

How Important is the SharePoint 2010 Term Store? When and **When Not** to use Third Party Tools.

Prepared by:
Concept Searching
8300 Greensboro Drive, Suite 800
McLean, VA 22102
USA
+1 703 531 8567

9 Shephall Lane
Stevenage
Hertfordshire
SG2 8DH
UK
+44 (0)1438 213545

Martin Garland
President
+1 703 531 8567
marting@conceptsearching.com

October 27th 2011

Author Information

Martin Garland has over 20 years' experience in Search, Classification and Enterprise Content Management within the broader information management industry. He has focused on sales, marketing and general management, and has expertise in both startup and turnaround operations throughout Europe, the US and Asia Pacific. One of the founders of Concept Searching, Martin is responsible for both Marketing strategy and North American and International operations.

Overview

SharePoint 2010 adoption is rapidly accelerating and organizations now have a better understanding of the new features and capabilities that have transformed SharePoint from a content repository to an enterprise class platform that enables management of content as a business process, improves governance, and ensures compliance. According to a 2010 Gartner report, improving business processes ranks as the top CIO priority.¹ As the understanding of SharePoint 2010 has matured, organizations are now looking to leverage the native capabilities to capture the value of content and manage it in the context of business processes to meet the priorities of the organization.

Is the Term Store the Panacea to Better Manage and Leverage Content?

The Term Store in SharePoint 2010 is Microsoft's initial building block to enable organizations to natively build taxonomy structures and facilitate the management of large amounts of content, much of which is strategic or of high value to the organization. According to a 2011 Market Survey by Global 360, "The increase in volume of mission-critical content points to SharePoint's increasing credibility as an enterprise capable platform. At the same time, it points to an increased need of business to manage the content as part of mission-critical business processes, improve their governance of the environment, and protect the assets housed within the platform."² Organizations need to capitalize on the strengths of SharePoint and selectively evaluate third-party tools where SharePoint can be leveraged to gain maximum return on investment, both from a tactical and long-range approach.

The Nuts and Bolts

In SharePoint 2010, Microsoft has initially limited a Term Set to 30,000 terms and the Term Store to up to 1 million terms. The practicality is that for most all SharePoint deployments this limitation will be a non-issue. For the majority of SharePoint 2010 deployments, there are significant advantages in utilizing the Term Store including the protection of the organization's investment in SharePoint and achieving business objectives more quickly by working natively in the SharePoint environment. Non-integrated third-party solutions often add complex extraneous layers of overhead, cost, and resources reducing the ability to quickly adapt to changing business requirements. For the small subset of organizations that have large taxonomies, taxonomy tools, such as those from Concept Searching can be used outside of the Term Store to eliminate scalability challenges.

Coupled with the Term Store the Managed Metadata Services support the sharing of taxonomies and terms across multiple site collections, all within the SharePoint environment. This promotes the consistency of content tags based on organizational nomenclature and aligns with policies and procedures that will be required to enforce Term Store governance, in fact any type of taxonomy governance. That said, SharePoint 2010 only offers a manual pick list approach to document classification and minimal user or administrator management of the term sets within the Term Store which is a major deficiency when deploying in an enterprise setting.

Another key component in SharePoint 2010 is Enterprise Content Types which are basically content types published from a central Site Content Type gallery which is called the Content Type Hub, to content type galleries on other site collections (even on other web applications or across farms). If you are looking at SharePoint from an Enterprise Content Management perspective including management of documents and records management, Content Types are instrumental in overcoming end user adoption issues and enforcing records management compliance.

¹ Gartner, The Business unit CIO's 2010 Agenda, 19 February 2010

² Global 360, Inc. 2010 and 2011, www.global360.com

Taxonomies – Not a Top Priority – Think Again

A July 2010 SharePoint survey by AIIM³ indicated that taxonomy management was the 5th top third-party application being considered for SharePoint. As organizations are now more familiar with SharePoint 2010, perhaps it's time to revisit the inherent capabilities and select third-party tools that let organizations extend SharePoint across diverse business functions, while protecting their investment and leveraging the core functionality.

Taxonomies are a key component to improve many business processes. Gone are the days of a siloed approach to improve a single application, such as search. With the SharePoint and FAST search solutions, organizations now have world class search products available basically out of the box. Improving search can be accomplished by complimenting the Microsoft search products through add-on products that enhance the search outcomes but do not replace a best-of-breed search solution with non-standard structures from third party vendors.

Utilizing the Term Store, Records Managers can create taxonomy and define Content Types, workflow, and retention rules from one central location and then push these Content Types across site collections or farms. This can also be used for other applications beyond Records Management following the same steps. For example, to identify personally identifiable information (PII), protected health information (PHI) or any information the organization defines as confidential and route to a secure server so it is unavailable for unauthorized use. Taxonomies can also be used to aid in migration as well as Enterprise 2.0.

When to Use Third Party Tools and Why?

Consider when to use native SharePoint 2010 or a necessary proprietary solution. Clients often mandate a wish to want to deploy Taxonomy Software integrated with SharePoint 2010. Often consideration as to what this statement actually means is given all too little thought and it can be read to mean a number of things.

There are a number of software solutions available in the market today that offer taxonomy solutions that integrated with SharePoint 2010 however the degree of integration varies widely. At one end of the spectrum we see solutions that utilize all of the taxonomy management capabilities that Microsoft has included in the standard SharePoint 2010 product and then add the missing auto-classification features in an incremental fashion.

We also see solutions that ignore the standard SharePoint 2010 taxonomy features and implement proprietary alternatives, just as they did for SharePoint 2007. Most vendors offer one approach or the other.

Why does this matter? The main advantage of the proprietary approach, which ignores the SharePoint 2010 taxonomy features, is that the limitations of the SharePoint Term Store can be overcome: limit of 30,000 preferred terms per term set; limit of 1million terms per term store and the no support for polyhierarchies and certain relationships, as discussed previously. In practice, we find that these limitations are not critical for most customers however vendors with only one solution to offer, a proprietary solution latch onto these deficiencies in an attempt to raise them higher on the evaluation agenda.

The main advantage of the approach that builds upon the SharePoint 2010 taxonomy features is that it is more compatible with SharePoint 2010, from the user's perspective, a developer's perspective and also from the Administrator's perspective. SharePoint users will use the built-in SharePoint facilities to view and edit taxonomy metadata which works with the Microsoft and FAST search engines out of the box.

³ AIIM SharePoint Strategies and Experiences Survey 2010
<http://www.aiim.org/Research/Industry-Watch/SharePoint-2010>

SharePoint developers can customize the SharePoint interface and write applications using the native Managed Metadata controls and not have to learn about non-Microsoft controls. SharePoint Administrators can configure taxonomy metadata by creating list columns and mapping these to content types and sites using native facilities, not proprietary mapping tools.

In summary, unless there is an absolute clear cut necessity for having to adopt the proprietary approach think again as the approach “re-invents the wheel” in a way that will increase risk and costs.

The proprietary approach is likely to involve: additional end-user training, lengthen development timescales increase administration burdens and a lock-in to a proprietary product.

What Can Concept Searching offer? Unlike most vendors in the space, Concept Searching’s conceptClassifier for SharePoint can offer users either a solution tightly integrated with SharePoint 2010, utilizing all of the “out of the box functionality” including the ability to read and write in real time to the Term Store or, alternatively, a proprietary solution that does not integrate or use the Term Store and relies on Concept Searching’s third party extensions to SharePoint.

In a recent Forrester report⁴ Concept Searching’s conceptClassifier for SharePoint was outlined as a strategic purchase decision for corporations looking for Taxonomy and auto classification solutions integrated with SharePoint.

A solution tightly integrated with the SharePoint 2010 Term Store. conceptClassifier for SharePoint 2010 is unique in its integration with the SharePoint 2010 Term Store. Taxonomies (Term Sets) may be built out in the term store and those term sets are read natively and in real time to the conceptClassifier Taxonomy Manager where automatic concept extraction from client’s own content is performed and those natural language concepts are offered to the taxonomist (SME) to assist in the building out of the taxonomy classes and the auto-classification clues. As the taxonomy is built out the structure (classes) are automatically written back to the SharePoint 2010 term store along with the auto-classification clues which are held as custom properties.

As content is uploaded or ingested into SharePoint it is automatically classified to one or more term sets and the metadata is displayed together with the documents in the SharePoint document library.

This approach is fully compatible with existing SharePoint taxonomy facilities. Users and Administrators can modify the structures (subject to access rights) using either the Term Store or from Concept Searching’s Taxonomy Manager. Users can also modify the automatically generated metadata (subject to access rights) using the native SharePoint editing facilities. And the Microsoft and FAST search engines will utilize the metadata in the standard search refinement panel without any need for further development or third-party web parts.

Neither Microsoft Search nor FAST Search for SharePoint offers a browse taxonomy feature. conceptClassifier has a multi-select tree view control web part that allows users to browse the taxonomy when using either of the Microsoft search engines.

⁴ FORRESTER. SharePoint Partners Can plug Taxonomy Gaps, January 2011.

The downside to Solution One

- Polyhierarchies are not supported by Microsoft SharePoint 2010 term store and individual taxonomies are limited to no more than 30,000 preferred terms.

The upside to Solution One

- This approach builds upon what Microsoft provides, adding the missing features in a fully compatible manner.
- As Microsoft enhances the Term Store (for example, to allow polyhierarchies and to increase the taxonomy size limits) then these enhancements are immediately available to customers.
- This is the natural approach for SharePoint 2010 Users, Developers and Administrators.

A Proprietary Solution that does not utilize the SharePoint 2010 Term Store. Concept Searching can also deploy conceptClassifier “de-coupled” from the term store, similar to our solution for SharePoint 2007 and most other vendor’s 2010 implementations.

This approach performs all taxonomy management outside of SharePoint 2010 and publishes the resulting metadata into a proprietary property format, ignoring the existing Managed Metadata Controls.

In this scenario a considerable amount of custom code must be provided because the native formats are not being used. The custom code provided by all vendors who adopt this approach must include: metadata viewing and editing web parts; search enhancements to replace the native refinement panel which will not work with the proprietary controls.

As in Solution One, clients may also deploy the multi-select tree view control web part and the content type updater application.

The downside to Solution Two

- The SharePoint Term Store is completely bypassed.
- Configuration and management of taxonomies will have to be done outside of SharePoint using third party tool sets.
- Linkage between Taxonomies (Term Sets) and Managed Metadata Properties will be lost.
- Integration of Managed Metadata Properties into the Microsoft SharePoint Search and FAST Search for SharePoint search refinement panel will be lost.
- Editing of Taxonomy (Term Sets) metadata using the native SharePoint editing facilities will be lost. Again third party web parts will be required to be used from within SharePoint.
- All Taxonomy (Term Sets) manipulation will need to use custom property types instead of the built-in Managed Metadata Properties.
- Greater risk for clients in that additional third party web parts are required to deliver the solution and all bespoke application development will be forced to utilize non-Microsoft APIs.

All of this will be confusing for users and means extra work for developers and Administrators.

The upside to Solution Two

- Polyhierarchies and very large taxonomies not able to be supported by the Term Store are supported in this configuration.

Going beyond Findability. Concept Searching's offering for Compliance, Governance and Data Asset Protection? Concept Searching recently released to the market a unique application and extension to the conceptClassifier for SharePoint product suite, the contentTypeUpdater (CTU).

Developed as a joint venture with the Microsoft Public Sector Team and Concept Searching to demonstrate full utilization of SharePoint 2010's Information Management and Information Rights Management policy capabilities, the CTU will enable the client to use the metadata associated to specific data assets to drive the automatic application of SharePoint 2010 Content Types. Once documents have the appropriate Content Type based upon content contained within the document, workflows can then be automatically started to ensure proper document preservation and the application of security templates.

In most organizations that use SharePoint documents are often placed in the wrong location, have inappropriate metadata applied, and lack measures to control access and rights management for individual data assets.

Content Types in SharePoint 2010 enables organizations to take advantage of the workflow capabilities that can enhance organizational performance while driving down costs. The only obstacle with content type application is that individuals have to decide which Content Type applies to every document ingested by SharePoint. For organizations with a lot of content this is no trivial matter.

Regulatory guidelines associated with records management, information security, and e-discovery drive the requirement for workflow. Organizations without automated processes that enable records declaration, data transparency, and information security find themselves at increased financial risk when it comes to storing, preserving, securing, controlling, and exposing information.

In SharePoint 2010, semantic, records retention, and security metadata contained in the term store has to be applied to every data asset in order to ensure data transparency, records declaration, and to control access and apply digital rights management. Leveraging conceptClassifier for SharePoint organizations are able to automatically apply semantic, records retention, and security metadata.

Using Concept Searching's contentTypeUpdater organizations can then use metadata associated to data assets to drive the automatic application of Content Types. Once documents have the appropriate Content Type based upon content contained within the document, workflows can then be automatically started to ensure proper document preservation and application of security templates.

Scenarios include:

- Any document that has a digital social security number and/or vocabulary associated with how the organization collects personally identifiable information will be automatically moved to another document library where Information Rights Management will be able to control how the document is used by those who have access to those documents.
- Documents that have been tagged with retention codes are automatically moved to a Records Center library which can then be a staging area for transfer of documents to an end-state records management system.
- Documents that are for example company confidential can have content types automatically applied to enable workflow and processing around retention, reservation, disposition and security.

For a more information and a live demonstration, watch the [webcast](#).

Summary

Organizations are looking for ways to improve governance, ensure compliance and in turn to capture the value of content and manage it in the context of business processes to meet the priorities of the organization.

SharePoint is maturing and with SharePoint 2010 The Term Store is Microsoft's initial building block to enable organizations to natively build taxonomy structures, facilitate and automate the management of large amounts of content, much of which is strategic or of high value to the organization.

The initial implementation of the Term Store has its limitations, although these are rarely important to any but the largest enterprises and third party tools can deliver functionality overcoming these limitations with the down side that these tools often add complex extraneous layers of overhead, cost and resources reducing the ability to quickly adapt to changing business requirements.

Taxonomies in the mainstream have come of age with taxonomy Management being the ranks as fifth top third-party application being considered for SharePoint however Microsoft and most third party application vendors are still focusing taxonomy as a way to improve search and findability. That is however changing with enlightened forward thinking enterprises using taxonomy to address real issues around compliance, Governance and Risk.

Concept Searching is Microsoft's only managed partner in the Taxonomy and Auto classification space and the only third-party vendor that can offer a Proprietary taxonomy solution, for SharePoint 2010, a Taxonomy solution fully integrated with the Term Store and a taxonomy and content type integrated solution that can tag content based upon its vocabulary, apply an automatic content type and drive information rights management automatically. In this way Concept Searching is the only Third-party vendor able to deliver a taxonomy and auto classification solution that can be applied to Compliance, Governance and risk business issues.

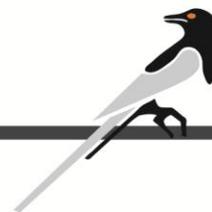
About Concept Searching

Founded in 2002, Concept Searching provides software products that deliver conceptual metadata generation, auto-classification, and powerful taxonomy management from the desktop to the enterprise. Concept Searching, developer of the **Smart Content Framework™**, provides organizations with a method to mitigate risk, automate processes, manage information, protect privacy, and address compliance issues. This information governance framework utilizes a set of technologies that encompasses the entire portfolio of information assets, resulting in increased organizational performance and agility.

Concept Searching is the only platform independent statistical metadata generation and classification software company in the world that uses concept extraction and compound term processing to significantly improve access to unstructured information. The Concept Searching Microsoft suite of technologies runs natively in SharePoint 2010, FAST, Windows Server 2008 R2 FCI, and in Microsoft Office applications.

Headquartered in the US with offices in the UK, South Africa and Canada, Concept Searching solves the problem of finding, organizing, and managing information capital far beyond search and retrieval. The technologies are being used to drive intelligent migration, enable effective records management, identify and lock down sensitive information and enhance governance. For more information about Concept Searching's solutions and technologies please visit <http://www.conceptsearching.com>

© 2012 Concept Searching



Americas
+1 703 531 8567
info-usa@conceptsearching.com

Europe
+44 (0)1438 213545
info-uk@conceptsearching.com

Canada
+1 703 531 8567
info-canada@conceptsearching.com

Australia
+61 (0)2 8006 2611
info-australia@conceptsearching.com

New Zealand
+64 (0)4 889 2867
info-nz@conceptsearching.com

Africa
+27 (0)21 712 5179
info-sa@conceptsearching.com

Marketing and PR
International: +1 703 531 8564
Europe: +44 (0)1438 213545
marketing@conceptsearching.com

Follow us on Twitter
[@conceptsearch](https://twitter.com/conceptsearch)
www.conceptsearching.com