

LANGZEITARCHIVIERUNG

Lang ist relativ kurz

DIE DATENARCHIVE IM NEUEN RECHENZENTRUM – EINTAGSFLIEGEN AUS SICHT DER ANALOGEN LANGZEITARCHIVIERUNG.



**Die Bandbibliotheken
im Daten- und Archiv-
raum.**

VON WERNER BAUR

Ein komplettes Stockwerk im neuen Rechnerwürfel des Leibniz-Rechenzentrums – eine Halle mit 500 m² Stellfläche – wird ausschließlich zur Datenhaltung genutzt. Die Halle wird im internen Sprachgebrauch des Leibniz-Rechenzentrums DAR genannt. DAR steht für Daten- und Archiv-Raum. Das klingt nach Beständigkeit, Stille, nach langen Regalreihen, in denen Tausende Bücher und Aktenordner im Halbdunkel jahrzehntelang friedlich verstauben. Nur ab und an kommt ein Archivar vorbei und sieht nach dem Rechten oder holt sich einen Ordner. Was das Halbdunkel und die langen Regalreihen anbelangt, ist dieses

Bild gar nicht so falsch. DAR hat keine Fenster und die Notbeleuchtung verbreitet nur schummriges Licht. Es gibt tatsächlich lange Regalreihen, in denen viele tausend Magnetbandkassetten lagern und darauf warten, von einem Roboter abgeholt zu werden. Der Rest des Bildes bedarf jedoch einiger Richtigstellungen.

Zehnjähriges Jubiläum

Das Archiv- und Backupsystem des LRZ ist in seiner heutigen Form seit Januar 1996 in Betrieb. Zwar dürfte ein Teil der Daten im DAR durchaus zehn Jahre alt sein, die Medien, auf denen diese Daten gespeichert sind, sind es definitiv nicht: die ältesten Kassetten in den Regalen

des DAR wurden 2003 beschafft. Von „jahrzehntelangem Dahinschlummern“ kann also kaum die Rede sein, auch dann nicht, wenn kein rüder Rechenzentrumsumzug die Archive wachrüttelt.

Im scharfen Kontrast zu traditionellen Informationsspeichern (z. B. dem Buch) spricht man bei der Lebensdauer von Speichermedien in der IT-Welt faktisch nicht über Jahrzehnte und schon gar nicht über Jahrhunderte, sondern einen Zeitraum von wenigen Jahren: Bei keinem Speichermedium kann hundertprozentig sichergestellt werden, dass es nach mehr als 10 Jahren noch gelesen werden kann. Dabei ist nicht allein die physische Haltbarkeit des Informationsspei-

Modernste Klimatechnik sorgt für konstante Temperatur und Luftfeuchtigkeit in den Rechenräumen.



Festplatten (*disk caches*), auf denen die Daten zwischengelagert werden, gehören. Der entscheidende Punkt dabei ist, dass jede Speicherkomponente nicht fest an einem Rechner hängt, sondern weitgehend dynamisch über das Netzwerk dort eingesetzt werden kann, wo sie gebraucht wird.

Ständige Erneuerung

Nicht nur die Datenträger selbst und die zugehörigen Schreib-/Lesegeräte sind dem schnellen technischen Wandel unterworfen, auch Rechner und Platten müssen in kurzen Abständen, etwa alle 5 Jahre, durch leistungsfähigere Geräte ersetzt werden. Nur die Roboter und ihre Schränke (*Libraries*) haben eine etwas höhere Lebenserwartung: hier können durchaus zehn Jahre und mehr erreicht werden. Beim Auszug aus der Barer Straße wurden beispielsweise fünfzehn Jahre alte *Libraries* zurückgelassen. Damit die *Libraries* nicht nach ein paar Jahren aus allen Nähten platzen, wird eine hohe Ausbaufähigkeit („Skalierbarkeit“) schon von vornherein mit eingeplant. Die neueste der drei *Libraries* im DAR bietet in ihrer heutigen Ausbaustufe Platz für 4.500 Kassetten und 64 Laufwerke und kann auf 10.000 Kassetten, im Verbund mit zusätzlichen Systemen auf 300.000 Kassetten erweitert werden.

Gasflaschen mit Argon, das im Brandfall jedes Feuer im Keim erstickt.

Kalt und warm

Im vergangenen Jahrzehnt waren Bänder und Kassetten sicherlich das am weitesten verbreitete Speichermedium für große Datenmengen. Wahrscheinlich, aber keineswegs sicher ist dies auch für dieses Jahrzehnt noch der Fall. Die magnetische Festplatte etwa verfügt heutzutage über ein ähnliches Speichervolumen wie die Kassetten des DAR.



Warum also nicht die auf den ersten Blick gar nicht so viel teureren Festplatten statt Bänder verwenden? Eine Antwort darauf ist die monatliche Heizungsrechnung. Um Energiekosten zu sparen ist es

im DAR - nicht etwa sehr kalt, im Gegenteil, es ist sehr warm. Die im Raum installierte Hardware produziert diese Wärme. Eine starke Absenkung der Raumtemperatur würde die Klimakosten deutlich in die Höhe treiben. Daher herrscht im DAR die maximale Temperatur, in der sich die Kassetten noch „wohl fühlen“.

Würde man das gleiche Datenvolumen auf Festplatten ablegen, würde man jahrein jahraus doppelt zahlen: einmal für die Stromkosten, die der Betrieb der Platten erfordert und einmal für die Klimakosten, die entstehen, um die von den Platten produzierte Wärme wieder abzuführen.

Kein Raum für Katastrophen

Der Albtraum jeder Bibliothek ist ein Brand im Allerheiligsten. Noch nicht lange zurück liegt der Brand in der Herzogin Anna Amalia-Bibliothek in Weimar. Was das Feuer nicht vernichtet, zerstört das Lösch-

wasser. Hier stößt auch die Lebensdauer der gedruckten Information, ansonsten der Lebensdauer der digitalen Datenträger weit überlegen, an die gleichen Grenzen. Selbstredend wäre Feuer und Löschwasser

auch für die Daten im DAR fatal. Ein ausgeklügeltes System von Sensoren sorgt dafür, dass ein ausbrechender Brand sofort bemerkt wird. Gelöscht wird mit Argon, das den Luftsauerstoff so weit reduziert, dass jedes Feuer im Keim erstickt wird. Das Verfahren ist für die Datenträger völlig unschädlich, die Kosten sind allerdings erheblich. Sollte es trotz aller Vorsichtsmaßnahmen aber doch einmal zu einer völligen Zerstörung des Rechenzentrums kommen, bleiben immer noch die Nachbarn: Die wichtigsten Daten werden schon seit Jahren an das einige hundert Meter entfernte Rechenzentrum der Max-Planck-Gesellschaft in Garching kopiert.

Um sich mögliche Ursachen für den Datenverlust vorzustellen, müssen übrigens nicht gleich die Naturgewalten bemüht werden. Es genügt die Zerstörung der Datenbanken, in denen für jede der über zwei Milliarden Dateien Buch darüber geführt wird, auf welcher Kassette welche Version von welchem Benutzer von welchem Rechner aus zu welchem Zeitpunkt gespeichert worden ist. Diese Datenbanken bilden somit das sorgsam behütete Herzstück des Systems. Zahlreiche IT-technische Maßnahmen sorgen dafür, dass die Besitzer der Daten und vor allem die verantwortlichen Systemadministratoren nachts ruhig schlafen



können. Beispielsweise verkräftet das System durch eine Kombination verschiedener Schutzmechanismen bis zu vier gleichzeitige Plattenausfälle, ohne dass Daten verloren gehen. Ersatzplatten übernehmen dabei automatisch die Funktion der defekten Platten.

Gestern noch viel, heute ein Nichts

Die Datenmenge, die vor gut zehn Jahren zur Speicherung eine ganze Bibliothek am LRZ erfordert hatte, passt heute spielend auf ein bis zwei Kassetten. Müssten wir umgekehrt die Datenmenge, die heute in den Archiven des DAR liegt, auf Medien mit der Kapazität von damals unterbringen, würden die 500 m² des DAR nur für einen Bruchteil der benötigten 1,2 Mio. Kassetten ausreichen.

Die technische Entwicklung eilt mit den gleichen Riesenschritten

voran, mit denen auch der Speicherbedarf wächst. Nur deshalb ist es überhaupt möglich, die entstehende Datenflut zu bewältigen.

Der Terminus „Bibliothek“ passte übrigens 1995 noch ganz gut für die elektronischen Archive: eine Kassette entsprach in etwa dem Informationsgehalt eines Buches. Heute passt der Inhalt von 5000 Büchern auf eine Kassette.

	1996	2006
Datenmenge im Archiv des LRZ	1.000 GB	1,2 Mio GB
Kapazität einer Datenkassette	10 GB	500 GB
Aufzeichnungsgeschwindigkeit	9 MB/sek	120 MB/sek

Regalreihen einer IBM-Library gefüllt mit mehreren Tausend Kassetten.

Der Weltgeist in Garching

Werfen wir zum Schluss noch einen Blick in ein anderes Datenarchiv, nämlich in das des Zeitmagazins. Beim Stöbern stößt man dort auf einen Artikel, demzufolge der Weltgeist ein Speichervolumen von 1.300 Petabytes oder 1,3 Milliarden Gigabytes hat. Stetigen linearen Fortschritt der Speichertechnologien vorausgesetzt, lässt das Raum für Visionen: In wenigen Jahren ist der Weltgeist in einem schummrigen Datenraum in einem kleinen Ort bei München, Oberbayern, unterzubringen.

Der Autor leitet die Gruppe „Datei- und Speichersysteme“ am Leibniz-Rechenzentrum.



Ein Roboter-Greifarm holt Kassetten aus einer Library.

