

## 2 Grundstruktur eines Dokumenten-Management-Systems

### 2.1 Dokumente und Dokumenten-Management

Die Funktionen eines DMS sind wesentlich von der Art und dem Zweck der zu verwaltenden Dokumente bestimmt. Dazu ist es wichtig zu erkennen, welche Parameter ein Dokument beschreiben.

#### 2.1.1 Der Zweck des Dokumentes

Auch wenn eine Vielzahl unterschiedlicher Dokumente existiert, gibt es in Bezug auf den Zweck zwei Grundtypen:

- Das Dokument soll als **Nachweis** einer Tatsache oder potenzielles **Beweismittel** dienen: Typische Beispiele hierfür sind Rechnungen, Verträge, Lieferscheine, Protokolle, Untersuchungsberichte, Gutachten etc. Umgangssprachlich bezeichnet man diese meistens als **Belege**, auch wenn dieser Begriff juristisch so nicht existiert. Es muss nachgewiesen werden können, dass das Dokument nicht verändert wurde, seit es den Status eines »Beleges« erreicht hat. Dabei wird oft auf gesetzliche Regelungen wie »Revisionssicherheit« Bezug genommen. Was das genau bedeutet, wird weiter unten erläutert.
- Das Dokument dient als **Träger von Informationen**: Dies können z. B. sein: Fachaufsätze, Produktblätter, Preislisten, Pflichtenhefte, usw. Die Information im Dokument muss hier möglichst einfach – maschinell – zugänglich sein. Das heißt, man will die Inhalte finden, verändern, kopieren und anderweitig weiterverwenden. Damit es weiter bearbeitbar ist, muss es als CI-Dokument<sup>1</sup> gespeichert werden.

Es gibt natürlich auch Dokumente, die beiden Zwecken dienen: Ein Angebot z. B. ist zum einen ein verbindlicher Beleg, der nicht mehr verändert werden soll. Zum anderen kann ein Angebot auch Vorlage für ein anderes Angebot (z. B. für einen Kunden mit ähnlichen Anforderungen) sein. Dann möchte man es kopieren und

---

1. Coded Information: Ein CI-Dokument wird bearbeitbar gespeichert (z. B. als Word-Datei).

die Kopie inhaltlich anpassen können. Dies kann natürlich auch zur Folge haben, dass es zweimal (veränderbar und nicht veränderbar) gespeichert wird.

### 2.1.2 Erscheinungsformen von elektronischen Dokumenten

Bezogen auf die Art ihres logischen Inhalts müssen zwei grundsätzliche Erscheinungsformen elektronischer Dokumente unterschieden werden: Dokumente, die keine kodierte Informationen enthalten (NCI-Dokumente<sup>2</sup>) und Dokumente, die kodierte, d. h. durch DV-Systeme auswertbare Informationen enthalten (CI-Dokumente).

Die Erscheinungsform ist in der Regel von der ursächlichen Entstehung der Dokumente abhängig. So ist ein gescanntes Papierdokument als NCI-Dokument einzustufen, denn die erzeugten elektronischen Daten repräsentieren lediglich ein fotografisches Abbild des Dokumenteninhalts. Eine Auswertung, z. B. der Textinhalte, ist nicht möglich. Wird hingegen ein vorhandenes NCI-Dokument mit einer Schrifterkennungssoftware (OCR-Software) oder einer Vektorisierungssoftware zur Nachdigitalisierung von Grafiken bearbeitet, so können ein oder mehrere CI-Dokumente erzeugt werden, die parallel zum ursprünglichen NCI-Dokument existieren.

Hingegen handelt es sich beispielsweise bei einem XML-Dokument, welches in einem E-Commerce-System erzeugt worden ist, um ein CI-Dokument, denn die Daten des Dokuments sind ohne weitere Bearbeitung durch IT-Systeme auswertbar.

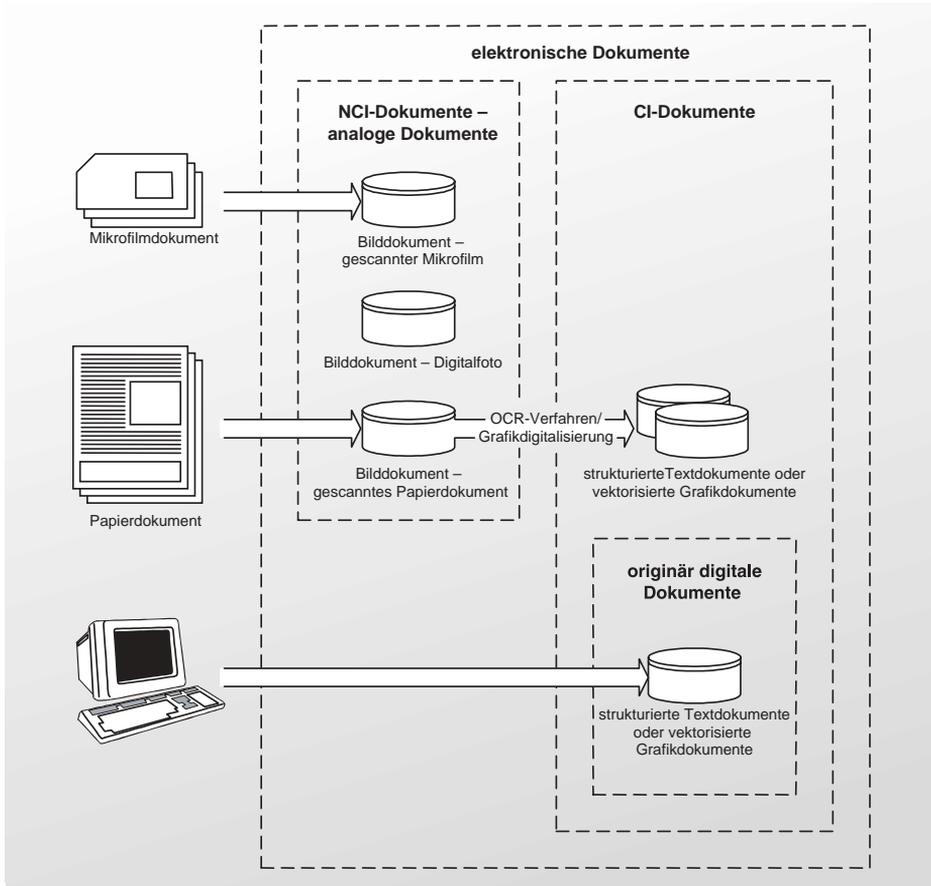
An dieser Stelle müssen auch die Begriffe »analoges Dokument« und »originär digitales Dokument« eingeführt werden. Hierbei handelt es sich um rechtliche Kategorien, mit denen die juristische Relevanz elektronischer Dokumente und die Verwendung elektronischer Signaturen beschrieben wird. Der Begriff des analogen Dokuments ist gleichbedeutend mit dem NCI-Dokument und der Begriff des originär digitalen Dokuments umfasst CI-Dokumente, die vor ihrer Einstellung in ein elektronisches Archivsystem bzw. vor der elektronischen Signierung bereits als CI-Dokumente vorlagen.

Hier soll auf Folgendes kurz eingegangen werden:

- Wenn ein Dokument von extern als Papierdokument in das Unternehmen kommt und dieses als Beleg dienen soll, dann muss es als analoges Dokument abgespeichert werden. Eine spätere Reproduktion (z. B. auf dem Bildschirm oder als Ausdruck) muss eine unveränderte bildliche Wiedergabe gewährleisten.

---

2. Non-Coded Information: Ein NCI-Dokument wird als Bild gespeichert (z. B. im TIFF-Format).



**Abb. 2-1** Erscheinungsformen von elektronischen Dokumenten

- Wenn man selbst Dokumente verschickt und man die eigene Kopie nur elektronisch vorhält, dann muss bei einer Reproduktion nur die unveränderte inhaltliche Wiedergabe sichergestellt sein. Eine exakte bildliche Wiedergabe ist nicht nötig.
- Seit kurzem gibt es auch die Möglichkeit, elektronische Dokumente mit einer elektronischen Unterschrift zu versehen und zu verschicken, die juristisch mit eigenhändig unterschriebenen Papierdokumenten gleichgesetzt sind. Dies sind »originär digitale Dokumente« die als Datei unverändert gespeichert werden müssen. Eine bildliche Wiedergabe ist in Bezug auf eine »Belegeigenschaft« irrelevant, wichtig ist allein, dass die Datei mit dem Dokument nicht verändert wird.

Im Detail werden diese Aspekte in Kapitel 8 behandelt.

### 2.1.3 Dokumentenlebenszyklus

Dokumente sind auch nicht statisch, sondern haben einen **Lebenszyklus** – sie entstehen, verändern sich, werden »eingefroren« und irgendwann vernichtet. In einer Organisation sieht man oft nur einen Ausschnitt aus dem Gesamtleben eines Dokumentes. Gerade wenn ein Dokument von einer Organisation zu einer anderen wechselt, ist es oft für beide Beteiligten wichtig, den Zustand des Dokumentes im Sinne eines Beleges festzuhalten. Die Vernichtung von Dokumenten ist nicht immer trivial, teilweise werden aus Gründen des Datenschutzes hohe rechtliche Anforderungen an die Art und Weise der Dokumentenvernichtung gestellt.

Der Lebenszyklus der Dokumente ist in der Regel eng mit den Geschäftsprozessen verknüpft:

- Geschäftsprozesse kreieren, verändern und vernichten Dokumente.
- Dokumente begleiten und steuern Geschäftsprozesse.

Daher gibt es bei den meisten Dokumenten-Management-Systemen inzwischen auch Workflow-Komponenten. Ein wichtiger Aspekt der Geschäftsprozesse ist nämlich das Dokument. Die oft auch sehr unterschiedlichen Formen von Dokumenten sind das eigentliche Objekt der Bearbeitung. Dokumente treten als Auslöser, Zwischenprodukte und Ergebnisse von Geschäftsprozessen auf. Ohne Dokumente ist in vielen Fällen der Prozess nicht zu verstehen. Die Erstellung des Dokumentes ist oft selbst ein ganzer Workflow – gerade im publizistischen Bereich. Dokumente verkörpern das Ziel des Prozesses. Im Prozessablauf sind Dokumente (egal ob elektronisch oder auf Papier) Träger der fachlichen und oft auch der steuernden Informationen!

### 2.1.4 Daten und Dokumente

Ein weiteres Thema ist die Abgrenzung der Dokumente von den Daten. Was unterscheidet nun Dokumente von Daten? Unter Daten versteht man stark strukturierte Informationen, während Dokumente eher als unstrukturierte oder schwach strukturierte Informationen angesehen werden. Dies stimmt zwar tendenziell, aber eine klare Trennungslinie lässt sich damit nicht formulieren.

Unstrukturierte Informationen lassen sich in herkömmlichen Datenbanken als so genannte »BLOBs« (= Binary Large Objects) abspeichern. Umgekehrt können Dokumente sehr gut strukturiert sein. Angefangen von reinen Images (NCI-Dokumente; NCI = Non-Coded Information) über Textdokumente, bis hin zu EDIFACT-Dokumenten, die vollautomatisch auswertbar sind, wächst der Grad der Strukturierung. Es lassen sich daher nur die Extreme klar bestimmen, der Übergang ist fließend. Im Einzelfall ist oft aus der Praxis heraus zu entscheiden, welcher Kategorie die einzelnen Informationen zuzuordnen sind. Oft sind Mischformen vorzufinden.

Die eindeutigste Unterscheidung ist wohl eher technischer Natur: Dokumenten-Management-Systeme teilen im Allgemeinen die Informationen in zwei Archive: Dokumentenarchiv und Meta- bzw. Indexdatenbank. Im Dokumentenarchiv werden die einzelnen Dokumente in ihrer Gesamtheit abgelegt. Beschreibende Informationen (wie Index, Schlagwörter etc. – also so genannte Metadaten) zu den Dokumenten werden in einer herkömmlichen Datenbank mit einem Verweis auf das jeweilige Dokument gespeichert. Zusätzlich gibt es unter Umständen noch Strukturinformationen (z. B. bei XML-Dokumenten), die etwas über den Aufbau des Dokumentes verraten. Datenbanksysteme legen dagegen alle Informationen in einer Datenbank ab.

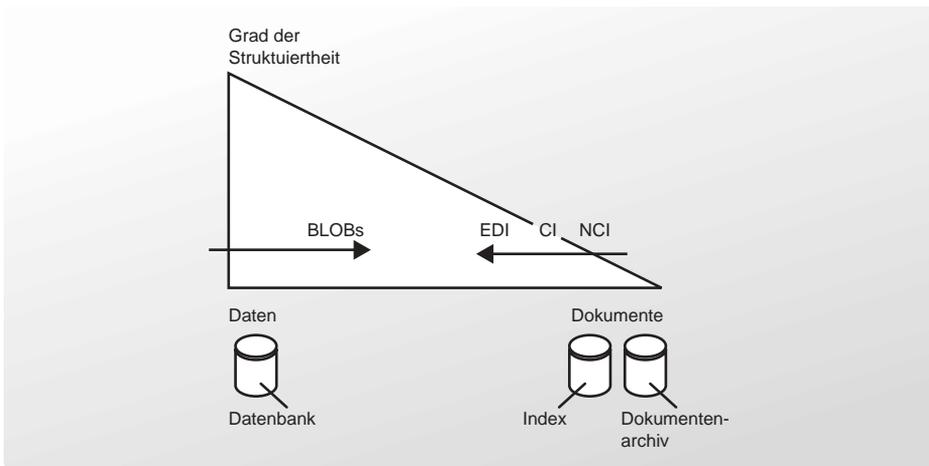


Abb. 2-2 Daten versus Dokumente

### 2.1.5 Die Beschreibung des Dokuments

Wenn ein DMS ein Dokument verwalten soll, dann werden unterschiedliche Elemente und Informationen benötigt:

1. **Das eigentliche Dokument**

Das Dokument wird in seinem Ursprungsformat und/oder konvertierten Format übernommen und gespeichert.

2. **Schlüssel (key)**

Für jedes Dokument gibt es einen eindeutigen Schlüssel (wie Rechnungsnummer oder künstlicher Schlüssel), mit dem es gefunden werden kann.

3. **Metadaten zum Dokument**

Dies sind beschreibende Attribute zum Dokument, die bei der Suche nach dem Dokument helfen können.

#### 4. **Strukturinformationen**

Für manche Dokumente gibt es Strukturinformationen über den inneren Aufbau des Dokumentes.

#### 5. **Regeln**

Lebensdauer und -zyklus, Zugriffsrechte und andere Verhaltensweisen können mit Regeln festgelegt werden.

Die Punkte 1. und 2. sind unabdingbar. Allerdings ist dies dann eher als einfaches Archivsystem zu bezeichnen.

## 2.2 Die Funktionen eines Dokumenten-Management-Systems

Für Dokumenten-Management-Systeme existiert eine Vielzahl von Anwendungen. Es gibt aber kaum ein System auf dem Markt, das sämtliche denkbaren Anforderungen optimal erfüllen kann. Ein Unternehmen, das ein Dokumenten-Management-System einsetzen möchte, muss deshalb zunächst einmal prüfen, ob die einzelnen Funktionen und Komponenten des Systems zu den jeweiligen eigenen Anforderungen passen. Grundsätzlich sollte ein Dokumenten-Management-System, um zukunftssicher zu sein, alle wesentlichen Belange berücksichtigen, die in Unternehmen, Behörden etc. für die **Erzeugung**, den **Empfang**, die **Ablage**, das **Verwalten**, das **Wiederauffinden** und die **Weiterverarbeitung** eines Dokumentes relevant sind.

Dokumentenerzeugung erfolgt heute überwiegend computerunterstützt. Briefe, Protokolle, Dokumentationen usw. werden mit Textsystemen (z. B. MS-Word) geschrieben; für die Rechnungsschreibung (Fakturierung) ist ein ERP-System zuständig. Bilder, Charts und Präsentationen werden entweder automatisch aus dem Zahlenmaterial einer Anwendung generiert oder mittels eines Grafikprogramms erstellt. Für die Weiterverarbeitung der Dokumente ist es sinnvoll, wenn sie in ihrer kodierten Form vorliegen, dagegen ist es jedoch für eine Langzeitarchivierung besser, die Dokumente im TIF-Format vorliegen zu haben; eine Aufbewahrung in kodierter Form erzwingt das jahrelange Mitführen der Software in der jeweils vorliegenden Erstellungsversion. Dass ein Dokumenten-Management-System heute mit seinen Prozessen nicht am Posteingang beginnt, sondern schon beim Kunden, Lieferanten oder anderen Partnern, zeigen ein- und ausgehende Dokumente in elektronischer Form (z. B. Electronic Mail, Datenträgeraustausch, EDI). Vor allem ist im Rahmen von E-Commerce mit seinen B-to-B- und B-to-C-Geschäften XML als Datentransfer für Dokumenteninhalte von großer Bedeutung. Der elektronische Austausch von Dokumenten beinhaltet darüber hinaus noch weitere Vorteile, wie Schnelligkeit, Zuverlässigkeit usw. Jedoch muss dies bezüglich der rechtlichen Möglichkeiten entsprechend untersucht werden. (vgl. Kapitel 8 – Rechtliche Aspekte).

Dokumenten-Management-Systeme sind nicht nur unter dem Gesichtspunkt der »Archivierung« der Dokumente zu betrachten, sondern auch unter dem Aspekt »Was geschieht davor und danach?« Folgende Punkte sind z. B. zu behandeln:

- Bearbeitung des Dokuments vom Entwurf bis zur Fertigstellung
- Prüfung und Genehmigung des Inhalts
- Weiterleitung des Dokuments
- Ablage und Archivierung
- Wiederauffinden des Dokuments

In jedem dieser Punkte stecken Arbeitsgänge, die noch häufig manuell ausgeführt werden. Genau hier liegt der Ansatz eines Dokumenten-Management-Systems. Aufgrund der differenzierten »Lebensstile« und »Lebenszyklen« können Dokumente unterschiedliche Zustände einnehmen. Solche Zustände sind z. B. »im Entwurf«, »in Bearbeitung«, »warten auf«, »fertig«, »wird geändert«. Ein Dokumenten-Management-System sollte daher – wenn solche Punkte wichtig sind – auch ein Statuskonzept aufweisen! Ein weiterer Punkt ist die Lebensdauer von Dokumenten: Wie lange sind sie zu archivieren und was passiert danach?

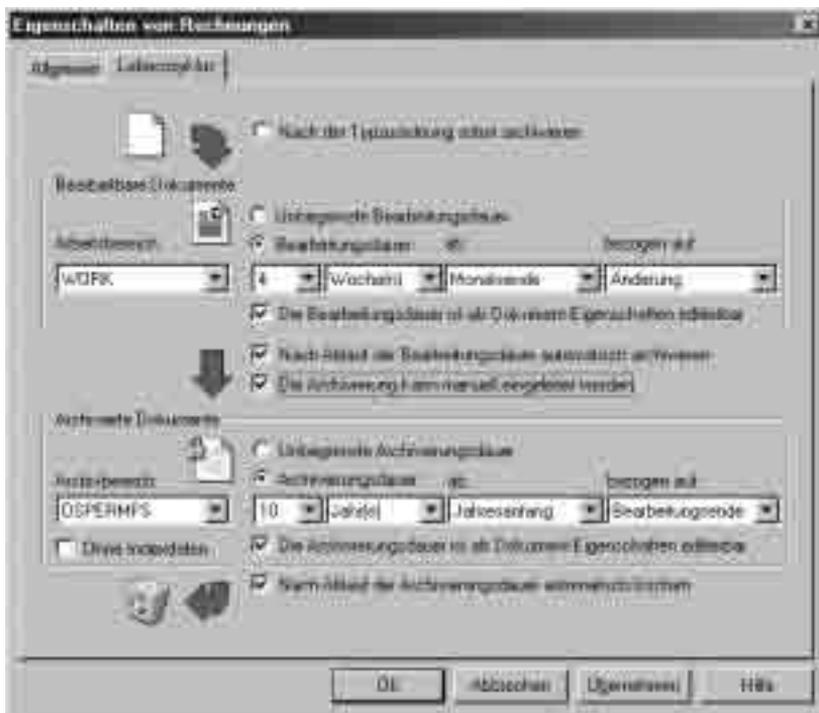
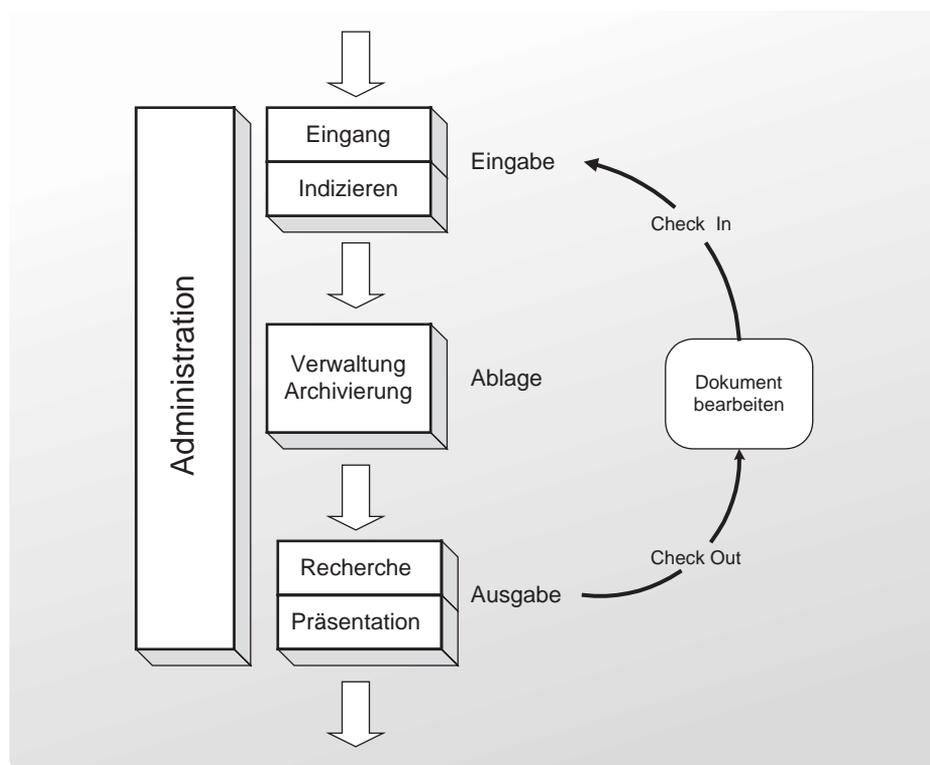


Abb. 2-3 Definition des Lebenszyklus von Dokumenten (Quelle: Windream GmbH)

Jeder Dokumententyp ist entsprechend der heutigen Technik in ein »föderatives<sup>3</sup>« Dokumenten-Management-System einzugliedern, d. h. jeder Dokumententyp steht für sich selbst innerhalb eines solchen Systems. Jedes Dokument soll nur ein einziges Mal gespeichert sein. Trotzdem ist zu gewährleisten, dass es

- für jeden Berechtigten (und nur diesen),
- jederzeit,
- sofort,
- vollständig,
- aktuell
- und verknüpft mit zugehörigen im Kontext stehenden Dokumenten zur Verfügung steht.



**Abb. 2-4** Grundstruktur eines Dokumenten-Management-Systems

3. Ein föderatives Datenbank- bzw. Dokumenten-Management-System besteht aus lose gekoppelten heterogenen lokalen Datenbanken, die über ein Netzwerk integriert sind. Im Gegensatz zu verteilten Datenbanken besteht kein einheitliches Schema, sondern mehrere unterschiedliche Schemata werden miteinander verbunden – es entsteht sozusagen von unten nach oben (bottom up)! Um diesen Verbund herzustellen, kommt meist eine so genannte »Middleware« zum Einsatz (z. B. auf Basis von DMA).

Die derzeit angebotenen Dokumenten-Management-Systeme sind aufgrund unterschiedlicher Voraussetzungen auf den Markt gekommen. Sie entsprangen den Bedürfnissen bestimmter Branchen oder wurden als Erstanwendungen bei großen Unternehmen entwickelt. Daher ist eine Anforderungsanalyse im eigenen Haus wichtig. Hier kann sich schnell herausstellen, dass unterschiedliche (Teil-) Systeme ggf. zu kombinieren sind, um die gestellten Anforderungen zu erfüllen.

Im Zuge des Einsatzes von Dokumenten-Management-Systemen sind verschiedene Formen von Dokumenten systematisch zu unterscheiden. Bei der Vielfalt der Dokumente ergeben sich zwangsläufig eine Reihe von Fragen: Soll es für mein Unternehmen nur **ein** Dokumenten-Management-System geben, das **alle** Anforderungen erfüllt? Gibt es überhaupt so ein System? Wenn ja, ist es nicht zu komplex und damit anwenderunfreundlich? Muss mein Unternehmen mehrere Systeme einsetzen? Wie sieht es dann mit den Schnittstellen aus und wird es nicht doch wieder insgesamt zu komplex? Um diese Fragen zu beantworten, sollte man sich zunächst mit den grundsätzlichen Funktionen und Komponenten eines Dokumenten-Management-Systems befassen.

Die Abbildung 2–4 zeigt die prinzipielle Grundstruktur eines Dokumenten-Management-Systems. Die Funktionsbereiche Eingabe und Ausgabe untergliedern sich immer in einen physischen und einen logischen Teil: Eingang des Dokuments und Indizierung bzw. Dokumentenrecherche und Reproduktion (Präsentation) des Dokuments. Die Details zu den Funktionsbereichen sind in Kapitel 7 nachzulesen. Auf den nachfolgenden Seiten sind die Funktionen nur grob erklärt.

## 2.3 Der Funktionsbereich Eingabe

Der Bereich Eingabe hat die Aufgabe, dass sämtliche Arten von Dokumenten, die in einem Dokumenten-Management-System zur Verfügung stehen müssen, dem System auch zugeführt werden. Dabei können die Dokumente ganz unterschiedlichen Ursprungs sein. Sie kommen z. B. von externen Partnern (Kunden, Lieferanten). Das kann über den herkömmlichen Postweg gehen (Papierdokumente) oder über eine Datenleitung (z. B. beleglos über EDI oder XML) geschehen. Ein Fax in Papier- oder elektronischer Form gehört ebenfalls dazu. Dokumente, die per E-Mail eintreffen und solche, die mit PC-Anwendungen erstellt wurden, muss das DMS übernehmen können. Dokumente, die in EDV-Anwendungen entstehen (z. B. Lieferscheine, Rechnungen für Kunden, Bestellungen an Lieferanten), sind dem Dokumenten-Management-System ebenfalls zuzuspielen. Auch hier kann es sich um beleghafte und beleglose Dokumente handeln.

Der Bereich Eingabe lässt sich in zwei Funktionen unterteilen (vgl. Abb. 2–4), die in den folgenden Abschnitten erläutert werden.

### 2.3.1 Die Funktion Dokumenteneingang

Der Dokumenteneingang befasst sich mit der physischen Übernahme des Dokuments in das Dokumenten-Management-System. Hier können unterschiedliche Techniken zum Tragen kommen:

Sämtliche Dokumente, die in beleghafter Form vorliegen, sind zu scannen. **Scanner**, vor allem die Hochleistungsscanner, gehören mit zu den wichtigsten Komponenten. Beim Scan-Vorgang wird das Dokument in ein elektronisches Bild (Dokumentenformat = Image) verwandelt. Es ist dann ein NCI-Dokument (Non-Coded Information).

Soll der Inhalt dieses Dokuments elektronisch weiterverarbeitet werden, ist es in ein CI-Format umzusetzen. Das kann mit Hilfe von OCR-Verfahren geschehen. Dokumente, die mittels eines elektronischen **Fax-Eingangs** in das System gelangen, haben prinzipiell den gleichen Charakter wie gescannte Dokumente und sind entsprechend zu behandeln.

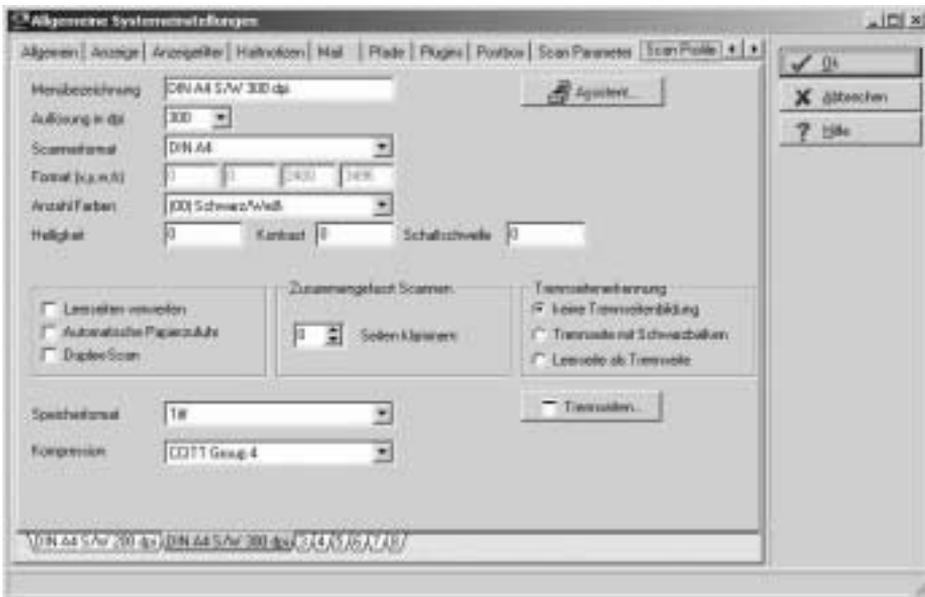


Abb. 2-5 Scanner-Einstellung (Quelle: ELO GmbH)

Wenn das Ursprungsdokument schon in elektronischer Form (also als CI-Dokument) vorliegt, kann es für weitere Be- und Verarbeitungsprozesse im Dokumenten-Management-System in diesem Format beibehalten bleiben. Für die Langzeitarchivierung ist es in ein NCI-Dokument (z. B. TIFF) umzuwandeln. Es kann durchaus auch sinnvoll sein, dass ein Dokument im Dokumenten-Management-System in zwei Formaten existiert. Für die Überleitung der Dokumente aus HOST-Anwendungen kommt das so genannte **COLD-Verfahren** (Computer Output on **L**aser **D**isc) zum Tragen.

COLD ist im Prinzip die neue Version der alten **COM**-Technologie (**Computer Output on Microfilm**). In manchen Bereichen fallen große Mengen an Dokumenten an (zum Beispiel Rechnungen im Versandhandel), die zwar vorgehalten werden müssen, deren Ausdruck und Archivierung für den Eigenbedarf aber sehr unwirtschaftlich ist. In diesen Fällen ist eine Kopie für den Eigenbedarf direkt auf einen optischen Datenträger zu schreiben, der bei Bedarf reproduziert werden kann.

Es ist noch zu unterscheiden, ob das gesamte Dokument (z. B. mit Formularrahmen oder Briefkopf) oder nur der Inhalt abzuspeichern ist. Im letzteren Fall ist festzuhalten (vom System unterstützt), um welche Art von Dokument es sich handelt, um bei der Reproduktion den passenden »Hintergrund« dazuspielen zu können.

Beliebige CI-Dokumente treffen täglich im Unternehmen ein: E-Mails oder EDIFACT-Dokumente, Dokumente aus Datenträgeraustausch, HTML-Dokumente, die mit XML übertragen wurden, etc. Alle diese Dokumente können in das Dokumenten-Management-System übernommen werden. Es ist aber darauf zu achten, dass das Dokumenten-Management-System die Dokumentenformate der importierten Dateien selbst interpretieren kann. Ist das nicht der Fall, so ist zum Reproduzieren auf die Originalanwendung zurückzugreifen. Oft können aber auch beliebige Viewer in das DMS integriert werden.

### 2.3.2 Die Funktion Indizieren

Das **Indizieren** bewirkt, dass die Dokumente mit Attributen versehen werden, die ein rasches Wiederfinden des Dokuments ermöglichen. Ziel einer Recherche ist es, dass nach Eintippen eines oder mehrerer Suchbegriffe die Dokumente sofort zur Verfügung stehen (z. B. am Monitor angezeigt werden). Für die Index-Erfassung gibt es eine allgemein gültige Regel: Je mehr Begriffe und Werte für das Dokument eingegeben werden, umso gezielter ist es später wieder auffindbar. Erhält ein Dokument nur einen oder wenige Indizes, so kann ein »Blättern« in mehreren Dokumenten erforderlich sein, um das richtige zu finden. Mittels der Indizes wird eine Datenbank aufgebaut, die zu den einzelnen Dokumenten die Suchbegriffe speichert und bei einer Recherche die gefundenen Dokumente anzeigt. Diese Indizes gehören zu den »Metadaten« der Dokumente. Es ist zu unterscheiden zwischen manueller und automatischer Indizierung. Mit automatischen Verfahren versucht man die Indizes aus vorhandenen Quellen zu generieren. Dies ist z. B. möglich bei den Verfahren, die mit COLD, Volltext, EDI/EDIFACT und Barcode arbeiten. Da diese Verfahren in der Regel zuverlässiger und billiger sind als das manuelle Indizieren, sollte man versuchen, diese so oft wie möglich einzusetzen.

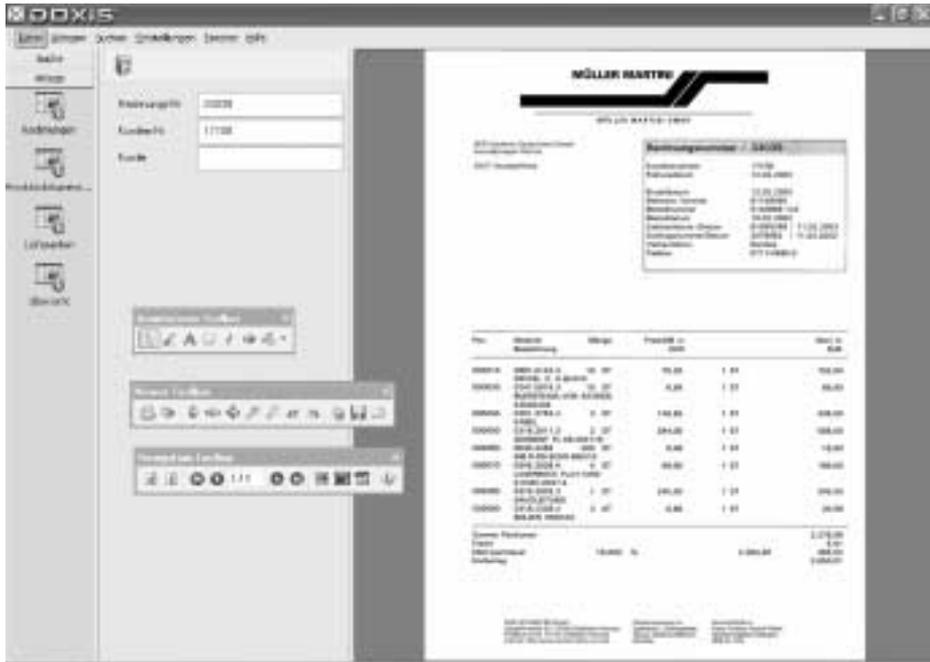


Abb. 2-6 Ablegen der Dokumente mit DOXIS (Quelle SER GmbH)

Weitere detaillierte Informationen zur Indizierung sind in Kapitel 7 zu finden.

Sehr robust ist die Anwendung von Barcode, da dieser sehr zuverlässig zu lesen ist. Praktisch alle Systeme haben heute ein Barcode-Modul (siehe Abb. 2-7).

Zurzeit entstehen Systeme auf der Basis von wissensbasierten Ansätzen oder neuronalen Netzen, die eine Klassifizierung der Dokumente automatisch durchführen können. Die Idee bei wissensbasierten Systemen (auch regelbasiert genannt) besteht darin, dass das System mittels umfangreichen Wissens über das Umfeld der Dokumente diese automatisch richtig einordnet und indiziert. Die Eingangspost eines Unternehmens soll z. B. automatisch richtig weitergeleitet werden. Dazu sind die Dokumente korrekt zu klassifizieren (z. B. Rechnungen, Aufträge, Beschwerdebriefe), der passende Vorgang zu ermitteln (z. B. Versicherungsnummer, Bestellnummer) und der Empfänger (Abteilung, Sachbearbeiter) festzustellen. Das System muss daher sehr viel über das Unternehmen und seine Umwelt wissen, um diese Aufgabe zu erfüllen. Dazu gehören Informationen über Geschäftsprozesse, Kunden, Produkte, Lieferanten, Gesetzgebung etc.

Der neuronale Ansatz geht anders vor: Zuerst indiziert ein Mensch verschiedene Dokumente und trainiert dabei das System. Das heißt, das System versucht anhand der Beispiele selbst Kriterien zu finden, nach denen die Dokumente indiziert werden können. Nach einer gewissen Trainingszeit arbeitet das System selbstständig.



Abb. 2-7 Konfigurieren der Barcode-Erkennung (Quelle: FileNet)

Diese Verfahren bringen oft schon erstaunlich gute Ergebnisse – sind aber praktisch nie fehlerfrei. Daher kann man sie immer dann gut nutzen, wenn ein Mensch das Dokument erhält und noch die Chance einer Korrektur besteht (wie z. B. bei der Eingangspost). Wenn das nicht gegeben ist, z. B. um die Endablage zu automatisieren, muss man dies sehr gut testen, ob es wirklich immer funktioniert.

## 2.4 Der Funktionsbereich Ablage

### 2.4.1 Die Funktion Verwaltung

Eingehende Dokumente müssen entsprechend ihrer Klassifizierung abgelegt und die Schlagwörter (vgl. Indizierung) in der Datenbank mit einer Referenz zum Dokument gespeichert werden (**Check-in**). Wichtig ist, dass nur über diesen »offiziellen« Weg Dokumente in die Ablage kommen.

Möchte sich ein Benutzer Informationen aus einem Dokument holen, dann startet er mittels Recherche einen Suchvorgang. Wird es für eine Weiterbearbeitung benötigt (**Check-out**), dann treten ähnliche Probleme wie bei einer Datenbank auf: Das Dokument ist für andere Nutzer (zumindest für Änderungen) zu sperren. Wenn die Bearbeitung zu einer neuen Version des Dokuments führt, dann ist die **Versionsverwaltung** gefordert. Hier gibt es die Möglichkeit, entweder immer das gesamte Dokument in jeder Version abzuspeichern oder nur den Unterschied zur Vorgängerversion.



Abb. 2-8 Historienverfolgung (Quelle: Windream GmbH)

Oft ist es erforderlich, den **Status** eines Dokuments festzuhalten. Manche Dokumente verfügen über einen richtigen Lebenslauf. Das Dokumenten-Management-System muss hier immer festhalten, in welchem Status es sich befindet, und wissen, wie die Statusübergänge erfolgen. Dies ist wichtig, weil z. B. Zugriffsrechte auf Dokumente von deren Status abhängen können.

Eine weitere Anforderung kommt hinzu, wenn komplexe Dokumentationen zu verwalten sind. So besteht z. B. bei vielen technischen Systemen, die in unterschiedlichsten Versionen in viele Länder geliefert werden, das Problem, dass die Dokumentation hierzu immer von Fall zu Fall variiert. Um dies mit vertretbarem Aufwand durchführen zu können, ist ein Dokumenten-Management-System mit einem **Konfigurationsmanagement** nützlich. Hiermit kann immer die gerade notwendige Dokumentation automatisch – in Abhängigkeit von der Konfiguration des technischen Systems – generiert werden. Bei diesen Funktionen kommt man allerdings schon in den Grenzbereich zum Content Management.

Analog zu den verteilten Datenbanken können auch **verteilte Dokumenten-Management-Systeme** aufgebaut werden. Hier sind in einem Netzwerk die Doku-

mente an verschiedenen Orten gelagert, um z. B. Zugriffszeiten zu optimieren. Von jeder Stelle im Netz besteht aber Zugriff auf alle Dokumente im System. In einer weiteren Stufe ist vorzusehen, die Dokumente mehrfach an bestimmten Orten abzulegen. Damit immer alle elektronischen Exemplare eines Dokuments den gleichen Stand haben, sind **Replikationsmechanismen**<sup>4</sup> anzuwenden.

Oft bilden verschiedene Dokumente unterschiedlicher Art und Herkunft eine logische Einheit und sind in so genannten **Mappen** oder **Containern** zusammengefasst. Diese sind gleichfalls zu verwalten. Damit erhält das Dokumenten-Management-System neben der Klassifikationshierarchie eine weitere hierarchische Struktur, die mehr an den Geschäftsvorfällen oder anderen anwendungsorientierten Kriterien ausgerichtet ist. Der Übergang zu einem **Workflow-System**, das oft mit derartigen Vorgangsmappen arbeitet, ist hier schon vorgezeichnet (vgl. Kapitel 4).

#### 2.4.2 Die Funktion Archivierung

Die **Archivierung** umfasst das Thema Speicherung und Wiedergewinnung der Dokumente und deren Metainformationen. Wichtig ist hierbei, den sicheren und schnellen Zugriff auf gewünschte Informationen zu gewährleisten. Dafür stehen verschiedene Komponenten und Techniken bereit:

- Datenbanken
- Schnelle Festplattensysteme
- Cachespeicher
- Jukeboxen mit WORM, MO- oder CD-ROM-Datenträger<sup>5</sup>
- Spiegelung zur Datensicherung (RAID-Systeme)

Die Administration des Archivs umfasst die Einrichtung der Jukeboxen, Verwaltung der Medien und der Datenbank, die Einrichtung des Caching und die Organisation der Sicherung. Weitere detaillierte Angaben zur Funktion Archivierung sind in Kapitel 7 »Archivierung und Verwaltungsmodul« beschrieben.

## 2.5 Der Funktionsbereich Ausgabe

Ähnlich wie der Bereich Eingabe ist auch die Ausgabe in zwei Teile zu gliedern: die Suche (= Recherche) nach einem Dokument und die Reproduktion mittels unterschiedlicher Ausgabemedien.

- 
4. Automatischer Kopiervorgang, der in einem Netzwerk alle Exemplare einer bestimmten Information (wie Datensatz oder Dokument) auf den gleichen Stand bringt. Bekanntestes Beispiel für ein derartiges System ist Lotus Notes.
  5. Es werden auch schon DVD-Laufwerke angeboten. Da aber bei den DVDs noch kein stabiler Standard erkennbar ist, sollte man hier noch vorsichtig sein.

### 2.5.1 Die Funktion Recherche

Die **Recherchefunktion** ermöglicht die gezielte Suche nach Dokumenten. Es gibt im Prinzip zwei Arten von Suchmöglichkeiten: die Suche über eine Struktur und die Suche über Volltext.

Die **strukturierte Suche** stützt sich auf die vorher fest eingegebenen Suchbegriffe (Klassifikation, Schlagwörter, Attribute). Wenn die Indizierung eine gute Qualität hat, dann kann diese Suche zeitlich sehr effizient und inhaltlich genau sein.

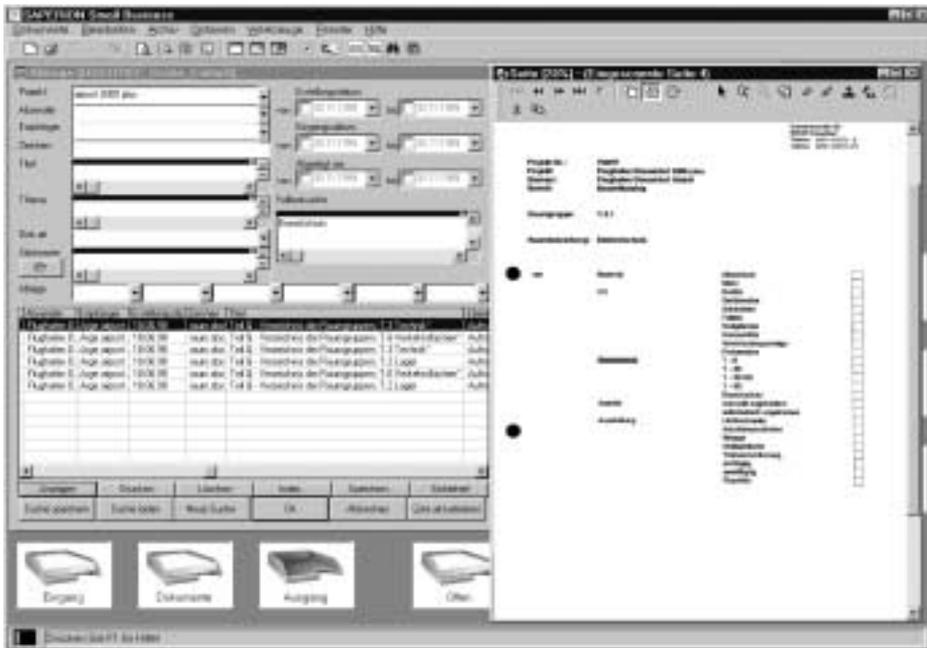


Abb. 2-9 Beispiel für Recherche-Bildschirm (Quelle: SAPERION AG)

Bei der **Volltextsuche** wird der Inhalt des ganzen Dokuments nach den Suchbegriffen abgesucht<sup>6</sup>. Vorteil ist, dass keine Abhängigkeit von der Qualität der Indizierung besteht. Das Dokument muss dazu natürlich in maschinenlesbarer Form als CI-Dokument vorliegen. Reine Images sind so nicht absuchbar.

Verfeinern lässt sich das System noch, wenn automatisch auch nach gängigen Synonymen gesucht wird und Begriffe in einem logischen Kontext stehen (z. B. nach »Mutter« ist nur zu suchen, wenn sich die »nähere Umgebung« in der Textfundstelle mit Schrauben befasst).

6. Eigentlich wird schon bei der Aufnahme des Dokuments in das DMS ein vollständiger Index vom gesamten Dokument angelegt.

Dokumenten-Management-Systeme stellen oft beide Formen integriert zur Verfügung. Sie bieten dem Anwender damit die Möglichkeit der intuitiven Suche. Die vielfach geäußerte Meinung, dass bei einer Volltextrecherche mit langen Suchzeiten zu rechnen ist, trifft heute nicht mehr zu, vor allem, wenn das System mit einem Thesaurus arbeitet.

Die Formulierung einer Suche kann dabei beliebig komplex sein: Es sollen z. B. alle Dokumente gesucht werden, die entweder eine Bestellung oder Anfrage nach einem bestimmten Produkt beinhalten, die in einem bestimmten Zeitraum erfolgten, aber nicht aus Österreich kamen. Hierfür steht die Logik der Boole'schen Algebra zur Verfügung, mit deren Hilfe derartige Abfragen formuliert werden können. Ein normaler Nutzer kann dies aber meistens nicht fehlerfrei formulieren. Daher muss ein Dokumenten-Management-System über eine entsprechend komfortable Oberfläche verfügen, die es ermöglicht, einfach und intuitiv derartige Abfragen zu starten.

Weitere Details hierzu sind in Kapitel 7 unter »Ausgabemodul« beschrieben.

### 2.5.2 Die Funktion Reproduktion

Ein gefundenes Dokument ist dem Anwender bzw. dem Nutzer wieder zu präsentieren. Hierzu soll das Dokument angezeigt, gedruckt, elektronisch weitergeleitet oder anderweitig reproduziert werden.

Dokumenten-Management-Systeme verfügen über eine Vielzahl von Möglichkeiten, Dokumente dem Anwender zur Verfügung zu stellen. Dokumente können gedruckt, am Bildschirm angezeigt, per Fax oder E-Mail versendet werden. Bei CI-Dokumenten kann auch sofort die passende Applikation mit dem geladenen Dokument gestartet werden. Die meisten Systeme verfügen inzwischen auch über einen Web-Client. Dadurch kann man dann von beliebigen Orten aus über Internet oder Intranet im DMS recherchieren. Die Techniken der Dokumentenanzeige mit Hyperlinks auf eingebettete Objekte (Grafiken, Fotos, Dokumente aller Art) stehen dem Anwender zur Verfügung.

Bei der Auswahl der Drucker sind insbesondere die gewünschte Qualität, das Druckvolumen und Formate (evtl. auch Farbe) zu berücksichtigen.

Bei rechercheintensiven Arbeitsplätzen und solchen zur Indizierung sind Monitore mit höchster Qualität gefragt. Am Budget sollte hier nicht gespart werden. Wichtig ist, dass der Mitarbeiter das Dokument möglichst in Originalgröße und flimmerfrei in hoher Auflösung sieht. Die einschlägigen Richtlinien für Bildschirmarbeitsplätze sind unbedingt zu beachten. Zu bedenken ist auch, dass selbst die besten Bildschirme mit der höchsten Auflösung noch nicht die Auflösung erreichen, die mit Scannern und Laserdruckern standardmäßig erzielt wird. Das bedeutet, dass der Mitarbeiter am Monitor immer eine schlechtere Bildqualität vor sich hat!



Sind Dokumente aus beliebigen Anwendungen zu importieren, ist zu unterscheiden, ob das Dokumenten-Management-System das Dokumentenformat (z. B. Winword-Dokument) kennt, oder ob zur Reproduktion die Originalanwendung zu aktivieren ist. Letzteres bedingt, dass die Originalanwendung (wie Winword) überall dort, wo die Dokumente gedruckt oder angezeigt werden sollen (zumindest in einer Viewer-Version), zur Verfügung steht. Möglicherweise sind für ältere Dokumente auch die alten Versionen der Anwendungen (z. B. wegen Makros) einzusetzen. Das erfordert eine zusätzliche Administration.

## 2.6 Der Funktionsbereich Administration

Ein DMS muss auch administriert werden. Der Funktionsbereich **Administration** befasst sich mit der Installation und Pflege von:

- Zugriffsberechtigungen
- Benutzereinstellungen
- Komponentenanschlüssen
- Schnittstellenintegrationen mit anderen Anwendungen
- Statistischen Auswertungen für
  - Protokolle
  - Datenträgerbelegung
  - Zugriffshäufigkeit auf
    - Datenträger
    - Dokumentenarten
    - Attribute
- Daten- und Dokumentensicherung
- Backup und Recovery

Dabei müssen nicht immer alle Administrationsfunktionen in der IT-Abteilung angesiedelt sein: Je nach Thema können sie auch in speziellen Rollen der Fachabteilung zu finden sein, wie z. B. die Vergabe von Zugriffsrechten auf Dokumente, Ordner etc. oder die Einrichtung von neuen Ordnern bzw. Veränderung der Ablagestruktur. Systemnähere Customizing-Funktionen, wie die Definition der Metadaten oder Einrichtung von Erfassungsmasken sind in der Regel Aufgaben der IT.

Zugriffsberechtigungen in einem Dokumenten-Management-System haben im Allgemeinen einen mindestens ebenso hohen Stellenwert wie bei herkömmlichen IT-Anwendungen. Die Informationen, die in einem Dokument enthalten sind, können sehr sensibel und personenbezogen sein. Es ist darauf zu achten, dass es bei der Zugriffsberechtigung eine Unterscheidung zwischen dem Zugriff auf Indizes bzw. Daten und Dokumente gibt. In einem Landratsamt z. B. darf ein Sachbearbeiter, der für das Grundschulwesen zuständig ist, auf alle Dokumente eines Bürgers zugreifen, die sich auf notarielle Beglaubigungen eines Grundschuleintrages und auf die Hypothekendarlehen beziehen, nicht aber auf die Einkommenssteuererklärungen bzw. -bescheide. Die Recherche auf den Bürger ist

möglich, nicht aber der Zugriff auf alle unter seinem Namen archivierten Dokumente.

Mit Benutzereinstellungen sind die Komponenten gemeint, die einem Arbeitsplatz zur Verfügung stehen. Ein Recherche-Arbeitsplatz benötigt beispielsweise keine Scannerkomponente(n). Dem Administrator obliegt auch die Überprüfung der richtigen Installation von Komponentenanschlüssen. Er muss z. B. dafür sorgen, dass der Arbeitsplatz, der mit OCR arbeiten soll, auch das OCR-Programm aktivieren kann.

Die Integration des DMS mit einer Host-Anwendung ist entweder standardmäßig vorgegeben, wie z. B. zwischen SAP R/3-Anwendungen und von SAP zertifizierten Dokumenten-Management-Systemen über die SAP-ArchiveLink-Schnittstelle, oder sie muss individuell erstellt werden.

Die statistischen Auswertungen dienen unterschiedlichen Zwecken. Protokolle für Dokumente, die dem HGB bzw. der AO unterworfen sind (siehe auch Kapitel 8 »Rechtliche Aspekte«), muss das DMS ausgeben können. Die Entstehung, Indizierung und Benutzung der Dokumente ist zu protokollieren, um damit den Nachweis der ordnungsmäßigen Führung derselben erbringen zu können. Dieser Punkt ist beim Anbieter des Dokumenten-Management-Systems besonders zu erfragen, da nicht sämtliche DMS-Hersteller diese gesetzliche Anforderung realisiert haben.

Die Statistiken für die Häufigkeit des Zugriffs auf Datenträger, Dokumentenarten und Attribute können helfen, Trendaussagen zu erkennen, so dass z. B. aufgrund der Größenordnungen des Zugriffs entsprechende Dokumenten-Pools zu bilden sind, die durch geeignete technische Maßnahmen höhere Zugriffsgeschwindigkeiten erlauben.

Daten- und Dokumentensicherung sind administrativ sorgfältig zu handhaben. Wenn es zu einem schwerwiegenden Fehler bei der Daten- und/oder Dokumentenspeicherung kommt, kann dies zu einer fatalen Situation des Unternehmens führen. In diesen Speicherbereichen liegt das gesamte Firmen-Know-how. Steht es nicht zur Verfügung, kann das im schlimmsten Fall einen Unternehmenskollaps verursachen.

Archive für Dokumente sind aus der Sicht des Betriebssystems oft nur eine einzige große Datei, die mit den herkömmlichen Datensicherungssystemen nur in ihrer Gesamtheit gesichert werden kann. Dies führt zu sehr großen, oft weitgehend redundanten Sicherungsbeständen. In diesem Fall ist es besser, wenn das Dokumenten-Management-System selbst eine Sicherungsprozedur mit den bekannten Möglichkeiten der Differenzsicherung bietet.

Das **Customizing**<sup>7</sup> ermöglicht individuelle Anpassungen. Erfassung- und Recherchemasken müssen z. B. nach den Anforderungen des jeweiligen Arbeits-

---

7. Das Thema Customizing ist als Grenzfall zwischen Administration, Entwicklung bzw. Weiterentwicklung des DMS zu sehen und muss nicht unbedingt der Administration zugeordnet werden.

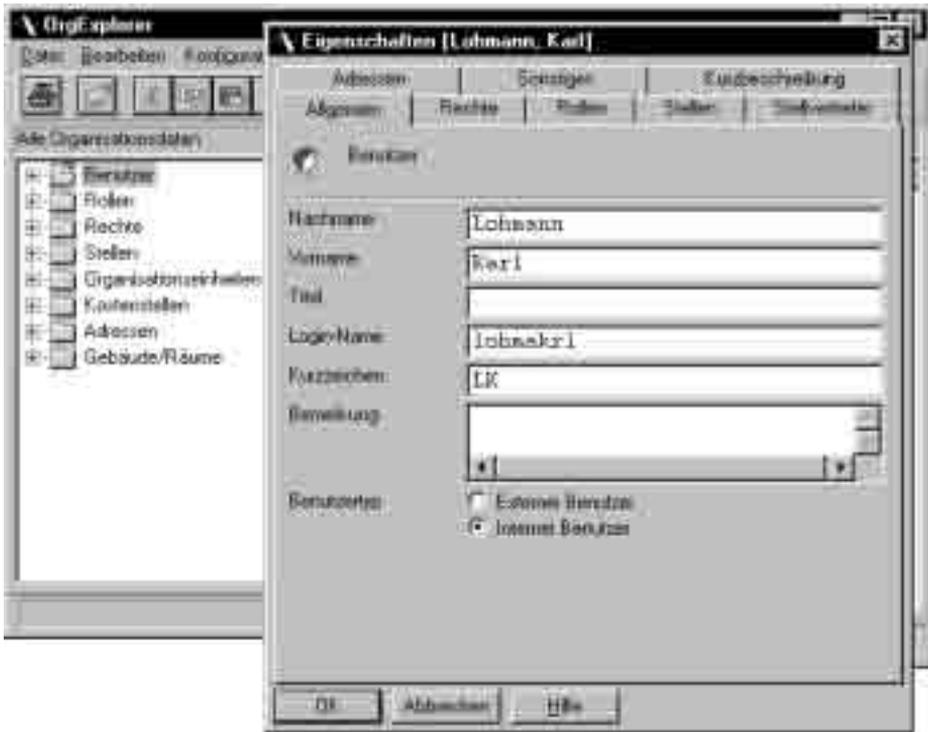


Abb. 2-11 Beispiel einer Maske für Benutzerverwaltung (Quelle: COI)

platzes gestaltbar sein. Maskengeneratoren oder von den DMS-Herstellern oft eigengefertigte Skriptsprachen erlauben solche Anpassungen (vgl. auch Kapitel 7). Mit diesen Tools lassen sich Datenfelder z. B.

- individuell positionieren,
- mit Defaultwerten belegen (feste Wertvorgabe),
- als Mussfelder festlegen,
- mit einer Eingabesperre definieren,
- mit Plausibilitätsmerkmalen prüfen,
- mit Prüzfiffernroutinen ausrüsten,
- Bereichszugehörigkeiten kontrollieren (von: bis:)
- und Abhängigkeiten zu anderen Datenfeldern bestimmen.

Die Maskengeneratoren und Skripts führen auch fertige »Buttons«, die speziell für Dokumenten-Management-Systeme gefertigt wurden. Dazu gehören Objekte wie »Scannen«, »Suchen«, »Anzeigen«, »Drucken«, »Hinzufügen«, »Ändern«, »Löschen« usw.

Das Customizing beinhaltet auch Verknüpfungen mit anderen Anwendungen auf PC-Ebene, z. B. mittels DDE (Dynamic Data Exchange) und/oder OLE (Object Linking and Embedding). Mit DDE erfolgt ein Datenaustausch zwischen

den Anwendungen. Über OLE finden Verknüpfungen zwischen Objekten statt. So kann ein Dokument mit anderen Dateien (Text, Tabellen, Grafiken), die nicht zum ursprünglichen Teil eines Dokumentes gehören, verknüpft werden. Diese Objekte bleiben eigenständig. Es gibt auch die Möglichkeit, Objekte zu einem Teil eines Dokumentes zu machen, indem sie über OLE in das Dokument einzubetten sind.

Eine weitere Technik des Customizing ist das Arbeiten mit APIs (Application Programming Interfaces). APIs sind fertige Funktionsobjekte einer Programmiersprache, die immer wieder im selben Programm oder in anderen Programmen genutzt werden können. So ist das Objekt »Scannen«, das beispielsweise in Visual Basic erstellt wurde, in allen Visual-Basic-Programmen aufrufbar.

Ein DMS sollte so komfortabel sein, dass die anwenderbezogenen Anpassungen über Parameterangaben stattfinden. Es wird zu einem teuren Produkt, wenn die meisten Customizing-Anforderungen neu zu programmieren sind.

## 2.7 Weitere Komponenten

Über die Basiskomponenten hinaus sind weitere Komponenten zu beachten, vor allem dann, wenn Anforderungen an elektronische Postverteilung und an Funktionen der Vorgangsbearbeitung (vgl. auch Kapitel 4) zu stellen sind. In diesen Fällen sollte das Dokumenten-Management-System folgende Komponenten unterstützen:

- Zuordnung oder Verteilung eines Dokumentes sowohl/als auch nach
  - Postkörben
  - Personen
  - Gruppen
  - Vorgängen
  - Mappen/Akten/Ordern
  - Sachgebieten
  - Prozessen
- Elektronische Signatur
- Prioritätensteuerung von Vorgängen/Prozessen (automatisch und individuell)
- Führen von Bearbeitungsstati (z. B. Neuer Vorgang, Auf Wiedervorlage, Aktiv, Ruht, Abgeschlossen)
- Automatische Wiedervorlage zum Arbeitsplatz
- Signalgebung bei Eintreffen eines Dokumentes im Postkorb (von Personen, aus Wiedervorlage)
- Nachrichtenausgabe an Vertreter oder Führungskraft, wenn nach einem bestimmten Zeitintervall ein Dokument nicht bearbeitet wurde
- Protokollierung der Aktivitäten im DMS

Diese Aufstellung ist nur beispielhaft und erhebt nicht den Anspruch auf Vollständigkeit. Eine weitere Komponentengruppe, die ebenfalls im Rahmen von Dokumenten-Management-Systemen zu berücksichtigen ist, betrifft hauptsächlich die Funktionen der unterschiedlichen Dokumentationserstellungen:

- Technische Dokumentation
- Montageanleitungen
- Gebrauchsanweisungen
- Bedienungsanleitungen
- usw.

Zusätzliche Tools, die Funktionen wie

- Konzeptentwurf
- Entwicklung der Dokumentation
- Prüfen der Dokumentation
- Übersetzen
- Publizieren
- und Ausliefern

unterstützen, sollten im DMS integrierbar sein (siehe Kapitel 3.3.2 »Die technische Dokumentation folgt dem Produkt«).

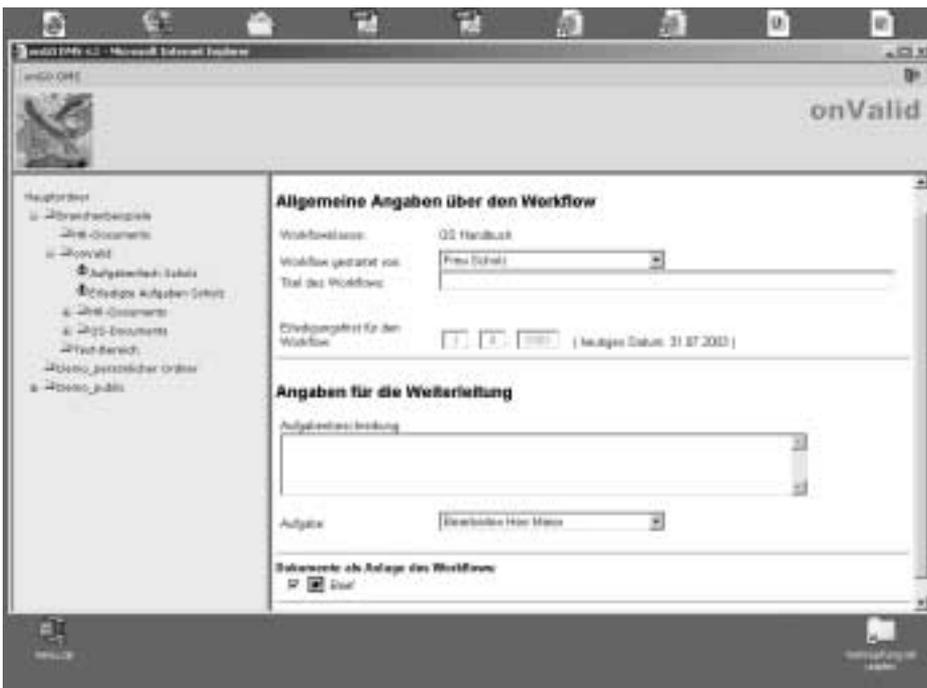


Abb. 2-12 Beispiel für Workflow-Komponente zur Vorgangsteuerung (Quelle: UNIPLEX)