

PhUSE EU Connect 2018

Paper SI11

CDISC SDTM - An Automated Approach

Binitha Sathyabhama, Cognub Decision Solutions, Thiruvananthapuram, India
Sreedevi Menon, Cognub Decision Solutions, Thiruvananthapuram, India

ABSTRACT

The CDISC Study Data Tabulation Model (SDTM) is a mandatory standard for regulatory submissions of clinical trial data. This paper introduces an effective tool to automate the entire process of migration to SDTM standardization, including the creation of define.xml from legacy data. The automation is performed based on some functionality oriented meta-data and logic checks. The solution is characterized by an easy to use and simple interface that can be used by anyone. The robust architecture is designed to handle multiple studies irrespective of the study phase and therapeutic areas. This comes with a scalable system design that supports adapting to different versions of the CDISC standards very easily. The implementation of this automated approach aims at a significant reduction of the time and resource requirement for an SDTM migration task. This approach can be further extended to accomplish ADaM migration.

INTRODUCTION

This paper introduces an automated tool based approach to the migration of legacy data to CDISC SDTM standards. and is aimed at readers who are familiar with the SDTM mapping process. With FDA mandating submission of clinical trial data in CDISC compliant format, the scope of this approach is to devise a methodology that makes the process of data standardization faster, automated, error-free and cost effective. It focuses on stakeholders of the clinical research process who are involved in the CDISC migration and regulatory submissions to the US Food & Drug Administration (FDA) and the Japan Pharmaceutical and Medical Devices Agency (PMDA), and soon to be followed by the European Medicines Agency (EMA).

SDTM AUTOMATION

The traditional approach of migrating legacy data to SDTM standards involves different steps:

- 1) Developing the specification document, also called a mapping document. (usually in an MS Excel format in line with the guidelines)
- 2) Importing of raw data to SAS format (if the raw data is not in SAS format)
- 3) Generating the datasets using SAS programming techniques
- 4) Developing the define.xml document
- 5) Preparing reviewer's guidelines

Different approaches have been taken by different people to automate one or more steps in this process. Over the years, SAS macros have been developed to support faster generation of datasets. We have tools like Pinnacle 21 which provide an interface to generate the define.xml documents which once used to be generated via xml programming.

Here, we are aiming to integrate our software development and AI capabilities with our clinical research expertise to develop an intelligent system that automates most of the above-mentioned steps. The system consists of a simple user interface wherein the raw data and relevant documents like the CRF, implementation guidelines etc can be uploaded to the system and a specification document with automatically configured specification data and algorithms is generated. With a minimal amount of effort from the specification developer, the mapping file can be finalized. The dataset and define.xml generation is taken care through backend processing and generated once the specification document is finalized.

A built in QC and QA process is also planned to be incorporated into the system. Further, all the technical controls for 21 CFR Part 11 compliance has been built into this system.

PhUSE EU Connect 2018

The data flow diagram explains how the system proceeds through the sequence of automated steps to generate the final data for submission.

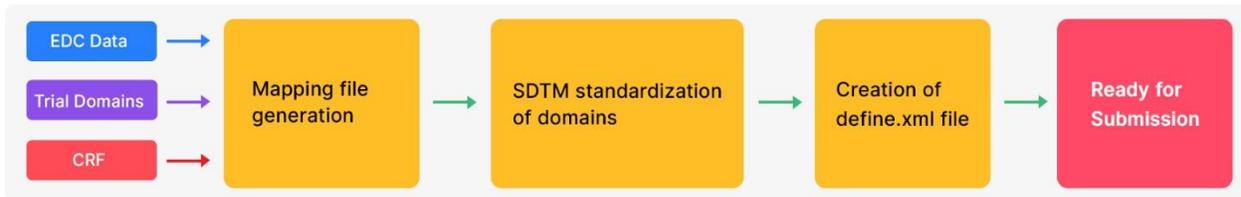


Figure 1. Data Flow Diagram

FUNCTIONALITY

The system under discussion envisions the automation of the various steps involved in the CDISC SDTM migration effort. A set of controls, both from technology and procedure aspects have been incorporated to meet the closed system concept of CFR Part 11 requirements for data privacy/protection. Here, we are trying to elaborate on the various functionalities supported by this tool and how it will help in revolutionizing the process of data standardization.

USER PROFILE, ROLE BASED ACCESS & SECURITY

The system has a provision to add studies, study details and users with different privileges in line with the information flow structure practised in the clinical research industry. Proper authentication procedures are incorporated wherein only a designated authority of an organisation will be able to add studies, users and also assign the users to the different studies in respective roles. Various roles are incorporated that help in ensuring that the standard requirements related to quality and process have been met. A user friendly interface further helps in easy communication between the various roles. The application will be hosted on a private cloud environment with multi-level authentication checks to ensure data / user integrity. Audit trails help backtrack user activity, maintain access logs, mitigate risks and help device fallback mechanism in need. The production versions will have geographically dispersed cloud mirroring capabilities, enabling disaster management and near to 100% uptime at all scenarios.

SPECIFICATION DOCUMENT GENERATION

Specification document generation is one of the most critical steps in the CDISC migration process as the dataset and define.xml generation is basically an execution of the specification document. It requires good knowledge of the implementation guidelines, the study protocol and a good amount of insights to commonly used best practices to be able to develop a comprehensive mapping file. Such aspects, thus, made it quite challenging to automate this step.

PhUSE EU Connect 2018

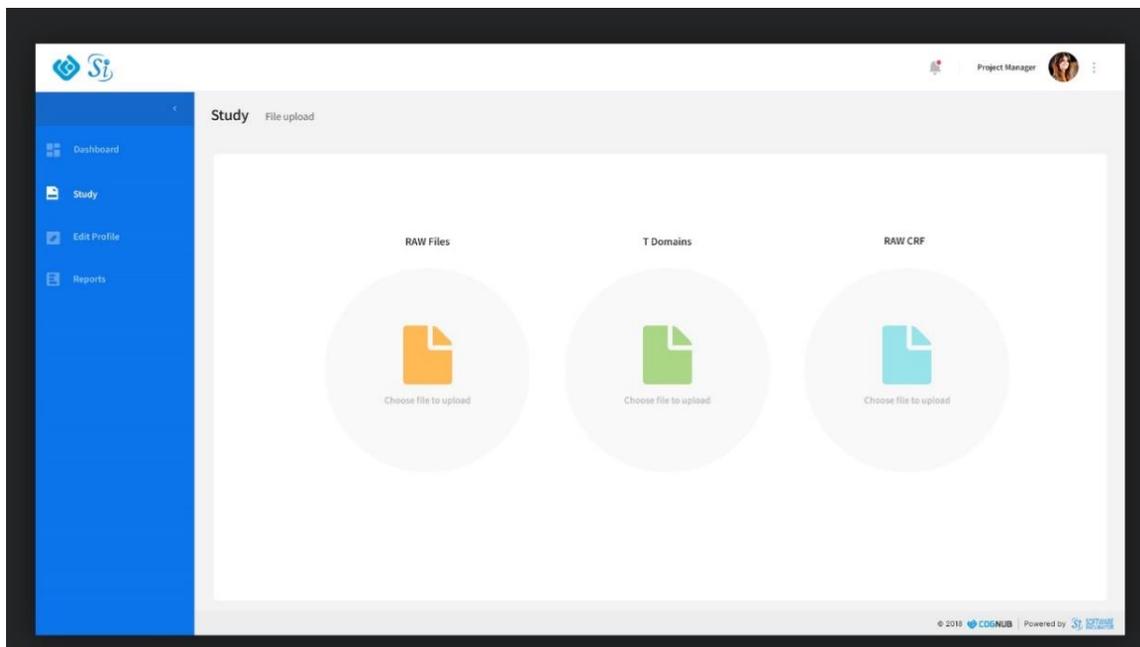


Figure 2. The Input Files Upload Interface

The tool gives a provision to upload raw data files, trial domains and the raw CRF into the system and generates a specification document/mapping file with about 50 - 80% of information already configured depending on the complexity of the study.

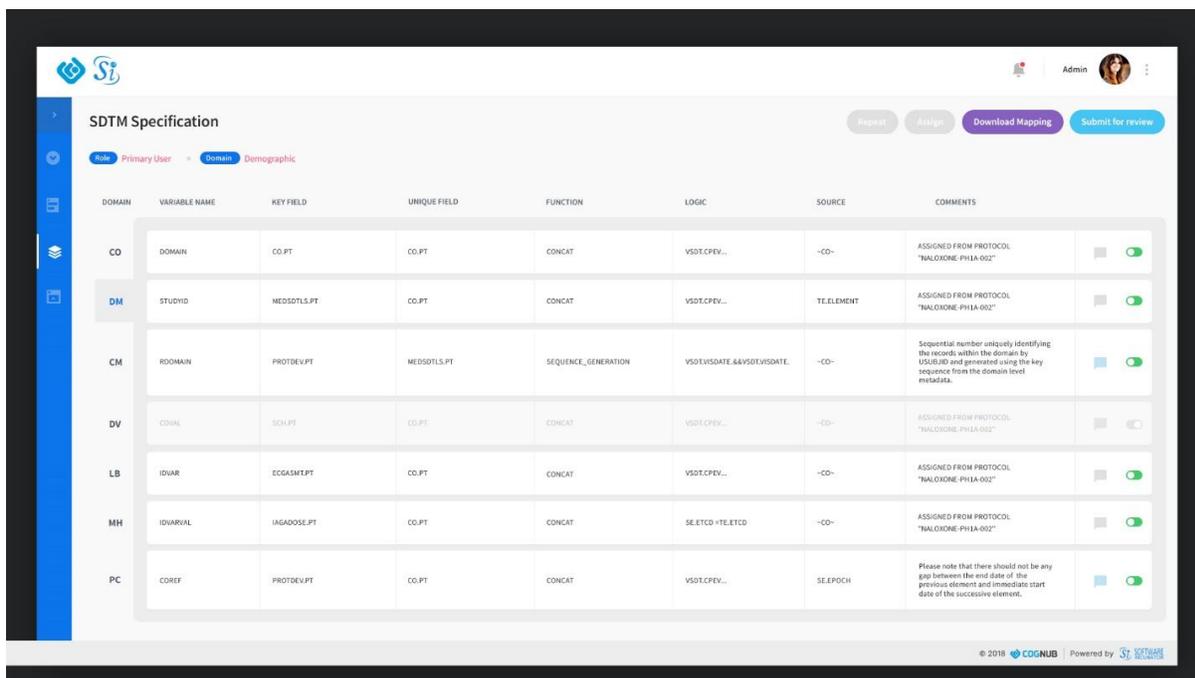


Figure 3. SDTM Specification Window

The specification document can be completed with minimal human intervention of editing and re-assigning of variables. Using the built in quality control modules, the document can be passed through an iterative QC process before finalising the same. Even though the interface handles all the domains at one go, it allows working on separate domains in parallel by different users, based at different locations.

PhUSE EU Connect 2018

The validation stage is managed in a strategic way to allow the user to review the specification both electronically and manually. The tool has the provision to notify the user about any conflict in the developed specification and allow user to update the specification to resolve conflict.

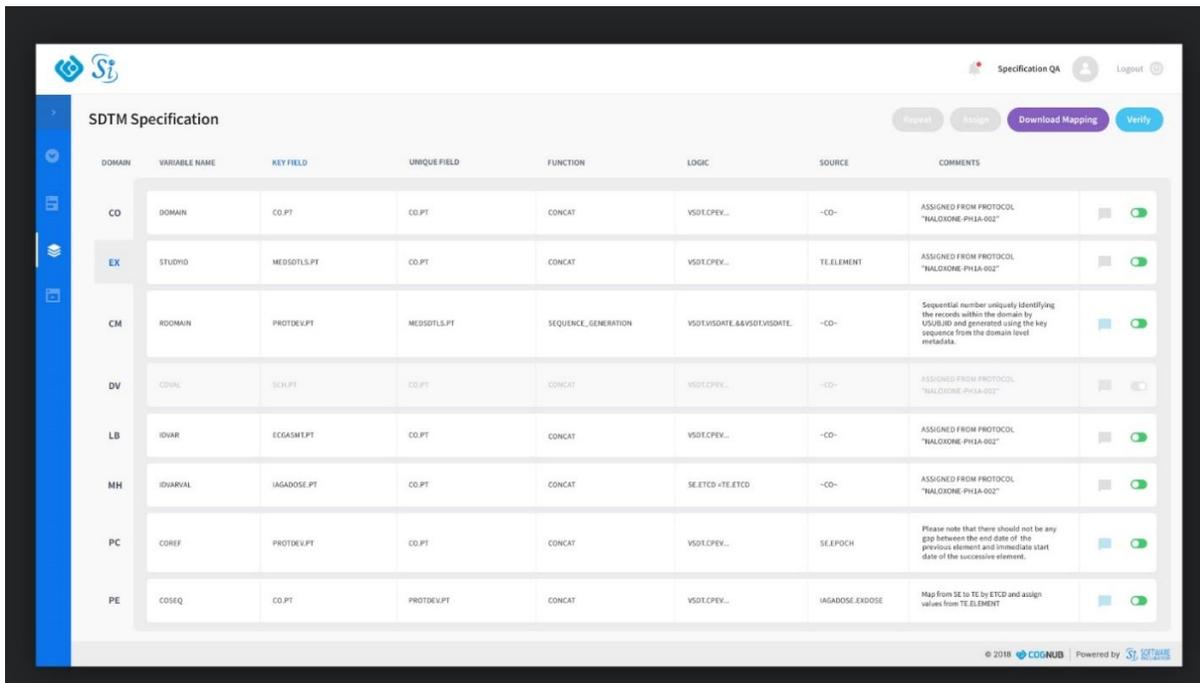


Figure 4. SDTM Specification Verification Screen

The QC module with notification feature has been designed such that communication between the developer and QC person happens from within the system and all this is automatically stored in retrievable format to meet the documentation requirements from a process audit perspective. The approval on any matter (if required) is also handled from within the system. A chat bot based approach as shown below helps in easy communication.

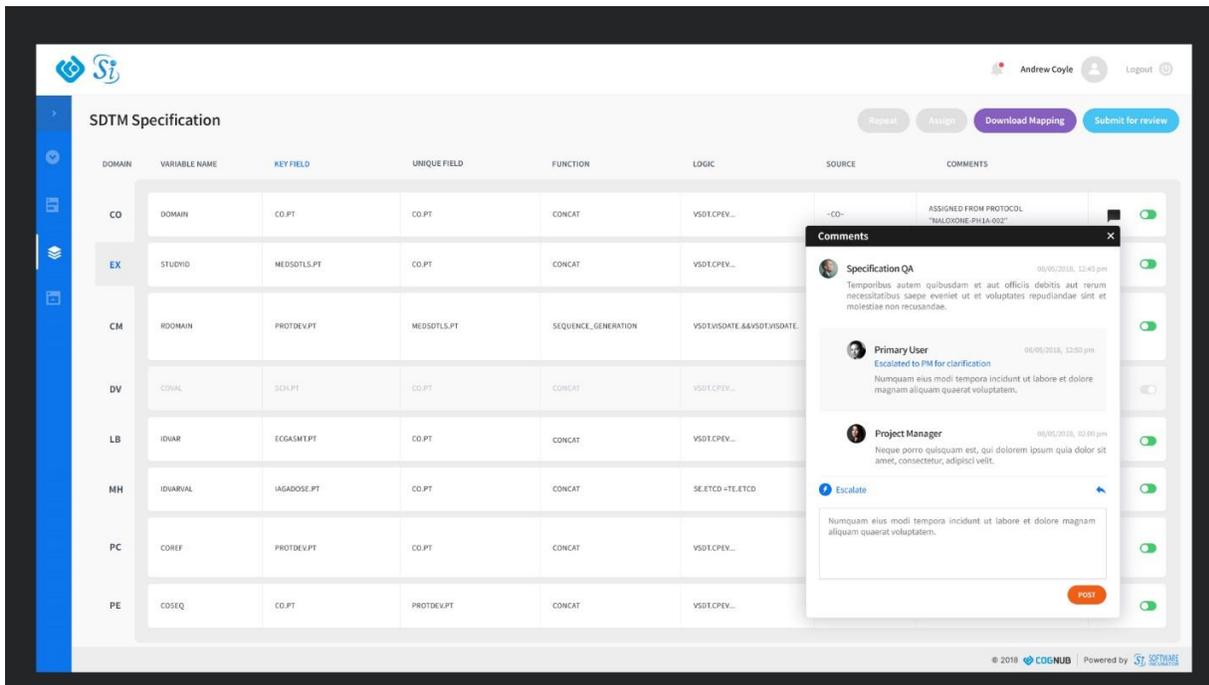


Figure 5. The Specification QC Comment Screen.

PhUSE EU Connect 2018

CRF ANNOTATION

The later versions of the tool will have an automated system in place to perform the annotation of the CRF. Our machine learning expertise will be employed to fine tune this module such that the annotation is performed with the help of a predefined algorithm when a blank CRF is provided as input.

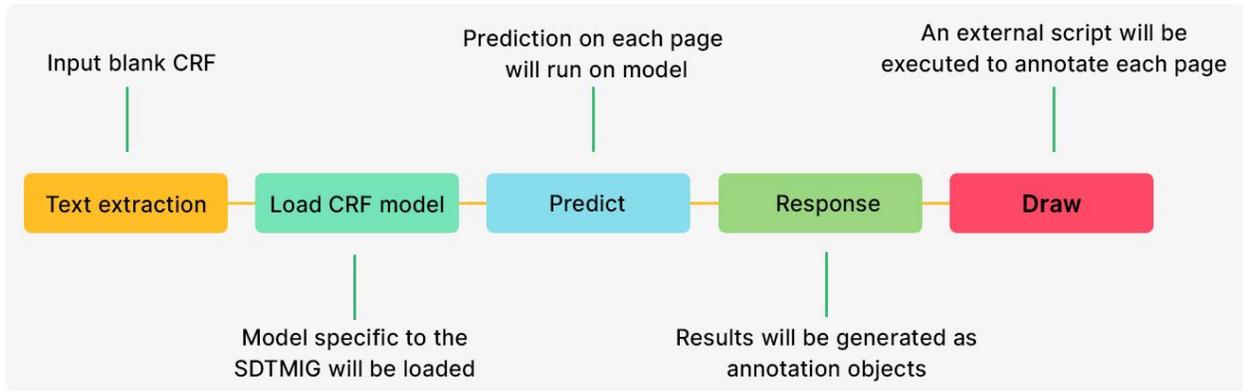


Figure 6. The CRF Annotation API

INITIALIZE THE STANDARDIZATION

Once the QC comments are incorporated and the specification is approved by the reviewer for all the domains, the specification is considered ready to be moved to the next step, i.e. generation of the datasets or in other words “initialisation of the standardization process”.

The generation of the datasets happens as a backend process. On the click of a button the datasets are generated in a downloadable format with reference to the specification document, controlled terminologies and raw data.

DOMAIN	STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTM	RFENDTM	RFXSTDTM	RFXENDTM	RFLCDTM	RFPENDTM	DTHDTM	DTHFL	SITEID
CD	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02019	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02020	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
EX	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02031	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02032	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
CM	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02033	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02034	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
DV	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02035	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02036	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
LB	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02037	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02038	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
MH	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02039	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02030	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
PC	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02031	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02032	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
PE	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02033	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR
	RD14 IFRS PIQC-002	DM	RD14 IFRS PIQC-002 VACR-02034	2019	2014-12-02	2014-12-19	2014-12-02T08:00:00	2014-12-18T08:00:00	2014-11-17T08:28:00	2014-12-22			VACR

Figure 7. Standardized SDTM Data

PhUSE EU Connect 2018

The standardized data is also created as an xpt file which forms input to the generation of the define.xml document. This again happens as a back end process. The define.xml file that is thus generated can be passed through the different validation processes to ensure the quality of the same.

DOCUMENTATION AND AUDIT REQUIREMENTS

The system is developed such that all the technical controls as per the CFR Part 11 compliance is built in. Further, the process oriented approach adopted to meet the workflow makes the system and its outputs audit ready from a process audit perspective.

CONCLUSION

In summary, implementing an AI based software solution to automatically create a standard SDTM domain and define.xml for submission, helps users to identify and resolve issues early on, without impacting development workflows or delivery timelines. This is a better and more intelligent approach towards cost containment with better quality deliverables.

REFERENCES

www.cdisc.org

Study Data Tabulation Model Implementation Guide (SDTMIG): Human Clinical Trials, Version 3.2

ACKNOWLEDGMENTS

I sincerely thank all my colleagues who reviewed the paper and provided valuable comments and suggestions. Additional gratitude goes to my colleagues at Software Incubator for working through the details of implementing CDISC/SDTM standards.

CONTACT INFORMATION

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

Binitha Sathyabhama

Cognub Decision Solutions Pvt. Ltd.

T4, 7th Floor, Thejaswini Building, Technopark

Thiruvananthapuram - 695 581, India.

Work Phone: +91-471-2527640

Email: binitha.s@cognub.com

Web: www.cognub.com

Sreedevi Menon

Cognub Decision Solutions Pvt. Ltd.

T4, 7th Floor, Thejaswini Building, Technopark

Thiruvananthapuram - 695 581, India.

Work Phone: +91-471-2527640

Email: sreedevi.menon@cognub.com

Web: www.cognub.com

Brand and product names are trademarks of their respective companies.