conceptSearching

The ROI of Intelligent Metadata Enabled Solutions
Intelligent Migration

Intelligent metadata enabled solutions are implemented based on Concept Searching's Smart Content Framework™. This enterprise infrastructure is based on a metadata repository where semantic metadata is automatically generated and auto-classified to one or more taxonomies.

The taxonomy component provides organizations with the ability to test, validate, and manage one or more taxonomies resulting in an intelligent metadata infrastructure that address search, compliance, information lifecycle management, sensitive information protection, migration, text analytics, and Enterprise/Web 2.0 challenges.

Enterprise Metadata Management | Search | Records Management | Compliance | Policy Enforcement | Migration
Information Governance | Data Privacy | Text Analytics | Social Networking | Cloud Computing | eDiscovery

Intelligent Metadata Enabled Solutions

The Typical Migration Approach

Organizations are requiring more sophisticated techniques to ensure compliance objectives are met, and a typical loop hole is in the migration process. Simply moving documents from one repository is not enough, as content that was typically unmanaged will remain unmanaged, continuing to expose an organization to risk. Information cannot be managed from inception to deletion without comprehensive metadata associated with the content, and incorporating the multiple channels and origination points from which content was received.

Migration of unstructured content can be a laborious and time consuming project. The challenge is that documents can exist in multiple places at the same time, different revisions of the same document exist, some documents should be deleted, and others should be archived. There may be records that were never declared, as well as confidential or privacy information that will not be identified when migrated. The ability to mass move content is relatively straight forward. However, from an information governance approach, mass moving content results in the same problem of mismanaged content.

The Costs

84% of data migration projects fail *(Bloor)*

72% of organizations delay migration because it is too risky *(Bloor)*

70% of projects reported schedule overruns of about 30% while 64% reported average budget overruns of 16%
(Hitachi Data Systems)

In an Enterprise Strategy Group survey, 39% perform data migration on a weekly or monthly basis

Major risks identified in migration included unexpected or extended downtime, budget overruns, customer impact
(Hitachi Data Systems)

Survey respondents rely on end users to validate whether their data migration was successful or not.
(Enterprise Strategy Group)

The Hidden Costs of Migration

"At the 2012 Compliance, Governance and Oversight Counsel (CGOC) Summit, a survey of corporate CIOs and general counsels found that, typically, 1 percent of corporate information is on litigation hold, 5 percent is in a records-retention category and 25 percent has current business value. This means that approximately 69 percent of the data most organizations keep can – and should – be deleted."

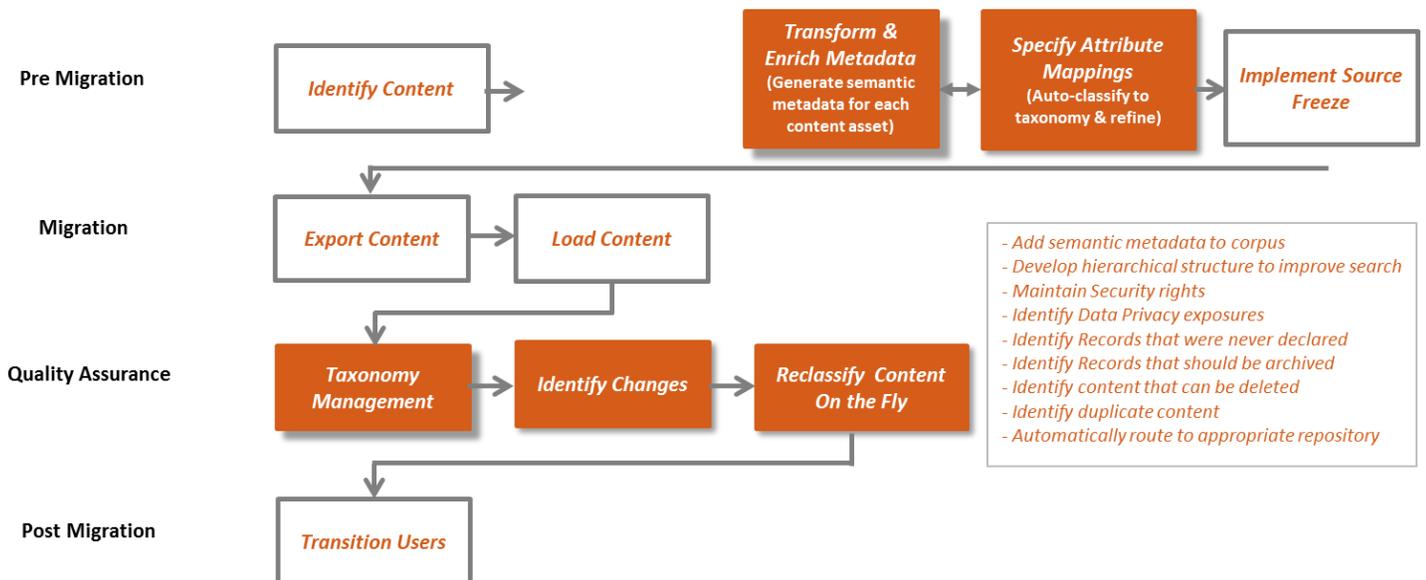**Compliance, Governance and Oversight Council (CGOC)**

The Intelligent Migration Approach

To migrate document collections effectively, the text content of each document needs to be searched to determine its value. This cannot be done manually, as the volume is too high, and the consistency of human review and decision making is unreliable as well as costly. If manually processed, the security rights of the documents as they are moved to their new location must be applied. General migration tools cannot safeguard document confidentiality because they do not make intelligent taxonomy workflow decisions based on the text content of the individual document.

Using Concept Searching technologies, an intelligent approach to migration can be achieved. As content is migrated it is analyzed for organizationally defined descriptors and vocabularies, which will automatically classify the content to taxonomies, or optionally the SharePoint Term Store, and automatically apply organizationally defined workflows to process the content to the appropriate repository for review and disposition.

The approach includes:

- Index content

  - File Shares to File Shares, File Share to SharePoint

  - SharePoint to SharePoint

  - Custom Action – from any other repository (.NET code and Web services)

  - Plug in architecture to custom develop content sources and destination sources

- Connect to Concept Searching taxonomies or the SharePoint Term Store

- Train system to accurately classify content using clues, multi-word concepts, rules, and metadata clues (file properties, file path, keywords, dates, etc.)

- Set-up rules for workflow

- Automatically generate semantic metadata, auto-classify and route to appropriate SharePoint site, Library, or Folder

Pre Migration	*Identify Content*	**Transform & Enrich Metadata** (Generate semantic metadata for each content asset)	**Specify Attribute Mappings** (Auto-classify to taxonomy & refine)	*Implement Source Freeze*
Migration	*Export Content* → *Load Content*			
Quality Assurance	**Taxonomy Management** → **Identify Changes** → **Reclassify Content On the Fly**			
Post Migration	*Transition Users*			

- Add semantic metadata to corpus
- Develop hierarchical structure to improve search
- Maintain Security rights
- Identify Data Privacy exposures
- Identify Records that were never declared
- Identify Records that should be archived
- Identify content that can be deleted
- Identify duplicate content
- Automatically route to appropriate repository

The Bottom Line

Information governance best practices should be applied to the migration of unstructured content. This approach enables rapid document migration as well as the ability to evaluate each document as it is migrated. The end result is a highly effective approach to cleanse irrelevant or unnecessary documents, as well as to identify records that may not been declared or have potential privacy exposures.

conceptSearching

Intelligent Platform Components

The Concept Searching technology platform is comprised of a Service Oriented architecture (SOA) based search and classification technology, a browser based taxonomy management technology, and a tightly integrated feature set that operates with any platform. Industry unique *compound term processing* technology enables the rapid creation of semantic metadata, which can be classified to organizationally defined taxonomies. The tagging and auto-classification of content can be aligned to business goals and the semantic metadata generated can be easily integrated with any third party application or platform that can interface via Web services.

The Concept Searching suite of products is platform agnostic and includes conceptSearch, conceptClassifier and conceptTaxonomyManager. The Microsoft suite of products includes conceptClassifier for SharePoint and conceptClassifier for Office 365. The products use a single code base, able to be deployed in SharePoint 2007, 2010, 2013, and Office 365, providing clients with the choice of on-premise, cloud based, or hybrid environments. The Microsoft products fully integrate with Windows Server 2012, SQL Server 2012, and the former Microsoft FAST products. conceptClassifier for SharePoint operates bi-directionally, in real time with the Term Store, to simultaneously reflect changes to the taxonomy and the Term Store thereby delivering an enterprise class term store management capability.

conceptTaxonomyWorkflow serves as a strategic tool managing migration activities and content type application across multiple SharePoint and non-SharePoint farms and is platform agnostic. This add-on component delivers value specifically in migration, data privacy, records management or any application or business process that requires workflow capabilities.

The technologies also support a variety of search solutions including Solr, Autonomy, SharePoint, Google Search Appliance, and IBM Vivisimo.

About Concept Searching

Concept Searching specializes in semantic metadata generation, auto-classification, and taxonomy management and is a Microsoft Gold ISV and Managed Partner. Concept Searching has a current Enterprise Authority to Operate (ATO) US Air Force, a current Enterprise Certificate of Networthiness (CoN) US Army, and has been deployed on the SIPR, NIPR, and DISA networks.

The technologies encompass the entire portfolio of unstructured information in on-premise, cloud, or hybrid environments. Clients are using the technologies to improve search, records management, data privacy, migration, and text analytics.