8 Steps to Develop a Taxonomy

As the foundation for all activities within the corporation relating to documents, a taxonomy can further a wide range of corporate objectives, such as enabling business processes, protecting intellectual property, and building the foundation for compliance. Developing a taxonomy requires the guidance of records management, IT, legal, compliance and the involvement of every business unit.

Carol E.B. Choksy, Ph.D., CRM

A taxonomy is a high-level, hierarchical classification for documents and records that facilitates the management of recorded information throughout its lifecycle. Its focus is on the purpose of information within and how it is used across the organization. Each organization requires a different taxonomy because each has unique processes, organizational configurations, core competencies, and histories.

A taxonomy is a living document that changes as the work within the company changes. It is never final because organizations constantly change their processes and organizational structures, sometimes due to mergers, acquisitions of new business units, or expansion into other countries. A good taxonomy should be flexible enough to handle changes so it never has to be recreated.

As the foundation architecture for managing documents within an enterprise, a taxonomy is the foundation architecture for records management.

Step 1: Select the Taxonomy Team

Information technology (IT), legal, compliance, and records management (RM) must be part of the taxonomy team because of the special knowledge each brings to the project:

• IT knows what the current technology environment can accomplish with regard to documents as well as what the planned future technology environment should accomplish. Part of the process of creating a taxonomy is to discover other technology opportunities or limitations that should be addressed.

• Legal knows about current litigation management problems the corporation is experiencing as well as current and potential statutory and regulatory impacts on specific documents and records series.

• Compliance knows what impacts the current and proposed regulatory environment will have on entire domains of documents and records series.

• Records management knows how end-users actually perform their work and what work-arounds have been created to avoid compliance with records management good business practices as well as gaps within the records management program.

At the Core

This article

▶ Explains what a taxonomy is and its role in corporate strategy
▶ Provides a step-by-step method for developing a taxonomy
▶ Tells how to overcome barriers to successful project completion

Step 2: Determine Role in Corporate Strategy

As the foundation for all activities within the corporation relating to documents, the taxonomy is both a tool and an opportunity (See Figure 1.) The motivation for the taxonomy can be tactical, as with a departmental content management program, or strategic, as with making proprietary business processes more efficient.

Assign the resources, champions, operational and IT leads, and team members who will be interviewed in the process. At a minimum, a
Define the Corporate Objectives to Be Addressed

To ensure the success of the taxonomy and subsequent compliance, the taxonomy creation project must be motivated by corporate objectives. The implementation of an enterprise taxonomy can further a wide range of corporate objectives. For example:

- Taxonomies make business processes more efficient, thus making employees more effective. Most business processes involve human interaction with documents, and the right person having the right documents at the right time is one of the foundations for efficient processes.
- Taxonomies protect intellectual property by identifying those information assets that constitute intellectual property and defining where they intersect and interface with business processes.
- Taxonomies enable confidentiality, security, and privacy by identifying the affected documents and their intersection with the business process.
- Taxonomies are the foundation for corporate compliance by creating a document grid for determining what statutes, regulations, standards, and business practices apply to each.

Include in communications about the project the definition of the corporate objectives that the organization expects to be addressed by the taxonomy initiative.

Define the Corporate Values to be Addressed

A taxonomy provides an opportunity to assign responsibility and accountability for each set of documents. It also provides a basis for training, auditing, and quality control methods. When taxonomy programs are well-implemented, employees develop higher expectations about document handling for each other, creating a self-reinforcing esprit de corps.

Include the corporate values to be addressed by the taxonomy initiative in communications about the project.

Step 3: Determine Business Purposes and Requirements for Taxonomy

This step involves conducting a round of surveys and interviews with representatives of the business units to determine what records-, document-, and content-related problems exist throughout the organization.

Clarify Role in Corporate Strategy

Enunciating a clear strategy in which the taxonomy plays a major role – including within the legal, compliance, IT, and RM functions – is vital to ensure buy-in from individuals who will be asked to participate in subsequent activities of the project, whether in surveys, interviews, or review cycles of draft versions of the taxonomy.

The strategy will also clarify which aspects of records/document/content management will be addressed (e.g., retention schedule, phases of the life cycle, user access, vital records) by the taxonomy initiative so that the survey can be tailored to those functions. Figure 2 on page 34 lists the phases and functions to be considered.

Clarify Information Technology Initiatives to Be Addressed

Content, documents, records, forms, and reports are all captured, created, or generated using IT systems. Each part of the life cycle for each of these categories as well as litigation management, compliance, and audits can also be problematic. The survey and interview process of Step 5 is the opportunity to address these specifically.

Determine Individuals to Interview to Define Requirements

The roles of the legal, compliance, IT, and RM functions will have been outlined as part of the definition of how the taxonomy will fit into the corporate strategy. Representatives of each of these groups must be interviewed for their input into the pressing needs of the organization.

The list of interviewees should include employees from various workgroups that are experiencing severe problems, as well as those that have found workable solutions. Employees from support workgroups should be interviewed in addition to employees from line-of-business workgroups.

Create and Distribute Survey

Before proceeding with interviews, develop a survey to begin to identify the records-, document-, and content-related problems that exist within the various departments and workgroups and to familiarize the interviewees with the types of issues to be discussed in the interviews.
Review Survey Results

The survey responses usually have a clear focus, such as information in silos, Sarbanes-Oxley concerns, too many documents taking up too much space, and difficulties in finding documents.

Use the surveys to prepare for interviews with the representatives of the various departments and workgroups. Knowing a group’s major document- and record-related issues helps form the basis for more detailed questions in the interview sessions.

Interview Individuals

The interview sessions are likely to serve a dual purpose: 1) education concerning the taxonomy, its intended purpose, and what the taxonomy can do for the corporation; and 2) information-gathering toward the definition of the business requirements of the desired taxonomy.

The interviewer should introduce him- or herself, the purpose of the project, and, as briefly as possible, what a taxonomy is, and then permit the interviewee to do most of the talking.

Figure 2: Phases and Functions

| 1. The Phases of the Life Cycle to Be Addressed | 1.1. Policies  
1.2. Creation  
1.3. Capture  
1.4. Distribution  
1.5. Storage  
1.6. Decommissioning  
1.7. Destruction  
1.8. Cultural Archive  
1.9. Procedures  
1.10. Audits  
1.11. Quality Control |
|-----------------------------------------------|----------------------------------|
| 2. Document Control | 2.1. Version Control  
2.2. Copy Control  
2.3. Draft Control  
2.4. Security, Privacy, Confidentiality Control |
| 3. Litigation Management | 3.1. Response to Subpoena  
3.2. Litigation Hold  
3.3. Release Hold |
4.2. Create Procedures for Business Continuity |
5.2. Assigning Accountability, Responsibility |
| 6. Compliance | 6.1. Statutes and Regulations  
6.2. Standards  
6.3. Good Business Practices |

The wording of the survey questions is critical, as many employees tend to identify “records” with paper. Questions should address “electronic documents” and “content” so that the full range of documentary problems within the group can be discerned. General questions such as, “What is the most pressing problem facing your workgroup?” should not be placed in conjunction with questions concerning records because respondents’ answers will address records only, rather than wide-ranging issues such as employee turnover. Such issues nonetheless have an impact on the success of the taxonomy initiative; being able to determine that employee turnover is a problem will help diagnose the requirement: a need for clear accountability and procedures.

The survey should also include explanatory paragraphs alongside questions that reference unfamiliar terms or concepts – for example, a definition of “taxonomy” – to ensure that the survey respondent understands terms and concepts.

The team should allow at least one business week for completion of the survey. Posting it electronically helps ensure the receipt of electronic results.
Analyze Survey and Interview Results

The results of the surveys and interviews will indicate the employee-identified records-, document-, and content-related problems. The taxonomy team should list these problems and concerns on a grid, by employee role and employee concern. The result will be a clear pattern of concerns that will help the team determine which functions and phases of records/document/content management should be emphasized (i.e., phases of the life cycle, document control, litigation management, vital records/business continuity, business process management, compliance). Some functions and phases will not be necessary for some groups.

Review Findings with Individuals and Define Requirements

Review these findings with key individuals to clarify and prioritize business requirements with regard to the corporate strategy.

Review Findings with Information Technology

Technology problems and opportunities, as well as change management and governance issues, may be diagnosed through the interviews, or gaps in the survey may be identified that could be of benefit later in the project. This discussion also helps to focus the project on a prioritized list of potential IT projects.

Step 4: Gather and Review Pre-existing Information

In this step of the taxonomy creation process, the taxonomy team compiles a number of pre-existing materials that will serve as inputs to the enterprise taxonomy.

Gather Taxonomy-related Materials

Information should be gathered from four sources: Step 3 interviews with internal, corporate, subject matter experts such as the records manager, IT managers, compliance manager, and internal lawyers; the Step 5 survey; interviews with the employees of each workgroup or department; and pre-existing materials. The best opportunity to gather pre-existing information is in the interview. Having reviewed the survey at that point, employees have a better idea what might be useful.

Pre-existing Materials

Pre-existing materials include

- Records management materials such as retention schedules, existing file plans, and policies
- Content/document management materials such as file plans and listings of document types
- Business process documentation such as process flows and process maps

Most of the materials will be valuable in either creating or completing the taxonomy, but none will substitute for completing the survey. Make certain to ask questions about the context for the creation of the materials to understand what the material represents and how directly it will translate into the taxonomy.
The Information Management Journal

Step 5: Conduct Survey and Interviews

Determine Individuals and Groups to Interview to Identify Elements of the Taxonomy

The information necessary to create a taxonomy comes primarily from the employees in each workgroup. Only the employees can describe what they do, what they need, how each document and records series interfaces with the business process, and issues and concerns they have regarding documents and records in their area. Often a single person in a division who seeks clarification and pays attention to detail can be more valuable in the interview than a large number of high-level managers who know how things should be done, but not the reality of what is being done. More experienced employees, people who tend to be the go-to individuals for finding things, and people with experience from other organizations are also good candidates for interviews in the Step 5 round of surveys.

RM, IT, and compliance and legal must also be interviewed, but they provide issues that affect the taxonomy, not the content of the taxonomy:

1. The records manager can describe the current programs, policies, and procedures and can also describe the current situation with storage.
2. The IT managers can describe applications and storage management.
3. Legal and compliance can describe current compliance requirements and litigation issues the corporation must address.

Create and Distribute Survey

When the previous steps have been completed, creating the survey is fairly straightforward. Besides identifying information for each individual interviewed (name, title, unit, e-mail), identify the unit’s function to create a context for the information being gathered. The follow-up interview is the opportunity for employees to elaborate on this function including a brief description of the segment of each business process in which they participate.

Interview Individuals and Groups to Explain Taxonomy

The interview is the corporation’s opportunity to educate the employees about the purposes of the taxonomy, to flush out potential issues that will affect how the taxonomy is created, and to help employees understand the scope of the taxonomy project.

The interviewer also must do a great deal of listening and probing about what the employees believe are problems within their areas concerning creation of the taxonomy. These issues may have

IT Systems Materials

Request IT systems materials such as technology architecture and clarification of each electronic repository, its purpose, and section of life cycle managed. Knowing what the technology architecture is will determine what taxonomy solutions are possible.

Analyze Materials

Review the compiled materials with the taxonomy strategy in mind. For example, a retention schedule created by lawyers from different business units and without the input of users or RM professionals may be of no value if it does not capture the categories or the structures of documentary information actually in use by the employees.

Assess Corporate Readiness for Taxonomy

The state of the pre-existing materials and the interviews will provide a clear indication of how ready the corporation is for the taxonomy. This will also help to point up areas of the survey for Step 5 that will require greater detail. Interviewees will have provided information about how accountable each department is, giving evidence about whether surveys can be completed and returned in a timely way. They will also provide information about whether they will conform to the taxonomy after it is created.

Perform Gap Analysis

The gap analysis is performed by comparing the survey structure and final, desired taxonomy with the pre-existing materials. This essentially identifies which departments will have difficulty completing the survey and may require further meetings and which can be assisted through phone calls and e-mails.

Create Explanation for Other Materials Required

Employees create documents about procedures with great care and pride. Although some documents may be inappropriate for the taxonomy, they are appropriate for the individual purposes for which they were created. Therefore, explanations about the inappropriateness of material should not address “deficiencies,” but rather focus on the difference in purposes. Clear descriptions must be presented for what is additionally needed to create the taxonomy.
nothing to do with the taxonomy, but addressing them (not just discussing them) helps ensure the employees believe they are being taken seriously and gives the taxonomy credibility. Without credibility, the taxonomy will not be used and documents will not be well-managed.

While the survey should be sent out ahead of the interview, it should not be completed before the interview to allow the interviewer to provide needed context and definitions. Few individuals have enough knowledge of information management to clearly understand what is being asked of them. The purpose of the project, the goals of the survey, and the purposes of records management must be explained.

Review each slot in the survey with the interviewee and explain difficult concepts, such as records series, document groups, and metadata.

Explain that the goal is to capture 80 percent of their records series and document groups, that the taxonomy is a living document that will change over time, and that more and more records series and document groups will be captured just as new records series will be created and old ones retired.

**Review Interview Results for Potential Barriers to Completion**

The interviews will help reveal what barriers to completion may exist. One problem is simply the provincial nature of work. Many employees answer that the business process is what happens at their desks and may not know from whom documents come nor to whom they go. The interviewer must weave this together.

Another problem is the insistence of some employees that their documents are special and cannot be part of the planned program. Confidential documents, for example, may have to be handled differently, but there are no sets of documents that cannot be included in the process.

Finally, some managers answer that their documents are not part of the program because they do not have “records” or their documents do not have to be “retained.” The answer is that all documents are part of the program and the program addresses destruction and litigation management as much as it addresses retention.

**Determine Issues Related to Policies and Procedures**

The interviews will clarify the specific policy issues, governance, and procedures that must be addressed. Depending upon how compliant and knowledgeable employees are, these issues may need to be more elaborated than a generalized statement. Policies, governance, and procedures should be as brief as possible to ensure they can be changed or added to when necessary. This is also the time to create an outline for these policies, procedures, and governance.

**Determine Opportunities for Improvements Using Information Technology**

By the end of the interviews, the technology needs and opportunities of the different workgroups should be fairly clear. All requirements – business, IT, vendor, and functional – should be easily collected and stated. Functional requirements should also extend to the need to automate business processes, to include forms management, reports management, and what type of repository may perform best for each workgroup. This is the time to create a content/document/records management IT strategy.

**Step 6: Create Inventories**

**Analyze Results of Survey and Interview**

As surveys arrive, they must be reviewed for completeness, and incomplete surveys should be returned promptly with their deficiencies explained.

Review them also to ensure the contents are in alignment with the business unit interview. Where findings are of concern or interest, or if they point to an opportunity, create a table describing these. Any gap must also be filled by asking the business unit or workgroup.

**Determine Taxonomy Issues**

Every organization has challenges to creating its taxonomy that affect how the hierarchical structure of the taxonomy is to be created. Where there is a single exception – and most corporations have at least one exception – a determination must be made how dramatic an effect that single exception is and the team members with experience creating taxonomies must determine the best course of action.

**Creating the Inventory and Records Series**

To create the inventory, the completed information from the workgroup survey is loaded into a single table or set of tables and combed for consistencies and inconsistencies. Inconsistencies that will be problematic for creating the classification are pinpointed. Information is not simply reformatted and fed back to the workgroup; part of the process of creating the classification described in the next step is actually begun here, particularly in creating the records series.

The records series is one of the more complex concepts required in creating a retention schedule. In reviewing the combined survey information, scrutinize each group of documents identified as a records series carefully to ensure it is truly a records series and scrutinize called-out documents to ensure they are truly records series and documents. Descriptions provided by the workgroups usually help determine this.
features of security, privacy, and confidentiality are addressed. This review will address the organization’s requirements for each document and records series as well as limitations and capabilities of the current and planned IT environment.

Departments Review Their Respective Inventories

By returning the inventory to the workgroup using a slightly different format, the employees have a chance to review their input and correct it. Most often this brings the percentage of data collection up to 100 percent. Often workgroups realize there are documents and records series that no longer serve a function within the business processes they manage, and they may want to remove items from the inventory because they no longer reflect the way business is done. This should be noted, but the records series or documents cannot be removed from the retention schedule portion of the project because the records in question could be around for years, and possibly decades. The records series or documents should be removed from the electronic content/document/records management component as this will eliminate unnecessary work.

Incorporate Departmental Revisions

The lead will review the changes and determine whether they are consistent with the project. Most of the time, the changes will be clarifications to the survey. Sometimes the changes will not be consistent with the survey and the programs intended by the survey. These must be discussed with the workgroup and, occasionally, an executive must be brought in to explain why a particular change cannot be made.

Many of the changes made at this point that do not work for the programs are peculiar to the workgroup and either do not need to be included or are a workaround for technology problems. For the taxonomy to work, it must be as consistent and rational as possible, so identifying these issues before the classification is created is very important.

Quite often workgroups will either delete or add metadata because they have a better understanding of what they want out of the process. All metadata must be carefully maintained with the appropriate records series and documents to ensure hand-off to other projects such as the retention schedule and implementation in a content/document/records management application.

Step 7: Rationalize Classification

Create Structure of Classification

Through the interviews, surveys, inventory, analyses, and problem identification sessions, the structure of the classification should be very clear. (For a visual representation of a taxonomy structure, visit www arma.org/pdf/journal/taxonomy.pdf) The taxonomy creator should avoid creating more than four hierarchical levels:

- Function
  - Process
  - Records Series
  - Documents

Rarely does the need to create another level arise. However, when the occasion arises, as in document-focused processes such as loans, architecture, and law, do not hesitate. The additional level should indicate how the business process is divided as the “case” moves through the organization. The point in the classification at which this level sits would have to be determined during the Step 5 interviews. This will assist each workgroup to create a unique set of documents required for its part of the overall process while respecting that all the documents belong to a particular project.

Yet another level may be of assistance in creating a proprietary business process, particularly in document-intensive business processes. No two projects are the same; each has its own particular concatenation of details. There may be sets of projects that bear a strong “family resemblance,” which may be in the circumstances of the project or deal, in the industry type or size of corporation, or in the sets of countries in which the project or deal is to be executed. When those resemblances are identified in the workgroup, they are handled a particular way. Those projects having a particular “family resemblance” and the special handling may become the hallmark of a single workgroup to whom they are always assigned. Creation of the taxonomy is the opportunity to collect this information and formalize it, making it a proprietary business process.

Create Classification of Functions/Processes/Records Series

This cannot be performed before survey information has arrived from every workgroup and the inventory been approved by the workgroup.

Functions are key activities performed by the organization. Functions are not a reflection of the organization chart, but of what the company actually does. Each organization has its own special configuration of functions. For example, every organization has an accounting function, but how that function is configured varies.

Payroll may be only a small process within human resources because payroll activities are primarily outsourced. Another organization may create tax returns or outsource this to an accounting firm. How clearly articulated these functions are determines whether they are functions, business processes, sub-processes, or records series.

Within those functions there are also clearly articulated activities. Where an activity is outsourced, such as payroll, the business process may be to manage the outsourced activity. More than likely,
within accounting there will be accounts payable and accounts receivable, but these may be distinguished from other types of activities where monies are received and disbursed by the organization.

Often a process involving documents moves from one clearly identified activity to another. For example, documents are created in sales that then may move to design and then production. More often than not, the entire movement of documents does not need to be identified as a single process with sub-processes. Where document processing is the major activity to perform the activities of the corporation, this does need to occur.

Because each organization fills a special niche, no set of functions and business processes can be used for every company within an industry sector. In addition, interviewees may be unaware that an activity has become articulated to a point to become a function rather than a process. Occasionally interviewees believe an activity rises to the level of a function when it is only a records series or a business process.

Sometimes changes affect the classification enough that a miniature process of rationalization must take place again.

The classification must also not be confused with billing information. Identifying a document or records series with a client name is not a level of the classification, it is billing information. The customer is known throughout the business process, rather than at the end of production. The unique identifier for a particular file may include a billing number or name for the client and project being performed, but that is a unique identifier, it is not a part of the classification.

Workgroup information must always be retained for each inventory item so that individualized taxonomies can be created for each. The workgroup does not approve the entire taxonomy, it approves only its own. When the time comes to implement the taxonomy in either paper or electronic form, this is done by workgroup. This also means, however, that any changes a workgroup wants to make will affect the taxonomy for the entire organization.

**Rationalize Taxonomy**

Rationalizing the taxonomy requires:

1. Sorting and combing the information to remove departmental identifiers from the names of records series and documents and to provide process differentiators for documents and records series. Quite often workgroups differentiate their work by calling a document “Workgroup 1 Change Order” or “Workgroup 2 Change Order.” The descriptions and metadata for these are identical, so these are both “Change Orders” and should be listed as a single item used by two different workgroups. Conversely, documents that have the same name and are used in different workgroups, but serve a different function should be differentiated.

2. Ensuring the functions, processes, and other levels are concise, consistent, and frugal. Because employees and workgroups do not have the advantage of perspective, they are unaware of where their work fits into the larger corporation. Often the names of processes make work appear different when they are similar or even the same. The interviews are most important for determining where the differences and consistencies are within the classification.

3. Testing the taxonomy for usability by end-users. Sample reports are printed to determine whether the classification will suit a particular workgroup and whether records series and documents can be found easily.

**Distribute Department Components and Receive Comments**

After the classification has been rationalized, each workgroup classification is printed and distributed for review by the workgroup.

Review the workgroup’s returned comments carefully to determine the source of the discrepancies. Where the comments address content provided from the survey, carefully compare the survey, inventory, and classification to see if confusion was introduced at any of these levels.

**Adjudicate Discrepancies**

Where comments from the workgroup indicate the classification is in question, carefully review the role of the records series and documents to ensure the intent of the workgroup was understood. The issue may simply be terminological, in which case the issue should be placed in the thesaurus as an issue. The issue may be that the workgroup does not like the name of the function or the process, so if it can be accommodated, do so.

Where the workgroup cannot be accommodated, but the difference is one of consistency within the taxonomy and does not truly affect the workgroup except aesthetically, work with the workgroup to help the group understand that the needs of the enterprise outstrip its need for internal consistency.

Sometimes changes affect the classification enough that a miniature process of rationalization must take place again. This may affect other workgroups. Where that is the case, those workgroup classifications must be re-circulated.

**Create Draft Taxonomy**

When the changes have been incorporated and the classification has been rationalized and adjudicated, the entire first draft taxonomy can be produced.
Step 8: Finalize Taxonomy  
Departments Review Their Components

Departments make a final review of their classifications and submit final changes. This may actually take several rounds of review. Most of the changes at this point are to fix metadata and clarify descriptions.

Adjudicate Final Discrepancies

Again, any changes that affect the classification materially must be thoroughly reviewed and rationalization must occur when necessary.

IT, Legal, Compliance, and RM Review Taxonomy

IT, legal, compliance, and RM must now review the taxonomy to ensure it is a realistic representation of the environment and that it calls out special handling, such as security, privacy, and confidentiality, as well as litigation management, statutory and regulatory requirements, and lifecycle management issues.

Create Final Taxonomy and Thesaurus

The final taxonomy is then created and reviewed by team members for any defects. This classification must be a fair representation of how business is performed, constrained by the fact that the enterprise must enact controls on how business is conducted.

The thesaurus cannot be finalized until this point because it is part of the rationalization and adjudication process. When companies acquire other companies, a new set of terminology is also acquired. At times the differences in terminology are not relevant, the most frequent example being one part of the corporation using the term “agreement” and another using the term “contract.” There are legal reasons for using either one or the other, but those legal reasons are rarely the basis for these usages. Determining whether those terminological differences make a difference or whether each references the same thing is part of the taxonomy process. The thesaurus should be the list of “approved” terms with clear definitions of what the terms mean and when the terms should be used.

Next Steps

After the taxonomy is developed, the team must create a governance structure that includes creating policies and procedures, assigning responsibilities and accountabilities, and creating training before implementation. For information about these steps and to see samples of the survey and interview questions and a piece of a taxonomy, visit [www.arma.org/pdf/journal/taxonomy.pdf](http://www.arma.org/pdf/journal/taxonomy.pdf).

Carol E.B. Choksy, Ph.D., CRM, PMP, is CEO of IRAD Strategic Consulting Inc. and Adjunct Professor in the School of Library and Information Science at Indiana University, Bloomington. She is the President Elect of ARMA International. She may be reached at cchoksy@iradconsulting.com.