

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

In the next issue of
EDC Today:

Improving
Communications

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 Today® 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

Data Marts

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, some of us here in EDC Management were involved in an effort to set up data entry systems where one of the major design considerations was integrating those systems with the corporate Data Mart (and using standards evolved with the Data Mart in mind). Being curious, we investigated the reasons why the Data Mart was formed and what its intended uses were. What we learned was food for thought.

In this issue, we discuss what a Data Mart is and explore some of their potential applicability to Biopharmas and specifically how data collected during clinical trials might be used within a Data Mart. We will consider the possible value of a Data Mart beyond that of using one to assist data standardization efforts, which may well be its biggest return on the investment required to establish a mart.

Introduction

Computer applications (commonly known by the overworked term of “computer systems”) fall into two basic, broad classifications: Operational Systems and Informational Systems. Operational Systems are those used to run and conduct business, for example an accounting and payroll application. On the other hand, there are other functions that go on within the enterprise that have to do with planning, forecasting and managing the organization. The computer applications that support these functions are generally known as Informational Systems.¹

The distinction is not terribly important but does help one understand the definition of Data Mart. A Data Mart², also loosely known as a Data Warehouse³, is a repository formed from multiple data sources for “mining” use. Mining is the process of finding “value” in the data or making information out of data. Therefore one can think of a Data Mart as an Informational as opposed to an Operational System. Furthermore, the “multiple sources of data” used to create the Data Mart repository tend to be data stored within Operational Systems.

(continued on page 2)



What Use Is A Data Mart In The Clinical Arena?

The fact that data stored in clinical data management systems is not Operational in nature raises a troublesome warning flag to us at EDC Management. Putting this aside, we begin by looking into the clinical trials world and contemplating the formation of Data Marts using all the data collected by all the clinical trials run by a Biopharma. The potential uses for these Data Marts are shown in Table 1.

Table 1. Uses for a Clinical Data Based Data Mart

1. Create Patient Registries to locate subjects eligible to enter a new clinical trial/study. One might look for subjects eligible for new studies amongst the patients that were screened or even previously enrolled in other studies. Unfortunately, as nice as this might sound, privacy concerns like those posed by The Health Insurance Portability and Accountability Act of 1996 (HIPAA) make this usage unlikely.
2. Look for unexpected therapeutic reaction(s) and/or effects. In particular, a reaction not noticed until patient population was large in statistical terms. A sufficiently large patient population might be formed as in the result of conglomerating all clinical trials for a drug, class of drug, or drugs of some characteristic that allows sensible “grouping”.
3. Look for unexpected safety issues. For instance, combinatoric effects of medications, not noticed before the creation of a large enough data set.

A Data Mart need not be strictly confined to the data captured during clinical trials; a Biopharma might use a Data Mart to combine clinical trials data with a Clinical Trials Management System (CTMS), an Operational System used to manage clinical projects and personnel, along with payments to the parties involved. One could integrate the data from a CTMS with clinical trials data in order to determine how much a study or its components cost on basis not available using the CTMS alone.

Leaving clinical trials data aside altogether, a Biopharma might use a Data Mart in its most commonly defined role in the marketing and manufacturing arenas. A Data Mart might be used to manage overall business performance, identify new sales opportunities, improve existing sales efforts, match supply and demand, or improve inventory and/or customer service levels. In this case, people that work primarily with clinical data would not be involved in building or using the Data Mart.

One Data Mart middleware vendor, Silvon, lists a number of Pharmaceuticals / Chemicals firms that use its product.⁴ Amongst these firms are Biopharmas such as Alcon Labs, American Pharmaceutical Partners, Bristol Myers Squibb, Freedom Drug, Invitrogen Corporation, J & J Medical, Mallinckrodt, Inc., Novartis Nutrition, Nutramax Products, and Pharmavite. It appears that in each of these cases, the Data Mart does not have a clinical trials data component.

(continued on page 3)



Data Mart Software

Much as it is with EDC software, the marketplace for Data Mart applications is rather chaotic, in turmoil, and replete with confusing nomenclature and offerings. A Data Mart is created using an application that is used to build a repository from multiple data sources, commonly known as “middleware.” Table 2 shows an incomplete listing of Vendors offering middleware products for Data Marts.

Table 2. Vendors with Middleware Products for Data Marts⁵

Vendor	Website	Comments
Ascential Software	http://www.ascential.com	
Carleton Corporation	http://www.oracle.com	Bought out by Oracle
Evolutionary Technologies Int'l	http://www.eti.com	
Informatica	http://www.informatica.com	
Platinum Technology	http://www3.ca.com/products	Bought out by Computer Associates International
Praxis International	http://www.praxisint.com	
Prism Solutions		Website no longer exists
Red Brick	http://www-306.ibm.com/software/data/informix/redbrick	Bought out by IBM
Reliant Data Systems	http://www.compuware.com	Bought out by Compuware
Sagent Technology	http://www.sagenttech.com	Bought out by Group1 Software
SAS	http://www.sas.com	
VMARK	See row 1.	Bought out by Ascential Software

As one can see from the table, there is an ongoing Vendor shakeout much like the one currently going on amongst EDC vendors. Unlike the turmoil amongst EDC vendors, which might be caused more by technological and regulatory changes, turmoil amongst Data Mart vendors appears to be due more to marketing excesses. That is, over-hyping the product along with the fact that the nature of the product is at best ambiguous in reasonable or measurable results has led to a perceived under-performing of what usually is an expensive customized application built from the product's tools.

(continued on page 4)



Creating A Data Mart

A largely overlooked consideration is that Data Marts take a lot of thinking, planning and hard work for uncertain benefits. While the nature and value of what one might find while “mining” is not known beforehand, it is necessary to carefully define the goals of the Data Mart before it is built in order to ensure that the repository that is created can provide the basis for the information being sought.

If a Biopharma does not have a number of very old legacy systems and/or “foreign” database(s) acquired via merger partners, their data should be fairly accessible in a “single” database with common metadata - especially if they only have ever used products from a single CDMS vendor. In this case, a Data Mart may be fairly easy to create, even without any middleware product.

More recently, due to generational changes in CDMS and EDC products, however, Biopharmas are changing vendors. The shift from mainframe host to client-server systems started the migration to new products and in some instances to different vendors; the “web enabling” of software has further encouraged Biopharmas to replace legacy systems, many of them developed in house, with newer systems purchased from a software vendor. The result of this is the proliferation of databases used to store clinical trials data, making the creation of a Data Mart more difficult, but potentially more beneficial.

Given the possible uses for a clinical data based Data Mart (as listed in Table 1), the multiple data sources used to build a repository could conceivably include all the places clinical trials data is stored. Typically, clinical trials data is currently stored in four or five places at a sponsoring Biopharma and these are shown in Table 3.

Table 3. Clinical Trials Data Sources

Data Source	Notes
Old legacy mainframe databases and/or files	These files are usually of hard-to-use proprietary format.
Relational Database Management System (RDMS) such as Oracle	The database utilized by the most current CDMS/EDC product/application(s).
Other CDMS	For example, a database built by one or more merger partner(s).
SAS datasets	Typically these would be stand-alone databases and not reporting databases derived from CDMS.
Archives	On-line where Data Marts need them in order to work, perhaps in CDISC/XML form.

(continued on page 5)



At the end of the clinical trial data life cycle, it has in the past, usually been archived off-line. In recent years, however, online storage has become very inexpensive and the need to “archive” data has probably reached a point where it is not necessary at all, let alone off-line, except in the form of an off-site, disaster recovery back up. This means archives could be increasingly left on-line. With the advent of CDISC’s Operational Data Model (ODM), clinical data stored in archives might be left (or brought) on-line in the form of modern ASCII-based XML files. The obvious benefit of XML is that it can be used to retain important metadata, whereas ASCII files can only retain such data with effort.

In summary, implementing a Data Mart might be best done when CDMS, EDC, and if any, merger partner systems are expected to remain unchanged for several years. EDC Management, however, is not convinced that Data Mart and clinical trials data is a useful combination, largely because combining clinical trials data from multiple studies can be easily performed within SAS which would normally be used to look for unexpected therapeutic reaction(s) and/or effects as well as to explore “hidden” safety issues.

Caveats

Also troubling is the fact that data stored in clinical data management systems is not Operational in nature and that Data Mart products are developed primarily to integrate operational system databases. Misapplication of any software product used to develop custom applications is not a good idea. An analogy that might illustrate what we mean by this is using an EDC product developed for drug trials to conduct device trials. Although somewhat similar in nature, the underlying requirements are very different. This misuse of the product raises the failure risk factor significantly.

However, our outlook is not all bleak. Some EDC vendors might consider why their EDC product is a difficult sale and realize that maybe it is because it is mispositioned and sold to the wrong people. Perhaps it should be re-configured and sold to hospitals and physician sites in order to make use of some of the Data Mart technology to extract data normally collected by hospitals and other investigational sites. Doing so, and by having “add-on/drop-in” modularity to provide scheduling, capture and data basing of additional study required data, edit checks, and meet regulatory requirements for audit trails, good clinical practices, good computer practices, and electronic signatures and so on, one might create a new generation of Electronic Data Capture applications that build a clinical trials database that is a data repository created from multiple data sources.

(continued on page 6)



Conclusion

The recent involvement by some of us here at EDC Management in setting up data entry systems where one of the major design considerations was integrating those systems with the corporate Data Mart posed some interesting questions. Data Marts biggest benefit may lay in their implicit support of data standardization. Data Marts require careful planning and implementation. They guarantee no return on the expense and effort expended on them.

Data Mart vendors are undergoing the same changes that EDC vendors are going through, leaving the marketplace unsettled and filled with vendors that may not be around tomorrow. Data Mart technology on the other hand may promise some EDC design insights. Data Mart products are certainly based on technologies that are worth watching, as they may eventually become a break-through product of significant merit.

Readers are encouraged to share recommendations that they may have for how vendors can improve their products. We would also like to encourage feedback and suggestions to this issue, and welcome suggestions of topics for future issues.

References

¹ <http://www.kenorr.com/dwpaper.html>

² http://www.webopedia.com/TERM/d/data_mart.html

³ http://www.webopedia.com/TERM/d/data_warehouse.html

⁴ <http://www.silvon.com>

⁵ <http://www.infogoal.com/dmc/dmcdwh.htm>

Resources

http://www.its.state.ms.us/et/datawarehouse/et_dwh.htm



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