

# Development of the BIMO Evaluation Support Tool (BEST) – Using Computer Vision and OCR to Enhance On-Site and Remote Evaluations of FDA-Regulated Research Sites

Catherine Li<sup>1</sup>, Godfrey Amadi<sup>1</sup>, Divine Agbale<sup>1</sup>, John-William Declaris<sup>1</sup>, Melkamu Getie-Kebtie<sup>2</sup>, Kara Scheibner<sup>2</sup>, Yiyue Zhang<sup>2</sup>, Sean Kassim<sup>2</sup>, Seongeun Julia Cho<sup>2</sup>, Sanya Gupta<sup>3</sup>, Vivek Roy<sup>3</sup>, Miriam White<sup>3</sup>, Dilip Philipose<sup>3</sup>, Collin Lobb<sup>3</sup>

<sup>1</sup> FDA CDER OTS Office of Computational Sciences, <sup>2</sup> FDA CDER OTS Office of Study Integrity and Surveillance, <sup>3</sup> Deloitte Consulting LLP

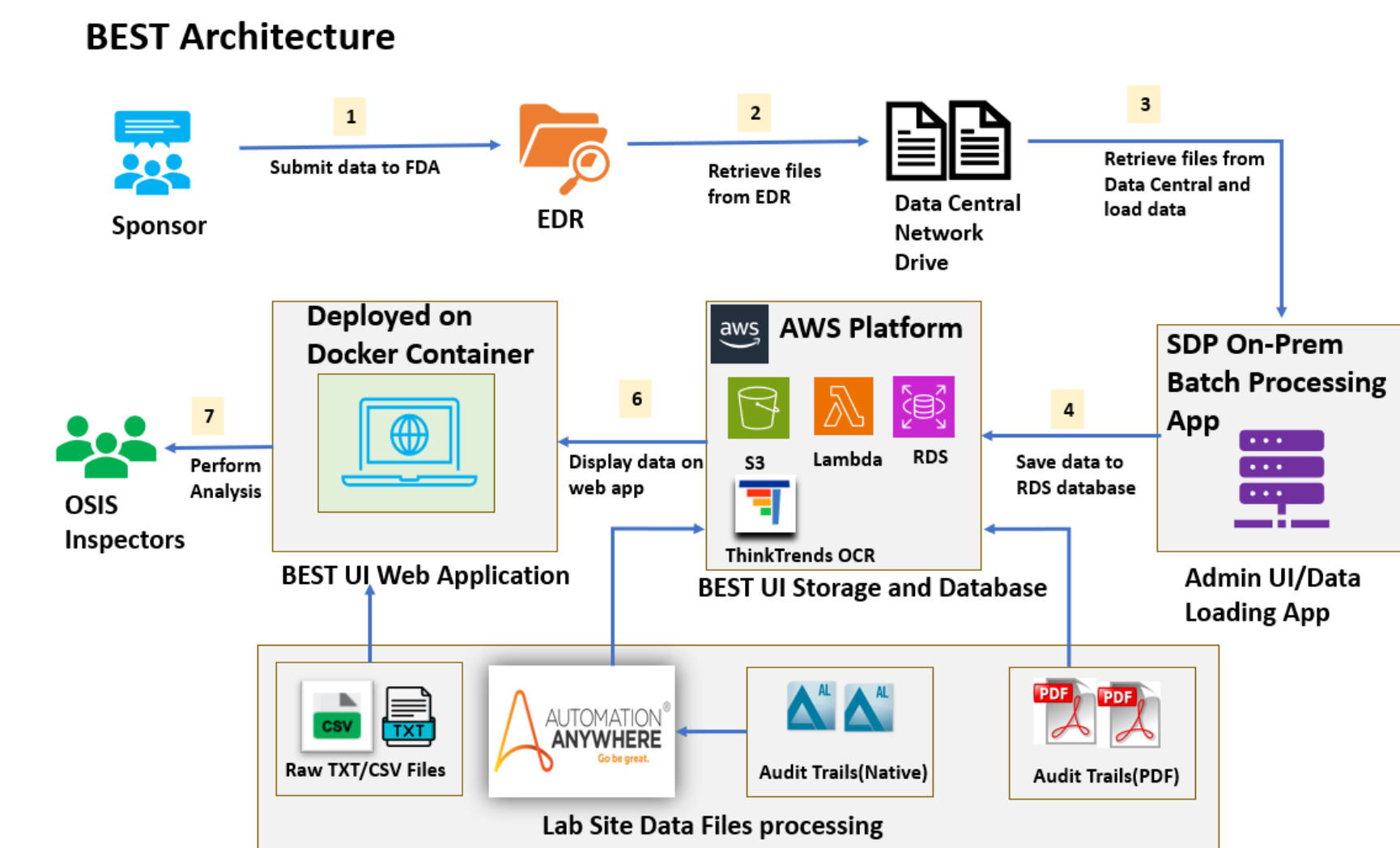
PP20



## ABSTRACT

The Office of Computational Science (OCS) at FDA, in collaboration with the Office of the Study Integrity and Surveillance, developed the BIMO Evaluation Support Tool Platform (BEST) using computer vision and OCR technologies to enhance FDA’s evaluation of study sites (onsite and remote).

BEST platform contains two modules: Data Comparison Tool (DCT) and Audit Trail Analysis Tool (ATAT).



### Tech Stack –

Java, Angular, REST, Spring, Oracle DB, AWS Services (S3, Lambda, RDS), Docker Container, OCR/computer vision, and RPA.

### High Availability –

BEST is deployed on Docker containers for high availability and reduced downtime.

### Cloud-based Services –

BEST uses Amazon Web Services (AWS) S3, Lambda, RDS, etc. for storage and processing.

### Optical Character Recognition (OCR) –

BEST uses ThinkTrends OCR technology for data extraction from PDF files.

### Robotic Process Automation (RPA) –

BEST uses Automation Anywhere RPA bots to process native audit trail files.

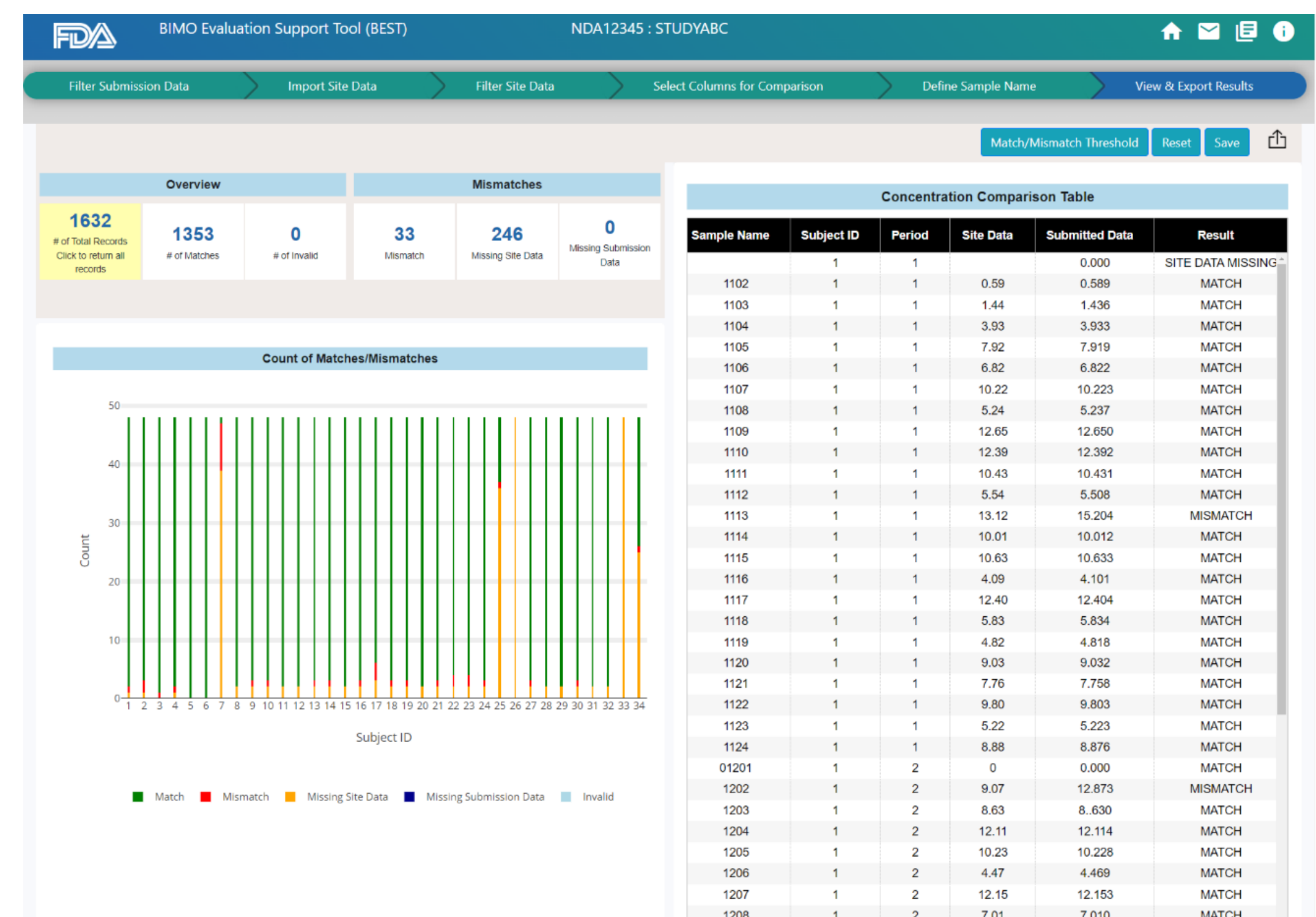
## ACKNOWLEDGEMENTS

The authors would like to acknowledge colleagues at the Office of Study Integrity and Surveillance. We would also like to thank ThinkTrends for collaborating with us on the BEST Platform.

## Data Comparison Tool (DCT)

DCT enables data authentication by comparison of data in FDA submissions to source data provided by sites. DCT replaces significant manual work and provides users with enhanced data analytics through automated comparisons, visualizations, and data export.

The tool provides an efficient workflow that allows inspectors to filter for data of regulatory interest and dynamically join site data with FDA submitted study data for analysis.



The View & Export Results Page displays the joined results between the submission and site data (*mock data shown*). Filters throughout each element of the page can be applied to create a custom data table and graph to be directly exported into Excel. The results can be saved to the Previous Analyses Repository with notes to be accessed at any later time.

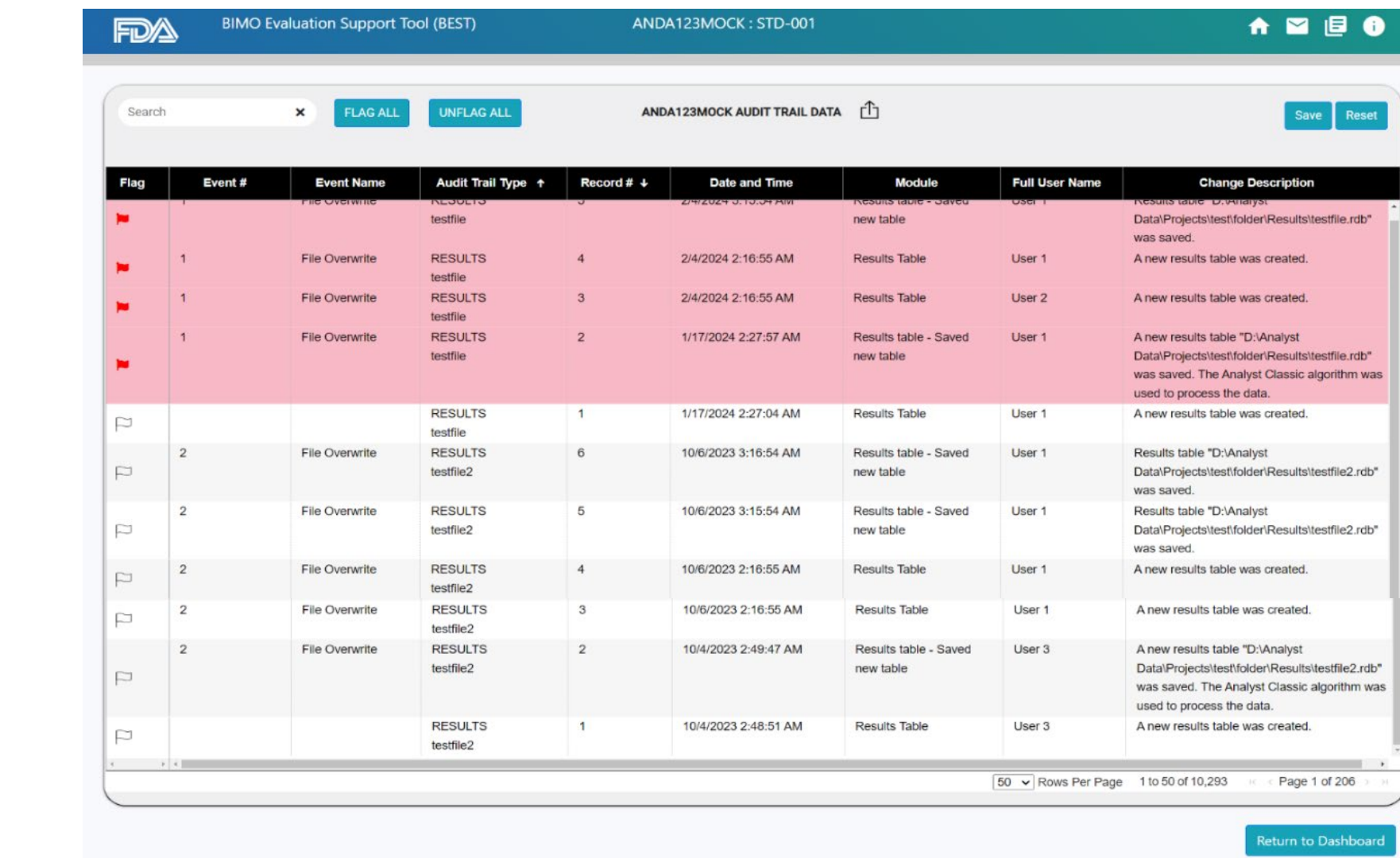
## BUILT ON STUDY DATA PLATFORM

- OCS Study Data Platform (SDP) is a clinical and nonclinical trials data warehouse application for CDISC SDTM, and ADaM data that facilitates automated processing, management, and reviewer access to standard subject-level clinical and nonclinical trials data to meet a variety of regulatory review and analysis needs. SDP is used to house the study data within BEST.
- SDP is designed especially to support analytical and reporting functions involved in the regulatory review of drug applications.

## Audit Trail Analysis Tool (ATAT)

ATAT uses RPA, Computer Vision and OCR technologies to convert mass spectrometry software audit logs to a machine-readable format to allow users to view/sort/filter/search audit logs and flag records of interest. ATAT module is based on two types of data extraction:

- Extraction from Native Analyst Audit Trail files.* Extraction is done from native files using Automation Anywhere RPA bots and SciEx Analyst software.
- Extraction from PDF Audit Trail files.* Extraction is done from PDF files using ThinkTrends OCR technology and computer vision.



The ATAT module presents a consolidated table (*mock data shown*) with all audit logs for a study. Inspectors can efficiently analyze the data across all the audit log files for a study (10+ files, 9000+ rows of data) through search, filter, and sorting functionalities. ATAT automatically identifies groups of records which meet an explicit OSIS defined business rule (i.e., “File Overwrite” in example above).

## IMPACT

- BEST has implemented automations to simplify the process to request and load a study. These automations have decreased the study load time from 2-6 hours to approximately 15 minutes.
- DCT has decreased the time required for data verification by 4+ hours/analyte, which is a major reduction of time required for review and verification.
- DCT allows inspectors to perform a comprehensive review of site and submission data through automation.
- ATAT improves efficiency and quality in analyzing audit trail logs over manual review of PDF documents.