



# DOK magazin

Technologien, Strategien & Services für das digitale Dokument

„Die Zukunft des Dokuments“ Was ist - im Zeitalter von Enterprise 2.0 - ein Dokument?  
Einführung Dokumenten-Management-Systeme Standortübergreifende Prozessoptimierung  
Semantic Desktop Informationsmanagement trifft Web 2.0



## Web-Harvesting: Webseiten archivieren

Special: Posteingang

# Was ist eigentlich – im Zeitalter von Enterprise 2.0 – ein Dokument?

Dokumenten-Management, Dokumentenformate, Wikis, Blogs, Tags, Compliance, Web-Content-Management, Aktenverwaltung, Geschäftsprozesse

Was ist eigentlich ein Dokument? Eine simple Frage, die wir vor einigen Jahren noch sehr gut beantworten konnten. Heute, im Zeitalter von Enterprise 2.0 mit Wikis, Blogs und großen Mengen an multimedialen Daten, lässt sich diese Frage nicht mehr so einfach beantworten.

## Linear

Doch gehen wir kurz zurück an den Anfang. Ein Dokument ... was sind seine wesentlichen Charakteristika? Es liegt in Papierform vor und kann praktisch von jedem gelesen werden. Die Richtung des Lesens ist klar durch den Kulturkreis definiert. Der Inhalt des Dokuments lässt sich (zumeist) gut beschreiben. Was aufbewahrt werden muss, d.h. das Trägermedium, ist klar definiert. Nun zu digitalen Dokumenten, Version 1.0. Die Kennzeichen für diese Ära sind:

1. Das Dokument wurde elektronisch erstellt.
2. Das „gedankliche“ Zielformat war Papier.

Was ist damit gemeint? Alle erstellten Dokumente, E-Mails, PDFs, Word-Files usw. sind im Prinzip dafür gedacht, zum Lesen ausgedruckt zu werden. Aus Kostengründen und weil moderne Computer kleinere Dokumente gut darstellen können, wurde das vielfach schon vermieden. Niemand druckt mehr E-Mails aus. Die Papiersimulation durch den Computer genügt. Die Vorteile dieser Art von digitalen Dokumenten sind hinlänglich bekannt. Und die Nachteile? Obwohl diese Art von Dokumenten schon seit mehreren Jahrzehnten existiert, sind viele Probleme noch völlig ungelöst. Ein Stichwort hierfür ist die Langzeitarchivierung – mittlerweile kann jeder die Datenbytes gut konservieren, aber was wird gemacht, wenn die Anzeigeprogramme nicht mehr existieren oder auf den aktuellen Betriebssystemen nicht mehr lauffähig sind?

[www.hyperwave.com](http://www.hyperwave.com)

**Peter Luttenberger** ist Director Product Management bei der **Hyperwave AG** in Graz. Das Unternehmen ist Anbieter von Collaborative Information Management (CIM) Lösungen mit einer integrierten E-Learning und Web-Conferencing Umgebung. Hyperwave beschäftigt in Österreich, Deutschland, UK und den USA insgesamt rund 80 Mitarbeiter.

## Nicht linear

Wie sieht das bei den Dokumenten bzw. Dokumenttypen aus, die wir heute verwenden? Wenn man sich die Informationen, die in einem üblichen Intranet abgelegt sind, anschaut, so sind meiner Erfahrung nach bereits mehr als die Hälfte der Informationen nicht in klassischen Word- oder PDF-Dokumenten enthalten, sondern in HTML/XML-Dateien, die die Informationen beinhalten. Diese HTML-Dateien sind nicht mehr linear aufgebaut. Sie bestehen aus Verlinkungen von verschiedenen Informationsquellen. Das bedeutet praktisch, dass ich die Informationen so, wie ich sie gelesen habe, gar nicht mehr ausdrucken kann. Ich habe viele Links verfolgt, mich vom Ausgangspunkt weit entfernt und dabei einen Weg eingeschlagen, der mir als Leser richtig erschien. Wenn ich jetzt gefragt werde, welches Dokument ich gelesen habe, ist es unmöglich, eine genaue Referenz auf ein Dokument oder eine URL anzugeben. Auch den Autoren ist nicht bekannt, in welcher Reihenfolge ich die Informationen konsumiert habe. Ich habe somit das Inhaltsverzeichnis meines virtuellen Buches selbst erstellt. Wie kann der Autor nun ermitteln, welche Teile in welchem Zusammenhang für mich wichtig waren? Vielleicht war es nur ein Link auf ein Video oder einen Podcast.

Die Zusammenfassung, was moderne Dokumente sind, ist die, dass sie erstens nicht (linear) ausgedruckt werden können und zweitens auch beliebige multimediale Inhalte enthalten können. Beispiele für solche Dokumenttypen sind die bekannten Blogs und Wikis. Ein Blog ist ein chronologisches Dokument, es gibt keine Inhaltsstruktur außer dem Datum der Veröffentlichung, und die ist natürlich unabhängig vom Inhalt oder anders formuliert: Aus dem Inhaltsverzeichnis eines Blogs – sprich dem Kalender – kann ich nicht erkennen, worum es inhaltlich in dem Blog geht. Vielfach kann ich als Leser sogar selbst die Inhaltsstruktur erstellen, beispielsweise über die Navigation nach Attributen (Tags).

Auch für diese (modernen) Dokumenttypen sind die Vorteile hinlänglich bekannt. Wie sieht es jedoch mit den Nachteilen aus? Alles, was bei linearen Dokumenten bezüglich Langzeitarchivierung gilt, gilt hier natürlich auch, wobei erschwerend die Vielzahl an verschiedenen Datenformaten hinzukommt. PDF/A, der zurzeit aktuellste Standard für Langzeitarchivierung, deckt diesen Anwendungsfall nicht oder nur sehr bedingt ab. Zusätzlich zu diesen Problemen kommt noch die Frage, wie man Nachvollziehbarkeit (Compliance) sicherstellt, wenn die Informationen nicht in einem oder mehreren Dokumenten enthalten sind, sondern aus beliebig vielen Informationsstücken zusammengesetzt sind und deren Zusammensetzung wiederum beliebig variierbar ist. Wenn ich ein derartiges Dokument „zitieren“ möchte, wie mache ich das? Den Google-Suchbegriff, der verwendet wurde, als Quellenangabe speichern?

## Akten

Unter dem Begriff Akte kann sich jeder wie unter dem Begriff Dokument etwas vorstellen. Akten lassen sich archivieren, elektronische Akten kann man mit Records-Management-Systemen verwalten. Aber wie kann ich nicht-lineare Dokumente archivieren? Wie muss ein dafür passendes Records-Management-System aussehen? Reicht es eigentlich überhaupt noch, einzelne Dokumente als Records abzulegen? Wie ist das mit einer Übersichtsseite, die dem Kunden Preis und Verfügbarkeits-Informationen live aus dem ERP-System, Produktdaten aus dem Informationssystem der Entwicklungsabteilung und Beschreibung von Einsatzszenarien aus dem Marketing-Wiki anzeigt? Vielleicht wird auch ein Demovideo angezeigt, das auf YouTube abgelegt wurde. Wie wird jetzt im etwaigen Streitfall nachgewiesen, was zum Zeitpunkt des Vertragsabschlusses wirklich zu sehen war? Welche Informationen haben sich noch hinter Links zu AGB usw. verborgen? Einfaches Ausdrucken oder Speichern als PDF funktioniert nicht.

Hier gibt es zwei mögliche Lösungsansätze:

1. „Einfrieren“ – dabei werden alle Informationen gespeichert, d. h. es wird eine komplette Kopie der Website einschließlich der verlinkten Informationen erzeugt. Vermutlich nur praktikabel, wenn sich die Inhalte nicht häufig ändern.
2. Komplettes Auditing aller Informationen bei gleichzeitiger Versionsverwaltung aller Daten. Über ein Auswertungstool lässt sich der Zustand zu einem beliebigen Zeitpunkt rekonstruieren.

Außerdem scheint es sinnvoll, nicht mehr von Dokumenten zu sprechen, weil mit diesem Begriff eine klare Vorstellung über die Darstellungsform verbunden ist, und einen derartig fest verankerten Begriff so umzudefinieren, dass er moderne Dokumenttypen inkludiert, ist sicherlich nicht möglich und würde mehr Verwirrung stiften als Erklärung liefern. Nur welcher Begriff umfasst alles? Content?

## Zukunft des Dokuments

Zuerst eine Beschreibung, was heute schon möglich ist, aber noch nicht überall verwendet wird. Zuerst die „intelligenten Dokumente“. Viele Dokumente sind nach wie vor linear, wie dieser Artikel zum Beispiel. Dennoch kann man auch diese alten Dokumenttypen mit „Intelligenz“ bzw. mit Automatismen versehen, die unsere tägliche Arbeit erleichtern. Zum Beispiel: Warum ist es notwendig, wenn ich einen Urlaubsantrag ausfülle, eine Reiserechnung abgebe oder ein Angebot annehme, das zugehörige Dokument an der richtigen Stelle abzugeben? Das Dokument kann doch wissen, dass es eine Reiserechnung

ist, und den entsprechenden Arbeitsablauf anstoßen. Warum muss ich wissen, wer heute im Versand Dienst hat bzw. an welche Versandabteilung ich das Dokument zur Bearbeitung weiterleiten muss. Alle diese Dinge kann ein Dokument entsprechend seiner Dokumentklasse automatisch erledigen. Ähnlich ist es mit der Ablage: Wenn ich einen Vertrag ablege, warum muss ich mir überlegen, in welchem Ordner ich das Dokument ablegen soll?

Das System kann automatisch anhand des Typs und unter Berücksichtigung verschiedener Inhaltselemente den Speicherort auswählen bzw. das Dokument mit Metadaten anreichern, um Compliance-Auflagen zu erfüllen. Das System kann beispielsweise automatisch das aktuell bearbeitete Projekt erkennen – über den aktuellen Aufgabenschritt, für den ich zuständig bin, oder über den E-Mail-Verteiler.

### „Aktive Dokumente“

Diese Dokumente können mir Antworten liefern, wenn ich Fragen zum Inhalt habe. Wer ist der Experte auf diesem Gebiet? Gibt es neuere Dokumente, welche Projekte wurden anhand dieser Beschreibung bisher realisiert usw. Einzelne Datentypen in Dokumenten können automatisch Verknüpfungen herstellen. Eine Telefonnummer könnte als solche erkannt werden und eine Verknüpfung mit dem CRM-Eintrag herstellen – Analoges gilt für Artikel oder Bestellnummern.

### „Lebende Dokumente“

Dokumente müssen nicht mehr statisch sein, sondern können sich automatisch aktualisieren. Diese automatische Aktualisierung kann erstens wie eine Online-Zeitung immer den aktuellen Leitartikel enthalten oder die aktuellen Geschäftszahlen. In diesem Sinn könnte zum Beispiel ein Portal als „lebendes Dokument“ bezeichnet werden oder ein RSS-Feed. Oder der Inhalt des Dokuments könnte abhängig vom Vorwissen des Lesers erstellt werden. Sozusagen „Embedded E-Learning“, die richtige Information für den jeweiligen Leser. Ebenso können Such-Ergebnisse als aktive oder lebende Dokumente bezeichnet werden. Man bekommt keine Liste von Such-Ergebnissen, sondern ein Dokument, das mir personalisiert eine Antwort auf meine Frage liefert. Das Antwortdokument kann dann, wenn die Information nicht ausreicht, weiter befragt werden (aktives Dokument).

### Vorteile der neuen Dokumententypen

1. Reduktion des Suchaufwands. Die meiste Zeit verbringen Mitarbeiter mit der Suche nach den richtigen Informationen. Gesucht werden Informationen, deren Qualität für Entscheidungen ausreicht.
2. Vermeidung von Arbeitsduplizierung. Viele Präsentationen und Dokumente werden doppelt und dreifach erstellt, da die Vergessenskurve in Unternehmen sehr hoch ist bzw. die Daten auf Abteilungs-Shares oder Notebooks einzelner Mitarbeiter liegen. Für unterschiedliche Gruppen werden ähnliche Dokumente erstellt, ohne dass die einzelnen Autoren voneinander wissen.
3. Verknüpfung von zusammenhängenden, aber unterschiedlichen Informationsquellen bzw. neue Daten mit bestehenden verbinden. Vorschrift X ersetzt Vorschrift Y oder Dokument A ergänzt Dokument B.
4. Automatische Klassifizierung. Arbeitsprozesse können automatisch gestartet werden, es gibt keine falsche Zuordnung von Dokumenten zu den falschen Prozessen.
5. Personen mit Dokumenten verbinden. Vermutlich der wichtigste Aspekt überhaupt. Eine der grundlegendsten Informationen über ein Dokument ist die Angabe des Erstellers. Zur besseren Einschätzung der Relevanz bzw. Verlässlichkeit der enthaltenen Informationen sind weiterführende Daten zur Person des Erstellers unerlässlich. Mit den modernen Dokumententypen kann man den Erstellprozess (wer war involviert, warum wurde es erstellt) ermitteln und ebenso alle Prozesse, in denen das Dokument verwendet wurde. Es findet ein zeitversetzter Wissenstransfer statt, der nicht vorgeplant werden muss.
6. Erfassung von Wissen ist ein impliziter Vorgang. Es müssen nicht nachträglich FAQs oder Lerndokumente erstellt werden. Die Informationen aus den Prozessen können direkt weiterverwendet werden bzw. stellen eine unmittelbar verfügbare Informationsquelle dar. Sozusagen ein Abfallprodukt, das bei der Anwendung eines Wissensmanagement-Systems automatisch entsteht.
7. Das Wissensmanagement-System kann systemisch Informationen/Dokumente generieren, die der Anwender zum entsprechenden Zeitpunkt braucht. Vorstellbar ist das nicht nur für die oben erwähnten „lebenden Dokumente“, sondern auch für Hilfe-Informationen oder automatische Aufgabenlisten basierend auf üblichen zusammengehörenden Tätigkeiten. Am ehesten mit dem System vom Amazon „Kunden, die x kauften, haben sich auch für y interessiert“ vergleichbar.

>> weiter auf Seite 26“

8. Optimierung der Geschäftsprozesse: Aktive Dokumente, d. h. die Fragen, die an Dokumente gestellt werden, liefern Informationen über die Verwendung der Dokumente und ermöglichen so gezielt Verbesserungsmaßnahmen. Das geht über die klassische Frage „war diese Information hilfreich?“ weit hinaus, da sowohl die vorhergehenden als auch die nächsten Schritte des Anwenders mitberücksichtigt werden können und der gesamte Kontext der Aufgabe dem System bekannt ist.
9. Anwender-Selbstservice bzw. automatische FAQs: Wenn dem System die bewerteten Antworten bekannt sind, kann es diese Informationen vorab bereitstellen oder die entsprechenden Antworten liefern, ohne dass ein Mitarbeiter eingreifen muss. Nicht vom System beantwortbare Fragen können, mit dem gesamten Kontext, an entsprechende Betreuer zur Beantwortung weitergeleitet werden und stehen dann neuen Anwendern sofort zur Verfügung.
10. Compliance-Anforderungen können leichter erfüllt werden. Sind alle Verwendungsstellen eines Dokuments und deren Zusammenhang bekannt, können Neuerungen (AGB, neuer E-Mail-Footer usw.) entsprechend eingeführt bzw. alle betroffenen Personen und Projekte informiert werden.

## Die Zukunft

Wie geht es weiter? Bill Gates hat vor kurzem in einer Rede an der Carnegie Mellon Universität mit dem Titel „Bill Gates Unplugged: On Software, Innovation, Entrepreneurship and Giving Back“ ([www.cmu.edu/homepage/multimedia/billgates.shtml](http://www.cmu.edu/homepage/multimedia/billgates.shtml)) das Ende der Tastatur angekündigt, und das impliziert nicht nur die Spracheingabe für Word und Suchanfragen, sondern ermöglicht ganz neue Dokumenttypen, die vermutlich dem Menschen weit näher sind als bisherige Dokumente. Die zwischenmenschliche Kommunikation läuft nämlich nicht in Form von Dokumenten ab.

Wenn wir uns unterhalten, geschieht das nicht durch Vorlesen von Texten, sondern wir sind weit interaktiver und berücksichtigen sehr viele Aspekte der Kommunikation. Gesprochene Worte sind nur ein, wenn auch wesentlicher, Teil des Ganzen. Jeder kennt die Gespräche, in denen das Nicht-Gesagte wichtiger ist als das, was explizit gesprochen wurde. Erkennbar ist das entweder am Kontext des Gesprächs oder an der Mimik der Teilnehmer. Um nun alle Aspekte der Kommunikation zu erfassen, muss neben Audio auch das zugehörige Video der Gesprächsgruppe erfasst werden. Ein zurzeit noch nicht gelöstes Problem ist die automatische Indizierung (abgesehen von zeitlichen Markierungen) eines Videos. Experimente wie am Know-Center (Projekt Mistral) mit fixen Meeting-Szenarien mit fest eingestellten Kameras und Mikrofonen haben schon erstaunlich gute Ergebnisse gebracht. Aber um damit ein Gespräch eines Mediziners mit seinem Klienten oder ein Kundengespräch vor Ort aufzunehmen und daraus ein Dokument zu generieren, das dann als nachvollziehbares Beweismaterial im Streitfall herangezogen werden kann, ist noch ein weiter Weg.

Für uns als Lösungsanbieter für Collaborative Information Management stellt die Verbindung von Personen mit Informationen, egal in welcher Form sie heute oder zukünftig vorliegen, unter Einbindung der Geschäftsprozesse das entscheidende Kriterium für den nachhaltigen Erfolg eines Unternehmens dar. Fest steht jedenfalls, dass wir uns vom Begriff Dokument als Beschreibung des Informationsträgers trennen müssen. ■