Paper SD02

Automated SDTM creation and discrepancy detection jobs: The numbers tell the tale

Joris De Bondt, SGS Life Science Services, Mechelen, Belgium

ABSTRACT

The FDA is advising use of data standards as early as possible in the study lifecycle. As a result, Data Management centers are using the Study Data Tabulation Model (SDTM) to drive operations from First Patient In till Database Lock. Many tools on the market allow for the creation of SDTM datasets via intuitive user interfaces. However, targeted tools are needed to manage nightly jobs taking care of data source downloads (eCRF, ePRO, Lab, etc), data uploads in a staging database, converting to SDTM and running edit checks before the Clinical Data Manager arrives in the morning. Each step requires complete error handling (missing data sources, network connectivity issues) to guarantee a database ready for use by the Data Manager. Visualising the performance metrics of the nightly jobs in intuitive graphs is key for spotting and resolving glitches, allowing for a well-tuned database environment that hosts a large set of small phase I to large phase III trials.

SETTING THE SCENE

eCRF/eSource and CDISC standards have offered numerous benefits to the Clinical Research industry in the past decade. The Clinical Data Interchange Standards Consortium (CDISC) finally made global, platform-independent data standards available to the industry, providing invaluable time gain throughout the acquisition, exchange, submission and archive of clinical research data and metadata. eCRF/eSource are now a widely accepted alternative for paper trials, and have numerous advantages which include: immediate access to both trial data and trial status reports via the internet, improved data quality by auto-queries that are sent by the EDC system as soon as data is entered, and dynamically triggered items / forms / visits / treatment arms.

At the same time, almost unnoticed, Data Management centers evolved from using Clinical Data Management Systems, allowing only entry of paper study CRF data in a fixed proprietary data format, towards fully automated platforms supporting data coming from a wide range of eCRF/eSource vendors, as well as ePRO, laboratories, imaging centers, and any other data coming from external vendors or third parties. Since downloading data for large phase III studies can take more than an hour, it is essential that this task is running overnight to allow a refreshed database when the study team is entering the office in the morning. Network connectivity glitches are unfortunately still a reality today, and should be handled via multiple download retries to avoid a disappointed study team because of a non-refreshed database. Of course, IT-staff could monitor the entire process flow overnight. However, for small to mid-size CROs having overnight-staff is often not an option; so smart systems are in place to handle potential errors instead. On top of that, performance management is key for long running studies; conversions to SDTM are tested based on a limited number of test patients, while the actual study will have often 100 times more data! Since all processes are running overnight, error logging combined with detailed performance metrics are the only tool for the IT-staff to look at when investigating process flaws.

CONTACT INFORMATION

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

Author Name: Joris De Bondt
Company: SGS Life Science Services
Address: Generaal De Wittelaan 19 A bus 5
City / Postcode: Mechelen / 2800
Work Phone: +32 15 29 93 40
Email: joris.debondt@sgs.com
Web: www.sgs.com/cro

CONVERSION CENTER

A central tool to manage the end-to-end flow for refreshing the Data Management study databases unattended overnight

for all kinds of data

with complete visibility every step of the way

DOMAIN	START_DATE	END_DATE	DURATION	STATUS
CLINTRIAL	2014-08-25 04:44	2014-08-25 04:48	00:03:21	FINISHED
A2O_SYNC	2014-08-25 04:40	2014-08-25 04:44	00:04:48	FINISHED
REPORTS_PAGES2LOCK	2014-08-25 04:39	2014-08-25 04:40	00:00:03	FINISHED
CHECKS	2014-08-25 04:30	2014-08-25 04:32	00:01:19	FINISHED
REPORTS	2014-08-25 04:30	2014-08-25 04:39	00:09:12	FINISHED
METADATA	2014-08-25 04:30	2014-08-25 04:30	00:00:00	FINISHED
AFTERCONV	2014-08-25 04:27	2014-08-25 04:29	00:02:50	FINISHED
CO	2014-08-25 04:24	2014-08-25 04:26	00:01:54	FINISHED
RELREC	2014-08-25 04:14	2014-08-25 04:24	00:10:45	FINISHED

allowing for targeted performance monitoring and tuning from day 1 till database lock