

A wide banner featuring a panoramic view of the Shanghai skyline at dusk or dawn. The Oriental Pearl Tower is the central focus, with other skyscrapers visible in the background. The sky is a mix of light blue and orange.

**2024** CDISC CHINA INTERCHANGE **SHANGHAI**

30-31 AUGUST: CONFERENCE & EXPO | 28-29 AUGUST: TRAININGS

**aCRF and SDTM Integrated Automation Tool based on NLP + ML**

Shijia Wang, Senior Manager, Data Science, Tigermed-Macrostat



## Meet the Speaker

Shijia Wang

**Title:** Senior Manager

**Organization:** Tigermed-Macrostat

Shijia has 9 years of experience in the pharmaceutical industry, specializing in clinical programming. He is proficient in programming and tool development, with expertise in SAS, Python, and VBA to boost work efficiency and quality.



# Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*
- *The author(s) have no real or apparent conflicts of interest to report.*



# Agenda

1. Background
2. Introduction of a New aCRF Automation Tool
3. Demo



**Background**

# What is aCRF and Why We Need aCRF

## NMPA, 药物临床试验数据递交指导原则(2020):

- 临床试验数据相关的申报资料通常包括数据库及其相应的数据说明文件、数据审阅说明、程序代码和注释病例报告表(annotated Case Report Form, aCRF)

## FDA, Study Data Technical Conformance Guide (2024):

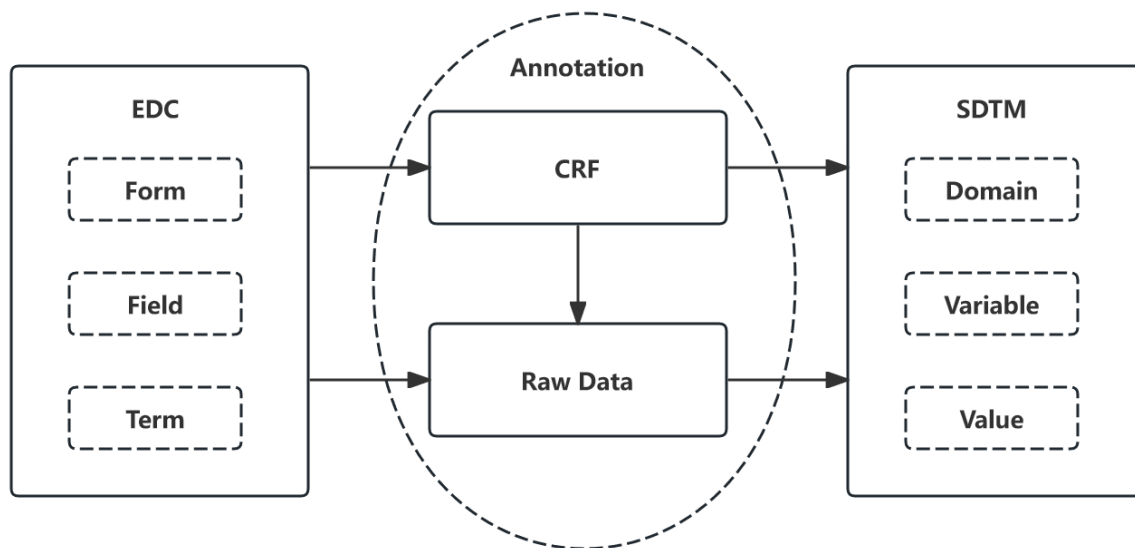
- aCRF should be submitted preferably at the time a protocol is submitted.
- The aCRF should be provided as a PDF with the file name 'acrf.pdf.'

## PMDA, Technical Conformance Guide (2024):

- The file format of the Annotated CRF, in principle, should be a PDF, as specified in the eCTD v4 notification, and the file name should be 'acrf.pdf'. In principle, it should be stored in the same folder as SDTM datasets

# What is aCRF and Why We Need aCRF

aCRF is The Fundamental of SDTM Process





# Current Method of Annotation

- **Manual Work**
  - Time-consuming and Labor-intensive
- **VBA/SAS + FDF/XFDF**
  - Semi-Automation
- **Is there a simpler, more straightforward, and what-you-see-is-what-you-get (WYSIWYG) way to generate aCRF?**





# Introduction of a New aCRF Automation Tool

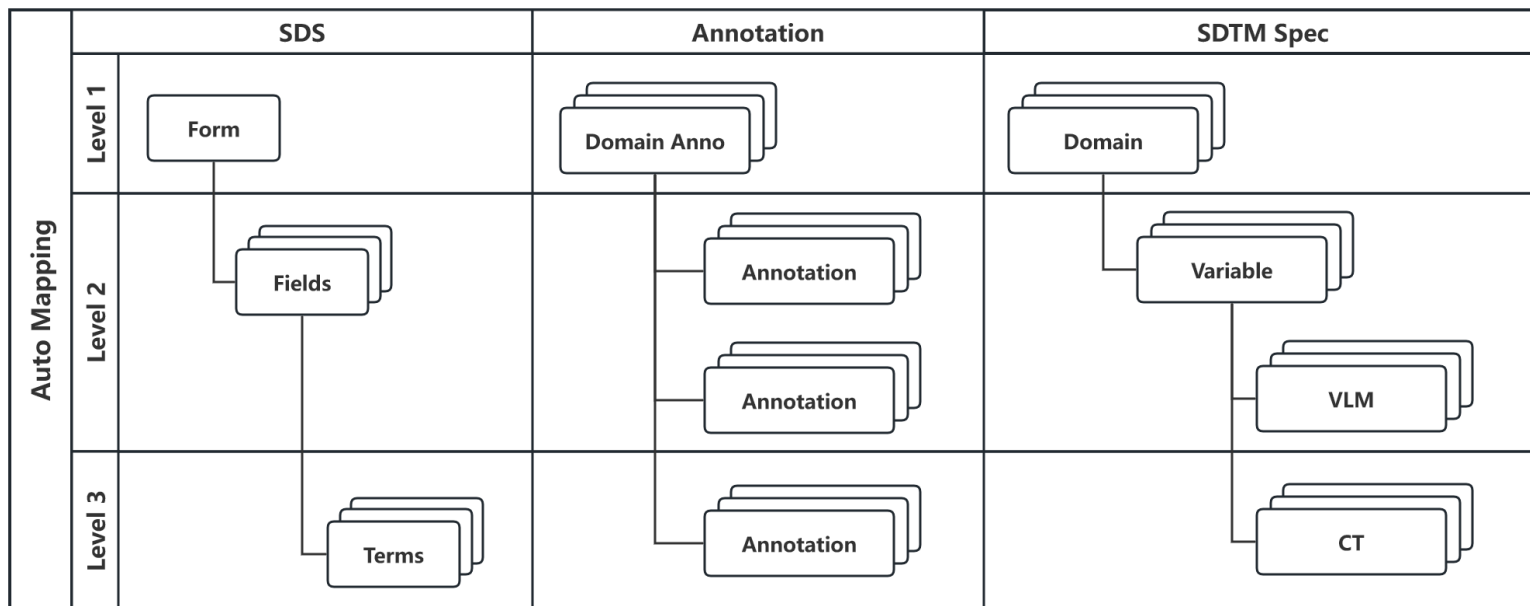


# Target of New Tool

- **Automated Mapping**
  - Automatically Map EDC and SDTM
  - High Confidence Recommendation through NLP + ML
- **Dynamic Annotation Generation**
  - WYSIWYG Generates Annotations
  - Contents + Format + Position + Page
- **Streamlined Workflow**
  - Within Project: Annotation + SDTM Spec + SDTM Programming
  - Beyond Project: Experience Share and Use

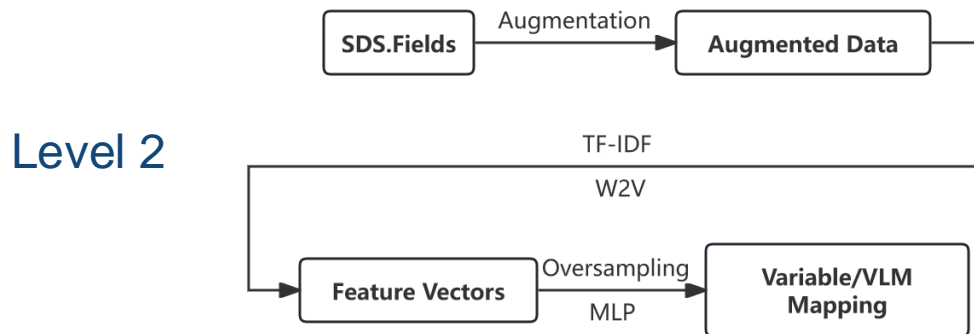
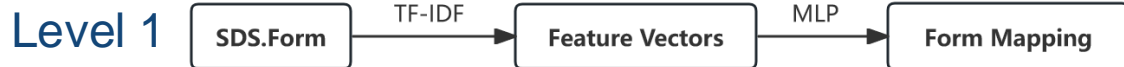
# Automated Mapping

- A bridge from EDC to SDTM: SDS (Study Data Specification)
- Relationship Between SDS, Annotation and SDTM spec



# Automated Mapping

- NLP and ML models developed for Level 1 and Level 2 Mapping



Precision	Recall	F1 Score
0.94	0.95	0.94
0.93	0.92	0.91



# Dynamic Annotation Generation

- Annotation Types

Type	Examples
Variable	[ <i>Variable</i> ]
Variable Value	[ <i>Variable</i> ] = [ <i>Value</i> ]
Result with Condition	[ <i>Variable</i> ] when [ <i>Condition Variable</i> ] = [ <i>Condition Value</i> ]
Result value with Condition	[ <i>Variable</i> ] = [ <i>Value</i> ] when [ <i>Condition Variable</i> ] = [ <i>Condition Value</i> ]
SUPP	[ <i>Variable</i> ] in [ <i>SUPP--</i> ]
RELREC	RELREC when [ <i>Variable</i> ] = [ <i>Condition Domain</i> ]. [ <i>Condition Variable</i> ]
Comments	[ <i>Comments</i> ]

# Dynamic Annotation Generation

- What-You-See-Is-What-You-Get
- Contents + Format + Page + Position
- Operation in Web Platform instead of PDF reader/editor

The screenshot displays the aCRF 自动化工具 (aCRF Automation Tool) web interface. The main window shows a form titled "DM (Demographics)" with various fields and their corresponding data types. The form includes fields for Birth Date (BRTHDAT), Age (AGE), Age Units (AGEU), Sex (SEX), Ethnicity (ETHNIC), and Race (RACE). The right-hand panel shows a mapping table with columns for Reference, New, Clear, Mapping History, and Save. The table lists the following mappings:

Reference	New	Clear	Mapping History	Save
<input type="checkbox"/>	No	Coding	Description	Operation
<input type="checkbox"/>	1	BRTHDAT	Birth Date	

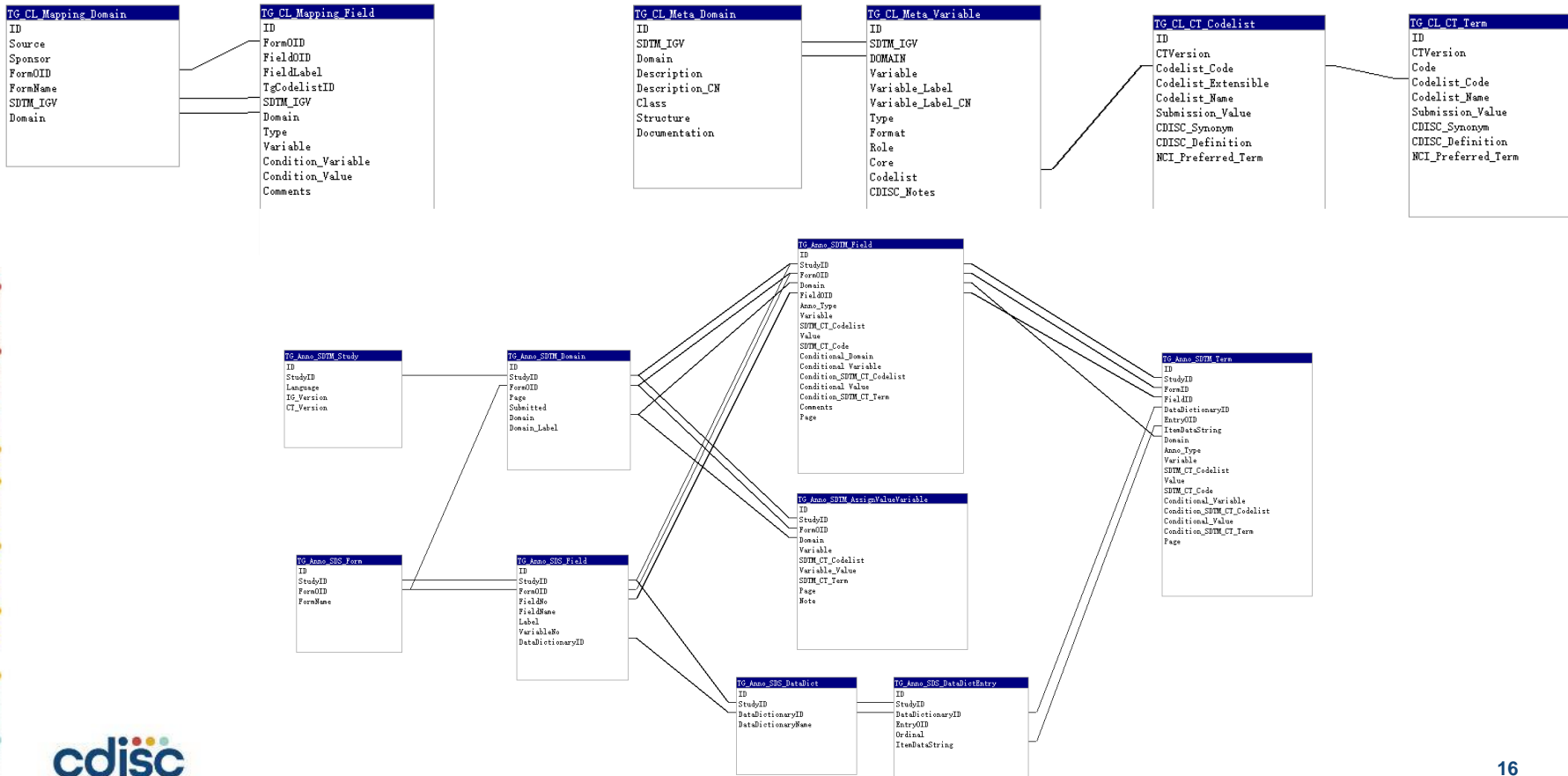
  

Domain	Type	Variable	Value	Conditional Domain	Conditional Variable
DM	ML	Variable	BRTHDT1		

<input type="checkbox"/>	>	2	AGE	Age	
<input type="checkbox"/>	>	3	AGEU	Age Units	
<input type="checkbox"/>	>	4	SEX	Sex	
<input type="checkbox"/>	>	5	CHILD	If Female, does the su...	
<input type="checkbox"/>	>	6	ETHNIC	Ethnicity	
<input type="checkbox"/>	>	7	RACE	Race	

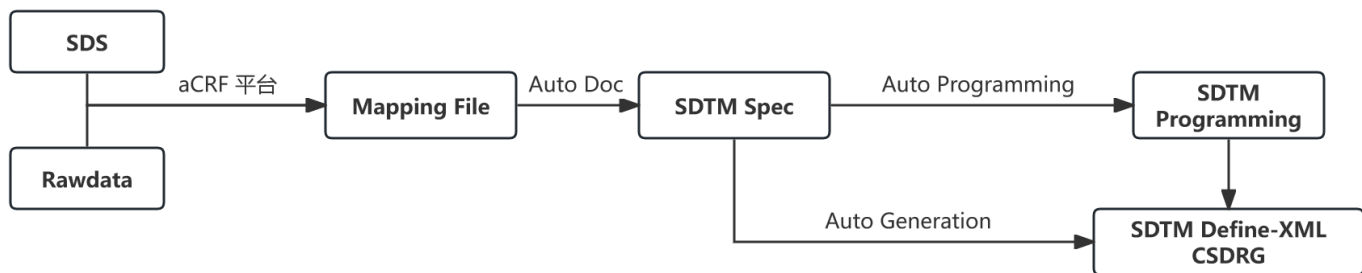
# Dynamic Annotation Generation





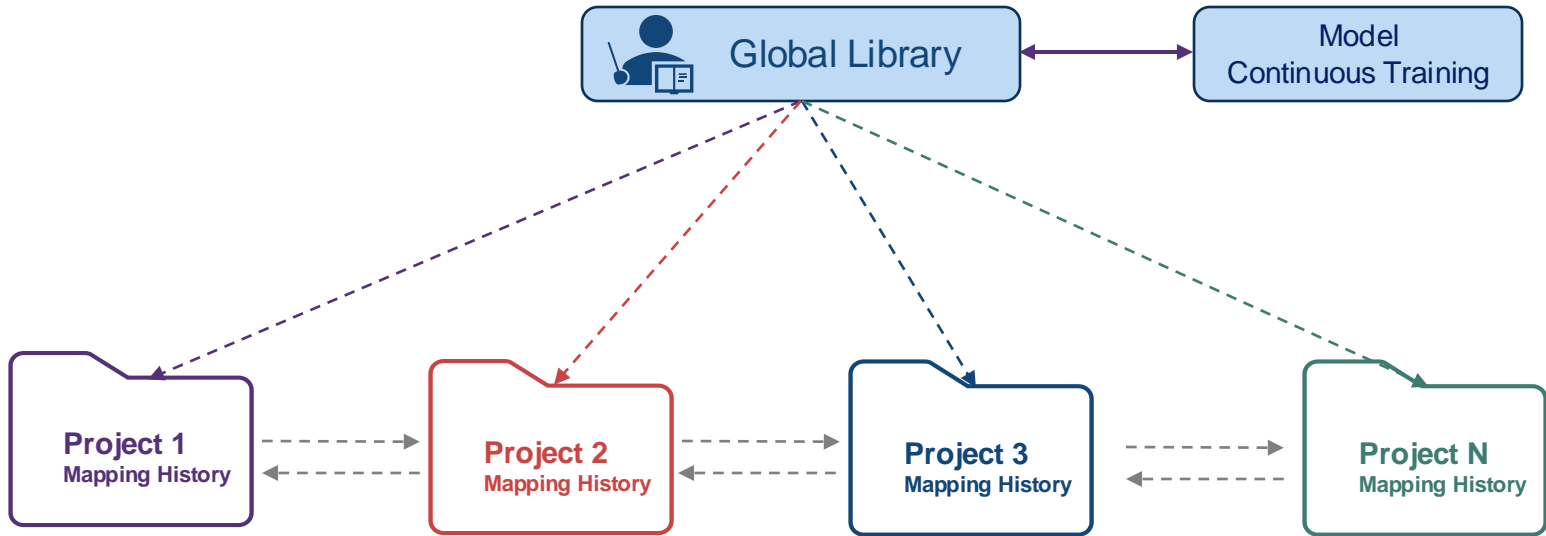
# Streamlined Workflow

- Within a Project:



# Streamlined Workflow

- Beyond Project:

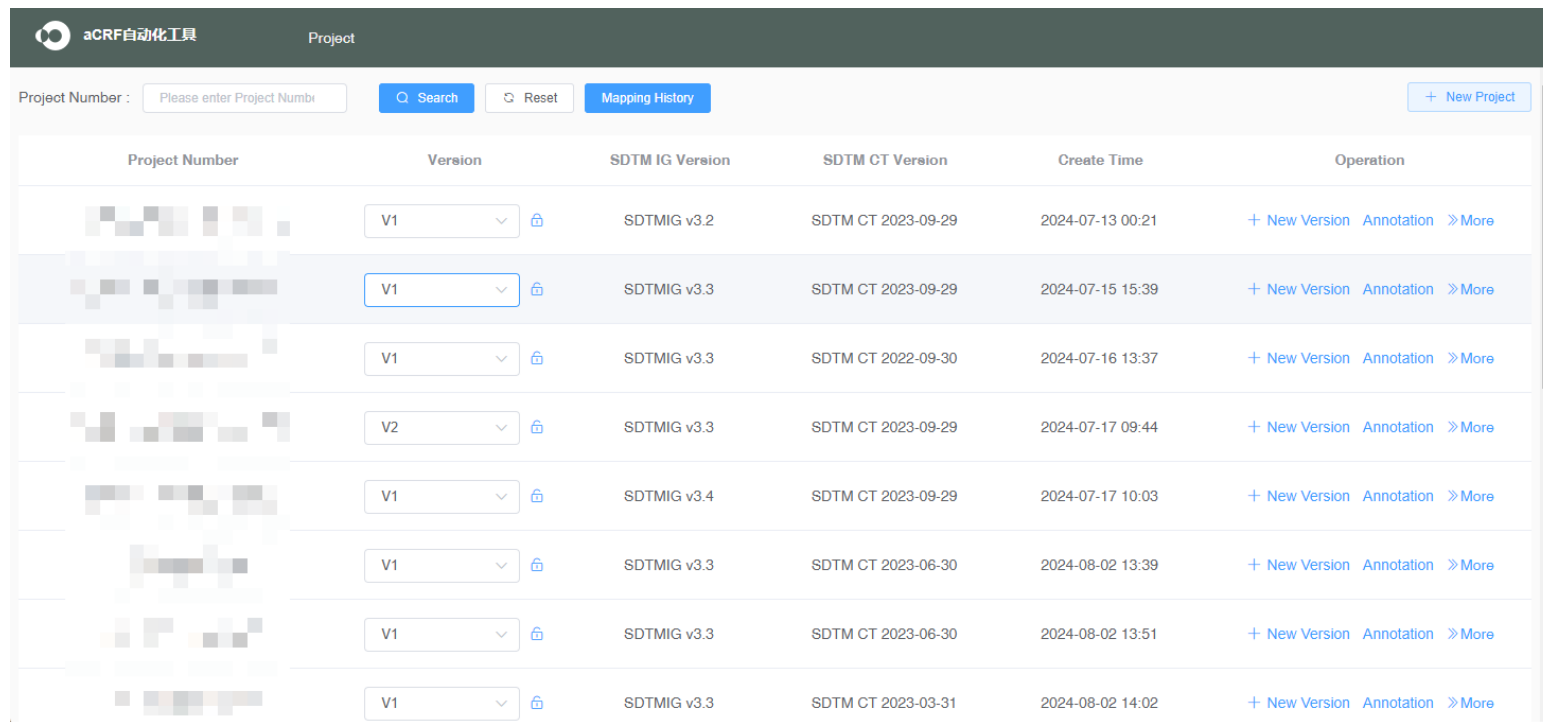




Demo

# aCRF Automation Tool

- Project List:



The screenshot displays the 'Project List' interface of the aCRF Automation Tool. The header includes the tool's name 'aCRF自动化工具' and the section 'Project'. Below the header, there is a search bar with the placeholder 'Please enter Project Number', a 'Search' button, a 'Reset' button, and a 'Mapping History' button. A '+ New Project' button is located in the top right corner. The main content is a table with the following columns: Project Number, Version, SDTM IG Version, SDTM CT Version, Create Time, and Operation. The table contains eight rows of project data.

Project Number	Version	SDTM IG Version	SDTM CT Version	Create Time	Operation
[Blurred]	V1	SDTMIG v3.2	SDTM CT 2023-09-29	2024-07-13 00:21	+ New Version Annotation > More
[Blurred]	V1	SDTMIG v3.3	SDTM CT 2023-09-29	2024-07-15 15:39	+ New Version Annotation > More
[Blurred]	V1	SDTMIG v3.3	SDTM CT 2022-09-30	2024-07-16 13:37	+ New Version Annotation > More
[Blurred]	V2	SDTMIG v3.3	SDTM CT 2023-09-29	2024-07-17 09:44	+ New Version Annotation > More
[Blurred]	V1	SDTMIG v3.4	SDTM CT 2023-09-29	2024-07-17 10:03	+ New Version Annotation > More
[Blurred]	V1	SDTMIG v3.3	SDTM CT 2023-06-30	2024-08-02 13:39	+ New Version Annotation > More
[Blurred]	V1	SDTMIG v3.3	SDTM CT 2023-06-30	2024-08-02 13:51	+ New Version Annotation > More
[Blurred]	V1	SDTMIG v3.3	SDTM CT 2023-03-31	2024-08-02 14:02	+ New Version Annotation > More

# aCRF Automation Tool

- Form - Domain Mapping:

The screenshot displays the aCRF Automation Tool interface for configuring a form domain. The top navigation bar shows the tool name and the current project/form details: "aCRF自动化工具", "Project / Form", "Form: DM", "Form Name: Demographic Information", "PageNumber: 5", and "Is Annotation: Yes". A "+ Field" button is visible in the top right.

The main content area is split into two panels. The left panel shows a preview of the form titled "DM (Demographics)". It includes fields for "Birth Date", "Age", "Age Units" (with a "YEARS" label), "Sex" (with radio buttons for Male, Female, Unknown, Undifferentiated), "If Female, does the subject have childbearing potential?" (with radio buttons for Yes, No), "Ethnicity" (with radio buttons for Hispanic or Latino, Not Hispanic or Latino, Not reported, Unknown), and "Race" (with radio buttons for American Indian or Alaska Native).

The right panel is titled "Need Submit" and has radio buttons for "Yes" (selected), "No", and "NA", along with a "Save" button. Below this are two tabs: "Domain" (selected) and "Mapping History".

The "Domain" tab contains a table with the following data:

No	Domain	Description	Operation +
1	DM	Demographics	-

Below the domain table is the "Assigned Value Variables" section, which contains a table with the following data:

No	Domain	Variable	Value	Operation +
1	Please select			-

# aCRF Automation Tool

- Field – Variable Mapping:

The screenshot displays the 'aCRF 自动化工具' (aCRF Automation Tool) interface. The main window shows a form titled 'DM (Demographics)' with fields for 'Birth Date', 'Age', 'Age Units', 'Sex', and 'Race'. Each field is mapped to a variable name: 'Birth Date' to 'BRTHDTC', 'Age' to 'AGE', 'Age Units' to 'AGEU', 'Sex' to 'SEX', and 'Race' to 'RACE'. A red arrow points to the 'Race' field label.

On the right, a 'Mapping History' table lists the mappings. The table has columns for 'Reference', 'New', 'Clear', 'Mapping History', and 'Save'. Below these are 'Need Submit' options (YES/NO) and a table with columns: 'Domain', 'Type', 'Variable', 'Value', 'Conditional Domain', and 'Conditional Variable'.

Reference	New	Clear	Mapping History	Save	
<input type="checkbox"/>	No		Coding	Description	Operation
<input type="checkbox"/>	1		BRTHDAT	Birth Date	
Need Submit <input checked="" type="radio"/> YES <input type="radio"/> NO					
Domain	Type	Variable	Value	Conditional Domain	Conditional Variable
DM	Variable	BRTHDT	YEARS		
<input type="checkbox"/>	> 2	AGE	Age		
<input type="checkbox"/>	> 3	AGEU	Age Units		
<input type="checkbox"/>	> 4	SEX	Sex		
<input type="checkbox"/>	> 5	CHILD	If Female, does the su...		
<input type="checkbox"/>	> 6	ETHNIC	Ethnicity		
<input type="checkbox"/>	> 7	RACE	Race		

# aCRF Automation Tool

- Mapping History:

The screenshot displays the 'aCRF自动化工具' (aCRF Automation Tool) interface. The top navigation bar includes 'Project / Form / Field' and a search bar with '5' and '113' results. A 'Mapping History' button is visible in the top right. The main content area is titled 'Mapping History' and features a table with columns: Domain, Assign Value Variables, Field, and Term. The 'Field' column is currently selected. The table lists several entries with columns for Project Number, Version, FormOID, FormName, FieldOID, and FieldLabel. A 'Quote' tooltip is shown over the first row's Project Number. On the right side, a 'Conditional Variable' panel is partially visible.

Domain	Assign Value Variables	Field	Term
		Quote	
	Project Number Q	Version Q	FormOID Q
	FormName Q	FieldOID Q	FieldLabel Q
		1	DM
		1	DM
		1	DM
		2	DM
		1	DM
		2	DM
		1	DM



# aCRF Automation Tool

- **Efficiency Improvement**

aCRF: **25.15%** increase from baseline

Entire SDTM Process: **21.10%** increase from baseline





**Thank You!**

**cdisc**