

On the Record with SharePoint

Bridging the Gap between Retention and Retrieval

This month, we explore the relationship between the classification and retention of content by users and the retrieval of content to satisfy business and legal requirements. Susan Cisco, PhD, and Jonathan Brandenburg, a Technical Architect, both work for Gimmel on ECM and RM projects. One project in particular was the basis for this article.

How can 9,532 unique types of content created or acquired by an organization with 15,000 information workers be categorized consistently for retention and retrieval? It would be tempting to develop one categorization system to meet both purposes, but Gimmel knew that the “one size fits all” approach would be impractical based on the large volumes of physical and electronic records at this organization. The purpose of this article is to identify an approach for bridging the gap between the requirements for categorizing content for retention and requirements for categorizing content for retrieval.

Enterprise Retention

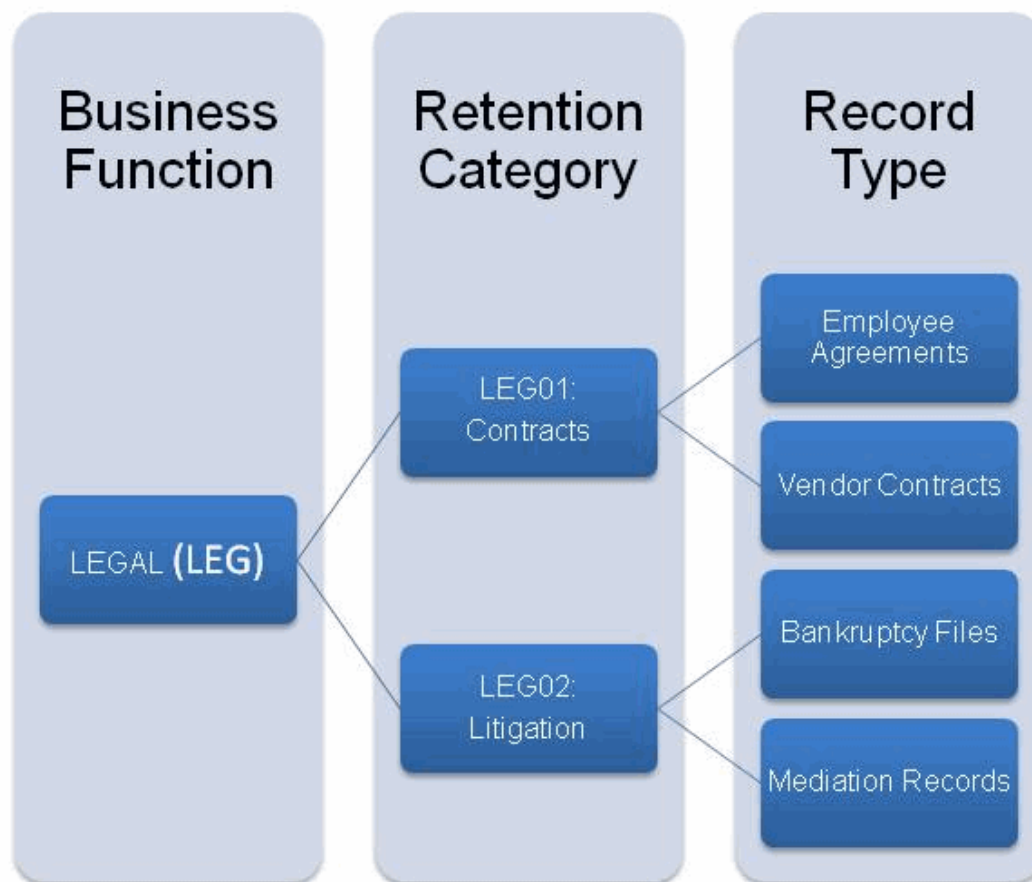
Terminology – In this article, we use the term “content” to include both official records and documents/non-records

A retention schedule (RS) is a formal business policy that lists the types of records an organization or enterprise creates and acquires and how long they should be retained according to legal, regulatory, and business requirements. An up-to-date approved RS addresses legal requirements for all jurisdictions in which an organization operates. A RS typically needs to be updated every 18 to 24 months, especially in highly regulated industries.

Organizations want to make it as easy as possible to apply their RSs and categorize records into the correct retention categories (also known as record series) so they adopt a “big bucket” strategy for streamlining their RSs. A big bucket RS serves one purpose and perspective – defining retention periods. The big bucket strategy simplifies a RS by consolidating record types related to the same business function or process and with similar retention requirements into bigger buckets of retention categories (usually 100-150 for a large organization). With fewer buckets resulting in fewer retention choices, information workers and auto-categorization tools are more likely to consistently categorize records for retention, which ensures better compliance with an organization’s record retention requirements. This, in turn, reduces risks associated with keeping records too long and reduces costs for maintaining and responding to e-discovery demands for large volumes of unneeded records.

Because the sole purpose of the big bucket RS is to define retention periods, the retention categories are structured to be compatible with how laws and regulations relate to a particular business function (e.g., accounting, finance, human resources, and tax) and not necessarily with how information workers retrieve records for business and legal reasons. The categorization system/taxonomy within a big bucket RS is usually based on the international standard for records management, *ISO 15489 Information and Documentation — Records Management*. The 15489 standard notes that the framework for a records management classification system is usually hierarchical and is related to an organization’s business

functions, activities, and transactions. Within such a hierarchical structure, records related to contracts would be categorized following the model below:



Enterprise Retention

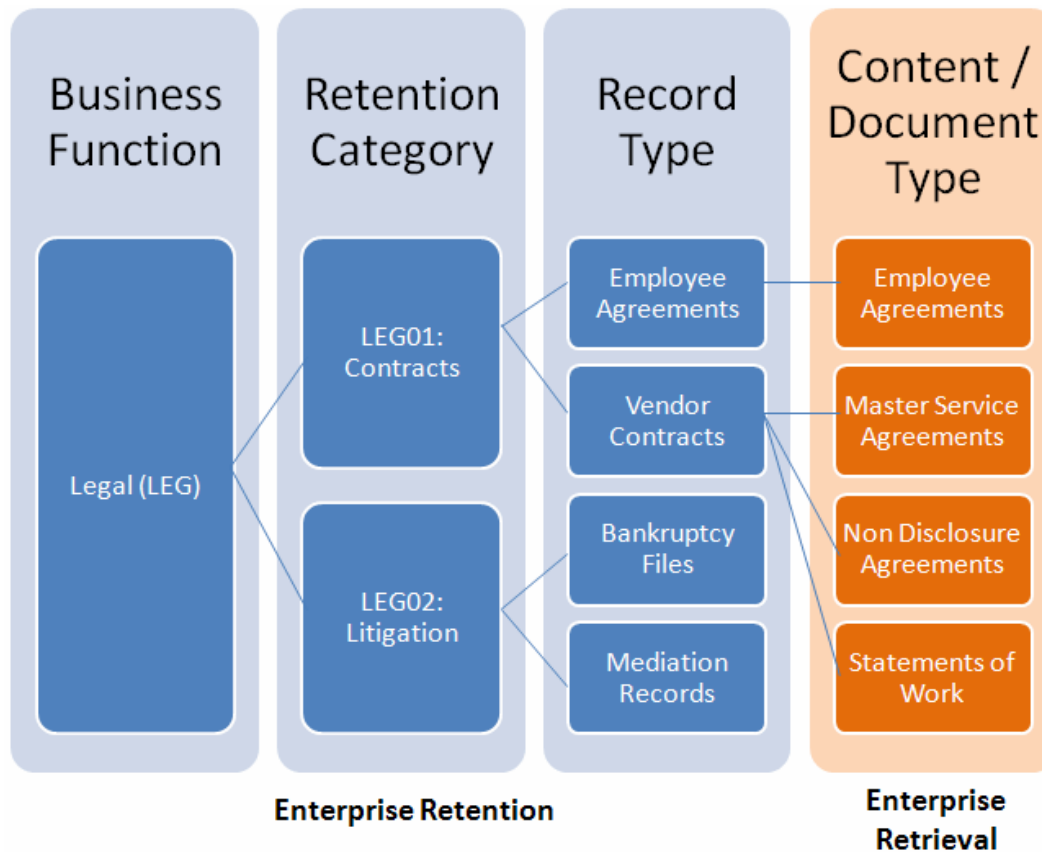
- **Business Function** - Reflects the business function and not the name of a business unit.
- **Retention Category/Record Series** - Is based on the activities constituting the function; retention periods are assigned, and final disposition is managed at this level.
- **Record Type** - Further refinements of the activities or groups of transactions that take place within each activity.

The legal retention research, including laws and regulations for all relevant jurisdictions, is also organized by business function into legal groups. This approach brings together the legal retention requirements applicable to the same types of records, regardless of jurisdiction or source. In *Recordkeeping Requirements*, Donald S. Skupsky, JD, suggests that 15 to 20 legal groups usually suffice to cover just about every record in an organization. Mapping legal groups to retention categories produces the enterprise RS. Due to its complexity, the process of conducting legal research across domestic and international jurisdictions, and mapping it to an enterprise RS is frequently outsourced to outside counsel.

Enterprise Retrieval

We define the term "retrieval taxonomy" to mean a structured, often hierarchical categorization system of concepts and subject categories developed for retrieving content, including both records and documents/non-records. This use, in fact, highlights the key difference between an enterprise RS and an enterprise retrieval taxonomy. While an enterprise RS is focused on the retention and eventual destruction of records, an enterprise retrieval taxonomy deals with finding those records in support of business processes. It is helpful to examine the three primary purposes of an enterprise retrieval taxonomy:

1. **Search for content and support enterprise policies across business processes** - An enterprise retrieval taxonomy defines attributes that are applicable to all content in the enterprise. As these attributes cross business process boundaries, they allow information workers to search for content across the entire enterprise. Examples of such searches include responses to regulatory or legal queries. To select the global attributes for an enterprise retrieval taxonomy, the attributes first must be applicable to a vast majority of the documents in an enterprise. Second, the attributes should support enterprise business processes and policies. Finally, the selected attributes should impose a limited workload on the organization's information workers. It is this third criterion that is most often overlooked. If an enterprise taxonomy requires that information workers manually provide the values for many attributes, the taxonomy will not be effective because the information workers will be unable remain productive. Thus, if an attribute cannot somehow be automatically provided for most content, its inclusion in and value to the taxonomy must be carefully considered. As a starting point for the global attributes, many organizations select attributes from the Dublin Core Metadata Initiative (<http://dublincore.org>) or Department of Defense 5015.2 Electronic Records Management Software Applications Design Criteria Standard (<http://jitc.fhu.disa.mil/recmgt>). Typical organizations select between five and 15 global enterprise attributes, including "Title", "Author", "Creation Date", "Department", and "Retention Category".
2. **Search for content and support information workers within individual business processes** - An enterprise retrieval taxonomy also supports individual business processes. For example, the taxonomy may require that all contracts have certain common attributes such as contract parties, signatories, duration, and approvers. These attributes would apply to all contracts regardless of the corporate department involved in the contract. Thus, in an organization in which departments manage their own contracts, a contract related to information technology and a contract related to marketing would consistently have these attributes even though two different departments were the source of the contracts. These attributes are often called "local" attributes because they are local to a particular type of document, such as the contract in this example. Such local attributes are usually defined by performing a mapping of the business process involving the document or record and identifying attributes necessary to support that process. Interviews with information workers may also be helpful in identifying attributes that support information workers' day-to-day jobs.



3. **Support the Enterprise RS** –The retrieval taxonomy supports the enterprise RS by providing attributes that assist in identifying retention categories for records. Retention categories may be selected directly from a list by the information worker, but, ideally, the information worker naturally identifies the content/record type during the record's information lifecycle (<http://www.aiim.org/infonomics/on-the-record-with-sharepoint.governance.aspx>). For example, when an information worker is creating an invoice as part of a business process, the worker can identify the document as an invoice and the system can automatically assign the retention category for invoices. Conversely, the enterprise RS often supports the enterprise retrieval taxonomy by providing a high-level document categorization system. That is, the record types in an enterprise RS are often the same content/document types found in an enterprise retrieval taxonomy

Then, what is the difference between a content type/document type and a record type? In a relatively simple enterprise retrieval taxonomy, there is little difference. The record types in an enterprise RS provide the catalog of content/document types in an enterprise taxonomy. As an example, consider the retention category "LEG01 - Contracts and Agreements - General". This category contains common contract-related record types such as "Employee Agreements" and "Vendor Contracts" which could very naturally correspond to "Employee Agreements" and "Vendor Contracts" document types in an enterprise retrieval taxonomy with these document types having local attributes such as "parties" and "duration". An enterprise retrieval taxonomy can also contain additional detail. For example, under "Vendor Contracts", it may present specific types of vendor contracts, such as "Master Services

Agreements", "Non-Disclosure Agreements", and "Statements of Work". It is this additional detail in the enterprise taxonomy that allows the taxonomy to support individual business processes and searching within particular content/document types. The relationship between an enterprise RS and an enterprise taxonomy is symbiotic, with the enterprise RS providing a high-level structure for the enterprise taxonomy and the enterprise taxonomy ensuring attributes are present to categorize a record in the enterprise RS.

Conclusion

The categorization systems within enterprise RSs and retrieval taxonomies are related because they reference the same content; however, one system rarely satisfies requirements from both the retention and retrieval perspectives:

- An enterprise retention schedule is primarily concerned with the appropriate categorization of records for retention and ensuring their timely disposition.
- An enterprise retrieval taxonomy is primarily concerned with finding and retrieving content, both records and documents/non-records.

By recognizing the importance of perspective and by striving to understand information worker requirements through research and testing, organizations can organize and categorize enterprise content so that it can be leveraged in digital and physical recordkeeping environments such as SharePoint, Documentum, and OmniRIM.

In our next article, we discuss how to address retention compliance in electronic recordkeeping environments with a sustainable policy and process. We will highlight what other organizations are doing when destroying content and provide a list of eight elements/practices that surfaced repeatedly across organizations and industries.

[Susan Cisco](#) is a Solutions Director in Gimmal's ECM/RM services organization and brings more than 25 years of experience in the records and information management field as a practitioner, educator, and consultant. Susan holds an M.L.S and Ph.D. in Library and Information Science from The University of Texas at Austin. She is a member of ARMA International, and in 2000 was named as a member of ARMA's Company of Fellows.

[Jonathan Brandenburg](#) is a Technical Director with Gimmal and has over twenty years of experience with ECM/RM systems and emerging technologies. In his role as a Technical Architect, Jonathan has assisted organizations with the selection and implementation of technology components, including Microsoft SharePoint, supporting business needs related to Document Management and eDiscovery.

[Mike Alsup](#) is a Sr. Vice President with Gimmal Group (www.gimmal.com), an ECM and RM systems integrator. He blogs at (kqj109.wordpress.com). He welcomes comments or scathing remarks at malsup@gimmal.com.

Originally published February 2010, in AIIM Infonomics Magazine