Dana-Farber Cancer Institute Speeds Medical Research with Advanced Data Warehouse

Dana-Farber Cancer Institute
Boston, MA
www.dana-farber.org

Industry:
Healthcare

Annual Revenue:
US$665.7 million

Employees:
3,184

Oracle Products & Services:
Oracle Healthcare Transaction Base
Oracle Fusion Middleware
Oracle Identity Management
Oracle Access Manager
Oracle BPEL Process Manager

Implementor:
Oracle Consulting

Oracle Partner:
InforSense Inc. www.inforsense.com

“Oracle Healthcare Transaction Base enabled us to securely integrate our clinical, sample, and genomic data—helping us to maximize the use of this information in our quest to develop a better understanding of and treatments for serious diseases.”
– John Quackenbush, Ph.D., Professor of Biostatics and Computational Biology, Dana-Farber Cancer Institute

In the quest for more effective cancer treatments, researchers face a fundamental challenge: how to make sense of mountains of scientific and clinical data and find clues buried in them. The Human Genome Project, an unprecedented program that resulted in a map of the human genome and opened up limitless opportunities for exploring the genetic mechanisms behind diseases, also resulted in considerable challenges for researchers. Scientists can now tap literally tens of millions of pieces of evidence locked in DNA sequences, which offer exciting research potential but also bring about a new challenge of how to effectively manage and analyze all this data to advance research.

Scientists at Dana-Farber Cancer Institute, one of the leading cancer research and care centers in the United States, are pioneering a new approach to manage massive amounts of data to accelerate medical research. Armed with US$1 million from an Oracle Commitment Grant, the Boston-based institute has built a new translational research infrastructure, creating a data warehouse the pulls together clinical, sample, and genomic data.

With an initial focus on multiple myeloma, a deadly, difficult-to-treat cancer that strikes plasma cells and eventually bone marrow, Dana-Farber seeks to maximize the value of clinical research and help improve disease understanding and patient care.

Leveraging Oracle Healthcare Transaction Base, a comprehensive data repository and service infrastructure that enables data interoperability, and Oracle Fusion Middleware components as its foundation, the institute has pioneered a data infrastructure that is uniquely efficient, effective and secure.
Key Benefits:

- Combined sample, clinical, and genomic data in one database to enable advanced querying and accelerate cancer research
- Delivered real-time querying, enabling researchers to find answers in days rather than weeks
- Shortened the study design process from months to minutes
- Ensured the highest level of security and privacy to meet industry standards
- Provided an easy-to-use interface that requires little training
- Delivered a scalable infrastructure that the institute can extend to other research projects and partners

The new Oracle-based data infrastructure helps researchers investigate clinical data more thoroughly, make complex queries and more complete data analysis, and improve experiment design. It also enables more rapid queries, providing researchers with answers in minutes to queries that previously required days to process.

Seeking Genetic Clues

John Quackenbush, Ph.D., professor of Biostatistics and Computational Biology, who joined Dana-Farber after working on the Human Genome Project, explained the challenges the institute faced.

“The genome project spawned a whole host of technologies that enable us to create a DNA fingerprint. What we are trying to understand is how those fingerprints vary between individuals, and how those fingerprints vary between people who are likely to develop diseases and those who might be resistant to them,” he said.

Quackenbush explained that while the data surrounding gene expression and variation is huge, the metadata is even more complex, including information about the patients, when and how samples were collected, and many more variables.

“All of this data is becoming very complex and rich, and what it really requires—if we are going to use that data to our advantage—is an integrated, well thought-out approach to collecting and managing all of that information,” Quackenbush said.

Dana-Farber is leading the way in correlating data from patients—everything from tumor biopsies to clinical data—with the vast stores of information scattered across various internal and external systems.

“If we want to analyze genes effectively and really understand what they’re doing in the context of developing disease, we need to be in a position to layer on a great deal of metadata about those genes,” Quackenbush said.

Quackenbush explained, “If you could look for genes whose expression correlates with a favorable outcome, or a poor outcome, you could actually identify patients ahead of time who are most likely to benefit from a particular therapy. This approach
“Bringing information together gives you the ability to find associations that you couldn’t find otherwise. We’re seeing more and more that the Oracle Healthcare Transaction Base is an incredibly powerful research tool.”

John Quackenbush, Ph.D.
Professor of Biostatistics and Computational Biology
Dana-Farber Cancer Institute

could be a big step toward one of the holy grails in current biomedical research—personalized medicine.”

Dana-Farber’s translational research infrastructure, built on Oracle Healthcare Transaction Base, enables researchers to bring together clinical data with sample data, as well as third-party data from multiple sources, including GenBank, a repository of all publicly available DNA sequences, which the National Institutes of Health maintains.

“We realized early on that we needed to be in a position to integrate internal information stored in various locations as well as data available in the public domain. Oracle Healthcare Transaction Base enables us to achieve this goal, efficiently and securely,” Quackenbush said.

**Building the Foundation**

Dana-Farber’s translational infrastructure project began in the multiple myeloma group where researchers were collecting samples from patients enrolled in clinical trials and using DNA array technologies to create a database of these samples. At the same time, they were collecting from each patient corresponding clinical data on blood counts, bone marrow counts, response to treatments, etc.

The clinical data was in one database, existing in one domain, and the sample data was in another database, controlled by another group in another domain. “There was no direct, simple or automated way to bring this information together in a way that would enable us to analyze complex interactions and begin to identify patterns,” Quackenbush said.

The process of bringing together clinical and sample data could take months and did not always yield the desired results. To secure clinical data, researchers submitted paper forms to one group and another form to another group to gain access to the samples. When they finally had access, they often found too few samples available. The back and forth frustrated researchers and added months to projects.

However, by using Oracle Healthcare Transaction Base to pull information out of both databases and link them together in a data
repository, Dana-Farber has enabled query capabilities not previously possible and opened up a new dimension of research. With the Oracle system, the multiple myeloma research team can pull data from each system—the live clinical database, the bio-sample database, and other sources—and combine and correlate the information in a manner that conforms to industry-standard Health Level 7, version 3 (HL7v3) procedures for exchanging medical information.

“Bringing information together gives you the ability to find associations that you couldn’t find otherwise. We’re seeing more and more that the Oracle Healthcare Transaction Base is an incredibly powerful research tool,” Quackenbush said.

**Ease of Use**

To make sure the system was easy for a researcher or principal investigator (PI) to use, Dana-Farber worked with InforSense Inc., an Oracle partner, to design a user-friendly interface that enables users to access clinical and sample data using a single platform and seamlessly connect it with experimental data.

Dana-Farber leverages InforSense’s ClinicalSense product, a Web-based tool for cohort selection and study design that enables users to analyze summary clinical data, compare sample relevant characteristics, and select appropriate sample subsets to optimize study design.

Once the researchers have designed studies, they can analyze the resulting experimental data using the platform’s statistical tools or easily export data to other software tools to identify biomarkers or to develop more complex predictive models.

The combined Oracle and InforSense solution delivers immediate feedback and data control, enabling Dana-Farber researchers to capture and investigate new subtleties and insights they might otherwise miss.

**Accelerating Research**

Dana-Farber’s new data infrastructure helps researchers investigate clinical data more thoroughly, make complex queries and more complete data analysis, and improve experiment design. It also enables more rapid, real-time queries, providing researchers with answers in minutes to queries that previously required days to process.
Further, the infrastructure accelerates the process of initially creating a study.

“If we want to design a study and need to know how many patient samples in our freezers meet a specific set of clinical criteria, it previously would have taken months to find this information,” explained Quackenbush. “Now we are able to sit down and, in a matter of minutes, design a study because we know exactly what samples are available with each clinical criterion.”

**Ensuring Privacy**

With a project of this nature, security and confidentiality of the medical data associated with patients are huge concerns.

To help ensure compliance with HIPAA requirements, Oracle infrastructure software enables Dana-Farber to match clinical data and samples from specific patients and then remove identifying information from the data before presenting it to a researcher.

Dana-Farber also deployed Oracle Identity Management and Oracle Access Manager, components of the Oracle Fusion Middleware suite, to enable researchers to securely access data from any location.

Further, to increase access while maintaining stringent security, Dana-Farber uses Oracle Fusion Middleware to build and manage Web services that facilitate integration with third-party applications for advanced analytics and data mining across the various data sets. For example, the institute leverages Oracle BPEL Process Manager to manage complex clinical data transfers securely, as well as replicate processes that require multiple deployments across the infrastructure.

In addition, Oracle Healthcare Transaction Base supports sophisticated user authentication techniques, including dual-identification methods such as tokens and smartcards.

**Extending Research Tools to Other Facilities**

Dana-Farber’s layered architecture is designed to scale to meet its needs today and in the future. The initial project, the multiple myeloma database, is relatively small with just a few hundred patients and approximately 20 to 30 users. However, Quackenbush noted it was a wise project to start with because it represents “some of the stickier challenges in terms of how we track and manage data here at Dana-Farber.”
With the initial project’s success, Quackenbush and his team plan to expand the system to benefit other areas within the institute.

“The Oracle and InforSense system has generated a great deal of interest based on word of mouth about what this warehouse will allow researchers to do. I think there is great potential in expanding it,” Quackenbush said.

He added that since Oracle Healthcare Transaction Base gives Dana-Farber the ability to draw data from multiple sources, it makes it possible to work securely with third-party partners, which is key to the success of several future projects.

“We’ve demonstrated with the multiple myeloma project a solution that really is scalable, flexible, and doesn’t require any particular institutional infrastructure. We could take the solution and set up an instance for any hospital or research organization in the country, because it does not rely on a particular database or a particular format. It is really a standalone integration center that addresses critical biomedical research needs,” Quackenbush said.

**Why Oracle?**

To address its need to combine sample, clinical, and genomic data, Dana-Farber chose to build its warehouse using Oracle Healthcare Transaction Base. Using this unique solution, Dana-Farber can pull information from its live clinical database, the biosample database, and other sources, while ensuring compliance with HL7v3 procedures.

“We didn’t want to reinvent and maintain a live clinical database,” Quackenbush said. “With Oracle Healthcare Transaction Base, we could reach into the production databases and pull out the types of clinical data we needed.”

Meanwhile, the Oracle data-integration software also performs crucial data-cleansing tasks, such as data de-duplication and normalization, removing any ambiguity in meaning between database records.

**Implementation Process**

The project began in 2007 when researchers in Dana-Farber’s myeloma group approached Quackenbush with the problem of combining sample and clinical data. Conceptually, the solution seemed straightforward.
“We didn’t want to reinvent and maintain a live clinical database. With Oracle Healthcare Transaction Base, we could reach into the production databases and pull out the types of clinical data we needed.”

John Quackenbush, Ph.D.
Professor of Biostatistics
and Computational Biology
Dana-Farber Cancer Institute

“If you have two databases that won’t talk to each other, then why not pull information out of both data bases and link them together in a third repository that you could query?,” Quackenbush said.

However, to make this a reality, the institute had to carefully select a technology infrastructure to support its unique needs.

Dana-Farber met with Oracle Consulting to evaluate existing Oracle products and technologies and share its needs and challenges.

“We developed a sensible plan for how we could combine our internal expertise in integrating research information with Oracle’s expertise in areas where we didn't feel we were best positioned to attempt solutions, such as encrypting data and matching information across clinical records where the data is sometimes imprecise,” Quackenbush said.

Ultimately, the institute selected Oracle Healthcare Transaction Base and Oracle Fusion Middleware as its foundation for developing a prototype warehouse. The conceptual development and use-case development took about one year.

In May 2008, Dana-Farber held a series of demonstrations in which it presented use cases to its stakeholders. A large focus of these presentations was proving that it had met the security and privacy concerns for the system. Following the successful demonstrations, in the summer of 2008, the institute went through a final internal beta test of the software. Then, in September, the system went live, just five months after completion of the design phase.

*Dana-Farber Cancer Institute provides expert, compassionate care to children and adults with cancer, while advancing the understanding, diagnosis, treatment, cure, and prevention of cancer and related diseases. Founded in 1947, it is a principal teaching affiliate of the Harvard Medical School and is among the leading cancer research and care centers in the United States. It is a founding member of the Dana-Farber/Harvard Cancer Center (DF/HCC), designated a comprehensive cancer center by the National Cancer Institute.*