Leveraging metadata when mapping to CDISC Standards with SAS® machine learning in a Results-as-a-Service Model (RAAS)

Preetesh V Parikh, SAS Institute Inc., Cary, North Carolina, USA
Sandeep Juneja, SAS Institute Inc., Cary, North Carolina, USA
Ben Bocchicchio, SAS Institute Inc., Cary, North Carolina, USA

ABSTRACT

Mapping raw Electronic Data Capture (EDC) data to a Study Data Tabulation Model (SDTM) or Analysis Data Model (ADaM) standard is a very time consuming process. As the evolution of standards continues to develop, the ability to collect and reuse metadata on how raw variables get mapped into standardized variables can help reduce the time to standardize subsequent studies. The reduction in time and effort comes from the ability to reuse the mapping by analyzing the raw and standardized metadata.

Using the power of Machine Learning algorithms can further help derive and predict possible mappings. Instead of a customer having to install software locally and making the necessary configurations for it to run, SAS can offer this as a hosted solution. This means that customer can upload data and standards, run the process, and results of the process will be SAS code containing the mapping information that can be utilized in the customer environment.

INTRODUCTION

Mapping raw data to standards is one of the most challenging process in the healthcare industry. Reusing or reapplying the information collected during mapping processes from previously mapped studies and building upon that knowledge inference is the most important part of the mapping process. Most companies struggle with building knowledge inferences and reapplying them through the data process efficiently. In addition, sometimes, there are multiple tools/programs required to go through full-cycle of data mapping. Therefore, it is difficult for the standard user to know all the different versions and tools, and use them correctly throughout the data mapping process.

The SAS Data Mapping Tool is a web-based tool that provides a user-friendly interface for everything from mapping raw data to generating SDTM standards (including domain templates). Simple User Interface (UI) and click-away concept design provides access to all the required information on a single screen. Auto-mapping and smart-mapping features in the SAS Data Mapping Tool, which are based on knowledge inference derived from machine learning algorithms, reduce time and effort for the user. This leads to improvements in quality, efficiency and consistency.

Display 1. Data Mapping Process Flow
SAS DATA MAPPING TOOL’S SIX MAJOR COMPONENTS:

1. STANDARDS/CONTROLLED TERMINOLOGIES

To begin mapping, users need to register Standards/Controlled Terminologies (CTs) for use in mapping the study. Users can import or add Standard/Controlled Terminology to the tool and register it. This allows the user to build reports and provide comparisons among Standards/CTs. As more studies go through the mapping process, the tool allows users to see associated studies with the same Standards/CTs.

Display 2. Register Standards

2. STUDIES

Studies are added under departments. This keeps the study information restricted to within departments and controls user access to this information. Once the standards are defined for the study, users set up a new study by filling out the required study-related information. This includes the study start date, end date, phase, location of output template programs, and study setup file, etc. Real-time reports are available at every step of the process. A study reset option is also available. This allows users to clear content, but maintain the study structure.

Display 3. Study Setup
3. DATA

Users define references for data sources. Once the data sources are added, the tool can provide snapshots of the metadata associated with the raw data. This allows the user to familiarize themselves with the data before mapping. By default, the tool also shows the first 100 records from selected raw datasets. Users can restrict extraneous variables from displaying, while mapping, by using the Ignore functionality of the tool, thereby allowing them to focus on only the mapping variables.

4. MAPPING

Once the raw data is input into the tool, users link source datasets to destination domain mappings. It can be one-to-many or many-to-many mappings. There is also an option to ignore the raw data source completely if it is not used within the study. After identifying tables, users proceed with variable mapping. If a user is mapping the first study, then the process is manual. However, if at least one study is published, than the user can use the auto-mapping or smart-mapping tools. The auto-mapping tool allows users to use the knowledge inference from published libraries to populate tables mapping and variables mapping. If similarly sounding or spelled variables do not map through auto-mapping, users can run smart-mapping on variables to get suggestions based on variables used in other studies. Auto-mapping and smart mapping is only in effect once users have published at least one study into the library. If users pick more than one study to scan for auto-mapping or smart-mapping, the order that information is retrieved is based on ranking. Lower ranked studies get precedence over higher ranked study in terms of obtaining information through the scanning process. At minimum, users need to have the required variables with the mapping definitions. Otherwise, the tool turns those variables a RED color with the designated “Not Mapped” status.
5. SAS PROGRAM SHELLS

The SAS Data Mapping Tool allows users to generate SAS Program Shells. These shells contain the automated SAS code, based on the mapping information. It is a non-validated code, which may need further modification depending on the users’ needs.
Leveraging metadata when mapping to CDISC Standards with SAS® machine learning in a Results as a Service Model (RAAS), continued

6. PUBLISH LIBRARIES

Once a user is through the mapping process, he or she can publish mapping information to a library. As users go through more and more studies, these libraries grow and the user can leverage this information in making mapping decisions for future studies.

Throughout these mapping components, real-time reports are available from the variable-level to the standards level. With the help of these reports, users can stay well-informed about the study progress, allowing for improved resource allocation. The tool is not limited to only SDTM. It can support ADaM or any other industry or custom standards as well.

Results-As-A-Service (RaaS) Model:

The SAS Data Mapping Tool can be used by customers as a RAAS model in which the customer provides the data and specifications to SAS and SAS delivers the final mapping solutions using this tool.

CONCLUSION

The SAS Data Mapping Tool is an easy-to-use tool, which provides functionalities required to map raw data to defined standards all in one place. The ability to leverage metadata and reuse knowledge inference makes it a very powerful tool.

CONTACT INFORMATION

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

<table>
<thead>
<tr>
<th>Name</th>
<th>Company</th>
<th>Address</th>
<th>Phone</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preetesh V. Parikh</td>
<td>SAS Institute Inc.</td>
<td>100 SAS Campus Drive, Cary, NC</td>
<td>(919) 531-3737</td>
<td><a href="mailto:Preetesh.parikh@sas.com">Preetesh.parikh@sas.com</a></td>
</tr>
<tr>
<td>Sandeep Juneja</td>
<td>SAS Institute Inc.</td>
<td>100 SAS Campus Drive, Cary, NC</td>
<td>(919) 531-0541</td>
<td><a href="mailto:Sandeep.juneja@sas.com">Sandeep.juneja@sas.com</a></td>
</tr>
<tr>
<td>Ben Bocchicchio</td>
<td>SAS Institute Inc.</td>
<td>100 SAS Campus Drive, Cary, NC</td>
<td>(919) 531-3704</td>
<td><a href="mailto:Ben.Bocchicchio@sas.com">Ben.Bocchicchio@sas.com</a></td>
</tr>
</tbody>
</table>
Leveraging metadata when mapping to CDISC Standards with SAS® machine learning in a Results as a Service Model (RAAS), continued

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.