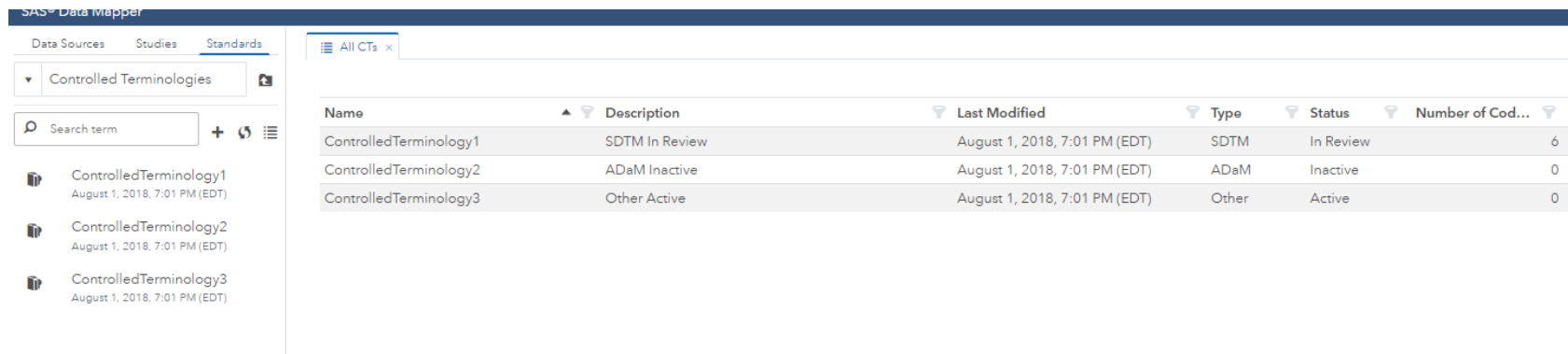
- View Domains Information
- View Domain Variables Details

<div> <div>All Data Standards</div> <div>DataStandard4</div> <div>DataStandard4/AE x</div> </div>												
Metadata		Variables										
Order▲	Name	Label	Description	Type	Length	Format	Origin	Role	Core			
1	STUDYID	Study Identifier	Unique identifier for a study.	CHAR	200		CRF	Identifier	Required			
2	DOMAIN	Domain Abbreviation	Two-character abbreviation for the domain.	CHAR	200	AE	ASSIGNED	Identifier	Required			
3	USUBJID	Unique Subject Identifier	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.	CHAR	200		DERIVED	Identifier	Required			
4	AESEQ	Sequence Number	Sequence Number given to ensure uniqueness of subject records within a domain. May be any valid number.	NUM	8		ASSIGNED	Identifier	Required			
5	AEGRPID	Group ID	Used to tie together a block of related records in a single domain for a subject.	CHAR	200		DERIVED	Identifier	Permissible			
TA	Trial Arms		STUDYID, ARMCD, TAETORD									
TE	Trial Elements		STUDYID, ETCD									
TI	Trial Inclusion/ Exclusion Criteria		STUDYID, IETESTCD									
TV	Trial Visits		STUDYID, VISITNUM, ARMCD									
VS	Vital Signs		STUDYID, USUBJID, VSTESTCD, VISITNUM, VSTPTREF, VSTPTNUM									

# Data Mapper

## Knowledge Base – Controlled Terminology (CT)

- Register Controlled Terminology by Importing
- View list of Registered Controlled Terminologies



The screenshot shows the SAS Data Mapper interface. The top navigation bar includes 'Data Sources', 'Studies', and 'Standards'. The 'Standards' tab is active, and 'Controlled Terminologies' is selected in the left sidebar. A search bar is present above a list of three controlled terminologies. The main panel displays a table titled 'All CTs' with the following data:

Name	Description	Last Modified	Type	Status	Number of Cod...
ControlledTerminology1	SDTM In Review	August 1, 2018, 7:01 PM (EDT)	SDTM	In Review	6
ControlledTerminology2	ADaM Inactive	August 1, 2018, 7:01 PM (EDT)	ADaM	Inactive	0
ControlledTerminology3	Other Active	August 1, 2018, 7:01 PM (EDT)	Other	Active	0

# Data Mapper

## Knowledge Base – Controlled Terminology (CT)

- View Codelists Information
- View Codelist Values

All CTs							
ControlledTerminology1 x							
Metadata Codelists							
All CTs							
ControlledTerminology1							
ControlledTerminology1/AgeBuckets x							
Code	▲ ▼ Codelists Code	▼ Codelist Name	▼ CDISC Submission Value	▼ CDISC Synonyms	▼ CDISC Definition	▼ NCI Preferred Term	▼
AGERANGE-1	QA2	AgeBuckets	<2		AGERANGE-1-CDISC DEFINITION	AGERANGE-1-NCI PREFERRED TERM	
AGERANGE-2	QA2	AgeBuckets	2-5		AGERANGE-2-CDISC DEFINITION	AGERANGE-2-NCI PREFERRED TERM	
AGERANGE-3	QA2	AgeBuckets	6-12		AGERANGE-3-CDISC DEFINITION	AGERANGE-3-NCI PREFERRED TERM	
AGERANGE-4	QA2	AgeBuckets	13-20		AGERANGE-4-CDISC DEFINITION	AGERANGE-4-NCI PREFERRED TERM	
AGERANGE-5	QA2	AgeBuckets	21-64		AGERANGE-5-CDISC DEFINITION	AGERANGE-5-NCI PREFERRED TERM	
AGERANGE-6	QA2	AgeBuckets	>64		AGERANGE-6-CDISC DEFINITION	AGERANGE-6-NCI PREFERRED TERM	

Standards Consortium  
Study Data Tabulation  
Model.

# Data Mapper

## Knowledge Base – Mapping Collection (Under Development)

- Define Mapping Collections by Importing
- View list of defined Mapping Collections



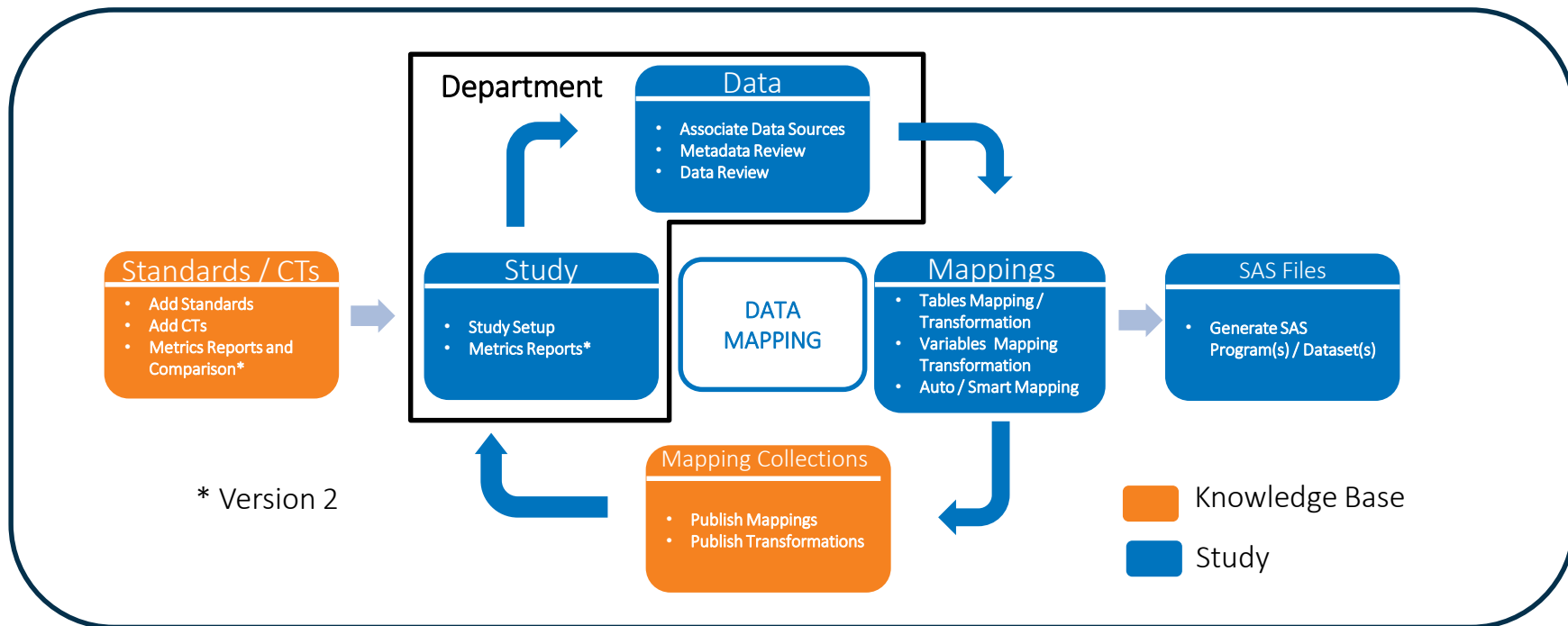
# Data Mapper

## Knowledge Base – Mapping Collections (Under Development)

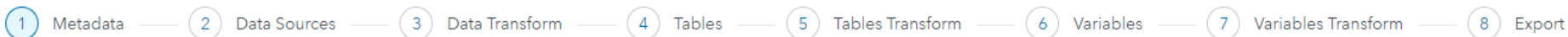
- View Mapping Collection Information

# Data Mapper

## Study Flow



# Study Mapping Process



## 1. Metadata –

- Define study name and description
- Mapping Library
- Associate Standards
- Associate Controlled Terminologies

**Name\***


**Description**

**Shared Mapping Collection**


Please select a Shared Mapping Collection ▼

**Status\***

Active ▼

**Data Standard** 

DataStandard4 ▼

**Controlled Terminology** 

ControlledTerminology3 ▼

+ –

# Study Mapping Process



## 2. Data Sources –

- Associate Data Source(s) defined at Departments with study

### Available

custom\_formats  
dataSourceB  
dataSourceC  
empty  
qatest  
rwe\_1k  
sample

### Selected

dataSourceA  
NICSAH\_DATA

# Study Mapping Process

1 Metadata — 2 Data Sources — 3 Data Transform — 4 Tables — 5 Tables Transform — 6 Variables — 7 Variables Transform — 8 Export

## 3. Data Transform –

- Provide ability to read non-SAS datasets and convert them into SAS datasets.
- Live SAS Session with ability to view log and output datasets

Code Log Data Output



```
LIBNAME DS1 "/tla/warehouse/content/data_sources/dataSourceA" ;  
LIBNAME DS5 "/tla/warehouse/content/data_sources/NICSAH_DATA" ;  
LIBNAME DERIVED "/tla/warehouse/content/studies/study-1/derived" ;
```

```
1 proc export  
2   data=ds5.fube  
3   dbms=csv  
4   outfile="/tla/warehouse/content/data_sources/NICSAH_DATA/labs.csv"  
5   replace;  
6 run;  
7  
8 PROC IMPORT OUT= DERIVED.labs  
9   DATAFILE= "/tla/warehouse/content/data_sources/NICSAH_DATA/labs.csv"  
10  DBMS=CSV REPLACE ;  
11  GETNAMES=YES;  
12 run;  
13
```

Code Log Data Output



Number of Observations (Displayed): 100 Total columns: 13

Table: labs.sas7bdat

STUDYNO	RPTCTR	RPTHOSP	RPTINV	PTINIT	CTRREGNO	SAHDAT
11001	1	11	011A	DHW	1	10144
11002	1	11	011A	HL	2	10146
11003	1	11	011A	CG	4	10174
11004	1	11	011A	ABH	30	10195
11005	1	11	011A	VBL	31	10201
11006	1	11	011A	DTT	29	10208
11007	.	.	011A	LS	45	10224

# Study Mapping Process

1 Metadata — 2 Data Sources — 3 Data Transform — 4 Tables — 5 Tables Transform — 6 Variables — 7 Variables Transform — 8 Export

## 4. Tables —

- Provide ability to map Source datasets to Destination Domains.
- List all associated data sources
- Mapping count is maintained.

Source: View All

Please select a Data Source

View All

dataSourceA

Derived

NICSAH\_DATA

labs  
Derived 1

admin2  
NICSAH\_DATA 0

adverse  
NICSAH\_DATA 0

angio  
NICSAH\_DATA 0

Destination Tables

Filter destination tables

EG  
ECG Test Results

EX  
Exposure

IE  
Inclusion/Exclusion Criteria Not Met

LB  
Laboratory Tests Results

labs  
Derived

fube  
NICSAH\_DATA

# Study Mapping Process

1 Metadata — 2 Data Sources — 3 Data Transform — 4 Tables — 5 Tables Transform — 6 Variables — 7 Variables Transform — 8 Export

## 5. Tables Transform –

- Provide ability to realign source datasets and derive datasets for better alignment for variable mapping.
- Live SAS Session with ability to view log and output datasets

Standard Table: LB

Code Log Data Output



```
LIBNAME DS1 "/tla/warehouse/content/data_sources/dataSourceA" ;
LIBNAME DS5 "/tla/warehouse/content/data_sources/NICSAH_DATA" ;
LIBNAME DERIVED "/tla/warehouse/content/studies/study-1/derived" ACCESS=READONLY ;
LIBNAME MAPPING "/tla/warehouse/content/studies/study-1/mapping" ;
```

```
1
2 proc sort data=ds5.fube out=work.fube;
3   by studyno;
4 run;
5
6 proc sort data=derived.labs out=work.labs (drop=RPTCTR RPTHOSP FUBEDISA);
7   by studyno;
8 run;
9
10 data MAPPING.LB;
11   merge fube(in=A) labs(in=B);
12   by studyno;
13 run;
14
```

Standard Table: LB

Code Log Data Output



t

Number of Observations (Displayed): 100 Total columns: 13

JMBER	REPORTING CENTER	REPORTING HOSPITAL	REPORTING INVESTIGATOR	PATIENT INITIALS
01	011	011A	DHW	
01	011	011A	HL	
01	011	011A	CG	
01	011	011A	ABH	

# Study Mapping Process

1 Metadata — 2 Data Sources — 3 Data Transform — 4 Tables — 5 Tables Transform — 6 Variables — 7 Variables Transform — 8 Export








## 6. Variables —

- Provide ability to map Source variables to Destination Domain variables.
- List all associated data source tables variables per domain
- Mapping count is maintained

Standard Table: LB

















Source Variables

Filter source variables

 STUDYNO PATIENT STUDY NUMBER	1
 RPTCTR REPORTING CENTER	1
 RPTHOSP REPORTING HOSPITAL	0
 RPTINV REPORTING INVESTIGATOR	0
 PTINIT PATIENT INITIALS	0
 CTRREGNO CENTRAL REGISTRY NUMBER	0
 SAHDAT SAH DATE	0

Destination Variables

Filter destination variables

 LBORRESU Original Units	
 LBSTRESU Standard Units	
 LBTPNUM Planned Time Point Number	
 STUDYID Study Identifier	
 USUBJID Unique Subject Identifier	
 STUDYNO PATIENT STUDY NUMBER	
 RPTCTR REPORTING CENTER	
 VISITNUM Visit Number	



# Study Mapping Process

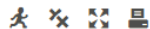
1 Metadata — 2 Data Sources — 3 Data Transform — 4 Tables — 5 Tables Transform — 6 Variables — 7 Variables Transform — 8 Export

## 7. Variables Transform –

- Provide ability to add any derivations/macro calls for Domain variables.
- Live SAS Session with ability to view log and final output Domain

Standard Table: LB

Code Log Data Output



```
LIBNAME MAPPING "/tla/warehouse/content/studies/study-1/mapping" ACCESS=READONLY ;
LIBNAME OUTGOING "/tla/warehouse/content/studies/study-1/outgoing" ;
```

```
1 data outgoing.LB;
2 set mapping.LB ;
3
4 *** BEGIN: DOMAIN ***;
5 DOMAIN = 'LB';
6 *** END: DOMAIN ***;
7
8 *** BEGIN: STUDYID ***;
9 STUDYID = 'NICSAH';
10 *** END: STUDYID ***;
11
12 *** BEGIN: USUBJID ***;
13 USUBJID = CATS(STUDYNO,"-", RPTCTR);
14 *** END: USUBJID ***;
```

Standard Table: LB

Code Log Data Output



Number of Observations (Displayed): 100

Table: lb.sas7bdat

DOMAIN	STUDYID	USUBJID
LB	NICSAH	11001-01
LB	NICSAH	11002-01
LB	NICSAH	11003-01

# Study Mapping Process

1 Metadata — 2 Data Sources — 3 Data Transform — 4 Tables — 5 Tables Transform — 6 Variables — 7 Variables Transform — 8 Export

## 8. Export —

- Ability to export mapped domains.
- Ability to publish mapping to Library

[Publish Mappings to Library](#) [Download](#)

<input type="checkbox"/> IE : Inclusion/Exclusion Criteria Not Met
<input checked="" type="checkbox"/> LB : Laboratory Tests Results
<input type="checkbox"/> MH : Medical History
<input type="checkbox"/> PE : Physical Examination
<input type="checkbox"/> SC : Subject Characteristics
<input type="checkbox"/> SUPPDM : Supplemental Qualifiers for Demographics
<input type="checkbox"/> SUPPDS : Supplemental Qualifiers for Disposition
<input type="checkbox"/> SUPPEX : Supplemental Qualifiers for Exposure

# REST APIs

<b>sasSubmissions</b> SAS submissions		▼
GET	/sas/submissions/{submissionId}	Gets a SAS submission based on ID
DELETE	/sas/submissions/{submissionId}	Cancels a SAS submission based on ID
POST	/sas/submissions/executeCode	Executes provided SAS code and creates a SAS submission
<b>studyMappings</b>		▼
GET	/departments/{departmentId}/studies/{studyId}/mappings	Get the study mappings
PUT	/departments/{departmentId}/studies/{studyId}/mappings	Updates the study mapping by ID
<b>Models</b>		▼
Credentials ▼ {		
username	string	the username to authenticate with
password	string(\$password)	the password to authenticate with
}		
TokenResponse ▼ {		
version	integer	The API resource version

# Study Mapping Process (Under Development)



## 4. Tables AutoMap –




- Provide ability to Automap Source datasets to Destination Domains based on previous mapping information available in published libraries



Help Automap Reset Save Next




Source: View All

Filter source tables

 cars dataSourceA	0
 labs Derived	1
 admin2 NICSAH_DATA	0

Destination Tables

Filter destination tables

 AE Adverse Events	⬆
 CM Concomitant Medications	⬆
 DM Demographics	⬆

# Study Mapping Process (Under Development)

1 Metadata — 2 Data Sources — 3 Data Transform — 4 Tables — 5 Tables Transform — 6 Variables — 7 Variables Transform — 8 Export

## 4. Tables AutoMap –

- Smart Mapping function is under development for improvement & enhancement.
- The following screen shot was from previous version of Data Mapper.

Suggested Mappings						
Domain (Raw)	Domain (To)	Variable	Variable Type	Confirm Suggested Mappings	Usage	Ignore
ADVERSE	AE	AEOUT	Num	<input type="checkbox"/> AE.AEOUT (Char)	AEOUT maps to AE.AEOUT 1 time in 1 study	<input type="checkbox"/>
ADVERSE	AE	TRTGRP	Char			<input type="checkbox"/>
				<input type="button" value="Update"/>	<input type="button" value="Cancel"/>	

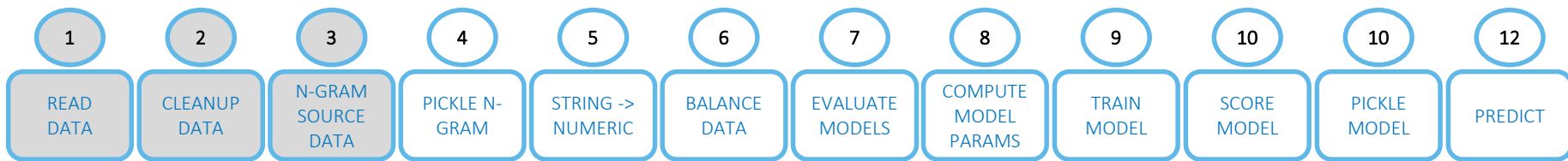
# Data Mapper

## Mapping Types

MAPPING TYPE	DESCRIPTION	SOURCE			DESTINATION		
TABLES	1-1 MATCH (Top Similarity Value)	DATASET			DOMAIN		
VARIABLE (AUTOMAP)	1-1 MATCH (Top Similarity Value)	DATASET	VARIABLE		DOMAIN	VARIABLE	
VARIABLE (SMART MAP)	1-3 MATCH (>0.25 Similarity)	DATASET	VARIABLE		DOMAIN	VARIABLE	
CONTROLLED TERMS	1-1 MATCH (Top Similarity Value)	DATASET	VARIABLE	VALUE	CONTROLLED TERM		VALUE

# Data Mapper

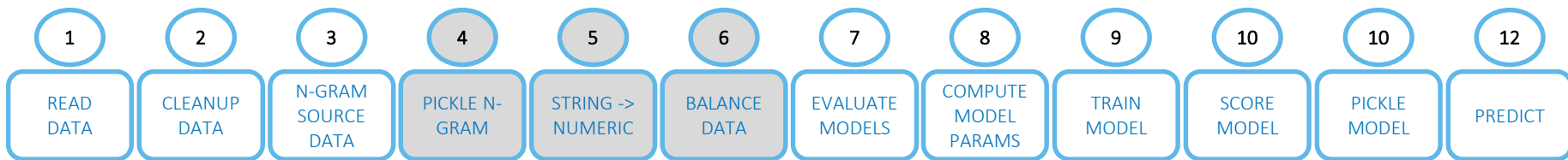
## Design Flow



STEP	DESCRIPTION	MODULES	PROCESS
1	Read Data	Pandas Dataframe	Read data into Pandas Dataframe – A tabular data structure in Python
2	Cleanup Data	Pandas Dataframe	Drop NaN Records, convert to lower case, combine columns, filter categories based on MIN count.
3	n-gram Data	NGRAM	n-gram data using character n-gram with limit of 2. This gives us more data to work with when comparing similarity in variable names.

# Data Mapper

## Design Flow

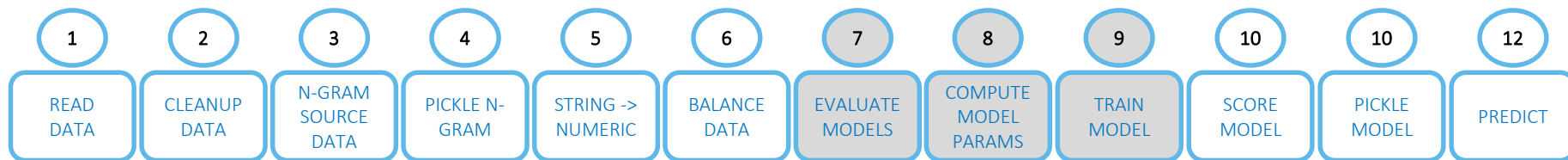


STEP	DESCRIPTION	MODULES	PROCESS
4	Pickle n-gram	Pickle	n-gram pair of (Source, Destination) value and pickle it. The pickle file is the python data object after serialization, and is an efficient way to store and use any python object data.
5	String to Numeric Value conversion	TfidfVectorizer, MinMaxScaler	Generate Dictionary with TF-IDF values and re-scale it using MinMaxScaler. For machine learning to work, the values must be numeric between 0 and 1.
6	Balance Data	SMOTETomek	Balance data to handle under-sampling and over-sampling. If the data we have is skewing our results, we want to normalize closer to 0.5.



# Data Mapper

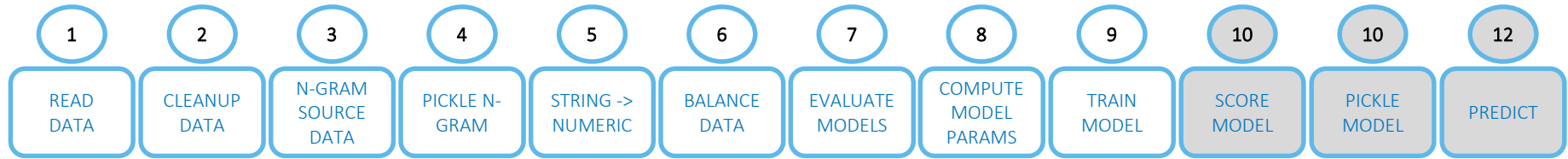
## Design Flow



STEP	DESCRIPTION	MODULES	PROCESS
7	Evaluate Models	sklearn	Plenty of different models to evaluate: LogisticRegression, DecisionTreeClassifier, RandomForestClassifier, GaussianNB, MultinomialNB, OneVsRestClassifier (LinearSVC), OneVsRestClassifier (SGDClassifier)
8	Compute Model Params	Sklearn, GridSearchCV	GridSearchCV attempts multiple combinations on the model parameters, tweaking them to find the best parameter values to use.
9	Train Model	sklearn	Fit model.

# Data Mapper

## Design Flow



STEP	DESCRIPTION	MODULES	PROCESS
10	Score Model	sklearn.metrics	Classification_report – Precision, Recall, F1 Score. This tells us how much data was relevant to the suggested mappings, and gives a value on how accurate the model thinks the mapping is.
11	Pickle Model	Pickle	Save the trained model in pickle file.
12	Predict	sklearn	Predict based on trained model. With user interaction with the predicted mapping (Supervised Machine Learning), the model will become smarter and more accurate over time.



[sas.com](https://sas.com)