COMPOSITE
— S O F T W A R E —

Data Accessibility Allows Quicker Drug Development

ABOUT PFIZER

Founded in 1849, Pfizer is the world's largest research-based biomedical and pharmaceutical company. The company reported $68 billion in revenues for 2010, and invested $9.4 billion in R&D. Pfizer has delivered to market 46 major pharmaceutical and animal health products.

CASE STUDY BACKGROUND

The industry average to bring a new drug to market is 15 years. Each year of discovery and development entails billions of dollars in research and testing. Pharmaceutical companies like Pfizer must make informed decisions about which prospective drugs show the most promise, and focus their resources accordingly, to be competitive.

In Pfizer's Informatics Division, research scientists and drug portfolio managers must access and analyze a complex set of data from multiple, disparate sources. This information comes from scientific databases that contain diagrams of compound molecular structures, test results from prior research, and details on a compound's lineage through various stages of development, as well as business-side information such as the number of researchers involved in a project (HR databases), the cost of personnel and other resources (finance systems), and the number of "stage gates" each project has passed through (project management databases).

THE PROBLEM

Data integration represented the biggest bottleneck to effectively provide information to researchersand managers during the drug discovery process. Pfizer has traditionally used three approaches: custom coding between sources and consuming applications; replication of file extracts; and Extract, Transform and Load (ETL) to create data marts and warehouses. Yet, proliferating data silos and increasing complexity had minimized the effectiveness of hand coding. Data replication had only increased the burden on IT of maintaining the quality and security of the replicated data in addition to the original source data. While ETL effectively integrated data for large-scale, multi-dimensional analysis, its development typically required weeks or months to complete, with the resulting data often out-of-date. Between the administrative complexity and difficulty in getting at the underlying sources, Pfizer averaged three to four months to complete its design, development, test, and deployment phases.

Further, none of these traditional approaches fit easily into Pfizer's new service-oriented architecture (SOA) strategy that emphasizes creating data objects for reuse.

> *"Composite is helping our group literally shave years off the drug R&D process."*
>
> Dan Eng,
> Project Manager,
> Pfizer Global Research and Development,
> Informatics Division

AT-A-GLANCE

Industry
Pharmaceuticals

Business Problem Solved
The challenge of accessing data from a complex system of silos was delaying efficient data sharing, lengthening the drug discovery cycle.

Data Integration Patterns
Pfizer established a more effective method for prototyping complex analytics solutions by creating a portal environment using the Composite Information Server to allow quicker access to data from multiple, disparate sources.

Composite Software Products
Composite Information Server 4.0.1

Data Sources
BEA Weblogic Dashboard supporting 75 business managers and analysts

Data Consumers
Research scientists and drug portfolio managers

Platform
Linux

ALTERNATIVES CONSIDERED

Before the Composite Information Server was implemented, Pfizer required a sequential development process to build ETL scripts, data marts, delivery scripts and a portal, resulting in lengthy development times. The manual coding, maintenance of data marts, and coordination with IT operations to set up data mart infrastructures created lengthy delays and backlogs.

THE SOLUTION

The Composite Information Server was implemented, allowing Pfizer's overall data integration capabilities to expand to data virtualization, data abstraction and data federation across both SOA and non-SOA environments. The scientist workbench can now support both build and run-time requirements.

With the establishment of an easy-to-use data modeler and code generator, data can be abstracted in the form of relational views for reporting as well as Web services for SOA initiatives. The high-performance query engine securely accesses, federates and delivers the diverse, distributed data to consuming solutions in real-time. Users now gain on-demand data for quicker decision making

With this solution, new capabilities and benefits are available to Pfizer's Research Scientists Workbench including:

• Automated data-level development, freeing developers to work on application-level development while reducing the total development time in half.

• Drag-and-drop development environment, built-in security and automated generation of Web services, requiring fewer specialized skills.

• SOA-compliant WSDL data services providing data in the form needed by portal developers.

• Loosely coupled data services that are easier to maintain than ETL and data delivery scripts when changes are made to the underlying data sources or the portal.

• Reusable data service assets.

A research data virtualization layer based on the Composite Information Server.
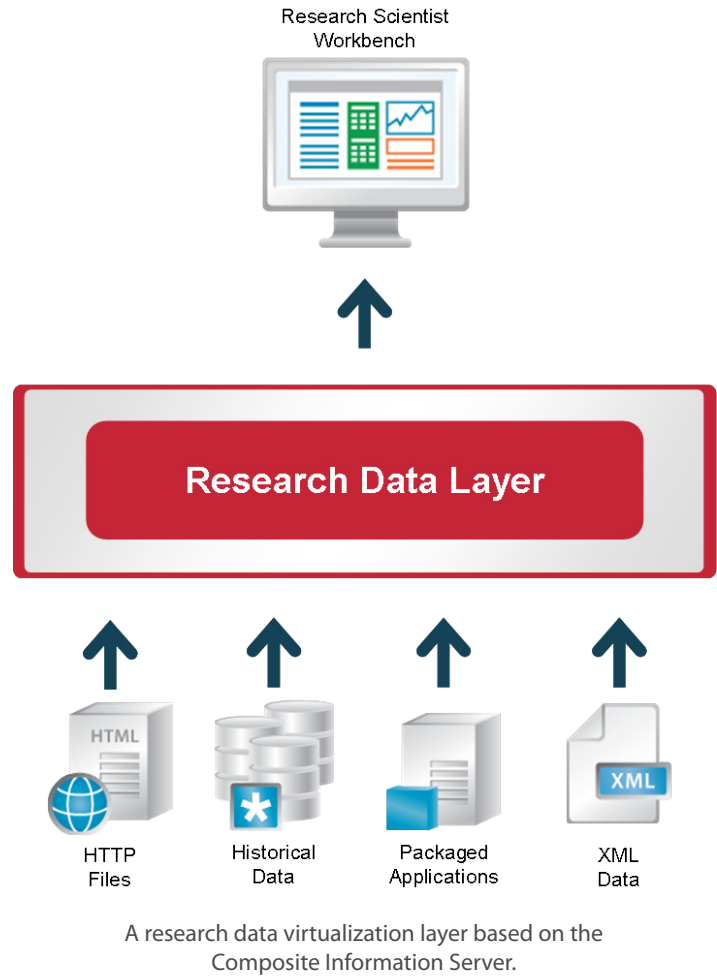
THE RESULT

• 50% reduction in time required to develop DI projects, from 3-4 months to 6-8 weeks

• 5% improvement of data quality by using data services to achieve SOA compliance

• 10% reduction in operational infrastructure costs

• 60% reduction in risk of drug discovery projects missing launch dates