

Single Day Event

<AI-powered Data Transformation>

Zhiheng Chen,
Boehringer Ingelheim



Disclaimer

The views and opinions expressed in the following PowerPoint slides are those of the individual presenter and should not be attributed to PHUSE, its directors, officers, employees, volunteers, members, or any organization with which the presenter is employed or affiliated.

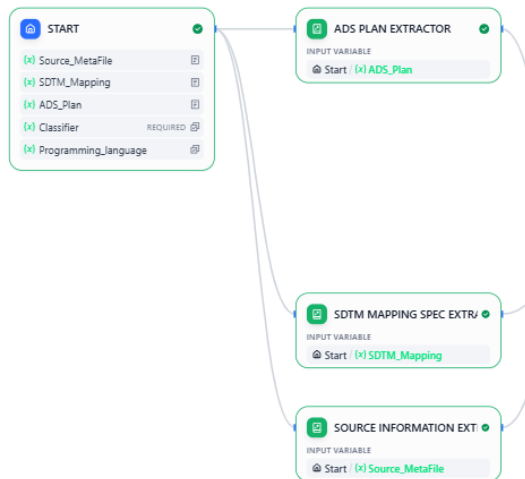


AI-powered Data Transformation





DataNexus



Information Extractor

Add description...

SETTINGS LAST RUN

(x) Set variable

SYSTEM

2702 Jinja (x)

You are given three meta information representing the structure and derivation logic of clinical data sets: Source, SDTM, and ADaM. Each meta information may contain multiple tabs, with each tab corresponding to a single dataset. Each tab contains multiple columns describing variable attributes (e.g., variable name, variable length, variable label, etc.). From the previous step, these three files have been extracted. The source information is (x) text . The SDTM information is (x) text . And the ADaM information is (x) text . Since there are dependency among these metafiles, please read through ADaM first, and then SDTM, and then source. When you read ADaM file, you will find the derivation of variable may based on SDTM variable. The names of ADaM dataset all start with AD. And the names of SDTM dataset are usually two levels, such as DM, AE, etc.

Key rules and context:

Data Derivation Hierarchy:

ADaM datasets are derived from SDTM datasets.

SDTM datasets are derived from Source datasets.

Derivation Logic:

Each variable has a 'Derivation / Source' column specifying how it is derived.

In derivations, references to other variables use the format XX.YYYY, where XX is the dataset name and YYYY is the variable name (e.g., DM.AGE refers to the AGE variable in the DM dataset).

Variable Attributes:

Each variable may have additional attributes such as length, label, type, etc.

Your tasks:

Summarize Relationships:

Summarize the relationships among variables across Source, SDTM, and ADaM based on the derivation logic.

Question Answering:

When the user asks a question related to these meta files, provide a concise and accurate summary based on the information from the meta files. Here is the user's question (x) sys.query

Code Generation:

For each dataset, generate code (e.g., in SAS or Python/pandas) that implements the derivation logic as specified in the 'Derivation / Source' column, taking into account the variable attributes and references to other datasets/variables.

Input:

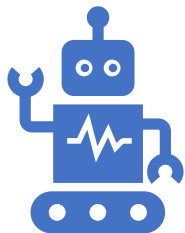
You will receive aggregated meta information from all three files.

Output:

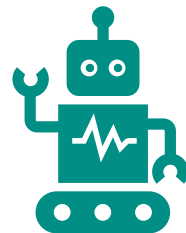
A summary of variable relationships.

+ Add Message

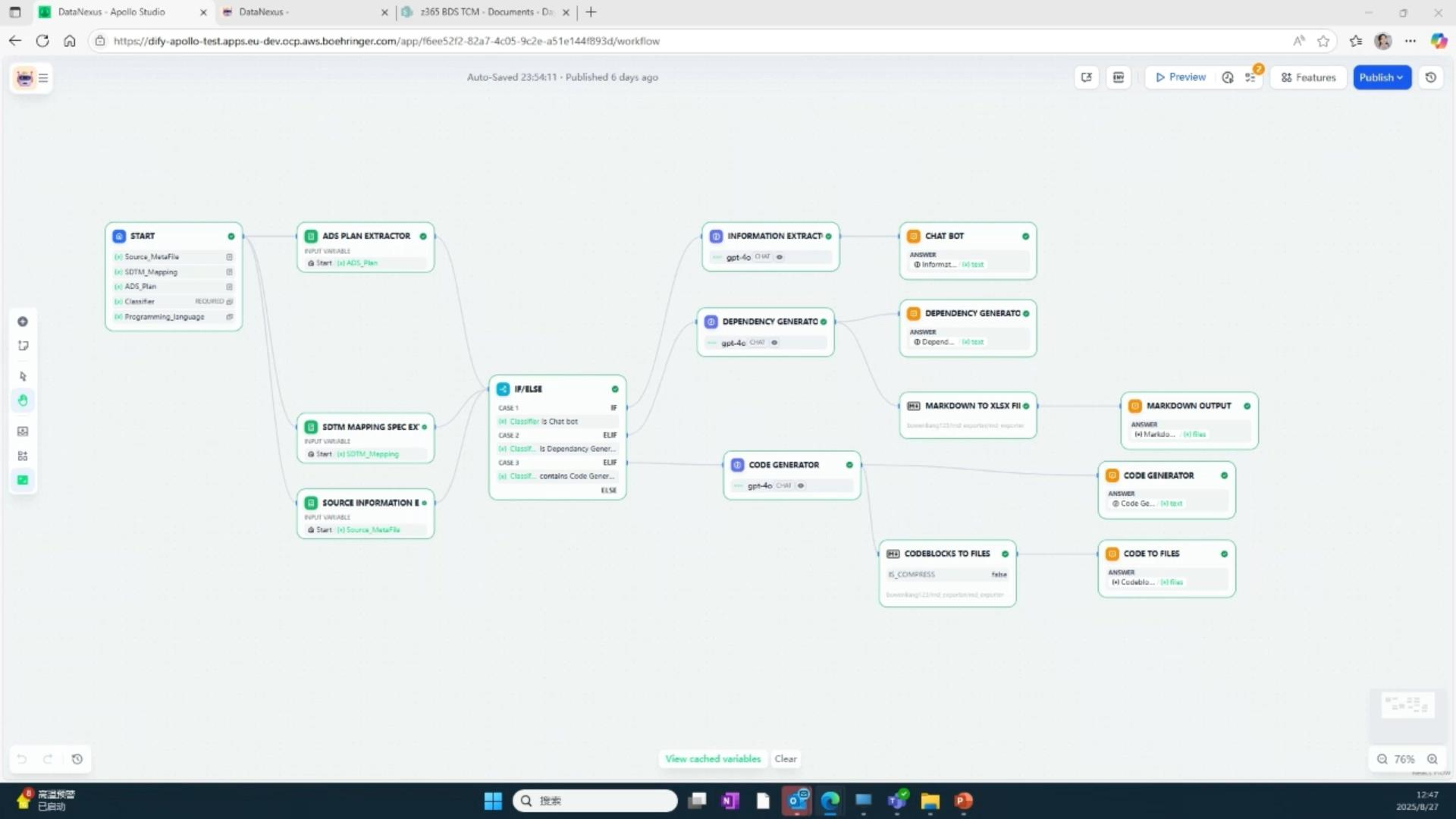
Demo



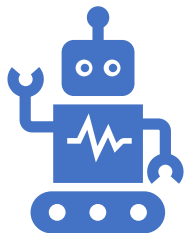
DataNexus 



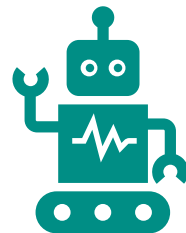
Dependency Lab 



Demo



DataNexus



Dependency Lab



Dependency Network Visualization

Upload
Dependency File

Choose Excel File

Browse...

No file

Upload an Excel file
with two columns:

- Column 1: Variable names
- Column 2: Dependencies (comma-separated)

Welcome to Dependency Network Visualizer!

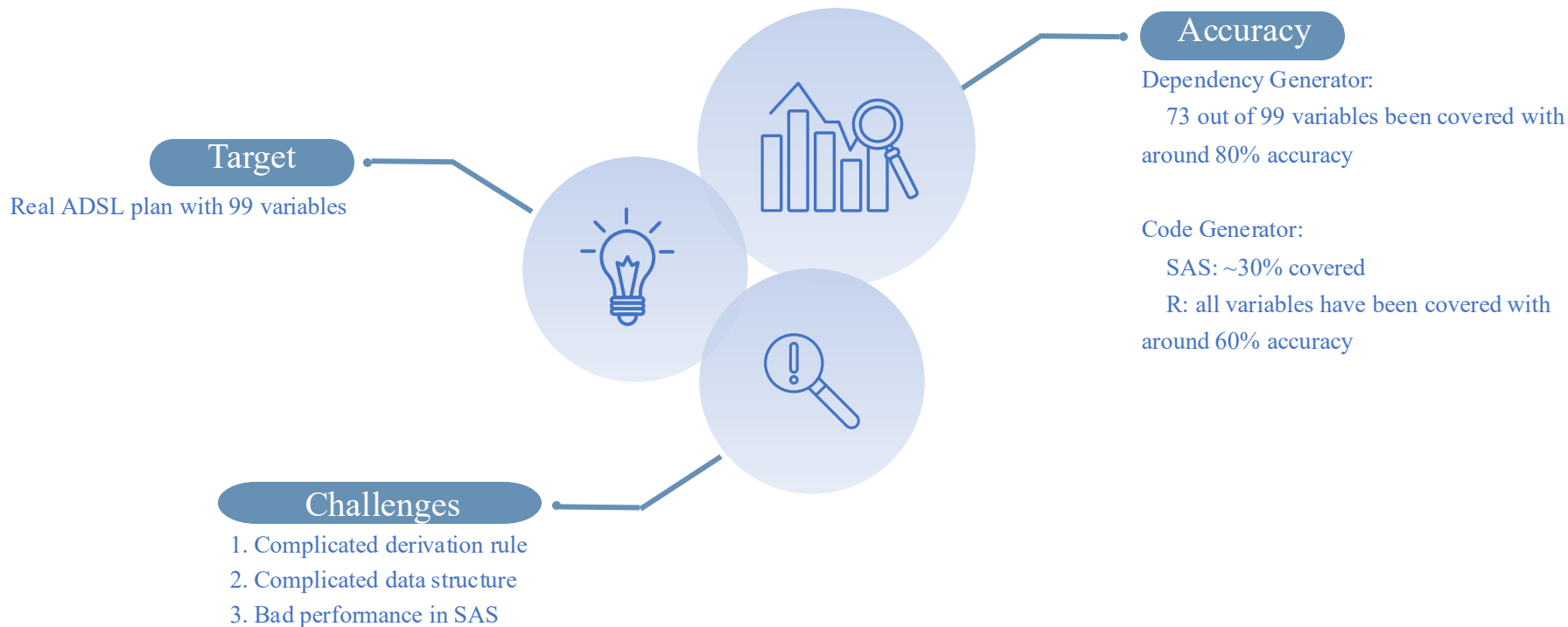
Upload your Excel file to visualize the dependency relationships.

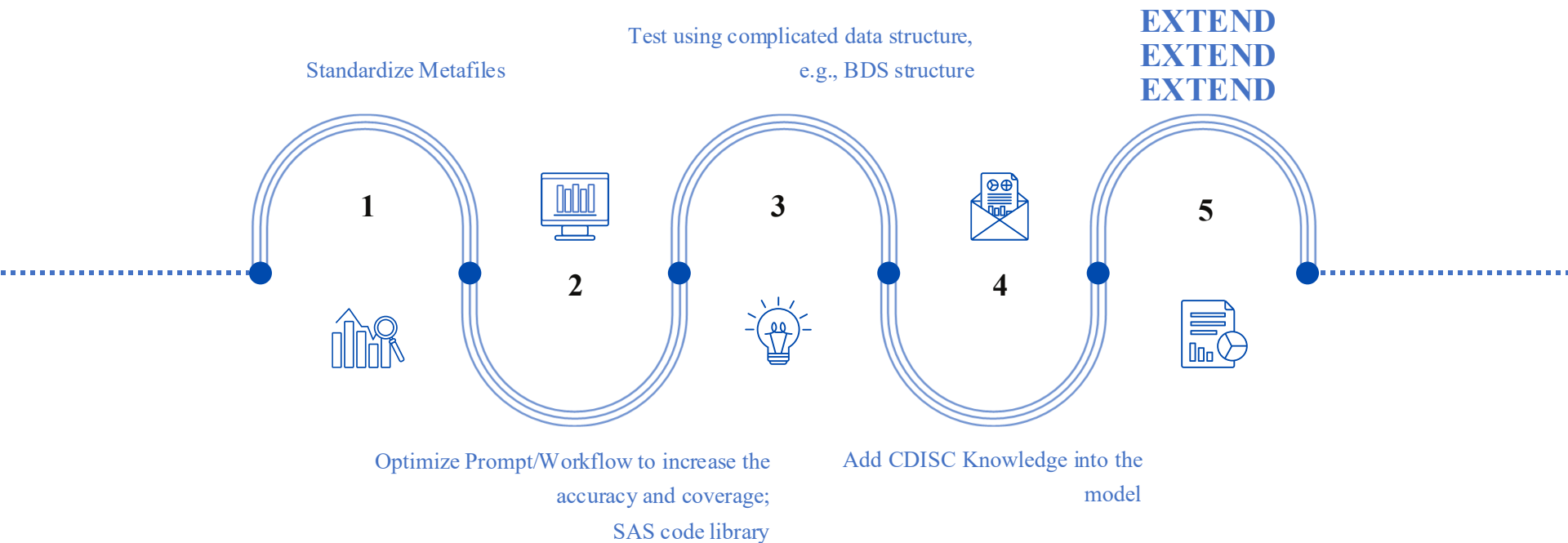
Upload File

Please upload a file to
get started

Please upload a file to
get started

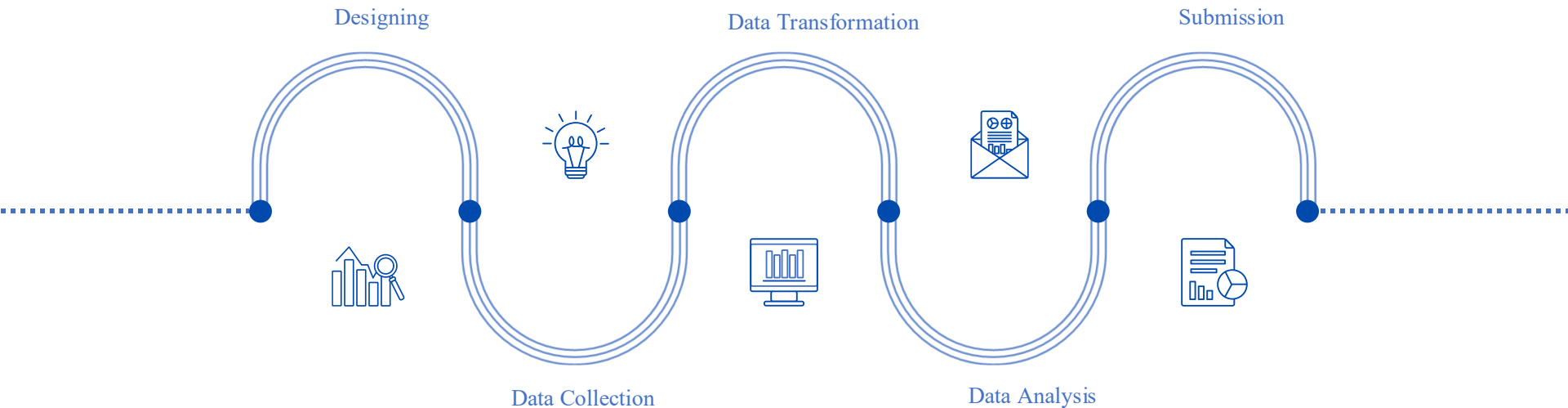
Testing Report







Metadata Driven Extension



Thank You