

# Elektronische Archivierung von Website-Inhalten

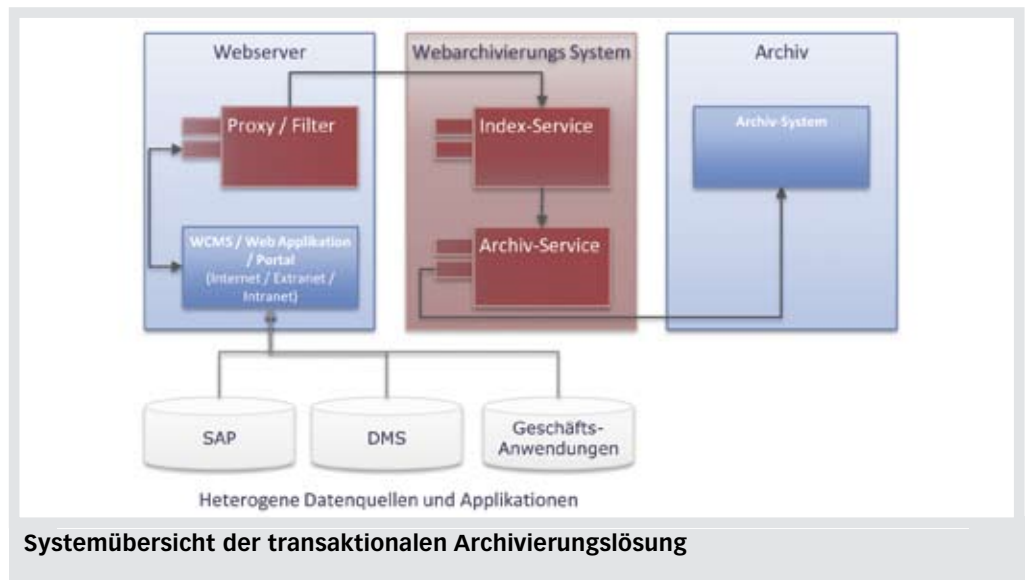
Für Inhalte, die im Web publiziert werden, gelten dieselben regulatorischen Archivierungsanforderungen wie für traditionelle Dokumente. In der Praxis werden diese Regelungen noch kaum umgesetzt. Drei Lösungsansätze behandeln die Thematik, wobei sich ein klarer Favorit herauskristallisiert hat. *Jürg Truniger*

Das Thema Website-Archivierung hat in Deutschland in den letzten Monaten eine regelrechte Kontroverse ausgelöst. Ein im Oktober 2008 durch die Nationalbibliothek erlassenes Gesetz verpflichtet die Website-Betreiber zur Ablieferung ihrer öffentlichen Netzpublikationen. Viele Betroffene üben daran massive Kritik, vor allem wegen des entstehenden Aufwands sowie der angeordneten Geldbussen. Die Kritik, die dieser Gesetzesartikel in Deutschland auslöste, darf nicht darüber hinwegtäuschen, dass Informationen, die auf Websites publiziert sind, grundsätzlich denselben Aufbewahrungsrichtlinien unterstehen wie Printmaterialien oder elektronische Office-Dokumente.

Auch Schweizer Organisationen sind dazu verpflichtet, die regulatorischen Richtlinien bezüglich Compliance und Archivierung zu erfüllen. Auf Websites veröffentlichte Informationen sind darin eingeschlossen. Die Regulierungen umfassen sowohl branchenübergreifende (z.B. die Geschäftsbücherverordnung) wie branchenspezifische Richtlinien (z.B. Anforderungen der FDA oder aus SOX), aber auch interne Anweisungen aus einem Managementsystem oder Qualitätshandbuch. Zudem können Bestimmungen zur Überlieferungssicherung beziehungsweise Bewahrung des firmenhistorischen Interesses relevant sein. Das Thema Website-Archivierung ist besonders bei Organisationen von Belang, die einer hohen Regulierungsdichte ausgesetzt sind, wie Finanz- und Pharmaunternehmen oder behördliche Institutionen.

## Webinhalte und ihre speziellen Herausforderungen

Papierdokumente sind formal in sich abgeschlossene, statische Dokumente. Bei Webpages hingegen handelt es sich um dynamisch erstellte, komplexe Konstrukte, die aus unterschiedlichen Elementen bestehen (HTML-Code, Stylesheets, Javascript usw.) und aus verschiedenartigen Quellen kombiniert werden. Webdokumente lösen den Zusammenhang zwischen Inhalt, Struktur und Format bewusst auf. Sie lassen sich auf un-



terschiedliche Arten aufbereiten, zum Beispiel für einen speziellen Browsertyp oder ein Smartphone, und können mit geringem Aufwand personalisiert werden. Eine zusätzliche Herausforderung bezüglich Records Management ist die gezielte Anreicherung mit Metadaten. Dies betrifft im Besonderen Attribute, über die eine Zuordnung zu einem Geschäftsfall oder Dossier hergestellt werden kann.

Auf Websites publizierte Informationen weisen eine steigende Bedeutung und Verbreitung auf. Aufgrund von Effizienz- und Kostenoptimierungsüberlegungen verlagern Organisationen Informationen ins Internet und bauen dazu heterogene Systemlandschaften auf. Die Webpräsenz einer Organisation besteht in der Regel aus unterschiedlichen Plattformen. Neben einer Corporate Website zählen dazu Intranet und Extranet, Portale, E-Shops und weitere Webapplikationen. Es ist durchaus üblich, dass Teile davon von einem Outsourcing-Partner betrieben werden.

Die Herausforderungen an die regelkonforme Archivierung, die durch die dynamische Informationsaufbereitung sowie die heterogenen Systemlandschaften entstehen, sind erheblich und bei der Gestaltung einer Lösung von zentraler Bedeutung.

## Drei unterschiedliche Ansätze – ein klarer Favorit

In der Praxis werden drei unterschiedliche technologische Ansätze zur Archivierung von Webinhalten unterschieden.

### • Serverseitige Archivierung

Die heutigen modernen Web-Content-Management-Systeme (WCMS) bieten die Möglichkeit, Inhalte zu versionieren. Eine regulatorisch konforme Verwaltung der Informationen wird damit aber nicht erreicht. Zur Archivierung dient ein kontinuierliches Back-up der Datenbank, die vom WCMS genutzt wird beziehungsweise von dessen Repository. Dieser Vorgang wird als «serverseitige Archivierung» bezeichnet.



**Jürg Truniger**  
ist Senior Consultant bei der Unic AG. Als Spezialist für ECM verfügt er über langjährige Projekterfahrung mit Web Content Management, Records Management und Archivierungssystemen.  
[juerg.truniger@unic.com](mailto:juerg.truniger@unic.com)

► Was als einfache Lösung erscheint, erweist sich bei weiterer Abklärung als enorm aufwendig, vor allem wenn die Langzeitarchivierung der Webinhalte zu gewährleisten ist. Die verschiedenen WCMS verwalten ihre Daten auf unterschiedliche, meist proprietäre Art. Dies führt dazu, dass bei einer Migration auf ein neues System – teilweise auch bei Release-wechseln derselben WCMS-Software – die alte Legacy-WCMS-Infrastruktur weiterbetrieben werden muss. Nur so lassen sich die Datenbank-Records wieder als Webseiten aufbereiten. Zudem können Webapplikationen und dynamisch generierte Inhalte aus Drittsystemen auf diese Weise nicht archiviert werden.

#### • Clientseitige Archivierung

Eine vom Ansatz her attraktive Methode ist der Einsatz von Crawler-Technologien. Die Inhalte werden dabei von «extern» auf sehr ähnliche Weise abgefragt, wie dies auch die Endnutzer oder Kunden tun. Deswegen wird diese Archivierungsvariante auch «clientseitig» genannt. Die so abgerufenen Inhalte lassen sich zu einem «Snapshot» zusammenfassen und in jedes vorhandene elektronische Archiv beziehungsweise Records-Management-System schreiben.

Obwohl diese Methode der serverseitigen Archivierung grundsätzlich überlegen ist, existieren auch hier relevante Nachteile. Wenn Webpages während zweier Crawler-Durchgänge mehrfach geändert werden, lassen sich die Modifikationen nicht mehr lückenlos nachvollziehen. Bei sehr umfassenden Websites kann es durchaus sein, dass ein Crawler-Vorgang einige Zeit in Anspruch nimmt. Je länger dies dauert, desto grösser ist die Wahrscheinlichkeit einer mehrfachen Editierung einer einzelnen Seite. Zudem sind Crawler aufgrund nicht durchgängiger Verlinkung und insbesondere auch von Login-geschützten Bereichen nicht in der Lage, 100 Prozent der Inhalte zu durchlaufen. Das grösste Problem entsteht allerdings dadurch, dass applikatorische Elemente (wie Shops, Formulare, Personalisierung) aufgrund der theoretisch möglichen Menge an Parameterkombinationen durch Crawler nicht erfasst werden können.

Ein aus Sicht des Records Management weiterer relevanter Nachteil ist nur begrenzt mögliche Vergabe von Metainformationen.

#### • Transaktionale Archivierung

Der dritte Ansatz wird als transaktionale Archivierung bezeichnet. Analog der clientseitigen Methode verfolgt sie das Ziel, Webinhalte exakt so zu archivieren, wie diese dem Endnutzer präsentiert werden. Erreicht wird dies durch die Aufzeichnung und Archivie-

rung des gesamten Datenstroms (der Transaktionen) zwischen Webserver und Webbrowser. Dazu wird ein eigens dafür vorgesehener Proxy oder Filter eingesetzt.

Nur die transaktionale Archivierung ist in der Lage, Lückenlosigkeit zu gewährleisten. Obwohl ihre Mächtigkeit in der Fachwelt grundsätzlich unbestritten ist, wurde ihr gegenüber immer wieder Kritik laut, vor allem bezüglich des vermeintlich unkontrollierbaren Wachstums der Archivdaten. Hierbei wurde mittlerweile ein fundamentaler Durchbruch erreicht: Ein filigran einstellbarer Indexdienst bezieht die vom Proxy aufbereiteten Daten und verarbeitet diese auf intelligente, auf Wunsch zeitlich asynchrone Art. Auf Basis des Modifikationsdatums und mittels Prüfsummenchecks testet er, ob eine Webseite (HTML-Seite inklusive Bildern usw.) bereits im Archiv abgelegt ist und fügt lediglich dann eine neue Version hinzu, wenn eine relevante inhaltliche Änderung festgestellt wird. Der Indexdienst ermöglicht zudem die gezielte Anwendung von definierten Archivierungs- und Records-Management-Richtlinien. Damit wird erreicht, dass nicht sämtliche auf einer Website verfügbaren Inhalte aufbewahrt werden müssen, sondern nur diejenigen, die gemäss den existierenden Richtlinien entsprechend zu behandeln sind. Erfolgen kann dies auf unterschiedliche Weise:

- Archivierung von spezifischen Navigationsbereichen: In der Praxis sind aufbewahrungspflichtige Inhalte mit hoher Wahrscheinlichkeit innerhalb dedizierter Bereiche der Site-Navigation aufgeführt, zum Beispiel unter «Investor Relations». Die transaktionale Archivierung kann gezielt auf die relevanten Website-Bereiche angewandt werden.
- Festlegung archivierungsrelevanter Bereiche auf einer einzigen Webseite: Informationen auf Webpages lassen sich gezielt «ausschneiden» und sind somit von der Archivierung ausgenommen. Es kann durchaus sinnvoll sein, auf einer Webseite aufgeführte Werbe-Teaser nicht mit dieser ins Archiv abzulegen, wenn diese inhaltlich nicht miteinander verknüpft sind. Dasselbe gilt für rein formatoptimierte Aufbereitungen von Informationen wie die druckoptimierte Ausgabe.

Der Indexservice bietet eine frei definierbare Anreicherung der Webinhalte mit Metadaten. Eine automatische Klassifizierung wie die Vergabe einer Records-Management-Klasse oder die automatische Zuordnung eines Webinhalts zu einem Master-Dossier oder Ordnungssystem (z.B. GEVER) lässt sich damit gewährleisten. Sowohl Proxy- und Indexservice lassen

sich unabhängig vom eingesetzten Web-Content-Management-, E-Shop- oder Portalsystem auf den Webservern installieren. Die Kommunikation mit dem zentralen Archiv oder Records-Management-System erfolgt mittels verschiedener Schnittstellentechnologien wie Webservices oder direkte Integration in Java oder .Net. Während das Archiv im Normalfall innerhalb der Firmen-Firewall steht, können die Webserver «irgendwo» im Netz positioniert sein. Eine aus verschiedenen (also auch ausgelagerten) Systemen bestehende Webpräsenz einer Organisation lässt sich dadurch in einem einzigen elektronischen Archiv beziehungsweise Records-Management-System verwalten. Dessen Funktionalitäten (wie Versionierung, Erstellung von Langzeit-Repräsentationen wie PDF/A, die Gewährleistung von Authentizitäts- und Integritätsanforderungen) stehen vollumfänglich zur Verfügung. Damit ist die transaktionale Methode der Website-Archivierung in der Lage, die regulatorischen Anforderungen bezüglich Datenaufbewahrung umfassend und auf kosteneffiziente Weise abzudecken.

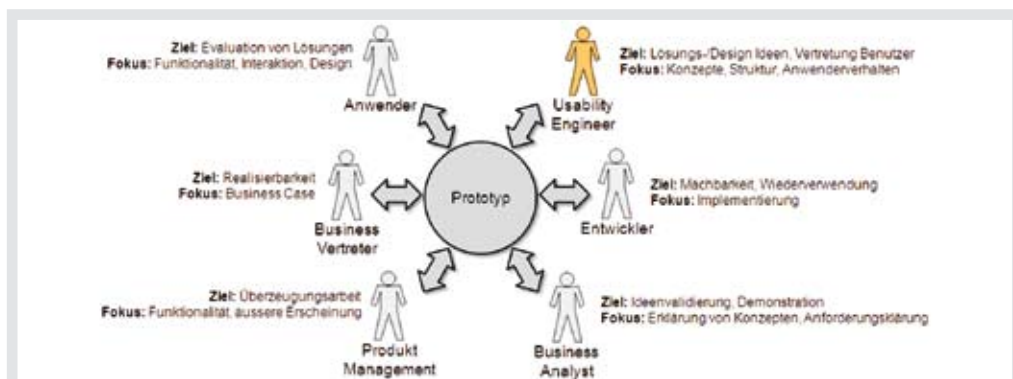
#### Website-Archivierung in der Praxis

Archivierungsaufgaben sind nie allein durch den Einsatz technologischer Mittel lösbar. Die wesentlichen Erfolgsfaktoren sind organisatorischer und prozessualer Natur. Ein Vorhaben zur Archivierung von Website-Inhalten startet am besten mit einem Review der in einer Organisation existierenden Records Management Policy beziehungsweise der Weisung zur Archivierung von Informationen. Das Ziel dieses Prüfungsvorgangs ist es, die Inhalte zu identifizieren, bei denen eine Publikation auf Webplattformen möglich ist. Anhand von unternehmensspezifischen, aber auch generischen Bewertungskriterien (z.B. die Klärung der Frage, bei welchen dieser Inhalte das Web das primäre Publikationsmedium darstellt) wird eine Richtlinie zur Behandlung von Webinhalten erstellt. Darin sind auch für die Webredaktoren relevanten prozessualen Elemente beschrieben.

Unter Anwendung dieser Richtlinien wird eine gezielte Klassifizierung der Inhalte bezüglich ihrer Archivierungsrelevanz sowie der zur Anwendung gelangenden Records-Management-Klassen durchgeführt. Dies kann sowohl bei bereits bestehenden wie auch bei erst zu entwickelnden Websites durchgeführt werden. Die im Einsatz stehende transaktionale Website-Archivierungslösung sowie das elektronische Archiv lassen sich im Anschluss exakt gemäss diesen Richtlinien und Klassifikationen konfigurieren. Dadurch wird der Prozess abgeschlossen und die regulatorisch konforme Behandlung von Website-Inhalten ist erreicht.

# Schräge Prozesse, eitle Ingenieure, dumme Nutzer?

Usability von Enterprise-Content-Management-Lösungen wird oft in Sales-Präsentationen als wichtige Produkteigenschaft betont. Demgegenüber stehen häufige Klagen von Anwendern über merkwürdige Arbeitsabläufe oder Funktionen. Ist Usability tatsächlich nur der Puderzucker für schöne GUIs? *Christoph Kunz*



**Use-Case-Prototypen zeigen schnell Lösungsvarianten auf und dienen als gemeinsame Sprache.**

Während sogenannter Contextual Inquiries – Prozess- und Aufgabenanalysen durch Beobachtung und Interviews von Anwendern im Arbeitskontext – wurde Folgendes beobachtet: Andi Assistent soll die Stammdaten von Frau Schweizer ändern, die umgezogen ist. Dazu füllt er ein elektronisches Änderungsformular aus, das per E-Mail an die zentrale Mutationsstelle geschickt wird. Im tatsächlichen, beobachteten Arbeitsfluss nutzt Andi Assistent einen Formular-Builder und druckt das ausgefüllte Formular aus. Den Ausdruck legt er in den Scanner und schickt sich den Scan als E-Mail an sich selbst. Zurück am Arbeitsplatz leitet er die E-Mail an die Mutationsstelle weiter. Die Begründung: Diese Arbeitsweise sei einfacher, als das Formular als PDF abzuspeichern, eine neue E-Mail zu erstellen, das PDF im Dateisystem wiederzufinden und es schliesslich via Drag&Drop in die E-Mail zu ziehen. Eine simple Funktion «Senden als E-Mail» im Formular-Builder würde den Arbeitsfluss deutlich effizienter gestalten und zusätzlich Papierkosten sparen. Die im PDF-Reader äquivalente Funktion wurde übersehen.

## Wie entsteht Usability?

Die genannte Problematik entsteht durch gar keinen oder durch zu späten Einbezug der Anwender in den Entwicklungsprozess. Modernes Usability Engineering setzt genau hier an und übersetzt die Sprache des Anwenders und des Business in die Sprache der IT-Entwick-

lung. Usability Engineering ist eine ingenieurmässige, planvolle und methodische Tätigkeit! Diese bestimmt vor der eigentlichen Entwicklung den Umfang und die Eigenschaften eines Produkts in der Art, dass dieses im gegebenen Kostenrahmen optimal in spätere Arbeitsweisen eingebettet ist. Dazu werden nicht nur die Bedürfnisse der Anwender (im Gegensatz zu explizit geäusserten Wünschen) identifiziert, sondern diese auf die Businessziele und auf Entwickelbarkeit der Produkteigenschaften abgestimmt. Usability entsteht nicht zufällig, sondern ist vielmehr Ergebnis eines systematischen Prozesses. Dieser beruht im Wesentlichen auf zwei Prinzipien.

- Iterative Lösungsfindung: Nicht immer trifft der erste Schuss ins Schwarze. Auch mittels aufwendigem Requirements Engineering werden manche Anforderungen vergessen oder ungenügend spezifiziert. Agile Methoden wie Use Case Prototyping präzisieren die Anforderungsspezifikationen und visualisieren Lösungsvarianten.
- Einbezug des Anwenders: Interaktive Produkte und Softwarewerkzeuge werden nicht zum Selbstzweck entwickelt, sondern sollen Kunden und Anwender bei der Erfüllung von Aufgaben unterstützen. Wichtige Fragestellungen in der Lösungsfindung müssen daher unter Einbezug der Anwender beantwortet werden.

Durch diesen Engineering-Prozess lassen sich nicht nur fehlende, sondern auch über-

flüssige Produkteigenschaften identifizieren. Während fehlende Funktionen lästig sind, können sie mittels Workarounds ausgeglichen werden. Ungenutzte Funktionen sind verschwendetes Geld!

## Use Case Prototyping

Die Effektivität eines Produkts, das heisst das Vorhandensein der «richtigen» Eigenschaften hängt stark von einer guten Anforderungsanalyse ab. Die Anwendung der oben genannten Prinzipien fällt im traditionellen Requirements Engineering nicht immer leicht, da hier das Entwicklungsziel in möglichst formalen Anforderungen besteht. Diese sind jedoch für Anwender und Businessauftraggeber oft schwer zu verstehen und daher zu bewerten.

Diese Lücke wird durch Use Case Prototyping geschlossen. Dies bedeutet nicht das teure Implementieren eines Softwareprototypen, sondern das schnelle Kreieren und Verwerfen von verschiedenen visuellen Lösungsvarianten. Die Wiedergabetreue solcher Varianten reicht von Papierprototypen bis hin zu voll interaktiven GUIs, die mit speziellen Prototyping-Tools erstellt werden. Der niedrige Abstraktionsgrad solcher Skizzen erleichtert die Diskussion der Lösungsvarianten mit allen Projekt-Stakeholdern, insbesondere mit den Auftraggebern und Anwendern. Durch Prototyping wird sowohl die Qualität der Anforderung selbst als auch die Geschwindigkeit der Anforderungsanalyse und der Abstimmung über Anforderungen deutlich gesteigert.

Erfüllbar wird der Anspruch guter Benutzbarkeit durch die richtige und vor allem projektangepasste Anwendung von Usability Engineering. Zusätzlich werden Projektkosten und -risiken minimiert.



**Dr. Christoph Kunz** ist Projektleiter und Usability Experte bei der Business Solution Group Technology Innovation.

# Erfolgreiches Marketing durch individuelle Kundenkommunikation

Die Kommunikation mit Kunden hat sich in den letzten Jahren drastisch verändert. Der Kunde erwartet von Unternehmen über verschiedene Kommunikationskanäle und -formen informiert zu werden. Dazu gehören unter anderem E-Mail, SMS, Mailings oder klassische Briefe mit personalisierten Inhalten. *Achim Schuch*

Standardisierte Produkte wie Flyer wandern heute immer öfter ungelesen in den Papierkorb. Deshalb setzt sich die individuelle Kundenkommunikation zunehmend durch und führt Analysten zufolge zu einer bis zu 500 Prozent höheren Response-Rate. Es ist notwendig, dass die Unternehmen das oftmals teuer erworbene Wissen über den Kunden anhand von definierten Prozessen umsetzen und verwerten.

## Perfekte Zusammenarbeit aller Systeme

Eine regelmässige Kommunikation mit den Kunden findet in der Regel über Dokumente wie Kontoauszüge oder Rechnungen statt. Die Kunden erwarten und fordern zu Recht einen besseren Service, mehr Aufmerksamkeit und sofortige Reaktionen auf Anfragen. In geografisch verteilten Märkten tätige Unternehmen benötigen den Zugriff auf marktrelevante und der Compliance entsprechende Dokumente aus Zweigstellen, von Vertretern oder Wiederverkäufern und dies nicht nur aus dem lokalen Netzwerk. Diese Problemstellung ist nur zu lösen, wenn die involvierten Systeme vom Dokumenteneingang via Post, E-Mail oder dem Internetportal, über das Dokumentenmanagement

bis hin zur automatisierten Erzeugung der Dokumente (Content Publishing) perfekt zusammenarbeiten. Dies ist entscheidend für das Marketing und damit für die effiziente und einfache Verbreitung von Botschaften.

## Dynamische Lösungen nutzen variable Daten

Dynamische Content-Publishing-Lösungen ermöglichen druckreife und sofort nutzbare Vorlagen für die Produktion von massgeschneiderten Begleitmaterialien. Dazu gehören Angebote, Newsletter, personalisierte HTML-Mails und personalisierte Internetseiten für das Web. Diese dynamischen Content-Publishing-Lösungen nutzen variable Daten und Logik bei der Zusammenstellung und Gestaltung dieser Dokumente. Andere wichtige Marketinganwendungen für die Bereitstellung als Printmaterial, per E-Mail und im Web können ebenfalls schnell und effektiv erstellt werden.

Content-Publishing-Lösungen müssen die Möglichkeit bieten, schnell auf Veränderungen am Markt reagieren zu können. Ein proprietärer Design-Client führt zu hohem Aufwand an Schulung, Konvertierung von Dokumenten inklusive entsprechender Layoutänderungen und längeren Laufzeiten. Es kann Tage oder gar Wochen dauern, bis eine Botschaft die Kunden erreicht hat oder eine neue Corporate-Identity-Richtlinie durchgesetzt ist, müssen doch Änderungen über die IT-Abteilung implementiert werden. Der optimale Zeitpunkt ist schnell verpasst und man ist zu spät. Deshalb lohnt es sich bereits in der Projektphase darauf zu achten, dass Standardprodukte wie Microsoft Word, Adobe Indesign oder Adobe Dreamweaver eingesetzt werden, um automatisch generierte Dokumente zu gestalten.

## Individuelle Kundenkommunikation erhöht Erfolgchancen

Die individuelle Kundenkommunikation hat aus den erwähnten Vorteilen innerhalb der Unternehmen auch ausserhalb der IT-Abteilungen an Bedeutung gewonnen. Nicht mehr

nur die IT-Verantwortlichen, sondern auch auf den Chefetagen und in den Marketingabteilungen nimmt sie heute einen hohen Stellenwert ein. Doch welches sind die Kriterien für die Auswahl eines entsprechenden Dokumenten-Management-Systems?

Die Skalierbarkeit und Ausfallsicherheit ist ebenso von Bedeutung wie die Bereitstellung von Diensten über beliebige Netzwerke (serviceorientierte Architektur). Dokumente werden nicht mehr nur für den Massendruck am Abend oder in der Nacht erzeugt (Batch), vielmehr ist die Bereitstellung dynamischer Dokumente «on-the-fly» eine immer wichtigere Anforderung an die Systeme (Realtime oder Near-time).

Für welchen Kanal das Dokument verwendet werden wird, muss zum Zeitpunkt der Erstellung noch nicht definiert sein. Deshalb muss ein modernes System in der Lage sein, alle Ausgabekanäle sowohl in Batch als auch On demand aus derselben Vorlage heraus zu bedienen.

## Realtime-Erzeugung von Dokumenten stellt hohe Anforderungen

Besonders die Realtime-Erzeugung von Dokumenten stellt an viele Systeme grosse Anforderungen. Diese Funktionalität wurde oft nachträglich implementiert, was zur Folge hatte, dass häufig Performance-Engpässe entstanden.

Im Interesse der Unternehmen lohnt es sich, bei der Auswahl des Systems auf einen global aufgestellten Anbieter zu setzen. Dokumenten-Management-Systeme müssen sich möglichst nahtlos in bestehende und zukünftige Infrastrukturen einpassen.



**Achim Schuch**  
ist Account Director bei EMC Document Sciences.

## ECM-Forum

Das kommende **ECM-Forum** am Donnerstag, **30. April 2009** im **Kongresshaus Zürich**, steht unter dem Fokusthema: «**Innovationen für ein nachhaltiges Enterprise Information Management – die Verbindung von Dokumenten, Content- und Wissensmanagement**».

Themen wie Business Process Management, Collaboration, Dokumenten Lifecycle Management, Technische Dokumentation, Records Management werden von Referenten und Gästen kritisch hinterfragt und reflektiert.

Anmeldung unter: [www.ecmforum.ch](http://www.ecmforum.ch)