Zeitschrift: Bulletin Electrosuisse
Herausgeber: Electrosuisse, Verband für Elektro-, Energie- und Informationstechnik
Band: 102 (2011)
Heft: 6

Artikel: Langzeitarchivierung von Geschäftsunterlagen
Autor: Jörgens, Ulrich
DOI: https://doi.org/10.5169/seals-856821

Nutzungsbedingungen
Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation
L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use
The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 08.07.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Langzeitarchivierung von Geschäftsunterlagen

Potenzial von Enterprise-Content-Management-Systemen

Es gibt für Unternehmen diverse berechtigte Gründe zur Archivierung von Geschäftsunterlagen. Die elektronische Archivierung bietet zahlreiche Vorteile (z. B. die komfortable Suche), bringt aber auch Herausforderungen mit sich: Fragen zu Datenformat-Veränderungen, Sicherstellung der Authentizität und Datenschutz stehen dabei im Mittelpunkt. Ein Blick auf Enterprise-Content-Management-Systeme und wie sie mit solchen Fragen umgehen.

Ulrich Jörgens

Der Begriff «Archivierung» ruft meist Assoziationen mit Staub und Dunkelheit hervor. Auch bei genauer Betrachtung bleibt der Begriff oft unscharf. Während für die einen mit dem Ablegen von Dokumenten in einer Hängeregistratur die Archivierung bereits abgeschlossen ist, gehen andere sehr viel weiter und verstehen darunter den gesamten Prozess, der folgende Schritte umfasst:

- Aufbereiten von Dokumenten.
- Klassifikation (d.h. Vergeben von Schlagworten für den Rückgriff und zur Bestandsverwaltung).
- Aufbewahrung in speziell vorgesehenen Räumen.
- Rückgriff bzw. Reproduktion.
- Letztlich auch die Vernichtung nicht mehr benötigter Dokumente.

An die Aufbewahrung werden dabei in der Regel besondere Anforderungen gestellt. So muss das Archivgut sowohl gegen unbefugten Zugriff als auch gegen äussere Einflüsse wie Witterung, Feuer, Flut etc. geschützt werden.

Ziel der Archivierung ist dabei vor allem der Erhalt von Authentizität, Integrität und Vollständigkeit der Informationen.

Bei der elektronischen Archivierung kommt allerdings erschwerend hinzu, dass elektronische Dokumente in zahlreichen Formaten existieren, die sich zum Teil noch über ihren Lebenszyklus ändern können. Zudem ist das unbemerkte Verändern elektronischer Dokumente sehr viel einfacher als bei Papier. Die elektronische Archivierung muss daher Verfahren bereitstellen, welche die Authentizität und Integrität von elektronischen Dokumenten sicherstellen. Einfacher gesagt, wenn etwas verändert wird, dann muss das bemerkt werden können und nachvollziehbar sein. Unbeabsichtigte und nicht autorisierte Änderungen müssen wirkungsvoll unterbunden werden können.

Dieser Artikel behandelt die elektronische Archivierung von Geschäftsdokumenten. Spezielle Anforderungen, wie sie zum Beispiel die historische Archivierung stellt, werden bewusst ausgeklammert.

Archivierungshintergründe

Geschäftsunterlagen werden meist nicht freiwillig aufbewahrt. In der Schweiz sind Unternehmen aus unterschiedlichen Gründen zur Aufbewahrung ihrer Geschäftsunterlagen verpflichtet:

- Obligationenrecht, insbesondere die Geschäftsbücherverordnung (GeBüV).
- Steuergesetze (vor allem Mehrwertsteuergesetz).
- Zahlreiche branchenspezifische Gesetze und Verordnungen (vor allem im Finanz- und Pharmaumfeld sowie im Gesundheitswesen).
- Qualitätsmanagement.
- Gewährleistungsanforderungen.

Systeme, die die Anforderungen der Geschäftsbücherverordnung respektive des Mehrwertsteuergesetzes erfüllen, werden auch als «revisionssicher» bezeichnet.

Datenschutz

Nicht zuletzt stellt auch der Datenschutz besondere Anforderungen, welche Informationen wem zugänglich gemacht werden dürfen und wie lange diese aufbewahrt werden müssen. So müssen bestimmte Personalunterlagen nach Ausscheiden eines Mitarbeiters fristgerecht vernichtet werden. Vor allem auch im Gesundheitswesen ist der Umgang mit sensiblen Daten heikel und stellt besondere Anforderungen an den Datenschutz.

Hinzu kommen Unternehmensinteressen, etwa die Konservierung von Know-how und die Transparenz und Nachvollziehbarkeit von Geschäftsprozessen.

Konfliktpotenzial bei Anforderungen

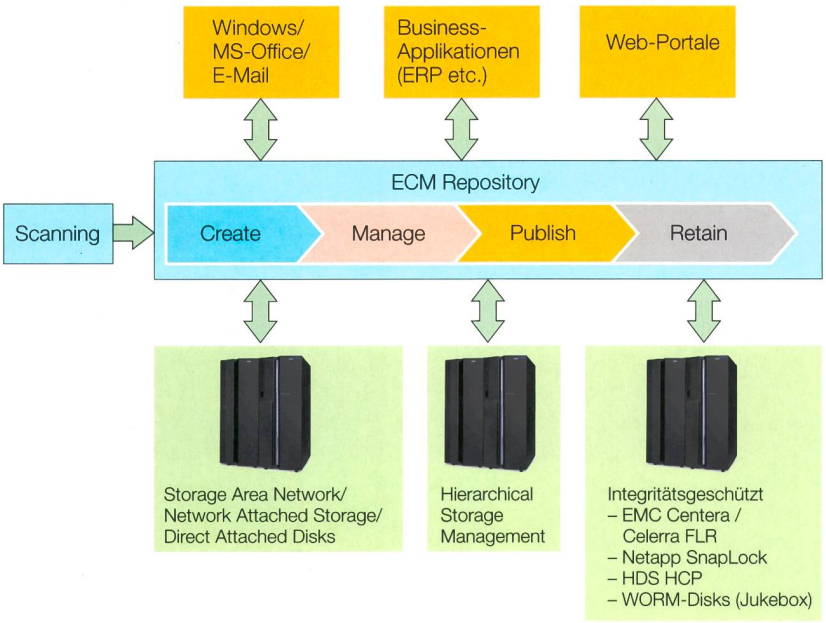
Die Vielzahl der Anforderungen macht die Komplexität der Materie deutlich. Oft sind die Anforderungen an die Archivierung sogar widersprüchlich. Daten müssen wegen eines schwebenden Verfahrens unbegrenzt aufbewahrt werden, obwohl sie aus Datenschutzgründen vernichtet werden müssten.

Häufig gelten für ein und dasselbe Dokument folglich unterschiedliche Aufbewahrungsregeln, die zudem zum Zeitpunkt der Archivierung nicht immer bekannt sind.

Moderne elektronische Archivsysteme verfügen inzwischen über ausreichende Flexibilität, um diese Anforderungen erfüllen zu können. Eine rein technische Lösung gibt es für diese Problematik aber heute leider nicht. Hier ist ein nicht zu vernachlässigender organisatorischer Aufwand erforderlich, um Dokumentenklassen (z. B. Lieferantenrechnung, Projektvertrag, Kaufvertrag) zu bilden und dann für jede Klasse den Lebenszyklus und die Zugriffsberechtigungen festzulegen.

Jedes Dokument wird dann klassiert, das heisst einer (oder mehreren) Klassen zugeordnet, wodurch die Aufbewahrungsregeln und Berechtigungen festgelegt werden.

Auch heute erfolgt die Klassifikation noch oft manuell, etwa beim Scanning oder beim «Hochladen» eines Word-Dokuments ins elektronische Archiv. Dies ist aufwendig, für den Benutzer lästig und zudem fehlerträchtig.



Elca

Bild 1 Schema einer unternehmensweiten ECM-Plattform.

Enterprise Content Management

Eine der wesentlichen Aufgaben bei der Einführung (oder Modernisierung) von elektronischen Archivsystemen ist, diese Aufgabe so weit wie möglich zu automatisieren. Dabei geht es prinzipiell darum, so früh wie irgend möglich mit der systematischen Erfassung zu beginnen, also praktisch den gesamten Lebenszyklus eines Dokuments von der Erstellung bzw. vom Eintritt in das Unternehmen bis zur Vernichtung systematisch zu verwalten. Für diese ganzheitliche Sicht werden auch regelmässig die Begriffe «Enterprise Content Management» oder «Information Lifecycle Management» (sowie eine Fülle weiterer) verwendet, auf deren Definition hier aber verzichtet wird.

Bild 1 zeigt eine schematische Übersicht, wie eine ECM-Plattform als zentrale Ablage in die IT-Landschaft eines Unternehmens integriert wird.

Erfassung und künftige Lesbarkeit

Die Idee dahinter ist, aus dem Bearbeitungsprozess heraus Informationen zum Geschäftskontext zu sammeln und somit Suchkriterien praktisch ohne Zusatzaufwand zu gewinnen. Dies reduziert mühsames manuelles Erfassen bei der Archivierung und verhindert zudem die Mehrfacherfassung von Informationen.

Insbesondere für das Scanning von Dokumenten stehen heute ausgereifte Lösungen für die Datenextraktion zur Verfügung (Stichwort OCR, Formular- und Freiformerkennung). Diese Systeme können beispielsweise Lieferantenrechnungen automatisch erkennen und die wesentlichen Informationen bis hin zu den Rechnungspositionen extrahieren. Genug Informationen, um nicht nur den Buchungsvorgang zu automatisieren und einen Freigabeworkflow zu steuern, sondern auch für die automatische Archivierung des Dokuments sowie seiner Vernichtung 10 Jahre nach Ablauf des Geschäftsjahres.

Neben dem Scanning gewinnen natürlich andere Dokumentenquellen zunehmend an Bedeutung. Die wichtigsten darunter sind:

- Dokumentenoutput aus Fachapplikationen (z. B. Ausgangsrechnungen, automatisierte Kundenkorrespondenz etc.), die gedruckt oder via Web beziehungsweise E-Mail publiziert werden.
- MS-Office-Dokumente, welche oft in Teams gemeinschaftlich erstellt und dann in einem Workflow freigegeben werden.
- E-Mails, die aus rechtlichen Gründen oder zur Entlastung der E-Mail-Infrastruktur archiviert werden.
- Web-Inhalt, der reproduzierbar bleiben soll, auch wenn er nicht mehr online verfügbar ist.

Bei der Archivierung von Dokumentenoutput kann ein sehr hoher Automationsgrad erzielt werden, denn die Suchkriterien (Kundennummer, Rechnungsnummer oder ähnliche) können in der Regel von der Applikation gleich mitgeliefert werden. Kritische Aspekte sind hier vor allem das Volumen (welches zum Teil mehrere Millionen Dokumente pro Tag erreichen kann) sowie die Formatumwandlungen (etwa von einem Druckformat wie AFP in PDF).

Die Problematik der Formatumwandlung gilt gleichermassen für E-Mails und MS-Office-Dokumente. Hier muss das Archiv sicherstellen, dass die Dokumente auch nach 10 oder 25 Jahren (bei personenbezogenen Dokumenten oft sogar noch länger) noch originalgetreu reproduziert werden können. Ob aber zum Beispiel das MS-Office-2000-Format in 25 Jahren noch problemlos reproduziert werden kann, ist nicht vorhersagbar.

Adobes PDF-Format hat sich daher (neben TIFF für gescannte Dokumente) als Standard weitestgehend durchgesetzt. Mit PDF/A gibt es neuerdings auch einen ISO-Standard, der auf die Belange der Langzeitarchivierung ausgerichtet ist.

Physische Speicher

Schliesslich stellt sich noch die Frage, wie die Dokumente nach dem Erfassen gespeichert werden. Traditionell wurden elektronische Daten auf sogenannten WORM-Medien (CD, DVD, magneto-optische Disks) gespeichert. Diese haben den Vorteil, dass Daten nur einmal geschrieben und danach nicht wieder verändert werden können, wodurch Integrität und Authentizität der Daten sichergestellt sind. Diese Datenträger werden in sogenannten «Jukeboxen» eingesetzt. Dies sind geschlossene Systeme, die über zahlreiche Ablagefächer, mehrere Laufwerke und einen Roboterarm zum Medienwechsel verfügen. Die wesentlichen Vorteile sind mit den günstigen Investitionskosten und der Revisionssicherheit bereits aufgezählt. Dem steht eine Reihe von Nachteilen gegenüber: Der Zugriff ist langsam, die Fehleranfälligkeit aufgrund des hohen Mechanikanteils hoch (Störungen verursachen Betriebskosten) sowie die Flexibilität bezüglich Lebenszyklussteuerung tief (dies muss immer für ein ganzes Medium erfolgen).

Softwarelösungen als Alternative

In den letzten Jahren sind Soft-WORM-Lösungen immer populärer geworden. Diese basieren auf «normalen» Festplattensystemen, wie sie auch in Fileservern eingesetzt werden. Im Gegensatz zum Fileserver können die gespeicherten Daten allerdings auf Dokumentenebene für eine einstellbare Periode schreibgeschützt

werden. Nur der Hersteller kann diesen Schreibschutz in der Regel umgehen.

EMC war mit der Centera der Vorreiter in dieser Technologie. Heute haben praktisch alle führenden Hersteller entsprechende Systeme im Angebot. Die folgende – unvollständige – Liste nennt die wichtigsten Vertreter dieser Gattung:

- EMC: Centera, Celerra FLR
- Hitachi Data Systems: HCP (Bild 2)
- NetApp: SnapLock

Charakteristisch für diese Lösungen ist u.a. eine Steuerung der Aufbewahrungsfrist auf Datei- oder Speicherbereichsebene, eine Deduplikation (identische Dokumente werden nur einmal gespeichert, alle weiteren werden referenziert), hohe Schreib- und Lesegeschwindigkeit (analog zu Fileservern), hohe Fehlertoleranz sowie eine verbesserte Datensicherheit durch Spiegelung an einem anderen Standort.

Nach Ablauf der Aufbewahrungsfrist wird das Dokument freigegeben und kann modifiziert, gelöscht oder erneut schreibgeschützt werden.

Wie bei einer Jukeboxlösung ist für die Steuerung des Archivierungsvorgangs, für die Inventarisierung der archivierten Dokumente sowie natürlich für die Suche eine separate Archivsoftware erforderlich. Diese Software übernimmt in der Regel auch die erforderlichen Formatkonvertierungen und verwaltet eine sogenannte Metadatenbank (typischerweise auf Basis der gängigen Datenbanksysteme wie Oracle, MS-SQL-Server, DB/2).

Drei Suchstrategien

Praktisch alle marktüblichen Lösungen unterstützen dabei drei wesentliche Suchstrategien. Die erste Strategie ist die **strukturierte Suche** über Metadaten (vom System vergeben, wie Archivierungsdatum, Autor, über spezifische Metadaten wie Kundennummer, Dokumentenklasse. Die zweite ist die **unstrukturierte Suche** (die Volltextsuche im Content). Drittens wird noch das **Navigieren in Ordnerhierarchien** (analog zum Fileserver) eingesetzt.

Für die Volltextsuche haben viele Hersteller leistungsfähige Suchmaschinen eingebunden. Die Erfahrung lehrt allerdings, dass der Wert der Volltextsuche oft überschätzt wird. Mit zunehmendem Füllgrad werden auch die Trefferlisten immer länger. Das Finden des richtigen Dokuments wird dadurch oft zur Glückssache.

Die strukturierte Suche wird gerne vernachlässigt, weil beispielsweise das Erfassen der Suchkriterien mühsam ist und oft manuell erfolgen muss. Ausserdem erfordert die Definition der Suchkriterien eine aufwendige Analyse von Unternehmensprozessen, der darin genutzten Dokumente sowie des Benutzerverhaltens.

Für eine längerfristig gut funktionierende Suche ist die Definition eines möglichst schlanken Satzes an Schlüsselkriterien von essenzieller Bedeutung. Soweit möglich, sollten diese Schlüsselkriterien automatisch gewonnen werden können, etwa durch Abgleich mit Datenbeständen aus Geschäftsapplikationen oder durch Datenextraktion per OCR und Formularverarbeitung.

Ein wesentliches Problem von hierarchischen Ordnerstrukturen ist deren mangelnde Flexibilität. Möchte man Lieferantenrechnungen nicht nur über den Lieferanten auffinden können, sondern auch alle Lieferantenrechnungen zu einem Projekt, so muss in einer Ordnerstruktur oft manuell verlinkt werden.

Vorteile dynamischer Verfahren

Diesem Manko treten dynamischere Verfahren entgegen. Immer weitere Verbreitung finden in den letzten Jahren Suchmethoden, wie die sogenannte «Facetted Search», die es ermöglichen, über die Schlüsselkriterien Suchfilter zur Einschränkung der Suchresultate zu bilden.

Hitachi Data Systems

Bild 2 Ein Beispiel einer Soft-WORM-Lösung: Die Hitachi Content Platform (HCP).

Diese Technik wird zum Beispiel auch von Ebay oder Amazon zum Auffinden von Artikeln angeboten.

Diese Schlüsselkriterien dienen oft auch dem Bilden und Strukturieren von virtuellen Dossiers (welche dann für den Benutzer wieder wie Ordner aussehen) wie auch dem Direktzugriff aus Geschäftsapplikationen (z.B. SAP) auf archivierte Dokumente. Letzteres erfordert eine Integration mit der jeweiligen Geschäftsapplikation. Hier geht der Trend verstärkt in Richtung serviceorientierte Architekturen, welche die Integration stark vereinfachen und die Entwicklungs- und Betriebskosten stark senken können.

Eine tiefe Integration mit den Geschäftsapplikationen ist der Schlüssel zu effizienten und ergonomischen Archiv-

Résumé **L'archivage à long terme des documents professionnels**

Le potentiel des systèmes de gestion du contenu d'entreprise

Les motivations des entreprises concernant l'archivage des documents professionnels sont diverses et légitimes. L'archivage électronique offre de nombreux avantages (par exemple, une recherche aisée). Toutefois, elle représente aussi des défis. En effet, les questions relatives aux modifications du format de données, la garantie de l'authenticité et la protection des données sont au centre des préoccupations. Comment les systèmes de gestion du contenu d'entreprise répondent-ils à de telles questions et quels avantages apportent-ils lorsqu'on les utilise sous forme de vastes systèmes d'informations ? A cette occasion, l'archive même est réduite à la fonction de dépôt, lequel reste en grande partie caché à la vue de l'utilisateur. No

lösungen, da der Benutzer für den Archivzugriff die Applikation nicht wechseln muss. Ausserdem erübrigt sich eine Suche, da aus dem Geschäftskontext das richtige Dokument direkt referenziert und angezeigt werden kann.

Damit lässt sich dann auch der Nutzen eines Archivsystems als umfassendes Informationssystem voll ausschöpfen. Das Archiv selbst wird dabei zu einer Lagerstätte reduziert, welche dem Benutzer weitgehend verborgen bleibt.

Neben dem Nutzen als Auskunftssystem und der Beschleunigung von Geschäftsprozessen kann ein Archivsystem auch zur Entlastung und Optimierung der Infrastruktur genutzt werden. Erfahrungsgemäss werden 80–90% der gespeicherten Informationen nie wieder verändert. Gleichwohl belegen Sie Kapazität auf Speichersystemen und müssen in Backup und Recovery eingebunden werden. In vielen Unternehmen führt die schiere Informationsflut zu Datenmengen, deren Wiederherstellung Zeitfenster von mehreren Stunden bis Tagen benötigt.

Elektronische Archivsysteme können so ausgelegt werden, dass Risiken des Datenverlusts weitgehend ausgeschlossen werden können und somit die Archivspeicher nicht zwingend in das tägliche Backup einbezogen werden müssen.

Elektronische Archivlösungen werden also immer stärker fester Bestandteil der Unternehmensinfrastruktur. Neben lästigen Pflichten bieten sie durchaus messbaren Nutzen, der allerdings oft nur schwer im Voraus quantifiziert werden kann.

Um das Potenzial umfassend nutzen zu können, ist eine sorgfältige Analyse und Konzeption erforderlich. Die Fallstricke liegen oft in kleinen Details verborgen, die ohne Erfahrung in der Implementierung von ECM-Projekten kaum zu erkennen sind, den Projekterfolg jedoch gefährden können.

Nicht zuletzt betrifft ein derartiges Projekt praktisch jeden Mitarbeiter im Unternehmen. Die Einführung einer ECM-Lösung ist daher hauptsächlich als Organisationsprojekt anzusehen. Um den vollen Nutzen aus einer Lösung ziehen zu können, sind Änderungen in der Arbeitsweise und den Prozessen unvermeidlich. Es ist daher empfehlenswert, einen erfahrenen Implementierungspartner an Bord zu holen, welcher auch die organisatorischen Aspekte berücksichtigt.

Angaben zum Autor

Ulrich Jörgens beschäftigt sich seit seinem Studium der Informatik mit der Automation von Geschäftsprozessen mittels Enterprise Content Management Lösungen. Er ist bei Elca Informatique SA als Senior Manager im Bereich ECM für den gesamten Lebenszyklus von ECM-Implementierungsprojekten von der Analyse bis zur Realisierung verantwortlich.

Elca Informatique SA, 1211 Genève 26
ulrich.joergens@elca.ch

Anzeige