

H. Völz, Thesen zu:

Was muss künftig gespeichert werden?

Dieser Text besteht aus zwei Teilen. Zunächst wird provisorisch versucht, die wahrscheinlich wichtigsten Fragen zu erfassen. Dann folgt ein Auszug aus meinem in Arbeit befindlichen Band 3 „Handbuch der Speicherung von Information“, der einen speziellen Bereich erfasst, dafür aber bereist gründlicher ausgearbeitet ist. An Diskussionen und Hinweisen zu den Fragen und Aussagen bin ich sehr interessiert. Bitte per email: voelz@zedat.fu-berlin.de.

Voraussetzung: Im Prinzip ist es heute technisch **möglich, Alles** was auf der Welt elektronisch geschieht, zu speichern. Damit ergeben sich neue Fragen zu dem was **gespeichert werden sollte**. Die folgenden Listen sind unvollständig und müssen ergänzt und verändert werden.

Inhalte erfragen

Was **wofür** (Utility. Anwendbarkeit)?

Was für **wen** (individuell)?

Was für die **Menschheit** (Notwendigkeit, Nutzen)?

Was für **Kultur und Geschichte** (Menschenbild, Religion usw.)?

Hierfür müssen **Kriterien** gefunden werden.

Hinweise wären möglich aus

Was kommt in **Lehrbücher, Lexika, Publikationen** usw.?

Gewiss nicht mehr alles, was schon einmal gespeichert wurde; Lehre, Ausbildung,

Was interessiert den **Historiker**?

Nur die Quellen oder auch deren alte Interpretationen

Was kommt in die **Museen, Archive, Bibliotheken** usw.?

Museen heben wohl nur Materialisiertes, Dinge usw. auf, aber keine Ideen usw., daher

Was speichern **Rundfunk** und **Fernsehen**?

Was ist für die **Archäologie** bedeutsam?

Folgerungen und Vertiefungen

Wie wird man **Urhebern** von Ideen gerecht? Frage der Zitate, Plagiate usw.

Wann sind **Doppelungen zur Sicherheit** notwendig? (Backup)

Wie könnte ein universelles **Weltarchiv** entstehen? Wissen ist Macht!

Moralischer Umgang mit **destruktiver** Information, Lügen usw.

Was muss vor **Missbrauch**, vor **Zerstörung** usw. geschützt werden?

Wie sind **Geheimhaltung** und **Geheimdienste** zu berücksichtigen?

Brauchen wir z.B. alle Audioaufnahmen, benötigen wir alle Filme?

Dies könnte ein besonders wichtiger Aspekt der Speicherung werden.

Daher: Wie sind die Archive usw. vom unbenötigten **Ballast** zu befreien?

Begriffe **Redundanz** und **Relevanz** wären hierfür neu zu definieren

Technische Fragen

Wie ist ein **Universalformat** zu finden?

Wie schnell muss der (technische) **Zugriff** zu verschiedenen Daten erfolgen?

Was sind hier neue technische Varianten? assoziative Speicherung?

Originale, Formeln in welchem **Format der Zukunft** ablegen?

Kompression ja/nein Zugriff/Speicherkapazität (Kosten)

Was wird mit veralteter, nostalgischer Technik?

Wie oft, wie viel, individuell ↔ offiziell

z.B. Schallplatte und mechanische Musikinstrumente

Sonderfälle beachten: z.B. Verpackung, Werbung usw.

Vergleich von **inhaltlichen** und **ökonomischen Grenzen**

Hinweis

Allein eine Vervollständigung und inhaltliche Begründung dieser und weiterer Fragen erbrächte auch ohne Antworten nützliche Aussagen!

Teil 2

Hierarchie, Backup und Archiv

In der Anfangszeit der Rechentechnik, tauschten der Rechner direkt mit dem Magnetbandgerät Daten, Programme usw. aus. Zur Verkürzung der langen Zugriffszeiten entstand der Start-Stoppbetrieb. Sie wurde dann durch zusätzliches Einfügen von Disketten, Wechsel-, Festplatten usw. weiter gesenkt. Es entstand die typische Speicherhierarchie, auf die u.a. im Abschnitt 1.3.4 und in Bild 1.18 eingegangen ist. Heute existieren aus der Perspektive des Magnetbandes drei recht unterschiedliche Anwendungen der Magnetbandtechnik. Sie betreffen den Einsatz des Magnetbandes und führten zu z.T. recht unterschiedlichen Lösungen, die im Folgenden nacheinander behandelt werden:

1. innerhalb der *Speicherhierarchie*,
2. für das Sicherheits-*Backup*.
3. als *Archiv*-Medium und

Heute können heute nahezu beliebig große Datenmengen gespeichert werden. Dabei ändern sich mit der Datenmenge je nach dem benutzten Speichermedium insbesondere der Preis je Bit, die typische Zugriffszeit und das notwendige Archivierungsvolumen. Vor Allem haben begrenzte ökonomische und räumliche Ressourcen dabei für große Datenmengen eine **komplexe Hierarchie** mit verschiedenen Medien erzwungen. Sie beginnt bei den statischen RAMs im Rechner und endet beim Magnetband und bei sehr großen Magnetbandbeständen über Autolader bei den mit Robotern bestückten Bandbibliotheken. Doch in vielen Fällen ist es nicht zulässig, dass hierbei für einige Daten eine sehr lange Zugriffszeit auftreten kann. Dem Anwender muss auch ein riesiger Datenbestand so erscheinen, als ob er zu allen Daten einen schnellen Zugriff hat. Grundsätzlich verlangt das eine Lösung, die dem Cache der Halbleiterspeicher ähnelt (vgl. Abschnitt 2.3.3). Doch für die auf Magnetbändern gespeicherten Daten gelten andere Bedingungen als für Programmdateien der Rechner. Die nacheinander angeforderten Daten befinden sich meist *nicht* in Nachbarschaft der aktuellen Daten. Außerdem existieren bei Magnetbändern keine Adressen wie bei den RAMs. Die nächsten Daten können immer nur durch Suchen auf dem Band gefunden werden. Somit ist praktisch kein „Vorhalten“ von wahrscheinlich angeforderten Daten möglich. Erst in den letzten Jahren sind für diese Probleme neuartige Lösungen entstanden. Sie sind durch Begriffe (bzw. Schlagwörter) mit nur z.T. verschiedenen Inhalten gekennzeichnet (s.u.). Die Situation ist z.Z. noch recht unübersichtlich und vorwiegend durch firmenspezifische Hard- und Software-Lösungen gekennzeichnet. Auch einem ersten Buch mit zehn Kapiteln [GRA03] gelang es kaum, die angewandten Methoden eindeutig herauszuarbeiten. Das gilt genauso für die dort übernommene Hype-Kurve der Gartner Group (vgl. Bild 1.30) demonstriert bestenfalls die inhaltliche Unsicherheit dieser Gebiete. Deshalb werden hier nur wenige, vielleicht die wichtigsten Hinweise gegeben.

Die Daten eines Unternehmens hängen von mehreren Parametern ab. Dies berücksichtigt das „**Information Lifecycle Management**“ (ILM), das jedoch kein Produkt, sondern eine Strategie ist, bei der die Datenklassifizierung entscheidend ist. Mit ihm werden Parameter wie Speicherkosten, Datenmenge, Notwendigkeit/Wichtigkeit für den Betriebsablauf, Änderungsgeschwindigkeit, Gültigkeitsdauer, Vertraulichkeit, Geheimhaltung, Zugriffshäufigkeit, Zugriffszeit, Datenrate, zulässige Wartezeit, Archivierungspflicht und Betriebsgeschichte auf ein Modell der verschiedenen Medien abgebildet. Dann wird ein Programm geschrieben, das (nahezu) automatisch und kosteneffizient die Daten bezüglich der genannten Parameter verwaltet und dabei automatisch in einer mehrstufigen Hierarchie umspeichert. Dabei soll den Nutzern das System so transparent erscheinen, dass er das Gefühl erlangt, er habe zu allen benötigten Daten einen direkten Zugriff. Beispiele für die Organisation geben die folgenden Beispiele. Z.B. müssen unternehmenskritische Daten immer sofort verfügbar sein, weil sonst die Geschäftsprozesse zum Stillstand kommen. Sie müssen auf den Speichern mit kurzer Zugriffszeit (z.B. schnelle Festplatten) verbleiben. E-Mails sind in der Regel nur einen Tag aktuell und können daher auf sehr langsamen Speichern archiviert werden. Konstruktionsdaten werden während der Konstruktion dauernd gelesen und geändert. Doch nach dem Abschluss werden sie nur noch für die Fertigung und später für Service-Zwecke benötigt. Es gibt auch Daten, die zum Vergleich immer wieder saisonal benötigt werden. Bei vielen Anwendungen wurde festgestellt, dass auf die Hälfte der neu erzeugten Daten nach einem halben Jahr nicht mehr zugegriffen wird. Daher können sie in der Hierarchie schrittweise zu Speichern mit größerer Zugriffszeit verlagert werden. Schließlich gibt es auch

Daten, die im üblichen Betriebsablauf nicht mehr relevant sind aber aus rechtlichen Gründen aufbewahrt werden müssen. Sie können ähnlich wie Daten zur Geschichte des Betriebes ganz ans Ende der Hierarchie, zu den Bändern und Bandbibliotheken oder gar offline ausgelagert werden.

Andere Firmen bezeichnen ähnliches Vorgehen als „Hierarchical One Volume Manager“ (OVM-H) oder „Hierarchisches Speicher-Management“ bzw. „Highend Storage Management“ (HSM). Das „Storage Area Networks“ (SAN) versucht u.a. zusätzlich und integrierend zu berücksichtigen, dass jährlich die Kosten für die Hardware um 35 % fallen, aber die für das Speicher-Management um 15 % steigen. Beim „Direct Attached Storage“ (DAS) wird für solche Prozesse sukzessiv eine vernetzte Infrastruktur aller Speicher geschaffen. Dabei automatisiert eine Storage-Management-Software weitgehend Datenlokalisierung, Backup und Recovery. Eine spezielle Komponente bei den genannten Prozessen – aber auch beim Backup (s.u.) – ist die „**Tape-Virtualisierung**“. Sie ist eine Variante der allgemeinen Speicher-Virtualisierung, bei der mehrere, z.T. unterschiedliche physikalische Speicher unter einer gemeinsamen Oberfläche (Schicht) zusammengeführt werden. Sie erscheinen dem Anwender dann transparent als ein einziger Speicher. Bei der Tape-Virtualisierung wird der Magnetbandspeicher zunächst auf einer Festplatte simuliert. Erst wenn die dort gesammelten Daten ein Band füllen, werden sie dorthin gespeichert (Imigration). So können Bänder – statt wie sonst nur zur Hälfte – fast vollständig ausgenutzt werden.

Theoretisch sollte ein **Backup** als zusätzliches Aufheben eigentlich überflüssig sein, denn *gespeicherte* Daten sollten – das ist der wesentliche Zweck der Speicherung – beständig und damit stets vorhanden sein. Doch in der Wirklichkeit gibt es viele Ursachen, welche die gespeicherten Daten unzugänglich machen, sie verändern oder gar zerstören. Die Gründe dafür sind vielfältig und lassen sich wie folgt einteilen:

- **Hardwareausfälle** (≈20 %) u.a. durch defekt gewordene Festplatten (head crash), Bauteile und Kabel sowie fehlerhafte Datenträger, störende Magnetfelder (z.B. Lautsprecher, Farbfernseher, Motoren, Netztransformatoren), Überhitzung, Staub, Rauch und unkontrollierte Veränderungen.
- Zu den **Softwarefehlern** (≈15 %) zählen u.a. Rechnerabsturz, fehlerhaftes Backup und Restore (Rückholen der Daten aus dem Backup).
- **Menschliches Fehlverhalten** (≈50 %) erfolgt z.B. durch Bedienfehler, insbesondere versehentliches Löschen oder Überschreiben.
- **Naturkatastrophen** und höhere Gewalt (≈7 %) betreffen Überschwemmungen, Brände Erdbeben usw., z.T. auch Stromausfall.
- **Absichtlich böswillige Attacken** (≈8 %) bewirken Viren usw. oder gehen aus Rache, Zerstörungswut, Sabotage, Kriminalität und Diebstahl hervor. Erstaunlich ist dabei, dass etwa 80 % hiervon auf Mitarbeiter zurückgehen. Zuweilen wird empfohlen wenigstens drei Virencanner anzuwenden. Denn laut Statistik findet jeder nur zwischen 85 und 97 %.

Die in Klammern eingefügten Prozente sind nur Richtwerte mit z.T. beachtlicher Streuung. Als ein negatives Musterbeispiel zum nachlässigen Umgang mit wertvollen Daten ist u.a. in Chip 4/02 auf Seite 272 zu lesen¹:

„Wie man mit wichtigen Daten nicht umgehen sollte, demonstrierte ausgerechnet die amerikanische Luft- und Raumfahrt-Behörde NASA: Sie hat in einer Nacht- und Nebelaktion sämtliche Baupläne und Produktionseinrichtungen für die Einweg-Trägerrakete Wernher von Brauns vernichtet – zu Gunsten des wiederverwendbaren Space-Shuttle. Dreißig Jahre nach der letzten Landung auf dem Mond wäre die NASA nicht mehr im Stande, die weltweit stärkste Trägerrakete nachzubauen und so die Voraussetzung für eine weitere Mondfahrt zu schaffen.

Nicht weniger folgenreich war die Nachlässigkeit, Millionen von Magnetbändern unkatalogisiert in Pappkartons zu verpacken, in Lagerhäusern abzustellen – und dort zu vergessen. Den Magnetschichten auf den Bändern blieb genug Zeit, sich von der Trägerfolie zu lösen oder sich

¹ Es wird hier bewusst nicht auf Aussagen Bezug genommen, welche u.a. hieraus folgern, dass die Mondlandungen gar nicht stattgefunden hätten. Vgl. [WIS05]. Es ist aber auch zu bedenken, dass in der 70-jährigen Geschichte des Magnetbandes Totalverluste von Aufnahmen eine extrem seltene Ausnahme sind. In den meisten Fällen handelte es sich dabei um durch Brand unbrauchbar gewordenenes Magnetband. Vgl. Vortrag Werner Singhoff: „Sachgerechte Langzeitlagerung von Analog- und Digitalbändern“ 1992 [langzeitlager.pdf] **wird noch genauer gesucht!!!** Andererseits ist bekannt, dass die amerikanische Regierung in den 80er Jahren mehr als eine Million Dollar für das Reaktivieren von *Geräten* und *Programmen* ausgeben musste, um die noch fehlerfrei vorhandenen Daten der alten Volkszählung wieder nutzbar zu machen.

stellenweise auf der Spule durchzukopieren. Dokumente aus drei Jahrzehnten US-amerikanischer Raumfahrt gingen verloren, weil die NASA keine Backups erstellt hat.“

Verliert eine Firma alle ihre Daten, so belegt die Statistik, dass sie nach zwei Jahren bestenfalls noch mit 50 %er Wahrscheinlichkeit existiert. Verlor sie die Daten gar durch Naturkatastrophen, wie Feuer oder Erdbeben bzw. durch Diebstahl mit folgenden Missbrauch der Daten, so überlebt sie kaum das nächste Halbjahr. Wie teuer eine neue Erzeugung von Daten ist, weist die **Tabelle 33** von Ende 1997 aus. Daher ist es erstaunlich, dass die notwendigen Backups leider oft nur fehlerhaft durchgeführt werden. Es gibt sogar die Aussage, dass die Mehrzahl der individuellen Computernutzer überhaupt kein Backup durchführt. Selbst Sprüche wie „wer ein Backup macht, ist feige“ kursieren. Nach der Firma Ontrack existieren sogar für 80 % der dort bearbeiteten Datenverlustfälle Backups. Doch leider stellt sich dann sehr allzu oft heraus, dass sie sich in keinem verwertbaren Zustand befinden. Nur um 30 % der Backups sind wirklich verwertbar. Deshalb wird empfohlen, des Öfteren eine Probewiederherstellung (Verifizierungslauf) auf eine andere Festplatte oder in ein anderes Verzeichnis durchzuführen. Ferner ist zu beachten, dass es bei digitaler Technik – ganz im Gegensatz von zum z.B. Papier und zu analogen Medien – kaum eine Möglichkeit gibt, den Beginn von Ausfällen zu erkennen. Wegen der daher üblichen „schleichenden“ Fehler wird oft empfohlen, regelmäßige Backups zumindest im Abstand von vier Wochen bis zu einem Jahr durchzuführen. Besser ist natürlich das häufig empfohlene GVS-Prinzip (**Großvater-Vater-Sohn**). Die Söhne sind fünf Bänder (-Gruppen), auf denen umschichtig die differentiellen Tageskopien von Montag bis Freitag vorgenommen werden. An jedem Wochenende erfolgt immer ein volles Backup umschichtig auf fünf Väter-Bänder. Schließlich kommen noch zwölf Großvater-Bänder hinzu, auf denen sich die vollständigen Backups der zurückliegenden Monatsenden befinden. Nach jedem Backup ist darauf zu achten, dass der Schreibschutz aktiviert wird. Nur so ist versehentliches Löschen oder Überschreiben zu vermeiden. Die Lagerung aller Backup-Bänder muss an einem Ort erfolgen, der recht weit vom Rechner entfernt ist. Dort und beim Transport dürfen keine Magnetfelder auf die Bänder einwirken (s.o.). Erst nach Einhaltung dieser Vorkehrungen gilt schließlich die Aussage: Die einzigen Daten, die bei einem Computer jemals verloren gegangen sind, sind Daten, von denen es kein Backup gab. Dem muss aber noch hinzugefügt werden, dass zum Backup auch immer ein in gewissen Abständen erprobtes Recovery bzw. Restore (Wiederherstellung) gehört. Neben dem eigentlichen Backup erlangen in der letzten Zeit auch immer mehr spezielle Methoden zur Wiederherstellung an Bedeutung. Sie ermöglichen es u.a. einzelne Dateien, die durch versehentliches Löschen, Rechnerabstürze oder schädigende Fremdeinwirkung (u.a. Viren) vernichtet oder geändert wurden, zu rekonstruieren.

Tabelle 33. Kosten und Zeit für die Wiedererzeugung von 20 MByte Daten [ntz 12/97; S. 26]. Ein erfolgreiches Restore dauert dagegen selbst bei einem Totalverlust und sehr viel größeren Datenmengen erfahrungsgemäß „nur“ wenige Tage.

Bereich	Kosten	Zeit
Vertrieb/Marketing	25 TDM	19 Tage
Rechnungswesen	29 TDM	21 Tage
Produktion	40 TDM	32 Tage
Forschung	147 TDM	42 Tage

Ein Backup hat auch Nachteile. Es erfordert auf mehrfache Weise Zeit und bedingt Kosten für die notwendigen Speichermedien. Die **Zeit** betrifft dabei nicht nur den Aufwand des Operators. Dieser kann sogar durch weitgehend automatische Routinen verringert werden. Z.T. wird beim Backup auch vorübergehend der Zugriff zu den jeweiligen Daten verhindert und kostet dann Arbeitszeit der Nutzer. Ein lang andauerndes Backup ist auch nicht mehr eine Momentaufnahme (Snapshot, Schnappschuss) des aktuellen Speicherzustands. Zwischen Anfang und Ende können Dateien verändert, gelöscht oder hinzugekommen sein. Daher sollte ein Backup recht schnell und möglichst außerhalb der eigentlichen Arbeitszeit erfolgen. Die Zeit bis zur Wiederherstellung bei einem notwendigen Restore wird neben anderen Einflüssen (s.u.) vor allem die Datenmenge bestimmt.

Zur Senkung der **Kosten** gibt es mehrere Backup-Varianten:

- Beim **vollständigen** (totalen) Backup (Snapshot, Gesamtsicherung) werden alle Daten eines Laufwerkes, Servers usw. auf das Backup-Medium, fast immer Band übernommen. Dabei werden die Archivbits der Dateien zurückgesetzt.

- Ein **differentielles** Backup (Zuwachssicherung) speichert infolge der inzwischen gesetzten Archivbits nur die seit der letzten Sicherung veränderten oder neu entstandenen Daten. Ihr Archivbit wird zurückgesetzt.
- Ein **individuelles** Backup (Auszug, Selektivsicherung) speichert nur einzelne Dateien und/oder Pfade. Das Archivbit bleibt unverändert. Dieses Backup wird vorwiegend für besonders wichtige Daten oder spezielle Zwecke angefertigt. U.a. lassen sich so auch redundante und „private“ Daten ausschließen. Das bewirkt meist eine erhebliche Senkung des Speicheraufwandes und Verkürzung der Backup- sowie Restore-Zeit gegenüber einem vollständigen Backup.
- Ein **komprimiertes** Backup ist bei den drei o.g. Methoden zusätzlich zur Reduzierung des Speicherbedarfs möglich. Dabei werden alle zu sichernden Dateien verlustfrei verdichtet in einer oder wenigen Spezialdateien zusammengefasst. Außerdem kann hierdurch beim Restore Zeit eingespart werden. Die Zeit für das Dateilesen ist nämlich meist deutlich größer als die für das Expandieren im Arbeitsspeicher.

Die ersten beiden Verfahren stellen eine gute Kombination zur regelmäßigen Sicherung dar. Ein Beispiel hierfür ist das o.g. GVS-Prinzip. Es hat aber auch Nachteile. Wenn bei einem Ausfall ein länger zurückliegendes vollständiges und mehrere danach erfolgte differentielle Backups vorliegen, müssen sie nacheinander in der richtigen Reihenfolge zurückgespielt werden. Das verlangt hohe Aufmerksamkeit. Außerdem können dabei auch Daten hergestellt werden, die zur Ausfallzeit des Rechners bereits gelöscht waren und damit eigentlich nicht mehr benötigt werden.

Für ein Backup sind Sicherheit, Kosten und Zeit entscheidend. Dabei werden oft Festplatte und Magnetband gegenübergestellt. In den Kosten/Byte ist das Magnetband um ein Vielfaches günstiger. Auch seine Sicherheit ist bei sachgemäßer Lagerung sehr hoch. Selbst nach über hundert Jahren (spezifiziert sind meist *nur* 30 Jahre) existiert es in einem einwandfreien Zustand (s.o.). Festplatten fallen dagegen viel häufiger aus. Typisch sind wenige Jahre. Hinzu kommt, dass kein Hersteller bereits ist, Aussagen zur Zuverlässigkeit von *gelagerten* Festplatten zu machen. Es gibt aber mehrere Anzeichen dafür, dass sie im Dauerbetrieb sicherer sind. So wird empfohlen, *gelagerte* Festplatten jeweils nach etwa jeweils einem Jahr auf ihre Funktionstüchtigkeit hin zu überprüfen. In diesem Zusammenhang wird immer häufiger auf die Gefahr der Auslagerung von Daten auf externe Geräte, wie z.B. USB-Festplatten hingewiesen. Wegen Preisdruck genügen sie meist nur geringen Anforderungen. Damit liegt der einzige Vorteil der Festplatten bei der Zugriffszeit. Daher sind Lösungen entstanden, die beide Techniken kombinieren. Hierbei lässt sich auch ein typischer Mangel beim Backup und Restore vermeiden. Vielfach ist heute nämlich die Datenrate der Magnetbandtechnik so hoch, dass Server zuweilen nicht folgen können. Dann muss das Bandgerät warten. Vor einem neuen Start ist dann das Band zurücksetzen. Nur dann kann es mit der erforderlichen Sollgeschwindigkeit wieder an der richtigen Stelle fortfahren. Dieser Prozess heißt u.a. **backhitch** (*englisch back* zurück und *to hitch* rücken, hinken, anhängen) oder **shoe shining** (*englisch* etwa Schuhe putzend). Er bewirkt einen hohen Verschleiß, verlängert die Zeit und ist mit einem unangenehmen Geräusch verbunden. Gewünscht ist dagegen ein **Nonstop-Streaming**. Es kann u.a. durch einen sehr großen elektronischen Puffer erreicht werden. Einige Streamer besitzen dafür auch die Möglichkeit, die Bandgeschwindigkeit variabel oder in Stufen anzupassen. Doch besonders effektiv ist es, zunächst das Magnetband auf einer Festplatte zu simulieren. Auf ihr werden solange Daten gesammelt bis sie ein Band füllen (Zwischensicherung). Dann werden sie kontinuierlich auf dem Band gespeichert (s.o. Tape-Virtualisierung). Hierdurch ist es auch möglich, die Speicherkapazität des Bandes nahezu vollständig zu nutzen, während sonst oft nur 50 % erreicht werden. Außerdem entfällt das sonst oft kritische Zeitfenster.

Zwischen Backup und **Archivierung** gibt es Ähnlichkeiten und Unterschiede. Ähnlichkeiten betreffen einige Aufgaben und Ziele, die aus der folgenden Aufzählung ersichtlich sind

- Um Datenverlusten vorzubeugen, müssen wichtige Daten zusätzlich und außerhalb des Computersystems am sicheren Ort gelagert werden.
- Auch aktuell oder im Arbeitsprozess eigentlich überhaupt nicht mehr benötigte Daten, müssen – z.B. aus gesetzlichen und historischen Gründen – prinzipiell verfügbar sein.
- Zwischen entfernten und nicht vernetzten Rechnern – u.a. Notebooks – ist häufig ein Datenabgleich (Synchronisation) notwendig (nur Backup-ähnlich).
- Sensible Daten müssen gegen unbefugten Zugriff gesichert werden (bedingt eine spezielle Speicherung).

Unterschiede betreffen vor allem *Zeitfaktoren* sowie *Veränderbarkeit* und *Originaltreue* der Daten. Während beim Backup immer die benötigte Zeitdauer einschließlich des sicheren Restores möglichst kurz sein muss, zählt bei der Archivierung vor allem die prinzipielle Verfügbarkeit, wobei dann die Zeit bis zum Erhalt der gespeicherten Daten meist unkritisch ist. Im Gegensatz zum Backup verlangt die Archivierung vor allem eine sehr lange Verfügbarkeit der Daten. Dies wird zuweilen durch Begriffe wie Langzeitarchivierung und erforderlich Lebensdauer der Daten unterstrichen. Unter allen elektronischen Speichermedien hat hier das Magnetband – gute Lagerung und Pflege (s.u.) vorausgesetzt – die besten Eigenschaften (s.o.). Sie überstehen fast alle negativen Umwelteinflüsse, wie hohe Temperaturen, Feuchte und Strahlungen aller Art. Lediglich für die sehr viel teureren magnetooptischen Medien (MO) werden etwas längere Archivierungszeiträume garantiert. Dennoch ist zu beachten, dass nach den bisherigen Erfahrungen keineswegs Jahrhunderte (wie bei alten Schriftstücken oder speziellen Photographien) oder gar Jahrtausende (wie bei tönernen und steinernen Dokumenten) möglich sein dürften. Hier wird dringend nach Lösungen gesucht. Dabei halten sich die Medien – von selbstverschuldeten Ausnahmen abgesehen – sogar wesentlich länger als die dazugehörige Hardware und die notwendigen Programme (s.o.). So sind vielfach alle vorhandenen Daten in bestimmten Zeitabständen (z.B. alle fünf Jahre) auf neue Techniken zu konvertieren. Es wird dann von Migration gesprochen. Eine andere Möglichkeit besteht darin, die alten Systeme (eigentlich ja nur Programme) auf der neueren Technik zu emulieren. Doch dies ist nicht periodisch fortsetzbar. Solche Probleme bestehen z.B. beim Mikrofilm nicht. Er ist immer mit einem typischen Mikroskop lesbar. Deshalb versuchen z.Z. einige Firmen ein möglichst universelles Datenformat zu finden. Auch an eine indirekte Beschreibung der Daten über Algorithmen bzw. Formeln wird gearbeitet. In einigen Firmen werden daher sogar alle eingehenden Daten selbst auf Kosten ihrer automatischen Textlesbarkeit zu dem (angeblich) universell handhabbaren Bildformat „tiff“ konvertiert. (**Details raussuchen!**)

Für gewisse Daten wird gefordert (s.o.), dass nach ihrer Speicherung/Archivierung – z.B. aus rechtlichen, gesetzlichen oder historischen Gründen – es nicht mehr möglich sein darf, sie zu verändern. Dies ist bei WORMs (**w**rite **o**nce **r**ead **m**ostly), z.B. bei CDs, CD-Rs usw. automatisch erfüllt. Jedoch besitzen sie andere Nachteile, auf die im Abschnitt ### eingegangen ist. Doch Magnetbänder, die ansonsten vorteilhafte Eigenschaft, sind dagegen immer wieder neu beschreibbar. Das ermöglicht ein unzulässiges Manipulieren der Daten. So wurde in den letzten Jahren eine WORM-Funktionalität mittels einer speziellen Software entwickelt, die in die Hardware integriert ist. In einigen Fällen genügt sie sogar den höchsten staatlichen Forderungen und ist daher auch für entsprechende Daten zulässig geworden. Ein Beispiel sind DLT-Laufwerke (s.u.) mit „DLT-Ice“. Schließlich sei erwähnt, dass bei Archivierung vor allem zwei Varianten der *Nutzbarkeit* unterschieden werden: *Bewahrung der Unversehrtheit* in ursprünglicher Form und *Sicherung der Authentizität* im Sinne von Vertrauenswürdigkeit.

Für die Langzeitlagerung von Magnetbändern sind einige Besonderheiten zu beachten. Leider ist es für sie noch nicht allgemein üblich, dass sie, wie andere Archivmedien, z.B. Filme, Schallplatten, Bild- und Buchoriginale, unter günstigen Bedingungen gelagert werden müssen und einer ständigen Pflege bedürfen. Temperatur- und Feuchtebedingungen sind weniger kritisch als bei anderen Medien. Ausreichend sind etwa 0 bis 30 °C (QIC sogar 5 bis 45 °C) und 5 bis 80 % relative Feuchte (QIC nur 20 bis 80 %; Riemen!). Selbst übliche Magnetfelder, sofern sie nicht sehr groß werden (Abstand beachten!), stören nicht vorhandene Aufzeichnungen. Gute Beschriftung und Ordnung im Archiv sind jedoch, wegen der nicht unmittelbar „sichtbaren“ magnetischen Aufzeichnung, wichtiger. Beim Aufwickeln des Bandes wird immer etwas Luft mit eingeschlossen. Sie entweicht langsam. Dann bilden sich im Wickel falten- bzw. schlaufenförmige Lücken und Verdichtungen. Dieser Vorgang wird meist **chinch** genant (*englisch chinch* Wanze). Er kann das Band mechanisch verformen und z.T. überdehnen. Er kann auch zu seitlich hervorstehenden einzelnen Bandwindungen führen, die leicht zu beschädigen sind. Bei älteren Bandtypen ist ein Verkleben der Lagen möglich. Diese Mängel lassen sich durch regelmäßiges und vollständiges Umspulen in Zeitabständen von einigen Jahren völlig vermeiden. In jedem Fall ist ein Umspulen vor der Wiedergabe nach langer Lagerung sinnvoll. QIC-Bänder (s.u.) sollten etwa alle halbe Jahre umgespult werden. Infolge des elastischen Riemens steht ihr Band unter besonderer Spannung. Selbst wenn die Unterlage bei alten Bändern spröde geworden ist, existiert – ganz im Gegensatz zu anderen Medien – die informationstragende Magnetisierung der Aufzeichnung noch in guter Qualität. Dann hat die Wiedergabe lediglich mit einem vorsichtigen Transport zu erfolgen. Wichtig sind auch Sauberkeit und Staubfreiheit in den Lager-, Aufzeichnungs- und Wiedergaberäumen. Teppichböden sind deshalb wesentlich schlechter als Parkettfußböden geeignet. Auch Rauch ist

besonders kritisch (Rauchverbot in Archivräumen!). Anderenfalls werden kleine Partikel in den Bandwickel aufgenommen, drücken sich in die Magnetschicht ein und führen zu Dropouts. Zur Pflege der Geräte werden häufig Reinigungsbänder und -Disketten angeboten. Sie sollten jedoch meist sehr viel seltener angewendet werden, als es das Begleitmaterial empfiehlt. Denn ihre Anwendung bewirkt auch Abrieb bei Köpfen und Führungen, der die Betriebszeit der Geräte verkürzt.

Anmerkung: Mehr aus meinem 3. Band sind per Passwort herunterladbar. Es wird bei Begründung und Zusicherung der Urheberrechte (späterer Druck) individuell mitgeteilt. Details und Inhaltverzeichnis der vorhandenen Abschnitte: siehe „Bücher von mir“ linke Rangierspalte.