

How Do Enterprise Content Management Systems Capture Content?

Along with Content Storage, Preservation, and Delivery, Capture is one of the key components of Enterprise Content Management. This article will explore the ways content is captured in ECM systems.

Capture typically consists of acquiring raw data and then processing it in some way.

Data Capture

Data can be captured manually by ECM systems from:

Paper documents that can either be scanned for their images, or for essential details within the content of the document to be transcribed into an electronic data-entry form

Electronic office documents such as correspondence, spreadsheets, presentations, and so on created originally in an electronic form

E-mails sent or received

Multimedia objects like audio or video content, animation, and interactivity

Microfilm

Data can also be arranged to be captured automatically from EDI or XML documents, ERP applications, and other line-of-business applications like Accounting or CAD. Automated interfaces can be built with these sources.

Preliminary Processing

Scanned documents and digital faxes are not readable text. To convert them into machine-readable characters, different character recognition technologies are used. At present, these include:

Optical Character Recognition - OCR - used to convert typed document images into text documents with readable and editable characters

Handwritten Character Recognition - HCR - used to convert handwriting or lettering into text characters. The technology has not yet been perfected

Optical Mark Recognition - OMR – use to read markings in checkboxes and other pre-defined fields in forms, etc.

Standardized barcodes, allowing the extraction of information using barcode readers

Both OCR and HCR have been continually improved using artificial-intelligence features such as comparison, logic, and reference lists.

Document-imaging techniques help improve the quality of scanned images by enhancing legibility and adjusting images that have been captured in an awkward angle.

ECM can understand data captured through external forms if the capture system knows the structure and logic of the forms.

Aggregation and Indexing

Enterprise Content Management systems capture content in various formats from numerous sources. The content is then aggregated and indexed so that it can be retrieved in meaningful ways.

The indexing logic of ECM is on its own, and not dependent on any indexing logic of original sources, if the content had been indexed there.

The Enterprise Content Management system needs to develop a structure of its own that will allow accommodating the varied categories of content it accommodates.

Captured Content is Input to the Later Stages

The content captured from different sources by the Enterprise Content Management system is "managed" so that it can be processed and used, or archived.

Separate articles will identify the components of managing databases, authorizing access, and the developing the stages of storage, preservation and

delivery.

Conclusion

Content capture is the first step in Enterprise Content Management. Considering the varied nature of the content to be captured, ECM has to use varied technologies to do it. Scanning paper documents, creating interfaces to capture electronic documents from other applications, converting document images into machine-readable/editable text documents, using imaging technologies to improve the quality of captured images, etc. are examples of the technologies available.

The captured content goes to a common repository where it's indexed under meaningful categories. The content then passes into subsequent phases of management, storage, preservation, and delivery.

About the Author

Based on user experience, Ademero's flagship product, Content Central™, is a browser-based document management software, [document imaging software](#), and [enterprise content management software](#) that provides a convenient way to capture, retrieve, and manage documents and other content.

Source: <http://www.articletrader.com>