

“Everything we need to know about a patient’s participation in a trial is, for the first time, all in one place, which increases the productivity of our clinical staff and puts us in a better position to guide patients through the trial process.”

— David B. Agus, M.D., Research Director,
Louis Warschaw Prostate Center
Cedars-Sinai Medical Center



Cedars-Sinai Medical Center

Rapidly building flexible applications that leverage large amounts of clinical content to improve physician productivity and patient care.

Cedars-Sinai Medical Center is one of the largest non-profit hospitals in the western United States, with more than 8,500 employees and over 1,800 affiliated physicians across virtually all medical disciplines. It is ranked among the top ten non-university hospitals nationwide in terms of research, and its staff conducts hundreds of studies at a state-of-the-art facility.

Project Goal

A vast amount of useful information on drug trials and patients is untapped because it exists in different formats, scattered among different systems. Cedars-Sinai envisioned a better way for healthcare professionals and researchers to seek answers about various therapies and tailor these therapies to each patient. Their goal was to create a central, clinical content repository that they could easily and concisely search, cross-correlate, and update.

Content Challenge

The form and nature of medical research information and a desire to collaborate with other institutions dictated Cedars-Sinai’s technology choice.

First, most clinical data is actually content, such as patient consent forms and self-reports, medical charts, and scientific research. It is created and stored in a variety of formats, and the different bodies of content are not linked. In addition, gaps often exist in the information due to variations in the way it is collected. Information is also often redundant because different researchers use different terms with the same meaning.

Second, Cedars-Sinai wanted to collaborate with other research facilities, which meant they needed a system

that supported XML and the associated standards for exchanging medical data.

Third, they expected to create a very large database, including information from hundreds of trials and thousands of patients, as well as numerical data from blood and tissue analyses. With advancing technology in spectrometry, gene sequencing, and other innovative approaches to clinical data gathering, the database’s size could reach into the terabytes.

Why Mark Logic

MarkLogic Server, the industry’s leading XML content server, enabled Cedars-Sinai to rapidly develop applications that accommodate a large amount of clinical content. A pure XML system that combines powerful full-text search with the W3C-standard XQuery language, MarkLogic Server provided the capabilities Cedars-Sinai required.

Open XML Standards

MarkLogic Server stores data in XML, allowing Cedars-Sinai to exchange medical data in industry-standard formats.

Flexibility

Cedars-Sinai can define a wide variety of clinical information through metadata, thus they can collect data for many studies with one application. In addition, MarkLogic Server accepts content “as is” from many sources. It does not require that information first be normalized to a static database schema. Together, these features enable Cedars-Sinai to support a broad array of clinical and research activities with a single, metadata-driven application.



Scalability

MarkLogic Server is massively scalable in both storage and performance. It delivers millisecond response times against terabyte-scale contentbases, allowing Cedars-Sinai to gather as much clinical and research data as they need, with no degradation in performance.

Developer Productivity

MarkLogic Server provides a very natural, seamless application development environment. First, XQuery and XML interact in a straightforward manner, which allows developers to write very precise and effective queries. Second, MarkLogic Server tightly fuses the content with the metadata, so everything a developer needs is in one place. Cedars-Sinai built their first application in a few weeks, not months, as would be the case with a relational database.

Applications

Historical Study Repository

This application allows physicians to build and maintain their own repository of content from drug trials that have been reported over several years. Users can author new metadata, add new medical terms and records, define fields and questions, and create web forms, all in real time.

Using a relational database to build this type of application would have been labor-intensive to maintain and update, because the required schema would have been too rigid and incapable of accepting the varying types of data. Moreover, once complete, changing the application would be at best, time-consuming and at worst, impossible.

With MarkLogic Server, physicians can change, update, and add to the repository to support new research projects and their unique data requirements, without disrupting the content or having to modify the application.

Regulatory Correspondence Tracking

This application monitors all regulatory correspondence associated with patients and trials, patient enrollment information, and certain clinical data. It was originally developed over several months using Microsoft® Access and involved numerous tables, queries and forms, but had limited capabilities. In just days, the original application was implemented on MarkLogic Server, and now it does more with less application code: all the forms, queries and tables are tied up neatly together in XQuery, and it also provides search engine functionality.

Clinical Data Repository

The clinical data repository leverages a schema-driven, “instant data collection” application. Written in just three weeks, it collects data for a wide variety of studies. Researchers can add new types of content, such as study descriptions, to the repository without disrupting existing information. To allow for sharing information across institutions, the study descriptions are stored following XML standards from the National Cancer Institute. The application generates web pages and data entry forms, enabling users to view and input information for each study.

In the future, the repository will include definitions of clinical concepts and their relationships to one another, clinical trial protocols and processes, and even gene sequencing information. This content will be used to drive applications for diagnosis, clinic scheduling, machine learning, and more. It will also be the first repository of protein expression data, which is a key to understanding how cancer grows.

“The repository can serve as infrastructure for numerous studies and facilitate the work of many institutions,” according to Dr. David B. Agus, Research Director for the Louis Warschaw Prostate Cancer Center at Cedars-Sinai. “Sharing this knowledge among the wider medical research community, ultimately improves the lives of patients.”

Results and Benefits

The Cedars-Sinai team estimates that using conventional tools to complete everything they built with MarkLogic Server would have required four full-time developers and one or more full-time database administrators. And even then, the applications would not be flexible enough to allow physicians to add new terms and change forms on the fly.

With MarkLogic Server, two part-time developers created the core modules which drive all of the applications in about six weeks. The beauty of Cedars-Sinai’s “proof of concept” approach is that the first two applications became the building blocks for the clinical data repository, shaving an estimated two months of development time off the project.

By building state-of-the-art applications without diverting resources from direct patient care, Cedars-Sinai is able to continue to advance state-of-the-art clinical research.

Mark Logic Corporation
www.marklogic.com
Headquarters
999 Skyway Road, Suite 200
San Carlos, CA 94070
+1 650 655 2300
New York
+1 646 378 2104
United Kingdom
+44 (0) 207 643 1712