PhUSE US Connect 2019

Paper AR05

Data Preparation and Publishing Best Practices for a Successful RBM Implementation

Sreekanth Gudapati, eClinical Solutions, Boston, USA

ABSTRACT

A key to successful RBM implementation is having the right tool kit and applying best practices specifically with data preparation and publishing, as the number of data sources required for analysis continues to increase. Deploying an end to end technology platform for RBM helps with user engagement while speeding the implementation, providing value along the way.

This presentation will include a case study reviewing key considerations for both sponsors and CROs:

- Data Ingestion and Integration: Tools and best practices needed to blend data across different systems and formats (CTMS, EDC, Spreadsheets etc.) for real-time RBM analytics.
- Data Aggregation: Tools and best practices on normalizing and aggregating for review and risk identification across different levels including sites, regions or cohorts.
- Visualize and Take Action: Easy ways to visualize data, detect outliers and raise issues resulting actions all through audit able interactive UI is pivotal for a successful RBM implementation.

INTRODUCTION

Data preparation and publishing comprise the major technical work involved in RBM implementation. As the complexity of clinical trials increases and various data points are captured across disparate data sources such as Electronic data capture systems (EDC), ePRO, mobile questionnaires, wearables, and more, it is a constant challenge to get all the data together in one place and automate integration for any risk assessment and data aggregation needs. Easy to use visualizations become a foundation for the user adoption and return of value from the risk based approach. Gathering the right tools for both these functions and following best practices determines the success and adoption of the Risk Based Monitoring Implementation.

There are a few key areas that impact the implementation in a significant way: Data Ingestion and Integration, Data Aggregation, and the final step, Visualizing and Taking Action. In this paper we outline the best practices within these key areas that eClinical Solutions has developed while working through different implementations.

DATA INGESTION & INTEGRATION

This covers the first step of processing the data from a source - irrespective of its format - and loading the data into the user's platform. Due to the variance of data formats used in the industry, it is important to choose a file format and metadata agnostic tool that speeds up the data prep. The tool or processes should supplement tracking of any change in metadata. Without such tracking, there is a high probability that undetected source structure changes from the numerous sources, vendors and systems can break any integration process.

CHALLENGES

- Manual data sources: Users often generate the source files and a data manager or vendor then uploads a file to SFTP. It is important to push for having a system-generated, automated file in order to avoid manual dependencies and file inconsistencies.
- Timing of different data inputs: The datasets from different sources often arrive at different times and frequencies. Any timing differences must be addressed in the schedule and the users would have to be educated about this difference.

Example 1: To identify the sites at risk that have patients who are behind on ePRO completion for more than two assessment periods.

Here the data points were spread across three different sources: the Rave EDC system, the ePRO file sent weekly as a delimited file, and static Excel files provided by the user.

- EDC: Rave offers rich web APIs to help extract the data. Leveraging these APIs puts the schedule in users' control. EDC data was needed to determine whether the subject was still active on the study and at what stage of the visit schedule.
- ePRO: ePRO data was integrated through FTP on a weekly basis as a delimited file. This included ePRO questionnaires filled by the subjects.
- Static files: An Excel file was provided by the user to drive the grace period allowed after each visit. Having this sourced from a file allowed changes of the grace period as needed.

Example 2: To flag and do additional review for patients that meet renal safety risk criteria: a change of 50% from baseline to the next two visits.

- EDC: EDC data was received through Medrio web API calls on a daily basis. EDC data was needed to supplement the data received from the vendor during review once the subjects with associated risks were identified.
- Lab data: Central lab data through SFTP was delivered by the vendor once every 2 weeks. This served all the data points needed for the risk assessment.

In both the examples, having the ability to extract and process the data from a variety of data sources helped accomplish the data ingestion and integration faster and helped make the data aggregation easier.

DATA AGGREGATION

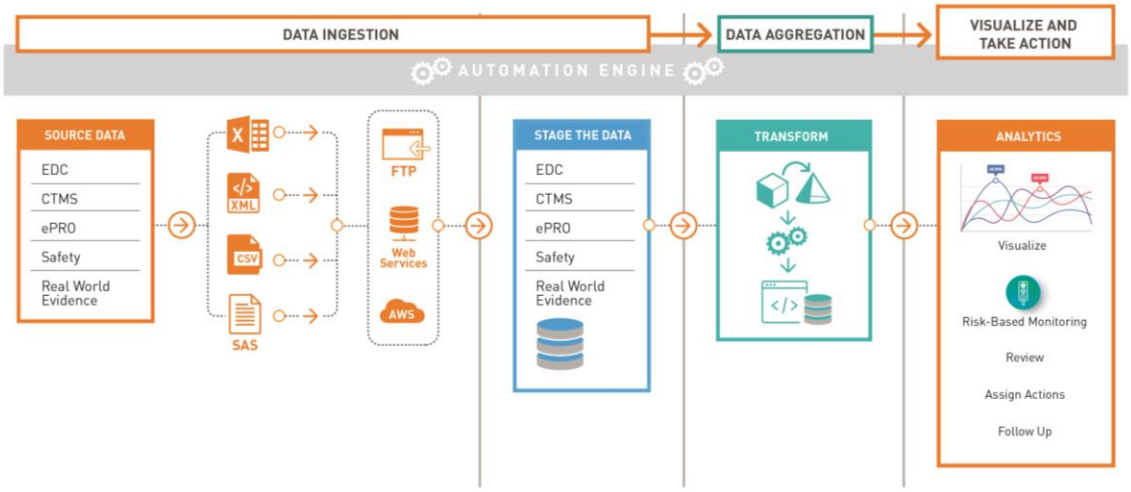
For data aggregation, users will need a data standardization/ETL tool which can access all the data sources ingested into the platform during step one (data ingestion and integration), and transform the data into normalized structures that can be used for risk assessment and RBM reporting. Data ingestion and aggregation should have an automated workflow to avoid manual intervention and provide the latest data to the user.

While defining structures during data normalization it is important to factor in these items:

- The history of each KRI is captured to see the trend
- There is an ability to register an action for each event
- There is a capability to track the action recorded by the user on a triggered event

CHALLENGES

- Automation of data aggregation with ingestion and report generation: Having independent tools creates multiple failure points and increases the implementation time.
- Aggregation could be simple to very complex, driven by the KRI requirement: Having a robust library of reusable components either directly through the tools or built by the programming group helps expediate the development cycle and reduce maintenance overhead.



VISUALIZE AND TAKE ACTION

Once the normalized data points are available with the calculated risk scores, the next significant step is to provide the user with easy steps to review and take action on the events. Here are a few things that drive the adoption.

During data normalization, while defining structures it is important to factor in below items:

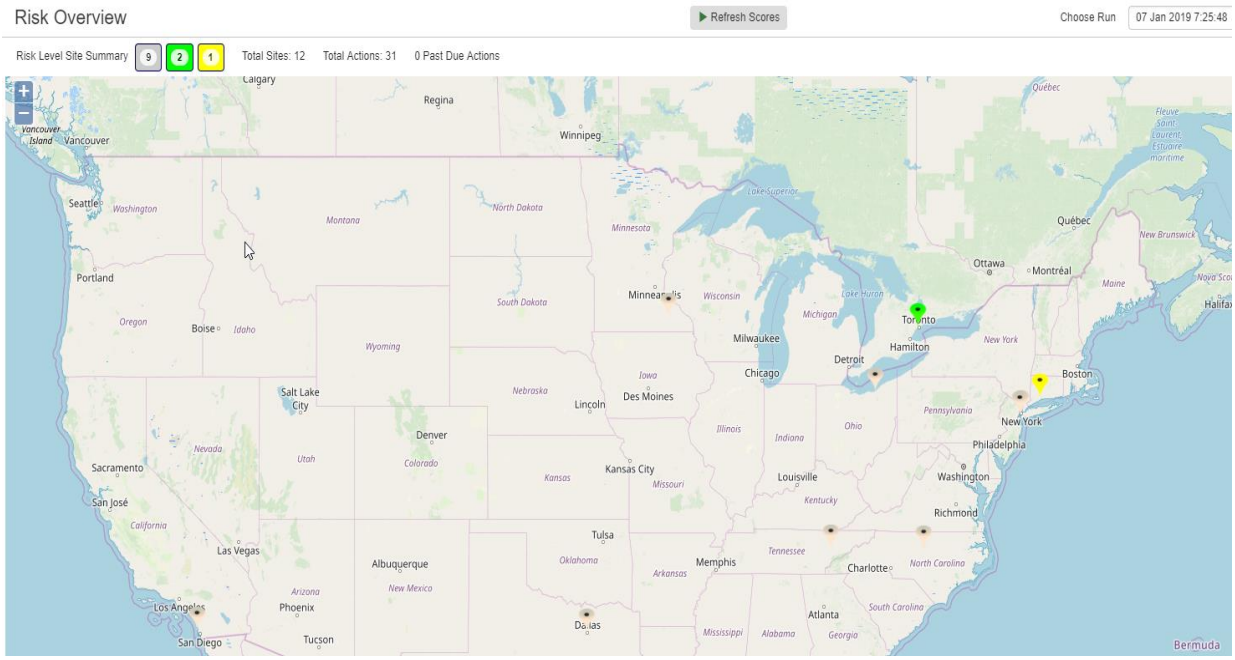
- Easy to use visualizations with supplemented data listings. Data listings help re-assure the observation the user is making from a visual. Our observation was that providing listings with data relevant to the risk indicator helps self-service.
- When a KRI meets the defined criteria of risk and a site/subject or cohort is flagged, the user should have a way to review the data and mark to closure, or a means to assign it to the right individual to bring the issue to closure.
- Snooze the user action to suspend the alerts on a known issue and have the privileges to set a required date.

CHALLENGES

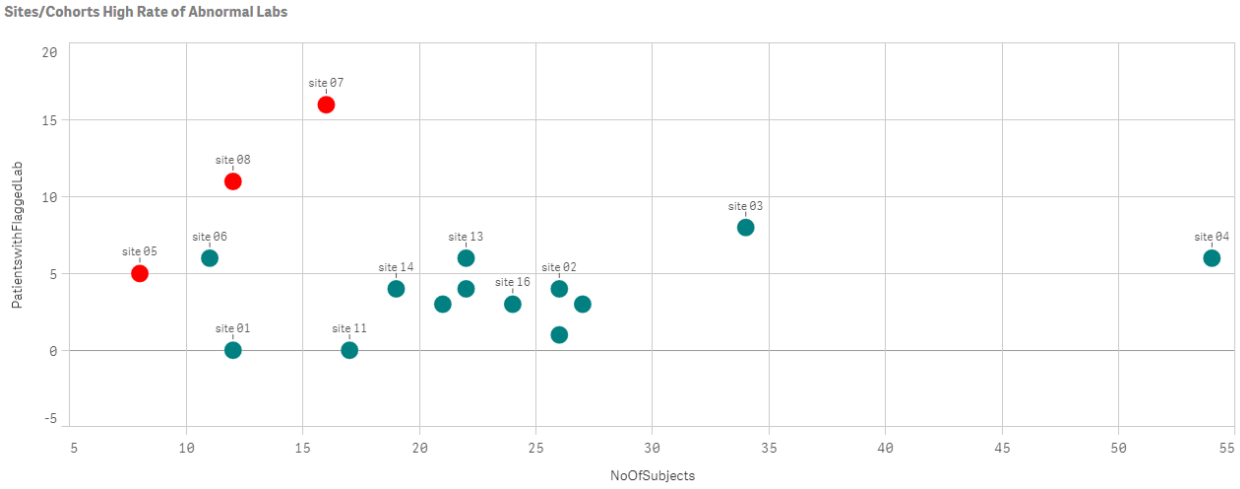
- Often it is easy for the user to become overwhelmed with the information on the visualizations. Keeping the visualizations simple to use helps keep the user engaged and focused on what is meaningful.
- Presenting all the data points required for review in one place so that the user does not have to jump through different systems is a common challenge in this step.

EXAMPLE:

1. The screen shot below is a visualization presenting the sites for the study on a map color coded with risk scores. This gives quick insight into any sites that the user would have to pay attention to.



2. The screen shot below is a visualization with masked data presenting the Sites/Cohorts color coded by risk qualifying criteria. The user has the flexibility to choose between Sites/Cohorts, simplifying the access to different slices of data.



CONCLUSION

As data collection expands to wearables and other devices, data ingestion and integration challenges will only compound and become a hurdle for time sensitive implementations. This highlights the significance of having the right tools to enable short implementation timelines by identifying the risks earlier in the study.

CONTACT INFORMATION

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

Sreekanth Gudapati
eClinical Solutions LLC
603 West Street
Mansfield, MA 02048, USA
Work Phone: 866.961.3542
Email: sgudapati@eclinicalsol.com
Web: www.eclinicalsol.com