Paper SI01

Automate to Accelerate: Saama's Smart Auto Mapper

Mike Collinson, Saama Technologies, London, UK

AI-DRIVEN CLINICAL DATA TRANSFORMATIONS

Inspecting and mapping complex datasets to CDISC standards can be tedious and time-consuming. Saama's Smart Auto Mapper shifts the paradigm by using Artificial Intelligence to overcome standard data onboarding and transformation challenges. Smart Auto Mapper automatically generates data mappings using out-of-the-box and custom functions that transform data into submission and analytics-ready datasets.

Smart Auto Mapper leverages Machine Learning algorithms to automatically propose Dataset, Column and Functional mapping logic based on experience of other studies, standards, and by looking into the data values themselves. Accuracy is ensured with human-in-the-loop feedback controls to refine mappings at a study level and continually improve the algorithmic suggestions.

This presentation will demonstrate how Smart Auto Mapper uses Machine Learning to match non-standard (and often messy and incomplete) source data formats to globally managed target standards, describe how reviewers can approve and manage automated mapping suggestions, and explain how this feedback improves the quality of the automapping suggestions over time.

QUALITY, STANDARDIZED DATA IS ESSENTIAL FOR REGULATORY SUBMISSION

From its inception in 1997, the Clinical Data Interchange Standards Consortium (CDISC) has recognized the need for the establishment of standard data models to improve the process of electronic acquisition and exchange of clinical trials information for the benefit of all medical and pharmaceutical stakeholders.

CDISC's Study Data Tabulation Model (SDTM) is a standard for submitting data tabulations to the FDA in support of marketing applications. In 2004 this standard became a Study Data Specification, referenced in the eCTD Guidance.

Regulatory agencies around the world have adopted the SDTM data model for clinical data submissions:

Agency	Requirement	Date
FDA- CDER & CBER (USA)	Required	17 Dec 2016
PMDA (Japan)	Required	01 Apr 2020
NMPA (China)	Preferred	Sep 2019

Manual mapping typically employs a package of programming and visualization tools, which requires significant effort to build for each study. Significant effort is also required to review transformation logic for inconsistencies and anomalies and move the programs to an approved state. Domains must also be verified using specific validation rules, split according to size and translated into an open-source format before submission.

DATA AVAILABILITY AND CONSISTENCY

A typical clinical study involves 30–50 raw CRF datasets with 40–60 variables each, along with several external sources and data streams from devices etc. Study deliverables vary greatly, due to the variation in protocol directions and data collected.

Transforming this data into submittable SDTM is critical to downstream functions such as analysis, reporting and submission. However, protocol differences and variability of domains included in the standard mean that the transformations must be quality checked and approved, leading to delays.

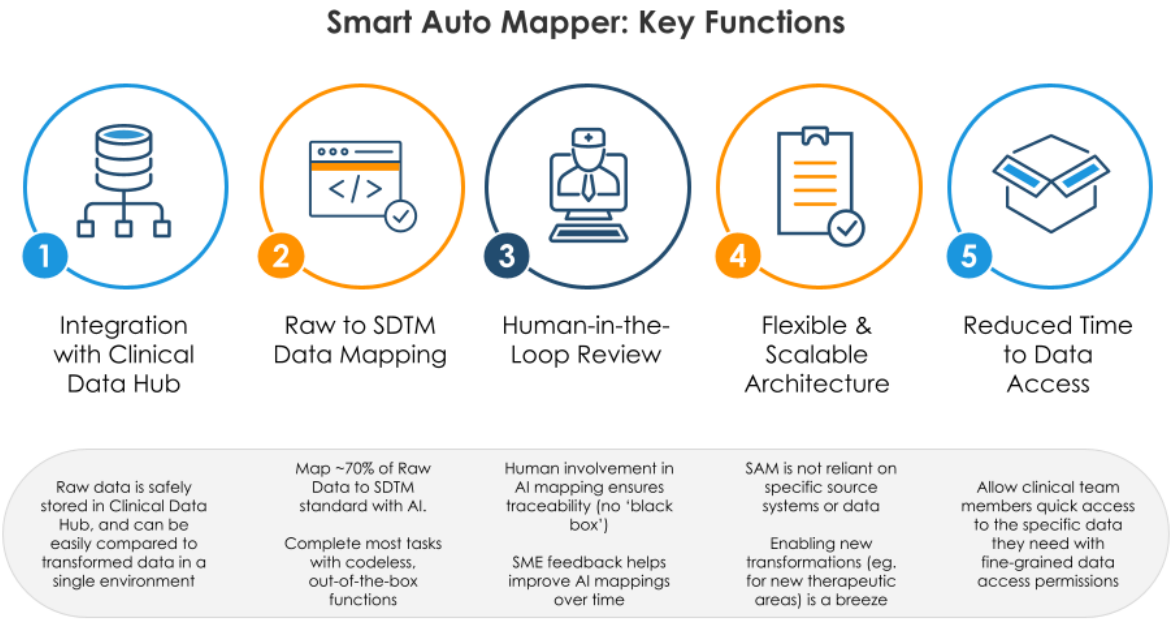

Figure 1 Smart Auto Mapper: Key Functions

AI CAN MAP DATA FASTER AND IMPROVE CONSISTENCY

Machine learning models trained on historical clinical trial data can identify clinical data transformations more efficiently and consistently.

The AI technology is made possible by the ability to ingest data from all your source systems, bring it together in a unified study data model, and update it in real time. This is the foundational capability of Saama's Life Science Analytics Cloud (LSAC).

Saama's machine learning data mapping solution, Smart Auto Mapper, gives you all the components you need to transform your data into submittable formats, including SAS V5 xpt files.

By learning patterns from historical data, the machine learning model automatically propose Dataset, Column and Functional mapping logic based on experience of other studies, standards, and by looking into the data values themselves.

Experienced users then accept or reject the mappings, in the process submitting feedback which is used to further refine the algorithmic suggestions. Accuracy is guaranteed with human-in-the-loop review and acceptance of mappings at a study level and integration with Pinnacle 21 to complete the domain validation rules.

MAP DATA MORE EFFECTIVELY WITH SMART AUTO MAPPER

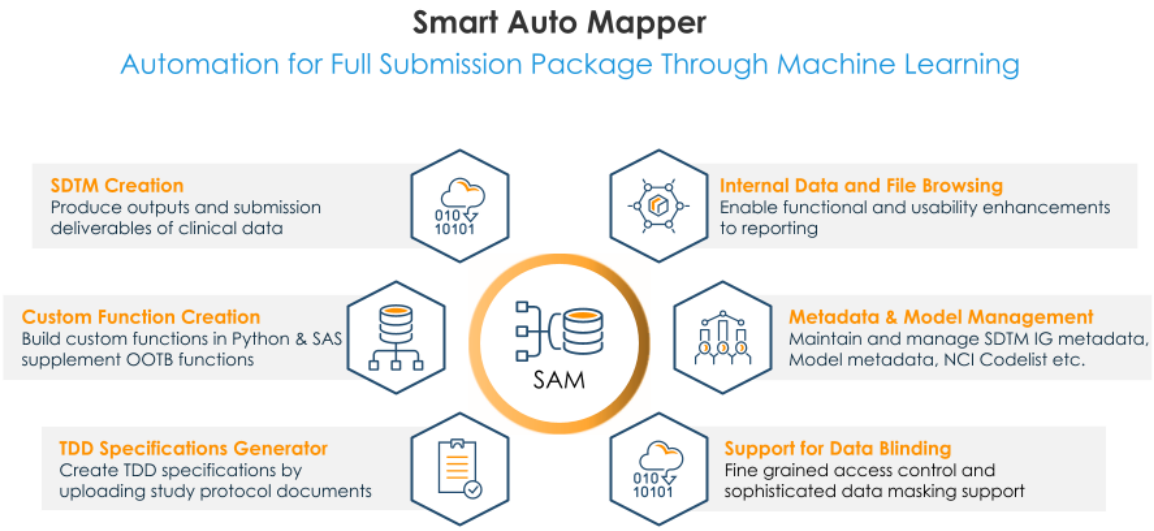

Figure 2: Automation for Full Submission Package Through Machine Learning

SMART AUTO MAPPER SOLUTION COMPONENTS

Saama's Smart Auto Mapper consists of three modules:

1. **Hub Administration** is used to load EDC and external data into the platform
2. **Global Administration** is used to manage CDISC Standard Versions, Variables / Data Elements and Controlled Terminology
3. **Study Administration** applies a CDISC Standard and Terminology Version to a study, and automatically assists users in mapping by suggestion source data transformations to the CDISC target

Hub Administration	• Adapters for pulling data and metadata in near real time from EDC and CTMS • Intelligent file watcher utility for pulling in third-party files for metadata checks
Global Administration	• Global data model management for additional therapeutic areas and indications • Reference data management include SDTM NCI Controlled Terminology • Global Standards for mapping activities • Global Custom Programs

Study Administration	• Global data model management for additional therapeutic areas and indications • Automated SDTM mappings, SAS xpt file generation and validation rule integration and reports • Study Standards such as Laboratory Unit standardization and Standard Results • Interactive Data Browser • Study specific Custom Programs
Out of the Box Functions	• Out of the Box functions exist in Smart Auto Mapper to complete mapping at 2 layers, Dataset and Column • These Out of the Box functions can be written in python or SAS, and take complex parameters • Out of the Box functions and their parameters are specified in full by the Automapping engine
Custom Functions	• Out of the Box functions can be supplemented by custom programs written in python or SAS • A coding standard gives a uniform appearance to the code which improves readability and maintainability • Custom programs can be used from the global level, or demoted to allow study specific changes
Lifecycle	• Comprehensive object lifecycle across global and study level objects • Full version history, including roll back to any previous version • Only approved versions of standards can be used in transformations

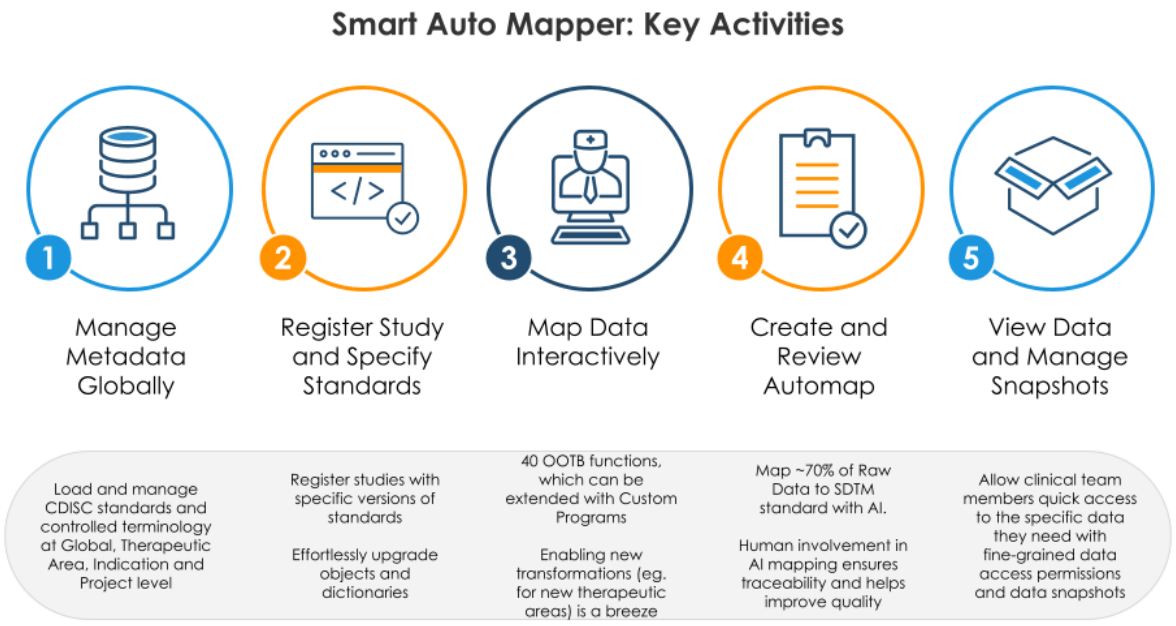

Figure 3: Smart Auto Mapper: Key Activities

MACHINE LEARNING MODEL DEVELOPMENT AND IMPROVEMENT

The heart of the system is a Machine Learning (ML) SDTM Industry Model (SIM) that is pre-trained based on SDTM Implementation Guides, SDTM Models, Therapeutic Area User Guides and FDA Conformance Rules. Using this ML-based approach, the improvement of the map prediction will be continuous. Users are able to create & maintain target standards (SDTM Metadata and NCI Codelist) at global, TA, Indication, Project/Compound level and then propagate them to be used at study level.

Users will be able to generate maps automatically and use them as a blueprint for data transformation to Conformed SDTM. Appropriate users are able to approve these mappings and incorporate them in jobs, while being able to determine lineage and traceability. Users are also able to generate XPT files for the purpose of running Pinnacle21 reports within the Smart Auto Mapper interface.

AI-DRIVEN ADVANTAGE

BEFORE SMART AUTO MAPPER

- Manual metadata and terminology management
- Homegrown programming processes with limited re-use or consistency
- Reliance on data from third-party providers such as CROs for creating submittable data packages
- Long delays in delivery of data to downstream consumers

AFTER SMART AUTO MAPPER

- Streamlined metadata and terminology management
- Sophisticated data mapping assistance including laboratory standardization
- Centralized programming processes with full change control and approval
- Comprehensive package of Out of the Box mapping functions
- Human in the loop mapping review, along with increased consistency from automapping suggestions
- Self sufficiency in creating fully documented, submittable data packages
- Early delivery of data to downstream consumers

CONCLUSION

Machine learning models trained on submission standards and historical mappings can save time when transforming data for regulatory submission and deliver data sooner. Accuracy is ensured with human-in-the-loop feedback controls to refine mappings at a study level that continue to improve the quality of the automapping suggestions over time, while the automated mapping suggestions increase consistency.

CONTACT INFORMATION

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

Mike Collinson
Saama Technologies Ltd

To learn more about Smart Auto Mapper and arrange a demo, contact Saama at info@saama.com or 408-371-1900, or visit us online at www.saama.com