

Written and researched by:
Dr. Kirk Mousley
and Karl Mousley

In the next issue of
EDC Today:

Mobile Computing -
Understanding
the Basics

About EDC Management:

EDC Management is the leader in Clinical and Data Management and Electronic Data Capture (EDC) consulting services for the biopharmaceutical industry. EDC Management publishes well-researched and timely information about Electronic Data Capture technologies and processes through EDC TodayTM and EDC In Depth. We do not sell or endorse any specific EDC software application or vendor. Improve process today; position for tomorrow.

EDC Management

P.O. Box 384
Conshohocken, PA 19428
484-530-0300 (voice)
610-567-0357 (fax)
info@edcmanagement.com
www.edcmanagement.com

ECS and SAS: How They Interface

EDC Today is an independent publication on current information and issues in Electronic Clinical Systems (ECS) strategies and technologies for the Biotechnology and Pharmaceutical (Biopharma) industry. Each month we examine topics related to ECS theory, technology, practice, or implementation.

Recently, a number of readers have requested that we write about ECS (which consists of Electronic Data Capture (EDC), Clinical Data Management Systems (CDMSs), and SAS along with other components) and how it has impacted what Biopharmas expect from SAS. SAS was at one time an acronym for 'Statistical Analysis System' but has since entered the lexicon on its own right, referring to the array of statistical analysis software marketed by the company also named SAS.

In this issue, we discuss ECS and SAS and how they interface. In the past, a boundary separated the clinical database and the data used for statistical analysis and reporting. We investigate whether or not this separation is still required given recent technological changes in the clinical trials data arena. We also investigate claims that a Biopharma only needs SAS to manage its clinical trials.

Introduction

Oracle's Relational Database Management System (RDMS) has been used for nearly 30 years by Biopharmas – it is the most widely used repository for clinical data. SAS has been in use for about as long - its primary purpose is for statistical analysis and presentation (i.e., "reporting") of clinical trial data and study results.

Traditionally, CDMS's have been built on Oracle's RDMS (e.g., Phase Forward's Clintrial, eResearch Technology's EDM, Oracle's Oracle Clinical), and SAS has been used on data "extracted" from Oracle. There are a number of reasons for this, chief amongst them is that early versions of SAS had no real data input mechanism, no data security (e.g., Oracle roles), no built-in data management functions (e.g., Adverse Event verbatim autoencoding), no transactional ability (i.e., rollback of changes), and no database recovery mechanisms. Therefore, SAS was not a good choice for either data capture or database management. Likewise, Oracle was not a good choice for statistical analysis and reporting since it didn't provide easy-to-use tools for these purposes.

(continued on page 2)



While fundamental technical differences once existed and still exist today, some have begun to question whether these differences are beginning to blur due to continued changes in technology. Figuring prominently in this discussion is the development of new clinical data capture, trial management, and trial analysis and reporting applications; the most significant being EDC, Data warehousing, and Portals.

In loose terms, EDC and Portals can be thought of as software module(s) that lie between investigator site personnel and the sponsor's or EDC vendor's RDMS, which is almost always Oracle! SAS can be thought of as software module(s) that lies somewhere between the RDMS and the FDA reviewer. Just where it is positioned is largely a matter of configuration and implementation – both of which are heavily dependent on the Biopharma's business practices as well as practical considerations regarding the strengths and weaknesses of the software being employed.

Recently, SAS has made significant changes to its software, and introduced “SAS Drug Development”. This latest, pharmaceutical-slanted offering from SAS has an enhanced single “point of secure access” (i.e., a Portal) that allows users to view disparate data sources — along with journal articles, “competitive analyses” (e.g. both the report code that is used to generate reports and the results of the reports), and other documents — from a more intuitive user interface. The software also includes significant additions to its support for “21 CFR Part 11” regulatory compliance.¹

So now you can have a “front-end” portal (i.e., with EDC) and a “back-end” portal (i.e., with SAS). You may now be wondering what this means in terms of the boundary (or interface) between the two.

Why Is SAS So Important?

SAS is the de facto standard for producing data analysis and transporting data within the Biopharma industry, as well as the definitive standard for submitting data to the FDA.² SAS is used in 40,000 customer sites worldwide by some 3.5 million users.³

According to SAS:

“SAS software is widely used for analyzing and reporting clinical trials data, and since 1999 the FDA has specified the use of SAS file formats for data sets included in electronic regulatory submissions to the FDA. Nearly 600 pharmaceutical companies and divisions worldwide use SAS' wide variety of solutions for the life sciences market, including Bristol-Myers Squibb, Eli Lilly and Novartis Pharmaceuticals.”

And they have recently concluded a deal with the FDA:

“The agreement with the FDA enhances a five-year, \$21 million enterprise license signed in 2000 by the U.S. Department of Health and Human Services.”⁵

So it is clear that for nearly every Biopharma, having a SAS license is, for all practical purposes, required.

(continued on page 3)



Interface: Reporting Database

The RDMS database is used to store data captured from “live” (or ongoing) clinical trials. This database is usually fairly “normalized” (See sidebar on “Normalization”). It holds some unclear or even inconsistent data and its contents are changing as the clinical study progresses. Furthermore, it may even be undergoing structural (e.g., new, modified, or deleted table columns) or definitional (e.g., codelist additions) changes. All of these attributes tend to make both the writing and generating of reports used for analysis and presentation difficult.

Sidebar – Normalization

A normalized database eliminates duplication (i.e., functional dependencies) in the data so that updating the database is easy and efficient. A fully normalized database would have many small tables, and any reporting query would require several joins. These many relations and joins make updating the data much easier, but they slow down most queries.⁶

So traditionally, “reporting database(s)” have been created. Data is extracted from the CDMS (which usually sits on Oracle tables) into files easily readable by SAS (usually SAS Datasets) using the tool provided by the CDMS (e.g., Clintrial “Retrievals”). The extraction process might consist of a “data dump” or it may be more elegant and provide for a number of desired data transformations, derivations and reorganizations such as de-normalization, creation of items (values) not stored in the RDMS and regrouping of data. SAS itself is then usually used to complete the extraction or transformation process, further transforming, deriving and reorganizing the data into a form that lends itself to statistical analysis, listing, graphing and other forms of reporting.

Creating a “reporting database” is done to meet a number of demanding requirements. One reason is to create a “snapshot” of the database at a particular point in time in the lifecycle of the study (e.g., interim analysis) as well as to make a storable/retrievable archive of the data used for reporting. Another of these requirements is the aforementioned transformations (e.g., standardization of units), derivations (e.g., true “date” types, e.g., 12-DEC-1997, from date pieces, e.g., ‘12’, ‘DEC’, and ‘1997’), and reorganization (e.g., “flattening” the structure of relational data, a process called “denormalizing”) which can greatly ease the report programming and generation process.

Additionally, sub-setting of the data can be done for ease of reporting and to provide a “stable” (and with planning) “clean” database for reporting programmers writing new programs. Furthermore, a single program that performed any desired transformations would need to be validated once and run once to create the record(s) subsequently used in analysis and reporting instead of that code being embedded in possibly a large number of analysis and reporting programs, executing each time an analysis or report was performed or generated. Thus one can gain in efficiency in two ways: less work in program validation and less processing work at analysis and reporting time. Lastly, data stored in Oracle is not easily transferable (or portable) to the FDA for the purposes of a submission whereas SAS has a number of simple transport formats including one recommended by the FDA for use with New Drug Applications (NDAs) and Biologic License Applications (BLAs).

(continued on page 4)



What isn't clear to us at EDC Management is exactly how records electronically signed in "live" RDMS databases should retain or gain their signatures when they are transformed into SAS dataset records or in data included in a NDA/BLA submission to the FDA. The FDA states that it is "...developing procedures for archiving documents with electronic signatures. Until those procedures are in place, documents for which regulations require an original signature, such as certifications, must be accompanied by a paper copy that includes the handwritten signature and the submission identifier..."⁷

We believe it is reasonable to think that the person that creates the record is the person of "authority" and whose signature should initially be associated with the record and that the audit trail should start at this point for this record. Therefore, it is our belief that the person of "authority" at the laboratory would electronically sign records created at a central laboratory. When, and if, these records are "imported" into a CDMS, the person doing the importation would assume authority and their signature would be affixed to the newly created records. When data within records are extracted and imported into a SAS database, once again, the person doing the extraction and importation would assume authority and their signature would be affixed to the newly created records.

Why So Many Databases?

In Issue 11 of *EDC Today, Technology Old and New: Integrating Legacy Systems with EDC* we discussed a number of clinical trials databases and their nature. They included EDC, CDMS, serious adverse event (safety), central laboratory, assay tracking, patient tracking, CRF page tracking, study agent (drug) accountability, autoencoding dictionaries, codelists, investigator site accounting and tracking information and finally, one or more submission databases.⁸ What should be apparent is that there are a number of diverse clinical trial-related information sources.

With their Drug Development offering, SAS positions themselves as a data warehousing provider. With one application, they strive to gather up into a single point of use, all the data in the clinical realm. Such is the promise of a "data warehouse". A data warehouse is a concept that has been around for some time, but has been seen relatively little in the Biopharma arena. While implementations vary, the warehouse is basically a consolidated copy of aggregated, sometimes summarized, transaction and non-transaction (e.g., codelists) data specifically structured (i.e., optimized) for query and analysis.⁹

Doesn't this sound like the old traditional "reporting database" to you?

Using only SAS

Occasionally, Biopharmas ask, "Why not just use SAS? After all, Oracle and CDMSs seem expensive to acquire, implement and use?" The "its expensive" argument is fallacious – a full cost-benefit analysis of the use of a CDMS will undoubtedly demonstrate the worth of such a system, but even if we should choose to ignore this, it should be apparent that a CDMS is a "core" application to a biomedical department, and that even if the software should cost upwards to a million dollars or more, the first successful NDA/BLA stemming from its use will more than pay for its cost.

(continued on page 5)



One could “just” use SAS but not even SAS is suggesting anyone do this. SAS positions their product as a data warehouse and statistical analysis and reporting application. It should be obvious that by using “only” SAS, one will be trading off the data capture optimizations and clinical trial management functionalities built into a CDMS and EDC product (e.g., discrepancy management) for the apparent benefits of a monolithic workspace and programming environment. The requirements that drive the creation and maintenance of both a “live” clinical trials database and a reporting database remain, and by removing the “physical” boundary between the two, one might have trouble creating and maintaining the virtual boundary.

Conclusions

Oracle’s Relational Database Management System (RDMS) and SAS have been used for nearly 30 years by Biopharmas. SAS has become the application of choice for statistical analysis and presentation (i.e., “reporting”) of clinical trial data, and study results, even with the FDA.

SAS has played and continues to play a very important role in the drug development process and new updates to SAS products offer easier data access and analysis¹⁰, as well as making it easier for a Biopharma to be in regulatory compliance.¹¹

The reasons for having a CDMS and SAS remain. There remains a need for separate data management and data reporting databases. New advances such as EDC over the web, improved communications and collaboration due to portals, and improvements in data warehousing have not made the “live” clinical database and “static” statistical database obsolete.

References

¹ <http://www.sas.com/industry/pharma/develop/index.html>

² *Ibid.*

³ <http://www.sas.com/start/exec.html>

⁴ <http://www.sas.com/news/feature/12may03/fda.html>

⁵ *Ibid.*

⁶ <http://www.4guysfromrolla.com/webtech/042699-1.shtml>

⁷ <http://www.fda.gov/cder/guidance/2867fnl.pdf>, page 12.

⁸ EDC Today™, Issue 11, “Technology Old and New: Integrating Legacy Systems with EDC”, pp 2-3.

⁹ Ralph Kimball, “The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses”, page 310.

¹⁰ http://www.sun.com/br/lifesciences_902/feature_sas.html

¹¹ SAS Drug Development — A solution for addressing 21 CFR Part 11 compliance; <http://www.sas.com/apps/whitepapers/whitepaper.jsp?code=O15>



Who's behind the research?

Our lead researcher, Kirk Mousley, PhD received BS and MS degrees in Electrical Engineering from MIT and a PhD in Computer Science from Lehigh University. He has been the President of Mousley Consulting, Inc. since its founding in 1993 and has directed the company's efforts in the areas of clinical database design, data editing/cleaning, document management, and submissions.

Karl Mousley received his BS in Mechanical Engineering from Rose-Hulman Institute of Technology and a MS in Computer Science from Villanova University. He has been a senior member of the technical staff at Mousley Consulting, Inc. since 1993. Among his significant accomplishments are the investigation, evaluation, and implementation of new computer technologies for clinical data management systems and developing strategic plans for integrating these technologies into current systems. He has extensive experience preparing Standard Operating Procedures (SOPs).



EDC Today and EDC In Depth

EDC Management publishes well-researched, timely information about EDC technologies and processes.

EDC Today is a free electronic technical bulletin.

Each month we examine topic areas related to Electronic Clinical Systems (ECS) theory, technology, practice, or implementation.

Each *EDC In Depth* research report comes with an executive summary and may be purchased individually for \$395 or as a group of related reports for \$975. Available via downloadable electronic version or paper version sent via mail.

To subscribe to ***EDC Today*** or purchase a specific ***EDC In Depth*** research report:

Order online at
www.edcmanagement.com

Email us at
info@edcmanagement.com

Call us at
1-484-530-0300