

11 Steps to Building a Knowledge Map

by Jason Bargent

The steps involved in building a Knowledge Map, or K-map, are logical and follow a typical life-cycle approach. The success of your K-map is dependent on aligning it to a business pain or problem, which enables the K-map to directly address that problem. You use an iterative process to create the K-map, initially selecting a small group of data (known as the training set), and then slowly adding to this training set (see, "Glossary of K-Map-related Terms," below, for definitions of terms used in this article).

Once you've spidered the training set and have initially created the K-map, members of your organization's business community (who know the data) use the K-map Editor to teach the Lotus Discovery Server (LDS) the correct category titles and positioning of the documents. As you add more data to the K-map, the less involved this effort becomes, resulting in the business community maintaining the K-map as a small ongoing task. (To learn more about LDS components and K-map security, see "[How to Build an Effective Knowledge Map](#)," May 2002.)

The following steps describe in detail how to build a successful K-map.

Step 1. Identify Requirements

Before you rush off to create a corporate taxonomy or K-map from your organization's actual data, you must clearly define the goal of the K-map. This typically involves identifying a set of business pains and then targeting data sources for the K-map that address those pains. To identify business pains for the K-map to target, you should determine whether

- people are having difficulty locating documents or experts in a particular area (e.g., a department, specific project, product)
- you have too many "islands of data" in your organization, in which only a few people know where everything is stored (e.g., too many departmental intranets, excess file servers with incorrectly filed documents)
- the data that you need to store is in legacy applications or relational database systems, potentially making it hard to access or share
- people know that information exists, but they have too many sources to go to for the information, which can result in confusion
- you're duplicating too much work and could benefit from reuse (this is particularly relevant in a project basis)

A K-map can directly address all these questions. However, these problems may occur across the entire business, so you'll also need to decide the scope and target audience. For example, do you build a K-map for a specific project or department, or do you build a corporate-wide K-map that spans all departments (a first attempt at building a K-map should almost certainly be small and manageable). Note that at present with Lotus Discovery Server v1.1, you can build only one K-map per physical server. Although this may change in future releases, it's important to remember that you can easily extend the scope of a K-map to cover additional communities or departments. If you do, however, want to provide multiple K-maps (e.g., one per department or project), each department and/or project will require its own server. Keep in mind that you build a K-map relevant to a business pain, not for individual departments or projects. At anytime, you can increase the scope of the K-map to add new data sources (you can often achieve results more rapidly by starting small and then expanding).

Step 2. Conduct an Information Audit

Once you've established the business pains and scope of the K-map (such as a specific project or department), you audit all the electronic information used by the target audience. This information can include internal and external Web sites, Lotus Domino databases (including Lotus Domino.Doc as a document management system), files, and

databases.

You conduct this audit to define a detailed list of every data source used that can potentially be targeted for spidering. It's often advantageous to capture the information in a Lotus Domino database so that you can easily sort the contents. A typical information audit database for K-map sources contains information such as

- classification (e.g., Web site/page, file system, Domino database, other database)
- link to source (e.g., URL if a Web page, Universal Naming Convention (UNC) if a remote file or remote file share/folder)
- type of information (e.g., a report, news feed, projects folder)
- whether the information is duplicated in another part of the organization
- who uses the data (e.g., a specific individual, groups of people)
- whether others outside this team benefit from the information
- importance of information to a job role
- how the information is used (e.g., reference, storing information)
- how often users access the information
- whether the information is internal or external (e.g., external may be news feeds)
- who owns the information (from a security/admin perspective)
- whether the information is security protected
- the life span of the information

Most organizations or departments don't have a definitive list of useful information sources, so this list can also act as valuable output of your Knowledge Management initiative, ensuring that people have access to a single catalog of relevant internal information to a specific job role or function.

Step 3. Define Information Sources to Use

Now that you've established a complete list of data sources for the target K-map, subject matter experts (people who know, understand, and work with the data sources) must closely review the K-map as follows:

- Remove useless sources: Almost every information audit will reveal irrelevant information.
- Remove duplicates: Because many departmental Web sites can contain duplicate information, this step is particularly important if your K-map extends across the organization.
- Evaluate each data source for its metadata: Quality metadata is fundamental to the result of an effective K-map. Metadata describes information such as the document title, author, and date. You use this information to show summary information in a K-map. Lotus Domino databases and document management systems are typically the best sources for metadata because behind each document are predefined fields in the forms. File system documents such as Microsoft Office or Lotus SmartSuite documents often aren't the best sources for high-quality metadata (just think how often you fill in the metadata of a word-processing document before you save it). If you use poor metadata, you'll likely need to clean up the metadata before you can begin building the K-map.
- Identify any access or security issues: You'll need to ensure that LDS has appropriate connectivity to reach the data sources. You must test this connectivity for each source and evaluate whether the spider has sufficient security rights to read the data (e.g., is a proxy server or Web page name/password required? Can you grant Manager access for the Lotus Discovery Server to file system sources?).
- Split the list: Prioritize the list based on the value results of one to four to effectively split it into two parts — those that are sufficient to consider adding to the K-map, and those that are insufficient (e.g., poor metadata, no security access).

Now, run another workshop with the same subject matter experts to define the initial set of training data (the training set). You'll use this training set to initially build the K-map, which you do by evaluating each information source and reviewing it against the ideal training set criteria. The training set should

- contain high-quality metadata
- include internal documents (these documents often contain technical terms)
- include lengthy documents (to ensure a wide spread of terms)
- include numerous technical or industry-specific terms

- contain few or no attachments (attachments are considered as separate documents to their containers)

The final training set should be small (a good guideline is 10 percent of your overall data size) and a true representation of the overall source information.

Step 4. Modify the Stopword List

The Lotus Discovery Server maintains its own list of words that are ignored when data sources are spidered. As a result, these words aren't included in the K-map generation, and therefore, the contents of the stopwords list can dramatically affect the results of a K-map.

A generic stopwords list of words (e.g., the, and, about) is supplied as a default starting point from which you should add other meaningless or industry-specific terms. For example, the terms "oil" and "rig" may be too general for a mining-related taxonomy and therefore worth adding to the stopwords list.

To modify the stopwords list, locate the language-appropriate stopwords list for your organization (the English stopwords list is named stopwords_en.txt) in the Discovery Server's data directory (\lotus\ds\data) and make a backup copy of this file. Using a text editor such as Notepad, open the file and enter any new words you want to be ignored ([Figure 1](#)). View some of the predefined stopwords and note how their synonyms are also included in the file. Once you've completed all the changes, save and exit the file (you should make a backup of this new list or document the modifications you made).

It's often useful to test the accuracy of the stopwords list to see whether you need to add any words. To do so, build a test K-map with the identified training source (making a backup before the test K-map is generated and restoring to the backup once you're ready to progress). Follow the instructions in steps 5, 7, and 8 and evaluate the category names generated in the K-map. If ambiguous words appear, you can add them to the stopwords list.

Step 5. Configure the Lotus Discovery Server

Before you can configure the data sources or generate the initial K-map, the Lotus Discovery Server must undergo some initial configuration. From this point on, you perform all configuration and setup of Lotus Discovery Server via the browser-driven Discovery Administrative interface. To access this tool, ensure that you've correctly installed and configured your server (as in the installation documentation in the \lotus\ds\data\help directory) and point your browser at the URL <http://servername.domain.com/discoveryadmin.nsf>. After authenticating with a Lotus Domino user name and password (that is a member of the KDSAdministrators group), you'll immediately see the welcome page ([Figure 2](#)).

To initially configure the Lotus Discovery Server, click Configure Server(s) in the left panel ([Figure 3](#)). A list of servers and their associated spiders will appear in the right window (at this stage, you won't have any spiders configured). Click the server name once and configure the server as follows:

- enable the desired spiders (this is associated with your information audit and the type of data identified as your training set)
- ensure that you've enabled the profile source spider (this spider generates people profiles from your organization's directory)
- enable the K-map Building, K-map Indexing, and Profile Synchronization services

Before saving and closing the window, you may want to configure each spider that you've enabled (based on your own needs). By clicking the spider settings link, you can change settings such as

- maximum size of files or attachments to spider
- file types or attachment types to exclude
- directories to exclude
- spider only specific domains (this is useful when a Web site links to other servers)
- proxy server authentication details

Once you've saved and closed the window, a message will appear indicating that the services will automatically start.

Step 6. Generate People Profiles

The Lotus Discovery Server generates its own directory that it can create from your organization's corporate directory (such as a Lotus Domino Directory or Lightweight Directory Access Protocol (LDAP) server). The entries within the Lotus Discovery Server directory are referred to as people profiles and consist of a person's general contact information, photo, past projects, and affinities as discovered by the K-map. People profiles are an important when working with the K-map. When a name appears as the author of a document, users can click the name and immediately see the author's contact details and known affinities.

The process of generating people profiles involves two steps: (1) generating the profiles from a master or authoritative source and (2) optionally updating the generated profiles from a secondary or supplementary source. The authoritative source can be an LDAP or Domino Directory; the supplementary source can be a Domino database or another directory (although it's easy to import information from a spreadsheet or non-Domino database such as SAP into a Lotus Domino database).

To set up your first profile source (the authoritative source), select Specify Sources under the Generate Profiles option on the Lotus Discovery Administrative console. By default, the Lotus Domino Directory of the domain is included as a disabled source. In this scenario, we'll use the Lotus Domino Directory as the authoritative source.

Our first task is to enable the source and define a schedule so that any changes in that directory are automatically reflected in the profiles. To accomplish this, click the check box to the left of the server name (a checkmark will appear), and then click once on the Actions menu at the top of the window. From the drop-down list, select Enable and define schedule. If required, change the default start time and days on which the profiles will be synchronized. Once you click Save, the profile is activated and the profile spider is scheduled to run at the specified time (default of 4 a.m.). To activate the spider now (outside the schedule), click the check box as before and choose "Add to queue immediately" from the Actions menu.

To verify the status of the profile spider, select Verify Generation under the Generate Profiles option. A window will appear from which you can see all profiles grouped by the first letter of their surname ([Figure 4](#)). By clicking the twisty, you can expand the category and view the generated profiles. Clicking a person's name displays the full person profile.

Most Lotus Domino or LDAP directories don't include a person's full contact, or they lack some of the information from a Discovery Server profile (e.g., past projects). As a result, most people profiles are supplemented by another source. In this scenario, a Lotus Domino database that stores human resource information supplements the profiles. A special technique known as field mapping maps the required fields from the human resource database to the profile form.

To set up a supplementary source, click the Add Source button in the People Sources view. A blank people source form will appear that requires you to complete the following information:

- **Title:** a brief description of the profile for easy identification (e.g., XYZ LDAP Server, HR Database)
- **Type:** the source type (Domino Directory, LDAP Directory, Domino database)
- **Use for:** defines whether the source is authoritative or supplementary
- **Import Scope:** defines whether to import all entries or a subset of entries
- **Field Map:** used to select a default field mapping (for Domino or LDAP directories) or to create a new field map to suite your own requirements; simply insert the Domino database field name next to the profile field name that you want to map
- **Spider Schedule:** enables or activates the spider with a defined running schedule

Note that owners of the people profiles can make changes themselves by clicking the edit icon in the top-right of their profile. Any such edits won't be made effective until you've enabled the metrics processing service.

Step 7. Set Up Data Sources for Spidering

Once you've successfully created the people profiles, you must set up the training set of data in the Discovery Server for the initial K-map to be created. To set up these data sources, select Specify Data under the Create K-map option. Because you have yet to add sources, you'll see the message No Documents Found. Click the Add Source button to display a blank data repository form ([Figure 5](#)), where you must complete the following information:

- **Title:** a brief description of the data repository for easy identification (e.g., HR Database, XYZ file share); because this description is also used in the K-map to indicate the source, make it as meaningful as possible for users
- **Type:** the spider type to use (Domino.Doc database, generic Notes database, Notes e-mail database, QuickPlace database, Windows compatible file system, Web); depending on the spider type, the following information is also required:
 - Domino.Doc database: the Lotus Domino server name, the required Lotus Domino.Doc database, the preferred viewer (Lotus Notes or browser), and the HTTP server name
 - Generic Notes database: the Lotus Domino server name, the required Lotus Domino database, the preferred viewer (Lotus Notes or browser), and the HTTP server name (to spider only specific documents in a Lotus Domino database, create a view called (\$DominoSpider) in the database - LDS will spider only documents that are in that view)
 - Notes e-mail database: the Lotus Domino server name and the required mail database; e-mail sources aren't indexed for use in the K-map but instead are used to identify only affinities
 - QuickPlace database: the Lotus Domino server name, the required Lotus QuickPlace database, launch attachments in the native application, and the server HTTP name
 - Windows compatible file system: the directory to spider in UNC format (\\<machinename>\<share>), the HTTP server name, and whether to include all subdirectories, only the identified directory, or a specified number of subdirectories
 - Web: the URL for the Web page, any necessary authentication details, the number of links to follow, and whether to follow links to other servers (critical for containing a spidered Web page)
- **Spider Schedule:** enables or activates the spider with a defined running schedule
- **Field Mapping:** defines how the source metadata is mapped to generic spider fields such as author, creation date, last read date, title, subject, revised by list, revision date list, body, keywords, and summary; typically, the default won't match every source you use, so you'll need to customize field mappings to ensure the correct metadata is used

As [Figure 6](#) shows, once you've enabled and saved the source, a message appears saying that the documents are being collected. Every source must be processed before you can move on to the next step — the creation of the K-map.

Step 8. Create the K-Map

Once you've set up and processed the training set, you're ready to create the K-map. You begin by selecting Begin K-map Creation under the Create K-map option. This process involves three steps:

1. **Add repositories to process:** Click the Add Repositories button and add the previously defined repositories (if one is missing, it's either because it failed in the previous step or because it must still be read by the spider).
2. **Review K-map settings:** This is where you define the K-map creation and categorization rules and who can edit the K-map. You can either accept the defaults or modify the settings for a maximum number of documents, maximum number of subcategory levels, and maximum number of categories in which a single document can be placed.
3. **Create K-map:** Once you're comfortable with the K-map settings and have added all the training sources, K-map generation can begin. Before clicking the Create K-map button, you should perform a full system backup because this step can't be undone.

To monitor progress of the K-map creation, select Monitor Creation under the Create K-map option. This will open the K-map building log and display the results of the K-map creation ([Figure 7](#)). This log will indicate the number of documents too small to include in the K-map, the number of categories created, and the number of documents added to those categories.

Step 9. Train the K-Map

Lotus Discovery Server doesn't automatically create a working K-map, so even after creation, the K-map isn't ready for use. Further training of the K-map is required. Subject matter experts must teach the system the correct category labels and document placement.

This training process is known as K-map editing. Because the K-map learns by example, the training process is an

iterative cycle. Subject matter experts incrementally added data sources and edit the K-map until all sources have been added. The objective of this editing process is to teach the system and ensure that meaningful category labels appear for easy browsing and descriptive affinities. Editing the K-map involves the following steps:

- **Understand the taxonomy:** Before you begin editing the K-map, it's useful to browse through the K-map using the K-map Editor or, if possible, print the K-map to help you understand the categories and document placement (e.g., which documents have been placed in which category).
- **Modify category labels:** Notice that the K-map's initial categories appear unusable, as [Figure 8](#) shows. To modify category labels, review the categories and the documents within them, renaming categories and creating new subcategories where appropriate. This process is not optional but an essential step as it provides easy navigation for users browsing the taxonomy in the K-map and lets people retrieve their affinity name for discovered affinities from these categories.
- **Move documents between categories:** Once you've renamed the categories, you must review the documents and move them into the correct category. Remember that the K-map is dynamic, so it will learn from every document that you move. For example, if you manually move 10 documents, the system remembers this manual placement. The next time you run the K-map indexing process, the system will position new or updated documents based on the manual editing.
- **Merge and delete redundant categories:** Finally, you rationalize the categories by combining similar categories and removing redundant one. Remember that you don't want your taxonomy to be too large; otherwise, it will be unusable.

You can perform all these tasks using the K-map Editor tool. This tool is a Windows application that you install on your workstation from the URL <http://<yourserver>/kmapeditorinstall.html> or by following the link from the Edit K-map under the Refine K-map option in the Discovery Administrative console. To launch the tool, click the icon on your desktop and enter your user name and password. Provided that you've been granted access as a K-map editor, you'll see the editor tool and your initial K-map ([Figure 9](#)).

In the left panel of the K-map Editor, you'll see the K-map's category labels. The right panel displays the documents within the selected category. Each document entry indicates the name, fit value (describing how well a document fits in the current category - the higher the number the better the fit), last date of change (e.g., a move), placement (automatic or manual), number of categories in which it exists, spider type, value (calculated by the metrics service), and author. To move documents, drag them into the appropriate category. Double-clicking the name of a document opens the document for viewing.

It's important to remember that no automatic process can create an exact taxonomy that represents your organization, so the iterative editing process is essential for teaching the Discovery Server how to position new data sources. Over time, you'll spend less effort editing the K-map because the system understands document placement in context to your organization's data. Carefully consider how many iterations you want performed and the size of the documents in each iteration. For example, consider choosing three iterations, starting with the training set and then adding two more iterations of the entire data source (however, depending on the size and quantity of data sources, you may want to use more or fewer iterations).

As practical example, suppose you have a total of 300 data sources, in which 20 of these sources are identified as the training set and used on the first pass of creating the initial K-map. Once you've edited these 20 sources, you then add about another 150 document from the original 300 data sources and then, following further editing, add the remainder of the 300 sources. Best practice shows that this approach is faster and more effective than spidering all 300 sources at once because the system learns from the iterations of edits.

Once you've created the K-map and complete all editing, you must assign the spidered documents to the K-map categories by enabling categorization. To do so, expand Categorize Data under the Refine K-map option and click the categorization button. Once you've enabled categorization, all new documents inserted into the K-map will be placed based on the categories and placement defined during the K-map editing process.

Step 10. Generate Affinities

Affinities reflect an individual's knowledge or expertise against K-map categories and are established by managers who can define affinities for specific users, automatic discovery based on documents that a person has authored, edited, and/or read, and self-nomination.

You use the metrics processing service to calculate and update all affinities, and you store the results in people profiles to let users easily view or search for affinities via the K-map.

Before generating affinities, you must first review the default settings and modify them as necessary. To open the settings form, select Specify Settings under the Generate Affinities option and click the Modify Settings button in the top-right corner. The following settings will be available for modification:

- **Affinity generation threshold:** This is the percentage of a person's knowledge about a K-map category than you can classify as experienced and therefore generate an affinity.
- **Affinity approval policy:** Depending on your requirements, you can establish whether a user agrees to an affinity (via e-mail approval), whether to publish the affinity if you don't receive an approval within a specific time, and whether to modify the wording of approval messages.
- **e-Mail mining policy:** You can clearly define whether to allow the use of people's e-mails to search for affinities and whether doing so requires approval from the user.
- **Affinity designators:** By designating users or groups of users as affinity designators, they can rate people for affinities. This automatically takes precedence over any automatically discovered affinities or self-declared affinities.

Once you've saved any modifications to the affinity settings, you must enable the metrics processing and collection services. The metrics collection service collects usage statistics about the data within the sources (such as the number of times the document has been opened with K-map links). The metrics processing service continually calculates document and affinity values.

To enable these services, modify the server settings from the Configure Servers option and click the check box for the appropriate services. The "run between" field identifies when the metrics services run and when affinities are calculated or updated. Change this to an appropriate time so that you can test the affinities.

Once you've saved the server settings, the metrics service will automatically enable and run against the identified schedule. Once the metrics service is running, you can see the status of its processing by selecting Monitor Generation under the Generate Affinities option. To view any discovered affinities, display people profiles by selecting Verify Generation under from the Generate profiles option.

Step 11. Test the K-Map

Now that the K-map has been generated and you've edited all iterations, you're ready to test the K-map. Be sure to test the K-map before deploying it to end users to ensure that all categorization and document placement has been completed.

You should perform the test against the actual K-map, not the K-map Editor. The success of your K-map is based on whether

- users can find the information they're looking for
- all missing categories are included
- categories are meaningful and logical
- groups of documents reflect real-life understanding
- affinities are accurate
- searching produces realistic results

Once You Have a Working K-Map

Once you've deployed the K-map, the spiders will automatically detect and reflect any changes to the identified data sources (e.g., additions are categorized according to their fit in the K-map). K-map editors must periodically monitor and edit the K-map to ensure the correct categorization of all new documents, addition of new categories where appropriate, creation of subcategories where required, and maintenance of views for uncategorized documents.

Deciding whether to add new data sources will also affect the maintenance time K-map editors require. Many K-maps

evolve over time as users establish new information requirements. K-map editors should repeat the editing process as defined in step 9 after adding new sources.

A Map to Success

Implementing a successful K-map requires an investment in time, but the results are always well worth the effort, producing tangible Knowledge Management results for your organization. The bottom-up approach Lotus Discovery Server uses to generate the K-map ensures that the resulting information is a true representation of your organization's data and expertise. You'll find other useful Knowledge Management references at <http://www.lotus.com> and at the Notes.Net KDS discussion forum (<http://www.notes.net/kdsforum.nsf>).

Jason Bargent is a solution architect for IBM Software Group Services in Melbourne, Australia.

Glossary of K-Map-related Terms

Affinity: The term that Lotus Discovery Server (LDS) uses to define relationships between people and information, indicating that someone is an expert or experienced in a particular subject. A person's affinity can be automatically discovered, nominated, or self-declared. Users can view affinities from the K-map or a person's profile.

Category: A group of associated information in the taxonomy. The category name is also used to derive the affinity of a person who is related to a document.

Metadata: The data (such as the author, title, or date created) that describes a document in some way. Metadata can be included in the file properties of popular products such as Microsoft Word (under File/Properties) or metatags in HTML code.

Spider: The process that the LDS uses to read and extract useful information from the identified data sources. A spider exists for different data sources (Lotus Notes, file system, a Web spider).

Taxonomy: The classification of organizations' knowledge, indicating relationships between information by grouping relevant or similar information into categories and subcategories. The taxonomy in LDS is represented in the Knowledge Map or K-map.

Training Set: The initial collection of data sources used to create the K-map. The training set is elected from the overall data sources as being a true representation of the data.

Copyright © 2001 - Penton Technology Media