

Informationsverlust durch die Digitalisierung

Balázs Bárány

1 Einleitung

Ein immer größerer Teil unseres Wissens und unserer Kultur ist digital. Das bedeutet, daß die geistigen Schöpfungen häufig gleich im Computer entstehen, meist am Computer weiterbearbeitet und wiederum über Computer(netzwerke) an andere Menschen weitergegeben werden.

Die digital gespeicherten Informationen sind sehr fragil; dies hängt mit vielen Eigenschaften der digitalen Technologie (bzw. ihrer heutigen Ausprägung) zusammen, und ist sehr schwer zu vermeiden.

1.1 Bekanntheit des Problems

Die wissenschaftliche Gemeinschaft wurde erstmals 1995 mit dem Artikel „Ensuring the Longevity of Digital Documents“ von Jeff Rothenberg im Scientific American auf das Problem aufmerksam gemacht; Rothenberg präsentiert dort auch einen Lösungsvorschlag. Seitdem werden in den damit befaßten Kreisen die unterschiedlichen Methoden der digitalen Langzeitverfügbarkeit diskutiert.

In der Öffentlichkeit und der Industrie ist jedoch kaum ein Bewußtsein für die Problematik vorhanden. Das behindert die Verbreitung besserer Technologien und Handlungsweisen, mit denen der Informationsverlust verlangsamt oder vermieden werden könnte.

Es sind bereits wichtige digital gespeicherte Informationen verlorengegangen. Gut dokumentiert sind vor allem Fälle in den USA: dort waren etwa Daten der Volkszählung von 1960 in den 1970-er-Jahren sowie Daten der NASA von verschiedenen Weltraum-Missionen nicht mehr lesbar.

1.2 Digitalisierung der Information

Digitale Daten sind ausschließlich mit technischen Hilfsmitteln lesbar. Doch die Lesbarkeit der Daten selbst bedeutet noch nicht, daß wir die enthaltene Information extrahieren und interpretieren können.

Digitale Datenspeicherung bedeutet, daß der abzubildende Aspekt des Originals in eine Zahlendarstellung umgewandelt (wenn er nicht ursprünglich schon aus Zahlen bestand) und so gespeichert oder weiterverarbeitet wird. Diese Zahlendarstellung kann als Eingabe für logische und interpretierende Prozesse dienen, und ihre Richtigkeit kann bei weiteren Kopierschritten sichergestellt werden, indem die Zahlen der Kopie mit denen des Originals verglichen werden. Fehler im Datenstrom sind erkennbar und unter Umständen korrigierbar.

Die Digitalisierung bietet große Vorteile, weil sie die Verarbeitung der Information vereinfacht und damit verbilligt, zur Konvergenz unterschiedlicher Informationsarten beiträgt und den Zugang zur Information stark verbessern kann.

2 Lebensdauer der Information

Für die Datenspeicherung im Computer brauchen wir Datenträger und zu ihnen passende Lesegeräte (manchmal sind beide zusammengebaut, wie bei Festplatten). Die Datenträger und

Lesegeräte bleiben unterschiedlich lange verwendbar. Die Strukturierung der Daten auf den Datenträgern beeinflusst die zukünftige Interpretierbarkeit der Information.

2.1 Datenträger und ihre Lebensdauer

Magnetische Datenträger basieren auf dem Prinzip, daß in einigen Metallverbindungen der Kontakt mit einem elektrischen Feld den magnetischen Zustand der Partikel ändert. Diese Änderung ist ziemlich permanent und stabil, solange kein anderes Feld wirkt. Die Magnetfelder lösen wiederum in anderen Metallen elektronische Zustandsänderungen aus. Auf diese Weise wird die Information wieder gelesen.

Magnetische Datenträger sind nicht sehr langlebig. Bei Disketten sind einige Jahre Nutzungszeit das Maximum, bei Festplatten ist es ähnlich. Magnetbänder und -kassetten sind etwas länger haltbar, aber auch nur etwa 10-15 Jahre. Die Datenträger sind durch Magnetfelder, elektromagnetische Strahlung, Licht und natürliche Entmagnetisierung gefährdet.

Bei optischen Datenträgern (z. B. CD, DVD) geschieht das Lesen mit Hilfe eines Laserstrahls. Der Strahl wird auf die Oberfläche des Datenträgers projiziert und dort entweder reflektiert oder abgelenkt. Die reflektierten Strahlen oder ihr Mangel werden als Daten interpretiert.

Industriell massengefertigte optische Datenträger werden gepreßt, die Lichtbrechung entsteht durch den Wechsel von Erhöhungen und Vertiefungen in der Datenträgerschicht. Selbst beschreibbare Datenträger hingegen enthalten meist organische Farbstoffe oder spezielle Metalllegierungen, die mit einem im Vergleich zum Lesen viel stärkeren Laserstrahl permanent (CD-R) oder immer wieder änderbar (CD-RW) dazu gebracht werden, ihre Reflexionseigenschaften zu ändern und damit den gleichen Effekt zu erreichen.

Kratzer, chemische Einflüsse, Licht, Hitze, hohe Luftfeuchtigkeit und Verformungen können allesamt die Lesbarkeit optischer Datenträger beeinträchtigen. Für die gepreßten Datenträger sind bei richtiger Lagerung um die 100 Jahre Lebensdauer möglich, einmal beschreibbare CDs dürften 10-15 Jahre halten, DVDs und wiederbeschreibbare Medien wahrscheinlich etwas kürzer. Wegen der extrem großen Verbreitung von CDs und DVDs scheint die Abwärtskompatibilität der neuen Lesegeräte zumindest für die nächsten Jahrzehnte gesichert.

Es gibt Materialien, die ihre elektrische Ladung permanent speichern können. Die auf diesem Effekt basierenden Datenträger heißen Flash-Speicher. Sie sind noch relativ teuer und langsamer beschreibbar und auslesbar als andere Datenträgerarten. Da Flash-Speicher aber eine permanente Datenspeicherung auf kleiner Fläche ohne bewegliche Teile bieten, werden sie mit fallenden Preisen und steigenden Kapazitäten immer beliebter.

Umwelteinflüsse beeinträchtigen Flash-Speicherkarten weniger als die anderen Datenträger, wodurch diese wahrscheinlich länger haltbar sind; allerdings gibt es mit ihnen noch keine Langzeiterfahrungen. Die Lesegeräte und die Medien entwickeln sich rasant weiter, weswegen die zukünftige Versorgung mit Lesegeräten für ältere Karten-Typen unsicher erscheint.

Für Lesegeräte sind Schnittstellen zum Computer notwendig. Bei verschiedenen Computertypen sind diese Schnittstellen häufig unterschiedlich, und sie ändern sich auch pro Computerplattform ca. alle 10-15 Jahre oder öfter. Aus diesem Grund sind alte Lesegeräte, selbst wenn sie noch funktionsfähig wären, mit aktuellen Computern selten einfach nutzbar.

Lesegeräte sind Präzisionsgeräte, die großen mechanischen und thermischen Beanspruchungen ausgesetzt sein können. Ihre Lebensdauer ist also auch begrenzt, sie dürfte in der Regel ein Jahrzehnt nicht oder kaum überschreiten.

2.2 Dateisysteme

Dateisysteme dienen dazu, Verzeichnisse und Dateien auf den Datenträgern abzulegen. Heute sind über 30 verschiedene Dateisysteme in Verwendung, früher gab es noch viele weitere. Betriebssysteme unterstützen selten eine größere Anzahl von Dateisystemen, aber ohne Unterstützung im Betriebssystem sind die Dateien innerhalb der Dateisysteme nicht einfach zugänglich.

2.3 Dateiformate

Es gibt eine große Anzahl von Dateiformaten, da unterschiedliche Anforderungen der Software und der Nutzung auch unterschiedlicher Vorgangsweisen bei der Strukturierung der Information bedürfen.

Bei Textdateien ist die Kodierung das häufigste Problem. Sie legt fest, welchem Zeichen eines Alphabets eine im Computer gespeicherte binäre Zahl entspricht. Heute stehen eindeutige und alle Alphabete abdeckende Kodiersysteme zur Verfügung, aber der Großteil der Dateien liegt immer noch in alten Kodierungen vor. Abgesehen von der Kodierung können auch Konventionen wie das Zeichen zur Markierung von Zeilenenden unterschiedlich sein.

Sogenannte binäre Dateiformate speichern die Daten sehr kompakt oder ermöglichen spezielle, besonders schnelle Arten des Zugriffs auf die Daten. Aus diesem Grund wurden sie früher, als Speicherplatz knapp und die Geschwindigkeit der Computer gering war, sehr viel verwendet; sie sind auch heute noch ziemlich verbreitet. Für den Zugriff auf binäre Dateiformate ist fast immer die Kenntnis der internen Arbeitsweise des dazugehörigen Programms notwendig, da die Datei selbst nicht genug Information über die Bedeutung der einzelnen Datenelemente enthält.

Bei Dateiarten, die viel Speicherplatz brauchen (Bilder, Videos, Musik usw.), sind reduzierende Methoden der Datenspeicherung verbreitet. Die Reduktion führt dazu, daß die entstehende Datei zwar kleiner ist als ohne Reduktion, aber die Daten werden geringfügig (für Menschen in der Regel kaum bis gar nicht feststellbar) verändert. Eine Wiederherstellung der ursprünglichen Information ist nicht mehr möglich. Aus diesem Grund sollten reduzierende Verfahren in der Langzeitarchivierung nicht verwendet werden.

Dateiformate sind dann für die Langzeitverfügbarkeit geeignet(er), wenn sie möglichst simpel, selbstdokumentierend (z. B. mit Hilfe von internen Strukturelementen), offen und verbreitet sind. Solche Formate dürften noch für eine längere Zeit verwendbar sein, allerdings liegen derzeit die meisten Informationen nicht in solchen Formaten vor. Heute besteht ein gewisser Konsens darüber, daß XML-basierte (eXtensible Markup Language, ein strukturiertes Textformat) Formate die längste Haltbarkeit von allen heute bekannten Formaten haben dürften.

2.4 Software

Jede nicht-triviale Software trifft Annahmen über die Umgebung, in der sie ablaufen soll. Wenn diese Annahmen nicht korrekt sind (etwa auf einem anderen Computersystem, unter einem anderen Betriebssystem oder wenn andere Voraussetzungen fehlen), läuft die Software nicht oder nicht richtig.

3 Lösungsansätze

Es wurden mehrere Lösungsansätze für die langfristige Verfügbarkeit der Information vorgeschlagen. Sie alle erfordern gute Kenntnisse über aktuelle und vergangene Computertechnologien und ständige Vorbereitung, Datenpflege und Überprüfung der alten und neuen Informationen.

Wenn eine Institution alte Computer, Zubehör und Software in einem „Computer-Museum“ aufhebt und durch ständige Wartung ihre Funktionsfähigkeit sicherstellt, kann sie den Zugang zu den Informationen so sichern, daß diese in ihrer ursprünglichen Umgebung erscheinen. Allerdings ist die Lebenszeit der Hardware selbst bei optimaler Wartung auf höchstens einige Jahrzehnte begrenzt. Alte Computer haben auch keine Schnittstellen zu aktuellen, wodurch die Daten in ihnen „gefangen“ sind.

Die Migration soll sicherstellen, daß Dateien in jeweils aktuellen Dateiformaten zur Verfügung stehen. Alte Dateiformate werden dafür in neue konvertiert. Allerdings eignen sich dafür nicht alle Dateiformate; bei solchen Dateien treten im Zuge der Umwandlung Verfälschungen auf. Bei vielen verbreiteten Formaten ist die Konversion jedoch leicht automatisierbar und verifizierbar, also für die Langzeitarchivierung geeignet. Da die migrierte Information in jeweils aktuellen, üblichen Formaten zur Verfügung steht, ist ihre Benutzung leicht.

Mit Hilfe der Emulation können neuere Computer die Funktionsweise von älteren abbilden. Auf diese Weise sind Original-Betriebssysteme und -Software auch ohne passende Hardware einsetzbar. In Emulatoren können ältere Emulatoren laufen, wodurch die Emulation auch in mehreren Schichten, bis hin zu den ältesten zu emulierenden Computern, möglich ist. Die Emulation ermöglicht also theoretisch, alle früheren Systeme ohne Zeitbegrenzung auf jeweils aktuellen Computern auszuführen. Allerdings gibt es in der Praxis einige Probleme (z. B. Bindung von Software an eine bestimmte Hardware; Online-Freischaltung der Nutzung von Software und Dokumenten), die dazu führen, daß für die betroffenen Informationen auch die Emulation nicht geeignet ist. Da noch keine besseren Verfahren gefunden wurden, kann es durchaus passieren, daß diese Informationen verlorengehen.

Für die Praxis dürfte ein kombinierter Ansatz aus Migration und Emulation am zielführendsten sein. Für unterschiedliche Dateitypen kann auf Grund von Erfahrungen und Tests festgestellt werden, ob die (wegen des leichteren Zugangs präferierte) Migration die Daten geeignet sichert oder ob die Vorbereitung einer Emulationsumgebung notwendig ist.

3.1 Rechtliche Probleme der digitalen Archivierung

Bei digitaler Information behindert das Urheberrecht wegen der Methoden der Emulation und Migration und der kurzen Lebensdauer die Sicherung der Langzeitverfügbarkeit viel stärker als bei herkömmlichen analogen Daten. Einige notwendige Schritte sind (in Österreich seit 2003) sogar mit Gefängnisstrafen bedroht. Dies wird, wenn für legitime Archivierung keine Ausnahmeregelungen geschaffen werden, zu weiteren, technisch nicht notwendigen Informationsverlusten führen. Andere rechtliche Regelungen wie Patente und Pflichtexemplar-Vorschriften können in einigen Fällen auch die Langzeitverfügbarkeit beeinflussen, sie sollten daher, wenn nötig, so geändert werden, daß sie die langfristige Verfügbarkeit digitaler Daten vorsehen und erleichtern.

4 Schlußfolgerungen

Die heute verbreiteten Datenträger halten ihre Daten nur einige Jahre oder wenige Jahrzehnte lang zugänglich. Die Lesegeräte werden häufig in noch kürzerer Zeit kaputt oder sie lassen sich nicht an aktuelle Computer anschließen. Die reinen Daten lassen sich zwar mit Hilfe von Netzwerken häufig von veralteten Computern auf neuere kopieren, aber ohne geeignete Software nicht darstellen oder interpretieren. Software ist jedoch an eine gewisse Umgebung angewiesen, die auf späteren Computern nicht einfach herstellbar ist. Die Methoden, all diese Probleme zu lösen, sind ziemlich aufwendig, und führen nicht einmal in allen Fällen zum Ziel.

Es hat für die Hersteller von Hard- und Software kaum Vorteile, wenn sie Langzeitverfügbarkeit in ihre Produkte einbauen, im Gegenteil: Wenn sie alte Formate und Standards „mitschleppen“ müssen, kann das ihre Kosten erhöhen und sie im sehr intensiven Wettbewerb am Markt behindern. Die Inhaltsproduzenten sind auch eher daran interessiert, daß die Mehrheit alter Inhalte verlorengeht.

Für kommerziell verbreitete Inhalte und mittlerweile auch für eigene Dokumente werden Technologien in den Markt gedrückt, die durch Verschlüsselung und ähnliche Verfahren den „unberechtigten“ Zugang zur Information unterbinden. Die Überprüfung der Berechtigung setzt komplexe Systeme voraus, deren langfristige Verfügbarkeit sehr fraglich erscheint, und deren Umgehung auch für Archivierungszwecke strafbar ist.

Selbst die Kombination aller vorgeschlagenen Verfahren löst nicht jedes Problem der Langzeitarchivierung digitaler Information. Die Migration eignet sich nicht für alle Arten von Dateien. Komplexe sowie zusammengesetzte Dateiformate sind auf ihre ursprüngliche Software-Umgebung angewiesen, diese kann in vielen Fällen mit Hilfe der Emulation geschaffen werden. Einige Elemente der Software-Umgebung können jedoch weitere, externe Anforderungen stellen. In diesen Fällen hilft auch die Emulation nicht immer weiter. Manche Lösungsansätze, die technisch möglich wären, fallen unter das verschärfte Urheberrecht und werden mit Gefängnisstrafen bedroht.