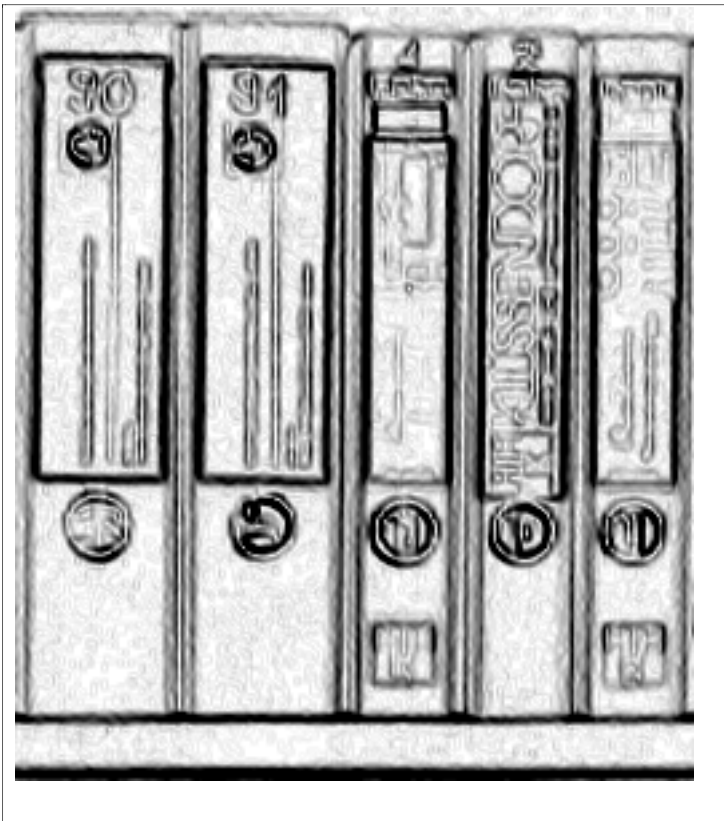


Von der Papierablage ins EDV-Archiv

^^

Wie aus Aktenbergen netzlesbare Daten werden

Wolfram W. Pichler, pichler communication berlin



Wie können wir Regalmeter an Altaktenbeständen einsparen und gleichzeitig unsere alten Papierdokumente online und onscreen verfügbar machen, also für Internet, Intranet und CD-ROM? In diesem Beitrag finden Sie Hinweise über die benötigten Werkzeuge und effiziente Verfahren.

Der Fortsetzungsbeitrag zeigt, wie man ab heute seine Dokumente und Anleitungen von vornherein so aufbaut, daß sie mit geringem Aufwand netzfähig werden.

Inhaltsübersicht

Seite

- 1 Die Papierfluten auf dem Aktenboden
- 2 Wie die Buchstaben vom Papier in den Rechner gelangen
- 3 Wie wir Speicherplatz auf dem Datenträger sparen
- 4 Wie wir unsere Daten wiederfinden
- 5 Tips zum Weiterlesen

- 1 Die Papierfluten auf dem Aktenboden

Als Beispiele für unsere Betrachtungen sollen uns ein Büro für Bauplanung und eins für Technik-Kommunikation dienen.

Stellen wir uns zunächst vor, wir hätten in einem kleinen Bauplanungsbüro für Haustechnik eine neue Anfrage zu bearbeiten. Normalerweise suchen wir uns den letzten ähnlichen Auftrag aus dem Archiv und kalkulieren entsprechend. Mit dieser Standardmethode sparen wir gewöhnlich eine Menge Arbeitszeit, weil wir die alte Kalkulation schnell auf die neue Anfrage projizieren können.

Nun erinnern wir uns: Der letzte vergleichbare Auftrag lag ungefähr fünf bis sieben Jahre zurück. Also gehen wir auf den Archivboden und suchen in den Aktenordnern um 1991 nach dem Auftrag, den wir unter einer bestimmten Baustelle finden, an die wir uns in diesem Zusammenhang ebenfalls erinnern.

Wir greifen uns einen Aktenordner, fegen den Staub ab, husten und - sobald wir wieder sehen können - blättern darin herum, bis wir die Unterlagen in diesem oder einem anderen Aktenordner gefunden haben. Eine nicht sehr angenehme Such/Finde-Situation, zumal der Archivboden über die Jahre auch noch ein Sümmchen an Miete verschlingt und sich die Papiermenge durchschnittlich alle vier Jahre verdoppelt. Anzutreffen ist diese Methode jedoch nicht nur in unserem angenommenen Planungsbüro, sondern auch in Zigtausenden real existierenden

- Konstruktionsbüros,
- Entwicklungsbüros,
- Fertigungsbüros,
- Verwaltungsbüros,
- Redaktionsbüros,
- Verlagen,
- Anwaltskanzleien,
- Detekteien,
- Behörden.

US-Beispiele für Capture/PDF aus Wirtschaft, Industrie und Verwaltung findet man auf der Adobe-Web-Page

<http://www.adobe.com/studio/spotlights/PDFS/Capture1.pdf>

2 Wie die Buchstaben vom Papier in den Rechner gelangen

Dieser Situation kann abgeholfen werden. Statt auf staubigen Aktenböden kann man seine alten Papiere genau so gut am eigenen staubigen Bildschirm suchen, finden, lesen, drucken - und das nicht nur im eigenen Büro, sondern an allen staubigen Bildschirmen dieser Welt, denen wir Zugriff auf unser Archiv einräumen. Dazu sind im Prinzip zwei Vorgänge nötig: Scannen und das Daten-Archiv organisieren, und zwar so, daß man die Daten jederzeit schnell wiederfindet.

Erstens müssen wir die Papierflut scannen. Zweitens müssen wir in unserer EDV ein Ablagesystem organisieren, das uns das Wiederfinden einzelner Vorgänge erlaubt. Entweder übernehmen wir die Struktur der Papierablage oder führen bei dieser Gelegenheit - wenn es etwas zu verbessern gibt - ein sinnvolleres System ein. Vielleicht stimmen wir es nun sogar ab auf die Erfordernisse der Qualitätsdokumentation nach der ISO-9000-Familie. Und wer sich den Luxus einer Volltextrecherche leisten will (oder muß), der verarbeitet die gescannten Papiere noch mit OCR-Software, um aus den Ganzseiten-Grafiken editierbare Texte zu reproduzieren.

Neu erstellte Papiere würde man dann von vornherein als netzfähige Dateien archivieren (vgl. Fortsetzungsbeitrag) und in der Papierablage nur bis zum Abschluß des aktuellen Projekts aufbewahren, es sei denn, gesetzliche Bestimmungen erforderten eine längere Aufbewahrungsfrist für die leibhaftig unterschriebenen Original-(Papier-)Dokumente.

So würde das Digitalisieren unserer Altbestände je nach Umfang des Archivs und Forderungen an die Ausgabequalität einen Rechnerarbeitsplatz mit Operateur einige Zeit voll in Anspruch nehmen. Andererseits läßt sich diese zeitraubende Routinearbeit auch an externe Dienstleister delegieren. Allein im Berliner Branchenfernsprechbuch sind 18 Firmen unter „Datenerfassung“ aufgeführt. Darunter versteht man teils noch - nach wie vor - das Einhacken von Daten durch Datatypist(inn)en, neredings aber auch: Scannen, Digitalisieren, Vektorisieren.

Zusätzlich zur Rechentechnik, die ohnehin in jedem Büro vorhanden ist,

benötigen wir an technischer Ausstattung (mit Richtpreisen):

a) Hardware

- Scanner (1 TDM)
- Platten-/Massenspeicher (1 TDM/1 MB)
- CD-Brenner und -Rohlinge (2 TDM)

Der Preis für CD-Brenner ist inzwischen von 8.000 Mark 1994 auf 1.500 Mark Untergrenze zusammengeschnürt.

b) Software (2 TDM)

- Makro-/Script-Software
- Scansoftware
- Grafikkonverter
- OCR-Software
- Adobe Acrobat
- HTML/WEB-Publisher
- CD-Toast-Software (für den Brennofen)

Die Makro- oder Script-Software benötigen wir, um die Verarbeitung für den Stapelbetrieb zu automatisieren. Würden wir jede Seite einzeln verarbeiten, dann würden wir nie fertig werden. Aber im Prinzip (also abgesehen vom Stapelbetrieb) geht die Verarbeitung einer Seite ungefähr folgendermaßen vor sich.

3 Wie wir Speicherplatz auf dem Datenträger sparen

Wir scannen eine Seite. Um die Datei möglichst klein zu halten, speichern wir nur den Bereich, der Information enthält, also nicht den weißen Rand. Wir sichern die Datei mit einer Auflösung von 300 dpi als komprimierte .tif-Datei und geben ihr einen Namen im Rahmen unseres Ablagesystems (vgl. Abs. 4). Der folgende Test veranschaulicht den jeweiligen Speicherbedarf. Nr. 1 bis Nr. 3 war das Inhaltsverzeichnis einer Betriebsanleitung, gescannt in 3 verschiedenen Auflösungen, Nr. 4 war Deckblatt und Inhaltsverzeichnis dieser Betriebsanleitung, also ein Zwei-Seiten-Dokument. Die gescannte Fläche einer DIN-A4-Seite betrug jeweils gut 18 cm x 27 cm.

Nr. 1	Inhaltsverzeichnis	gescannt mit 100 dpi	(1 Seite)
Nr. 2	Inhaltsverzeichnis	gescannt mit 200 dpi	(1 Seite)
Nr. 3	Inhaltsverzeichnis	gescannt mit 300 dpi	(1 Seite)
Nr. 4	Deckblatt und IV	gescannt mit 300 dpi	(2 Seiten)

Bei der Massenverarbeitung lohnt es sich, den Speicherbedarf der Dateien so gering wie möglich zu halten. Wenn Sie bei jeder Seite nur 1 kB einsparen, bringen Sie auf einer CD-ROM mindestens 10 MB mehr Daten unter. Der Speicherbedarf ist nicht nur eine Funktion der gescannten Fläche, sondern auch eine Funktion der Auflösung. Als Faustregel darf gelten: eine Erhöhung der Auflösung um 100 dpi verdoppelt den Speicherbedarf. Es lohnt sich also durchaus, Fläche und Auflösung nicht größer werden zu lassen, als einer späteren Ausgabe dient.

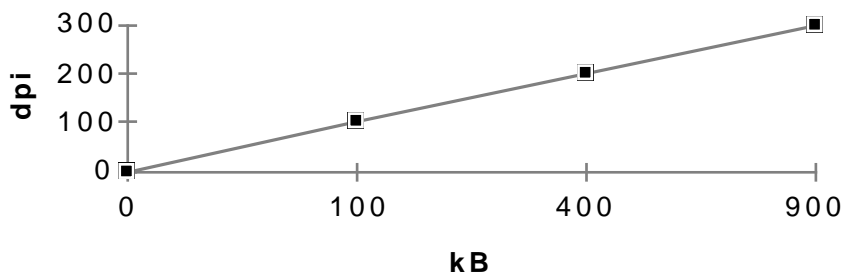


Bild 1: Mit jeder Auflösungserhöhung um 100 dpi verdoppelt sich der Speicherbedarf ungefähr.

Die verschiedenen Dateien nehmen im Test so viel Speicherplatz in kB ein:

Nr.	dpi	(u).tif	(k).tif	.doc	.ps	.pdf
1	1x100	102	33	-	-	-
2	1x200	406	78	-	423	46
3	1x300	913	143	-	930	65
4	2x300	1.736	195	117	1.749	78

dpi = dots per inch, Pixel je Zoll

(u) = unkomprimiert

(k) = komprimiert

.tif = tagged image file format

.doc = Word-Datei nach OCR-Prozeß
.ps = PostScript-Datei
.pdf = Acrobat-Datei, portable document format

Die Datei mit der Auflösung 100 dpi bringt keine Ausgabequalität, die auf Papier als lesbar gilt. Nur wer ausschließlich Seiten für das World Wide Web archivieren wollte, wäre mit 72 dpi gut bedient, weil die meisten Bildschirme ohnehin keine höhere Auflösung bieten. Für das Web muß man die .tif-Dateien noch in .gif-Dateien umwandeln, wenn der benutzte HTML-Publisher diesen Vorgang nicht automatisch durchführt. In Kürze wird das Grafik-Format .fif (fractal interchange format) zu Verfügung stehen, das noch einmal wesentlich weniger Speicherplatz verbraucht und eigens für den schnellen Transport und Aufbau im Internet entwickelt wurde. Mit diesem Grafik-Format bringt man dann ein Vielfaches der heutigen Datenmenge auf einer CD-ROM unter.

Für die Weiterverarbeitung mit OCR muß man beachten, daß manche OCR-Software nur unkomprimierte TIFF-Dateien mit exakt 300 dpi Auflösung verarbeiten kann. Die speicherfressenden .ps (PostScript)-Dateien kann man nach der Umwandlung in .pdf-Dateien wieder löschen. Sie werden ausschließlich zur Umwandlung in .pdf (das auf .ps basiert) benötigt, danach nicht mehr. Die Daten legen also vom Papier bis zur netzfähigen Datei etwa folgenden weg zurück:

Daten	Medium/Tool
Seite	Papier
Scan.tif	Scan-Software
(OCR.doc)	OCR-Software, wenn unvermeidlich
PostScript.ps	Text/Layout-Software
Acrobat.pdf	Adobe Acrobat Distiller

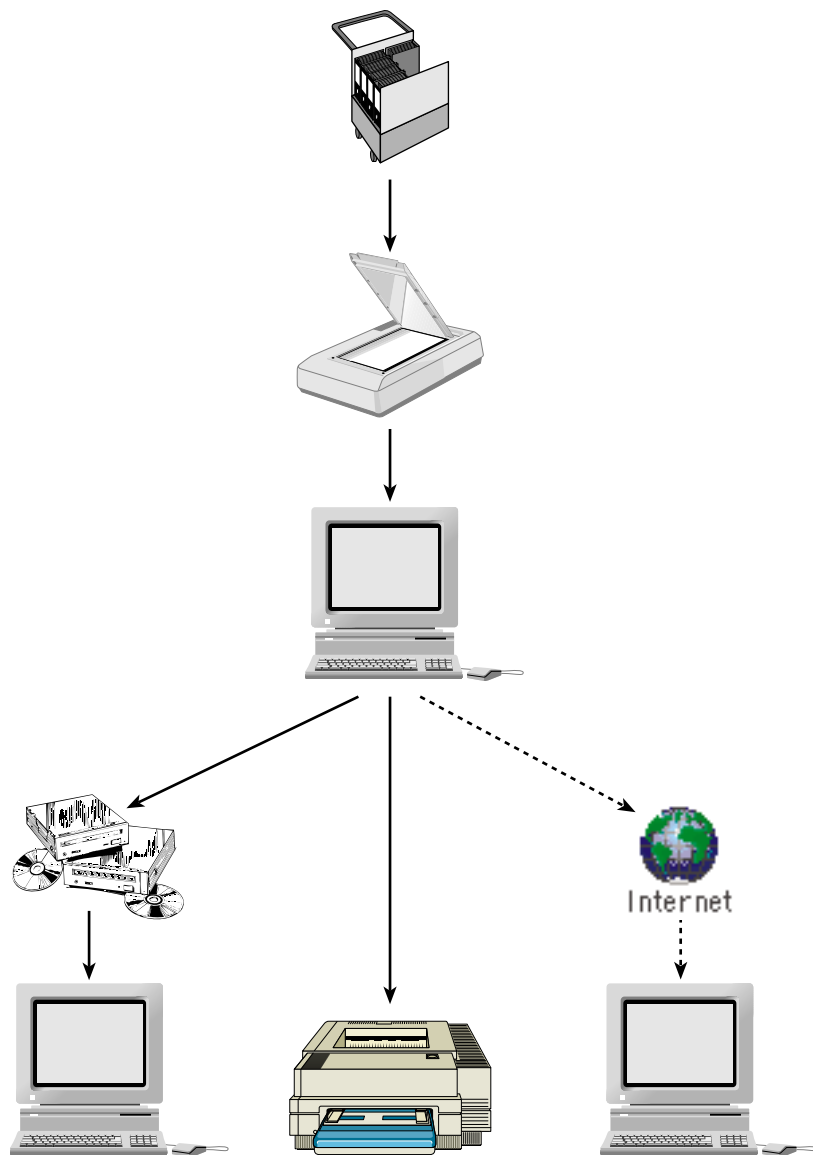


Bild 2: Wie die Buchstaben vom Papier ins Internet gelangen

Der dargestellte Test dauert eine gute Stunde. Zusätzlichen Zeitgewinn kann man sich portionsweise zukaufen durch

- a) Investition in den schnellsten Rechner, den es morgen Nachmittag zu kaufen gibt (z. B. CPU mit 500 MHz getaktet?),
- b) Übung und Routine bei der Einzelseitenverarbeitung und
- c) Makros oder Skripts zur Automatisierung von Stapelbetrieb.

Für 18 TDM gibt es eine Software, die alle Prozesse von der Scannersteuerung über Indizieren und HTML-Übersetzung bis zum

Retrieval voll durchautomatisiert. Ein großer holländischer Verlag für Fachzeitschriften hat damit seine gesamten Alt- und Neubestände an Artikeln ins Web gestellt und erlaubt registrierten Anwendern die Recherche mit Volltextsuche, die nach spätestens 3 bis 4 Sekunden jedes gesuchte Ergebnis liefert (falls das Internet mit seinen Übertragungsraten mitspielt).

4 Wie wir unsere Daten wiederfinden

Ein Dateiverwaltungssystem könnte den Abteilungen der jeweiligen Firma nachgebildet sein, wobei jeder Abteilung ein Verzeichnis (Win) oder Ordner (Mac) entspricht. Der Verzeichnisbaum verzweigt sich nach unten über mehrere Ordnerhierarchien bis zu den einzelnen Dateien. Jeder Ordner wird in jeder unteren Ebene mehre parallele Unterordner enthalten. Die oberste Hierarchie-Ebene eines selbständig operierenden Büros - beispielsweise eines externen Dienstleisters für Technik-Kommunikation - könnte so aussehen:

- A Büroverwaltung und -organisation
- B Einkauf
- C EDV
- D Forschung und Entwicklung
- E Qualitätsmanagement
- F Weiterbildung
- G Fachliteratur
- H Marketing, PR
- I Auftragsbearbeitung
- J Finanzbuchhaltung, Jahresabschlüsse
- K Kooperation
- L Lehre, Unterricht, Vorträge, Workshops
- M Fachtagungen, Reisen
- N Verbände, Mitgliedschaften

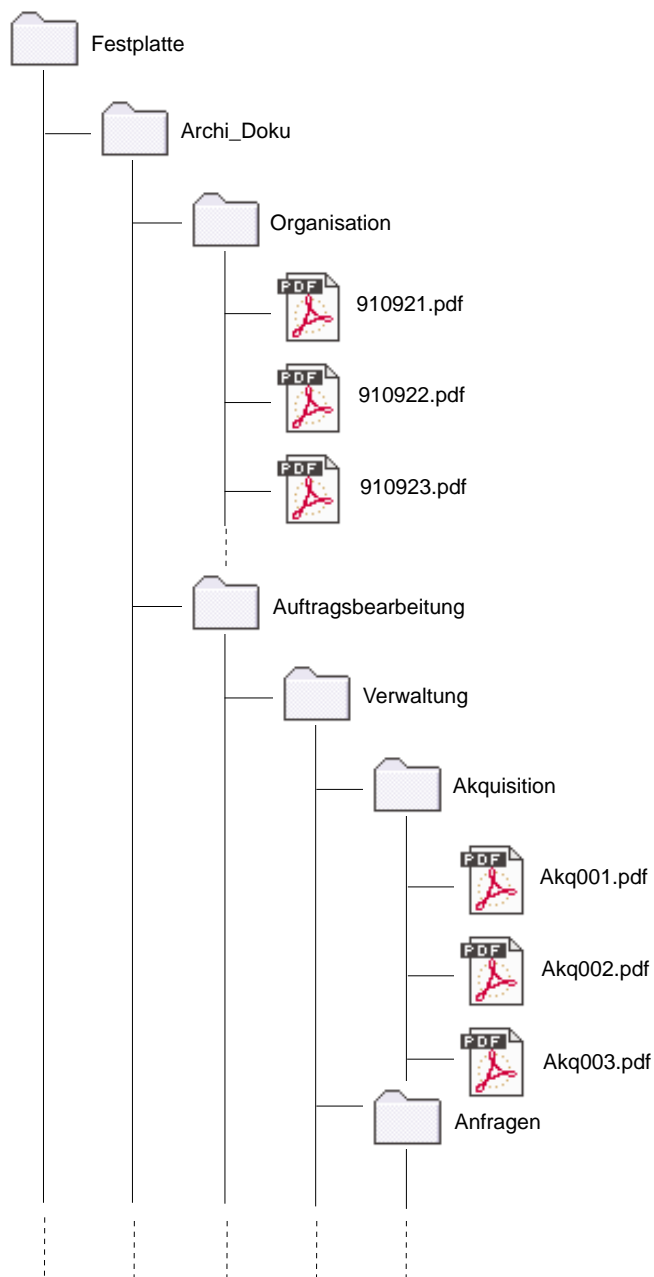


Bild 3: Eine denkbare Archivstruktur für ein Dokumentationsbüro

Dabei muß sich der Feinheitgrad der Unter-/Gliederung nach den individuellen Bedürfnissen der jeweiligen Organisationseinheit richten. Nicht zu vernachlässigen sind auch die in Jahren geprägten Such-/Findegewohnheiten der Mitarbeiter. In einem Redaktionsbüro könnte man zum Beispiel den Ordner I Auftragsbearbeitung folgendermaßen weiter untergliedern:

I1 Verwaltung (Akquisition, Anfragen, Kalkulation, Kostenvoranschläge, Aufträge, Nachkalkulation, Abrechnung, Mahnwesen)

- I2 Recherche (Material, Schriftwechsel, Entwürfe)
- I3 Betriebsanleitung
- I4 Version, Edition, Sprachvariante
- I5 Kapitel, Abschnitte

Beim eingangs angenommenen Bauplanungsbüro für Haustechnik sind die Gewerke so geordnet:

- P Elektro
- Q Heizung
- R Sanitär

Ein Ordner Bauvorhaben ist folgendermaßen untergliedert:

- Q1 Schriftverkehr
- Q2 Protokolle
- Q3 Besprechungsnotizen
- Q4 Ausführungszeichnung
- Q5 geprüfte Ausführungszeichnung
- Q6 BPU, Bauplanungsunterlage
- Q7 Baugenehmigungen
- Q8 Planung, Entwurf
- Q9 Leistungsverzeichnis, Aufträge
- Q10 Rechnungen
- Q11 Tagesablage (JJMMTTxx.pdf)

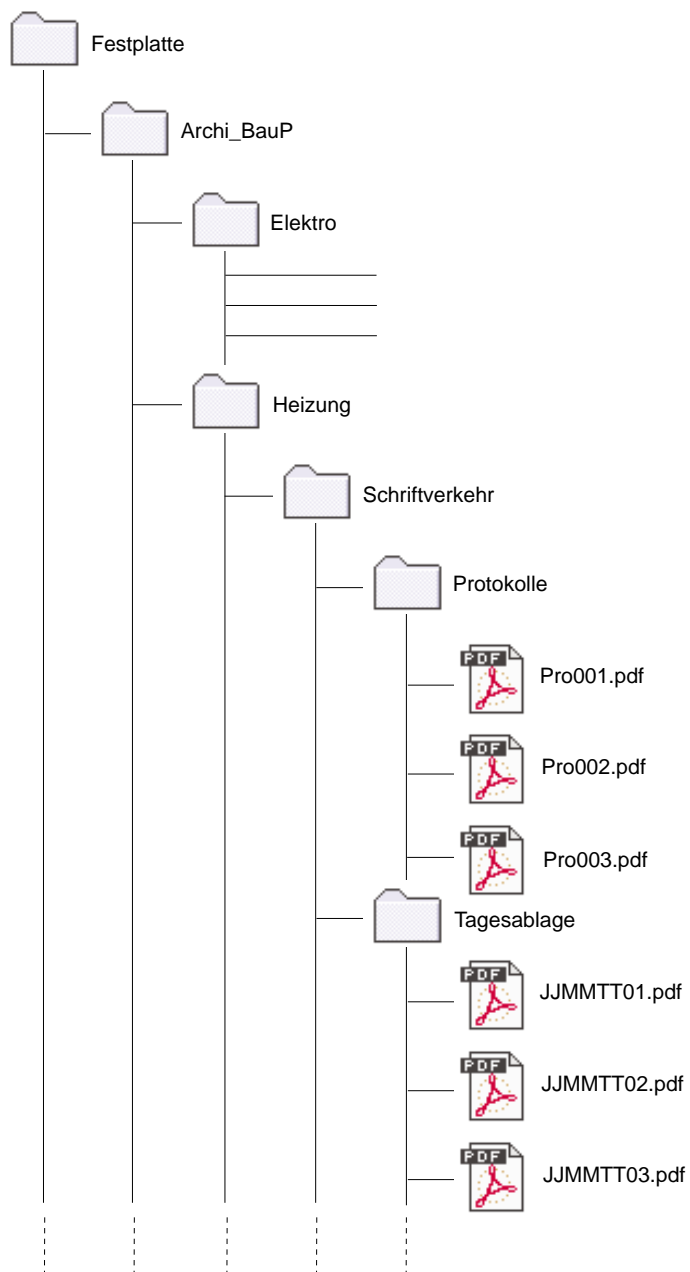


Bild 4: Eine denkbare Archivstruktur für ein Bauplanungsbüro für Haustechnik

Das Service- und Dokumentations-Archiv eines Elektronikunternehmens könnte so strukturiert sein:

1. Bedienungsanleitungen
2. Serviceanleitungen
3. Stücklisten
4. Serviceinformationen

5. Zubehörübersichten
6. Amtliche Papiere wie Zulassungsurkunden oder Konformitätserklärungen
7. Meß- und Prüfmittelkataloge
8. Schaltplansammlungen
9. Kombinationslisten "welches Gerät spielt mit welchen andern"
10. Anschlußhilfen und Buchsenbelegungen

Mit dem hier beschriebenen Verfahren kann man gut 10.000 DIN-A4-Seiten auf einer CD-ROM unterbringen. Da wird leicht klar, daß sich die ganzen ehemals staubigen Bestände eines Aktenbodens in einem kleinen Rollcontainer unterbringen lassen, der unter einen Beistelltisch paßt.

5 Tips zum Weiterlesen

Gabriel, Carl-Heinz: Rechtsverbindlichkeit elektronischer Produktdokumentation. In: *tekom nachrichten*, Fachzeitschrift für technische Kommunikation, Stuttgart, Nr. 1•2/96, S. 46-47

Kölling, Jürgen: Der Weg zum elektronischen Publizieren im DIN. In: *DIN-Mitteilungen* 75. Nr. 8/1996, Beuth Verlag, Berlin (4 Bilder), S. 547-550.

Sawall, Achim: Die eigene CD-ROM-Fabrik. *Computer*: Was beim Kauf eines speziellen Brenners zu beachten ist. In: *Berliner Zeitung*, Nr. 136 vom 13.6.96, Berlin

Siebke, Manfred: Digitale Dokumentenverwaltung. Erstellung, Verwaltung und Weitergabe von digitalisierten Zeichnungen, Normen und anderen Unterlagen. Hrsg.: DIN, 1996, 48 Seiten, Brosch., 45 DM. Beuth Verlag, Berlin. ISBN 3-410-13561-8

Schöll, Ulrich: Rechnergestützte Dokumentation in der Anwendung. In: *DIN-Mitteilungen* 75. Nr. 8/1996, Beuth Verlag, Berlin (7 Bilder), S. 551-553.

softline: Software zum Staunen für Windows und Mac. Katalog Nr. 11,
1996, Oberkirch, 212 Seiten, 10 DM. E-Mail = softinfo@softline.de WWW
= <http://www.softline.de/>

Tscheke, Tibor: Elektronischer Dokumentenaustausch - Stand der
Technik. In: DIN-Mitteilungen 75. Nr. 7/1996, Beuth Verlag, Berlin (2
Bilder), S. 490-493.