

## Real Time Data Cleaning – A Reality

Vidya Muthukumar, IQVIA, Durham, NC 27503

### ABSTRACT

With the constant emphasis and request by sponsors in clinical trials to "clean" data regularly and not wait until close to final database lock, CROs are looking for smart and cost-efficient ways to "clean" data. The exponential increase in different types of data sources along with the rising costs of clinical trials requires bolder, smarter, and leaner solutions to current manual data cleaning processes. When digitized, centralized data flows and standards connect together to provide an integrated, streamlined, and automated data management process, it transforms into a system called Real-Time Data Cleaning (RTDC). Continuous data ingestion and automated cleaning, results in shortened database locks, issue detection and mitigation. The increasing data source types together with the complexities of trial designs require new techniques to meet these challenges efficiently while maintaining high quality, without raising costs. The obvious solution is automation. Automation requires electronic repeatability, digitized consistency, standards, and compliance. Digitized processes leverage connected data and standards to drive data management, reducing manual processes, ensuring consistency, and increasing efficiency leading to faster identification of issues to enable companies to make faster informed decisions in clinical trials.

### INTRODUCTION

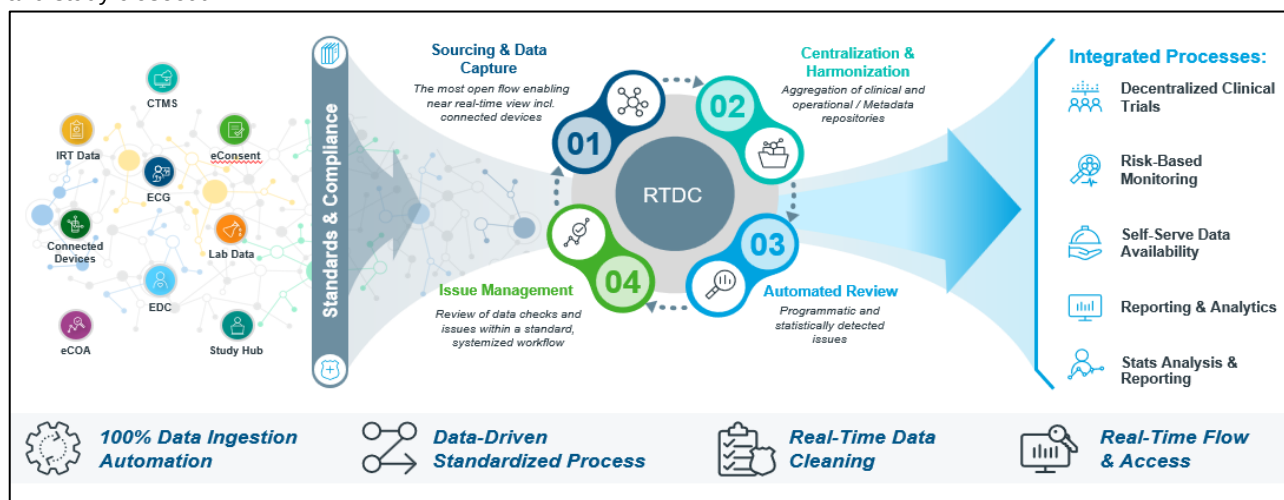
Real-Time Data Cleaning (RTDC) is a comprehensive IQVIA-owned connected data flow standards solution to provide an efficient real-time end-to-end data management process for a clinical trial. The environment has been designed to enable real-time flow of all data sources into a centralized Data Repository and features automated cleaning applications to reduce manual data cleaning activities. The connected environment enables data from multiple sources (such as different file types from single or multiple labs, blinded data, *if applicable*, and different vendor data types) to move seamlessly from electronic Case Report Form (eCRF) design to data review specifications, to data transfer programming and then to status reports, all with automation allowing data managers to focus on trends and decisions.

### WHAT DOES RTDC ENTAIL?

The combination of standards, data, operations, and reporting working in unison to provide an integrated, streamlined data management process.

- Standards – Use of IQVIA's (or Customer) Standards to design Electronic Data Capture (EDC), Edits and Study Data Tabulation Model (SDTM), Analysis Data Model (ADaM) datasets and Tables, Listings and Figures (TLFs).
- Run operational start-up and ongoing conduct within Data Management, including Data management Statement of Work (DM SOW), Data Management Plan (DMP) and approval, Standard Operating Procedures (SOP) lists, data flow setup, programming specifications, data review and issue management.
- Data Repository for Data Flow and Programming – EDC and vendor data ingested into the repository. Data refreshes as new files received, triggering Quality Control (QC) and cleaning flows.
- Reporting & Analysis from a data repository – Centralized data flow and provide reporting metrics. A Clean Patient Tracker (CPT) and Data Analytics reports are available to track study data metrics.

RTDC connects operational processes to a single data stream, enabling the data eco system to automate and manage real-time data cleaning activities. It promotes data cleaning readiness closer to FPI thereby enabling the Data Management (DM) team to focus on improving overall data quality and maintaining data integrity until final database lock and study closeout.



The combination of digitization, centralization and standardization enables RTDC to be the cutting-edge real-time data cleaning solution as a one-stop shop for global data management. Digitization eliminates physical documents by capturing information as systemized data in a centralized and standardized way, and then that data is used in the automated processes for dashboarding and reporting metrics using the EDC Status Reporting tool and the Clean Patient Tracker (CPT). With the single centralized data flow of all clinical data sources into a Data Repository, it can connect all consumers of that data to a consistent data ecosystem. With standardization, IQVIA uses a Metadata Repository (MDR) to build, connect, and house IQVIA and sponsor-specific metadata standards. These standards are grouped into connected libraries of electronic Case Report Forms (eCRFs), Edit Checks, Study Data Tabulation Model (SDTM) data sets, Analysis Data Model (ADaM) datasets, and Tables, Listings, and Figures (TLFs). These artifacts are maintained by standards libraries.

The RTDC process flow also entails, for example, an EDC status reporting tool that contains form level status, subject status, DOV date, randomization date, SAE details, missing visits, missing forms, incomplete forms, query status, SDV status, freeze status, signature and lock status customized for a project. Extracts from the EDC Status Report is then processed by Spotfire tool to create a Clean Patient Tracker (CPT) for efficient use and visualization in tabular format.

The CPT contains the following data:

- Summary- Summary of EDC status at a patient level.
- Missing Visits - Visits that are overdue and missing.
- Missing Forms - Forms missing.
- Incomplete Forms - Forms incomplete.
- Query Status - Open and Answered Status for various query action groups.
- Pending SDV Status - Forms pending SDV.
- Pending Freeze Status - Forms pending Freeze.
- Pending Signatures - Forms and casebooks pending Signature.
- Pending Lock Status - Forms and casebooks that are pending database lock.

## **BENEFITS OF RTDC FOR DATA MANAGEMENT**

- External source data validated by system upon ingestion into a data repository, freeing up programming teams for better data quality checks and review.
- Cleaning processes are in place closer to FPI, enabling issue/trend detection sooner.
- Cleaning processes triggered as data flows into the data repository, automatically presenting identified issues for review/resolution within the data review dashboard.
- Issues are reviewed and routed for resolution directly in dataflow system, eliminating manual transcription and tracking in Excel and email.
- Full issue traceability from generation to closure.
- Ability to standardize cleanliness reporting and tracking with minimal manual effort.

## **BENEFITS OF RTDC FOR SPONSORS**

- Cost-effective, efficient, state-of-art comprehensive on-stop shop solution.
- Complete data cleaning processes in place closer to First Patient In (FPI).
- Faster issue/trend identification provides information needed to make well-informed decisions about the study and patients in an expedited manner.
- Enables comprehensive, standardized tracking of data cleanliness real-time.

## **CONCLUSION**

Real-time data cleaning can expedite trial timelines by providing a seamless data management environment and near real-time access to queries and study results. By eliminating lags in data reviews, this automated solution can dramatically reduce manual data cleaning tasks and related errors, enabling faster review of results. What traditionally may take up to 12 weeks through manual data cleaning processes, with real-time data cleaning, timelines can be reduced 50% or more since IQVIA's RTDC is set up as a robust, comprehensive digitized, centralized and standardized one stop solution.

## **REFERENCES**

IQVIA RTDC SharePoint

## **CONTACT INFORMATION**

Author Name: Vidya Muthukumar  
Company: IQVIA  
Address: 2400 Ellis Rd Pod A, Level 4  
Durham, NC 27503  
Work Phone: +1-984-212-5831  
Email: vidya.muthukumar@iqvia.com  
Website: www.iqvia.com