

## ECM Industry Terminology

**API (Application Program Interface)** - A set of routines, protocols, and tools for building software applications. A good API makes it easier to develop a program by providing all the building blocks.

**ASP (Application Service Provider)** - A third-party entity that manages and distributes software-based services and solutions to customers across a wide area network from a central data center. In essence, ASPs are a way for companies to outsource some or almost all aspects of their information technology needs.

**BPM (Business Process Management)** - Business process management (BPM) is a holistic management approach that promotes business effectiveness and efficiency while striving for innovation, flexibility and integration with technology. BPM attempts to continuously improve processes (a.k.a. Process Optimization) through repeated process definition, measurement, and improvement.

**BMP (bitmap)** - The BMP File Format (Windows bitmap) handles graphics files within the Microsoft Windows OS. Typically, BMP files are uncompressed, hence they are large; the advantage is their simplicity and wide acceptance in Windows programs.

**Capture** - Capture means to save or convert the contents of a document or file into electronic format. Document capture may be performed using a variety of technologies, including scanners, electronic forms, electronic transactions, cameras, voice, and video.

**Classification** - A term used in document capture that refers to how capture software identifies one document versus another. In its simplest definition, intelligent document recognition automatically identifies document types from the layout and structure of the document and knows what information it should contain. Then, if it finds that information, it can extract it for use by the business.

**COLD (Computer Output to Laser Disk) / ERM (Enterprise Reports Management)** - The storage of data on optical disks, such as CD-ROMs. Storing large volumes of data on laser disk, as opposed to microfiche or microfilm, lets the user access and search this information on a computer, avoid the duplication and security costs of protecting physical documents or film, and more readily distribute information.

**Connected Device** - Any machine or component that attaches to a computer.

**Core 2/Dual Core** - A dual core computer has two processors on the same chip. The processor is the workhorse of the computer, so this makes for a faster, more power-efficient computer, particularly when doing multiple tasks at one time (multitasking).

**CRM (Customer Relationship Management)** - CRM is a business strategy built around the concept of being customercentric. The main goal is to optimize revenue through improved customer satisfaction by redefining interactions at each customer touch point.

**Data De-duping** - De-duping is the process of removing duplicate entries from your data. This term is often used when talking about removing multiple identical files across all network storage devices.

**Digital Signature** - Digital signatures validate the contents of a document and can be used to confirm the identity of the signer. Most often electronic documents are digitally signed to prove authorship and content integrity. The purpose of a digital signature is to guarantee the documents integrity has not been altered.

**DR (Disaster Recovery)** - The plans, procedures, and contingencies that enable your company to setup, reconfigure, and continue to work after a disaster or during a work disruption, such as a power outage.

**DPI (Dots Per Inch)** - DPI indicates the resolution (quality & cleanness) of images. The more dots per inch there are, the higher the resolution and the larger the file size will be.

**ECM (Enterprise Content Management)** - ECM is a structured approach employing policies, management practices and technology to manage the lifecycle of information for an organization with a goal of improving compliance, information reuse and sharing, and operational performance.

**E-Forms (Electronic Forms)** - Any type of form that is electronic. A form template in an electronic format allows designers to assign definitions, formulas and restrictions to data in fields for efficient processing and workflow.

**Electronic Signature** - An electronic signature is a paperless way to sign a document using an electronic symbol. The purpose of an electronic signature is to act as the electronic equivalent of a hand-written signature.

**ERP (Enterprise Resource Planning)** - Any software system designed to support and automate business processes. This may include manufacturing, distribution, personnel, project management, payroll, and financials.

**Extraction** - Using OCR/ICR technology, extraction involves intelligently / electronically recognizing data on a form or document and using that information for auto-indexing. This technique is used to minimize manual data entry.

**Firewall** - A firewall is a hardware or software solution used to enforce security policies. In the physical realm, a firewall is equivalent to a lock on a perimeter door or on a door to a room inside of a building – it permits only authorized users, such as those with a key or access card, to enter.

**Forms Processing** - The ability of software to accept scanned forms and extract data from the boxes and lines to populate databases. The software usually includes the ability to drop out the boxes and lines so that data recognition accuracy improves. Intelligent Document Recognition automatically identifies document types from the layout and structure of the document.

**GIF (Graphics Interchange Format)** - GIF is limited to an 8-bit palette, or 256 colors. This makes the GIF format suitable for storing graphics with relatively few colors such as simple diagrams, shapes, logos and cartoon style images. The GIF format supports animation and is still widely used to provide image animation effects.

**Group 4 Compression (G4)** - is a method of image compression used in Group 4 fax machines, defined in the ITU-T T.6 fax standard. It is only used for monochrome (black and white) images. G4 compression is also available in the TIFF image file format, as well as in the PDF document format.

**HASP (Hardware Against Software Piracy)** - A HASP is a hardware-based software protection system that prevents unauthorized use of software applications. The device (sometimes called a dongle) connects to a computer and protects software applications from piracy by prohibiting use if the key is not present.

**HTML (HyperText Markup Language)** - HTML is the authoring language used to create documents on the World Wide Web. It is also commonly used to create E-Forms.

**ICR (Intelligent Character Recognition)** - ICR is software that translates hand-written text from paper into a form that can be electronically manipulated.

**ILM (Information Lifecycle Management)** - ILM Policy consists of the overarching storage and information policies that drive management processes. Policies are dictated by business goals and drivers. Therefore, policies generally tie into a framework of overall IT governance and management; change control processes; requirements for system availability and recovery times; and service level agreements (SLAs).

**Indexing** - The process of marking a document with keywords so you can easily retrieve the file by an electronic search. It's like creating a table of contents for your documents.

**Keyword** - In database management systems, a keyword is an index entry that identifies a specific record or document.

**Legacy System** - An application in which a company or organization has invested that is being phased out (or retained only for historical purposes) in favor of newer technology.

**Lossless compression** - algorithms reduce file size without losing image quality, though they are not compressed into as small a file as a lossy compression file. When image quality is valued above file size, lossless algorithms are typically chosen.

**Lossy compression** - algorithms take advantage of the inherent limitations of the human eye and discard invisible information. Most lossy compression algorithms allow for variable quality levels (compression) and as these levels are increased, file size is reduced. At the highest compression levels, image deterioration becomes noticeable as "compression artifacting".

**Metadata** - Commonly known as "data about data," it is the data describing context, content and structure of records and their management through time.

**.Net / ASP .NET** - ASP.NET is a programming language used to create Web pages and Web services and is an integral part of Microsoft's .NET vision. ASP.NET technology takes an object-oriented programming approach to Web page execution. Every element in an ASP.NET page is treated as an object and run on the server. Because the code is run straight from the processor, pages load much faster than classic ASP pages, where embedded VBScript or JScript had to be continuously interpreted and cached.

**OCR (Optical Character Recognition)** - OCR is software that translates standardized text images into a form that can be electronically manipulated. For example, OCR can read a flat image PDF file to make the text searchable. OCR does not typically capture handwriting because of the lack of standardization.

**OMR (Optical/Object Mark Recognition)** - OMR is software that captures objects within forms. For example, OMR software reads forms to determine if a checkbox is checked or not, or if a fill-circle is filled in or left blank.

**PDF (Portable Document Format)** - By converting an electronic document to a PDF, you can capture formatting and layout information (like taking a picture of the document) in a very compact file size and ensure that when the person opens up the document, it looks exactly how you intended.

**PDF/A** - This standard defines a format for the long-term archiving of electronic documents and is based on the PDF Reference Version 1.4 from Adobe Systems Inc. (implemented in Adobe Acrobat 5). PDF/A is a subset of PDF, leaving out PDF features not suited to long-term archiving.

**Portal** - A portal is a website or service that offers a broad array of resources, such as email, forums, search engines, and online shopping malls. It is designed to function like a hub or a one stop shop where you can find everything you need.

**PPM/IPM (Page Per Minute/Images Per Minute)** - A measurement of printer or scanner speed.

**RAID (Redundant Array of Independent Disks)** - A technology that employs the simultaneous use of two or more hard disk drives to achieve greater levels of performance, reliability, and/or larger data volume sizes.

**SCSI (Small Computer System Interface)** - A small hardware connection device that provides for faster data transmission rates than standard serial and parallel ports.

**SATA (Serial Advanced Technology Attachment)** - An evolution of the Parallel ATA physical storage interface. SATA is a serial link. A single cable with a minimum of four wires creates a point-to-point connection between devices. SATA is primarily designed to transfer data between a computer motherboard and mass storage devices such as hard disk drives.

**SOA (Service-Oriented Architecture)** - SOA is a method for systems development and integration where functionality is grouped around business processes and packaged as interoperable services. SOA also describes IT infrastructure which allows different applications to exchange data with one another as they participate in business processes.

**SQL (Structured Query Language)** - SQL, pronounced either "SEE-quel" or as separate letters, is a standardized language used for requesting information from a database.

**Structured vs Unstructured Document** - These terms relate to the organization of data on documents. Information on structured documents is predictable. Unstructured documents lack syntax and order. The advantage of a structured document, such as a form, is that it allows automation of tasks involving the information.

**Taxonomy** - Taxonomy has been adopted as a term to describe the classification schemes using hierarchical groups and controlled vocabularies employed within Enterprise Content Management systems to make documents and data easier to identify, study, or locate.

**TIFF (Tagged Image File Format)** - TIFFs are one of the most widely supported file formats for storing bit-mapped images on PCs.

**USB (Universal Serial Bus)** - A connection port on a computer that is universally compatible with many types of devices. Most scanners recommend USB 2.0 as the standard connection.

**Validation** - Verification that something is correct or conforms to a certain standard. In data collection or data entry, it is the process of ensuring that the data entered fall within the accepted boundaries of the application collecting the data. Validation procedures are typically written into the program code and are invisible to the user.

**Vertical Market** - A term used to describe industries. For example, Insurance Vertical, Manufacturing Vertical, etcetera.

**WEBP** - WebP is a new image format that uses lossy compression. It was designed by Google to reduce image file size to speed up web page loading: its principal purpose is to supersede JPEG as the primary format for photographs on the web.

**Web Parts** - Web Parts are customizable components for corporate portals that plug into existing back-end systems and allow employees to access internal data, external feeds, and collaboration tools.

**Workflow** - The defined series of tasks within an organization to produce a final outcome. For example, in a publishing setting, a document might be automatically routed from writer, to editor, to proofreader, to production. At each stage in the workflow, one individual or group is responsible for a specific task. Once the task is complete, the workflow software ensures that the individuals responsible for the next task are notified and receive the data they need to execute their stage of the process.

**XML (Extensible Markup Language)** - XML is a programming language for the World Wide Web that allows designers to create their own customized tags, enabling the definition, transmission, validation, and interpretation of data between applications and between organizations.