

Qurie – Our own SDTM Bot

Deepak Ananthan, Zifo RnD Solutions, Chennai, India
Arvind Sri Krishna Mani, Zifo RnD Solutions, Chennai, India
Rajesh Yovan, Zifo RnD Solutions, Chennai, India

ABSTRACT

Studies are becoming more complex in terms of design and SDTM submission is being made mandatory across different regulatory authorities. At surface level this would automatically mean a need for more SDTM experts. But if we take a closer look we realize that there is only some truth in this and lot of SDTM mappings can be automated by applying machine learning and natural language processing. There are lots of patterns in the data that can be taken advantage of. Based on this thought we designed a self-learning bot which reads CRF data and predicts SDTM mapping. Just like humans, this bot is designed to self-learn as and when new mappings are encountered. This bot will greatly reduce the time spent on SDTM mappings thereby allowing the experts to concentrate on complex mappings and increasing the number of studies they can handle in parallel

INTRODUCTION

Clinical data review is the vital part of any trial to evaluate a study drug. To make data review more efficient and accurate regulatory authorities including FDA have mandated electronic data submissions to be in adherence with Standard Data Tabulation Model (SDTM) created by Clinical Data Interchange Standards Consortium (CDISC). Converting collected clinical data into SDTM conformant data is cumbersome and requires substantial expertise, time and effort making it a prolonged and expensive process. It is because of these difficulties, sponsors usually wait till the end of the study to decide on creating the SDTM datasets or not.

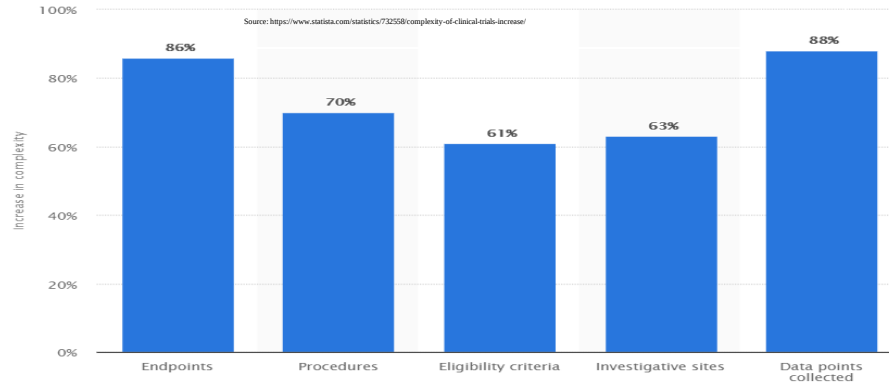
Naturally we would want to automate most, if not all of this process. Most of the existing SDTM automation tools deal only with automating the programming part and not the SDTM mapping. The crucial part that requires human intervention and SDTM expertise is the SDTM mapping. Qurie, our SDTM bot is designed to aid in automating the SDTM mapping for you. This paper will introduce Qurie to you.

WHAT MADE ME THINK OF QURIE

SDTM standard is not a new entity of the Clinical trials industry and has been around for a fairly long time. Ideally we should have conquered the learning curve and exponentially decreased the effort we spend on it. Despite the extensive literature in implementation, and a simple learning curve, the effort and time spent on SDTM mapping hasn't decreased by much. This stagnation in progress created a nagging curiosity within me and made me look deeper and understand the cause behind it.

The first obvious thing that came up when I was discussing with my colleagues is that the complexity of the clinical trials is going up all the time. And that is when I stumbled on this statistic that forever settled the debate on the increasing trial complexity. The following graph is a comparison of clinical trials complexity between 2001 and 2015 it can be safely assumed that the trend would have continued the same way. The % increase in some of the cases is quite staggering, especially with the endpoints, procedures and data points collected.

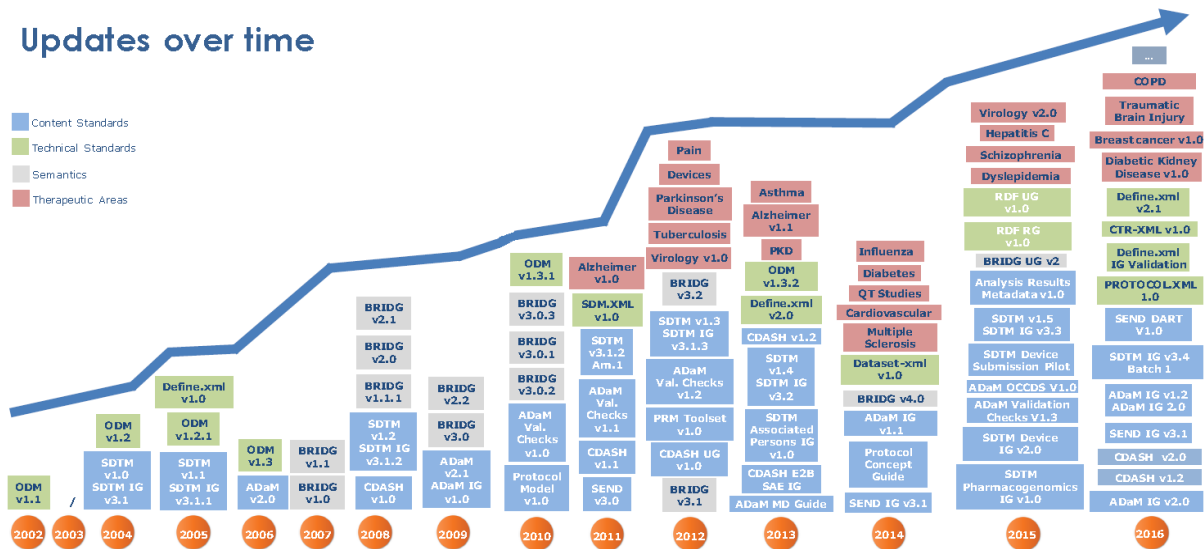
PhUSE US Connect 2018



Graph 1: Increase in clinical trials' complexity between 2001 and 2015

Another glaring question that was right in front of me was, even after years of working with standards, the industry has not been able to master this amazing set of ideas and concepts. This fact can be attributed to the fluidity of the standards, and how they have constantly been at pace with the changing complexity of clinical trials. As the complexity grew, the standards adapted, grew and evolved to the current CDISC standards that we have. The following graph shows a clear picture of how the number of standards published by CDISC has increased over the years.

Updates over time



Graph 2: Increase in number of new standards released by CDISC over years

Keeping pace with this evolution can at times be daunting, if not challenging.

Apart from the above there is also the increase in long term studies which brings in new set of problems. For example Oncology trials are commonly long term trials and we all know, on top of the increased complexity, they also undergo a lot of modifications in protocol and CRF designs. It is highly recommended to have the same SDTM programmer to work on this study through the end of trial. But practically it is very difficult to have the same programmer work on one study for years and using a new one every time will increase the time taken attributing to learning curve and knowledge transfer.

Likewise Sponsors generally conduct multiple trials on same drug, as they expect standardization of study data to be consistent across studies. In theory, when all these studies are converted to a single standard it should be consistent. But in reality, even when the same study is converted by different people within the same organization, there are always differences in the way the output is created. As there are little grey areas in the standard and differences in the interpretation of the standards by different people, there are always differences in the implementation and it is always the clinical programmer's decision to map CRF data to one the programmer feels appropriate in SDTM which results in inconsistencies across studies.

PhUSE US Connect 2018

Although SDTM makes the analysis of the clinical data and data review easier and effective, due to the time taken in implementation it is preferred only when it is being submitted to FDA. If we can reduce the dependency on the expertise and the time taken for the most critical data mapping part then SDTM data can be used right from the start of the study for any ongoing data review during the study conduct.

Unfortunately, because of all these reasons, even though we have been doing SDTM mapping for a long time now the time we spend on mapping a new domain still almost remains the same.

Sometime ago I stumbled on Sophia, a social humanoid robot. It's ability to mimic human conversations and think like humans got me thinking "How would it be if we have a bot for CDISC conversion?"

QUIE

This bot has the ability to read, analyze and predict the most accurate SDTM mapping for the collected raw data. One of the most fascinating aspects of this bot is its ability to remember the prediction it made for similar data and applying it when it encounters it again. In other words, just like how humans build up their knowledge by working on more studies, Quirie can learn by itself as and when it encounters new mapping. It can become better by increasing its accuracy and ability to predict upon every single use. In time it will become a companion for a clinical programmer reducing time the effort and time required to create a SDTM mapping for any study.

QUIE'S APPROACH

Quirie is designed to read CRF to predict SDTM mapping. It primarily uses the metadata and searches for a direct match in its SDTM knowledgebase. It does not stop there; it applies principles of natural language processing and finds all the potential matches for a single CRF field. The prediction among all the potential matches is made stronger by the repeated use in earlier studies for the same data. If there is a completely new CRF data then it seeks human assistance to learn. Once learned, the next time it encounters a similar CRF data, it will suggest the SDTM mapping that it learned. Quirie also learns when the user rejects its prediction so it also learns to understand the user's preference and needs and improve its accuracy of SDTM prediction.

WHAT CAN QUIE DO FOR US

Quirie, our SDTM bot with its exceptional abilities provides the best solutions to many of the challenges faced in SDTM implementation including the ones discussed above.

- Quirie is designed to automate the SDTM process which makes SDTM mapping possible with just a click from the User. This will free up SDTM expert's time enabling them to work on multiple SDTM studies in parallel.
- The fundamental function of Quirie is to automate transformation of CRF data to SDTM conformant data with just a click from the User.
- Quirie requires only CRF specification as input, which makes SDTM creation possible right after the EDC setup, i.e. even before the study goes live. This makes the interim analysis efficient and progress of a clinical trial more effective. So sponsor can decide on how to proceed with an ongoing study and make necessary amendments based on analysis results.
- Quirie, with its ability to learn like humans, remembers more than humans. It can be used anytime during the study. It does not matter when the study undergoes modification. With Quirie, SDTM mapping can be done instantly hence minimizing challenges faced on SDTM mapping due to study modifications at any time.
- Working with CRF amendments are made easy with Quirie as it can provide instant mapping for every new version of CRF. Programmers do not have to worry about the impact on existing mapping. Quirie can recognize and recreate the same mapping every time.

Quirie recognizes data from across studies using which it will provide consistent mappings every time it is used. This removes the necessity of having the same programmer work on all related studies to maintain consistency ensuring that consistency can be attained every time irrespective of the user. This plays a key role when SDTM conversion is necessary for several related studies or integration of legacy studies.

CONCLUSION

Quirie, right now needs a bit of handholding to take inputs from the user, learn and become familiar with consistent SDTM mapping. With enough data, Quirie can only become better in terms of efficiency, accuracy, and consistency. The future of Quirie would include it getting involved in every step of the SDTM process, where it will take the CRF datasets and convert them to completely conformant SDTM datasets without any human intervention.

PhUSE US Connect 2018

Once fully functional, it can lead the way for us to break the existing barriers of the SDTM transformation and lead us to the pinnacle of efficiency

REFERENCES

CDISC Japan Interchange Understanding the potential of Share - a visual business case and technical use case - 2016 (Graph 2).

CONTACT INFORMATION

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

Deepak Ananthan
Zifo RnD Solutions
21A, Anna Salai
Littlemount, Saidapet
Chennai / 600015
Work Phone: +91 44 43114002
Email:Deepak.a@zifornd.com

Arvind Sri Krishna Mani
Zifo RnD Solutions
21A, Anna Salai
Littlemount, Saidapet
Chennai / 600015
Work Phone: +91 44 43114002
Email:Arvind.m@zifornd.com

Rajesh Yovan
Zifo RnD Solutions
21A, Anna Salai
Littlemount, Saidapet
Chennai / 600015
Work Phone: +91 44 43114002
Email:Rajesh.y@zifornd.com