

DAS GROSSE VERGESSEN

DATENSCHWUND IM DIGITALEN ZEITALTER

CHRISTINE PLASS

Wenn es um das Verschwinden digitaler Daten geht, reicht die Spannweite der Emotionen von Gleichmut bis Panik. Gelassen sehen die einen einer Zukunft entgegen, in der ein Großteil unseres »Datenmülls« verschwunden sein wird, während andere vor dem katastrophalen Informationsverfall warnen, der schleichend und stetig voranschreitet und das kulturelle Gedächtnis auszulöschen droht. Manche glauben gar, dass die explosionsartige Entwicklung der Informationstechnologie im 20. Jahrhundert dazu führt, dass diese Epoche eine der am schlechtesten dokumentierten in der Geschichte sein könnte: ein dunkles digitales Zeitalter.

Dabei ist der Verlust von Dokumenten und kulturellen Errungenschaften so alt wie der Versuch, sie zu bewahren. Infolge von Kriegen, Katastrophen, der systematischen Zerstörung durch Menschen, aber auch durch den Zahn der Zeit, der an Schriftträgern nagt, kam es seit jeher zu unermesslichen Verlusten von Medien, die Wissen transportierten. Und so knüpft sich an die digitale Revolution, die dritte Kulturrevolution nach Erfindung der Schrift und des Buchdrucks, auch die Hoffnung, Dokumente vergangener Epochen, die jetzt auf Papier oder anderen vergänglichen Materialien zu zerfallen drohen, durch Digitalisierung zu bewahren. Ob dies gelingt, ist ungewiss.

Wie haltbar sind Datenträger?

Mit der Erfindung des Computers war der Verlust von digitalen Daten programmiert. Eben noch da Gewesenes ist von einem Mausklick zum nächsten unwiederbringlich verloren, sei es, dass man zu speichern versäumt, bevor das Programm abstürzt oder dass die Festplatte eines unerwarteten Tages ihren Geist aufgibt. Selbst wer konsequent und kontinuierlich Sicherheitskopien auf externen Speichermedien anfertigt, kann nicht sicher sein, dass diese Daten dem Verschwinden entrinnen werden. Über die Haltbarkeit von Datenträgern gibt es bis heute keine gesicherten Erkenntnisse, die man getrost für jedes Speichermedium annehmen

kann. Während eine CD-ROM in 200 Jahren noch lesbar sein mag, ist eine andere schon nach wenigen Monaten hinüber. Dies gilt ebenso für Disketten. Das Magazin *Speicherguide.de* räumt CD-ROM und DVD-ROM weniger als fünf und bis zu 200 Jahren, Disketten weniger als 30 und bis zu 100 Jahren Lebensdauer ein (vgl. Fröhlich/Wagner 2003). Dagegen gibt die Online-Enzyklopädie Wikipedia Disketten nur fünf bis zehn Jahre, CD-ROM bis zu 100 Jahren und Magnetbändern bis zu 30 Jahren (Wikipedia 2005).

Das eigentliche Problem besteht jedoch darin, dass die meisten Speichermedien schon obsolet sind, bevor sie unlesbar werden, weil kein System und kein Programm mehr in Gebrauch ist, das etwas mit ihnen anfangen könnte. Wenn ich wissen will, was ich vor gerade mal 15 Jahren über Adorno dachte, muss ich mir erst auf dem Flohmarkt einen alten Computer suchen, der 5¼-Zoll-Floppy-Disks lesen kann und mit einer Textverarbeitung ausgestattet ist, an die ich mich nicht mehr erinnere. In wenigen Jahren werden auch 3½-Zoll-Laufwerke verschwunden sein, schon jetzt gehören sie nicht mehr zur Standardausrüstung.

Dass bisherige Datenträger ihre Funktion nur unvollkommen erfüllten, ist nicht der Grund für ihre kurze Lebensdauer. Immer größer werdende Datenmengen erforderten schnellere Speichermedien mit größeren Kapazitäten. Auch in Zukunft werden wir es mit immer größeren Datenmengen zu tun haben; zudem wächst das Bewusstsein dafür, dass wir zuverlässige Datenträger benötigen. Hier könnte sich eine Technik anbieten, die schon vor 5.000 Jahren äußerst erfolgreich war: Das Eingravieren in Festkörper. Keine andere Schrift hat so lange überdauert wie die auf Ton oder Stein. Als Datenspeicher der Zukunft könnten Lithiumniobat-Kristalle fungieren, auf die mit Laser geschrieben wird. Ein Kristall von der Größe eines Zuckerstücks kann ein Terabyte Daten speichern (Denz 2003). Das entspricht etwa 1.500 CD-ROM, auf jede passt einmal der komplette Text der Bibel. Man schätzt, dass die gespeicherten Informationen bis zu 1.000 Jahre haltbar sein sollen.

Das kulturelle Gedächtnis einer digitalen Epoche

»Die Menschen werden mit Hilfe der Technologie in ein paar Jahren mehr Daten generieren als jemals zuvor in ihrer Geschichte. Damit wird die Kluft zwischen der Fähigkeit, Daten zu erzeugen und der Fähigkeit sie zu erhalten und zu verwalten immer breiter«, erklärt Neil Beagrie von der British Library (SAP-INFO 2004). Dies betrifft wissenschaftliche Dokumente, private und wirtschaftliche Internetseiten, Online-Magazine,

Netzkunst, Literatur – eben alle Bereiche, die es mit Wissen zu tun haben, das digital gespeichert ist. Aus Sicht von Bibliothekaren und Archivaren, die vor der Herausforderung stehen, dieses Wissen aufzubewahren, könnte man es auch so formulieren: Während Informationen billig und schnell erzeugt werden können, kostet ihre dauerhafte Speicherung viel Zeit und Geld.

Unter der Leitung der Library of Congress startete in den USA im Jahre 2000 das »Nationale Programm für Digitale Informationsinfrastruktur und Aufbewahrungen« (NDIIPP) [1]. Mit einem Budget von über 100 Millionen US-Dollar sollen neue Archivierungsmethoden und Strategien zum Umgang mit digitalen Materialien aller Art entwickelt werden: Internetseiten, digitale Zeitschriften und Bücher, digital aufgezeichnete Ton-, Film- und Fernsehdokumente. In Deutschland läuft seit Sommer 2003 das mit 800.000 Euro geförderte Projekt »Nestor« [2]. Es soll ein Kompetenznetzwerk zur Langzeitarchivierung und -verfügbarkeit digitaler Quellen aufbauen.

Die Vergangenheit des Internet in der Zukunft

Ein besonders schwierig zu archivierender Bereich digitalen Kulturguts ist das Internet selbst. Auch wenn das Internet gigantische Mengen an Informationen bereitstellt, ist es kein Wissensspeicher, sondern ein flexibles Netzwerk, in dem Informationen fortlaufend generiert, verbreitet, verändert und gelöscht werden. Die durchschnittliche Lebensdauer einer Internetseite soll 44 Tage betragen. Sie ist also ein sehr flüchtiges Phänomen und dazu auch noch komplex: »Jede Website verweist im Schnitt auf 15 andere Seiten und enthält fünf Objekte wie Bilder, Grafiken, Videos, Tondateien, Werbung« (Heuer/Kern 2005: 93).

Bisherige Strategien, Informationen zu bewahren, ließen von vielen Websites nur noch Text übrig. Aus medienhistorischer Sicht ist das fatal, macht doch die Verbindung von Grafik, Bildern und Texten das Medium Internet aus und nicht nur die Information. Die sind aber besonders aufwändig zu speichern. Ein weiteres Problem betrifft den Kontext von Websites: Künftige Historiker werden nicht nur vollständige Websites benötigen, sondern auch Kriterien, die es ihnen ermöglichen, die Relevanz, den Wert, die Herkunft und die Authentizität von Websites einzuordnen. Welche Organisationen dahinter steckten und wer die Seiten gelesen hat, wird nicht immer ersichtlich sein. Geben die verbreiteten Meinungen den Common Sense wieder oder eine von ganz wenigen geteilte Weltansicht? Angesichts von geschätzten 170 Terabyte Internet ist es undenkbar, alle gespeicherten Internetseiten mit Zusatzinformationen zu

versehen. Es ist ja nicht einmal möglich, alle zu erfassen. Zurzeit entscheiden Roboter darüber, welche Internetseiten der Nachwelt erhalten bleiben. Das sind Programme, die im Dienste ihrer Suchmaschinen das Web nach Inhalten durchsuchen. Dank seiner Roboter verfügt Google über ein riesiges Internet-Archiv mit einem Bestand von mehreren Milliarden Websites. Nach welchen Kriterien die programmiert wurden, bleibt Googles Geheimnis.

»The Wayback Machine« [3, 4], eine Initiative zur Erinnerung an das Internet, findet (laut Auskunft auf der Website) 30 Billionen längst vom Server genommene Internetseiten ab 1996 und macht sie kostenlos öffentlich zugänglich. Aber auch dieses Internet-Archiv ist nicht in der Lage, alle ab diesem Zeitpunkt existierenden Seiten wiederzugeben. Viele Internetseiten sind nicht verlinkt und werden daher von den Robotern nicht gefunden. Inhaber von Websites können diese von der Archivierung ausschließen, indem sie eine robots.txt-Datei auf ihrem Webserver installieren oder gegen die Speicherung ihrer Seiten Einspruch einlegen, was beispielsweise von Verlagen, die ihre Erzeugnisse kostenpflichtig ins Netz stellen, auch rege genutzt wird.

Was bleibt von mir?

Während man davon ausgehen kann, dass in Behörden, Firmen und Bibliotheken Strategien gefunden werden, um Dokumente für die Zukunft zu retten, bleibt das Bewahren der eigenen digitalen Informationen jedem selbst überlassen. Das betrifft längst nicht nur administrative Dokumente, Briefwechsel mit Behörden oder in Studium und Beruf erstellte Werke, sondern auch zunehmend Privates. Wer schafft es schon, regelmäßig Fotos von der Festplatte auf CD und von CD auf eine neue CD oder auf DVD zu kopieren oder gar in ein anderes Bildformat zu konvertieren? Und wer guckt sie sich überhaupt noch an? »Wer nicht regelmäßig Ordnung in seinen Datenmengen hält, wird in 20 Jahren hilflos vor seinem riesigen Fundus an Fotos und Filmen sitzen, die so geheimnisvolle Namen tragen wie p1020166 oder cimg0151«, mahnt Iris Rodriguez (2005: 95) in der Zeitschrift *Neon*.

Was soll dazu erst die Nachwelt sagen? Droht die nicht in der Vielzahl von Bildern über jeden möglichen und unmöglichen Moment unseres Daseins zu ertrinken? Falls sie überhaupt noch Software besitzt, mit der sich vorsintflutliche Bildformate lesen lassen. Und bereit ist, die schlechte technische Qualität von Bildern zu ertragen, die mit vorsintflutlichen Digitalkameras geschossen wurden. Ist es nicht eher tröstlich, dass eine Vielzahl von belanglosen, Überdruß erzeugenden Bildern, nur

mit Zahlen versehen und irgendwo auf der Festplatte abgespeichert, gnädig der Vergessenheit anheim fallen, dezent einfach verschwinden? Bilder sind nur ein Teil unserer Lebensäußerungen, ein Abdruck unseres Lebens und unserer Persönlichkeit, SMS und E-Mail ein anderer. Die meistgebrauchten schriftlichen Kommunikationsmittel unserer Zeit sind nicht nur für unsere Enkel, sondern auch für zeitgenössische Soziologen und künftige Historiker wichtige Informationsträger. Sie sagen etwas aus darüber, wie wir kommuniziert und gelebt haben, worüber wir uns gefreut und was wir gehasst haben, wie wir uns Liebe erklärt und Beziehungen beendet haben.

Offensichtlich bedarf es eines Programms, das in der Lage ist, einfach alles zu archivieren: Telefongespräche, Urlaubsfilme, E-Mails, Schnappschüsse und vieles mehr. Es heißt MyLifeBits [5] und befindet sich gerade in der Projektphase. Sein Erfinder, Microsoft-Senior-Researcher Gordon Bell arbeitet daran, persönliche Lebensäußerungen wie Briefe, Fotos, Telefongespräche, Filme oder E-Mails so zu digitalisieren, dass sie anhand von einfachen Stichwörtern gefunden werden können. Das Archiv soll als eine Art »Ersatzgehirn« funktionieren. Es soll Menschen geben, die verlegen ihre Briertasche und suchen sie dann bei Google. Mit MyLifeBits hätten sie zumindest die Chance zu erfahren, wo und wie sie sie bislang wiedergefunden haben. »Someday, MyLifeBits will allow people to google their own lives«, wird Bell in *Wired* (Scheeres 2002) zitiert.

Während sich die Ersten fragen, ob dies nun eine neue Ausgeburt des Zeitalters des Narzissmus oder der Vergesslichkeit sei, bleibt weitgehend undiskutiert, ob die in welchen Formaten auch immer gespeicherten Dokumente zehn Jahre später noch abrufbar sind? Oder was die Nachkommen damit anfangen sollen? Wird es ihnen, angesichts künftiger komfortablerer Suchmaschinen nicht viel zu umständlich erscheinen, sich durch das Leben ihrer Vorfahren zu googeln?

Überlebenschancen digitalen Wissens

Mit der Speicherung von Wissen in digitalen Dokumenten sind wir zum ersten Mal seit Erfindung der Schrift darauf angewiesen, Maschinen zu besitzen, die in der Lage sind, uns Zugang zu gespeichertem Wissen zu verschaffen. Jahrhundertelang hatte es genügt, dafür zu sorgen, dass die Nachkommen sprechen und lesen lernten und in der Lage waren, ein Buch zur Hand zu nehmen und Texte zu verstehen. Zukünftige Generationen werden überdies lernen müssen, wie sie Maschinen benutzen, die Daten entziffern können. Vielleicht werden sie sogar lernen, wie sie Datenträgern Informationen entnehmen, die sie nie gesehen haben? Solange

sich die Maschinen, die zum Entziffern digitaler Daten nötig sind, fortwährend und beschleunigt weiterentwickeln, sind wir darauf angewiesen, Daten in immer kürzeren Abständen von einem Dateiformat ins andere, von einer Software zur nächsten, von einem Betriebssystem zum nachfolgenden zu übertragen. Dabei ist es kaum sinnvoll, alles zu bewahren. Es spricht einiges dafür, dass wir E-Mails, die wir schon nicht lesen wollten, Datenmüll, der uns schon zur Verzweiflung brachte, die Informationsflut, in der wir zu ertrinken drohen, nicht auch noch nachfolgenden Generationen aufbürden können. Zurzeit besteht eine unserer größten Herausforderungen darin, uns individuell und gemeinsam zu überlegen, was wir kurz- und langfristig erhalten wollen. Weiterhin wird es darum gehen, bessere Techniken bereitzustellen, die es ermöglichen, Wissen zu bewahren und zu organisieren. Wenn es nachfolgenden Generationen dann noch gelingt, Lesen zu lernen, haben wir gute Chancen, dass ein wenig von unserem Wissen übrig bleibt.

Literatur

- Denz, Cornelia (2003): Volumenholographische Speicherung. [WWW-Dokument: www.uni-muenster.de/Physik/AP/Denz/?id=research_storage&lang=de - Zugriff: 28.01.2005].
- Fröhlich, Karl und Oliver Wagner (2003): Haltbarkeit von Datenträgern. [WWW-Dokument: www.speicherguide.de/magazin/special0603.asp - Zugriff: 28.01.2005].
- Heuer, Stefan und Thomas Kern (2005): Online in die Vergangenheit. In: Technology Rewiev, 1, 92-95.
- Rodriguez, Iris (2005). Weißt du noch? In: Neon, 1, 94-98.
- SAP-INFO (2004): Digitale Informationen werden niemals rein zufällig überleben. [WWW-Dokument: <http://www.sapinfo.net/public/en/index.php4/article/Article-3089140c577c931a92/de/articleStatistic> - Zugriff: 25.01.2005].
- Scheeres, Julia (2002): Saving Your Bits for Posterity. [WWW-Dokument: <http://www.wired.com/news/technology/0,1282,56734,00.html> - Zugriff: 28.01.2005].
- Wikipedia.de (2005): Langzeitarchivierung. [WWW-Dokument: http://de.wikipedia.org/wiki/Digitales_Vergessen - Zugriff: 25.01.2005].

Digitale Verweise

- [@1] www.digitalpreservation.gov
- [@2] www.langzeitarchivierung.de
- [@3] www.archive.org
- [@4] www.bibalex.org/english/initiatives/internetarchive/web.htm
- [@5] www.research.microsoft.com/barc/MediaPresence/MyLifeBits.aspx