TECHNICAL REPORT

ISO/TR 22957

Second edition
2018-12

Document management — Analysis, selection and implementation of of enterprise content management (ECM) systems

Reference number
ISO/TR 22957:2018(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 171, *Document management applications*, Subcommittee SC 2, *Document file formats, EDMS systems and authenticity of information*.

This second edition cancels and replaces the first edition (ISO/TR 22957:2009), which has been technically revised. The main changes compared with the previous edition are as follows:

— updates have been made to bring the document in line with current generation technologies;

— the references have been revised and updated throughout;

— terms and definitions have been added and the acronyms section has been removed;

— "electronic document management system (EDMS)" has been changed to "enterprise content management (ECM)" throughout;

— the wording has been improved throughout and the contents have been reorganized to provide clarifications.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

This document provides detailed information associated with the analysis, selection and implementation procedures associated with enterprise content management (ECM) systems. The development of this document is a result of organizational requests to receive vendor-neutral industry information associated with technology standards, technical reports and industry best practices for ECM projects.

Terms and acronyms associated with various aspects of ECM technologies commonly change over time, as technology developers and vendors update product lines and solutions to address customer requirements. In most cases, the new terms and acronyms reflect updates and changes to how these technologies are used, typically by incorporating additional levels of functionality and very rarely resulting in an entirely new core technology. This is important to note as the core ECM technologies are constantly maturing, and solution providers are identifying not only new approaches to addressing organizational issues and requirements, but also expanding the use of these technologies into areas previously unconsidered. As such, organizations are constantly challenged to keep pace with how an updated technology is currently being referenced, especially when the same core technology is referenced differently between vendors and, at times, various groups of suppliers.

For the purposes of this document, the terms “document management” and “content management” can be considered to be synonymous. As the ECM industry (previously referred to as the “document management” or “electronic content management” industry) has matured over the years the ability to store electronic information has greatly expanded from hard copy document scanning into digital images in the early 1980s to the management of any digital or electronic document that today is referred to as “electronically stored information (ESI)”.

Clause 4 provides detailed information describing each of these technologies, and how they operate and inter-operate.

Clause 5 provides detailed information associated with currently available industry standards and technical reports.

Clause 7 provides detailed information related to industry best practices associated with all the customary project phases for ECM technology analysis, selection and implementation. These project activities are considered to be industry best practices. It has been demonstrated over the past 10 years that organizations following all the recommended steps and activities have a much greater level of project success while greatly decreasing, and in most cases eliminating, unnecessary technologies, user licences, etc. This is very important, especially with most organizations carefully examining all expenditures related to all aspects of technology procurements.

This document provides detailed guidance to organizations considering the use of any of those technologies that comprise ECM [document imaging, document/library services, routing/business process management (BPM)/workflow, records management applications (RMAs), forms management, enterprise report management (ERM), etc.]. It should be noted and acknowledged that a complete records management programme set up against ISO 15489-1 is critical to any organization and is integral to any complete and thorough management plan associated with electronic information regardless of whether it is referred to as a “document”, “record”, “audio”, “video”, etc., internally by the organization.

Document management — Analysis, selection and implementation of of enterprise content management (ECM) systems

1 Scope

This document gives guidelines for a set of procedures and activities to be considered and/or performed by organizations when planning, designing and implementing various enterprise content management (ECM) technologies. The aspects or project phases range from initial business analysis through to vendor/integrator selection and technology implementation. The implementation of processes to manage electronically stored information (ESI) requires significant participation from the affected business units, if the content is stored and managed when created/received and controlled through the information life cycle following organizational policies and/or records retention and control policies are applied. As these efforts require multiple people with different disciplines, including technical teams, records managers and organizational management, this document has been prepared taking those perspectives into account.

This document is applicable to both in-house and outsourced systems, including cloud solutions. It can also be useful when dealing with specialized business systems. The term "enterprise content management (ECM)" (or "document management") used throughout this document is intended as an all-encompassing term referring to capture technologies [scanning, indexing, optical character recognition (OCR), forms, digital creation, etc.], management technologies (document services, workflow and other work management tools), and storage [primarily non-alterable or write once read many (WORM) technologies]. This document provides information to users related to the technical reports, guidelines and standards that have been developed for technologies commonly available in ECM systems.

This document is not intended to be an all-inclusive paper on electronic document or content management and does not attempt to influence any single technology or provide legal guidance or legal opinions. While there are storage technologies other than optical/magnetic currently available (i.e. microfilm, microfiche and hybrid storage systems) that are not included in this document, those technologies can be reviewed if determined to be appropriate by the end-user organization.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 12651-1, *Electronic document management — Vocabulary — Part 1: Electronic document imaging*

ISO 12651-2, *Electronic document management — Vocabulary — Part 2: Workflow management*

ISO 15489-1, *Information and documentation — Records management — Part 1: Concepts and principles*

ISO/TR 15801, *Document management — Electronically stored information — Recommendations for trustworthiness and reliability*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 12651-1, ISO 12651-2, ISO 15489-1 and ISO/TR 15801 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
library services
administrative components of the ECM system that handle access to information

3.2
expungement
process of removing a document from a system and leaving no evidence of the document ever having appeared on the system

3.3
semi-structured document
document that contains both some level of structure and organization along with completely unformatted text or text without any structure

EXAMPLE Letters, emails, forms with free-form text components, forms.

3.4
structured document
document that follows a strict structure or format

EXAMPLE Table, database record.

3.5
unstructured document
document that has no pre-defined structure or format and contains free-form text, which may or may not be similar to other documents of the same type, and varies greatly in structure, content, terminology and format

3.6
intelligent document recognition
IDR
technology that incorporates various methods of capturing and extracting information (or data) used to identify a specific type of document and data extraction with minimal (or no) user intervention required

Note 1 to entry: The most basic or traditional method of IDR has been in use since the late 1990s and incorporates the processing of barcodes, patch codes and other manual indexing methodologies. The current generations of this technology can also incorporate more advanced technology, techniques and algorithms to self-teach, i.e. to update the processing rules to classify or to extract data without user intervention, while others require linguistic algorithms, referred to natural language processing (NLP), to process unstructured content. IDR is a common marketing term used to describe the process of capturing content (scanned documents and digital born) and extracting content from the document within the limits of the technology being used.

3.7
natural language processing
NLP
technology used to determine and identify key words and phrases within processing audio data (e.g. call centres) and free-form text (e.g. the body of an email)

Note 1 to entry: This technology is able to reduce words to their base constructs and perform other actions, such as stemming, along with locating similar words or phrases without user intervention. This technology also varies greatly from standard IDR technology due to the ability to automatically update rules as determined by the users without the need for technical intervention. This technology is best suited for unstructured documents.

4 ECM technology

4.1 General

Many organizations still function in a hybrid environment of both electronic and hard copy environments. This environment is a direct result of the need of the organization to maintain information on all aspects of their business activities, as a business asset, to enable business efficiency and accountability. These requirements for organizational record keeping are often driven by government regulations and legislation to demonstrate accountability and implement quality control procedures. Organizations are also dealing with the exponential growth of digitally created content, ever increasing volume, diverging variety of formats, the acceleration of the velocity that content is arriving or being created and the growing doubt about the veracity of this content. These pressures lead to the disorganization of information as organizations try to keep up with the scale of change.

It is therefore a very important consideration for an organization evaluating or considering ECM technologies to first design, plan and implement the necessary foundational components, supporting the informational and organizational needs, and then expand and/or add additional functionality as required by the business units. This approach to phasing enables organizations to fully adopt new technologies while minimizing typically encountered issues adversely impacting day-to-day operations.

Organizations should also recognize that there is a wide range of stakeholders to be consulted, including senior managers, end users from business units, records managers and archivists, legal counsel and information technology (IT) staff.

4.2 Functional view of ECM systems

To help frame the concepts within this document, common terms used to reference these technologies should be discussed.

"ECM systems" has become an all-encompassing term referring to the integration of various underlying technologies, including:

— document imaging (used to convert hard copy documents into digital format);

— document/library services (used to manage digitally born documents);

NOTE Most ECM systems allow users to use this technology to also manage scanned documents, if desired.

— BPM/workflow (used to automate work processes, including the creation, routing, tracking and management of information being processed);

— ERM (used to store electronic formatted reports);

— forms processing (used to incorporate interactive forms and manage related forms data);

— OCR/intelligent character recognition (ICR) technologies;

— enterprise or faceted search;

— records management modules;

NOTE In some instances, this will be a module and in others an ECM will act as an integrated "front end" for an application that can be separate, such as RMAs, case management applications, legal discovery tools or web content management (WCM).

— metadata management;

— various applications (also considered add-ons), such as legacy system integration tools.

These systems provide users with greater access to ESI from common user interfaces, typically utilizing industry standard internet browser technology. One of the primary reasons users prefer web-based

client tools is the distributed functionality and ability to maintain standard desktop configurations for other office and business-related applications.

Figure 1 is a functional view of how ECM technologies, including records management, workflow, forms processing, etc., typically integrate within an organization. This view shows multiple applications integrated with and operating in conjunction with the ECM solution resulting in an overall trustworthy and reliable ECM environment. As Figure 1 shows, while there are many solutions available, it is important to remember that the policy layer is equally as important as the presentation and content layers.

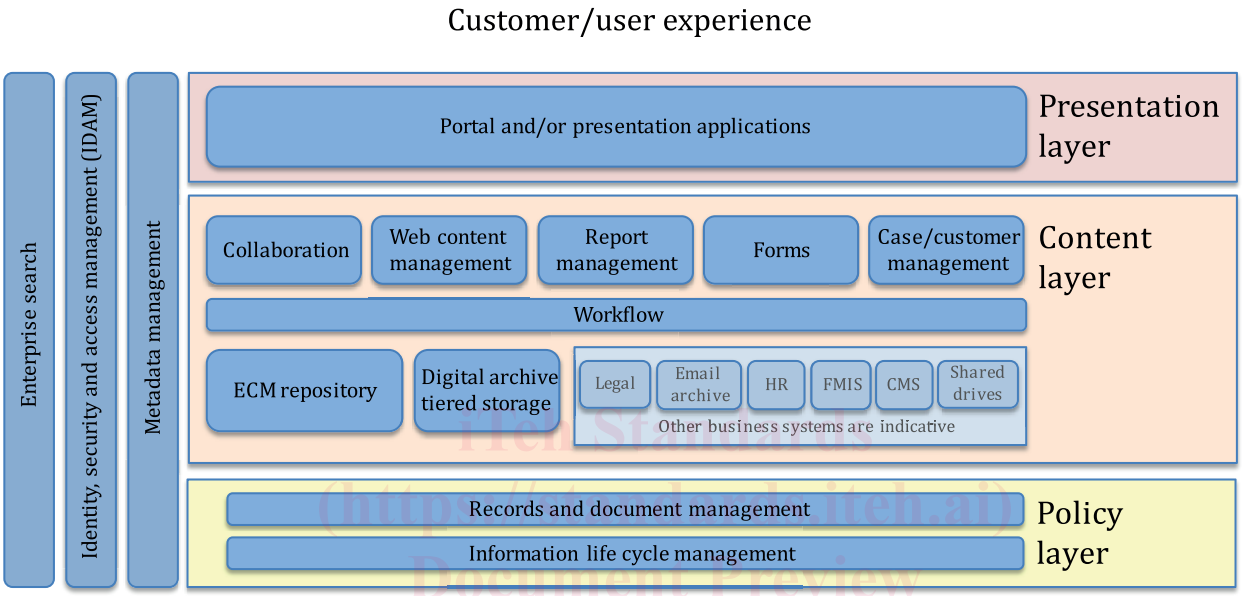

Figure 1 — ECM functional view

4.3 Technical view of ECM systems

4.3.1 Overview

From a technical perspective, the structure of ECM technologies can be viewed as a set of building blocks, as shown in Figure 2.

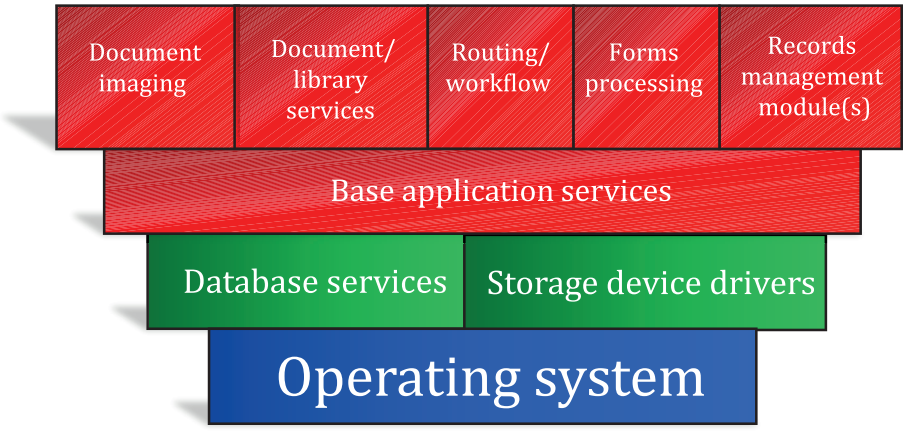

Figure 2 — ECM technology building blocks

The lowest level is the operating system. Database services and storage device drivers are installed onto the server as the second layer. The third layer includes the base ECM application components and services provided by the service provider. This layer typically includes the solution configuration tools, application programming interfaces (APIs), and application components integrating the core applications components with the database services along with providing the components that integrate the storage environment with the overall solution. The fourth layer incorporates the various core technologies of ECM technologies. Each of these core technologies (except for the operating system layer) are further described in the following sections.

4.3.2 Database services

There has been a significant shift from developing custom technology solutions at the database level to configuring/implementing commercially available software over the past 10 years. As the ECM industry and associated technologies matured, end-user organizations could shift from a "development" model to a "configuration" model for the base technological components. In fact, the selection of the database to be used is typically at the discretion of the organization but has become standardized using open database connectivity (ODBC) tools, which has resulted in the database components to be almost considered a "commodity item" rather than a specialized tool.

This is an important consideration for any organization evaluating ECM technologies from the perspective that almost all of today's ECM solutions have moved away from needing to have specialized database administrators to discouraging organizations from changing and/or modifying the ECM database table structures and configuration, which in many cases now results in the solution provider withdrawing solution support. While years ago it was important for the organization to hire dedicated ECM database administrators, this is no longer the case.

Since the very early 2000s, the ECM industry has effectively standardized the most commonly used database platforms enabling the solution providers to offer standardized support, thereby significantly reducing the organizational cost associated with hiring database specialists solely or primarily dedicated to managing the ECM solutions.

4.3.3 Storage device drivers

The storage device drivers or services are at the same level as the database services. The storage device drivers are used to connect the selected storage technology to the system and make the storage space available for the selected ECM solution.

4.3.4 ECM application services

The next layer in the "building block" is the ECM server application. Early in the development and maturity cycle of ECM technologies, end-user organizations were required to provide database administration and resources. During the late 1980s and early 1990s, ECM technologies had not matured to a level approaching commercial off-the-shelf (COTS) availability. While ECM technologies were maturing, end-user organizations were required to maintain the database along with the application.

Current versions of enterprise ECM solutions have shifted the database administrative functions back to the vendor/integrator with the end-user organization responsible for daily application maintenance and periodic server maintenance. Most enterprise ECM solution providers provide their products with technical support, including system installation, initial configuration, application updates/patches, etc. This major shift from requiring significant technical resources at the end-user level to vendor-supported solutions has resulted in an extensive amount of vertical market penetration. This has been achieved using standard technology components configured to address specific environments and business needs.

The application services are provided by the selected solution provider and include all the necessary components supporting the platform of modules being implemented. These services include the various application interfaces to other applications, along with providing the integration points between the operating system security and management features and the management of the data being stored in

the ECM solution. The application services also provide all the core and underlying functionality for all the solution components.

4.4 Core technologies and application-specific modules

4.4.1 Overview

There are various core technologies and application-specific modules that provide specific functionality, including some or all the following core technologies: document imaging, document/library services, workflow, forms processing, records management modules, etc. All enterprise ECM solutions have at least one of these core technologies and, in most cases, include multiple components. Furthermore, many solutions have integrated both the document imaging and document/library services components into a single application, while other solutions simply integrate these components as required by the organization.

A very important aspect of ECM technologies is associated with trustworthiness and reliability. Any ECM solution should ensure that any information stored and managed will be retrieved exactly as stored without any unauthorized access, modification or deletion. Not implementing a solution that can protect the managed information is rarely of much use to an organization and becomes simply another data silo. Trustworthy ECM solutions incorporate not only software and hardware, but require organizational procedures and acceptance.

There are two very distinct approaches seen in the industry related to integrating and implementing various ECM technologies. There are applications that are what is referred to as "image enabled", meaning that the legacy application has a process to save scanned documents and records along with digitally born content (e.g. office documents, emails, facsimiles). Typically, the electronic (or digital) information is then stored on a network device or directly within the application/database (only for much older applications, typically). It is important to note that this is not considered to be an ECM solution.

On the other hand, there are ECM solutions that go beyond basic "image enabling" and provide full and extensive levels of control over electronically stored documents or records. These solutions always incorporate the following functionality, with varying degrees of sophistication and capabilities that allow for:

— capture of enterprise content (via document imaging or importing, if already in electronic format);

— process automation (workflows, forms processing, process automation, application integration);

— content management, storage and retention controls (trustworthiness, retention schedules, security controls, access controls).

The following sections provide detailed information on each aspect of these technologies and how they function and/or inter-operate within an ECM environment.

4.4.2 Document/library services technologies

Document/library services technologies enable organizations to manage digitally born documents and have become the core component of almost every ECM solution available in today's marketplace. A key component of document/library services is the ability to generate, log and track all aspects related to users accessing content, creating content and all associated actions taken within the system. Having this history generation and logging capability of all user actions when searching, accessing and/or creating content is the cornerstone of any trusted ECM solution.

These technologies control the authoring, check-in/out and/or version control of documents being developed, managed or stored. This enables collaborative development when desired, along with a mechanism to store/manage digitally born document libraries. The basic capabilities of these technologies include allowing authorized users to:

— load or import digitally born documents directly into the system;

— enter relevant metadata associated with the document, according to a metadata's schema previously defined in the organization;

— create virtual folders linking various documents, according to a classification schema previously defined;

— check information in/out of the repository;

— make changes and check the modified information back into the repository;

— manage whether original documents are updated or replaced during the update operations;

— establish security levels for groupings of documents.

The management aspect of document/library services technologies includes the ability to restrict access to certain documents or groups of documents to only authorized users. Along with security controls, these technologies enable users to be granted different levels of access. For example, the author of a document might only grant "read" access to all users outside of a specific organization while granting "check-in/out" control to others who are working on updating the document. As the other users prepare to update the document, they would "check" the document out of the library, update the information and then "check" the document back in.

Document/library services technologies ensure that any other user attempting to check the document out would, first, not be allowed to check it out, and, second, would be notified that someone already has a copy being updated. Upon completion of the update cycle, the system automatically updates the version number of the document and makes it available to all authorized users.

Along with the abovementioned authoring controls, document/library services are commonly configured to integrate with records management technologies, establishing retention controls and enforcing the records retention policies in an automated and regulated fashion, rather than allowing users to select relevant record series during indexing. The linkage between the retention schedule and the enterprise content is referred to as the "retention bridge", where information from the retention schedule is associated as appropriate with the classification schema.

NOTE The linkage between the retention schedule and the enterprise content is usually established by assigning retention rules to the elements of the classification schema (file plan).

Additional information on the integration of records management technologies is discussed below.

4.4.3 Document imaging technologies

Before discussing document imaging in general, it is important to note that terms used throughout the industry are constantly changing as the ability to automate information classification and data extraction continue to mature. Since the mid-1990s, the industry began exploring ways to streamline the imaging process and the term "intelligent document recognition (IDR)" became a common marketing term within the content/document management industry. Initially, the critical need was to incorporate software with the ability to deskew, despeckle, rotate and ignore certain type of pages based on colour and then these technologies matured/expanded to include barcode, patch code and other more advanced recognition capabilities. Throughout this timeframe, various solutions used varying terms to describe their solution and those marketing terms continue in various forms today, but the underlying process to reduce user intervention remained the same goal.

From a basic technology perspective, document imaging technologies enable users to scan hard copy documents into a digital or electronic format. The components of document imaging consist of a hardware device used to scan/convert the document into electronic format and the software used to control the scanning device and (in some cases) allow users to adjust the image quality to improve legibility (many of these functions are described above). Other solutions and/or applications simply integrated document imaging functionality into their core applications. It is important to note that this is not the same as integrating document imaging to an ECM solution.

When integrated to an ECM solution, the technologies enable users to index or enter metadata directly into the system, which then stores and manages the information, including historical activities associated with the document itself. Much of this information comes from an IDR process to some degree, which is described in more detail below. This metadata allows users to identify (or index) this digital information allowing them to be retrieved later and all types of information required by the end-user organization to fully track all necessary metadata. Some ECM solutions support more efficient indexing through the support of a controlled vocabulary that regulate synonymies and polysemies.

4.4.4 Intelligent document recognition

During the late 1980s and 1990s, the entire capture portion of the ECM industry was based on converting hard copy documents into electronic (or digital) format for further processing. To decrease the need for fully manual indexing, the industry began integrating the ability to process "separator pages" between different documents, and printing barcodes (also considered patch codes) to simplify separating the end of one document and the beginning of another, while being able to extract some basic information from either the cover page or the barcode itself, if located on the pages. During that timeframe, users (or "preppers") would insert a coloured page or page with a barcode in between each document, allowing the software to recognize the beginning and ending of pages and documents, thus beginning further automation of the categorization of documents being ingested.

Processing or extracting content from documents with some level of intelligence has been a desired goal within some aspects of the capture portion of the ECM industry, with the goal of reducing otherwise extensive user interaction to get the desired content into the library for later searching or delivery to a line of business application for processing. This process is commonly referred to as "intelligent document recognition (IDR)", which began as a marketing term to broadly describe data extraction from the technical perspective, relying on the appropriate underlying technology required, e.g. OCR for machine generated text, ICR for handwritten text, optical mark recognition (OMR) for marks, and natural language processing (NLP) or self-learning (i.e. when the solution updates rules internally based on user responses) for free-form text and audio, such as calling customer service. It is important to remember that each of these underlying technologies is entirely different and requires integration efforts from highly trained resources.

These technologies cannot be simply installed and turned on, as significant design and configuration work is necessary for the more advanced technologies, such as NLP and/or AI self-learning solutions, in order to achieve any reasonable level of accuracy. It is also important to note that different technology solutions incorporate different combinations of these underlying technologies to process structured or semi-structured documents (where the desired data can be identified in a consistent fashion). The use of these technologies is far more advanced than the traditional data extraction method used for structured documents or those that can be processed via manual data entry and/or using zonal OCR, barcodes, etc., which may be sufficient to meet the organizational needs. None of these traditional methods of extraction produce accurate results when processing free-form text, which requires the integration of linguistic and NLP (e.g. complaint forms, free-form contracts that contain some similar topics but in random locations).

Following the advancement in technology related to NLP, traditional methods were replaced with software based on linguistic technologies, primarily voice recognition and response, rather than extracting data from forms and documents. Linguistic and NLP technologies are also excellent tools to process free-form text (e.g. body of an email, a general complaint letter). The capture portion of the industry has continued to expand functionality to software tools that are readily available to all organizations, enabling users to automate as much as required while reducing user intervention, while other organizations attempt to fully automate the data extraction. The same technologies would not be a good fit for both examples as not all organizations have the need for AI "self-learning" or NLP. Traditional information capture with basic index automation is typically sufficient for many organizations. The latter technology still produces errors that must be managed by an operator, but, in time, the level of accuracy will increase and most capture software vendors will begin providing this functionality as part of their core products, just as almost all capture software vendors support

