MarkLogic | 67 bricks

ADDRESSING ACADEMIC PUBLISHING CHALLENGES

MARKLOGIC & 67 BRICKS WHITE PAPER

Academic publishers need to create new revenues and products to meet the needs of today's authors, researchers, and librarians. Learn how they are using NoSQL and content enrichment to transform their data and content assets to address major publishing challenges.

SUMMARY

There are a number of major challenges facing academic publishers today in meeting the needs of authors, researchers, and librarians.

These challenges include how to manage content assets across journals and books, how to create content flows that can scale and are efficient, how to provide research communities with the most relevant and useful content and data, how to keep abreast of fast changing delivery channels, managing increasing pressure on the price of the content, how to add value to content to create new products, and how to make the content easily discoverable.

In addition, publishers need to create new revenues from diverse, disparate and traditional content silos.

A number of publishing organisations are turning to content enrichment and NoSQL technologies to transform their data and content to:

- Make the most of their content assets across the organisation including books, journals and reference material and eliminate content silos
- Enable all the content to be more discoverable by end users, as well as internal users to enable the efficient creation of collections and products
- Enable efficient content creation, editing, and delivery processes and workflows
- Enable new product creation and provide improved discoverability and SEO

The paper will show how this can be achieved whilst at the same time reducing overall cost and time to market, by providing an innovative and agile approach to addressing the pain points of publishers as well as delivering more value to researchers and consumers of the content.

CONTENT ENRICHMENT

Content enrichment is the process of transforming, enhancing, refining, adding structure to, or improving content to make it a more valuable asset. Content enrichment is becoming a core capability of publishers.

It is part of the role of a publisher to add value to the content to provide the required services to its community. The publisher is the trusted organisation to enrich the content to increase the quality of the content assets.

Content enrichment utilises capabilities like entity identification, content classification against a taxonomy, and creation of a content item semantic fingerprint. These capabilities can each drive multiple product features like the display of related content items, improved content notifications, taxonomic browse, faceted search, content collections, semantic search etc.

CHALLENGES

Knowing your content is essential in today's market. Whilst publishers' traditional role is to use their subject matter expertise to produce quality assured publications, they also have a critically important role to play in the integration, enrichment, and dissemination of research. Many publishing organisations are struggling to fulfil their role fully as their content and data are locked away in multiple systems and silos.

Organisations want to add new features to their services, improve discoverability, and enrich their content and often these activities are done manually through separate processes, maybe at the end of the life cycle. These legacy systems and processes slow publishers down in the rapidly changing environment they operate in; just adding extra processes to the end of the cycle does not fulfil all their goals and is an inefficient way of approaching the problem.

In addition, publishers are dealing with a variety of content types and formats – all regularly changing with new metadata requirements and enriched content and linking. This variety of content, formats and changes makes it difficult to quickly adapt the systems, processes, and data repositories to handle this.

Most of the thinking over the years has been based around the traditional printed-page model, and workflows and delivery geared around the printed model. Lots of process optimisation based around making these existing print publication processes more efficient means that changing is harder.

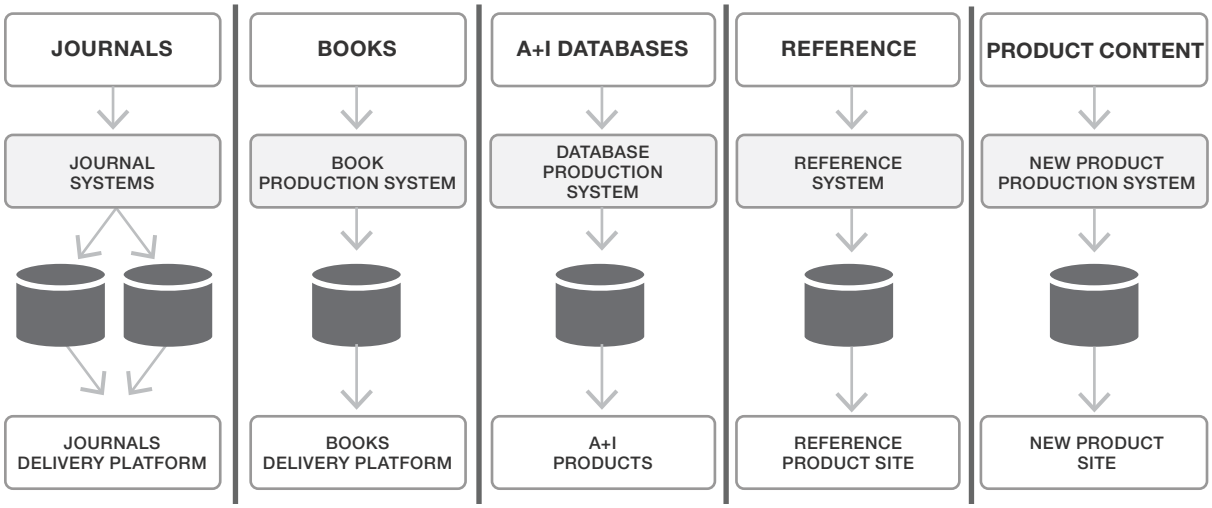

Figure 1: A typical scenario of a publishing organisation with siloed systems, content repositories, processes, and delivery systems.

The key to success is getting the content under better management, more flexible, granular, and integrated. That means managing it so as to be able to enrich it to satisfy the business's key needs today, whilst at the same time leveraging continuous improvement to drive content enhancement and new product creation.

This scenario, often seen in publishing organisations, makes it difficult to:

- Create cross-content collections
- Create new product slices or new products by reusing content
- Avoid duplicating work across multiple products and work streams
- Add enrichment across the publishing portfolio
- Show relatedness across content types
- Manage the content assets and improve discoverability
- Change metadata and add new content
- Increase functionality and content without increasing costs

Additional challenges facing publishers today are:

- How to manage change from traditional page-based models
- Accommodating new supply chain models rapidly
- Provision of open access content to agreed standards and mandates
- The need to add more value to the data associated with the research
- Finding and implementing new revenue streams
- Increase usage with more focussed content marketing and SEO
- Better and more relevant discoverability of content for the researcher
- Providing authors with the best service possible throughout the full life cycle of the publishing process

Publishing models are undergoing a major change and academic publishers are in a great position to be at the forefront at this exciting time: by using and implementing key capabilities such as content enrichment and NoSQL technologies they can be leaders in the new wave of publishing and reap the benefits in the years to come.

CONTENT ENRICHMENT AND NOSQL ARCHITECTURE

By introducing a content enrichment architecture and NoSQL content repository capability, publishers can address the challenges facing them today and provide future proofed capabilities to meet their current and future goals.

A content enrichment and NoSQL content store approach helps publishers develop an integrated capability across the full content life-cycle – from author services at submission through peer-review where appropriate, and into delivery of the enriched content across different channels and models and provision of highly discoverable quality research to users. Using

technologies such as MarkLogic®, the only Enterprise NoSQL database, provides a flexible data model to store, manage, and search the content without sacrificing any of the data resiliency and consistency features of relational databases.

Content enrichment is the process of adding additional structure, metadata, or context to content to provide applications and users with:

- Increased relevancy of content
- Improved discoverability across all content types
- Better linking by relatedness, topic browsing, inline linking to other sources of fact
- Better metadata to improve SEO and usage
- Added new value to content and data opening up new opportunities for new products and topics

Creating layers of enrichment and managing the metadata and associated content in new ways means the publisher can easily create new products, improve discoverability, unite disconnected data, and increase efficiencies in the production workflow.

This approach does not mean replacing your existing systems and processes or building one giant monolithic system. Adding the content enrichment capabilities and content repository into your architecture provides the missing pieces of the jigsaw to assist in solving the current and future challenges the publisher faces.

Creating a repository for enriched content gives the publisher ownership and control over its content assets and enables full use of the publishers' domain knowledge. It is in essence really a knowledge store of all content assets. It adds much more meaning and context to the publisher's existing content assets.

The introduction of content enrichment and NoSQL into an organisation requires clear business objectives, careful process change, planning and the right tools to do it properly and cost-effectively. 67 Bricks are experts in implementing content enrichment capabilities, content repositories and enriched content digital products. 67 Bricks have the required experience to help publishers develop an effective and appropriate strategy, design the technical architecture and implement the agreed solution.

MarkLogic is a new generation database that is built with a flexible data model to store, manage, and search data, without sacrificing any of the data resiliency and consistent features of the last-generation relational databases. MarkLogic is optimized for structured and unstructured data—allowing you to store, manage, query, and search across JSON, XML, RDF (semantics), geospatial data, text, and large binaries. This means faster time to value and an improved ability to handle change.

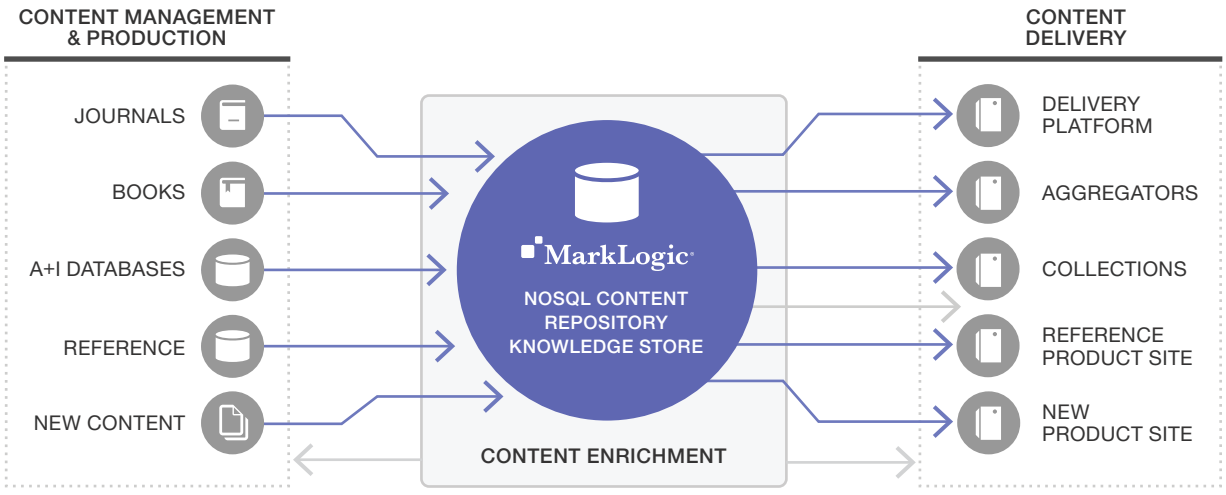

Figure 2: Including a content enrichment capability and NoSQL repository in an organisation's content work flow

Content enrichment architecture with MarkLogic provides publishing organisations with the capabilities for:

- Improved search and discovery across all their content
- Quick integration and deployment of heterogeneous content data (XML, HTML, JSON)
- Semantically enriching and linking content across different content types
- Providing flexibility and scalability
- Managing and delivering enriched content to improve SEO
- Collecting all information from across the entire organisation and organising it with ontologies and semantic relationships
- Discovering and re-using content for different products, collections, personalisation, geographies, and markets
- Measuring content usage and performance, feeding back into continuous improvement and efficiencies.

CONCLUSION

Implementing content enrichment capabilities and MarkLogic NoSQL technologies solves some of the key challenges facing publishers today and provides access to content across business units, streamlining business processes and helping to deliver new features and products rapidly to market. This gives publishers a competitive edge in delivering a better experience to their users.

MARKLOGIC CORPORATION
999 Skyway Road, Suite 200 San Carlos, CA 94070
+1 650 655 2300 | +1 877 992 8885 | www.marklogic.com | sales@marklogic.com