

PROVIDING SCALABILITY AND SPEED FOR A NATIONAL TREASURE

The U.S National Archives' Founders Online Project Increases Concurrent User Capacity by 50x



CUSTOMERS:

University of Virginia Press
National Archives

INDUSTRY:

Media

CHALLENGES:

Legacy design; Performance;
Minimal resources

CUSTOMER USE CASE:

Content management & delivery;
Data unification;
Search & discovery

BENEFITS:

Leverages existing IT resources;
Sub-second search; Scalability

COMPANY OVERVIEW

Founders Online is a free online tool commissioned by the U.S. National Archives and implemented by University of Virginia (U.Va.) Press that lets the public access the papers of six of America's Founding Fathers: Thomas Jefferson, Benjamin Franklin, George Washington, James Madison, John Adams and Alexander Hamilton. Funded by the National Historical Publications and Records Commission of the National Archives, Founders Online grew out of 50 years of scholarly efforts and gives unique insight into some of the brightest minds of the Age of Enlightenment. The website provides searchable access to over 150,000 documents, a number that's projected to grow to 175,000.

CHALLENGES

Since Founders Online grew out of decades of academic research, the preliminary challenge facing U.Va. Press was to turn a scholarly tool traditionally accessed by a limited number of researchers into a national resource capable of serving the public at large. It wasn't that U.Va. Press was behind the times – the organization chose MarkLogic early on, in 2004, to develop the XML-based platform that preceded Founders Online. But the original platform was never designed to handle concurrent users at scale; like many organizations, U.Va Press's needs had grown over time.

Increased load: On the old platform, system performance under load deteriorated quickly. When Founders Online was in development, testing suggested the architecture would only support 100 concurrent users.

Legacy design: The design imperative for the original system was to preserve the look and feel of print volumes. While this had a minimal impact on smaller files, longer, outlying collections stressed the system.

Redundant searches: With its focus on form instead of function, the old platform recreated each search from scratch. Even if a user had searched on "Jefferson" and "Independence" previously, the system ran a completely new search query each time a subsequent user employed the same terms. This reinvention of the wheel put an unnecessary burden on computing resources while slowing the system down.

Limited resources: To make matters more complex, the organization had the equivalent of just 1.5 full-time programmers to devote to the project.

“We’re a small shop. Leveraging MarkLogic, we were able to re-architect and re-code an online publication to scale up to a much larger audience. We could do that because MarkLogic is a single-stack platform where the data is mainly textual XML, and search, navigation and rendering are all built in.”

— David Sewell,
Editorial and Technical Manager,
University of Virginia Press

THE SOLUTION

As the only Enterprise NoSQL database platform, MarkLogic has long built content systems that can manage billions of data points quickly. It does so by allowing enterprises to populate data in whatever form they have, not just the inflexible rows and columns of traditional structured databases. A surprising aspect of that enterprise-grade flexibility, however, is that MarkLogic also enables small organizations to increase capacity quickly and accomplish great things without hiring an army of consultants.

WHY MARKLOGIC?

Using MarkLogic’s native search, navigation and rendering capabilities, U.Va. Press didn’t have to rebuild from the ground up to rescale its existing platform. It simply started thinking about queries in the aggregate, instead of on a document-by-document basis. For example, whereas a traditional structured database crawls through millions of rows and columns, MarkLogic uses data mapping to locate relevant documents quickly.

Sub-second search: The result for users of the site – the public – is a quick, Google-like search experience for a remarkable collection of documents written over 200 years ago. Using MarkLogic’s robust querying tools, the customer was able to cut the response time for a large, 90-page document from 19 seconds to just 1.86 milliseconds. When concurrent load increased to 5,000 users – or 50x projected capacity during initial testing – average response was still just 120 milliseconds.

Leverage existing IT resources: By using data mapping to look at aggregate groups of documents, the customer was able to avoid irrelevant results and information bottlenecks to return results quickly and accurately. It also built on previous programming, using existing switches to duplicate processes already in place.

Better data insight: The customer created a static index of previous search results, and prepopulated it before public launch with the most common search terms and links. Now, when a known search term is entered into the system, it simply serves up those existing results, instead of crawling through each stored document again. And the more people who use the system, the richer that stored search cache becomes.

Do more with less: At U.Va. Press, just two programmers – one working half time – were able to redesign the existing platform easily to dramatically shorten response times and increase concurrent load capacity. MarkLogic’s intuitive, user friendly interface allowed them to streamline existing processes and create a simpler, more elegant solution for the new task at hand. In other words, the business didn’t need to spend heavily on outside resources to make it all work.

Achieve customer objectives: “Using MarkLogic and putting it all together, we met all of our performance goals,” Sewell says. “Our endpoint was a completely new, public, open-access website, which we now know as Founders Online.”