Clinical Research Innovation through Shared Clinical Data Warehousing

Jerry Whaley – Pfizer

Jerry Whaley is senior director of development business technology at Pfizer and is involved in the implementation of Pfizer’s clinical data warehousing solution.

Jerry began his Pfizer career in 2001, as director, development informatics Ann Arbor site head. Before joining Pfizer, Jerry was an SAS programmer and supervisor at the Upjohn Company and a systems analyst, developer, project Leader and manager at Parke-Davis. Prior to returning to Pfizer Jerry was vice president at Advanced Systems Development with responsibilities for business development and client implementation management.

Increasing partnerships between pharmaceutical and biotechnology companies and/or service providers is a key emerging trend in the life sciences industry. Specifically in clinical trial management, there has been an overall re-assessment of what constitutes ‘competitive advantage’ with regard to data capture and management. The questions being asked include: does a ‘custom-developed’ electronic data management platform really provide competitive advantage? Could using a standard platform help minimize the many issues caused by variability of data? Where is it best to focus pharmaceutical company resources?

“Unsurprisingly, focus is all on data. If, as an industry, we can holistically understand our data better and more in-depth, so not just as Pfizer-specific data but, for example, healthcare as a whole, then that’s advantageous as it allows us to better analyze it.”

Pfizer’s clinical trials operational model has evolved over the years. The initial model progressed from conducting inhouse trials to outsourcing trials to 17 functional service providers. This has now evolved to the company’s current position of having two alliance partners (ICON and Parexel), or contract research organizations (CROs). This change has been driven by a business need to be more cost-effective and manage resources more efficiently. This model enables the company to leverage the CRO’s expertise in execution of clinical trials.
and allows Pfizer’s role to evolve, into a more oversight role. Pfizer’s oversight input on clinical trials also requires expertise and skills that need to be acquired over time. The consequence of this model is that it frees company resources to focus on analysis of trial data, rather than preparation of data. Such an undertaking requires a standardized data warehousing solution for data receipt, aggregation, access, and analysis.

Pfizer’s intention is to create a ‘road map’ to define and standardize processes for data integration and data sharing based on a communal data warehousing solution. In addition to enabling interactions with CROs, the visionary view for this type of clinical data warehousing is that it could also facilitate future interactions with multiple partners including other pharmaceutical/biotechnology companies, regulatory authorities, and companies absorbed through mergers and acquisitions.

**Pfizer’s Clinical Aggregation Layer (CAL) solution**

Pfizer’s vision is to create a cloud technology platform which facilitates efficient clinical trial operation for industry peers, to minimize duplication of effort in tool development, and drive process efficiency to accelerate new drug research. Company owned data handling tools and applications may not necessarily provide competitive advantage but do increase costs. In Pfizer’s view, as long as individual company data are secure and protected and there is appropriate legal and regulatory approval, data can be stored and processed from a central platform that is located externally to Pfizer, providing an opportunity for sharing technology across the industry.

Pfizer’s data warehousing solution, known as the Clinical Aggregation Layer (CAL), consists of three core components (Figure 1A and 1B):

1) A clinical and scientific data warehouse (CSDW) to manage, aggregate, and analyze clinical trial data
2) An operational data warehouse (ODW) for trial performance metrics
3) A custom-developed trial master file (TMF) tool to keep a comprehensive record of all clinical trial activities.

Data are loaded into CAL through various mechanisms, depending on data type and source (e.g. data exchange adapters, secure file transfer protocols, etc.) such mechanisms being based on industry standards. Data stored in a metadata repository, are also uploaded into CAL – Pfizer captures and maintains these metadata. It is imperative that input data are correctly referenced and indexed for such a solution to be effective.

“We [Pfizer] see this as sowing the seed of an industry infrastructure, that’s our vision. This is not just a Pfizer solution; we are trying to seed this solution with partners such as Oracle and Accenture.”
The CAL solution can be potentially both cost-effective and innovative and could drive the development of new tools available for a majority of users. Collective innovation may also result from being able to conduct in-depth analysis using shared data that is readily accessible from a standardized IT infrastructure. A shared platform may also be used to leverage trial data more broadly – for example, companies conducting trials in a single disease area could, in theory, share placebo data if patient recruitment criteria were similar thereby reducing costs for the placebo arm of a trial.

This model is possible due to sufficient evolution of data standards, services, technology and IT infrastructure. The convergence of technology and business needs, tighter business models (with regard to efficiency and cost), and stringent regulatory processes and requirements have further reinforced the premise of such a solution.

**Figure 1B: Components needed for the new Pfizer platform**

### Two Guiding Principles

<table>
<thead>
<tr>
<th>One version of the truth</th>
<th>Exchanged data based on defined data standards</th>
</tr>
</thead>
</table>

### Analysis and Review Tools

- **Contextual Data**
  - Reference Data Management (RDM)

### Operational Data Warehouse

- Integrated repository for data required to track and manage study and development program execution

### Clinical Scientific Data Warehouse

- Integrated repository of analysis-ready data from our partners, Pfizer, and relevant external sources

### Trial Master File

- Authoritative source of Essential Documents
The willingness of industry peers to be participants in such a warehousing solution is yet to be realized. To this effect, Pfizer is actively engaging in discussion with peer companies to gauge interest. Initial indications appear positive.

**Approach to implementing a shared clinical data warehouse**

The most evident differentiator of Pfizer’s clinical data warehousing solution is its accessibility of technology across the breadth of the life sciences industry. Pfizer is actively avoiding customization of its data warehousing solution, and is making every effort to maintain it as an ‘off-the-shelf’ solution to allow for broad applicability.

“We [Pfizer] are trying to stay rigid to the fact that these are commercially available, off-the-shelf solutions – Do not ‘Pfizerize’ them; Do not customize them – avoid this as much as possible. This approach allows reusability, ease of implementation and ease of support long-term.”

To further ensure this, Pfizer frequently engages in discussions with its IT provider to ensure that tools and applications remain generic, enabling easy upgrades, processing and most of all, seamless partnering with external parties. Although outsourcing operational functions (e.g. execution of clinical trials) and using shared data warehousing platforms and technologies is not a new concept, it was unprecedented for large companies like Pfizer. The long-term aspiration is to externalize most processes and maximize the use of external expertise to drive Pfizer’s healthcare goals.

Implementation of the Pfizer solution necessitates a phased-approach to make this complex and challenging undertaking a more manageable and viable prospect. Future releases include additional functionality within the CSDW, across other therapy area and clinical trial teams, followed by ODW-related operational data.

Stakeholders have readily championed and supported this project from its genesis. The CAL solution’s goal is to be instrumental in increasing Pfizer’s operational efficiency and re-focusing resources towards accelerating clinical research.

**Use cases supporting a clinical data warehousing solution and potential users**

A number of identified use cases supported the need for the CAL solution. The most immediate was associated with the new operational model of increased clinical trial externalization. This required consolidation of data to a single location to allow Pfizer easy access to trial-related information and data. In addition, the need to consolidate data gained through mergers and acquisitions. Pfizer identified the need to improve the ability to explore, analyze and mine both clinical trial and operational data so as to maximize its value. From a regulatory and compliance perspective, the availability of all trial data in a single location could facilitate a quicker response to queries.
Presently, the two critical primary users of the CSDW component of the CAL solution are clinicians and statisticians. Users can explore, analyze, and mine clinical trial data on a single platform and, as a result of data standardization, can leverage a broad range of tools to drive research. Other users are the company’s partners and service providers who upload trial-related data which Pfizer can use for data analysis, as well as monitoring trial progress. The ODW provides operational metrics which can be used to drive decisions (e.g. what geographical region may be suited to a trial in a particular disease area). In order to remain compliant with regulatory requirements, the TMF solution provides definitive proof and record of all clinical trial activities. Such factors contribute to cost-benefits, better time efficiency and management (through standardization) and, increased data value realization.

In addition, the CAL solution could simplify the role of industry regulators and auditors. For example, where previously there were different processes/systems for each company, with the CAL solution there is a single system to understand. Thus, processes such as auditing/inspecting could become more efficient based on the reduction of industry systems an auditor would need to be familiar with.

“This model can facilitate a more progressive and efficient life sciences industry in that regulators have only one infrastructure and set of applications that they need to understand, audit and ensure compliance.”

Planned trial throughput via the CAL solution

Moving ahead, Pfizer’s aim is to route as many clinical trials through the new operational model as possible. Typically the company runs approximately 800 trials in a given year – the intention is to transfer a proportion of trials (approximately 100) to the new model by the end of year one and to accelerate throughput to approximately 300 in year two. The longterm vision is to decommission all legacy processes and applications.

Although there is a company-wide effort to transfer clinical trials to the new system, there is recognition that this must be done in a controlled manner to maintain data integrity. To this effect, the Clinical Data Interchange Standards Consortium (CDISC) and Study Data Tabulation Model (SDTM) have played important roles. In addition, standardization facilitates amalgamation of data following mergers and acquisitions – an activity frequently associated with Pfizer.
“Standardization is an important piece of our [Pfizer] strategy because without it, aggregating these data would have a lot less impact and a lot less value.”

The past, the present, and looking to the future
In the past, Pfizer attempted to build and implement its own ‘in-house’ data warehousing solution but this proved challenging. Nonetheless, many lessons were learned from early efforts which have influenced the CAL solution including the value in using a generic, robust, commercial off-the-shelf technology tool-set acceptable to other industry peers as a shared data warehousing platform.

“We [Pfizer] felt we chose our tools wisely; it was important to choose a scalable and industry leading tool-set that others would embrace.”

The present implementation of the CAL solution has not been without challenges. Some of these include converting ingrained legacy business processes to new processes and systems, availability of required stakeholders/personnel for making implementation-related decisions and meeting stringent deadlines and, as with all ambitious projects, managing budgets effectively.

Looking to the future, Pfizer’s view is that this is merely the beginning of the journey. For the company, the CAL solution provides an innovation bed for managing, analyzing, accessing, exploring and extracting maximum value from data, whether legacy or newly generated. In addition, it provides a simple and standardized means of collaborating with multiple partners. Overall, the vision is that this joint data warehousing solution, which is a novel concept in the clinical trial space, will enhance innovation both from a scientific and technical perspective. To achieve this vision, Pfizer has leveraged the expertise of global IT service providers, including Oracle, to provide it with a clinical data warehousing solution that acts as an integration and collaboration platform to enable full-service, hybrid outsourcing, as well as to support internal processes.