The Clinical Data Warehouse – a New Mission-Critical Hub

Jonathan Palmer – Oracle Health Sciences

Jonathan Palmer is senior director for clinical warehousing and analytics at Oracle Health Sciences.

Jonathan first joined Oracle in 1997 where he has had various roles across business development, consulting, and product development. Since 2008, Jonathan has been involved in defining new product strategy targeted at developing innovative solutions for the life sciences industry.

Demands for clinical warehousing are increasing dramatically. From being viewed as a data gathering tool, a clinical data warehouse is now moving into a new phase of becoming a business-critical platform. Such a platform can support all clinical decisions across the clinical trial portfolio, be central to collaboration, and fundamental to the survival and agility of the business. To fully comprehend the meaning, applicability and relevance of a clinical data warehouse, and how it can potentially benefit a company, we must understand how it is different from a typical data warehouse, what has driven its need, and how its adoption can be maximized to drive clinical development.

Traditional warehousing versus clinical data warehousing

Traditional business data warehousing is a well-established IT discipline, the primary focus of which is often to deliver decision support capabilities to drive productivity and efficiency gains across a business. Typical examples can be horizontally focused, such as in inventory management, or in industries such as banking, telecommunications or retail. These warehouses are based on well-defined structures, data sources and goals. For example, a warehouse focused on inventory management allows a business to manage stock, assess sales, and answer well-defined questions to drive efficient stock management in response to sales activity.

Health Sci

First published March 2013

In contrast, the clinical trials segment of the life sciences industry is unique in its needs based on the variability of its data structures driven by trial design. The primary focus of a clinical data warehouse is to facilitate extraction of value from clinical data. This can aid study design, prove efficacy and safety of new products, and support regulatory queries. Productivity and efficiency gains, whilst important, are rarely the key focus of a clinical data warehouse.

Traditional warehouses are less focused on regulatory compliance and the need for full traceability of data life cycles is often less relevant, compared with the high demand for such capabilities from the life sciences industry.

As with all warehouses, a key feature of a clinical data warehouse is that it should allow a company to store and release value from data assets to drive better decisions. For example, it may enable a pharmaceutical company to realize value from the vast amount of clinical data generated from a single trial, all trials in a particular program, a therapy area or, indeed, all the trials in the company. The ability to standardize, pool, analyze, explore and mine data across large and disparate data sets has previously been challenging for the life sciences industry. Whilst a traditional warehouse uses only a finite set of data sources, there are potentially hundreds of data sources in the clinical space, each with a different structure, variability and frequency.

Drivers for change

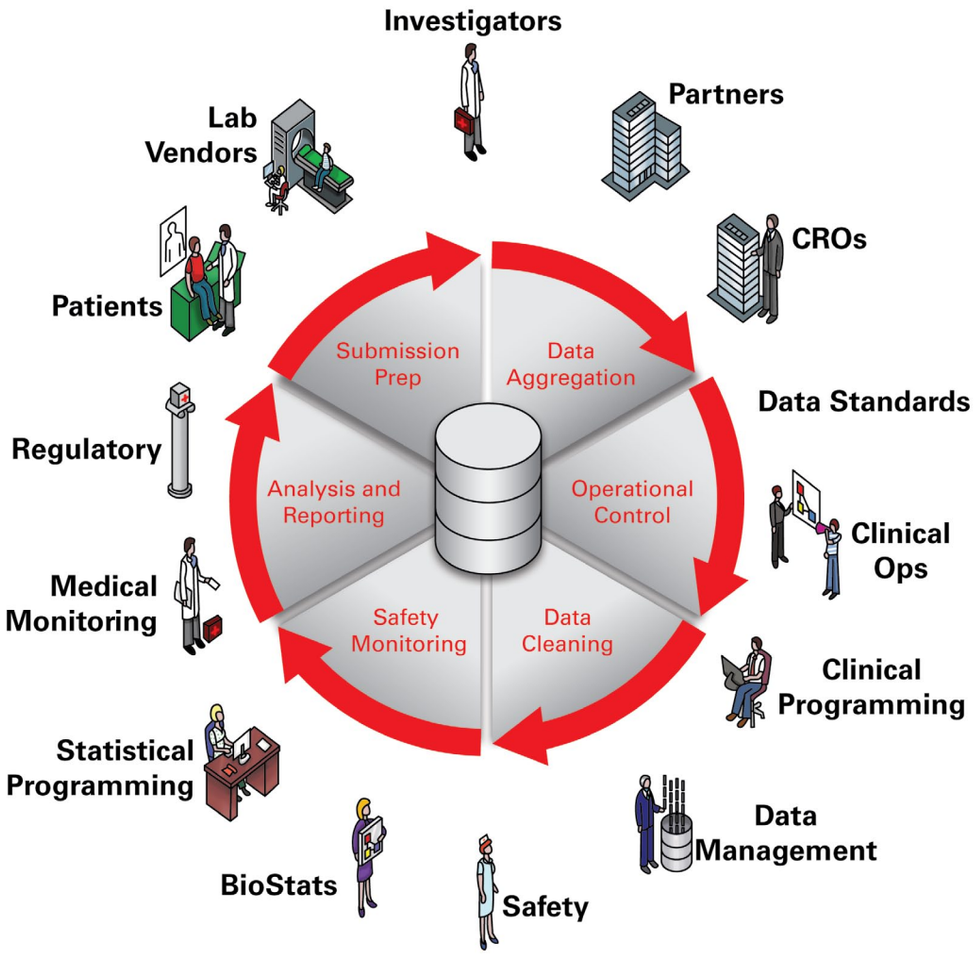
Historically, data has been managed and stored in distinct silos in a function-centric model – for example, data for the clinical data management, biostatistics or safety groups. Whilst synergies exist across these groups, often data in silos cannot be accessed by other teams. This often creates data lag, as data must be requested from one group to another, and 'manually' handed over. Transparency and availability of data across the breadth of the organization has been a challenge. As the industry has evolved, there has been an increasing need to access, combine and share data across multiple information domains. Further, there is the need to effectively align clinical and administrative data to provide a complete picture of study conduct, from an operational, safety and regulatory perspective.

There have been a number of key changes in the industry including:

1) Industry consolidation
2) Cost constraints
3) Globalization, outsourcing and virtual technology
4) Aggressive generic drug replacements
5) Increased scrutiny by governmental and regulatory bodies.

These drivers for change have focused the attention of senior management on the need for greater data visibility, transparency and availability. Information silos have become barriers to clinical innovation, and organizations are increasingly seeing clinical data warehousing as a platform to support more agile clinical research, and as an essential base to maximize the clinical portfolio.

Figure 1: A clinical data warehouse to enable collaboration through shared visibility



Clinical data warehousing is becoming ever more important. Acting as a central hub for information storage, collation and archiving, a clinical warehouse is essential to delivering a consistent, single view of the data assets across an organization. It is essential that all team members, both internal and external, have access to a consistent view of the data to drive clinical research (Figure 1).

Standardization and master data management

Standardization is fundamental to combining and sharing data. Whilst most industries are based on well-defined standards, the clinical trial industry has struggled to agree on a universal standard due to the broad nature of trial data. Standards organizations such as the Clinical Data Interchange Standards Consortium (CDISC) and Health Level Seven International (HL7) have made significant contributions over recent years. Complete adoption of standards across the clinical portfolio is now recognized as being fundamental to enabling efficient process design, and extensive usage of standard software tools for analysis, exploring, and mining without significant reliance on trial-centric programming resources. Submitting standardized clinical trial data to regulatory authorities (for example, the FDA or EMA) allows analysis and review of submissions using standardized methodologies and tools, which streamlines processes and regulatory reviews, potentially resulting in faster drug approval timelines. As with all standards, they are constantly evolving and expanding, and must make provision for data which cannot be standardized. Variation is typically focused around efficacy data, and platforms are available to support this variability.

Another consideration is management of master data, which typically relates to process-centric entities (for example, address details of principal investigators or trial sites). If master data are not recorded and managed correctly, critical contextual information may be lost, leading to expensive manual data cleansing and sourcing to complete the formal regulatory submission. Master data management, although well-known in general data warehousing, is a relatively new issue in the clinical trials data space.

"Of course the problem is, as soon as you find one data inconsistency, then everything is gone…the whole structure is compromized, and the research is at risk."

Evolving from clinical data management systems (CDMSs), to clinical data warehouses

Historically CDMSs have been used to capture case report form (CRF) data and support cleaning cycles. Over the years the role of the CDMS has changed. Organizations typically sought to import as much data as needed into CDMSs to support new product submission. Laboratory data were often imported into CDMSs, which evolved into an idealistic 'single source' or pseudo data warehouse. However, as clinical trials became more complex and data source variability became broader and more diverse, it became impractical to load all data into the CDMS. Conversely, the evolution of CDMS systems to electronic data capture (EDC) moved the focus from manual data processing to site-based data capture. Whilst succeeding in accelerating CRF data capture, the need to manage non-eCRF data sources remained unaddressed. Today many organizations pass the burden of data consolidation and integration onto data management programmers and/or biostatistics programmers, to manually combine a file-based data source with the eCRF-based data. This process is labor intensive, requires specialist programming skills, and can result in quality issues if data standards are not leveraged effectively.

Using CDMS systems as a pseudo data warehouse, augmented by file-based approaches, inherently fragments the data. For a single drug trial this approach may be adequate as it is possible to navigate through the different files and obtain the required data for drug approval. However, to view all trials in a therapy area (e.g. diabetes) and to explore across related trials becomes a complex undertaking as data are stored in separate files, requiring specialist programming skills to access and analyze.

"So has the industry survived? Sure. Companies have got new drugs to the market, but are often missing an opportunity to exploit huge inherent value embedded in their vast data stores. Why? Because in many cases the data are locked away, in a secret 'vault' that only the biostatistician or statistical programmer has access to."

As organizations have undergone this system evolution, it has become increasingly apparent that force fitting all data into data capture systems (CDMS or EDC) is not a viable solution. Furthermore, file-based stores are inherently difficult to search and mine across. This reinforces the role of the clinical data warehouse as a purpose-built relational data store, which can flex, expand and scale to meet the varied needs of clinical trials.

Globalization and outsourcing – a new paradigm

The life sciences industry is continually under increasing pressure to become leaner and reduce healthcare payers' costs, whilst ultimately increasing end value to shareholders. Ongoing industry consolidation is partly driven by the need to be more efficient and the need to find ways to expand portfolios, and defend against aggressive competition.

"Twenty years ago the industry was incredibly cash rich and blockbuster-centric. We've seen over recent years massive consolidation through mergers and acquisitions, huge patent expiries, and healthcare payer budget cuts. Clearly the industry needs to be a lot more agile and innovative."

The world is becoming smaller through more integration, virtualization and collaboration. Enabling technologies, like the Internet, have opened up new, previously unavailable business models. Workforces are becoming more geographically distributed across different time and language zones so as to lower costs. To meet the demands that these changes bring, the role of the clinical warehouse becomes ever more important. The clinical warehouse is evolving from an internal knowledge base into a hub for leveraging new business models. For contract research organizations (CROs), clinical warehouses had previously been irrelevant due to the study-centric processing model. However, many CROs are now focusing on building warehouses that can act as a centralized and standardized platform on which they can add tools for extracting value from data. This allows CROs to differentiate themselves from their competitors, from commodity vendor to strategic partner. Similarly, as pharmaceutical companies look to reduce internal costs and optimize processes, they are increasingly leveraging global service providers and using clinical data warehouses as an integration and collaboration platform to enable full service, and hybrid (using both internal and outsourced resources), outsourcing.

Challenges in implementing a clinical data warehouse

The challenges facing clinical data warehouse implementation are largely determined by the management's view of the implementing company. 'Forward-thinking' senior management view a clinical data warehouse as being fundamental to progression and essential for collaborative innovation i.e. the ability to adapt, be agile, acquire organizations, outsource, and generally be more efficient, cost-effective and competitive.

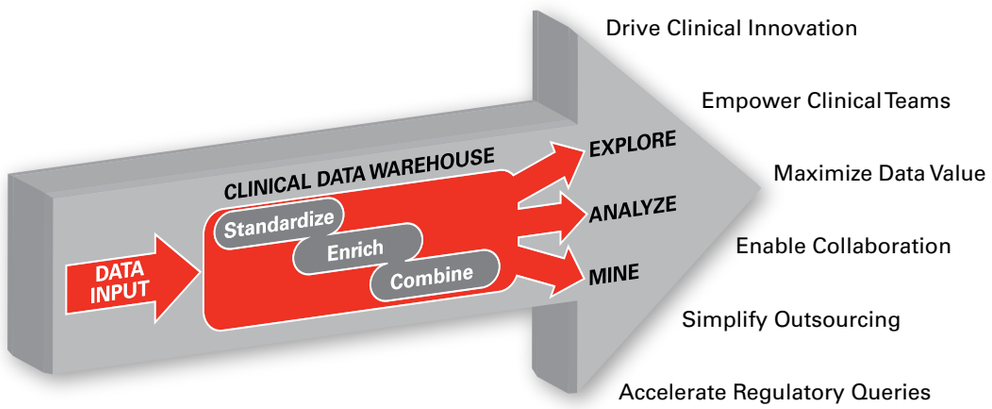
Demonstrating a clear return on investment (ROI) is always challenging as clinical data warehouses are often multi-year programs with abstract, cross-functional concepts. To be successful, a clinical data warehouse requires continuous senior management commitment and sponsorship.

It is important to build a broad, holistic picture for the organization rather than at a departmental level. Buy-in solely from one or two department heads is rarely sufficient. However, transition from batch-centric data preparation and programming to better solutions, more in keeping with the dynamic nature of the industry, requires a key company visionary, with the gravitas to communicate the overall benefits of a clinical data warehouse to the wider company.

The implementation phase of a clinical data warehouse can be challenging. This phase requires essential specialist skills in areas such as process design, data standardization, modeling, and system integration. However, once the clinical data warehouse is in place, fewer specialist skills are required.

Key emphasis must also be placed on management of process design and change, user training, adoption, and cross-departmental co-ordination to ensuring that investment in the clinical data warehouse is maximized. A continuous program of monitoring and driving user adoption, streamlining data flow, and extending and enhancing use cases and tools-sets for data exploration, visualization and analysis are key to ongoing return. These projects do not stop when they go live but evolve with continuous improvement and adaption (Figure 2).

Figure 2: Schematic depicting the general capabilities of a clinical data warehouse



Identifying use cases

For a clinical warehouse project to be successful it is essential that clear and specific use cases are defined before project initiation. A use case can be analogized to breaking up a 'big problem' into 'bite-sized' pieces, targeted at delivering specific business benefit. These could be focused on reducing data handoffs, standardizing data, accelerating statistical analysis, or simplifying medical and safety review. Without use cases to drive project goals and ROI, such projects can become complex IT architecture programs with poorly defined endpoints delivering little business value at completion. Focus on specific use cases allows a clear understanding of the likely process changes and an understanding of the potential benefits of changing these processes.

Once use cases have been identified, the right platform can be selected, along with associated technology and consulting services to assist companies in delivery.

"Phased-delivery is one of the key things. Due to the many interdependencies of the components in these projects some companies adopt a Big Bang approach of trying to do everything at once, but these often fail to deliver as return on investment is too long. It is critical to implement use cases that give incremental return."

A phased delivery, rather than trying to deliver universally on everything, is advisable. Defining specific use cases at the start, and phasing the implementation over several stages, allows the realization of tangible benefits, while successfully managing stakeholder expectations.

Extracting full potential

It is imperative to create a learning organization which can evolve, extend and expand, to extract ongoing value from a clinical data warehouse.

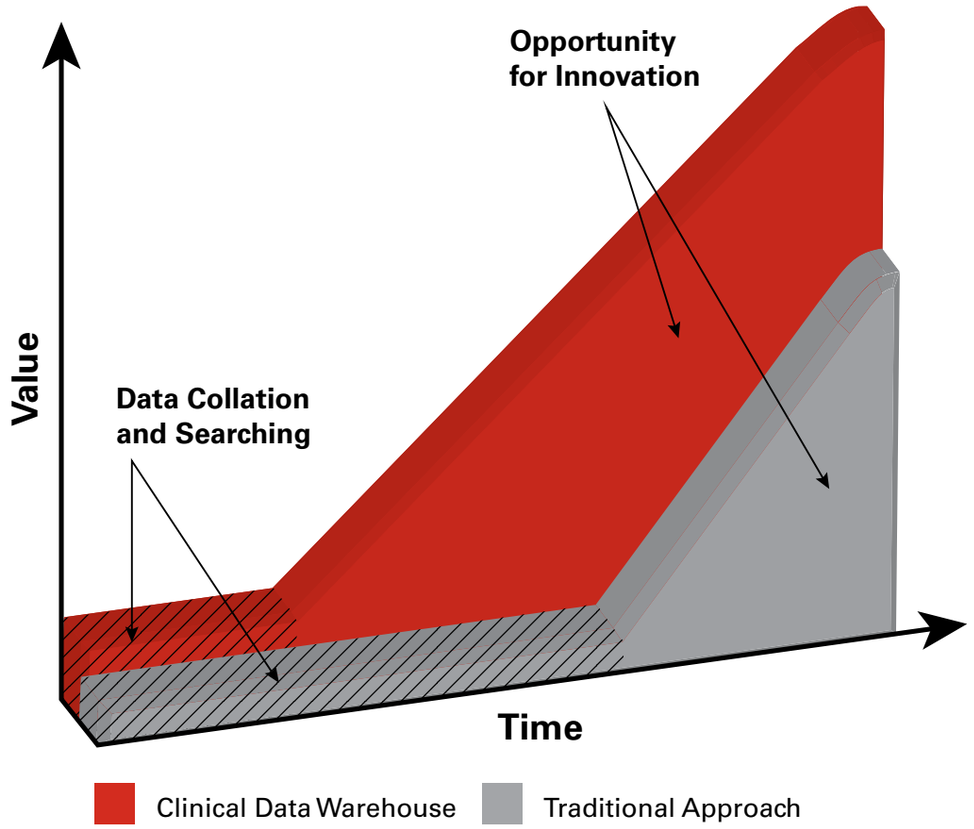
There must be a conscious effort towards creating alignment between IT and business units of an organization. The clinical data warehouse platform should be an enabling platform from which the business can gain significant, repeatable value, for example improving response time to regulators, identifying potential new therapies, or moving towards individualized medicine. This is the ultimate vision for clinical data warehouses.

Modeling and simulation is another area from which value can be extracted using a clinical data warehouse. For example, simulations may influence trial design in terms of identifying appropriate recruitment populations or helping to design efficacy parameters. Providing these groups with rich, collated, standardized data via a clinical data warehouse can enable them to effectively and precisely predict outcomes in various models.

"A good clinical warehouse can truly flip the '80:20 rule' for a modeling/simulation analyst. Instead of spending 80% of the day searching and cleaning data for analysis, and only 20% on analysis, they can leverage the data warehouse, find their data quickly, and spend 80% of the day analyzing to accelerate research."

The ability to easily mine and explore legacy data can also uncover hidden value, for example, a previously abandoned drug in one disease could be investigated for use in another disease. Likewise, mining and exploring data acquired through mergers and acquisitions could be used to augment current company data and thereby increasing data value (Figure 3).

Figure 3: Optimizing collation and search leads to increase opportunity for innovation



Emerging technologies and big data

There is increasing recognition of a need for improved management of big data in the clinical space, and efficient aggregation and integration with core clinical data will be key to successful clinical warehouses in the future.

Big data are unstructured compared with traditional (fully structured) or clinical (mostly structured) data formats. They can be obtained from sources or formats such as social media and include 'real world' use of prescription and overthe-counter drugs. In this scenario, patients may use social networks to relate their drug experiences, for example, in terms of safety or adverse effects. It is critical for the drug manufacturer to be able to mine these data, recognize potential issues arising, and address or manage them effectively, thus providing pharmacovigilance insights on a marketed drug. Other key sources of big data will come from advances in patient genomic profiling, as well as wearable medical monitoring technologies. Combining such huge data sources with well curated clinical trial data will be essential to delivering individualized medicine. An approach taken by IT vendors to address big data is to design technologies that combine both software and hardware to support large scale data sources and data warehouses, for example, Oracle Exadata. As the life sciences and healthcare industries converge around delivery of individualized medicine, the need for such database machines will be critical to delivering targeted treatments. By exploiting such high performance data management platforms the industry will transition from drug-driven clinical trials to patient-driven.

Final thoughts

In addition to, and as a consequence of, the overall move of the life science and healthcare industry to become more cost- effective, efficient and competitive, it must improve the overall value realized from its key asset, data. A clinical data warehouse provides a solution for life sciences companies to better access, mine, explore and use data across their trials and portfolios. Furthermore, clinical data warehouses can be viewed as an essential tool for speed of access to data when considering globalization, outsourcing and merger and acquisition activity. Most importantly it can be viewed as a platform for accelerating clinical innovation.

Clinical data warehouses represent an exciting area of current development and offer the potential to shape the way we utilize and manage clinical data. They will continue to evolve and, with the convergence of the life sciences and healthcare industry, will become a necessity in order to efficiently drive value from data and ultimately accelerate development of new therapies.