

**Headquarters**

97 Libbey Industrial Parkway  
Suite 300  
Weymouth, MA 02189  
Phone: 781-616-2100  
Fax: 781-616-2121  
E-mail: [info@infotrends.com](mailto:info@infotrends.com)  
[www.infotrends.com](http://www.infotrends.com)

**Europe**

3<sup>rd</sup> Floor, Sceptre House  
7-9 Castle Street  
Luton, Bedfordshire,  
United Kingdom LU1 3AJ  
Phone: +44 1582 400120  
Fax: +44 1582 411001  
E-mail: [euro.info@infotrends.com](mailto:euro.info@infotrends.com)

**Japan**

Hiroo Office Building  
1-3-18 Hiroo Shibuya-ku  
Tokyo 150-0012 Japan  
Phone: +81 3 5475 2663  
Fax: +81 3 5475 2710  
E-mail: [info@gsm.to](mailto:info@gsm.to)  
[www.gsm.to](http://www.gsm.to)

This material is prepared specifically for clients of InfoTrends, Inc. The opinions expressed represent our interpretation and analysis of information generally available to the public or released by responsible individuals in the subject companies. We believe that the sources of information on which our material is based are reliable and we have applied our best professional judgment to the data obtained.

## Dynamic Content Software Strategies Consulting Service

July 11, 2006

---

### Scaling the Content Delivery Process: The View from Oxford University Press

#### Introduction

This case study chronicles one organization's focus on streamlining its publishing process to facilitate the creation and delivery of new products. With the growing use of the Web and other electronic sources for information discovery and access, organizations like Oxford University Press are being pressed to deliver content through multiple channels. They must also create products that are aimed at specific market segments, often utilizing multiple information sources.

For Oxford University Press, each new online project meant starting from scratch. Building a standard publishing platform and using XML for its content format enabled Oxford University Press to quickly produce new products in much less time and re-use work from previous projects. This company provides an excellent example of an organization that:

- Continues to introduce new products targeted toward specialized market needs
- Understands that content for its products can come from multiple information sources, possibly with different content structures, characteristics, and attributes
- Recognizes the inefficiencies of building new workflows and systems for each content product
- Realizes the value that an XML-based publishing platform can provide
- Comprehends the importance of content aggregation and searching as core capabilities for building new products from the growing base of source information

To obtain these benefits, Oxford University Press built a publishing platform based on an XML content server from Mark Logic that enabled it to work with multiple content structures. To make the platform extensible for future needs, Oxford University Press developed an API (application programming interface) that would enable it to work with multiple development firms without major changes to the platform. Benefits are typically accrued across

a range of projects, but Oxford experienced significant results with its very first project.

## Company Overview

Oxford University Press (OUP) has an illustrious history. The company had its origins in the information technology revolution of the late fifteenth century, which began with the invention of printing from movable type. The first book was printed in Oxford in 1478, only two years after Caxton set up the first printing press in England.

Despite this early start, Oxford's printing industry developed in a somewhat haphazard fashion over the next century. It consisted of a number of short-lived private businesses, some that were patronized by the University. In 1586, the University itself obtained a decree from the Star Chamber confirming its privilege to print books.

This was further enhanced in the Great Charter secured by Archbishop Laud from King Charles I, which entitled the University to print "all manner of books."

In 1896, Oxford University Press opened an office in New York that followed all of Oxford's traditions of religious and academic publication. OUP USA is currently Oxford University Press's second major publishing center (after Oxford), annually producing nearly 500 titles.

Since 1896, OUP's development has been rapid in all areas. Music, journals, and electronic publishing have been introduced during the past 75 years, and ELT publishing (which started with books to teach English in Africa and India) has grown into a major international business. OUP is now one of the largest publishers in the U.K. and the largest university press in the world.

The project to create a publishing platform for its online content was led by New York-based OUP, Inc. This architecture and its resulting benefits would then spread out across Oxford University Press.

## Business Drivers

Since the onset of internet publishing, the introduction of each new online product (i.e. OED.com, American National Biography Online) required Oxford University Press to staff its own project team to create a new platform. Apart from shared best practices, therefore, OUP was unable to fully leverage its investments from one product to the next. This approach made sense while Oxford University Press was building flagship products like The Oxford Dictionary of National Biography, but OUP needed a more effective way to build niche products (i.e. a reference-based subscription product for the African American Studies market). This time, Oxford University Press wanted to produce a project in a way that would make it easier to handle similar projects in the future.

Oxford University Press understood that this could be achieved with the creation of a publishing platform. The platform needed to be flexible enough to handle varying functional requirements and diverse types of content, but scaleable enough to meet the needs of large and small projects.

### Oxford University Press at a Glance

**Headquarters:** Oxford, U.K. and New York, NY

**Founded:** 1478

**Annual Revenues:** £458 million (2003)

**Employees:** 3,000 worldwide

**Business focus:** Oxford University Press is one of the largest publishers in the U.K. and the largest university press in the world. The company publishes a range of products from journals to online study centers.

## Challenges

Oxford University Press made a strategic decision to develop its new platform simultaneously with the first product implemented on it. While this added complexity to the project, the company believed that this approach would accomplish two goals. First, it would mean that OUP could more quickly realize the fiscal and practical benefits of the platform. Second and perhaps more importantly, the platform would have a real-world example against which to test itself.

### The African American Studies Center's Requirements

The Oxford African American Studies Center (AASC), edited by Henry Louis Gates, Jr., was developed to be the online authority on the African American experience. To be the preeminent scholarly reference on African American studies, the AASC site needed to aggregate and publish a range of content and formats:

- Core content: Five major Oxford encyclopedias covering all aspects of the African American past, including *Encyclopedia Africana* (5 vol.) and *Black Women in America* (3 vol.)
- Additional source content from more than 18 Oxford reference sources, all editorially selected and reviewed
- Approximately 1,000 images at the product launch, with an ongoing image research program expected to add hundreds more images annually
- 100 primary sources, including the Emancipation Proclamation, Frederick Douglass's slave narratives, and text of speeches. Brief introductory essays accompany each primary source.
- More than 200 charts, tables, and graphs representing demographic information in areas like history, government and politics, business and labor, education, law and crime, and the arts
- Over 140 thematic maps illustrating demographics and history in a variety of areas
- A general timeline of African American history, and specific thematic timelines
- Learning center: Designed for the school market, this will be a growing area that aims to help users explore the content with ready reference tools like country profiles, suggested lesson plans by grade level, and study guides to match the curriculum

The online product must support two types of users:

- Knowledge professionals who want sophisticated searching
- Users who want simple searching

The AASC needed to support two types of users with different expectations for finding information:

- The primary users would be librarians, who are knowledge professionals. This class of users would demand complex search functionality, including boolean, proximity, field, thesaurus, stemming, and wildcard.
- Secondary users would be students, who would want simple searching capabilities in the model of Google.

## Platform Requirements

The two guiding principles for developing the platform, according to Alex Humphreys, OUP's Director of Online Engineering, were "flexibility and scalability." He goes on to say, "These two principles should each be realized in the areas of content creation and management, product development, and project management."

**Table 1: Platform Requirements**

	<b>Flexibility</b>	<b>Scalability</b>
<b>Content</b>	OUP wanted a model where content could be created once, and then used in multiple products. At the same time, the platform could not define the content format. It would need to use and integrate content in a variety of formats, from different sources, and using different DTDs.	The platform could not place any constraints on the volume of content that could be supported.
<b>Product</b>	It was critical that the platform did not dictate product features or capabilities. Instead, it needed to be extensible to support market-driven needs and requirements.	For the platform to be successful, the investment in features, hardware, and resource training needed to be leveraged between products. Each new product could not be a one-off proposition.
<b>Project</b>	The platform needed to provide multiple options for integration with other systems and capabilities. It needed to enable OUP to work with any number of different vendors, suiting the vendor selection to the specific needs of each new project.	The platform needed to provide a foundation that would enable OUP to work on multiple projects simultaneously. This meant limiting its dependence on any one vendor by allowing multiple vendors to interact with the platform using a defined API.

To ensure longevity and extensibility of the platform, OUP decided that the platform would need to be standards-based, utilizing XML to the greatest extent possible.

## Steps to Success

In March 2005, Oxford University Press selected Mark Logic's XML content server as the foundation technology upon which to build its new product platform. MarkLogic Server would provide three things:

- A flexible database for the content that OUP would include in AASC and in all future platform-based projects. As a database, it was specifically designed for the structured XML documents that constituted OUP's content.
- A powerful search engine that would enable customized searching on a per-product basis
- The ability to use the XQuery language, a W3C standard that OUP considers to be cutting-edge

Not long after selecting MarkLogic as the basis for its platform, Oxford University Press chose Boston-based Interactive Factory (IFactory) to build the platform and the AASC product. Having worked with IFactory on the Oxford Dictionary of National Biography ([www.oxforddnb.com](http://www.oxforddnb.com)), OUP was confident in its ability to understand the complex requirements of the platform and the product. The design phases of both projects began in June 2005.

Meanwhile, OUP was assembling the content for the African American Studies Center. The content took a variety of paths into the product:

- The core content from the main African American encyclopedias needed to be tagged into XML.
- Other content was used from other products such as Oxford Reference Online and Grove Music Online; this content required very little work to incorporate.
- Some content needed to be created specifically for the site. This set included articles and features as well as certain metadata regarding the rest of the content on the site.

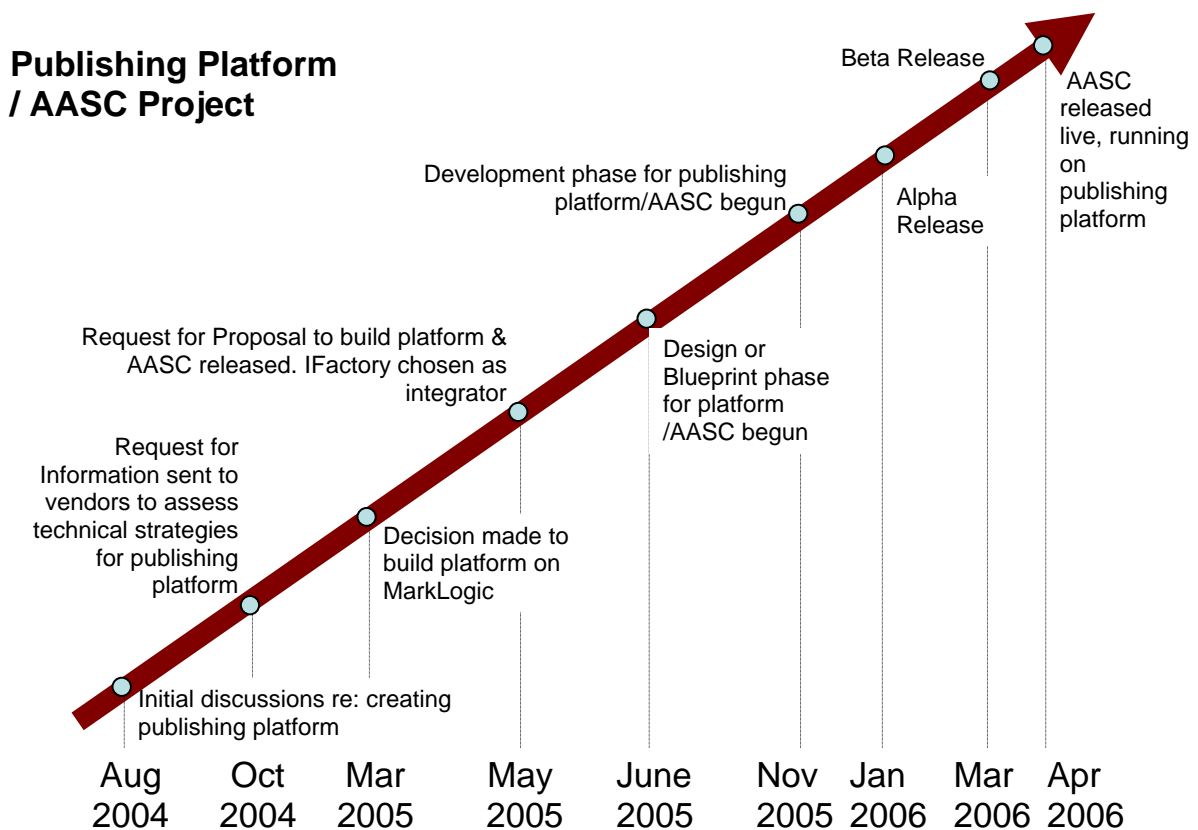
By late fall, OUP had solid designs for the product and the platform. The company also had the bulk of the content created and detailed plans for the remaining elements. With this, the development phase with IFactory began.

The platform was released and tested in an Alpha site in January, and a Beta site in March. The AASC product was launched in late April, in time to market it for the summer academic library season.

The teams responsible for the project consisted of a core group of four people from OUP and an additional four from IFactory. Those teams expanded, however, to include other people and roles when the project demanded it.

In all, OUP spent two years on this project. The first year was spent assessing opportunities and needs. Once MarkLogic and IFactory were chosen, the site and platform were designed and developed in less than eleven months. The schedule was aggressive, but ultimately successful.

**Figure 1: Publishing Platform and AASC Project Timeline**



## Impact of the Results

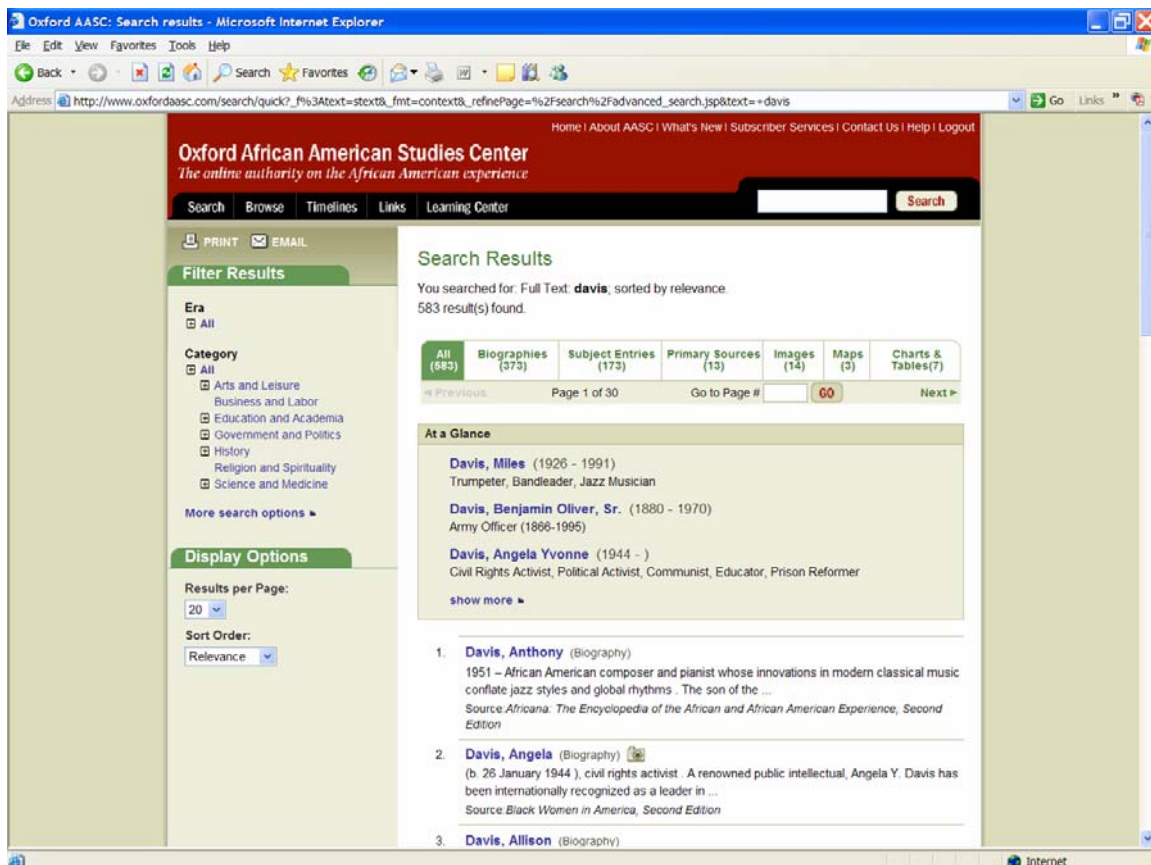
The product implemented on the platform looks similar to existing projects that were done the old way. This is a testament to the fact that OUP met its goal of not having the platform limit the product's design or functionality. The first reviews of AASC have been extraordinarily positive.

"Flexibility is the key unique value of this solution."

**Alex Humphreys**  
Oxford's Director of Online Engineering

The impact of the platform will be felt more significantly on future projects. For one thing, the new project team now has a resilient new platform on which to work. This should enable OUP to diminish the duration and cost of implementing new projects. Under the right circumstances (i.e. in the development of new Studies Centers), the time taken to develop new products could drop from 10 – 12 months to 4 – 6 months.

**Figure 2: Example from the Oxford African American Studies Center Site**



Going beyond the ability to develop and introduce new products faster and more affordably, Oxford University Press is also likely to see:

- More efficient utilization of technical staff, all trained on and fluent in one platform
- The ability to introduce new types of functionality and content in new and existing products, and the ability to combine various content types in ways that were once impossible
- Reduced hardware costs through shared servers and better utilization of resources
- The ability to add delivery channels such as print or even mobile

- Flexibility in choosing third-party vendors to develop new sites

As long as it is similar to an existing project—meaning that the content is of a similar structure (DTD) and that the functionalities of the Web sites are similar—a new project can be replicated fairly quickly. That is not to say that additional costs will not be incurred on future projects. Three things are likely to increase the cost of a new project:

- New types of content, specifically each new DTD "learned" by the platform. Alex Humphreys, Oxford's Director of Online Engineering, noted that "when and if lexical content is added, the platform would need to be taught how to handle it."
- New functionality, such as the ability to print to PDF or wiki-based interactivity
- The need for a new implementation vendor to implement a project. "We understand that there will be a learning curve, and therefore additional costs, when each new vendor implements a product using our platform's API," states Humphreys. "By being able to spread our projects across multiple vendors, we can avoid keeping all of our eggs in one basket."

Humphreys concludes, "The key value proposition of this solution is flexibility. The platform provides us with the flexibility to respond to specific project and product requirements. That could mean launching new products more quickly, or it could mean scaling the platform's hardware to handle a site much larger than AASC. We feel that this flexibility gives OUP a real advantage in responding to the rapidly evolving academic publishing marketplace."

Oxford University Press has three new products to add to the platform, and many more will follow.

## Lessons Learned

As with any project, a number of valuable lessons were learned by the Oxford University Press team as it developed its new platform and rolled out the African American Studies Center.

✓	<p><b>Know your content.</b> In addition to understanding the type of information and its structures, one must also understand how the content is interrelated. Possessing a deep understanding of the information enables you to fully define the content, its metadata, and its publishing requirements. You can then verify that the functionality you have planned properly leverages the content's granularity. The information and relationships play directly into the design of the final product.</p>
✓	<p><b>Understand that technology is not magic.</b> In the end, technology is not a silver bullet that can do everything for you—it is a tool that enables you to speed the development and publishing of information. In the end, it will not provide 100% automation, nor will it eliminate the need to work directly with the information. If used correctly, however, technology can greatly improve the pace and effectiveness of creating and delivering information.</p>
✓	<p><b>Utilize an XML architecture to help ensure flexibility.</b> Using XML helps to remove the dependency of information on any one particular technology, and this ultimately adds longevity to the information and platform. XML-based information can be aggregated with other information, automatically processed, and published in a more automated manner. It reduces the amount of direct interaction that is necessary to reuse or repurpose the content. An XML content server like MarkLogic Server can significantly improve this process.</p>
✓	<p><b>Plan for iterations.</b> Functionality, content, and metadata are often intertwined. For instance, metadata that was not used in the original content may be core to the online product. According to Alex Humphreys, "Iterations are about coordination. By cycling through these iterations, we can simultaneously mitigate risk while ensuring that the entire project team is working towards the same goal."</p>

## Product Reference: MarkLogic Server

Mark Logic Corp. helps its customers accelerate the creation of new information products. MarkLogic Server is an XML content server that enables organizations to build *content applications*, or applications that are based on content and that support business processes. With MarkLogic Server, companies can deliver products through multiple channels, integrate content from different sources, repurpose content into multiple products, build unique information products, and discover previously unknown information. MarkLogic has a complete implementation of XQuery, the W3C-standard for querying XML, and provides very fast and highly scalable content retrieval.

San Mateo, California-based Mark Logic is privately-held and backed by Sequoia Capital and Lehman Brothers. The company serves large, recognized companies in the publishing industry, public sector, and other industries.

### MarkLogic Server Overview

As MarkLogic Server is an XML content server, it is an ideal platform for content applications. With a native XML repository and a complete implementation of XQuery, MarkLogic provides robust support for:

- **Content loading:** Content can be loaded in its current format, converted to XML as necessary, and automatically indexed for later retrieval. Supporting any number of DTDs or XML schemas, MarkLogic Server is able to make sense of a disparate array of content.
- **Query:** With full-text and XML indexes accessible via XQuery, MarkLogic Server provides fine-grained search and retrieval based on document type, content structure, occurrence, and attributes.
- **Content manipulation:** MarkLogic Server provides the means to transform, update, and even assemble new content products based on specific needs.
- **Rendering content:** MarkLogic Server enables companies to format, publish, and deliver content products. If needed, MarkLogic can be integrated with other publishing technologies to facilitate the delivery of content through any number of media or channels.

MarkLogic Server enables organizations with high-volume content needs to automate the processes of content integration, content repurposing, content delivery, and custom publishing.

### Content Integration

Many organizations have numerous disparate silos of content due to unique project needs, mergers and acquisitions, licensing, or changing corporate standards. Much of that content is maintained in different formats. Until the content has been integrated, however, its true value cannot be realized.

MarkLogic Server bypasses the typical content normalization step by importing content “as is.” It uses the power of its query system to cut across all of the types and formats in a content base. MarkLogic supports loading, querying, manipulating, and rendering content in consistent ways even if there are no common DTDs, XML schemas, or formats. Furthermore, MarkLogic Server does not “shred” or “chunk” documents, so there are no problems with “round-tripping” documents to and from the content base. No information is ever lost.

### Content Repurposing

Content Publishers often need to repurpose content into multiple products to increase revenues, decrease costs, or accelerate time to market. The goal is simple—write once, use many. These organizations want to assemble content components to create customized, purpose-built documents and information products.

MarkLogic Server delivers content that is formatted and organized to end-users’ specifications. It can automatically create summary documents that extract relevant content across entire collections.

MarkLogic is the basis of self-service publishing systems that enable customers to assemble their own information collections.

Content repurposing is vital when multiple variants of a document are drawn from common materials. For example, a section of a research article may be combined with different content for Internet portals, print journals, grant applications, reference manuals, and textbooks. In another example, a master technical document about an aircraft may be excerpted to form manufacturing instructions, maintenance manuals, pilot briefing books, and training presentations.

## **Content Delivery**

When documents are ready to be published, syndicated, or otherwise delivered, they must go through an output formatting and rendering process. Preparing the same content for use in a textbook and a magazine means dealing with different document organizations, formats, and layouts. Meanwhile, publishing the same content for a Web portal, on-demand PDF, or handheld document reader requires further output rendering.

MarkLogic Server provides the services needed for delivery of content that is optimized for multiple output formats, media, and devices. It also provides a content processing framework that offers custom content processing pipelines to coordinate multi-step production processes. The content processing framework has a Web services interface for easy integration of third-party tools and applications, enabling MarkLogic Server to be seamlessly integrated into existing business processes to create efficient content delivery systems.

## **Custom Publishing**

Custom publishing is the on-demand assembly and production of documents from existing content. With a custom publishing system, users can build a custom document, book, article, newsletter, or collection of articles simply by “snapping” together content components from different sources.

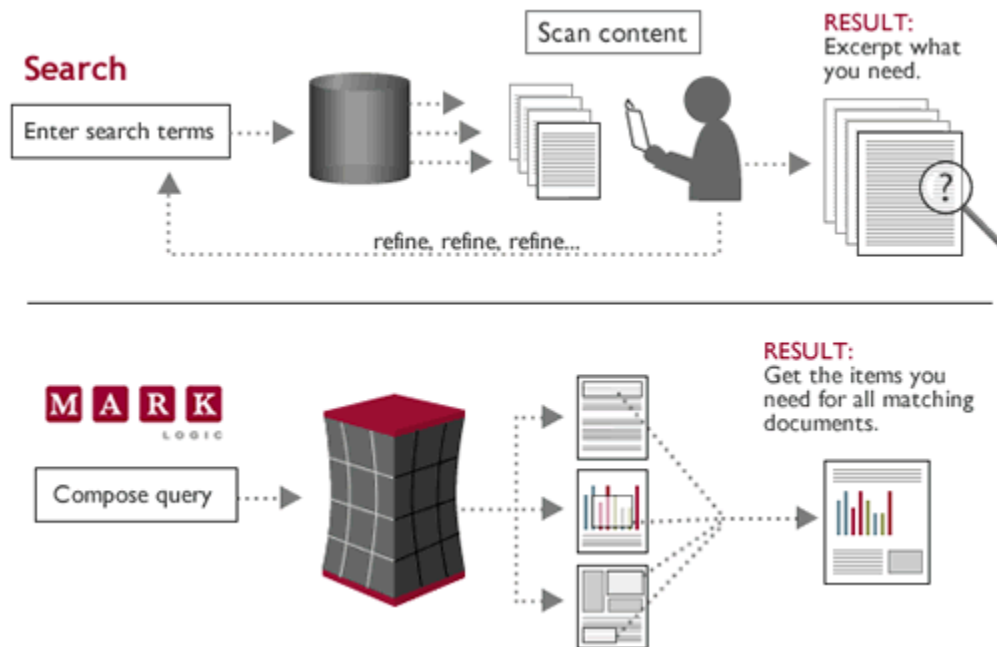
MarkLogic Server provides the services that enable content to be located, selected, and then assembled into a complete book or whatever format is required.

## Search and Discovery

Traditional enterprise search engines can often prove useful, especially when you know exactly what you are looking for. All too often, however, the potential matches are too numerous to even review. Furthermore, a user may not always know what he or she wants to find.

MarkLogic Server provides extensive enterprise search capabilities (i.e. phrase, Boolean, wildcard, proximity, parametric, range searches) and language processing features (i.e. stemming, thesauri, spell checking). All of these features are implemented in a platform that can scale to handle terabytes of content.

**Figure 3: MarkLogic Server vs. Enterprise Search Engine**



## InfoTrends' Perspective

The Oxford University Press story is an excellent example of a company that made an infrastructure investment in a standardized XML publishing system that enhanced its ability to quickly develop new products while also reducing costs. The Mark Logic-based publishing platform enabled OUP to develop new projects in less time than it had taken to produce one-off projects. This directly translates into a better utilization of resources, the ability to respond more rapidly to changing market conditions, and the flexibility to respond to new requirements and content in the future.

XML-based publishing solutions are gaining favor as customers and their vendors realize that information consistency is critical and metadata is paramount for a wide range of publishing projects. In fact, content-centric applications that assemble and distribute content through multiple touchpoints using open standards such as XML are no longer in the minority. OUP's willingness to grapple with its online publishing problem using an XML-based publishing approach that is built on MarkLogic Server will enable the company to better serve customers in the future as their information needs continue to evolve.

*Mike Maziarka, Craig Cline*