Paper SD04

EDC Raw Data to SDTM Curation, Mapping and Automation with Xbiom Tool

Sunil Gupta, Gupta Programming, Simi Valley, USA
Raja Ramesh, Pointcross, Bangalore, India
Rahul Madhavan, Pointcross, Foster City, USA

ABSTRACT

As an alternative to traditional SDTM programming and a CDISC-360 goal, important achievements in mapping and automation of SDTM curation are now possible. Non-programmers can access a simple, low-code interface for control and visualization of EDC raw data to SDTM mapping. With built in SDTM conformance rules, XBiom tool leverages EDC raw data and metadata attributes to strategically compare and map them to SDTM metadata and control terminology variables and derivations. Users then visually review and confirm Xbiom's recommendations of each SDTM variable mapping and data while XBiom 'remembers and learns' the clinical study's profile. This greatly helps when EDC raw data is refreshed. Clinical Study Managers and Biomarker Scientists, for example, can focus on continual building of a curated single-truth view of the clinical trial as it progresses since XBiom tool 'remembers' previous data corrections.

INTRODUCTION

Whether your role is a Director, Statistical Programmer, Clinical Scientist or Statistician, you have a vested role and interest in the data ingestion, curation, analysis and the submission process. A tool that leverages metadata from all sources, industry standards and control terminology with a user-friendly interface and drag-n-drop feature for mapping to SDTMs, ADaMs, Define.xml, SDRG/ADRG and TLGs is an important phase towards CDISC-360 mission compliance.

This paper introduces XBiom for data curation and for automatically mapping raw data to SDTMs.

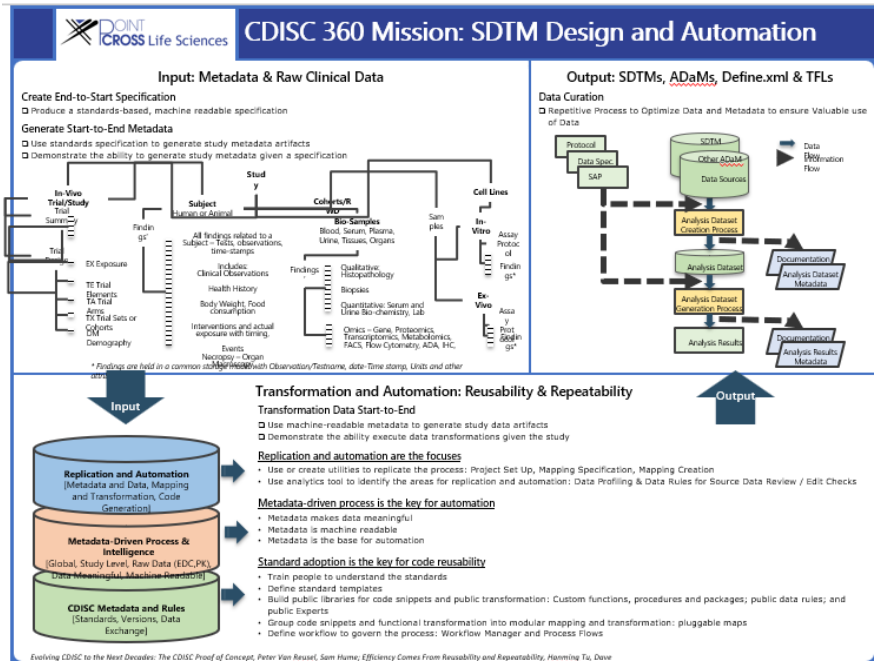
- Ready for Applying the CDISC-360 Mission
- Introducing the Xbiom Tool
- What is the Unified Data Model (UDM)
- What Does Data Curation Mean to You
- Understanding the Auto-Mapping Process

READY FOR APPLYING THE CDISC-360 MISSION

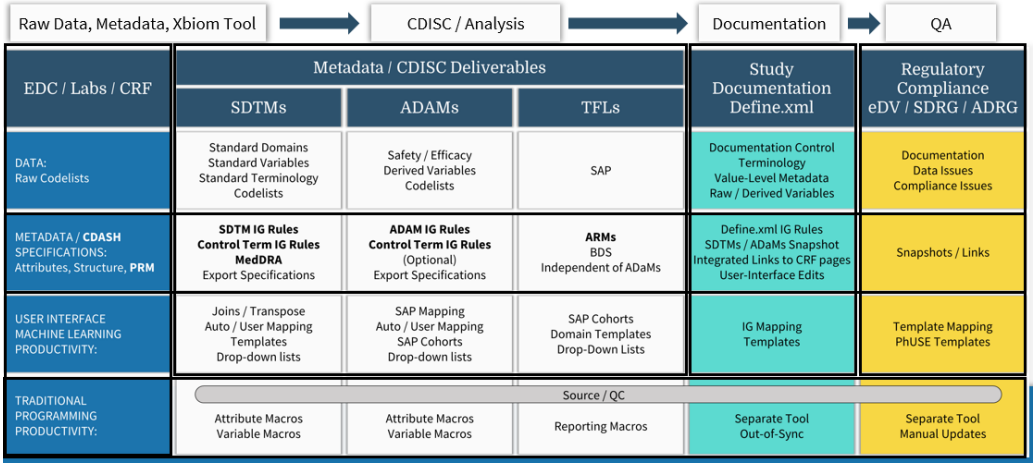
CDISC-360's mission is to apply the 80%/20% rule for SDTM generation.

"Apply the 80/20 rule to ensure the Project automates 80% of the end-to-end metadata and data processing needed to generate study artifacts suitable for a regulatory submission." Peter Van Reusel, Sam Hume, CDISC-360 Mission

The two process flow charts show the step-by-step process to transform raw data into SDTMs, ADaMs and TLFs. Throughout the process, metadata and standards should be applied to help automate the process. This is achieved by cross-referencing data structures, file names, variables, attributes and then data values. The many benefits of automation include SDTM optimization and reduced human errors.

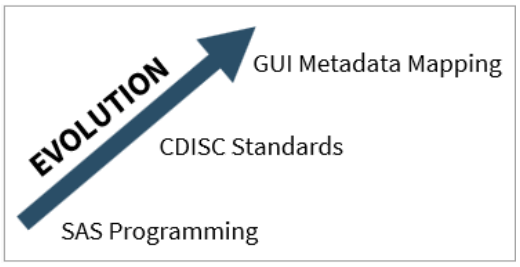


The second process flow chart includes all submission deliverables (define.xml and SDRG/ADRG). In addition, Xbiom tool exports SDTM/ADaM mapping specification excel file that can be read by R script to convert raw data to SDTMs/ADaMs. Without leaving Xbiom, SDTM and ADaM compliance checks can be performed with each cycle. In addition, define.xml specification is integrated within Xbiom so SDTM and ADaM are always in sync instead of snapshot versions. In the final version, define.xml, SDRG and ADRG can be finalized.



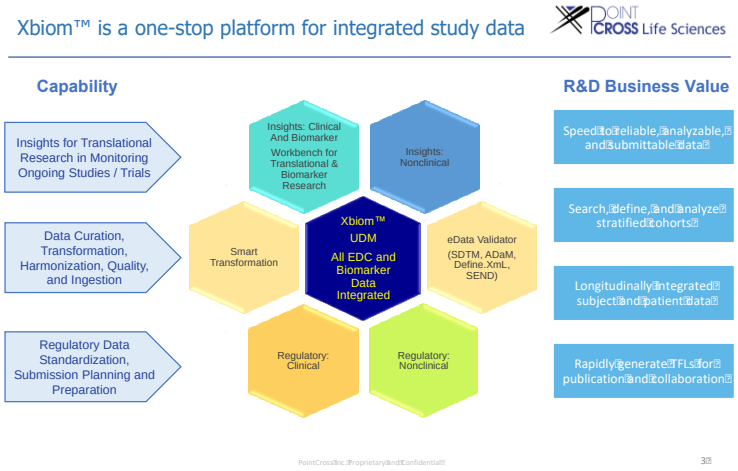
INTRODUCING THE XBIOM TOOL

The Xbiom tool is a very comprehensive metadata repository system within a statistical computing environment. Xbiom is based on the Unified Data Model (UDM). The UDM covers any eCRF, Biomarker Assay and other sub-studies into ONE unified data model that is F.A.I.R.; as opposed to a set of domain specific files (e.g. SDTM or path to SDTM). The idea is that UDM carries all the data of interest in a persistent permanent standardized format so that it can be a source for any SDTM-IG and CT of choice for exchange purposes without having to do any data wrangling with the as-collected data anymore. As a result, all data is ingested, cataloged, indexed and linked for user search and querying. The development of Xbiom has been an evolution process from traditional programming to CDISC standards to using GUI metadata mapping systems.

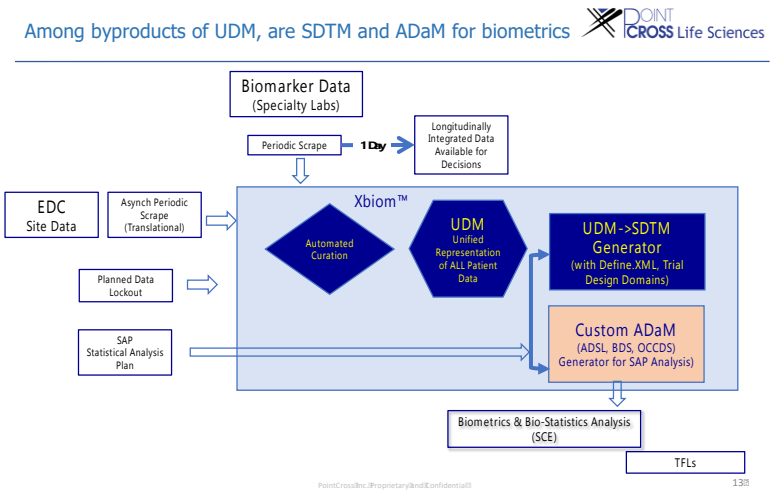


Xbiom consists of several key modules that are integrated as a complete and cohesive system. Both CROs and sponsors can benefit from this design to more rapidly ingest data for better monitoring of primary and secondary endpoints.

Xbiom tool enables users to reduce or eliminate data management chores by making each interaction for data curation a logged event. This then is a machine augmented smart transformation process that supports automatic generation of SDTM files based on selected IG and Control Terminology. A direct way is to apply the SAP driven cohort selection and statistical analysis on the UDM trial data. Then users can generate tabulations, figures and listings with annotations and tags (TFLs) for reporting. The next step is the generation of supporting traceable ADaM data sets that connect the SDTM to the TFLs.

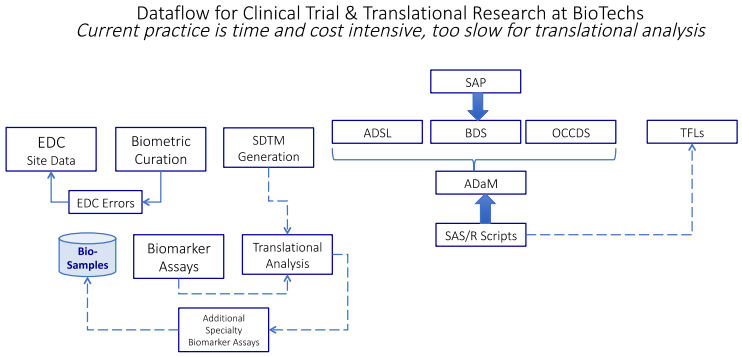


As a result of this non-linear model, SDTM generations are reduced to eight hour cycles. The UDM is the first model created before the SDTM automation process. With every raw data refresh, SDTMs can easily and quickly be refreshed as a simple operation since all of the mapping has been pre-defined.



With the SAP, cohorts can be created using the visual interface to drag-n-drop variables from any raw source since they are internally linked. Once cohorts are created, they can be applied as groups in pre-defined tables, lists and

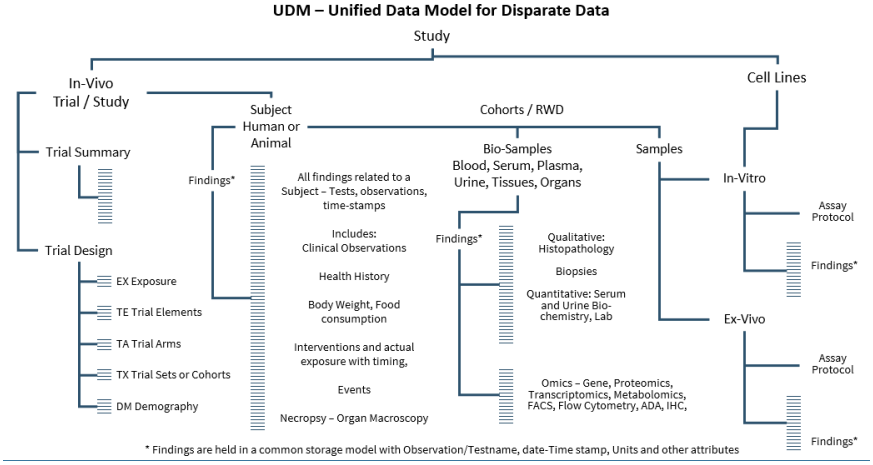
figures. Once tables, lists and figures are finalized, then ADaMs can be created based on derived variables. This unique approach results in better ADaMs since ADaMs are designed after reviewing the summary tables of efficacy and safety endpoints.



WHAT IS THE UNIFIED DATA MODEL (UDM)

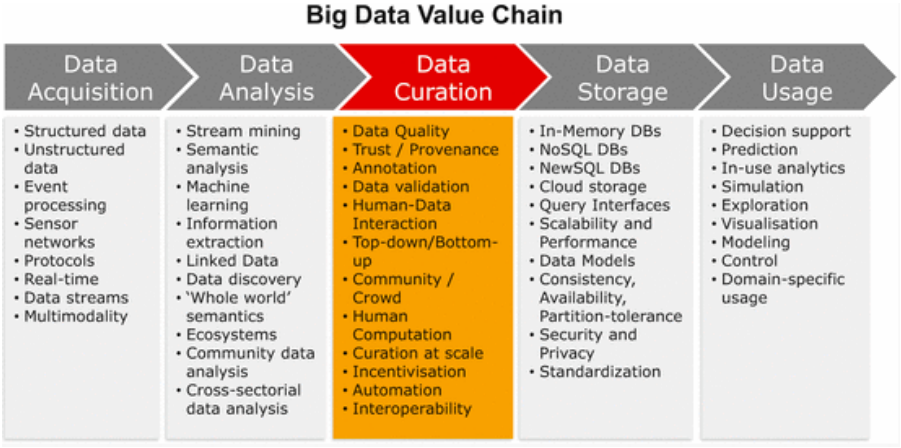
UDM is a repository standard that is built with a metamodel to represent all collected data in a study along its timeline, for each subject or its samples. It also includes all the study level metadata that is needed for a complete submission with SDTM as well as those needed to support analysis. UDM is ready for analysis, and ready for transformation into any exchange format. UDM cannot be readily exchanged unless transformed to an exchange standard and format.

In comparison, SDTMs represent an exchange standard organized into columnar tabulations or listings by subject level record in various data domains. Timepoints are a variable within such records. Analysis can be done only by re-assembling SDTM data to support the specific analysis planned. SDTM is readily exchanged but not readily analyzed.

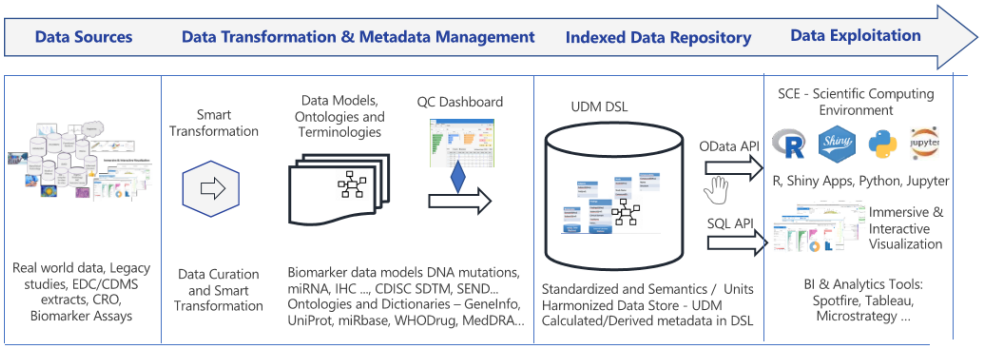


WHAT DOES DATA CURATION MEAN TO YOU

Smart organizations take time to document and fine tune their data ingestion and curation process. Because this is a repeated process, organizations realize that a small improvement makes a big long term impact. Data Curation is the repetitive process to optimize data and metadata to ensure valuable use of data. As shown in the figure below, data curation is a key component for better use of data.

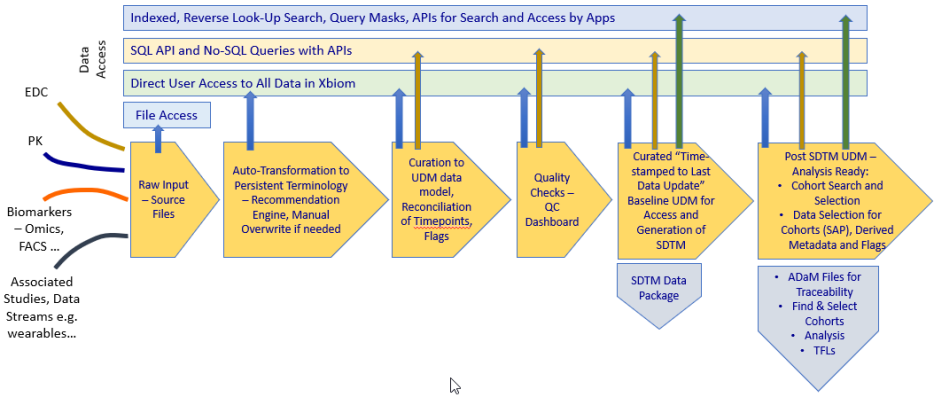


Within Xbiom, data curation is integrated within the system. From ingesting all data sources (EDC, Labs, biomarker, etc.), data transformation and metadata management to indexed data repository and data exploration, Xbiom has optimized the process of data input, management and analysis.



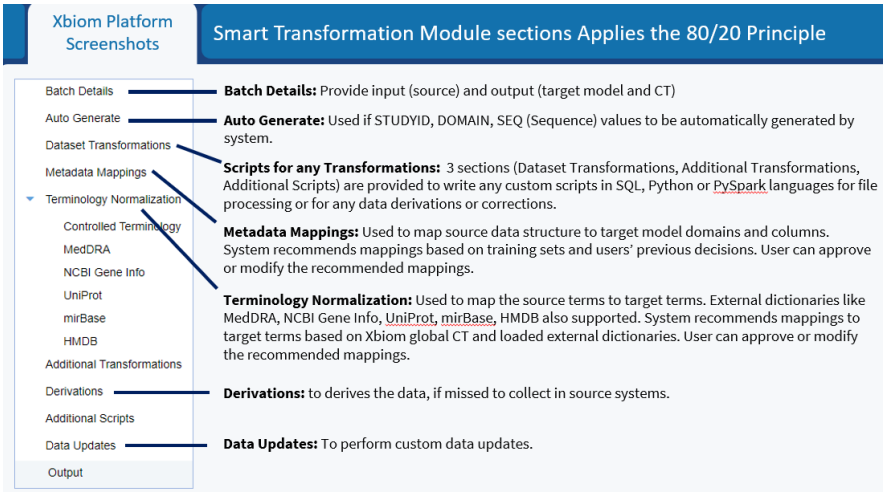
UNDERSTANDING THE AUTO-MAPPING PROCESS

Auto-mapping is part of the whole system of processing data from ingestion to standardizing using UDM followed by compliance checks and SAP based reporting and analysis. User interface panels with drop-down options help to specify study templates, models and dictionaries.



The design behind automapping is from a supervised learning model based on hundreds of clinical trials. The process of auto-mapping consists of a combination of pre and post batch processing, dataset transformations, metadata mapping, terminology normalization and derivations. This design incorporates the best of both worlds for applying a standard process to enforce rules as well as low-code programming for customization. After the system recommended mapping variables are reviewed and confirmed, users map the remaining raw variables using drop down options for selecting SDTM domains, variables and writing low code as needed.

With one cycle to auto-map and user-base mapping to SDTMs and ADaMs, the Xbiom tool documents and applies the variable mapping for repeating SDTM and ADaM creation cycles.



Within Xbiom, there are three sections (Dataset Transformations, Additional Transformations, Additional Scripts) which enable custom scripts in SQL, Python or PySpark languages. The is ideal for file processing or for any data derivations or corrections. The dataset transformation section allows users to enter program code to merge related datasets so that all variables are in a single dataset. In the variable mapping section, simple programming expressions and case when code block, for example, can be embedded. After the mapping section, there are boxes to select which standard variables should be derived.

CONCLUSION

The benefits of using Xbiom tool over traditional SAS programming teams are numerous. While many organizations may have a collection of SAS macros with metadata programming methods, very few actually have true auto-mapping feature that can learn and remember for similar clinical studies. With automation comes increased compliance and reduced cycle times. These are the KPIs for clinical study submission packags.

ACKNOWLEDGMENTS

We would like to thank Kishore Pothuri and Gayathri Mahadevan for accepting our paper in the software demonstration section.

RECOMMENDED READING

Harmonizing Raw Data with CDISC Standards to Streamline SDTMs, Sunil Gupta, Presentation
https://pointcrosslifesciences.com/harmonizing-raw-data-with-cdisc-standards-to-streamline-sdtms/

CDISC-360 Mission: SDTM Design and Automation
https://pointcrosslifesciences.com/wp-content/uploads/2023/01/CDISC-360-Mission-FINAL.pdf

End-to-End Clinical Study MetaData-Driven Process
https://pointcrosslifesciences.com/wp-content/uploads/2023/01/end-to-end-clinical-study-dataflow.pdf

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:
Sunil Gupta, Submission SME
Gupta Programming
213 Goldenwood Circle
Simi Valley, CA 93065
GuptaProgramming@gmail.com
SASSavvy.com, R-Guru.com

Raja Ramesh, Chief Architect (CTO)
Pointcross Life Sciences
Bangalore, India

Rahul Madhavan, VP – Strategic Programs
Pointcross Life Sciences
Foster City, CA