



DATA EXTRACTION FILE NAMING AND YOUR DOCUMENT REPOSITORY

File naming can be one of the most important decisions you make with respect to your scanning project. This document details different file naming conventions. How you intend to use the files going forward and the amount of time spent in the files will determine the amount of data you collect to name and track your files.

File Naming Rules of Thumb

Name your files to make it as easy as possible for your employees to locate the information they need quickly.

Ensure that the files you send your scanning vendor are all formatted the same way.

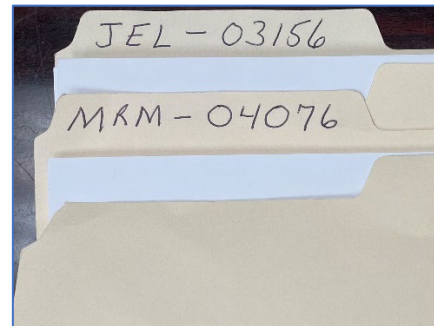
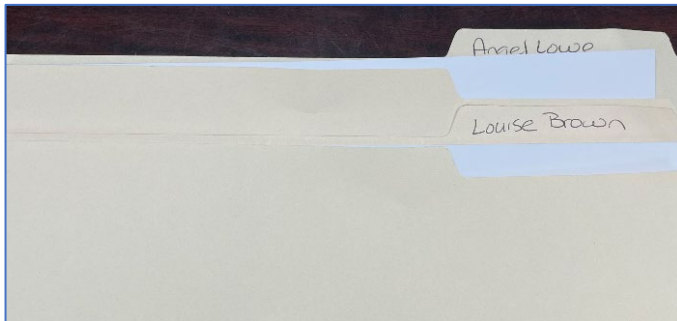
Check the formatting of the data from file to file and create a "rule" or set of "rules" about what to capture to name your files.

If you send your vendor files that have different formats then create separate jobs. One job for each type of file or document that you send to your vendor.

Folder Level Indexing

This is the least expensive and least flexible form of file naming, great for inactive documents)

In this example the information for naming a file comes off the tab of the folder itself. Everything in the folder/binder/redwell would be scanned as one PDF. Some examples are below...



Options would include setting up your files alphabetically by either last or first name. Using folder level indexing the above files would be named as follows:

Angel_Lowe.pdf
Louise_Brown.pdf
JEL_03156.pdf
MRM_04076

Placing those files on a network drive would have the documents sorting alphabetically by last name. You can search on the file using any piece of the name.

	Name	Date modified	Type	Size
AP Files				
Client Files				
HR				
MSDS				
	Brown_Angela.pdf	3/17/2023 12:22 PM	Adobe Acrobat D...	185 KB
	Harrison_George.pdf	3/13/2023 9:45 AM	Adobe Acrobat D...	398 KB
	Langley_Rachel.pdf	3/17/2023 12:23 PM	Adobe Acrobat D...	92 KB
	Makarov_Tristan.pdf	3/13/2023 9:45 AM	Adobe Acrobat D...	610 KB
	Smith_John.pdf	3/17/2023 12:19 PM	Adobe Acrobat D...	205 KB
	Stimson_Susan.pdf	3/13/2023 9:45 AM	Adobe Acrobat D...	255 KB

In the above example the underscore “_” separates the data. Make sure your vendor formats the file name this way. This structure will make it easier for you to manage file imports or exports in the future should you end up loading your files into a document management application.

Sometimes documents come to us in binders. Some companies would want you to use a “Folder Naming” methodology. They would scan all the documents from the binder into one pdf. Then name the PDF using the name on the binder.



	Name
AP Files	
HR	
MSDS	
	Acids.pdf
	Bases.pdf
	Bench Chems.pdf
	Current Flammables.pdf
	Equipment Cleaners.pdf
	Food Processing.pdf

Scanned this way you'd look for a specific document by opening up the PDF using Adobe Reader (Free app) and then searching on a chemical name (Example: e-Caprolactam). Text search would take you to pages that contain the name being searched.

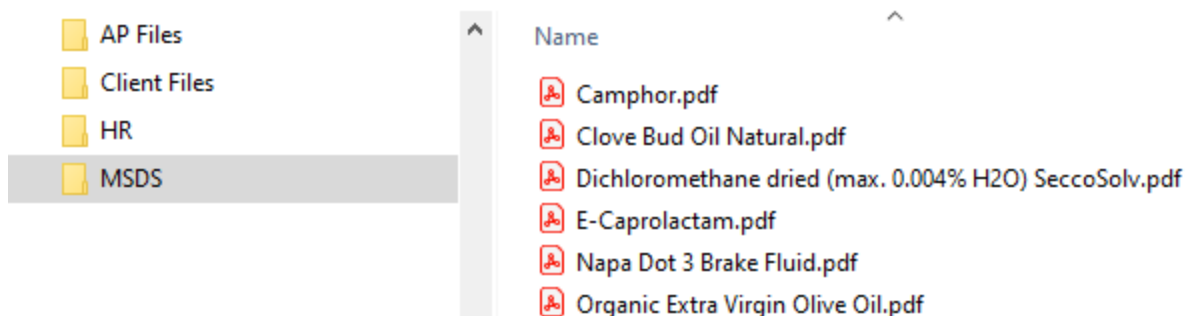
While certainly functional this type of indexing is limited with respect to managing your files moving forward. If you are interested in automating the management of your paper documents, you'll need a document management application and we would scan your documents differently. This different approach is called Document Level Indexing.

Document Level indexing

With this type of indexing we'd take the binder, scan it as 78 separate files, one for each document. From there you have some options. We would take the documents out of the sheet protectors...



Scan each one as a single pdf and, in this instance, name it using the chemical name...



This would make it easy to locate a specific file by using the "search" function built into your computer.

Document Level Indexing with Workflow Tools

Again, using the files in the binder as an example. Collecting even more data would allow you to manage your files using workflows. Again, using the above example, we would add Issue Date / Revision Date and an Update Date to the list of data being collected. That data could be part of the file name or exported into a document management application. This is quite common with human resource, accounting, medical, sales, certification, engineering and marketing files. Really, any file that is actively being managed would benefit from being managed in a document management application. Updating the document, adding information to the original document or tracking an expiration date. An example of an MSDS sheet in such an application is below...

The screenshot displays a document management interface. On the left, there is a form for adding a new note. At the top, it says "Start New Search" and "Back to Search Results". Below that are navigation icons. The form has a yellow box for "Enter New Note:" and a status "Doc ID: 1, 0 Non-Indexed Docs Remaining." Below the form are input fields for "File_Name: Camphor", "Issue_Date: 11/22/2016", "Revision_Date: 2/5/2018", and "Update_Date: 6/5/2023". There are buttons for "Update", "Delete", "Duplicate", and "Check Out", along with a "View Version History" link.

On the right, a preview of a Safety Data Sheet (SDS) for Camphor is shown. The SDS is titled "SAFETY DATA SHEET" and "FLAMMABLE STORAGE CODE RED". It includes the following information:

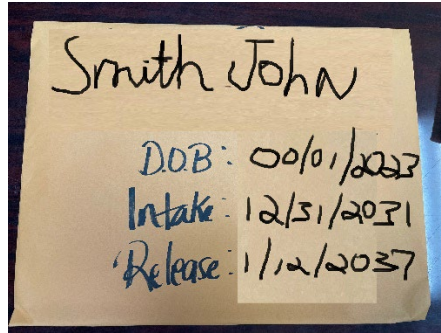
- Section 1: Chemical Product and Company Identification**
 - SDS No.: CC0160
 - Product: CAMPHOR
 - Synonyms: Borneol-2-one / Synthetic Camphor / Camphor Gum
 - Company: Ward's Science, 5100 West Hurontario Rd., Brossard, QC J4W 1K1, Canada
- Section 2: Hazards Identification**
 - Signal word: WARNING
 - Pictograms: GHS02 / GHS07 / GHS08
 - Target organs: Central nervous system
 - GHS Classification: Flammable solid (Category 2), Acute toxicity, oral (Category 4), Acute toxicity, inhalation (Category 4), STOT-SE (Category 2)
 - GHS Label information: Hazard statement: H228: Flammable solid, H302: Harmful if swallowed, H332: Harmful if inhaled, H371: May cause damage to organs.
 - Precautionary statement: P210: Keep away from heat/sparks/open flames/hot surfaces. No smoking, P240: Ground/bond container and receiving equipment, P241: Use explosion-proof electrical/ventilating/lighting equipment, P280: Do not breathe dust, P284: Wash hands thoroughly after handling, P270: Do not eat, drink or smoke when using this product, P271: Use only outdoors or in a well-ventilated area, P370+P378: In case of fire: Use dry chemical, alcohol foam, carbon dioxide or water spray to extinguish, P301+P312: IF SWALLOWED: Call a POISON CENTER or doctor if you feel unwell, P330: Rinse mouth, P304+P340: IF INHALED: Remove person to fresh air and keep comfortable for breathing, P312: Call a POISON CENTER or doctor if you feel unwell, P405: Store locked up, P501: Dispose of contents/container to a licensed chemical disposal agency in accordance with local/regional/national regulations.
- Section 3: Composition / Information on Ingredients**

Chemical Name	CAS #	%	EINECS
Camphor	76-22-2	100%	200-945-0

Collecting this type of data allows you to set up the system so that it will notify you to look for an update using the "Update Date" field. It also allows you to cross search on multiple data points.

OTHER TYPES OF STRUCTURED DATA

Sometimes we receive documents in envelopes. We can collect all or just select pieces of data from the folder. An example is below....



Actual document has been altered to protect sensitive information.



File Name: Smith_John_01-12-2037.pdf

Wide Format Indexing:

With respect to wide format drawings.

Typically, we'd grab information off the Title Box on each drawing. Then depending upon where the document is being stored (server or document management system) we'd provide you with a file name that would contain one or more or all of the index fields. If you're using a document management system that pulls up index fields then the data we supply could be different from the file name. If you're storing the document on a network server then typically the data

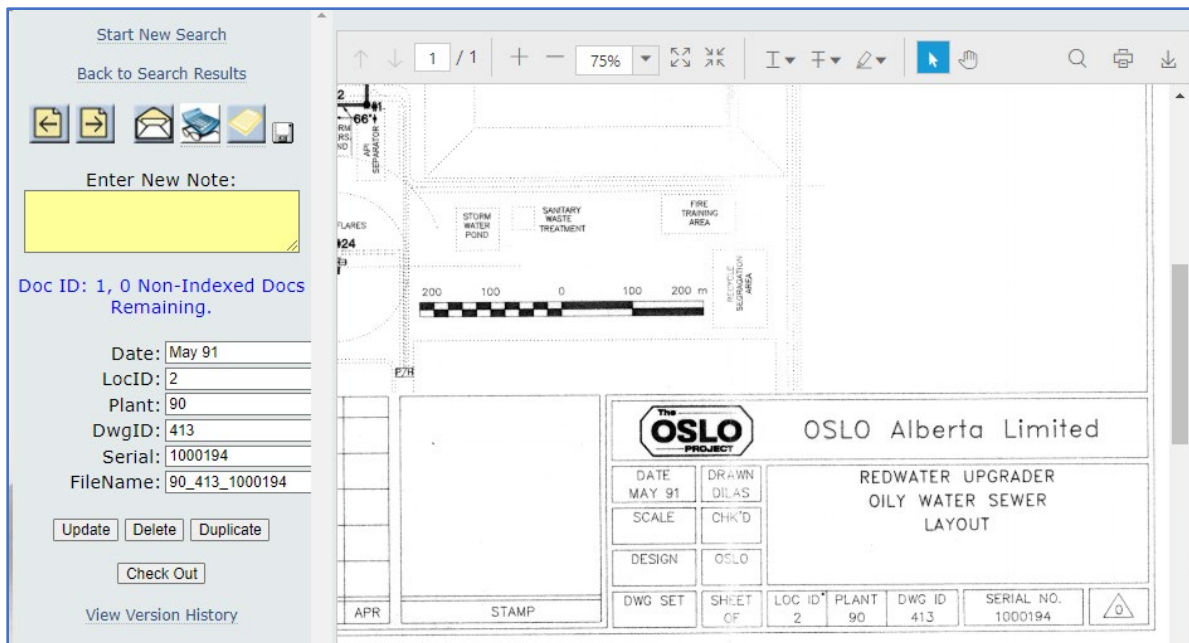
Date_LOC ID_PLANT_DWG ID_SERIAL NO.

		OSLO Alberta Limited				
DATE MAY 91	DRAWN DILAS	REDWATER UPGRADER OILY WATER SEWER LAYOUT				
SCALE	CHK'D					
DESIGN	OSLO					
DWG SET	SHEET OF	LOC ID 2	PLANT 90	DWG ID 413	SERIAL NO. 1000194	

Then we'd name the file using all the data if you're storing it on a local network server. The engineers can search for the drawing using one or more index fields.

In this instance the file name would look as follows.... MAY 91_2_90_413_1000194.PDF

Or if your drawing sits in a Document Management System we can shorten the name but provide the engineers with the data fields when they pull up the drawing. In the example below the document sits in a document management system and the engineers can cross reference all the data to pull up a file but the file name itself is a shortened version of that data. In this case just the plant number the drawing id and the serial number.



Part of creating a document scanning quote or scope of work includes defining how to name your files. It's not uncommon when scanning documents to capture sets of data to make it easy to find documents. Some of the major items to consider when deciding upon how to name your files...

- Typically documents sit in folders. Is there significant value in splitting the documents inside a folder into separate files/pdfs (common in HR / medical scanning) or would it be easier/better to scan everything in the folder as one document?
- Is the data you're using structured (the same from document to document or file folder to file folder)?
- Do you have a list of files in a database or spreadsheet that we can use as a QC tool when naming your documents? (Call or email if you have questions)
- Do you need your vendor to create a database for you?
- Where will the files end up. Are you putting them on a network drive (shared folder) or a software application (OnBase/FileNet/ImageFreeway)?

Below are some examples of the more common ways file names are created. ImageFreeway offers incredible flexibility with reasonable pricing vis a vie our competitors. As you read through this document, please keep that fact in mind. If you have a special situation, we can customize a program for you at a very affordable price.

GOAL: Name your files in a manner that's similar to the way you search for a particular document. Or use file naming to create an easier way to locate documents.