LTI
Let's Solve

A Larsen & Toubro Group Company

Whitepaper

Role of Intelligent Automation in Transforming the CSR QC Process

Introduction

The recent pandemic has altered the modus operandi for clinical trials in terms of planning, conduct and reporting. Acceleration of clinical trials with digital adoption has emerged as one of the key prerequisites for successful trials in the post pandemic era. While most of these digital adoptions are patient centric and focus on their recruitment and adherence, the quality and credibility of the data reported in Clinical Study Report (CSR) is equally important.

A Clinical Study Report (CSR) is the documented evidence to confirm that the clinical trial has been planned, conducted, analyzed and reported as per the applicable regulatory guidelines and the reported data are accurate and credible. It is the mandatory regulatory component of the Marketing Authorization Application (MAA) for all pharmaceutical drugs and class-III medical devices.

The quality of the CSR has significant impact on pre-marketing regulatory review and marketing approval. The quality check (QC) process systematically verifies the data reported in the CSR against source data to ensure that the results presented are accurate, the statements are supported by references cited, and the conclusions are consistent with the results. A poor-quality CSR not only triggers additional queries from the regulatory authorities leading to delay, but can also negatively impact trial credibility. The CSR-QC is therefore an inherent and the most crucial step of the CSR preparation process to ensure quality and accuracy.

CSR Submission:
Regulatory Time Frame and Impact on the Industry

CSR submission has very stringent regulatory timelines. The Food and Drug Administration Amendments Act (FDAAA) of 2007 and a subsequent 'Final Rule' to the Act implemented in January 2017, requires sponsors of most US-regulated clinical trials to register and report results on ClinicalTrials.gov within 12 months of primary completion, irrespective of whether the results are positive or negative. This 'Final Rule' introduced stringent reporting requirements with fines of up to US$10,000 a day for non-compliance. [1] , [2]

As per the recent statistical report published in LANCET (February 2020), after the enforcement of 'FDAAA Final Rule' in January 2017, only 40.9% (1722 of 4209) of clinical trials during March 2018 to September 2019 were compliant with the one-year timeline. The median delay from primary completion date to submission date was 424 days, 59 days higher than the legal reporting requirement of one year. [3]

FDAAA Trials Tracker has further reported that the US government could have imposed fines over US$ 12.6 bn to the pharma companies for this non-compliance. [4]

Apart from FDA non-compliance issues, delayed trial reporting also results in delayed product launches. It has been estimated that pharma companies have incurred huge losses – average of US $6 Million per day – for delayed submissions.

Manual CSR-QC Process: Conventional Challenges

The lengthy and complex CSR-QC process is one of the major contributors to delayed reporting of trial data. The current manual QC process faces significant challenges in terms of time, cost, quality and resource management.

Some of these conventional challenges are:

- Time involved in a manual quality check of the complex study report involving extensive numerical data
- High chances of observational errors leading to a prolonged CSR review cycle time, which increases the overall turnaround time for document generation
- The cost of these CSR QCs can be exorbitant, and go as high as US$6000 per report.

There are several checkpoints for a CSR-QC, of which the source data verification of the numerical values reported in the In-text Tables versus Source Tables is the most critical and time-consuming checkpoint.

Phase III CSRs, for instance, contain several hundred in-text tables manually created by medical writer. These in-text tables contain extensive numerical data, which needs to be verified against several hundred of source Tables, Figures and Listings (TFL) generated by the biostatistician. This is the most crucial part of CSR-QC, which has a direct impact on the clinical trial outcome as well as on patient safety. Currently it is being done manually, which is not only time consuming but also prone to observational errors. There are no known automation tools currently available for quality check of numerical data of the in-text tables.

The Need For Intelligent Automation of CSR QC

As mentioned earlier, the current CSR-QC process already has a flurry of challenges with respect to the complexity, time and cost. Additionally, COVID-19 has added some extra bottlenecks.

Huge volumes of new CSRs on COVID-19 clinical trials (drug, vaccines, medical devices) are on their way. Till September 2020, there were total 5892 clinical trials registered for COVID/ SARS-CoV-2, of which 3370 trials are registered on clinicaltrials.gov and an additional 2,522 trials are registered on WHO International Clinical Trials Registry Platform (WHO ICTRP).

Going forward, a sudden spike can be expected in clinical trials, especially on non-COVID and COVID-comorbid conditions.

Stringent timelines with overlapping submission deadlines may cause delayed submissions causing prolonged time-to-market for pharma companies and / or loss of CSR-QC business to service providers such as Contract Research Organizations.

Limited availability of trained resources for CSR-QC will add to these challenges.

Any delay in submission of CSRs may cause financial losses to organizations as it cuts down the monetization period during drug patent validity

Thus, it's imperative to automate the CSR-QC process to expedite the entire submission process. However, considering the complexity of the document, only conventional Robotic Process Automation (RPA) with or without optical character recognition (OCR) may not be helpful. Intelligent automation leveraging cognitive technologies such as artificial intelligence (AI) and machine learning (ML) along with RPA and Advanced OCR can help expedite and transform this cryptic process.

Keeping 'hyperautomation' at the core, LTI is developing an innovative solution for automating the entire CSR-QC process by leveraging the right blend of digital levers like RPA, OCR and AI to speed up the end to end QC process, ensuring timely CSR submissions, thereby helping pharma companies avoid heavy penalties and losses.

CSR Quality Check Solution

Conceptualization

The idea of CSR-QC Bot was conceptualized considering the complex and repetitive task of source data verification of extensive numerical data, which is susceptible to observational errors and data discrepancies.

We conducted an extensive review of CSRs from 2010 to 2018 to understand the possible data extraction pattern for in-text as well as source tables.

We noticed that an efficient, rule based, machine learning driven cognitive search algorithm can extract the data patterns from the Source and In-Text tables and compare them to present a pictorial dashboard of for the analysis of CSRs. We realized that such an automated solution could potentially save time and effort in the overall quality check process. And this gave birth to the LTI's CSR-QC Bot.

Though most of the CSRs are created as per ICH-E3 template and follow applicable Medical Writing Style Guide (e.g. AMA Style Guide), the structure and style of in-text tables and source tables largely vary as per study requirement. This variation interrupts the functioning of conventional rule-based automation.

The source tables are in SAS generated RTF format without any table grid, which makes it difficult for the automation tool to extract relevant data.

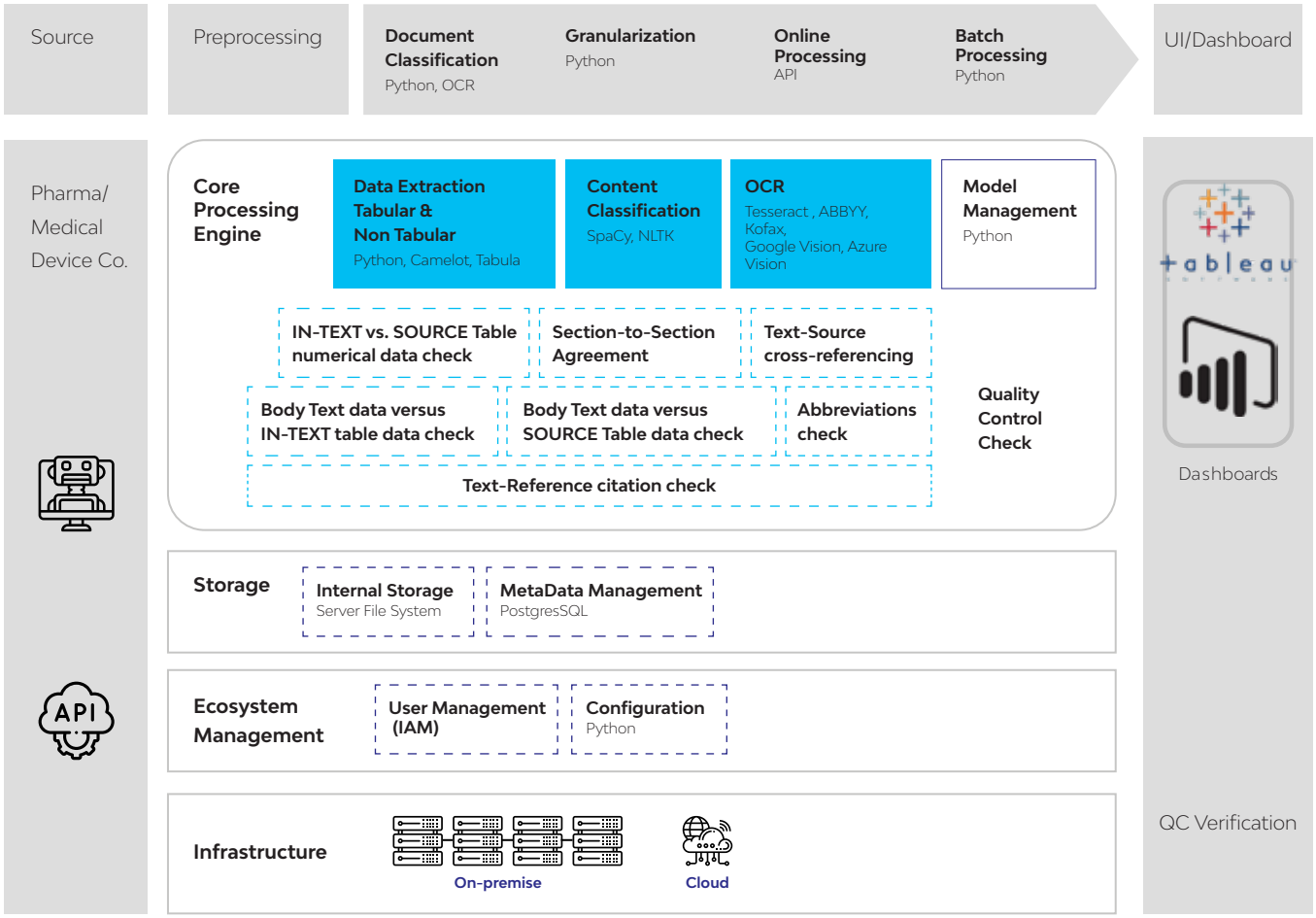
The complexity of the in-text tables further adds on to the challenges e.g. truncated in-text tables with certain values from the original source table; or multi-source in-text tables derived from more than one source table; or Shift in-text tables reporting logical value shifting from base line to various time points.

Apart from the data extraction challenges listed above, the inter-document spelling variation (American versus British spelling) is the biggest bottleneck for the tool to extract and compare the relevant data.

The conventional rule-based automation tools may not be competent enough to address those challenges. The solution needs to explore NLP driven cognitive algorithms to extract the data. Evolution of a rule-based algorithm will play a key role.

So How Does CSR QC Solution Work?

The CSR-QC Bot is envisioned to ingest documents via different channels like APIs, file servers, emails etc. Once a document is ingested in the system, it goes through a set of pre-processing activities like classification, granularization etc. Post this, the core processing engine works on the table detections, content classification, mapping of source data and actual QC process covering various scenarios. The solution also allows configuration of various setup information and other metadata details that are required upfront for the processing. This is also very useful to support multi-tenancy as required. The solution is also enabled to support rich visual interfaces for setup / configuration, validation, reporting, dashboarding etc.



How is the Model Trained?

Training of CSR-QC Bot is primarily based on the samples from the following document types and templates:

- **CSR Reports – various templates and functional variations**
- **In Text and Source reports**

Since the CSR-QC Solution is based on a cognitive machine learning based algorithm, it is imperative that more testing will improve the efficiency of the algorithm. More test data will build the machine learning database and will improve the predictability of the algorithm. Test data from various Clinical Research Organizations (CROs) and Pharmaceutical companies involved in Clinical Research play a vital role in fine-tuning the data extraction and comparison algorithm. This ensures that the algorithm identifies all the comparison errors in the reports.

Roadmap

In the current release of the solution, we've primarily targeted the effort intensive part of the CSR-QC process, which contributes to approximately 60% of the overall time i.e. the quality check of in-text vs. source data. The other functional aspects of the QC namely, Text Source Cross Referencing, Body Text data vs. Source / In-text data check, Citation checks, Abbreviations cross-references, Section-to-Section alignment etc. will be rolled out in subsequent releases.

An important aspect of CSR-QC Solution is its ability to connect to multiple, heterogeneous sources of data like PDFs, RTFs, proprietary formats etc. and effortlessly perform the comparison. It can reduce up to 70% of the average QC time per CSR for pharma companies resulting in significant cost savings and an increased throughput in

operations. Complete SLA adherence could lead to timely submissions thereby avoiding hefty penalties and loss to companies.

Apart from the core functionalities, we also intend to make it a holistic end-to-end integrated solution that caters to the complete workflow of the process. Also, from a hosting perspective, we would like to provide multi-tenant software as service model that can be subscribed by multiple CRO and Pharma companies as a cost effective alternative.

Conclusion

Automation as a stand-alone technology using levers like RPA has been in the market for more than half a decade now. A lot of standalone processes in the finance and HR verticals have been automated in these years. However, automation initiatives need to move up the value chain and discover the core industry processes within the business functions along with the horizontal functions. Compared to the regular use cases, these business specific use cases are cost intensive, as they demand time-consuming repetitive efforts. Automation of such use cases is the need of the hour. These use cases will need more than regular RPA like NLP, AI and ML and cognitive data extraction technologies to achieve end-to-end automation. CSR-QC is one such process that when automated, is bound to yield huge advantage to CROs and Pharma companies.

References

[1] https://jamanetwork.com/journals/jama/article-abstract/2553888

[2] https://clinicaltrials.gov/ct2/manage-recs/fdaaa

[3] THE LANCET. VOLUME 395, ISSUE 10221, P361-369, FEBRUARY 01, 2020

[4] http://fdaaa.trialstracker.net/

Authors:

Praveen Naregal

Associate Director - Intelligent RPA Practice, LTI

Praveen has 18+ years of extensive experience in product engineering, automation and delivery. He is involved in building multiple enterprise scale products & solutions for various industry verticals. He is passionate about building high-performing teams and solving complex customer problems by intelligent application of technology. He is currently responsible for delivery of intelligent RPA solutions to the Life Sciences, Consumer, Media & Technology, Manufacturing verticals.

Dr. Sachin Dighe

Senior Specialist - Business Analysis | Subject Matter Expert – Life Sciences, LTI

Sachin is SME for Life Science domain at LTI. He is a medical doctor with more than 15-year experience in pharmaceuticals research & development. His core areas of expertise are Clinical Trials (planning, conduct and reporting) and Regulatory Medical Writing & Quality Check.

Dr. Sachin drives LTI's digital transformation of pharmaceutical R & D. He is responsible for designing new digital solutions for accelerated drug discovery and clinical trials leveraging Artificial Intelligence, Machine Learning, Robotic Process Automation, and other advanced digital tools

Chandan Ghosh,

Automation Expert and COE Head - Intelligent RPA Practice, LTI

Chandan has represented LTI at different forums and winner of 'Automation Champion of the Year 2019' awarded by Global Sourcing Association (GSA UK & Ireland Professional Awards) for his Delivery Expertise, Service Innovation and Commitment in Robotics & Automation. Chandan has over 19 years of global IT experience, assisting marquee customers across major geographies, in their Large-scale Strategic & Transformational initiatives.

He has managed and delivered multi-million dollars in benefits by successfully implementing Automation Factory models & CoE roll outs in BPO & Shared Services for customers across Insurance, Manufacturing and Healthcare. Chandan's ability to visualize a solution with a potential to resolve industry specific problem help him standout.

LTI (NSE: LTI) is a global technology consulting and digital solutions Company helping more than 400 clients succeed in a converging world. With operations in 32 countries, we go the extra mile for our clients and accelerate their digital transformation with LTI's MosaicTM platform enabling their mobile, social, analytics, IoT and cloud journeys. Founded in 1997 as a subsidiary of Larsen & Toubro Limited, our unique heritage gives us unrivaled real-world expertise to solve the most complex challenges of enterprises across all industries. Each day, our team of more than 32,000 LTItes enable our clients to improve the effectiveness of their business and technology operations and deliver value to their customers, employees and shareholders. Follow us at @LTI_Global

info@lntinfotech.com