Elektronische Archivierung

Eine „never ending Story“?

Dr. Ulrich Kampffmeyer

PROJECT CONSULT
Unternehmensberatung
Dr. Ulrich Kampffmeyer GmbH

Hamburg, 2012

Elektronische Archivierung – eine „never ending Story“?

Über elektronische Archivierung wird seit Jahrzehnten geschrieben. Zahlreiche Produkte sind am Markt verfügbar und zahlreiche Anbieter sind auch schon verschwunden. Die langzeitige Aufbewahrung und Bereitstellung von Information ist eben ein besonderes Thema. Es läuft diametral zu allen anderen Trends, die von einer immer größeren Beschleunigung der Entwicklung gekennzeichnet sind. Elektronische Archivierung könnte der Ruhepol in der Informations- und Kommunikationslandschaft sein. Allerdings wird ihm zu wenig Achtung zu Teil und richtig konsequent haben nur die wenigstens das Thema Archivierung adressiert. Wir leben im Dunklen Zeitalter der frühen Informationskultur.

Kurz nach der Erfindung der Schrift wurden die ersten Dokumente geschaffen. Kurz darauf entstand der Bedarf, sie aufzubewahren. Das Archiv wurde erfunden. In der Antike wurden wichtige Dokumente der Stadt und ihrer Bürger im Archaion, dem Rathaus, verwahrt. Hiervon leitet sich unser Begriff „Archiv“ ab. Bisher orientierten sich Archive an den physischen Gegebenheiten der aufzubewahrenden Aufzeichnungen – Papier, Tontafeln, Papyri, Buchenholztafeln, Schnurbündel, Bücher, Ordner usw. Seit zwei Generationen, knapp 50 Jahren, verändert sich diese Tradition. Immer mehr Informationen und Dokumente werden digital geboren. Die elektronische Version ist das Original und der Ausdruck auf Papier ist nur noch eine mögliche Form der Repräsentation. Immer mehr Dokumente sind nicht mehr zur Aufbewahrung in Papierform geeignet.

Was ist ein elektronisches Archiv?

Der Begriff elektronische Archivierung steht für die unveränderbare, langzeitige Aufbewahrung elektronischer Informationen. Für die elektronische Archivierung werden in der Regel spezielle Archivsysteme eingesetzt. In Deutschland werden mit dem Begriff elektronische Archivierung unterschiedliche Komponenten zusammengefasst, die im angloamerikanischen Sprachgebrauch separat als "Records Management", "Storage" und "Preservation" bezeichnet werden. Der wissenschaftliche Begriff eines Archivs und der Archivierung ist zudem inhaltlich nicht identisch mit dem Begriff, der von der Dokumentenmanagementbranche oder bei der Datensicherung verwendet wird.
Elektronische Archivierung steht für die unveränderbare, langzeitige Aufbewahrung elektronischer Information. Für die elektronische Archivierung werden in der Regel spezielle Archivsysteme eingesetzt. Der Begriff Elektronische Archivierung fasst unterschiedliche Komponenten zusammen, die im angloamerikanischen Sprachgebrauch separat als "Records Management", "Storage" und "Preservation" bezeichnet werden. Der wissenschaftliche Begriff eines Archivs und der Archivierung ist inhaltlich nicht identisch mit dem Begriff, der von der ECM-Branche verwendet wird.

© PROJECT CONSULT GmbH 2012

Der Begriff der elektronischen Archivierung wird sehr unterschiedlich benutzt. Während heute Unternehmen schon Aufbewahrungsfristen von 10 Jahren für handelsrechtlich und steuerlich relevante Daten und Dokumente als nur sehr schwierig umsetzbar sehen, wird in historischen Archiven von einer sicheren, geordneten und jederzeit zugreifbaren Aufbewahrung von Informationen mit Speicherzeiträumen von 100, 200 oder gar 300 Jahre gesprochen. Angesichts der sich ständig verändernden Technologien, immer neuer Software, Formate und Standards, eine gigantische Herausforderung für die Informationsgesellschaft.
Die Aufbewahrung, Erschließung und Bereitstellung von Informationen ist eine Voraussetzung für die Arbeitsfähigkeit moderner Unternehmen und Verwaltungen. Mit dem exponentiellen Wachstum elektronischer Informationen wachsen die Probleme der langzeitigen Aufbewahrung, obwohl moderne Softwaretechnologien wesentlich besser geeignet sind, Informationen zu verwalten, als dies herkömmlich mit Papier, Aktenordnern und Regalen möglich war. Immer mehr Information entsteht digital und die Ausgabe als Papier ist nur noch eine mögliche Repräsentation des ursprünglichen elektronischen Dokuments. Durch den Einsatz elektronischer Signaturen erhalten elektronische Dokumente den gleichen Rechtscharakter wie ursprünglich manuell unterzeichnete Schriftstücke. Solche digitalen Dokumente existieren rechtskräftig nur noch in elektronischer Form. Diese Entwicklungen zwingen inzwischen jedes Unternehmen, sich verstärkt mit dem Thema elektronische Archivierung auseinanderzusetzen. Und – Archivierung ist nicht Datensicherung oder Kopien-Erstellen; Archivierung ist die systematische Aufbereitung von Information für eine langfristige Nutzung.

Revisionssicherheit und Aufbewahrung

Den Begriff „Revisionssicherheit“ mit seiner Ableitung „revisionssichere Archivierung“ gibt es seit 1992. Er wurde in einer Reihe von Artikeln und später in einem Code of Practice des VOI Verband Optische Informationssysteme e.V. vom Verfasser eingeführt.
Revisionssicherheit bezieht sich auf elektronische Archivierungssysteme, die den Anforderungen des Gesetzgebers, besonders HGB und Steuerrecht, entsprechen. Die Revisionssicherheit bezieht sich dabei nicht nur auf technische Komponenten, sondern auf die gesamte Lösung. Revisionssicherheit schließt sichere Abläufe, die Organisation des Anwenderunternehmens, die ordnungsgemäße Nutzung, den sicheren Betrieb und den Nachweis in einer Verfahrensdokumentation ein. Wesentliches Merkmal revisionssicherer Archivsysteme ist, dass alle gespeicherten Informationen datenbankgestützt wieder auffindbar, nachvollziehbar, unveränderbar und verfälschungssicher archiviert sind. Revisionssichere Archivierung ist so ein wesentlicher Bestandteil für die Compliance von Informationssystemen.
Revisionssichere Archivierung ist dann gewährleistet, wenn die Archivsystemlösung den Anforderungen des Handelsgesetzbuches §§ 239, 257 HGB sowie der Abgabenordnung und den GoBS an die sichere, ordnungsgemäße Aufbewahrung von kaufmännischen Dokumenten entspricht und die Aufbewahrungsfristen von sechs bis zehn Jahren erfüllt.

© PROJECT CONSULT GmbH 2012

Das Handelsgesetzbuch (HGB) und die Abgabenordnung (AO) geben hier die Grundlagen für die Speicherung, unabhängig ob in herkömmlichen Papierarchiven oder elektronischen Systemen, vor:

- *Ordnungsmäßigkeit*
- *Vollständigkeit*
- *Sicherheit des Gesamtverfahrens*
- *Schutz vor Veränderung und Verfälschung*
- *Sicherung vor Verlust*
- *Nutzung nur durch Berechtigte*
- *Einhaltung der Aufbewahrungsfristen*
- *Dokumentation des Verfahrens*
- *Nachvollziehbarkeit*
- *Prüfbarkeit.*

Auf die revisionssichere Archivierung beziehen sich auch die 10 Grundsätze des VOI Verband Organisations- und Informationssysteme (Version von 2009):

(1) *Jedes Dokument muss nach Maßgabe der rechtlichen und organisationsinternen Anforderungen ordnungsgemäß aufbewahrt werden.*
(2) *Die Archivierung hat vollständig zu erfolgen – kein Dokument darf auf dem Weg ins Archiv oder*
(3) *Jedes Dokument ist zum organisatorisch frühestmöglichen Zeitpunkt zu archivieren.*
(4) *Jedes Dokument muss mit seinem Original übereinstimmen und unveränderbar archiviert werden.*
(5) *Jedes Dokument darf nur von entsprechend berechtigten Benutzern eingesehen werden.*
(6) *Jedes Dokument muss in angemessener Zeit wiedergefunden und reproduziert werden können.*
(7) *Jedes Dokument darf frühestens nach Ablauf seiner Aufbewahrungsfrist vernichtet, d.h. aus dem Archiv gelöscht werden.*
(8) *Jede ändernde Aktion im elektronischen Archivsystem muss für Berechtigte nachvollziehbar protokolliert werden.*
(9) *Das gesamte organisatorische und technische Verfahren der Archivierung kann von einem Sachverständigen Dritten jederzeit geprüft werden.*
(10) *Bei allen Migrationen und Änderungen am Archivsystem muss die Einhaltung aller zuvor aufgeführten Grundsätze sichergestellt sein.*

Genaugenommen kann dieser Bereich der Archivierung nicht als „akademisch reinrassige Archivierung“ durchgehen. In den erwähnten Texten des Handelsgesetzes, des Steuerrechts, des Bürgerlichen Gesetzbuches, usw. findet sich nur der Begriff „Aufbewahrung“. Der Begriff Archivierung kommt lediglich im

© PROJECT CONSULT GmbH 2012

Bundesarchivgesetz und in den Landesarchivgesetzen vor. Das Bundesarchivgesetz oder Gesetz über die Sicherung und Nutzung von Archivgut des Bundes in Deutschland legt fest, wie das Archivgut des Bundes durch das Bundesarchiv auf Dauer zu sichern, nutzbar zu machen und wissenschaftlich zu verwerten ist. Nach § 5 „*steht mehr als 30 Jahre zurückliegendes Archivgut im Allgemeinen der Forschung offen. Archivgut des Bundes, das sich auf natürliche Personen bezieht, darf erst 30 Jahre nach dem Tode der Betroffenen durch Dritte benutzt werden. Ist das Todesjahr nicht oder nur mit unvertretbarem Aufwand festzustellen, endet die Schutzfrist 110 Jahre nach der Geburt des Betroffenen*“.
Eigentlich müsste man daher im gesamten Bereich der Archivierung von kaufmännischen Unterlagen, der revisionssichereren Archivierung, und in der öffentlichen Verwaltung, mit den sogenannten „Zwischenarchen“, nur von Aufbewahrung sprechen. In allen Gesetzen und Verordnungen wird immer von Aufbewahrung gesprochen. Aber weder bieten die ECM-Hersteller Aufbewahrungssoftware an noch kennt die Allgemeinheit die im akademischen und historischen Bereich eingeschränkte Bedeutung des Begriffes Archivierung.

Elektronische Langzeitspeicherung

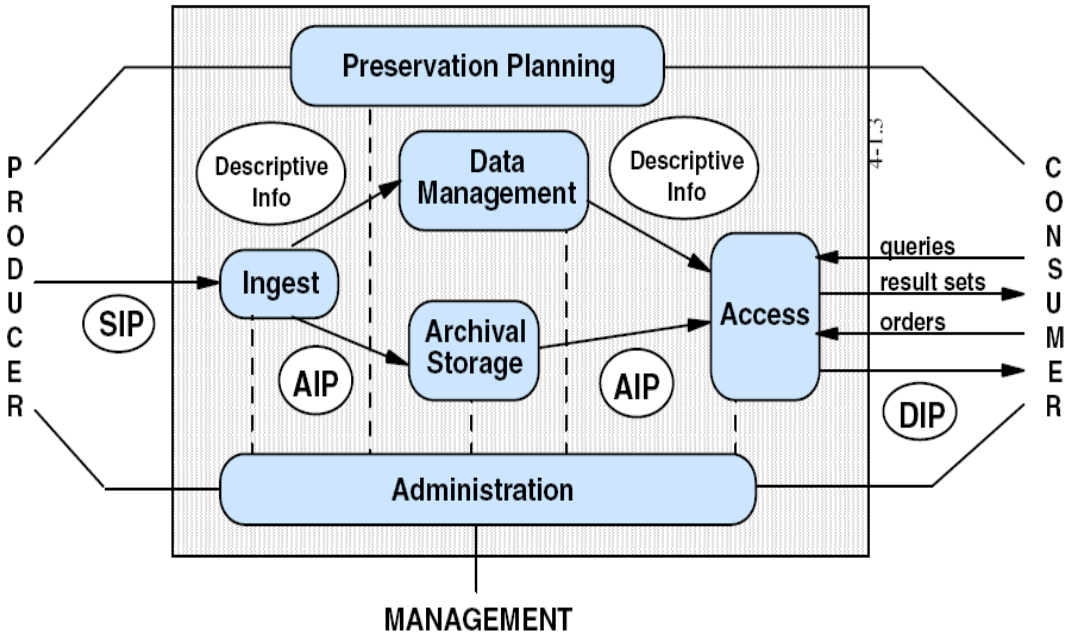
Um sich von dem allgemeinen Begriff „Archivierung“ abzuheben, wird daher im akademischen Bereich, in staatlichen Archiven, in Museen und Sammlungen der Begriff Langzeitarchivierung zunehmend verwendet. Der Begriff Langzeitarchivierung ist im Prinzip ein Pleonasmus, da Archivierung den Langzeitaspekt bereits impliziert und deutlich länger als ein Menschenleben anzusetzen ist. Während heute Unternehmen schon Aufbewahrungsfristen von 10 Jahren für handelsrechtlich und steuerlich relevante Daten und Dokumente als nur sehr schwierig umsetzbar sehen, wird in historischen Archiven von einer sicheren, geordneten und jederzeit zugreifbaren Aufbewahrung von Informationen mit Speicherzeiträumen von 100, 200 oder gar 300 Jahre gesprochen. Angesichts der sich ständig verändernden Technologien, immer neuerer Software, Formate und Standards, eine immense Herausforderung für die Informationsgesellschaft.
Eine wichtige Initiative in Sachen Langzeitarchivierung ist nestor, das deutsche Kompetenznetzwerk zur digitalen Langzeitarchivierung. In nestor arbeiten Bibliotheken, Archive, Museen sowie führende Experten gemeinsam zum Thema Langzeitarchivierung und Langzeitverfügbarkeit digitaler Quellen. Nestor definiert Langzeitarchivierung wie folgt:
„Langzeitarchivierung“ meint im digitalen Zusammenhang mehr als die Erfüllung gesetzlicher Vorgaben über Zeitspannen, während der steuerlich relevante tabellarisch strukturierte Daten verfügbar gehalten werden müssen. „Langzeit“ ist die Umschreibung eines nicht näher fixierten Zeitraumes, währenddessen wesentliche, nicht vorhersehbare technologische und soziokulturelle Veränderungen eintreten; Veränderungen, die sowohl die Gestalt als auch die Nutzungssituation digitaler Ressourcen in rasanten Entwicklungszyklen vollständig umwälzen können. Es gilt also, jeweils geeignete Strategien für bestimmte digitale Sammlungen zu entwickeln, die je nach Bedarf und zukünftigem Nutzungszenarium die langfristige Verfügbarkeit

© PROJECT CONSULT GmbH 2012

und Nachnutzung des digitalen Nutzerinteresses der Auf- und Abwärtskompatibilität alter und neuer Systemumgebungen nur dann eine Rolle, wenn dies dem Anbieter für die Positionierung am Markt erforderlich erscheint. „Langzeit“ bedeutet für die Bestandserhaltung digitaler Ressourcen nicht die Abgabe einer Garantieerklärung über fünf oder fünfzig Jahre, sondern die verantwortliche Entwicklung von Strategien, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können. Der Bedeutungsinhalt von „Archivierung“ müsste hier nicht näher präzisiert werden, wäre er nicht im allgemeinen Sprachgebrauch mit der fortschreitenden Anwendung der Informationstechnik seines Sinnes nahezu entleert worden. „Archivieren“ bedeutet zumindest für Archive, Museen und Bibliotheken mehr als nur die dauerhafte Speicherung digitaler Informationen auf einem Datenträger. Vielmehr schließt es die Erhaltung der dauerhaften Verfügbarkeit und damit eine Nachnutzung und Interpretierbarkeit der digitalen Ressourcen mit ein“.

Standards für Architektur und Funktionalität von Archivsystemen

Der immer noch wichtigste Standard für die Architektur und Funktionalität von elektronischen Archivsystemen ist die ISO 14721 „OAIS Open Archival Information System“, die auf einer Vorläuferversion der Weltraumbehörden basiert. Diese Form der Architektur ist keine technische Systemarchitektur sondern ein generisches Referenzmodell. Das als ISO 14721 verabschiedete Referenzmodell beschreibt ein Archiv als Organisation, in dem Menschen und Systeme mit der Aufgabenstellung zusammenwirken, Informationen zu erhalten und einer definierten Nutzerschaft verfügbar zu machen. Das Modell beschreibt im Detail, wie die von einem Produzenten hergestellte elektronische Information in ein Archivsystem gelangen soll, welche Bearbeitungsschritte für die langfristige Archivierung vorgenommen werden müssen und wie auf die im Archiv gespeicherte Information zugegriffen werden kann.



Das OAIS-Modell definiert sechs Hauptfunktionen:

(1) Preservation Planning (Archivierungsplanung)
- Verfolgen der Technologieentwicklung und Entwickeln von Empfehlungen in Bezug auf Archivierungsstandards und –richtlinien
- Überwachen der Archivierungsverfahren
- Ausarbeiten von Empfehlungen für die Erhaltung der Lesbarkeit der gespeicherten Information
- Planen von Datenmigrationen und Kopiervorgängen

(2) Ingest (Datenübernahme)
- Übernahme der vom anliefernden System erzeugten SIPs (Submission Information Packages)
- Überprüfung auf Vollständigkeit und Unversehrtheit
- Umwandlung der SIPs in AIPs (Archival Information Packages)
- Extraktion der beschreibenden Information für die Datenbank zum direkten Zugriff
- Übertragung der AIPs an den Archivspeicher
- Rückmeldungen an die Verwaltungskomponente Data Management

(3) Data Management (Datenverwaltung)
- Verwaltung der beschreibenden Informationen (Datenbank), die Archivbestände und Dokumente identifizieren, sowie weiterer Daten, die für den Umgang mit dem Archivgut notwendig sind
- Entgegennahme und Bearbeitung von Anfragen (queries) aus dem Nutzungsbereich sowie Bereitstellung von Navigation und Ergebnislisten

(4) Archival Storage (Archivspeicher)
- Aufbewahrung und Erhaltung der AIPs
- Erstellen von Backups
- Regelmäßige Prüfung der Datenintegrität
- Wiederherstellungsmechanismen für Notfälle (Recovery)
- Weitergabe von AIPs an Access für die Nutzung

(5) Access (Nutzung)
- Benutzerinterface
- Ermöglichen von Recherchen und Generieren von Antworten mit Beschreibung der AIPs und Angaben zu deren Verfügbarkeit
- Empfangen und Verarbeiten von Datenanfragen (requests), Umwandeln der AIPs in DIPs (Dissemination Information Packages) und Ausliefern der DIPs an die Benutzer
- Sicherstellen der Einhaltung von Zugriffsberechtigungen

(6) Administration (Verwaltung)
- Steuerung der Gesamtabläufe im OAIS und seiner Außenbeziehungen
- Konfiguration von Hard- und Software
- Vergeben von Zugriffsrechten

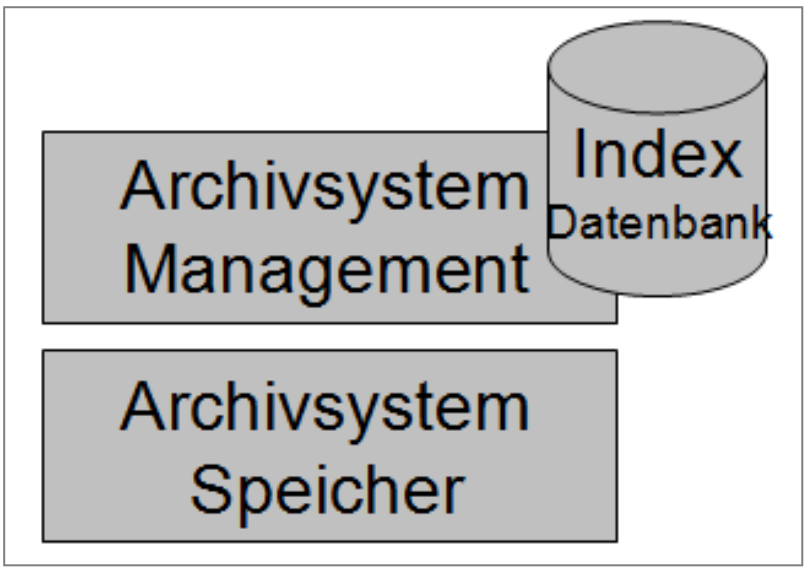
OAIS definiert die notwendigen Funktionsgruppen, jedoch keine technische Standards. So fehlen denn auch die Spezifikationen für die Schnittstellen zwischen den Komponenten und für die verschiedenen Objekte, Pakete genannt (SIP, AIP,

© PROJECT CONSULT GmbH 2012

DIP). Die ISO-Norm hat auch nicht den Anspruch eines technischen Standards zur Erzeugung von Kompatibilität und Austauschfähigkeit. Daher kann die Implementierung sehr unterschiedlich sein. Wesentlich ist aber die Trennung der verschiedenen Komponenten. Besonders Archivspeicher und Verwaltung der Objekte sollten getrennt sein, da sie unterschiedlichen Entwicklungszyklen unterliegen.

Technische Architektur von elektronischen Archivsystemen

Archivsysteme sind traditionell in einem Schichtenmodell aufgebaut, wobei häufig eine dynamische Ablage mit einem Langzeitarchiv kombiniert ist. Alle Bearbeitung, Bereitstellung und Verarbeitung wie Konvertierung findet auf der Ablage-Ebene statt. Auf der Archiv-Ebene sollten die Objekte möglichst statisch und langfristig stabil sein.



Beide Ebenen werden von der gleichen Datenbank mittels Metadaten und Pointern verwaltet, sodass für den Endanwender der Unterschied zwischen den Speicherorten kaum sichtbar wird. Diese Datenbank, häufig auch als Indexdatenbank bezeichnet, verwaltet im Regelfall nur Referenzen auf die separat gespeicherten Objekte. Man spricht hier bei Archivsystemen auch von einer „Referenz-Datenbank-Architektur“. Eine andere Möglichkeit der Speicherung wäre die BlOb-Architektur (Binary large Object), bei der die Objekte direkt in den Tabellen der Datenbank gespeichert werden. Dies entspricht allerdings nur eingeschränkt den Vorgaben von OAIS und kann auch zu Skalierungs- und Update-Problemen führen.

Funktionalität von Archivsystemen

Elektronische Archivsysteme zeichnen sich durch folgende eigenständige Merkmale aus:

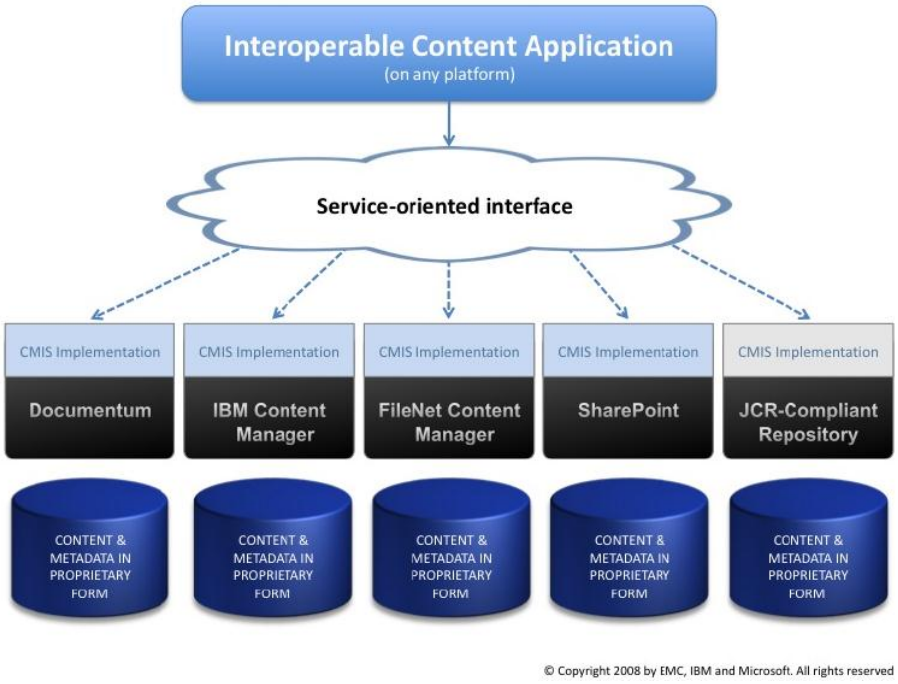
© PROJECT CONSULT GmbH 2012

- Programmgestützter, direkter Zugriff auf einzelne Informationsobjekte, landläufig auch Dokumente genannt, oder Informationskollektionen, z.B. Listen, Container mit mehreren Objekten etc.
- Unterstützung verschiedener Indizierungs- und Recherchestrategien, um auf die gesuchte Information direkt zugreifen zu können
- Einheitliche und gemeinsame Speicherung beliebiger Informationsobjekte, vom gescannten Faksimile über Office-Dokumente, komplexe XML-Strukturen, Listen, Webinhalten, E-Mails oder ganzen Datenbankinhalten
- Verwaltung von Speichersystemen mit nur einmal beschreibbaren Medien einschließlich dem Zugriff auf Medien die sich nicht mehr im Speichersystem direkt befinden
- Sicherstellung der Verfügbarkeit der gespeicherten Informationen über einen längeren Zeitraum, der Jahrzehnte betragen kann
- Bereitstellung von Informationsobjekten unabhängig von der sie ursprünglich erzeugenden Anwendung auf verschiedenen Klienten und mit Übergabe an andere Pro-gramme
- Unterstützung von „Klassen-Konzepten“ zur Vereinfachung der Erfassung durch Vererbung von Merkmalen und Strukturierung der Informationsbasis
- Konverter zur Erzeugung von langfristig stabilen Archivformaten und Viewer zur Anzeige von Informationsobjekten, für die die ursprünglich erzeugende Anwendung nicht mehr zur Verfügung steht
- Absicherung der gespeicherten Informationsobjekte gegen unberechtigten Zugriff und gegen Veränderbarkeit der gespeicherten Information
- Übergreifende Verwaltung unterschiedlicher Speichersysteme, um z.B. durch Zwischenspeicher (Caches) schnellen Zugriff und zügige Bereitstellung der Informationen zu gewährleisten
- Standardisierte Schnittstellen, um elektronische Archive als Dienste in beliebige Anwendungen integrieren zu können
- Eigenständige Widerherstellungsfunktionalität (Recovery), um inkonsistent gewordene oder gestörte Systeme aus sich heraus verlustfrei wieder aufbauen zu können
- Sichere Protokollierung von allen Veränderungen an Strukturen und Informationsobjekten, die die Konsistenz und Wiederauffindbarkeit gefährden können und dokumentieren, wie die Informationen im Archivsystem verarbeitet wurden
- Unterstützung von Standards für die spezielle Aufzeichnung von Informationen auf Speichern mit WORM-Verfahren, für gespeicherte Dokumente und für die Informationsobjekte beschreibende Meta-Daten um eine langfristige Verfügbarkeit und die Migrationssicherheit zu gewährleisten
- Unterstützung von automatisierten, nachvollziehbaren und verlustfreien Migrationsverfahren

© PROJECT CONSULT GmbH 2012

Schnittstellen

Für die Nutzung und Integration von Archivsystemen sind Schnittstellen von besonderer Bedeutung. So wurden für die Anbindung von Erfassungssystemen zum „Ingest“ verschiedenste Standards für Übergabeschnittstellen und SIP-Formate definiert. Einen universellen Ansatz verfolgt CMIS Content Management Interoperability Services. CMIS ist ein Web-Service-Standard für den Austausch von Content mit und zwischen verschiedenen Enterprise Content Management (ECM) Systemen. Es wurde zunächst von EMC, IBM und Microsoft entwickelt und auf den Markt gebracht. Die CMIS-Spezifikation wird von Anbietern wie Alfresco, Adobe Systems, EMC, IBM, Microsoft, OpenText, Oracle und SAP unterstützt und ist inzwischen eine OASIS-Norm. CMIS fördert die Interoperabilität von Systemen und bietet einen einheitlichen Zugriff auf proprietäre Repositories verschiedener Hersteller. Bei diesen Repositories handelt es sich vielfach um elektronische Archivsysteme. CMIS ermöglicht es so auch, Migrationen zu vermeiden, da verschiedene Systeme parallel betrieben werden können, z.B. die Älteren ohne Schreiben neuer Dokumente zwecks Ausalterung, jüngere als aktive Archive.



Speichersysteme

Galten in der Vergangenheit nur einmal beschreibbare digitale optische Speichermedien in Jukeboxen als erste Wahl, so ist die Ära inzwischen ausgeklungen. Die Referenzdatenbank-Architektur hat diese Medien optimal unterstützt und auch die Verwaltung von Medien, die nicht in direktem Zugriff sind, ermöglicht. Dies erlaubte auch gezielte Verteilungs-, Auslagerungs- und Sicherungsstrategien.

© PROJECT CONSULT GmbH 2012

Inzwischen sind Festplattenspeicher deutlich günstiger, größer und schneller als Jukeboxen. Hier hatte sich zunächst eine Gruppe von Produkten etabliert, die in Festplattensubsystemen exakt die gleichen Bedingungen wie auf einem WORM – Write Once Read Many – Medien geschaffen wurden. Aber auch diese sehr speziellen Subsysteme waren nur ein Übergang zur Einrichtung von Speicherbereichen, die durch Software gegen Veränderung geschützt werden. Im Rahmen der Virtualisierung von Speichersystemen setzt sich dieser Trend fort und findet sich auch als die vorherrschende Strategie in Cloud-Archivlösungen. Entscheidend ist dabei, dass die magnetischen Speicher so durch Software geschützt werden, dass sie die vorn aufgeführten Archivbedingungen erfüllen.
Auch für Archivspeichersysteme wurde versucht Standards einzuführen, so z.B. von der SNIA dem Dachverband der Speichersystemanbieter. Diese Standards konnten sich aber bisher nicht durchsetzen.

Meta-Daten-Standards

Für die Verwaltung der archivierten Dokumente in den Repositories sind Metadaten essentiell. Sie erlauben den Zugriff und verwalten den Lebenszyklus. Der Umfang der Metadaten bestimmt zugleich auch die Möglichkeiten des Findens. Für Metadaten im Archiv-Umfeld gibt es zahlreiche Standards und Normen. Zum Teil beziehen sich diese auf bestimmte Anwendungsfelder wie z.B. den bekannten Dublin Core Standard. Andere wie PREMIS oder LMER haben einen universelleren Anspruch. Vielfach orientieren sich aber diese Standards an den Anforderungen der historischen Langzeitarchivierung und unterstützen weniger die Nutzung des Archivgutes in aktuellen Geschäftsprozessen oder als Wissensbasis. Hier differieren die Anforderungen der allgemeinen elektronischen Archivierung mit denen der Langzeitarchivierung.

Objekt-Format-Standards

Ähnlich wie die Metadaten spielen auch die Formate der archivierten Objekte eine wichtige Rolle für die spätere Nutzbarkeit und die Konsistenz des Archivgutes. Bei Einzelobjekten gibt es klare Favoriten. Gescannte Seiten als komprimiertes TIFF oder als PDF, Office-Dokumente gewandelt in PDF, E-Mails zerlegt und in PDF gewandelt, Ausgabedateien in XML-basierte Listenformate, usw. . Der aktuell in diesem Bereich eingesetzte Standard ist PDF/A (PDF/Archive) in der Version 2. PDF/A-2 erzeugt stabile Formate ohne externe Links, kann Farbbilder einbetten und ist mit geläufigen Viewern anzeigbar. PDF/A ist inzwischen durch die als ISO 19005 normiert.
Problematisch wird es aber bei den immer häufiger aufkommenden komplexen Objekten, die nicht mehr für eine geschlossene „dokumentenartige“ Reproduktion geeignet sind. Im medialen Bereich ändern sich außerdem die Formate sehr schnell und besitzen eine große Vielfalt. Erinnert sein an Fotos, Videos, Multimedia-Objekte, Webseiten etc. Die Komplexität lässt sich beliebig steigern wie z.B. mit geografischen

© PROJECT CONSULT GmbH 2012

mehrschichtigen Modellen, 3-D-Visualisierung, Mash-Up usw. Hier steht die Archivierung vor immer neuen Herausforderungen.

Besondere Herausforderungen an die elektronische Archivierung

Archive für Office-Dokumente, für EDI-Daten, für Listen, für E-Mails, als Subsysteme für SAP oder Sharepoint, für gescannte Dokumente, steuerrelevanter Daten und Datensätze sind heute Standard. Die meisten Anbieter haben für die Anwendungsfälle standardisierte Lösungen im Portfolio. Es gibt sie als große Lösungen für Unternehmen, als Freeware für den privaten Anwender, im Hause installiert oder als Cloud-Anwendung. Die Lösungen sind weitgehend ausgereift und matur. Sie leiden lediglich daran, dass je Art der Information häufig eine separate Lösung angeboten wird, die wieder zu einem Informationssilo führt. Ziel sind jedoch Universal-Archive die alle Formen von Informationsobjekten nach den gleichen Kriterien verwalten, unabhängig von Ort, Anwender, ursprünglicher Anwendung und Zeit nutzungsgerecht bereitstellen. Für die klassischen Anwendungsfälle der Aufbewahrung in den Unternehmen gibt es solche Universal-Archive, die allen Anwendungen die archivierten Objekte entsprechend den Berechtigungen der Nutzer zur Verfügung stellen.
Die schnelle technologische Entwicklung, Mobility und Cloud, Social und neue Formate, setzen die Archivierung jedoch massiv unter Druck. Im Folgenden sollen einige akute Problemzonen der Archivierung näher beleuchtet werden.
Datenbankarchivierung
Hier geht es um die komplette Archivierung von Datenbanken einschließlich Nutzdaten, Anwendungslogik und Nutzungsmodell. Dies spielt besonders bei der Abschaltung von Altsystemen eine Rolle, wenn die ursprüngliche Anwendung lauf- und nutzungsfähig weiter bestehen muss. Nur sehr spezielle Archivsysteme erlauben hier eine entsprechende Lösung, die allerdings im Regelfall nicht dem OAIS-Modell entspricht. Ebenfalls in den Bereich der Datenbankarchivierung fallen die traditionellen Formen der Archivierung von Listen und anderen Ausgabeformaten. Hierfür gibt es zahlreiche Lösungsansätze, die unter Umständen, z.B. bei Archivierung steuerrelevanter Daten aus Compliance-Gründen, ausreichend sind. Datenstrukturen werden mitarchiviert und ein Suchen in den Tabellen ist somit gegeben.
Web-Archivierung
Mit der zunehmenden Abwicklung der Geschäftstätigkeit über das Web und der Flut von Web-Inhalten kommt der Web-Archivierung eine immer wichtigere Bedeutung zu. Hier sind jedoch auch sehr differente Anforderungen zu unterscheiden. Ein Anwendungsfeld ist die Archivierung von Web-Transaktionen wo es um die Nachweise geht, wer wann welche Bestellung abgeschickt hat und was er zu diesem Zeitpunkt auf seinem Bildschirm hatte. Komplizierter wird es schon, wenn mediale Inhalte oder Web-2.0-Funktionalität wie Foren, RSS, Blogs, Wikis, Video-Konferenzen usw. archiviert werden sollen. Vieles lässt sich nicht mehr auf Einzeldokumente reduzieren, wie dies bei der Archivierung von Transaktionen noch möglich ist. Es gilt komplexe Objekte und Strukturen mit ihrem Kontext zu

archivieren. Dabei sind häufig die Inhalte gar nicht mehr Bestandteil der jeweiligen Webseite sondern werden aus anderen Systemen geholt, generiert oder verlinkt. Eine weitere Dimension kommt hinzu, wenn es gilt ganze Webpräsenzen lauffähig zu archivieren. Hier sind besonders Personalisierung und Individualisierung kaum umsetzbar. Nur sehr wenige Softwareprodukte sind in der Lage, dieses Problem zu lösen. Angesichts der stürmischen Entwicklung und immer größerer Virtualisierung steigen hier die Anforderungen an die Archivierung weiterhin und universelle Lösungen für alle Archivierungsprobleme sind nicht in Sicht. Das Universal-Archiv bleibt so immer noch eine Zukunftsvision.

Cloud

Bei der Art, wie und wo man archiviert, gibt es inzwischen zahlreiche Alternativen. Produktiv-Archiv und Sicherheitsarchiv im eigenen Haus, das produktive System selbst im Griff, das Sicherheitsarchiv outgesourct, oder gleiche beide notwendigen Archive in der Cloud.

	Inhouse		Outhouse	
Eigen-Anwender	Archiv operativ	Archiv Sicherung		
Eigen-Anwender mit outgesourctem Sicherheits-archiv	Archiv operativ		Archiv Sicherung	
SaaS & Outsourcing			Archiv operativ	Archiv Sicherung

Die Archivierung ist für die Anbieter von Cloud-Storage und SaaS-Software-as-a-Service-Archivierung von großem Interesse, da sie eine enorme Kundenbindung mit wachsender Informationsmenge bewirkt. Allerdings ist das Thema Archivierung in der Cloud in Deutschland immer noch von Sicherheitsbedenken überschattet. Vorgaben, dass die Daten in Deutschland physisch liegen müssen, lassen sich auch nicht bei allen virtualisierten, global verteilten Cloud.-Angeboten umsetzen. Neben den psychologischen und rechtlichen Argumenten gibt es auch einige technische Restriktionen wie Bandbreite und Zugriffssicherheit. Ein weiteres Problem ist die Integration in die im Hause installierten Softwarewelten. Cloud funktioniert in sich geschlossen sehr gut aber es fehlen die Schnittstellen zu Integration mit lokal installierten Individualanwendungen. Als positiv ist zu sehen, dass sich die meisten Cloud-Archive über verschiedenste Plattformen synchronisieren können, auf unterschiedlichen Endgeräten nutzbar sind, sehr kostengünstig angeboten und über sehr hohe 24/7-Verfügbarkeitsstandards verfügen, die sich Inhouse nur mit großem

© PROJECT CONSULT GmbH 2012

Aufwand realisieren lassen. Zumindest die Sicherheitsarchive wandern inzwischen zunehmen in die Wolke. Auch das Angebot an speziellen Archivierungslösungen für E-Mail, Sharepoint und andere Anwendungsfälle wächst. Für den Privatmann ist heute die (mehrfache!) Archivierung in der Cloud schon fast der Regelfall da einfach, immer verfügbar, keine Betrieb und zu günstigen Preisen nutzbar.

Exotica

Neben den beschriebenen Herausforderungen Web-Archivierung, Datenbankarchivierung und Archivierung in der Cloud gibt es noch eine Reihe von Exotica bei der elektronischen Archivierung. Dies sind zum Teil sehr spezielle Objekte wie begehbare mehrdimensionale Räume und von innen besichtigbare, sich bewegende virtuelle Maschinen. Eine besondere Gruppe von Archivanwendungen nur in Deutschland beschäftigt sich (oder muss sich beschäftigen) mit elektronischen Signaturen. Diese einfach beim Entstehen oder Empfangen zu prüfen und dann im revisionssicheren Archiv „einzufrieren“ mit allen ihren Komponenten – das ist Standard. Aber in Deutschland wird an verschiedenen Stellen die Messlatte höher gelegt. Da sind einmal die nur in geschlossenen Systemen rechtskräftig existierenden De-Mails und E-Postbriefe. Will man diese „rechtsverlustfrei“ in eigene, damit externe Archive umlagern, gibt es Probleme. Die Prüfbarkeit und damit die sogenannte „Rechtssicherheit“ gehen im externen System verloren. Also wird es auch innerhalb der Systeme eigene Archive geben wie den De-Mail-Safe und als Anwender muss man dann halt wieder wissen, wo eigentlich die gesuchte Information ist – im De-Mail, im E-Postbrief, im E-Mail-Archiv-, im ERP-Archiv, im Sharepoint-Archiv usw. Die Vision eines übergreifend nutzbaren Universal-Archives wird hier unterlaufen. Noch komplexer wird es, wenn man an den Sonderfall der Archivierung elektronisch signierter Dokumente denkt, die über ihre Lebenszeit nachsigniert werden müssen, um die „Beweiskraft“ zu erhalten. Dieses sehr proprietäre Verfahren, dass es nur in Deutschland in dieser Form und mit diesem rechtlichen Anspruch gibt, erfordert das Bearbeiten der Datenbank mit Übersignieren oder gar das Übersignieren der Objekte mit anschließender Neuarchivierung. Damit werden die Prinzipien der Archivierung ad absurdum geführt. Wenn das Archivsystem nachweisen kann, dass seit der Speicherung keine Veränderung möglich war und auch keine Veränderung erfolgt ist, kann man sich das nachsignieren sparen. Dieses Modell der revisionssicheren Archivierung hat sich aber bei den Protagonisten von Signatur und Nachsignatur nicht durchsetzen können. Auch hier entfernt man sich von der Idee des stabilen Universal-Archivs da nun unterschiedliche Qualitäten in einem Archiv liegen und Teile der Archivverwaltung durch das Nachsignieren ständig manipuliert werden.

Ein leicht übersehenes Grundprinzip der Archivierung: Migration

Zum Abschluss der Betrachtung des Themas elektronische Archivierung soll noch einmal der Aspekt der Langzeitverfügbarkeit beleuchtet werden. Technik ändert sich schnell. Bei Archivsystemen muss man für Jahrzehnte, vielleicht Jahrhunderte

© PROJECT CONSULT GmbH 2012

planen. Dies heißt, bereits bei der Installation eines neuen Archivsystems ist bekannt, dass Hardware und Software bereits während der Aufbewahrungsfristen nicht mehr zur Verfügung stehen könnten. Daher gehört die Migration bereits in die Planungs- und Testphase eines neuen Archives. Archivsysteme sollten so angelegt sein, dass größere Datenmengen nicht auf einen Schlag umkopiert werden müssen, dass Kontext, Logik und Struktur verlustfrei in andere Umgebungen übertragen werden können, und dass dies in einem Tool-basierten, kontrollierten und dokumentierten Prozess erfolgt. Häufig ist es auch für den Betrieb eines elektronischen Archivsystems wirtschaftlicher, nach einigen Jahren die noch benötigten Informationen auf neue Medien oder andere Speichersysteme umzukopieren. Dabei sind zu unterschieden: weiche Migration, bei der nur die Verwaltungsinformationen migriert werden und die Dokumente unberührt bleiben; harte Migration, bei der Datenbank, Anwendung und Daten migriert werden; und die integrative Migration, in der ältere Systemkomponenten in eine neue Umgebung eingebunden und langsam ausgealtert werden. Migrationen sind aktiv zu planen, Änderungen in der IT-Landschaft sind auf Auswirkungen auf die Archive zu prüfen, und Anpassungen sollten vorgenommen werden, bevor Komponenten aus der Wartung fallen oder abgekündigt werden.

Electronic Archives are the Memory of the Information Society

Dieses Zitat von Erkki Liikanen, dem ehemaligen EU-Kommissars für die Informationsgesellschaft, aus dem Jahr 1999 zeigt die Aufgabe aber auch die Herausforderung und Problematik, der wir uns derzeit gegenübersehen. Das Gedächtnis der Informationsgesellschaft ist ungeordnet, überfrachtet und zeigt erste Ausfallerscheinungen. Einerseits werden wir von der „Information Flood“ überrollt, andererseits tut sich aber ein immer größer werdendes „Information Gap“ nicht mehr verfügbarer oder auswertbarer elektronischer Information auf. Gleiches gilt für die Archive in den Unternehmen, Verwaltungen und anderen Organisationen. Der Wert von Information für Unternehmen, Verwaltungen und Gesellschaft ist immer noch nicht richtig erkannt. Elektronische Archivierung ist nicht nur aus historischen oder Compliance-Gründen wichtig sondern stellt die Wissensbasis für die Unternehmen bereit. Gerade im Zeitalter der Veränderungen, des demographischen Wandels, wo ein Großteil des Wissens die Unternehmen verlässt, sind funktionierende Archive als Wissensbasis wichtig. Archivierung wird unterschätzt. Archivierung gehört nicht in den „Keller“. Archivierung ist ein strategisches Thema, dass in die Governance- und Geschäftstätigkeit intergiert sein muss. Ohne funktionierende Archive wird weder das Gedächtnis der Welt noch das Gedächtnis der Unternehmen funktionieren.

© PROJECT CONSULT GmbH 2012

Über den Autor

Dr. Ulrich Kampffmeyer ist Gründer und Geschäftsführer der PROJECT CONSULT Unternehmensberatung GmbH.
Er berät Kunden aller Branchen im In- und Ausland bei Strategie, Konzeption, Einführung, Ausbau, Migration und Dokumentation von Informationsmanagement-Lösungen (wie Records Management, Enterprise Content Management, Information Lifecycle Management, Wissensmanagement, Dokumentenmanagement, Archivierung etc.)
Von Fachzeitschriften wurde er zu den 100 wichtigsten IT-Machern Deutschlands gezählt. Er gilt als der Mentor der Dokumentenmanagement- und ECM-Branche in Europa.
Er beteiligt sich an der internationalen Standardisierung wie MoReq2 und ist als Kongressleiter, Referent, Moderator und Autor („Codes of Best Practice zur elektronischen Archivierung“, Bücher: „Dokumentenmanagement – Grundlagen und Zukunft“; „Dokumenten-Technologien: Wohin geht die Reise?“ und „ECM Enterprise Content Management") über die Grenzen Europas hinaus bekannt.

© PROJECT CONSULT GmbH 2012