Summary

For the 2012 Olympics, the BBC had an opportunity to re-work its infrastructure to handle online content delivery for its growing audience. Beginning in 2010, the BBC started moving away from traditional relational models and its static publishing solution to adopt an Enterprise NoSQL and semantics solution for use in driving dynamic semantic publishing.

The BBC chose MarkLogic’s Enterprise NoSQL database platform to build a content store. MarkLogic provided the BBC with a schema-agnostic database that integrates search, semantics and application services. Throughout the competitions, heterogeneous data (e.g., IOC data, tweets, images and video) continuously – and dynamically – streamed into the website from a wide range of organizations, channels and journalists. This unified, interactive repository was able to handle the crush of over 25,000 transactions per second, to create custom experiences that included 106 million requests for live and on-demand video.

Dynamic Content Delivery

During live-streaming users could choose different views to appear at the bottom of the application, called iPlayer. Here, athlete information populates the screen.
Dynamic Publishing Infrastructure
The Olympics is a dream and a destination – a culmination of years of planning, hard work and dedication that all plays out on a world stage, obviously for the athletes, and in this case, the British Broadcasting Corporation as well. With London the host of the 2012 Summer Olympic Games, the BBC was the host of coverage – and it wanted its online experience to be as dramatic as Danny Boyle’s opening and closing ceremonies.

"Our aspiration was that just as the Coronation did for TV in 1953, the Olympics would do for digital in 2012."

Phil Fearnley, General Manager News & Knowledge, BBC Future Media

The mission was lofty. There were more than 10,000 athletes to cover, a sophisticated audience that was relying more on social media, a need for continuous live coverage – and a demand for multichannel delivery. With only a dozen of journalists, the steadfast process of Static Publishing, which had been relied on for over 15 years, had to be retired in favor of a Dynamic Publishing infrastructure.

Dynamic Publishing has evolved to mean much more than to just dynamically serve content from a database and onto a page. Dynamic Publishing means creating a collection of related data elements and dynamically serving it as audiences demand. According to Matt Turner, CTO Media, MarkLogic Corporation, “getting the right content to the right user in the right format is key to keeping users engaged and maximizing the investment made in content. Like the BBC site, dynamic publishing can have huge payoffs with record-breaking breadth and depth of coverage.”

Non-Stop, Real-Time Delivery
In the case of the Olympic coverage, the flow of content was enormous, non-stop, real-time and went across every channel from web, mobile, tablets and broadcast and included:

- Event statistics and venue data updated in real time
- Athlete and event data from the International Olympic Committee
- Video streams from all 27 Olympic venues
- Feeds from the Press Association and other news organizations
- Stories from BBC journalists
- And of course, endless social media updates from around the world
One example of dynamic updates was the athlete page on über-swimmer Michael Phelps (one of 10,000 athletes), which included a journalist’s story about Phelps, as well as dynamically-generated information that updated real-time, including a box entitled “Phelps Against the World” (his medal count vs others), Phelps’ event performances and where he was swimming next.

It would have been impossible for the team of journalists and editors to maintain this level of detail for the star athletes, and certainly out of the question for every single other athlete that competed in the games – and it was just as impossible for the relational database that had been the workhorse since the 1990s to keep up as well. Dynamic Publishing required a new, flexible architecture that would allow the various data feeds to:

- Increase relevancy with content and data from multiple sources, delivered in real-time
- Make content available with asset search and discovery
- Assemble custom content and collections
- Deliver in multiple formats through multi-channels

**Semantics as a Solution**

So what changed? In a November 2012 webcast, Jem Rayfield, Lead Architect, BBC, explained there were two key components critical to the transformation: An Enterprise NoSQL content store and a “triple store.”

The combination allowed an unparalleled level of automation and dynamic delivery.

The triple store used linked-data technology to automate aggregation, publishing and repurposing of interrelated content objects – all driven by an ontological, domain-modeled information architecture. It’s an organizational system that conveyed that “Michael Phelps” was a member of the 2012 Olympic team, a member of the US swim team, of the men’s swim team, of the 4x200-Meter Freestyle Relay, competed in events and heats and won a variety of different “awards.” With each medal won, a dashboard depicted Phelps’ total medal count against all countries, and was automatically updated in real-time.

The triple store alone, however, could not process and store the massive amount of changing data. To handle the volume and ensure ability to scale, the BBC added MarkLogic to store all assets including stats, tweets, video metadata, images and articles. Video metadata included transcriptions to time-codes, so specific segments of video could be served.
Setting the Gold Standard

The resulting BBC Dynamic Publishing system produced a record number of content pages that maximized the editorial effort by creating the relationships and content and leveraging the system to create the index pages, team pages, schedules and additional content fed to users as they interacted with content.

For example, knowing that viewers were often engaged with a “second screen” – a mobile or tablet from which they could actively engage with social media, the BBC also served display-adjusting content to four channels: Interactive television, computer, mobile and tablet. By encouraging second-screen behavior, social media became an integral part of the coverage. Users could customize their feed by choosing content elements that would appear on the BBC’s iPlayer. They also introduced a new feature that allows users to restart a live feed within iPlayer.

The BBC’s new infrastructure not only provided an excellent user experience, it also helped them break records for viewership, and downloads:

- 106 million requests for BBC Olympic video content
- 55 million global browsers across the games
- 2.8 Petabytes of Data on a single day
- A daily record of 7.1 million UK browsers

After experiencing success with MarkLogic at the 2012 Olympics, the BBC has continued to set the gold standard for digital content delivery by expanding their use of Enterprise NoSQL and semantics into other areas of their business as well.

About the Author

A former journalist and digital media executive, Diane Burley is the Chief Content Specialist for MarkLogic Corporation.