How to Automate Records Capture and Classification

Records come in many forms:
- Electronic with documents, spreadsheets, email, social media communications
- Rich media such as video, audio or photographs

Capture is what takes place to get documents that are considered important into an ERM or ECM system. Scanning or digitizing the information is what happens to capture information for our information systems. Information that was originally created in digital format can through diligent management of the information, be incorporated into the information systems.

Step 1: Determine What to Capture

Once you have decided what you want to capture, you will need to determine the format of the information you wish to capture. Some of the information captured will be important for a specific department and/or individual.

AIM Tip

While you are reviewing your documents, you should also determine how you will archive records once they are captured. Always remember that some information is valuable based on business rules. Some information may not be valuable to the organization but may have value for a specific individual. It is also important to remember that not all information needs to be captured.

For more information:
- Capture at the Point of Origination: Enabling the Agile Enterprise
- What is Information Management?
- Capture and Imaging
- Determining Which Documents to Scan
- Digital Mailroom Explained
- To Scan or Not to Scan

Step 2: Explore Ways to Automate Records Capture

There are a number of ways in which records capture and classification can be automated. You can capture information by scanning in, or otherwise digitizing, paper and other physical records and capturing the images into the system. We can better
manage records if we streamline and automate their capture. Most content does not become a record until it is signed, approved or something else by an appropriate authority. Many electronic records management solutions can be programmatically connected to Microsoft Office, Adobe Acrobat, and other desktop productivity solutions. This gives the user the option to declare the document as a record from within the client interface and save the records and associated metadata to the repository.

**AIIM Tip**

There are a number of ways to automate the capture of records. Before you implement a method, make sure the users understand how to do it. You may find that you will need multiple ways of capturing records since everyone does not work in the same manner.

**For more information**

- IDR: Is it right for you?
- How to Automate Records Capture
- How to Automate Metadata Collection and Classification
- Easing the Burden of Records Management: Automating Records Capture
- Auto-classification - Where’s the Confusion?
- Auto-classification – Will cloud vendors get there first?

**Step 3: Use Metadata**

Metadata can be used in conjunction with bulk import tools to automatically capture some types of records. In order to use the import tools and metadata, it is important that you:

- Understand metadata and its uses
- Explore ways to automate the metadata capture process

**AIIM Tip**

Before you use metadata to classify and capture your information, make sure you have documented the metadata in your policies and procedures.
For more information

- Taxonomy and Metadata Training
- The Importance of a Metadata Model in Records Management
- There’s Metadata in Those File Names
- How to use Metadata and Taxonomy to Improve Access to Information
- Metadata and User Effort
- Metadata...Now and Zen

Step 4: Analyze your Content

Before you start to capture and classify your content, you need to better understand the content. This can be accomplished by:

- Identifying the records vs non-records
- Creating classification “buckets”
- Sorting or identifying information
- Searching for relevant information
- Creating word lists or taxonomies

**AIIM Tip**

You will want to automate the capture of records, the capture of the associated metadata and the classification of your records as much as possible.

For more information

- Auto-classification - where’s the confusion?
- Big Bucket Approach - How’s the view?
- Content Analytics: Automating Processes and Extracting Knowledge
- IG Policy vs IG Reality
- Content vs. Knowledge
- How to Make Content “Intelligent”
- Content Analysis = Business Intelligence + Document Management Solutions

Step 5: Physical Record Capture
Up to this point, we have been dealing with electronic records. We should not ignore or neglect the paper or physical records in the organization. It is possible to automate the capture of physical records. There are a couple of ways of doing this:

- Convert to digital (imaging)
- Implement the use of Barcodes or RFID

**AIIM Tip**

*Scanning records to electronic versions can improve or enhance the ability to find records when they are needed and ensure that only authorized users can access them. Make sure you are scanning records to the appropriate file format that will be easiest to use in your organization’s workflows and processes.*

**For more information**

- [10 Ways to Improve Capture OCR and Indexing](#)
- [Why Should I Scan my Records](#)
- [Paper Records: Keep or Destroy After Scanning?](#)
- [Leveraging Technology You Already Have to Create True Digital Experiences](#)
- [Shedding light on the dark data in your document capture processes](#)
- [The Less Paper Businesses Use, the More Problematic It Becomes: Resolving the Paper Paradox](#)
- [How We Captured 10.3 Million Pages of Documents – and Lived to Tell the Tale](#)

**Step 6: Automate Classification**

Automatic classification tools exist and provide value. These tools use the contents of documents or records, as well as their context in the form of metadata, to make decisions about whether something is a record or not and how it should be classified.

- Start by analyzing the content and review the retention schedule
- Establish classification rules and train the systems with examples
- Crawlers and recognition engines evaluate the content and generate a classification

**AIIM Tip**
Remember to get the best results you will have to train the analytic software or engine. It is best to have a number of examples of the various types of content in your organization to use to train the system. Make sure the analytics and auto-classification tools you use meet the needs of your organization.

For more information

- ECM, One Repository, and the Key Role of Classification
- Automating Information Governance
- Valuable Content or ROT: Who Decides?
- Using Automated Classification for Summarizing and Selecting Heterogeneous Information Sources
- Know What You are Capturing – Document Classification

Step 7: Determine Measures to Evaluate Accuracy of the Classification

Accuracy of classification can be broken down into two measures:

- Precision - the precision of the tool’s output is the percentage of the classified documents that a human expert would also place into that record category. If the classification process is functioning at 100%, all of the classified documents have been placed in the same category a human would have put them.

- Recall - recall is a distinct but related metric, often framed in terms of completeness. Recall is the percentage of all the valid document that the tool is able to classify correctly to a given document category. At 100% recall, all of the documents that meet the criteria for inclusion are found in the classification result.

AIIM Tip

Before you go live with your automated classification system, you should test the system with the use of a sample set of documents. This will help you to evaluate the accuracy and allow you to tweak the settings. It is important to realize that auto-classification is not perfect.
Step 8: Ongoing Maintenance

As with any system you will implement in your organization, you will need to train your staff on how the automated classification system works. You will need to monitor the system and periodically refresh the classification rules.

**AIIM Tip**

*When you start to notice the accuracy rate of the classification begin to drop below 60%, it is time to review the system and tweak the rules.*

For more information

- [Auto-Classification: Friend or Foe of Taxonomy Management?](#)
- [Electronic Records Management Course](#)
- [Information Governance Course](#)
- [Information Governance: too important to be left to humans](#)
- [The Best Way to Manage Your Documents is Staring You in the Face – Outlook!](#)
- [Capture Anywhere-to-Process](#)