Metadata-Enabled Hybrid Search as the Foundation for Information Governance

By Martin Garland, President, Concept Searching, Inc.

Martin Garland

One of the founders of Concept Searching, Inc., Martin Garland has over 26 years' experience in ECM. His understanding of the information management landscape and his business acumen provide a foundation for guiding organizations to achieve their business objectives using best practices, industry experience, and technology. Garland's expertise has been instrumental in assisting multi-national clients in diverse industries to understand the value of managing unstructured content to improve business processes.

The cloud delivers both benefits and challenges. Offering cost reductions, ubiquitous access, and unlimited storage, the cloud has changed the way companies do business. Internal issues, such as search effectiveness, security, compliance, and information governance, need to be addressed. While SharePoint on-premises feels familiar, cloud-based solutions may not be traditional or proven.

Enterprise information governance can't be siloed. All the way down the application value chain, co-existence and integration enable policy enforcement, and improve compliance, security, records management, data sovereignty, and search. It is only when the on-premises and cloud environments combine in a logical and consistent way that organizations can minimize risk. Cloud challenges can be met by a common element—metadata holds the key. Generally, metadata exists in isolation, is typically created by the end user, and is often erroneous, subjective, or just missing. Without accurate, meaningful metadata and proactive management of that metadata, organizations are unable to achieve multi-platform integration, connected search, or information governance, as nothing is commonly shared between environments.

These challenges require thoughtful evaluation of technologies, to find those that meet or exceed organizational requirements. Deploying unique solutions to address each challenge is costly and inefficient, resulting in siloed applications that might, at best, communicate across on-premises and cloud environments, but not necessarily with other applications within an organization. This approach delays or derails enterprise information governance.

The primary challenge faced within the cloud is security. More sophisticated organizations are well armed on the perimeter, but fall short in protecting knowledge assets. Most data exposures are caused internally, either deliberately or accidently. In either case, a data breach of personal information, customer information, or any data defined as confidential, means the content becomes available to access, download, or share. One breach can result in fines, the need for legal support or remediation, and brand damage. Hand in hand with security goes compliance. Remote or internal users storing undeclared records, either on-premises or in the cloud, can result in expensive and unpleasant audits.

Finally, what of search? Multiple search engines, multiple logins, and information scattered throughout an organization, increase costs. Users recreate existing content, and make decisions without data or based on incorrect information. Many just keep searching, increasing frustration and reducing productivity. Injecting multi-term metadata into the new cloud search service application overcomes these challenges and delivers true hybrid search—the key enabler in an information governance strategy. Conversely, if unusable or meaningless metadata is incorporated into the search index, search quality will degrade and the corpus of content will eventually become unmanageable and of no value.

> *"Without accurate, meaningful metadata and proactive management of that metadata, organizations are unable to achieve multi-platform integration, connected search, or information governance, as nothing is commonly shared between the environments."*

These challenges are just as prevalent on-premises as in the cloud—but why? We come full circle back to metadata. Unfortunately, 92% of SharePoint organizations still use manual tagging, either in multiple choice or free response form. The value of metadata still falls under the radar, business professionals ask "what's in it for me?" and IT professionals realize that, for the most part, manually created metadata is meaningless.

The answer is to automatically generate metadata. Automating tagging with end user involvement is ineffective. Capturing existing metadata is useful, but nuances, relationships, and concepts are lost, unless terms are manually added, which are highly subjective. The somewhat elusive, missing component is the ability to automatically generate metadata capable of identifying concepts within content, in the context in which the content was written, without end user participation. This opens the door to reuse the metadata to drive other processes that enforce information governance.

Many software products perform metadata capture and auto-classification, but SharePoint on-premises and SharePoint Online do not. Surveys indicate that the use of term sets is growing. Ideally, auto-classification of content to a taxonomy or to the SharePoint Term Store should be automatic and synchronized in real time. This approach enables modification and fine tuning of data in a robust, secure, enterprise-class tool, which automatically populates the term sets and eliminates the significant amount of development time typically needed. Concept Searching deploys such a solution and supports all versions of SharePoint—including SharePoint 2016, Office 365, and OneDrive for Business.

The benefits of recognizing the value of concept-based metadata generation is not to be underestimated. Metadata generation, classification, and taxonomy tools enable SharePoint organizations to face head-on their most pressing challenges, regardless of environment—on-premises, in the cloud, or both. Merely resolving the most basic metadata issues can significantly improve hybrid search, enable concept-based search, and provide intelligent migration. Enhanced data cleansing, text analytics, and records management can be achieved, as well as secure collaboration, while eliminating data privacy and confidential exposures before they occur. In other words, concept-based metadata puts organizations on the right path to achieving enterprise information governance. ∎