

ECM - Capture

Das Wichtigste zuerst

Dokumentenmanagement- oder Enterprise-Content-Management-Systeme (folgend: ECM) ganz gleich welcher Couleur ziehen ihren Nutzen immer aus den bereitgestellten Informationen. Aber bevor diese genutzt werden können, müssen sie den entsprechenden Systemen auch zugeführt werden. Deswegen ist die erste und wichtigste Stufe eines ECM-Systems diese Funktionalität. Man verwendet hierbei den Begriff „Capture“. Hierunter versteht man alle im Zusammenhang mit der Erfassung, Erkennung und Klassifizierung von Dokumenten oder besser Objekten durchzuführenden Tätigkeiten.



Von [Wolfgang Koch](#)

„Capture“ hat sich zwischenzeitlich zu einem eigenständigen Bereich entwickelt, der sowohl ECM-Systeme und Archive als auch operative Systeme (z.B. ERP-Anwendungen) mit

Daten beliefert. Die Erfassung der im ECM zu verwaltenden Daten ist ein wichtiger Bestandteil beim Dokumenten-Management.



Planung ist wichtig

Da hierbei zu einem wesentlichen Teil die späteren Nutzungsmöglichkeiten bestimmt werden, sollte man der Erfassung sowohl der Planung als auch später der Durchführung und der Kontrolle entsprechende Aufmerksamkeit widmen.

Die unterschiedlichen Dokumentenquellen - Papier, COLD, E-Mail, Office-Dokumente usw. - erfordern in der Regel auch unterschiedliche Erfassungs-, Attributierungs-/ Klassi-

fizierungs- und Prüf-schritte. Teilweise kommen bei der Erfassung zusätzliche rechtliche Anforderungen hinzu - etwa bei eingehenden elektronischen Rechnungen oder Belegen im Sozialversicherungsbereich.

Großes Spektrum

Das Spektrum der Erfassungsverfahren ist entsprechend groß. Die verschiedenen Verfahren verlangen teilweise nach recht unterschiedlichen Techniken. In vielen Unternehmen kommen dabei mehrere Erfassungsverfahren parallel zum Einsatz, um dem Anspruch des Enterprise-Dokument-Management-Anspruchs zu genügen und alle für das Unternehmen relevanten Dokumente möglichst integriert und weitgehend vollständig elektronisch zu erfassen und zu verwalten.

Der Flaschenhals der digitalen Informationsverarbeitung ist vor allem die schnelle Erfassung der Informationen. Sie gilt im besonderen Maße für existierendes Schriftgut, das mittels Scannertechnologie in ein elektronisches Informationssystem überführt werden soll. Dazu zählen Posteingang, sonstige Papierdokumente, eingehende Vordrucke etc.

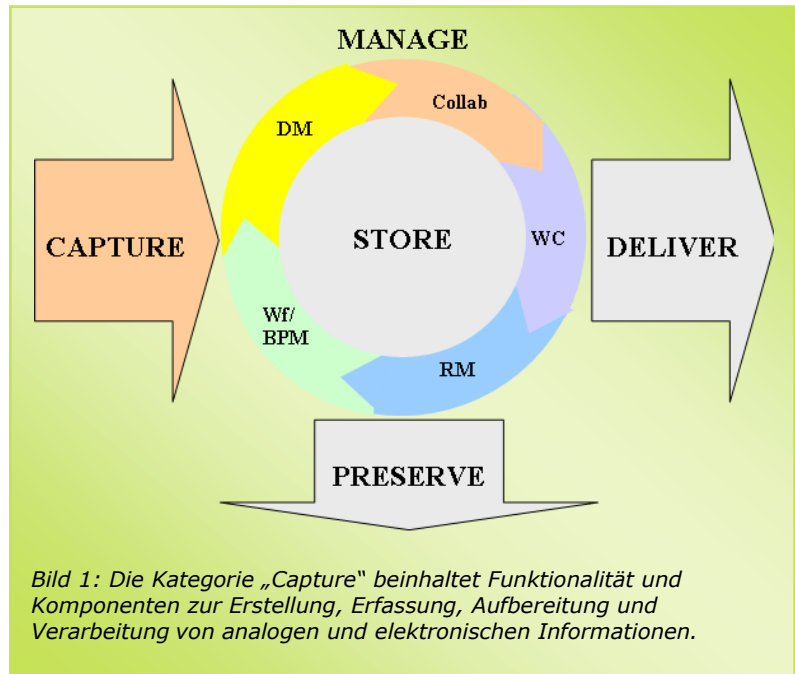


Bild 1: Die Kategorie „Capture“ beinhaltet Funktionalität und Komponenten zur Erstellung, Erfassung, Aufbereitung und Verarbeitung von analogen und elektronischen Informationen.

Quelle: Project Consult GmbH

Ein weiteres Problem liegt darin, diese NCI-Dokumente (NCI - Non Coded Information) mit Zugriffsinformationen zu versehen. Dies kann manuell beim Scannen, durch automatisches Erkennen von Text oder Barcode und durch Ergänzung fehlender Informationen aus bestehenden DV-Systemen geschehen.

In diesem Rahmen ist auch „Schriftgut“ zu sehen, das gar nicht mehr in Papierform erzeugt wird. Elektronische Dokumente erlauben eine einfache automatische Indizierung und gewinnen bei der Speicherung von eigen erstellten Dokumenten oder elektronischen, eingehenden Informationsobjekten immer mehr an Bedeutung.

Scannen

Die Entscheidung, welche Form des Scannens gewählt wird, ist abhängig von den Anforderungen an die Qualität und den Prozess der Erfassung.

Die Einrichtung einer eigenen zentralen Scanstelle kann dann in Betracht kommen, wenn relativ große Mengen an Dokumenten eingehen und diese taggleich elektronisch am betreffenden Sachbearbeiterplatz zur Verfügung gestellt werden sollen.

Scannen direkt am Arbeitsplatz oder abteilungsweises Scannen wird häufig in Anwendungsfällen mit kleinerem Volumen oder verteilt anfallendem Schriftguteingang eingesetzt. Als Sonderfall ist in diese Kategorie das dezentrale Scannen an entfernten Arbeitsplätzen und anschließende Übertragen der Daten per Leitung (z.B. von Baustellen) zuzuordnen, wobei diese Variante zusätzlich die vorgenannten noch ergänzen kann.

In vielen Fällen kann bei großen Belegmengen oder spezialisierten Aufgabenstellungen (z.B. bei der Indi-

zierung und Klassifizierung) die exklusive oder zusätzliche Inanspruchnahme von Scan-Dienstleistungen (Outsourcing) bei einem entsprechenden Dienstleistungsunternehmen stattfinden.

Indizieren

In den meisten Fällen nimmt die Indizierung beim Scan-Prozess nur eine erste Klassifizierung vor, die über die weitere Verarbeitung entscheidet. Während der Bearbeitung wird die endgültige Indexzuordnung ergänzt. Die Dokumente werden nach dem Scannen vernichtet oder zusätzlich im Original archiviert.

Das manuelle Indizieren kann auf Grund von Eingabefehlern des Erfassungspersonals ohne entsprechende Maßnahmen und technische Hilfsmittel zu einer mangelhaften Archivierung führen. Deswegen sollte immer dann wenn es möglich ist, ein entsprechendes Verfahren oder eine Kombination davon zum Einsatz kommen.



Eine wichtige Anforderung innerhalb des Scan-Vorgangs ist die revisions-sichere Protokollierung. Die revisions-sichere Protokollierung führt in Verbindung mit der Verfahrensdokumentation zu einem sogenannten „elektronischen Dokument hoher Qualität“. Protokollsätze sollen die Angaben von Benutzer, Signaturcode, Datum/Uhrzeit, Unique Identifier des Informationsobjekts, etc. enthalten.

Erkennungstechnologien

Automatisierte Erkennungstechnologien sollen die Sachbearbeiter vor zu vielen manuellen Eingaben bewahren. Dies kann manuell beim Scannen, durch automatisches Erkennen von Text oder Barcode und durch Hinzufügen fehlender Informationen aus bestehenden DV-Systemen geschehen. Für die automatische Extraktion von Zugriffsinformationen, der sogenannten Indexinformation, sind Techniken wie OCR (Optical Character Recognition) u.ä. erforderlich. Diese können nur unter bestimmten qualitativen Voraussetzungen der Dokumentenvorlagen sicher gewonnen werden.

Zur Verarbeitung von gescannten Belegen werden verschiedene Erkennungstechniken (Recognition - Mustererkennung) eingesetzt. Zu ihnen gehören:

- Barcode
- OCR (Optical Character Recognition):
- HCR (Handprint Character Recognition)
- ICR (Intelligent Character Recognition):
- OMR (Optical Mark Recognition):

Bei der Freiformerkennung müssen die gesuchten Attribute nicht mehr an einer festen Stelle stehen, sondern können irgendwo aufgedruckt sein. Über ein flexibles Layout und ein Dokumentenmodell wird definiert, welche Daten einem bestimmten Dokumententyp zugeordnet sind.

Man erkennt einen Dokumententyp anhand bestimmter Vorgaben und Konventionen. Typische Suchelemente sind Text, Trennlinien, weiße Lücken, Barcodes, Zeichenketten, Textfragmente, Objektsammlungen, Datum, Telefonnummern, Währun-



gen und sogar komplexe Tabellen.

Im Gegensatz zur Erkennung einzelner Elemente, kann durch eine Analyse der Elemente im Kontext die Genauigkeit gesteigert werden. Typische Anwendungen sind die automatische Rechnungseingangserfassung sowie die Klassifikation und Auswertung der Eingangspost.

Pattern Recognition

Im Gegensatz zu dem regelbasierten Ansatz sind in den vergangenen Jahren lernende Mustererkennungsverfahren entwickelt worden. Bei den Mustererkennungsverfahren werden der Anwendung viele Musterdokumente vorgeführt, aus denen die Anwendung bestimmte regelmäßige Strukturen entnimmt. Wird ein neues Dokument vorgeführt, wird die Ähnlichkeit zu den gelernten Dokumenten verglichen. Der Einsatz von lernenden Mustererkennungsverfahren wird meist begrenzt auf die Klassifikation und dient als Hilfsmittel der eigentlichen Datenextraktion.

Über automatische Datenbankabfragen lassen sich die erkannten Ergebnisse überprüfen und als „sicher erkannt“ markieren. Auch kann die Datenbank fehlende Indexmerkmale ergänzen. Die Korrekturen lassen sich so auf ein Minimum reduzieren. Nachgelagerte Datenbankabfragen und Konsistenzprüfungen innerhalb

der Erfassungslösung stellen zusätzlich sicher, dass die Daten korrekt sind.

RFID

Besondere Erwähnung in diesem Zusammenhang muss der Einsatz von RFID (Radio Frequency Identity) erwähnt werden, der sicherlich in naher Zukunft verstärkt im Umfeld von ECM-Anwendungen zu finden sein wird. Informationsobjekte im ECM-Umfeld können digitaler oder körperlicher Natur sein. Einige in körperlicher Form vorhandene Informationsobjekte, wie z.B. Dokumente in Papierform, lassen sich digitalisieren. Bei einer Reihe von Objekten, z.B. Blutproben, Gemälden, Postpaketen, Bauteilen etc. ist diese Digitalisierung



nicht möglich. In der Regel wird bei diesen rein körperlich vorliegenden Objekten nur der Standort im ECM abgespeichert.

COLD

Belege, die selbst erzeugt werden (und deren Aufbau und Datenstruktur damit bekannt sind), eignen sich besonders gut für eine automatische Zuführung in ein ECM-System (COLD-Verfahren).

Dabei geht es darum, angelieferte Ausgabedateien auf Basis vorhandener Strukturinformationen so aufzubereiten, dass sie unabhängig vom erzeugenden System indiziert und an eine Speicherkompo-

nente wie eine dynamische Ablage (Store) oder ein Archiv (Preserve) übergeben werden können. Die „Aggregation“ stellt einen Kombinationsprozess von Dateneingaben verschiedener Erstellungs-, Erfassungs- und zuliefernden Anwendungen dar.

Die COLD-Archivierung stellt eine sehr wichtige Komponente von Enterprise-Content-Management-Sys-

temen dar. COLD dient hier sowohl für die Aufbereitung der Ausgabe für verschiedene Formate als auch als Eingang für die Bereitstellung der Postausgänge in einer gesamtheitlichen Sicht auf alle elektronischen Dokumente (virtuelle Akte).

COLD-Verfahren werden ferner für Migration von Datenspeichern und die automatisierte Überführung von Bewegungsdaten in Dokumentenma-

nagement-Lösungen benutzt. Ebenso werden auf diesem Weg automatisch MS Office-Dokumente abgelegt.



Formatfragen

Die Speicherung der zu verwaltenden Objekte innerhalb eines ECM-Systems ist ein

sehr bedeutender Komplex, da das Format letztendlich die Verwendbarkeit und Darstellung des betreffenden Objekts bestimmt.

Da es eine Vielzahl verschiedener Formate gibt und mit Dokumenten nur dann umgehen kann, wenn man z.B. einen geeigneten Viewer besitzt, ist der Ruf nach standardisierten Formaten nur zu verstehen. Die Do-



kumentenformate lassen sich in folgende Gruppen zusammenfassen:

- **Rasterformate**
sind alle Formate, in denen der darzustellende Inhalt mit einzelnen Bildpunkten unterschiedlicher Farbe wiedergegeben wird. Beispiele sind TIFF und JPEG.
- **Auszeichnungsformate**
sind alle Formate, in denen das Layout der jeweiligen Objekte mittels lesbarer Kennzeichnungen (Tags) festgelegt wird. Beispiele sind XML und HTML.
- **Mixed Object-Formate**
beschreiben jeden Objekttyp (Text, Grafik, Barcode etc.) in einer eigenen, seine Besonderheiten berücksichtigenden Weise. Beispiele sind, PDF, PDF/A, PCL und AFP.

Die im Umfeld von ECM-Lösungen gebräuchlichsten Formate sind wohl TIFF, JPEG, XML und PDF bzw. PDF/A (ISO 19005).

Vor- und Nachteile verbreiteter Formate

Nicht immer stehen Objekte in dem Format zur Verfügung, welches in der aktuellen Anwendung verwendbar ist. Deswegen müssen in diesen Situationen Konvertierungen vorgenommen werden.

Es gibt verschiedenste Zeitpunkte und Gründe für die Anforderung, Dokumente in ein anderes Format zu konvertieren. So kann es vorkommen, dass für eine Anzeige am Bildschirm das Dokument konvertiert werden muss. Dies geschieht unmittelbar während des Viewing-Prozesses. Das Original bleibt dabei unverändert. Vielleicht muss eine Konvertierung für eine Langzeitarchivierung vorgenommen werden.

Hierbei wird das Originalformat verändert. In manchen Fällen müssen sogar mehrere Formate des gleichen Inhalts unter gleichem Index archiviert werden.

Automatische Klassifikation, Taxonomien

Die Nutzung von Dokumentenklassen ist eine der wesentlichen Eigenschaften von Systemen um Dokumente und Informationsobjekte zu schützen, zu strukturieren und zu ordnen, in geeigneter Form in elek-

tronischen Akten zu visualisieren und effizient zu verwalten.

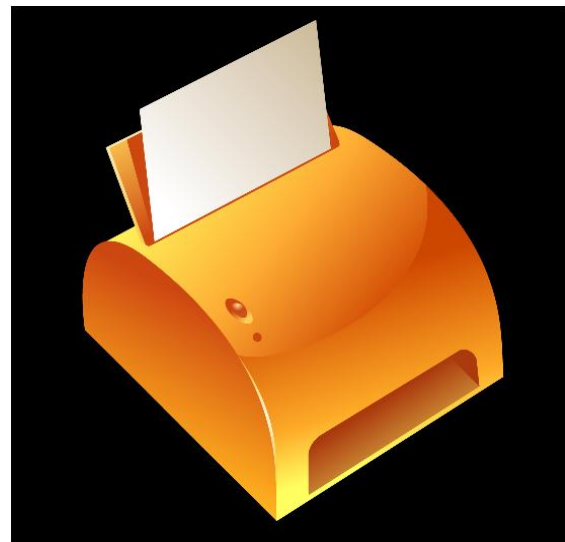
Es ist ein verständliches Anliegen der Anwender, diesen Prozess möglichst zu automatisieren und eingehende, elektronische und papierbasierte Dokumenten automatisch zu indizieren und zuzuordnen. Hierzu werden nach der Erfassung durch den Scanner die Faksimiles mit OCR/ICR-Techniken interpretiert und anschließend werden die Indexmerkmale nach vordefinierten Schemata herausgefiltert, geprüft und mit Stammdaten abgeglichen. Selbstlernende Programme erlauben die Generierung von Strukturen, Aufbau von Ordnungssystematiken und Zuordnungen anhand der Dokumentinformationen.

Solche Systeme können auf Basis vordefinierter Kriterien oder selbstlernend Informationen auswerten. Zur Anwendung können hierbei neuronale Netze oder regelbasierte Ansätze kommen. Neuronale Netze sind selbstlernende Mechanismen bzw. künstliche Intelligenz. Auf Grund einer definierten Lernmenge können alle weiteren Informationen automatisch zugeordnet werden. Der Effekt des Übertrainierens kann diese Systeme allerdings in unkontrollierbare Zustände versetzen.

Regelbasierte Ansätze sind alle Merkmale, die einen bestimmten Dokumententyp beschreiben, müssen durch eine Regel beschrieben wer-

den. Bei sehr heterogenen Dokumenten können erhebliche Administrationsaufwände für die Erstellung der Regeln entstehen.

Es ist verständlich, dass solche Automatismen nicht immer fehlerfrei ablaufen. Es können Erkennungsfehler auf Zeichenebene, Feldebene, Dokumentebene vorkommen. Die Fehlerrate nimmt mit jeder Ebene exponentiell zu.



In einer automatischen Nachverarbeitung kann unter Zuhilfenahme von Kontextwissen eine Korrektur von Fehlern vorgenommen werden.

Besondere Szenarien und Anwendungsbeispiele

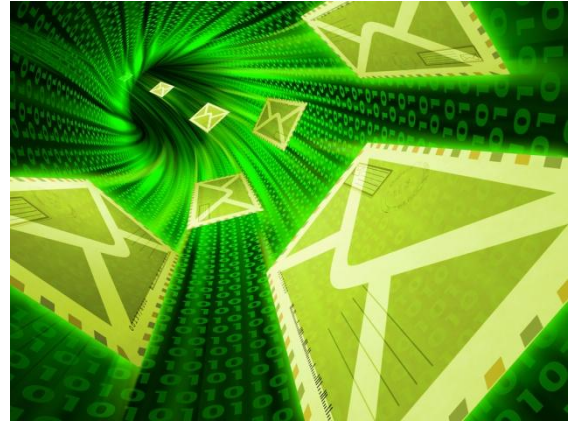
Die bisherigen Ausführungen zum Thema „Capture“ werden abgeschlossen mit einigen kurzen Szenar-

rien und Anwendungsbeispielen, wie sie uns in der Praxis begegnen.

Posteingangslösungen

Die mit am häufigsten anzutreffende Anwendung in ECM-Lösungen ist die Verarbeitung des Posteingangs. Sie ist auch eine der Anwendungen, die auf Grund der Verschiedenartigkeit der Belege mit die höchsten Anforderungen an das Capturing stellt. Automatisierte Posteingangserfassung mit maschineller Indizierung dient zur Überwindung des Flaschenhalses der Erfassung.

Bei diesen Anwendungen ist immer der Umfang der Automatisierbarkeit zu prüfen. Organisation besonders bei stufenweiser Einführung, Zeitpunkt der Bereitstellung, Abhängigkeit von der Verfügbarkeit und Prozentsatz „richtig“ erkannter Daten und Aufwand für Korrekturen sind Punkte, denen besondere Aufmerksamkeit geschenkt werden muss. Wo es möglich ist, sind alle Einflussmöglichkeiten auf selbst erstelltes Schriftgut (Vordrucke, Formulare, Individualbriefe) auszunutzen. Letztendlich sind alle rechtlichen Grundlagen wie Unterschriften, Papiervernichtung und Elektronisches Posteingangsbuch (zertifizierte Zeitstempel) sorgsam zu berücksichtigen. Ein „einheitlicher Postkorb“ ist das, was



die Anwender eigentlich wollen. Das bedeutet:

- Alle Nachrichten aus den Quellen Posteingang, Vorgangsbearbeitung, interne E-Mail, Internet-Mail, Fax, Datenbankrecherchen, Sprachaufzeichnung etc. in nur einem Posteingangsordner
- Der Medienbruch zwischen derzeit verschiedenen Ordnern („man muss wissen, wo die Information ist“) soll vermieden werden.
- Eine einheitliche, strukturierte Benutzeroberfläche mit Dokumentenmanagement-Funktionalität soll angeboten werden.

Rechnungseingangsverarbeitung

Eine Anwendung mit sehr viel Einsparungspotenzial ist die Verarbeitung von eingehenden Rechnungen. Hier werden nach dem Scannen die Rechnungen einer Dokumentenanalyse unterzogen, diverse Daten ex-

trahiert und abschließend in Verbindung mit dem bestehenden ERP-System die Rechnungen eingebucht. Als Daten werden extrahiert:

- Kopfdaten (Identifizieren des Buchungsvorgangs, Lieferant, Rechnungsdatum, Währung etc.)
- Positionsdaten (Artikelangaben, Menge, Einzelpreis, exakte Kontierungsangaben, Steuerangaben, Beschreibung, Zahlungsbedingungen, Skonto etc.)
- Zusatzangaben (Kostenstelle, Auftragsdaten, Projektdaten, Aufteilungsschlüssel etc.)

Schwierigkeiten, die eine 100%ige automatische Verbuchung behindern tauchen dann auf, wenn Rechnungen von noch nicht bekannten Kunden eingehen, noch keine Kunden-Nummer angelegt ist, das Rechnungs-Layouts sich geändert hat oder große



Sammelrechnungen geteilt werden müssen.

Scannen mit elektronischer Signatur

Generell werden elektronische Rechnungen durch die elektronische Signatur effizienter. Denn nur dadurch ist gewährleistet, dass die Vorsteuer abgezogen werden darf (UStG §14), dass umfangreicher Papieraustausch (teilweise zusätzlich zur elektronischen Rechnung) und Aufwendungen für Papierhandling (Portokosten) vermieden werden. Es sind Lösungen sowohl für Ein- als auch für Ausgangsrechnungen verfügbar.

Für die Nutzung der elektronischen Signatur beim Scannen wird verlangt, dass die korrekte Verarbeitung, Indizierung und Qualitätskontrolle jedes Einzeldokuments durch das Scan-Personal mit der persönlichen Signaturkarte bestätigt wird. Welchen Wert hat diese Signatur? Sie gilt nur als Bestätigung, dass vollständig und lesbar erfasst wurde. Es gibt keine Beziehung zum Absender bzw. Nutzer des Dokuments. Diese Vorgehensweise ist bei Organisationen, die der Sozialgesetzgebung unterliegen bindend.

Die elektronische Signatur verbreitet sich zunehmend, die Signatur findet zunehmend Akzeptanz, besonders bei elektronischen Rechnungen. Es

besteht bereits ein großes Angebot an Produkten. Die Öffentliche Verwaltung ist Vorreiter bei Fachanwendungen.

E-Mail-Management

E-Mail-Management rückt immer stärker in den Fokus der Anwender. Immer mehr eigenständige Produkte kommen auf den Markt. Noch beherrschen reine E-Mail-Archivierungslösungen den Markt, stellen jedoch eine Sackgasse dar. Langsam wird den Anwendern bewusst, dass sie weitere Insellösungen schaffen und eine Integration in vorhandene Ablagen in Kunden-, Vorgangs- oder Sachakten benötigen.

Die Bürokommunikationsprogramme quellen über und werden gleichzeitig immer größer und komplexer. Der Spam-Anteil bzw. der Anteil an unnötigen Mails und der Aufwand für die Systemadministration steigen. Bei E-Mails ist die Identifikation der aufbewahrungspflichtigen, der aufbewahrungswürdigen und der übrigen Dokumente besonders aufwändig, insbesondere wenn private Nutzung erlaubt ist. Häufig werden E-Mails und/oder Attachements gedruckt und dann abgelegt. Verschiedene Alternativen werden bei Anwendern eingesetzt. Die vollständige Archivierung liefert mit Sicherheit Konformität mit den



rechtlichen Erfordernissen, bietet die Möglichkeit der Prozessintegration und hat kaum Einfluss auf die Arbeitsweise der Nutzer.

Sie ist aber zugleich umfangreichste, aufwändigste und damit meist kostenintensivste Lösung. Die vollständige Archivierung mit Separierung der Anlagen führt zu wesentlicher Reduzierung der Mailboxgrößen, erhöht damit die Performance und benötigt weniger Hardware. Bei der selektiven Archivierung werden nicht immer alle rechtlichen Anforderungen erfüllt. Nach einem Crash des Mailsystems ist eine Zuordnung der archivierten Anhänge zu Mails meist schwierig. Ein Offline-Rückgriff auf die Anhänge ist nicht möglich. In der Realität finden sich natürlich auch Kombinationen der Verfahren. Alle diese Verfahren können anwender- oder systemgetrieben eingesetzt werden.

Scan-Outsourcing

Eine mittlerweile sehr gebräuchliche Weise Dokumente zu erfassen, ist die Vergabe dieser Tätigkeit außer Haus.

Doch ohne ausreichende Vorbereitung wird die Einbeziehung eines externen Dienstleistungsunternehmens nicht erfolgreich sein. Es sind grundlegende Fragen zu klären (z.B. Welches Ziel wird verfolgt? Ist es unternehmenspolitisch durchsetzbar? Ist das Volumen ausreichend? Sind die Prozesse überhaupt abgrenzbar?). Die eigenen Prozesse müssen mit Spezifikation der Schnittstellen, Angaben zur Dokumentenstruktur und zu Benutzergruppen dokumentiert werden. Generelle

Anforderungen und geforderte Service Levels müssen festgelegt werden. Eine Wirtschaftlichkeitsbetrachtung sollten durchgeführt werden und Angebote sollten eingeholt und verglichen werden.

Üblicherweise wird Outsourcing an einen Dienstleister vergeben, der die Aufgaben in seinen Räumlichkeiten durchführt. Es gibt aber auch die Variante des Inhouse Outsourcing. In diesem Fall „übernimmt“ der Dienst-

leister praktisch die Scan-Stelle des Auftraggebers mit seinem Personal oder er bringt sogar seine Scanner mit.

Altakten-Scannen

Bei der Neueinrichtung einer ECM-Lösung stellt sich oft die Frage, wie können „alte“ Bestände übernommen werden. Natürlich können diese eingescannt werden. Da es sich hierbei meist um sehr große Bestände handelt,

sollte dies konzeptionell gut durchdacht werden. Grundsätzlich bieten sich folgende Varianten an: der gesamte Altbestand wird gescannt, der Altbestand ab einem bestimmten Stichtag wird ge-

scannt, die Akten werden abhängig von geschätzter Zugriffshäufigkeit gescannt oder sie werden „on demand“ d.h. bei Zugriff gescannt.

Für die Übernahme großer Altaktenbestände sollte stets ein externer Dienstleister herangezogen werden. Es lohnt sich nicht, hierfür die Infrastruktur und Prozesse im Haus bereit zustellen (Hard- und Software, Personal, Raum).



Fazit & Ausblick

Abschließend sollen einige aktuelle Fragen aufgeworfen werden, die derzeit die Anwender und die Anbieter von Capture-Lösungen bewegen:

- PDF/A als DAS Langzeitarchivformat?! Viele Capture-Systeme liefern beim Scannen noch TIFF
- Ist Farbe beim Scannen rechtsrelevant?
- Benötigen wir den Begriff des „Elektronischen Originals“, wo doch in IT-Systemen meistens mit Kopien gearbeitet wird?
- Bleibt es bei nur zwei führenden internationalen Capture-Basistechnologie-Anbietern?
- Wird sich der Markt für automatische Klassifikation noch weiter konsolidieren?



- Wie lange bleibt die Erfassung von Papier noch wichtig, wenn der Trend zu immer mehr elektronisch originär entstandenen Dokumenten geht?
- Nur das Scannen outsourcen oder gleich die ganze Lösung außer Haus geben?

Fachlich kann festgestellt werden:

- Capture“ hat sich zu einer eigenständigen Disziplin entwickelt sowohl für die Belieferung von ECM-Systemen und Archiven als auch für die Lieferung von Daten an operative Systeme.
- Die automatische Klassifikation ist praxistauglich, besonders wenn die Ergebnisse mit vorhandenen Daten abgeglichen werden können.