Informationsmanagement in der Praxis

Vom Dokumentenmanagement zum ganzheitlichen Informations- und Wissensmanagement

Alle Rechte sind vorbehalten. Die Weitergabe des Zugangscodes/pdf-Dokumentes an unbefugte Dritte ist untersagt. Jede über die eigenen Zwecke des Erwerbers hinausgehende Verwendung, insbesondere die Vervielfältigung, Weitergabe und Veröffentlichung ist nicht zulässig.

Best.-Nr. 06 612-w

Version 1.0

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Informationsmanagement in der Praxis – Vom Dokumentenmanagement zum ganzheitlichen Informations- und Wissensmanagement / AWV – Arbeitsgemeinschaft für wirtschaftliche Verwaltung e. V. – Eschborn: AWV, 2003
ISBN 3-931193-44-6

ISBN 3-931193-44-6
© AWV-Eigenverlag, Eschborn 2003

Herausgeber:
AWV – Arbeitsgemeinschaft für wirtschaftliche Verwaltung e. V.
Postfach 51 29, 65726 Eschborn, Düsseldorfer Straße 40, 65760 Eschborn.
Telefon (0 61 96) 4 95-3 88, Telefax (0 61 96) 4 95-3 51,
E-Mail: info@awv-net.de, Internet: http://www.awv-net.de

Verantwortlich: Dr. Roland Wirth
Titelgestaltung und Layout: Karin Scheu

Verfasser:

Projektgruppe 5 des Arbeitskreises „Informations- und Dokumentenmanagement“ (Leiter: Dr. Burghard Nuhn), Fachausschuss „Informationswirtschaft“ (Vorsitzender: Dr. Hubert Salm).

Autoren des Arbeitskreises:

Jürgen Brunner	Volkswagen AG, Wolfsburg
Jürgen Klauenberg	Braunschweig
Dr. Burghard Nuhn	Triade GmbH, Neuss
Dr. Hubert Salm	MACH AG Düsseldorf
Martin Schlüter	SER GmbH, Hannover

Inhaltsverzeichnis

1 Vom Dokumentenmanagement zum ganzheitlichen Informations- und Wissensmanagement 1
1.1 Konstruktion und Evolution – Die Entwicklung von Industrieprodukten 2
1.2 Innovationszyklen und Evolution in der Informationstechnologie 3
1.3 Informationsmanagement in der Wende zum digitalen Zeitalter. 5

2 Strukturierungsmöglichkeiten 8
2.1 Automatisierte Dokumentenerfassung 9
2.2 Gewinnung von Metadaten aus den digitalen Objekten..... 9
2.3 Gewinnung von Metadaten, die das digitale Objekt beschreiben 13
2.4 Wissens-Management-Systeme verwenden keine Metadaten mehr 15
2.5 Strukturierung von Information (und Wissen) 16
2.5.1 Auswertung und Nutzung von Informationen 16
2.5.2 „Klassische“ digitale Objekte 17
2.5.3 Klassifikation der Nachrichten 18
2.6 Weitergehende Ziele der Wissenslogistik 19
XML-Technologie 24
Zusammenfassung und Ausblick 27

3 Ganzheitlicher Ansatz 29
3.1 Strategien 29

4 Der prozessorientierte Ansatz 31
4.1 Der Ablauf von Prozessen 31
4.2 Störfaktoren 31
4.3 Sollkonzept 33
4.3.1 Die Strukturierbarkeit von Prozessen 34
4.4 Automatisierung von Prozessen 36
4.5 Die menschliche Komponente 37
4.5 Wertschöpfung 38

5 Informationsmanagement und seine Anforderungen an Speichersysteme 39

1 Vom Dokumentenmanagement zum ganzheitlichen Informations- und Wissensmanagement

Der Siegeszug des Digitalen in allen Lebensbereichen hat uns eine Datenflut beschert, die in ihren drei Dimensionen Inhalt (Was), Format (Wie) und Historie (Wann) fast nicht oder nur mit Mühe organisierbar erscheint. Wie weit ist eine Informationswelt noch entfernt, in der die Verwertung der atomaren Informationseinheiten nicht mehr durch die Systemumgebung bestimmt wird, sondern durch den Nutzer? Die Weiterentwicklung der heutigen Informationsstrukturen, in denen spezielle Daten oder Dokumente ausschließlich über spezielle Anwendungen verfügbar gemacht werden können, ist ein vielleicht mühsamer, aber in jedem Fall lohnender Prozess.

IT-Hardware, Netzwerke, digitale und analoge Speichermedien wie Festplatten und Bänder, optische und fotografische Medien, spielen für die Organisation von Daten nur eine bedeutende Nebenrolle. Für die Erschließung der Goldadern und Abraumhalden unserer Datenüberflussgesellschaft sind wir auf die kontinuierliche Weiterentwicklung der Softwaresysteme angewiesen. Die Entwicklung des geeigneten Instrumentariums kann nur in den möglichst kurzen Zyklen zwischen Innovation und Markterprobung erfolgen. In dieser Prozesskette wird nicht nur von den Herstellern von Softwaresystemen erwartet, dass sie innovative Produktentwicklungen mit hohen Qualitätsstandards kostengünstig vorantreiben. Innovationsbereitschaft, Flexibilität der Organisation und Bereitschaft zu ständigem Lernen muss Teil der Unternehmenskulturen sein, die sich in diesem innovativen Feld behaupten wollen.

Dazu müssen bestehende Konzepte hinterfragt werden. In kurzen innovationsorientierten Produktzyklen müssen auf die dringlichsten der aktuellen Fragestellungen Antworten gegeben werden. Sind wir auf dem

richtigen Weg in die informationsgetriebene Welt? Diese Broschüre zeigt die wesentlichen Konzepte, die heute angewendet werden können, um eine eigene, erfolgreiche Strategie bei der Einführung von Systemen zum Informationsmanagement zu entwickeln und Informationsmanagementsysteme bedarfsgerecht einzuführen.

1.1 Konstruktion und Evolution – Die Entwicklung von Industrieprodukten

Um neue Technologien in Produkte zu verwandeln müssen einerseits vorhandene Konzepte aufgegriffen und weiterentwickelt werden, andererseits führt nur der unbefangene und kreative, oftmals auch fast spielerische Umgang mit neuen Technologien zur Erkenntnis neuer Möglichkeiten und damit zur Schaffung und Erschließung neuer Märkte.

Mercedes Omnibus aus dem Jahre 1895

Aus der Entwicklungsgeschichte unserer industriellen und postindustriellen Epoche ist das Automobil nicht wegzudenken, das, aufsetzend auf dem Konzept der Postkutsche, in einem über 120 Jahre dauernden Evolutionsprozess der gegenseitigen Anpassung das gesellschaftliche und wirtschaftliche Leben, die alltäglich erlebte Infrastruktur unserer Landschaften und das gesamte soziale Gefüge verwandelte.

VW Käfer aus dem Jahre 1938 mit „Trittbrett“ und „Kotflügel“

Über Jahrzehnte und die Einführung der „Fließbandproduktion“ durch Henry Ford hinweg ist im Automobil noch der Urahn „Postkut-

sche“ und das mit der Postkutschenära verbundene Verkehrswegesystem anzusehen.

In über einhundert Jahren konnte sich das Automobil zur heutigen Form und Funktion entwickeln – und wir dürfen heute immer noch viel Phanta-

sie darauf verwenden, uns die weitere Evolution und die Entstehung neuer Spezies aus dem Konzept Automobil auszumalen.

1.2 Innovationszyklen und Evolution in der Informationstechnologie

Von der Geburtsstunde des Automobils bis zum heutigen Entwicklungsstand haben etwa vier Generationen Automobilbauer das Gesicht unserer Landschaften verändert und mit dem Automobil den Innovations- und Wachstumsmotor schlechthin für unsere Industriegesellschaft geschaffen. Das Automobil im speziellen und die Verkehrstechnik im allgemeinen erfüllt einen Grundbedarf aller Menschen in allen Kulturen. Der Markt der produzierenden und der erdölexportierenden Länder ist global.

Einen fast noch größeren Einfluss auf den Gang der Weltgeschichte und die Geschicke der Menschen als die Erfindungen rund um die Mechanisierung und Automatisierung der Fortbewegung hatten die bahnbrechenden Erfindungen, die das Grundbedürfnis der menschlichen Gemeinschaft unmittelbar nach Essen und Trinken erfüllte: das Grundbe-

dürfnis nach Kommunikation, dem Austausch von Wissen und Erfahrung, der Weitergabe von Verbindlichkeit. Die Erfindung des Buchdrucks oder auch die Entdeckungen, die zur Technik der Telegraphie und des Telefonierens führten, haben die Menschheit im globalen Maßstab verändert.

Am Anfang der jeweiligen Entwicklungen stand immer auch die Euphorie über Perspektiven und neue ungeahnte Möglichkeiten ebenso wie auch die Ratlosigkeit über den Umgang mit und die Auswirkungen der gerade gemachten Entdeckung. Aber auch mancher Irrweg und viele vergebliche Versuche, die neuen Erkenntnisse und Fertigkeiten anzuwenden, begleiteten den jeweiligen „Innovationsschub“. Die Technikgeschichte bietet reichhaltiges Anschauungsmaterial über solcherlei Fehlversuche. Nachfolgenden Generationen erschienen und erscheinen diese oftmals lächerlich und unsinnig. Es ist das selbe Motiv, das uns heute eine gewisse melancholische Heiterkeit entlockt, wenn wir die Datentechnik der frühen neunziger Jahre betrachten.

Die erstaunliche Entwicklungsgeschwindigkeit der Basistechnologien findet ihren signifikanten Ausdruck beispielsweise im Leistungswachstum der zentralen Verarbeitungskomponenten von Rechnersystemen oder im Wachstum von Datenmengen, die pro Flacheneinheit untergebracht oder pro Währungseinheit gekauft werden können. Mit solchen Werten können wir die erstaunliche Entwicklungsgeschwindigkeit veranschaulichen, mit der die Informations- und Kommunikationstechnologie vorangetrieben wird. Aber so wenig wie die Motorstärke beim Automobil entscheidend für die Handhabung und den Nutzen eines Fahrzeuges ist, so wenig ist allein Prozessorleistung und Speicherplatz ausschlaggebend für die Handhabbarkeit der Informationstechnik und die Leistungsfähigkeit von Kommunikationssystemen. Nutzungskonzepte und deren Umsetzung in handhabbare Systeme, die sich nach Möglichkeit den Fähigkeiten der Menschen anpassen und sie in idealer Weise unterstüt-

zen, generieren erst den Nutzen, den Mehrwert, mit dem sich die Kosten für die Beschaffung und Einrichtung technischer Systeme rechtfertigen.

Nicht nur der Skeptiker der „Internetgesellschaft“, wenn es diesen überhaupt noch gibt, sondern gerade die Menschen, die in ihrem täglichen Arbeitsumfeld auf gut konzipierte Systeme und ergonomisch einwandfrei gestaltete Abläufe angewiesen sind, werden sich wohl oft so manches Mal mit Verwunderung fragen, ob die jeweilige Unterstützung eines „Systems“ für den Arbeitsvorgang wirklich ausgereift und optimal ist.[1]

Die Behauptung, dass die konzeptionelle Weiterentwicklung unserer Systeme heute ganz sicher nicht Schritt hält mit den technischen Möglichkeiten soll an einem Beispiel vertieft werden.

1.3 Informationsmanagement in der Wende zum digitalen Zeitalter

Das „papierlose“ oder „papierarme“ Büro ist die Zuspitzung des Konzepts einer technikunterstützten Arbeitsweise auf ein Schlagwort.

Dieses Schlagwort entwickelten Marketing-Strategen, die nach Anwendungsmöglichkeiten und der wirtschaftlichen Verwertung der damals gerade zur Marktreife entwickelten Technik der „optischen Speicher“ suchten. Die Ablösung der analogen „Datenträger“ Papier und Mikrofilm erschien zum Greifen nahe. Das Ausbleiben des „papierarmen“ oder gar „papierlosen“ Büros gibt Ursache für manchen Spott:

„Wer vor zehn Jahren das papierlose Büro prophezeite, der hatte sich geirrt, wie man sich schlimmer nicht irren kann.“ schreibt Georg Küffner

[1] Interessante Beobachtungen dazu von Axel Thallemer, High-Tech und Design. Design als Innovationsprozess in: 4:3 50 Jahre italienisches & deutsches Design. Bonn 2000.

in der FAZ vom 19. Februar 2001 unter dem Titel „Das Zeitalter des Papiers ist längst nicht vorbei – Trotz Internet der wichtigste Informationsträger."

Die Kontrahenten scheinen meist nicht bemerkt zu haben, dass die erwartete „digitale Revolution" bereits stattgefunden hat, dass dabei aber das Symptom „papierfrei" nicht eingetreten ist. Das Schlagwort vom „papierlosen" Büro gehört ebenso zu den überholten Relikten einer zurückliegenden Epoche der Technikentwicklung wie die 51/4" „Floppy" oder das 8" Magnetplattenlaufwerk. Während sich die Produktion und Verbreitung von „Schriftsätzen" schon längst auf digitale Mechanismen verlagert hat, sind wir für die kurzfristige Visualisierung und Speicherung und oft auch für die Verfolgung der Bearbeitung von Schriftsätzen auf das Papier angewiesen. Das Nebeneinander der digitalen und der analogen Verarbeitungsformen rührt nicht etwa daher, dass die technischen Plattformen nicht über ausreichende Kapazitäten verfügen. Es fällt uns einfach schwer, die über Jahrhunderte erprobten Bahnen zu verlassen und neue Konzepte zu entwickeln, die den analogen Verfahren und deren intuitive Zugänglichkeit zumindest ebenbürtig sind.

Solange die „Mensch-Maschine-Schnittstellen" wie Bildschirme und Tastaturen auch nicht annähernd für einen definierten Anwendungsbereich die Qualität von Papier und Bleistift haben, wird das Papier wesentlicher Informationsträger bleiben. Dennoch kann nicht bestritten werden, dass das Papier nicht mehr in der Form wie noch vor circa 20 Jahren unter der Sichtweise des Verfolgens schriftlicher Überlieferungen betrachtet werden kann. Da die entsprechenden überzeugenden digitalen Alternativen jedoch noch fehlen, geraten wir an manchen Stellen in die Gefahr, die alten, gültigen Mechanismen z. B. für die Nachvollziehbarkeit und Eindeutigkeit und damit der Überlieferung zu verlieren, ohne diese bereits durch geeignete digitale Formen ersetzen zu können. Die vielbeklagte „Informationsflut" ist nicht zuletzt darauf zurückzuführen, dass die

digitalen Produktions- und Kommunikationsmechanismen eine unerhörte Produktivität freigesetzt haben, denen aber kein Äquivalent bei der Verwaltung der Informationen entgegengesetzt wurde. Denn erstens erschienen die analogen Mechanismen ausreichend, solange sie funktionierten, und zweitens waren und sind die angebotenen digitalen Verfahren bezüglich ihrer Funktionalität und Wirtschaftlichkeit nicht immer überzeugend.

Für die Realisierung wirklich überzeugender digitaler Verfahren ist ein Geniestreich nicht in Sicht. Die Anforderungen an Software, die im Umfeld des Informationsmanagement gestellt werden müssen, sind anscheinend widersprüchlich:

Eine mächtige Funktionalität soll sich mit einfachster Handhabung verbinden, und sich dabei ständig neuen Organisationsformen anpassen, ohne veraltete Organisationsstrukturen zu zementieren. Sie soll den Anwender anleiten und steuern, jedoch nicht bevormunden. Sie soll sich an das persönliche Wissens- und Funktionsprofil anpassen und dabei gleichzeitig den Belangen des Daten- und Persönlichkeitsschutzes entsprechen.

Dabei sollte Software trotz der kurzen Innovationszyklen aus Investitionsschutzgründen langlebiger als andere Investitionsgüter sein.

2 Strukturierungsmöglichkeiten

Wie im vorstehenden Kapitel beschrieben, sind die bisherigen Verfahren zur Ordnung und Strukturierung von Informationen sehr nah an den damit klassischerweise verbundenen Medien wie Papier oder Mikrofilm ausgerichtet gewesen. Doch heutzutage wird ein überwiegender Teil der Informationen bereits elektronisch erzeugt. Dementsprechend sind nunmehr Verfahren gefordert, die die Verwaltung von digitalen Objekten zulassen, flexibel auf unterschiedliche Anforderungen eingehen ohne dabei aufwendig in der Pflege oder komplex in der Bedienung zu sein. Eine althergebrachte Methode der Strukturierung in Verzeichnissen, Unterverzeichnissen, Domänen, öffentlichen und geschlossenen Bereichen des Intranets oder der persönlichen Systemumgebung zeigt, wie sehr wir immer noch in gewohnter Weise die uns bekannten Verfahren nutzen, ohne uns bewusst zu werden, welche anderen Möglichkeiten heutzutage gegeben sind.

Während man bei elektronisch erzeugten, ggf. sogar selbst erstellten digitalen Objekten noch mit einfachen Methoden für eine entsprechende neue und flexible Organisationsform sorgen kann, wird spätestens bei der Betrachtung von eingehenden Papier- und Fax-Dokumenten deutlich, dass der Medienbruch zwischen analoger Welt (Papier, Fax, Mikrofilm) und der digitalen Welt der erste Hinderungsgrund ist, eine gemeinsame Basis zur Verwaltung aller Informationen zu schaffen.

Betrachtet man außerdem den hohen zusätzlichen Aufwand, den man betreiben muss, um ein analoges Dokument in ein digitales Objekt zu wandeln, scheint es fraglich, ob diese Mühe überhaupt lohnt. Doch es stehen vielerlei technische Verfahren der automatisierten Dokumenten-Erfassung zur Verfügung, die für unterschiedliche Einsatzbereiche und Informationstypen eingesetzt werden können. Eine Übersicht der zur Verfügung stehenden Methoden soll Hilfestellung bei der Auswahl geben.

2.1 Automatisierte Dokumentenerfassung

Um ein digitales Objekt mit beschreibenden Metadaten zu versehen, werden automatisierte Verfahren eingesetzt, die den Aufwand der Beschreibung minimieren und damit den Wert des digitalen Objektes unmittelbar steigern können.

Während bei manuellen Erfassungsverfahren häufig eine genaue Abwägung von Aufwand und Nutzen zu machen ist, da die Kosten für die Erfassung von Metadaten in Relation zum Wert des digitalen Objektes steht, kann bei automatisierten Erfassungsverfahren der Gesamtprozess auf Maschinenzeit umgelegt werden, so dass die Erfassungskosten signifikant sinken. Dabei ist muss nach wie vor, je nach Art der zu erfassenden Dokumente und Qualität der automatisierten Erfassungsverfahren, die manuelle Nachbearbeitung in eine Kostenbetrachtung mit einbezogen werden.

Bei der automatischen Erfassung unterscheidet man in zwei Arten von Metadaten:

a) Metadaten, die im digitalen Objekt enthalten sind (z. B. Worte, Werte, Fragmente)

b) Metadaten, die das digitale Objekt beschreiben, ohne im digitalen Objekt enthalten zu sein (z. B. Format, Größe, Autor, Quelle, Erstellungsdatum etc.)

2.2 Gewinnung von Metadaten aus den digitalen Objekten

Die einfachste Form der Gewinnung von Metadaten ist die direkte Extraktion aus dem digitalen Objekt. Die bekanntesten Verfahren sind die Verfahren zur automatischen Indizierung von Druckdaten, bei denen

man über feste Positionen Daten direkt aus dem Datenstrom extrahieren kann. Die Systeme orientieren sich dabei an absoluten Positionen auf dem Dokument (z. B. Zeile und Spalte) oder an Schlüsselbegriffen, denen die zu extrahierenden Metadaten folgen. Die Verfahren eignen sich insbesondere bei der Massendatenverarbeitung, da Sie vollkommen automatisiert und fehlerfrei ablaufen können.

Ein Sonderfall sind die heutzutage bereits weit verbreiteten Druckformate wie AFP, IPDS, EMF oder PreScribe, die neben den eigentlichen Nutzdaten auch beschreibende Formulardaten beinhalten. Hier werden zum Teil Konverter in den Hochleistungsdruckern genutzt, um eine möglichst flexible Formulargestaltung während des Druckvorganges zur Verfügung zu stellen. Daraus resultieren aber erhöhte Aufwände in der automatisierten Indizierung, da Nutzdaten von Formular- oder Steuerdaten separiert werden müssen. Somit wird auch zur Gewinnung der Indexdaten häufig ein zusätzlicher Konverter benötigt, um mit den vorstehend beschriebenen Verfahren zu einer automatisierten Übergabe von Schlüsseldaten zu gelangen.

Erhöhter Aufwand ist ebenfalls dort zu erwarten, wo die eigentlichen Metadaten des digitalen Objektes nicht an festen Positionen abzugreifen sind oder nicht in maschinenlesbarer Form vorliegen. Dies ist zum Beispiel bei gescannten Dokumenten der Fall, deren Inhalt zwar augenlesbar ist, die aber keinen tatsächlichen Text mehr beinhalten. Diese Dokumente werden über einen Konvertierungsprozess (OCR-Verfahren Optical Character Recognition – Optische Zeichenerkennung) in maschinenlesbare Schriftinformation umgesetzt, um anschließend automatisiert ausgelesen werden zu können.

Eine gebräuchliche Form sind die am Markt erhältlichen Formular-OCR-Verfahren, die an vordefinierten Stellen eines gescannten Dokuments Werte von Grafik in Text wandeln und ausgeben können.

Da diese Systeme sich an bestimmten Regeln festmachen, wo die Metadaten zu finden sind, reagieren sie sehr unflexibel bei Änderungen des Dokumentenaufbaus. Systeme, die basierend auf Mustererkennungsalgorithmen arbeiten, könnten hier Abhilfe schaffen. Diese Systeme sind in der Lage, unscharf zu suchen und aus dem Umfeld eines Dokumentes auf den zu extrahierenden Wert schließen. Somit sind diese, häufig auf lernenden Verfahren basierten Systeme geeignet, in schwächer strukturierten digitalen Objekten noch Metadaten zu finden. Das System lernt z. B. das in der Nähe einer Bankleitzahl das Wort Bankleitzahl auftritt und das die Schreibweise der achtstelligen Zahl dem Muster XXX XXX XX entspricht. Somit kann mit ausreichender Genauigkeit ein entsprechender Wert gefunden und extrahiert werden. Die Fehlerwahrscheinlichkeit steigt jedoch gegenüber den regelbasierten, bei sehr stark strukturierten Dokumenten zum Einsatz kommenden Verfahren, an.

Typische Einsatzgebiete sind daher Posterfassungssysteme, die eine Versichertennummer, eine Adresse oder eine Bestellnummer extrahieren sollen, die bei der hohen Inhomogenität von Posteingang über regelbasierte Systeme nicht gefunden werden können.

Um die Ungenauigkeiten bei der Umwandlung von gescannten Dokumenten in Zeichen zu umgehen, werden zum Teil wichtige Indexkriterien in maschinenlesbarer Form codiert und auf den Dokumenten aufgebracht. Dabei kommen in der Regel Barcodes zum Einsatz, die industriell genormt sind und mit sehr hoher Genauigkeit ausgelesen werden können. Je nach Barcode-Typ können numerische oder auch alphanumerische Metadaten abgebildet werden. Der hohe Platzbedarf der Strichcodeinformationen sowie der erhöhte Aufwand bei der Erstellung der digitalen Objekte setzt diesen automatischen Verfahren allerdings Grenzen.

Im Bereich von Anträgen finden sich häufig Ankreuzfelder, die mit Hilfe von OMR (Optical Markup Recognition) ausgelesen werden können.

Hier lässt sich aus der Position des Kreuzes eine eindeutige Aussage ableiten, die zur Beschreibung des digitalen Objektes herangezogen werden kann.

Volltext-Systeme werten den Inhalt von textlichen digitalen Objekten in der Form aus, das der beinhaltete Text gegen Stop-Wortlisten abgeglichen wird und die danach als sinntragend erkannten Worte in einer Volltext-Datenbank abgelegt werden. Anschließend können die Dokumente an Hand des eingegebenen Suchwortes aufgefunden werden. Eine typische Anwendung sind zum Beispiel Suchmaschinen im Internet, die den textlichen Inhalt von Web-Sites in Datenbanken vorhalten und bei Eingabe eines Suchwortes Treffer zurückliefern. Weitere Einsatzgebiete sind Pressearchive oder Verfahrensdokumentationen.

Der Vorteil von Volltextsystemen liegt darin, dass sie grundsätzlich ohne eine vorgegebene Metadatenstruktur auskommen und daher sehr flexibel sind. Über die Definition von Stopwortlisten kann verhindert werden, dass die Volltextdatenbank mit Worten aufgefüllt wird, auf die keine sinnvolle Suche möglich erscheint (z. B. Worte mit weniger als vier Buchstaben, Präpositionen, etc.)

Die Schwächen von Volltextsystemen werden dann deutlich, wenn Worte sprachlich mehrfach belegt sind (z. B. die Eingabe des Wortes „Bank“ führt zu Treffern aus den Bereichen Finanzinstitute, Sitzgelegenheiten und Sandanschwemmungen) und damit eine Eingrenzung schwierig wird. Weiterhin scheitern die Systeme daran, dass das eigentliche Suchwort im Text nicht enthalten ist, obwohl der Text sich mit dem gesuchten Thema beschäftigt (z. B. findet das Wort „Straßenbahn“ nicht die Dokumente, die das Wort „Trambahn oder Tram“ benutzen).

Thesaurussysteme sind dazu in der Lage, sprachliche Gleichheiten auszuwerten und damit zu einer höheren Treffergenauigkeit zu kommen. Somit werden alle Anfragen zum Suchwort „Wendel“ automatisch in das

Suchwort „Schnecke“ transformiert und auch hiermit der Bestand durchsucht. Die Pflege von Thesauri-Systemen ist allerdings sehr aufwendig und kann nur in einem bestimmten Umfeld eingerichtet werden, in denen die sprachlichen Entsprechungen klar definiert werden können.

Verfahren, die aus nicht textbasierten digitalen Objekten Metadaten extrahieren, sind in einigen, wenigen Spezialanwendungen zu finden. Auch hier werden Verfahren der Mustererkennung eingesetzt, um z. B. aus Audiodaten bestimmte Worte oder Geräusche zu identifizieren, oder aus einer Sammlung von Bilddaten diejenigen heraus zu suchen, auf denen ein Baum abgebildet ist. Der Rechenaufwand für die Extrahierung der Metadaten sind in diesen Verfahren immens, die Treffergenauigkeit abhängig vom Muster, das gesucht werden kann.

Die oben kurz beschriebenen Verfahren können sich, je nach Anwendungsgebiet, gegenseitig sinnvoll ergänzen. Im Informationsmanagement sind daher alle beschriebenen Verfahren verfügbar, um sie im jeweiligen Kontext der Organisation zielgerichtet einsetzen zu können.

2.3 Gewinnung von Metadaten, die das digitale Objekt beschreiben

Während die vorgenannten Verfahren Metadaten direkt aus dem digitalen Objekt beziehen, werden Metadaten, die das digitale Objekt beschreiben, auf andere Arten gewonnen.

Metadaten wie Verfasser, Erstellungszeitpunkt, Dateiformat, Dateiname, Speicherort, Zugriffsbeschränkungen können über das Dateisystem zur Verwaltung der digitalen Objekte bezogen werden. Inhaltsbeschreibende Daten sind oftmals auch Bestandteil des Dateinamens, der somit auch einen Rückschluss auf den Inhalt zulässt.

Weitere Informationen sind aus dem Aufbau von Dokumenten zu beziehen, so zum Beispiel die verwendete Softwareversion oder die erzeugende Quellapplikation. Mit Hilfe dieser Metadaten können digitale Objekte zusätzlich beschrieben und in Relation gesetzt werden. Ein weiteres, auf Grund der heutigen Rechnerperformance mögliches Verfahren, ist die Bestimmung von Ähnlichkeiten zwischen digitalen Objekten. Hiermit kann bestimmt werden, in wie fern digitale Objekte ähnlich im Sinne von inhaltlich deckungsgleich sind. Diese Verfahren eignen sich zum Beispiel zu Klassifikation von unbekannte digitalen Objekten in Klassen. Die Klassifikationsverfahren ermitteln die Ähnlichkeiten mit bereits bekannten digitalen Objekten und klassifizieren die unbekannten digitalen Objekt bei signifikanten Übereinstimmung als zu dieser Klasse zugehörig.

Moderne Wissens-Management-Systeme nutzen diese Technologien, um dem Benutzer Informationen zur Verfügung zu stellen, die in einem großen Informationsspeicher vorgehalten werden. In der Regel sind diese Informationsspeicher die Fileserver der Unternehmen oder elektronische Archiv- und Dokumenten-Management-Systeme.

Der Benutzer möchte eine möglichst einfache Art, an Informationen zu gelangen. Andererseits soll der Aufwand der Gewinnung von Metadaten klein gehalten werden. Insbesondere vor dem Hintergrund der schnell wechselnden Interessenlagen ist generell zu klären, in wie weit eine vorbestimmte Definition von Metadaten überhaupt sinnvoll ist. Sind Wissens-Management Systeme dazu in der Lage, digitale Objekte inhaltlich ausreichend nach unterschiedlichen Sichtweisen zu klassifizieren, kann ggf. ganz auf eine vorbereitende Erzeugung von Metadaten verzichtet werden.

2.4 Wissens-Management-Systeme verwenden keine Metadaten mehr

Das Dilemma, dass „nie die richtigen Informationen zum richtigen Zeitpunkt" zur Verfügung stehen, kennt jeder PC-Benutzer, der sich in einem zuvor festgelegten Ordnungsschema eines Datei-, Dokumenten-Management- oder Content-Management-Systems auf der Suche befindet. Das, was man sucht, findet man oft nicht oder erst nach langer und aufwendiger Suche. Das Problem liegt u. a. darin begründet, dass Ordnungsschemata, die zur Klassifizierung und Ablage von Informationen dienen sollen, veralten, und zudem Informationssuchende individuelle Interessenprofile haben, die durch ein globales Ordnungsschema nur unzureichend bzgl. spezifischer Bedürfnisse abgebildet werden können.

Was liegt folglich näher, als Informationen zeitnäher und nach individuellen Bedürfnissen zu klassifizieren? Moderne Softwarewerkzeuge ermöglichen einen solch neuartigen Umgang mit Informationen: Diese müssen nicht mehr manuell klassifiziert und archiviert werden, sondern können automatisch entsprechend individueller Interessenprofile und nach jeweils aktuellen Ordnungsschemata zugeordnet und in unstrukturierten Datenbeständen ,,fehlertolerant" recherchiert und letztlich (wieder-)gefunden werden.

Eine besondere Herausforderung bei der Konstruktion des ,,richtigen" Ordnungsschemas stellt die Dynamik der Informationsbestände und ihres jeweiligen Anwendungskontextes dar:

Ordnungsschemata verändern sich über die Zeit entsprechend der Interessen der Informationssuchenden und -verarbeitenden Personen.

Das Retrieval, also das Wiederauffinden der ,,richtigen" Informationen, gestaltet sich trotz der vorgegebenen Ordnungsschemata oft schwierig. Das inhärente Problem kennt jeder, der beispielsweise über einen län-

geren Zeitraum ein Dateisystem als ,,Austauschplattform" für die Dokumente einer Arbeitsgruppe organisiert oder in einem Archivbestand, der nach einer überholten Aktensystematik abgelegt ist, Informationen gesucht hat: Ordnungsschemata "veralten" mit zunehmender Lebensdauer bzw. verändern sich. Da große Informationsbestände auch nicht ,,mal eben schnell" manuell reorganisiert werden können, nahm man in der Vergangenheit von einer Anpassung, z. B. einer Reorganisation des Dateisystems, meist Abstand. An die Anpassung von Schemata an die Interessen einzelner Informationssuchender war dabei schon gar nicht zu denken.

Gerade der letzte Punkt, also die Ordnung von Informationen nach individuellen Bedürfnissen (,,persönliche Wissensportale"), stellt eine interessante Perspektive in unserer informationsüberfluteten Zeit dar. Hier ermöglichen es automatische Verfahren, sich von vorgegebenen starrer Ordnungsschemata zu lösen und statt dessen problembezogene Informationsfilter aufzubauen, die dem Informationssuchenden schnell eine individuelle Sicht auf Informationen bieten. Durch die Automatisierung der Klassifizierung wird es zudem möglich, dass Informationen aus dem täglichen Strom automatisch gefiltert oder, wenn man einem solchen System doch nicht ganz „trauen" will, zumindest nach Interessenlage gewichtet werden. Dadurch erreichen uns zuerst die Informationen (Push-Technologie), die auch tatsächlich unseren Interessen entsprechen.

2.5 Strukturierung von Information (und Wissen)

2.5.1 Auswertung und Nutzung von Informationen

Ein Ziel in der Informationswirtschaft ist die Bereitstellung von Verfahren und Methoden zur Unterstützung einer effektiven Erreichung eines bestimmten Geschäftszwecks. Eine Kernaufgabe ist dabei die Sicherstellung eines effizienten Informations-Austauschs zwischen allen betroffe-

nen Stellen (Organisationen, Firmen, Sachbearbeitern, Entscheidern, ...). Hierfür müssen die Informationen in Informationseinheiten (Objekten) strukturiert werden und der Austausch muss nach vorgangsspezifischen Regeln erfolgen.

Wir befassen uns an dieser Stelle zunächst mit den nachträglichen Strukturierungsmöglichkeiten von empfangenen unstrukturierten Informationen, Nachrichten und Dokumenten (digitale Objekte).

In den vorangehenden Abschnitten wurden verschiedene Verfahren beschrieben, mit denen Informationen aus Fremdsystemen, beim Empfänger strukturiert werden können. Das Ziel der nachträglichen Strukturierung besteht darin, entweder Nutz-Daten für die Weiterverarbeitung (Buchungssatz einer Rechnung erfassen) oder Meta-Daten für die Indexierung (Retrieval in einem DMS) aus den digitalen Objekten zu extrahieren und diese Daten soweit möglich und sinnvoll um zusätzliche bereits vorhandene Informationen (aus ERP, CRM, ...) zu ergänzen.

2.5.2 „Klassische“ digitale Objekte

Bei den Betrachtungen wurde davon ausgegangen, dass in den „klassischen“ digitalen Objekten (aus Textverarbeitung, Scan/OCR, EDV-Programmen) die Nutzdaten (Inhalte: Zahlen, Begriffe) mit den Auszeichnungen (Feldnamen, Tabellenbeschriftungen: Text) in einer einzigen binären Datei vermischt sind. Für die Analyse und Auswertung zur Strukturierung steht also ausschließlich diese eine Datei zur Verfügung. Für einen menschlichen Leser – für den diese „klassische“ Darstellungsart von Dokumenten ursprünglich eingeführt wurde - ist die manuelle „Nutzung“ der Information kein Problem. Für ein Computer-Programm ist die Aufgabe sehr komplex und ohne die Informations-Struktur exakt zu erkennen, nur begrenzt möglich.

Der Grad einer erreichbaren automatisierten Auswertung und Nutzung ist daher unterschiedlich und abhängig von der jeweiligen Komplexität der digitalen Objekte und den „beschaffbaren“ Zusatz-Informationen. Im Idealfall können alle erforderlichen Daten/Informationen extrahiert werden, um einen Vorgang unmittelbar auszulösen (Buchung, Überweisung). In anderen Fällen ist dies wegen zu hoher Komplexität und fehlender Zusatzinformationen zu aufwendig.

2.5.3 Klassifikation der Nachrichten

Es wurde aufgezeichnet, dass bei bestimmten Informationsarten und in bestimmten Anwendungsbereichen der Aufwand zur Gewinnung von Nutz- und Meta-Daten sehr hoch sein kann. In diesen Fällen ist zu prüfen, ob auf die Gewinnung der Metadaten gänzlich verzichtet werden kann und dafür heute verfügbare stochastische Klassifikatoren eingesetzt werden können.

Eine Abgrenzung von Einsatzgebieten in denen „regelbasierte“ oder „stochastische“ Klassifikations-Verfahren eingesetzt werden sollten, ist nur schwer möglich. Bei bestimmten Aufgabenstellungen kann es sinnvoll sein, beide Methoden zu kombinieren.

Nachfolgend werden Anhaltspunkte dafür gegeben, wann welche Methode sinnvoll eingesetzt werden sollte.

Wann strukturieren?

Eine nachträgliche Strukturierung von empfangenen nicht strukturierten Informationen (und der Einsatz von regelbasierten-Systemen) ist häufig dann sinnvoll, wenn die digitalen Informationen eindeutig identifizierbar sind (z. B. Rechnungsbelege, Bestellung, ...) und wenn der Inhalt (Daten) nach bekannten festgelegten Kriterien genutzt und weiter verarbeitet werden soll.

Die Aufgaben einer nachträglichen Strukturierung werden in der Zukunft an Bedeutung verlieren, da durch den Einsatz von Dokumenten-Standards wie EDI, Edifact und insbesondere XML die digitalen Informationen grundsätzlich strukturiert sein werden. Wegen der wachsenden Bedeutung von XML und seiner Verbreitung im Internet wird dieses Thema in einem separaten Abschnitt behandelt.

Wann nicht strukturieren?

Auf die Strukturierung kann und wurde bisher verzichtet, wenn die Informationen zunächst nicht für einen bestimmten Kontext erstellt oder abgelegt werden, wenn also die spätere Verwendung unbekannt ist oder nur ungenau oder nur mit sehr hohem Aufwand beschrieben werden kann.

Die bisherigen Ausführungen und Überlegungen beziehen sich auf die anfangs definierte Zielsetzung in der Informationswirtschaft. Darüber hinaus besteht Bedarf zur Nutzung von Informationen, die über die Automatisierung von Geschäftsprozessen hinausgeht. Nachfolgend werden daher weitergehende Aspekte zur Strukturierung dargestellt, die im Zusammenhang mit der Wissenslogistik stehen.

2.6 Weitergehende Ziele der Wissenslogistik

Ziel der Wissenslogistik ist die Unterstützung eines „intelligenten Verhalten“, d. h. die Unterstützung bei der Anwendung von Wissen, welches in einer bestimmten Situation gewonnen wurde, um Problemstellungen in einer anderen Situation zu lösen. Hierbei geht es im Wesentlichen darum, Ungenauigkeiten der natürlichen Sprache zu präzisieren, in dem eine verbindliche Verwendung der Sprache vereinbart und kontrolliert wird. Dies ist von Bedeutung, da der Autor und der Nutzer einer Information „normalerweise“ unterschiedliche Begriffe für vergleichbare Sach-

verhalte verwendet oder die Information in einer anderen Landessprache verfasst. Z.Tl. wurde bereits im vorangehenden Abschnitt hierauf hingewiesen.

Um ausgetauschte Informationen und das darin enthaltene Wissen im Hinblick sinnvoll nutzen zu können, müssen folgende Voraussetzungen geschaffen werden:

- Sinnvolles Ordnen von Informationen (Ordnungssysteme)
- Zweckmäßiges Aufbereiten des Wissens (Wissensdarstellung)
- Anpassung des Wissens bei Änderungen (Change-Management)

Daraus ergeben sich zwei zusätzliche Aufgaben für die Strukturierung von Informationen und Wissen:

Festlegung von Macro-Strukturen (=Ordnungssystem):

Für ein sinnvolles Ordnen werden die Informations-Objekte nach Ordnungskriterien gruppiert. Beispiele für Ordnungssysteme sind u. a.:

- Dezimal-Klassifikation (z. B. UDC)
- Lenoch-Klassifikation (Nach Fachgebieten)
- Thesaurus
- Semantisches Netz

Während die Dezimal-Klassifikation, die Lenoch-Klassifikation und der Thesaurus klassische Ordnungssysteme darstellen, sind semantische Netze bzw. semantische Datenbanken weniger weit verbreitet. Sie bieten z. B. durch die Einführung und Verwaltung einer metrischen „Distanz“ zwischen Begriffen eine weitergehende Charakterisierungsmög-

lichkeit als „Ober-, Unter- und benachbarter Begriff“ in den klassischen Ordnungssystemen.

Festlegung von Micro-Strukturen (=Begriffssystem):

Für die Zuordnung von Objekten zu bestimmten „Positionen“ innerhalb eines Begriffssystems müssen Begriffe mit zugeordneten Benennungen festgelegt werden. Dies erfolgt i.A. in

- Terminologie-Datenbanken (z. B. TermDat)
- Lemma-Dateien (= Eintrag oder Stichwort in einem Lexikon)

Die Leistungsfähigkeit und Flexibilität einer Terminologie-Datenbank wird im Wesentlichen durch den „terminologischen Eintrag“ bestimmt. Die Fähigkeit zur Verwaltung von Homonymen, Synonymen sowie der Mehrsprachigkeit werden hier festgelegt.

Verbindung von Macro- und Micro-Struktur:

Die Leistungsfähigkeit eines Objekt-Verwaltungssystems (Information oder Wissen) ergibt sich durch sinnvolle Kopplung eines geeigneten Ordnungssystems mit einem Begriffsystem: z. B. eines semantischen Netzes bzw. Datenbank mit einer Terminologie-Datenbank.

Software-Programme, die bei der Verwaltung von Informationen (Wissen) die Leistungsmöglichkeiten der Micro- und Macro-Strukturen nutzen, müssen bei der Inhaltsanalyse eines Dokumentes die s.g. Lemmatisierung, die Rückführung auf ein Lemma unterstützen.

Das hier aufgezeigte Thema ist Teilgebiet einer eigenständigen Wissenschaft, deren Erkenntnisse nur selten in den heute verfügbaren DMS-Produkten genutzt werden. Einige der hier aufgeführten Begriffe werden im Anhang zusammen mit weiterführenden Begriffe erläutert.

Objektorientierung (OO-Technologie)

Es wurde an verschiedenen Stellen der Begriff „Objekt“ und der Objektorientirerung verwendet. Obwohl die objektorientierte Sichtweise mehr als 25 Jahre alt ist, wird der Aspekt nachfolgend im Zusammenhang mit der „Strukturierungsmöglichkeit“ von Information (Wissen) noch einmal aufgegriffen.

Die Anwendung der Objektorientierung im DMS führt zur Strukturierung nach Objekt-Klassen, -Typen, -Variablen und -Interfaces. Durch das damit verbundene Prinzip des „Encapsulation“ wird die Unabgängigkeit der Informationen von Anwendungen erreicht und die einmal erfassten und verwalteten Informations- und Wissenseinheiten können von unterschiedlichen Anwendungen genutzt werden. Die Anwendungen können über die digitalen Objekte und deren Eigenschaften integriert werden.

Im Aufbau eines digitalen Objektes sollten u. a. die folgenden Komponenten enthalten sein. Es ist evident, dass die Wahl der Komponenten die Leistungsfähigkeit bei der Nutzung des Objektes und damit des Gesamtsystems bestimmt.

Object <name>

Class variables	Variable, die zu einer Objektklasse gehören: Vererbbarkeit
Instance variable	Spezielle Variable, die ein bestimmtes Objekt betreffen
Procedures	Regel für eine spezifische Bearbeitung des Objektes
Content	aktueller Inhalt des Objektes: Text, Daten, Bild, Ton, ...

Beispiele für Wissensdarstellung:

Klasse	Person	Stadt	Dokument
Name	Peter	München	KfZ-Brief
Variable	Größe	Einwohnerzahl	Gültigkeit
	Alter	Sprache	Rechte
	Geburtsort	Land	
Regel	Alter < 0	Nation = GB	EkSt-Gesetz
	=ungeboren;	Linksverkehr	Kreditregeln
	jedes Jahr		Anweisungen
	1 Jahr älter		
Content	Foto	Stadtplan	Text, Signatur
Interface	PAISY	XY-DB	DOXiS

Die Flexibilität bei der Nutzung eines entsprechend strukturierten Informations-Pools ergibt sich aus der Möglichkeit, die unterschiedlichen Objekte fallweise zu verknüpfen.

Die Ausprägung der Objektorientierung war zum Ende der 80iger und Beginn der 90iger Jahre häufig noch proprietär. Danach wurden verschiedene Formate in DIN (ODA/ODIF, EDI) und ISO (SGML) standardisiert.

Durch die Verbreitung des Internet wurden zunächst Formate wie HTML und nach 1998 das Format XML zum defacto-Standard.

Wegen der wachsenden Bedeutung von XML für die Dokumenten-Erstellung und Verarbeitung sowie der Präsentation sowie als der Basistechnologie für zukünftige „WebServices“ werden nachfolgend einige Aspekte dieses Standards dargestellt. Da das Thema sehr umfassend ist, kann die nachfolgende Darstellung nur einen ersten Einblick vermitteln. Der interessierte Leser wird an dieser Stelle auf die einschlägige Literatur verwiesen.

XML-Technologie

XML (eXtensible Markup Language) ist die Weiterentwicklung des bekannten HTML (HyperText Markup Language).

Der Ausgangspunkt von XML ist die Textverarbeitung, einer Softwareanwendung in der Text und die Formatierung in einer einzigen Datei integriert erstellt, verwaltet, gespeichert und auf Anforderung angezeigt oder gedruckt wird. Typische Vertreter dieser klassischen Dokumentenformate sind z. B. RTF oder Word. WYSIWYG bezeichnet das zugrunde liegende Konzept, das ursprünglich revolutionär war – aus heutiger Sicht des modernen XML müsste die Abkürzung etwas spöttich umgeschrieben werden: WYSIAYG: what you see is ALL you get. Man erhält also nur die Darstellung einer Information – und genau das ist das Problem beim DMS, das mit diesen klassischen Formaten arbeitet.

Durch die Abstraktion der Darstellungsformen wird die Vermischung von Daten/Text und Darstellungsinformationen vermieden. Datentypen können dadurch vielartiger weise einbezogen werden. SUN spricht daher im Zusammenhang mit XML auch weniger von Dokumenten sondern von „Portable Data“ oder „Portable Code“. Umgangssprachlich kann man dies übersetzen mit: XML führt zur „Verdokumentung von Daten“ oder „Verdatenbankung von Dokumenten“.

XML ähnelt mehr einer relationalen und objektorientierten Datenbank als einem Dokument aus einer klassischen Textverarbeitung. Es können sowohl Dokumente als auch Datenformen behandelt werden. Zusätzlich ist die Sprache durch anwendungsbezogenes Vokabular individuell erweiterbar. Dies ist ein Satz von Elementtypen wie z. B. /from, /to, /re, Typen, die zur Charakterisierung eines Document-Types „Nachricht“ sinnvoll sind.

XML-Dokumente sind Dokumente, die von Computer-SW für Computer-SW erstellt werden und die durch die interne Codierung auch von Men-

schen gelesen werden können. XML ist ein Baustein für die nächste Generation von Web-Anwendungen, die als Web-Services veröffentlicht werden können.

Im Zusammenhang mit XML sind folgende Komponenten zu betrachten, auf die im Rahmen dieser Übersicht jedoch nicht eingegangen wird:

- XML-Document (Text, Auszeichnung, Attribute)
- DTD (Document-Type-Definition)
- Schema (Verallgemeinerte DTD, selbst ein XML-Dokument)
- XSL, XSLT (StyleSheet, Translator)
- XML-FO (XML Formating Objects)
- RDF (Resource Description Framework)
- PICS, CDF (Metadaten-Beschreibung, basierend auf RDF)

XML wurde ausgehend von GML über SGML und HTML in den zurückliegenden 30 Jahren ständig weiter entwickelt und 1998 als Standard von W3C verabschiedet.

Eine kurze Übersicht über die XML-Historie:

1969	MGL, der Ursprung
1974	SGML, das Geburtsjahr
1986	SGML Standard ISO8879
	HyTime Darstellung von Hyperlink-Eigenschaften
	DSSSL Erzeugung von StyleSheets
1989	HTML
	CSS cascading style sheet: Notbehelf
1998	XML, Standard
	Xlink (aus HyTime)
	XSL (aus CCS und DSSSL)

In den Jahren nach der Veröffentlichung der Spezifikation hat neben der Java-Plattform keine andere Technologie einen stärkeren Einfluss auf die SW-Entwicklung gehabt als XML. Die Spezifikationen beschreiben

Regeln zur Konstruktion wohlgeformter XML-Dokumente. Sie bestehen – sehr vereinfacht dargestellt – im wesentlichen aus einer Menge geschachtelter Auszeichungswörter (Tags), die über Attribute verfügen und Textdaten einschließen können.

Ein wesentlicher Vorteil gegenüber „klassischen“ Dokumentenformaten besteht in der Möglichkeit zur Validierung ihrer Korrektheit. Mit der Spezifikation von XML wurden gleichzeitig Document-Type-Definitions (DTDs) eingeführt, die die Struktur eines Dokumentes definieren, und anhand derer automatisiert die Korrektheit eines Dokumentes geprüft werden kann. DTDs sind nicht als XML-Dokumente formuliert und lassen sich daher nicht mit Hilfe eines XML-Parsers validieren. Aufgrund dieser Einschränkung und zunehmend erkennbarer Unflexibilität des Konzeptes der DTDs wird das Konzept der XML-Schemata diskutiert und eingeführt.

XML-Schemata beschreiben wie DTDs die logische Struktur eines XML-Dokuments und stellen ihrerseits selbst XML-Dokumente dar. Sie können zusätzlich Informationen zur Typisierung des Inhalts enthalten.

XML und DTD sowie Schemata beinhalten keine Formatierungs-Informationen. Diese Informationen werden separat in StyleSheets gespeichert. Die Definitionen werden in einer eigenen Syntax, in der XSL Sprache formatiert. Die in XSL formulierten StyleSheets werden von XSLT Transformers auf Anforderung verarbeitet und es werden entsprechende Darstellungen des XML-Dokumentes erzeugt. So können z. B. EPS, PDF, TIF, ... Darstellungen generiert werden.

Seit kurzem ist die Art der Formatierung ebenfalls standardisiert. Dies ist in den XML-FO – Formatting Objects spezifiziert. XML-Dokument und die DTD bzw. Schemata definieren die Information und deren „Bedeutung“; FO definiert das exakte Layout, d.h. die Darstellung beim Empfänger.

Um auch Meta-Daten zu standardisieren wurde das „Resource-Description-Framework“ (RDF) eingeführt. Es handelt sich hierbei um einen Dokumenten-Type für Meta-Daten, der wie XML erweiterbar ist. Microsoft hat auf dieser Basis das „Channel-Definition-Format“ entwickelt; Netscape hat das „Meta-Content-Framework“ entwickelt.

Damit ist die Entwicklung von XML aber noch nicht abgeschlossen. Obwohl die Basistechnologien in wissenschaftlichen Instituten entwickelt wurden, sind die Spezifikationen für die Sprachen MathML und CML noch in Arbeit.

Weitere Zielsetzung ist die umfassende Abbildung von Dokumenten, mit der das Speichern, Bearbeiten, Durchsuchen, Übertragen, Anzeigen und Drucken automatisiert möglich ist. Hierzu müssen noch einige Punkte wie Attribute (z. B. cia-record-element für Sicherheitsstufen), externe HyperLinks (z. B. Xlink), Zeichendaten (z. B. UniCode, UTF8, ...) standardisiert werden.

Zusammenfassung und Ausblick

XML ist gleichzeitig ein Tool für Daten und Dokumente. Daher können mit XML sehr unterschiedliche Bereiche unterstützt werden. „POP”: Presentation Oriented Publishing (Content-Management) und “MOM”: Message Oriented Middleware. MOM bricht die klassische Verbindung zwischen Client und Server auf und die Middleware übernimmt die Funktion der Datenintegration aus unterschiedlichen Anwendungen.

Abschießend wird ein Ausblick auf das SOAP-Protokoll gegeben, mit dem verteilte Anwendungen kommunizieren und untereinander Dokumente und Nachrichten austauschen.

“SOAP is a light weight protocol for exchange of information in a decentralized, distributed environment”. Dieses Zitat aus dem Abstract der W3C Note zu SOAP, dem „Simple Object Access Protocol“, erklärt die Bedeutung des Protokolls für moderne DMS Projekte. Basierend auf XML und XML-Schemata bietet SOAP eine standardisierte und plattformunabhängige Möglichkeit, zwischen den einzelnen Komponenten des Systems Daten/Dokumente auszutauschen und RPC auszuführen.

SOAP besteht aus drei Teilen: Das SOAP-Envelope-Konstrukt definiert einen umschließenden Rahmen um auszudrücken, was eine Nachricht ist, wer sie zu behandeln hat und ob sie optional oder notwendig ist. Die SOAP-Kodierungsregeln definieren einen Seriaisierungsmechanismus, der benutzt werden kann, um Instanzen von anwendungsspezifischen Datentypen auszutauschen. Die SOAP-RPC-Repräsentation definiert eine Konvention, die genutzt werden kann, um Remote Procedure Calls und ihre Antworten zu repräsentieren.

Als ebenfalls plattformunabhängige Alternative zu SOAP existiert die Common Request Broker Architecture (CORBA), die als industrieller Standard etabliert ist. Im Gegensatz zu CORBA benötigt SOAP jedoch keine weiteren Programme, um aus einer Interface Definition Language (IDL) »Skeletons« und »Stubbs« zu generieren, die danach mit Funktionalität gefüllt werden. Eine Entscheidung für SOAP führt nicht zu einer Insellösung, da in W3C bereits ein CORBA/SOAP-Gateway diskutiert wird.

Als entscheidenden Nachteil eines plattformunabhängigen Protokolls wird allerdings eine Vergrößerung der übertragenen Datenmenge in Kauf genommen. Beispielsweise müssen Binärdaten in SOAP Base64-kodiert übertragen werden, was einen Größenzuwachs um 33 % entspricht.

Das W^3C verfolgt die Weiterentwicklung von SOAP seit September 2000 im Rahmen der »XML Protocol Activity.

3 Ganzheitlicher Ansatz

3.1 Strategien

Die Zeit der überwiegend papiergebundenen Informationsübermittlung mit Unterstützung unterschiedlicher Dokumente liegt in noch nicht allzu weiter Ferne. Lediglich Telefon bzw. Fax unterstützten die schnelle Informationsübermittlung. Oft jedoch war eine nachträgliche Schriftform erforderlich, wenn die Information rechtliche Beweiskraft erlangen sollte.

Auch bei der Datenverarbeitung durch Großrechner wurde das Ergebnis die verarbeitete Information, wieder auf Papier ausgegeben. Dieses ist vielfach bis heute der Stand der Dinge.

Die Weiterleitung und Bearbeitung von Vorgängen mittels Papierdokumenten ist zeit- und arbeitsaufwendig. Das für die Bearbeitung erzeugte und benötigte Papier diente bisher jeweils auch der rechtlichen Beweiskraft. Nach Beendigung der Vorgangsbearbeitung nach den vorgegebenen Ordnungskriterien wird der Datenträger Papier unter Beachtung der jeweiligen gesetzlichen Aufbewahrungsfristen abgelegt. Raum- und Sachmittel inklusive Personalkosten für Abheften und Suchzeiten erzeugen bei diesen Verfahren einen erheblichen Kostenblock.

Auch die analogen Medien und der Mikrofilm haben an dieser Stelle keine bahnbrechend Änderung herbeiführen können. Sie waren eher Mittel zum Zweck, wenn es darum ging, überfüllte Registraturen zu entlasten. Die „Archivierung“ auf Mikrofilm erfolgte erst dann, wenn die Vorgangsbearbeitung mit der Papierakte abgeschlossen war und die Dokumente gemäß Aufbewahrungspflichten zur Ablage gelangten.

Anfangs leisteten auch die „elektronischen Archivsysteme“ nicht wesentlich mehr. Die angebotenen Systeme waren zudem sehr teuer und personalaufwendig, denn für die Aufrechterhaltung eines reibungslosen Betriebes sind Administratoren erforderlich, die den technischen und organisatorischen Ablauf überwachen und gewährleisten.

Mit der einer rasanten Entwicklung der Elektronik, Kommunikationstechnik und den digitalen Medien entstand der Wunsch nach papierloser bzw. papierärmerer Sachbearbeitung. Die digitalen Techniken steigerten jedoch die Produktivität bei der Erzeugung, Kommunikation und Vervielfältigung von layout- und damit papierorientierter Information in einem solchen Maß, dass Neben der digitalen Informationsflut, auch der papiergetragene Informationsaustausch erheblich zunahm. Die strukturierte Informationsgewinnung und Bearbeitung wird zum heutigen Tag dadurch eher erschwert.

4 Der prozessorientierte Ansatz

Informationsprozesse laufen nicht immer nur streng reglementiert ab. Gerade infolge der Verbreitung von Personalcomputern im LAN, dem Anwachsen von Speicherplatz und durch die generelle Verfügbarkeit von Kommunikationsmechanismen wie E-Mail wird umfangreiche und unstrukturierte Kommunikation in einem nie gekannten Ausmaß erleichtert. Inwieweit müssen, sollen oder können diese Informationsprozesse nachvollzogen werden.

4.1 Der Ablauf von Prozessen

Für alle Prozesse in einem Unternehmen wie z. B. Entwicklung, Beschaffung, Produktion, Information, Zielfindung usw. werden Dokumente benötigt Die Erzeugung und Bearbeitung dieser Dokumente wollen wir hier betrachten.

Grundbausteine des Dokumentenmanagements sind die Erstellung, die Erfassung, die Bearbeitung, die Bereitstellung und die Archivierung. Jedes Dokument hat einen eigenen Lebenslauf. Betrachtet man die Erstellung der Dokumente, so erfolgt diese heute schon fast ausschließlich digital. In der Bearbeitungsphase werden sie gedruckt, mit Unterschriften versehen und weitergegeben. Dieser Medienbruch ist heute (leider) noch die Regel.

4.2 Störfaktoren

Wie in jedem Prozess treten auch in Dokumentenbearbeitungsprozessen häufig Störungen auf. Meist handelt es sich um folgende Faktoren:

- Der Prozess dauert zu lange

- Die Erfassung (Digitalisierung) dauert zu lange
- Die Indexierung dauert zu lange
- Die inhaltliche Erschließung der Dokumente ist zu aufwendig
- Die Liegezeiten der Dokumente sind zu hoch
- Dokumente werden nicht gefunden
- Die gefundenen Dokumente sind nicht lesbar

Bei der Entwicklung eines Konzeptes, zur Neugestaltung bzw. nachhaltigen Verbesserung von Prozessen müssen als erstes mögliche Störfaktoren zu ermittelt werden.

Wenn ein Prozess zu lange dauert, kann die Prozessgestaltung daran maßgeblich schuld sein. Ein sequentiell ablaufender Prozess wird immer länger dauern als ein Prozess mit parallelen Passagen. Ein Prozess, in dem analoge Dokumente per Post weitergereicht werden, wird auch immer langsamer ablaufen, als ein Prozess in dem die Dokumente in digitaler Form weitergeleitet werden.

Der Durchfluss der Dokumente muss mit allen Komponenten von der manuellen Vorbereitung des Schriftgutes (Heftklammern entfernen, Papier glätten) bis zu Kapazitäten von Scannern und Rechnern geplant werden.

Liegezeiten ergeben sich auch dann, wenn Informationen, die für die Bearbeitung eines Vorganges erforderlich sind, nicht verfügbar sind. Nicht selten sind die Liegezeiten eines Vorganges länger, als die Bearbeitungszeiten.

Gelingt es, Liegezeiten durch die Veränderung von Prozessen zu reduzieren (z. B. Parallelbearbeitung), so kann man damit auch die Gesamtlaufzeit des Prozesses wesentlich reduzieren.

Ein großes Risiko stellt auch of die Vorgabe durch die eingesetzte Software dar. Nur zu häufig wird Software nach individuellen Wünschen umgestaltet, weil man bestehenden Prozesse besser abbilden will. Wenn diese Modifikationen nicht updatefähig sind, müssen alle Konfigurationsschritte erneut vorgenommen werden und oftmals bleibt ein Risiko, dass die Prozesse nur noch in Teilbereichen oder eingeschränkt funktionieren.

Neben der Flexibilität der Software ist für die organisationsfreundliche Gestaltung der Abläufe auch eine ordnungsgemäße Organisationsanalyse erforderliche, die die abzubildenden Prozesse auf das wirklich notwendige Maß reduziert und verschlankt.

4.3 Sollkonzept

Selbst digitalisierte Informationsprozesse sind heute oft schon nicht in der Lage, die erforderliche Menge von Dokumenten bedarfsgerecht aufzubereiten und in vereinbarter Form termingerecht bereitzustellen.

Wir müssen unsere Prozesse weiter optimieren und unter Umständen auch weitergehend automatisieren.

Zur Herstellung von Prozessfähigkeit und Wirtschaftlichkeit ist oft auch die grundsätzliche Neuplanung der Prozesse erforderlich. Dabei wird der gesamte Prozess zur Diskussion gestellt. Nicht die Frage nach der Optimierung des bestehenden Prozesses steht im Vordergrund, sondern es wird überlegt, warum der Prozess überhaupt erforderlich ist und wie er idealer Weise gestaltet sein müsste.

Dokumentenmanagementprozesse sind immer als Dienst für übergeordnete Geschäftsprozesse zu betrachten. Der Dienst, den sie leisten orientiert, sich an den übergeordneten Prozessen.

Wenn man also von prozessorientiertem Dokumentenmanagement spricht, so heißt das, dass der Dokumentenmanagementprozess eine Ebene unter dem zugehörigen Geschäftsprozess liegt.

Als Dienst ist er nur in Ausnahmefällen Kern- oder Schlüsselprozess des Unternehmens (Ausnahme: Reproduktionsbetriebe). Er richtet sich aber an diesen Kern- Schlüsselprozessen aus.

4.3.1 Die Strukturierbarkeit von Prozessen

Starr ablaufende Prozesse bewirken, dass das Ergebnis eines Prozesses nicht immer optimal ist. Wenn beispielsweise bei der Lackierung von Fahrzeugen nach der Grundierung festgestellt wird, dass die Grundierung Fehler aufweist, macht es keinen Sinn, auch noch den Decklack aufzubringen.

Wenn es aber nicht möglich ist, das Fahrzeug aus dem Lackierprozess herauszuschleusen, kann der Prozess nicht ordnungsgemäß ablaufen.

Gleiches gilt für Dokumentenmanagementprozesse. Auch hier gilt es, Prozesse so zu strukturieren, dass sie fehlertolerant sind. Dies kann bedeuten, dass komplexe Informationsprozesse in elementare Segmente aufgeteilt werden müssen. Es kann aber auch erforderlich sein, Einzelprozesse so zu vernetzen, dass der Informationsfluss variabel gestaltet werden kann.

Es muss außerdem die richtige Abstufung zwischen statischem Workflow und interaktivem Prozess gefunden werden.

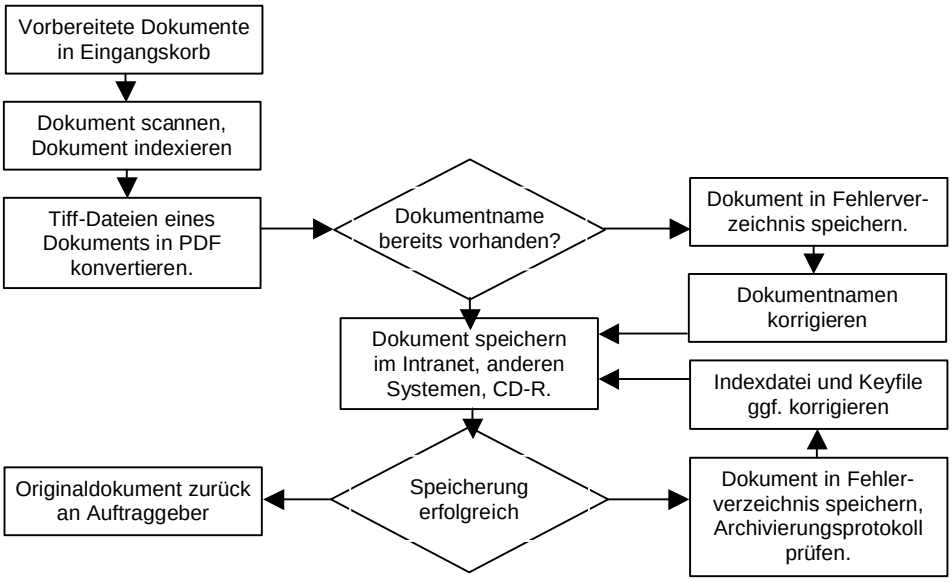
Es ist durchaus möglich, Teilprozesse als Workflow statisch zu modellieren, um dann zwischen zwei statischen Teilen einen interaktiven Teilprozess einzubinden.

Bei der Modellierung der Prozessstruktur ist es dafür als erstes erforderlich, die Schlüsselprozesse zu identifizieren.

Danach erstellt man mit diesen Schlüsselprozessen einen Prozessgliederunsplan, in dem alle für das jeweilige Unternehmensziel relevanten Schlüsselprozesse enthalten sind. Diese Schlüsselprozesse müssen dann in Teilprozesse, die auf anderen Ebenen ablaufen, heruntergebrochen werden. Diese Teilprozesse münden nach Ablauf dann wieder in die höhere Ebene.

Durch die Anordnung der einzelnen Teilprozesse im Gesamtprozess ergibt sich dann der neu gestaltete Prozess. Eine Prozessvisualisierung in Form eines Flussdiagramms kann erforderlich sein.

Beispiel:



Nun kann man beginnen, die einzelnen Teilprozesse aufzubrechen, um diese dann gegebenenfalls ebenfalls zu strukturieren. Nachdem die Struktur feststeht, muss diese dann ebenfalls wieder in einem Flussdiagramm abgebildet werden

4.4 Automatisierung von Prozessen

Eine Automatisierung von Prozessen oder von Prozessschritten ist immer dann sinnvoll, wenn große Dokumentmengen immer wiederkehrend und auf die gleiche Weise erfasst, bearbeitet und gespeichert werden müssen. Ziel ist es hier, Dokumentenmanagementprozesse so zu modellieren, dass sie in vorgegebenen Schritten immer wiederkehrend ablaufen können.

Ein Weg dazu ist die Automatisierung von Teilprozessen, wie das Auslesen von Metadaten für die Indexierung der Dokumente, oder das Auslesen von Indexdateien für die Speicherung von Dokumenten mittels eines Speicherassistenten. Die gewonnenen Daten lösen dabei automatisch Aktionen aus, die sonst manuell von einem Sachbearbeiter ausgelöst werden müssen.

Von der Modellierung eines Workflow bis zur Automatisierung ist ein weiter Weg.

Einzelne Bearbeitungsschritte können so miteinander verknüpft werden, dass Dokumente von Sachbearbeiter zu Sachbearbeiter zielgerichtet auf digitalen Wegen weitergeleitet werden.

Im Workflow kann man komplexe Prozesse so transparent abbilden, dass jederzeit ermittelt werden kann, wo sich bestimmte Vorgänge oder Aufträge befinden. Treten an bestimmten Stellen Staus auf, ist es möglich, die Ursache herauszufinden und gezielt gegenzusteuern.

Um einen Workflow auch in kritischen Situationen funktionsfähig zu erhalten, empfiehlt es sich, an kritischen Stellen (bottleneck, Problemzonen) bypässe einzurichten. Damit kann man einen Workflow auch dann funktionsfähig halten, wenn einige Funktionen versagen.

Durch die Modellierung von Workflows ist der Mensch gezwungen, taktgebunden zu arbeiten. Dies schränkt aber seine persönliche Entfaltung stark ein. Arbeiten in Workflows wird deshalb von vielen Menschen als Arbeiten in der Zwangsjacke empfunden. Unter bestimmten organisatorischen Rahmenbedingungen, wie sie z. B. häufig in öffentlichen Verwaltungen gegeben sind, kann daher der Einsatz von Vorgangsbearbeitungs- bzw. Workflow-Systemen sinnvoll sein, die auch zulasen, dass Prozesse sich dynamisch entwickeln („unstrukturierte Prozesse“ bzw. der Workflow nur eine „Grundsteuerung“ oder „-dokumentation“ bindend vorgibt, die vom einzelnen Bearbeiter sachgerecht ausgestaltet werden kann.

Nach Möglichkeit sollten also im Prozess interaktive Passagen eingebaut werden, damit die Kreativität nicht unnötig eingeschränkt wird.

4.5 Die menschliche Komponente

“Der Mensch ist ein Gewohnheitstier“. Zumindest was die Gewöhnung an Prozesse betrifft, stimmt dies häufig. Wenn jemand über längere Zeit in einen Prozess eingebunden ist, so kann es soweit kommen, dass dieser Prozess für ihn zu einem Ritual wird.

Das heißt dann, dass er im Denken so stark in dem Prozessablauf verwurzelt ist, dass es ihm kaum möglich ist, sich andere Abläufe vorzustellen.

Meist versucht man dann ‚vorhandene Prozessabläufe mit neuen Techniken besser ablaufen zu lassen. Die Gefahr ist groß, konventionelle Geschäftsabläufe digital nachzubilden. In den meisten Fällen führt dies zu einer “Verschlimmbesserung“.

Hier hilft nur eine völlige Neugestaltung des Prozesses.

Dafür wird heute von vielen Anbietern Standardsoftware angeboten, die zu einem hohen Prozentsatz in der Lage ist, die gewünschten Prozess-

abläufe abzubilden. Diese Software lässt aber Anpassungen an bestehende Prozesse oftmals nur in einem geringen Maße zu. Eine Anpassung der Organisation und der an ihr beteiligten Menschen an vorgaben durch eine Software ist in den meisten Fällen organisatorisch wenig sinnvoll und führt darüber hinaus oft zu erheblichen Akzeptanzproblemen.

Wenn zudem noch personelle Umstrukturierungen durch die Neugestaltung eines Prozesses notwendig werden, muss mit Konflikten gerechnet werden.

Es empfiehlt sich daher, alle betroffenen Personen so früh, wie möglich an den beabsichtigten Änderungen im Prozessablauf mitwirken zu lassen oder sie zumindest ausreichend über die beabsichtigten Änderungen zu informieren.

Ziel muss es sein, alle zum Mitwirken zu bewegen und so zu begeistern, so dass selbst dann, wenn im Laufe der Prozessumgestaltung eine gewisse Ernüchterung eintritt, die Akzeptanz nicht verloren geht.

4.5 Wertschöpfung

Alle unsere Konzepte müssen dazu führen, dass unsere Prozesse Nutzen für uns und unsere Kunden bringen. Dabei ist es gleichgültig, ob es sich um interne oder externe Kunden handelt. Wichtig ist es, dass wir durch neue Prozesse eine höhere Wertschöpfung erzielen.

Dabei wollen wir größtmögliche Sicherheit im Prozess bei maximaler Flexibilität zu erreichen.

5 Informationsmanagement und seine Anforderungen an Speichersysteme

Die heute am Markt angebotenen Archivierungssysteme zeichnen sich durch ihre lange Lebensdauer aus. Dabei ist jedoch zu beachten, dass Lebensdauer nicht gleichzeitig Verfügbarkeit der zum Schreiben und Lesen benötigten Technologie bedeutet. Das betrifft nicht nur die Hardware, sondern auch die Betriebssysteme der Rechner und die Anwendungssoftware.

Der Anwender steht dabei vor der Herausforderung die immer schnelllebigeren, leistungsstärkeren digitalen Infrastrukturen mittels Datenmigration zu nutzen und gleichzeitig für eine sichere, langlebige Archivierung zu sorgen.

Bevor also eine Medienentscheidung getroffen wird, sollte der Anwender folgende Fragen für seine Archivierungsstrategie stellen und bewerten:

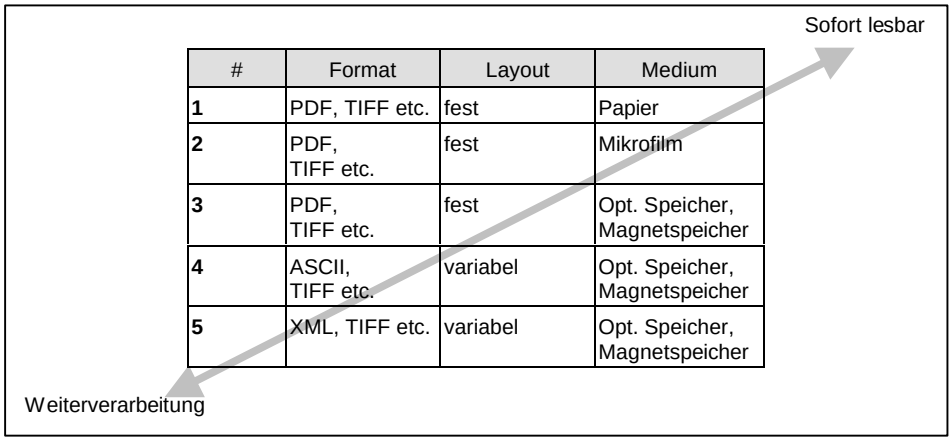
- Wie lange müssen die Dokumente aufbewahrt werden?
- Müssen die Dokumente über die steuerlich relevanten Zeiträume hinweg aufbewahrt werden? Beispiel: Maschinenakten, Stücklisten, Konstruktionsunterlagen, Unterlagen zur Produkthaftung, Personalakten, Lebensversicherungen, etc..
- Wann erfolgen die Zugriffe im Lebenszyklus der Dokumente und wie häufig werden, wann und wo zugegriffen?
- Welchen Wert (geschäftlich, historisch) haben die Dokumente während ihres Lebenszyklus?

Der Anwender muss sich deshalb bei seiner Entscheidung fragen, ob ein einzelnes Speicher- bzw. Archivmedium alle unterschiedlichen Anforderungen erfüllen kann. Es muss schnell und komfortabel sein, gleichzeitig jedoch sicher, langlebig und wirtschaftlich. Dabei ist zu be-

achten, dass Dokumente über ihren Lebenszyklus unterschiedlich (Rückgriffsfrequenz) genutzt werden.

Je nach Aufbewahrungsdauer der Dokumentklassen sollte unbedingt eine gegebenenfalls notwendige Migrationstrategie (Systemwechsel) mit allen erkennbaren Aufwänden berücksichtigt werden.

Das folgende Schaubild unterteilt den Lebenszyklus eines Dokumentes in eine operative, post-operative und eine Archivphase.



#	Format	Layout	Medium
1	PDF, TIFF etc.	fest	Papier
2	PDF, TIFF etc.	fest	Mikrofilm
3	PDF, TIFF etc.	fest	Opt. Speicher, Magnetspeicher
4	ASCII, TIFF etc.	variabel	Opt. Speicher, Magnetspeicher
5	XML, TIFF etc.	variabel	Opt. Speicher, Magnetspeicher

Die Lösung liegt also darin, in der unternehmensweiten Archivierungsstrategie vorbehaltlos analoge und digitale Technologien zu berücksichtige, gemeinsam zu nutzen und die Stärken jeder einzelnen mit den Stärken der anderen Technologie zu verbinden.

Die Medien stehen dabei nicht im Wettbewerb zueinander, sondern ergänzen sich gegenseitig. Das richtige “Medienmix” sorgt für eine optimale Gestaltung der Anwendung und seiner Archivbedürfnisse. Deshalb ist es wichtig, sich innerhalb neuer digitaler Infrastrukturen nicht nur mit DV-gestützten und -gesteuerten Arbeitsprozessen zu befassen, sondern auch mit allen möglich einsetzbaren Speicher- und Archivmedien.