

Der Müll der Zukunft: Datenmüll

**Autoren: Ana-Laura Thomat, Martin G. Bernhard
ECG Management Consulting GmbH**

Geringfügig überarbeitet erschienen in: IT Management, Dezember 2001

Auf fast jedem PC sind ca. 40 bis 60 Prozent Datenmüll. In jedem Projekt werden ca. 50 bis 90 Prozent an Datenmüll generiert, d.h. Dateien, auf die nicht mehr zugegriffen werden und die „Lagerplatz“ auf dem Datenträger beanspruchen. In den Unternehmen kommen p.a. zwischen 50 und 150 Prozent an neuen zusätzlichen Daten hinzu. Zur Zeit existieren über 500 Mio. E-Mail Postfächer, im Jahre 2005 werden dies weltweit 1,2 Mrd. Postfächer sein und pro Tag werden ca. 36 Mrd. E-Mails versendet. 7.500 Terabyte an Informationen in 550 Milliarden Einzeldokumenten sind heute allein im World Wide Web zu finden; das gab am 19.12.2000 BrightPlanet.com bekannt. Diese Informationen verdeutlichen erste Treibergrößen das enorme zukünftige Datenwachstum.

In Zeiten, in denen Computer noch nicht existierten bzw. nur in geringem Ausmaß in Unternehmen verbreitet waren, wurden auch nicht ständig neue Büroräume und Regale zugekauft und angebaut, um alle Belege und Ordner aufzubewahren. Jedoch werden heute die generierten und erworbenen Daten mitunter sehr lange Zeit aufbewahrt. Hierfür entsteht eine Besitzstand durch die Anwender. Die Antworten der Industrie gehen zudem in Richtung neuer hoch performanter Speichermedien, Storage Area Networks und Archivierungslösungen, d.h. Lösungen zur Aufbewahrung und Erweiterung der Datenbestände.

Der Speicherbedarf steigt permanent; wurde Anfang der achtziger Jahre noch in Kilobyte gerechnet, folgten bald Megabyte (= 1.024 Kilobyte), anschließend Gigabyte (= 1.024 Megabyte). Inzwischen sind Terabyte (= 1.024 Gigabyte) zu einer alltäglichen Größe geworden. Das Bundesministeriums für Bildung und Forschung ermittelte, dass ein PC, für den im Jahre 1999 3.000 DM ausgegeben wurde, entsprach einem Gegenwert von 3.000.000 DM im Jahre 1985.

Selbst Bill Gates fehlte es auch manchmal an visionärer Voraussicht. 1981 gab der Microsoft-Gründer anlässlich einer Pressekonferenz zu Protokoll, dass 620 Kilobyte Speicherplatz für Computeranwender völlig ausreichen:

„Niemand benötigt mehr.“

Die Content-Commerce Wertschöpfungskette wird erweitert

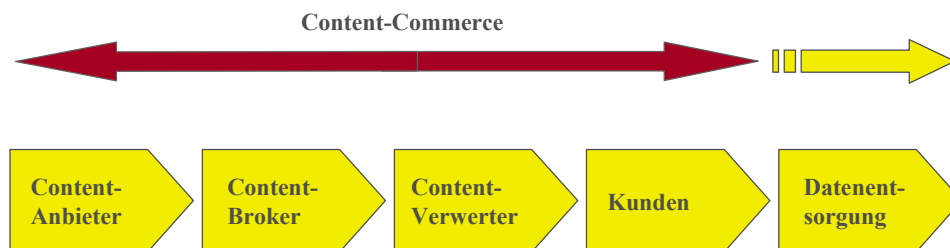
Um das Problem der mangelnden Datenentsorgung zu veranschaulichen, ist es notwendig, den Vorgang von der Datengenerierung/-produktion bis zur Datenvermarktung vorwegzunehmen: Die Wertschöpfungskette, die den Verlauf von der Content- Produktion über den sogenannten Content- Broker bis hin zum Kunden beschreibt, hat ihren Anfang beim Content- Anbieter.

Der Content-Anbieter (Universitäten, Verlage, Agenturen, etc.) produziert originäre Inhalte, beispielsweise solche von Nachrichtenagenturen, Verlagen und Universitäten. Die Halbwertszeit von Wissen und Informationen nimmt gerade heute schneller ab denn je! Diese Inhalte in Gestalt von gewichtslosen, digitalen Datenbanken und den verschiedenen Arten von Dateien können zu einen Content-Broker gelangen. Dieser vermittelt die Informationen an Verwerter (z.B. Xipolis oder Syndicate europe).

Sind die Informationen nun beim Content- Verwerter angelangt, werden sie veröffentlicht und auf Webseiten vermarktet. Beispiele hierfür sind yahoo, amazon.de, redmark.de, diplomarbeiten.de, hausarbeiten.de, competence-site.de.

Der beschriebene Vorgang der Wertschöpfungskette stellt den Prozess dar, welcher erforderlich ist, um ein Produkt, z.B. Präsentation zum Thema Wissensmanagement oder ein Skript einer Vorlesung auf einem Markt über eine Plattform im Internet zu platzieren, wo es dann jedem Kunden frei zugänglich ist.

Allerdings verfügt dieser Markt im Vergleich zu anderen traditionellen Märkten über einen bedeutsamen Unterschied: Der Bestand des angebotenen Produktes wird mit der Nachfrage nicht geringer, sondern er bleibt unverändert, er steigt sogar durch weitere Produkte stetig (bzw. Dateien), die angeboten werden. Denn bits kann man verkaufen und gleichzeitig behalten. Original und Kopie sind in diesem Fall nicht voneinander zu unterscheiden ... (N. Negroponte). Allerdings ist die Wertschöpfungskette mit dem erfolgreich abgewickelten Geschäft zwischen Content- Verwerter und Kunden noch nicht beendet



Quelle: ECG Management Consulting GmbH

Abb. 1: Die Wertschöpfungskette vom Content-Commerce bis zur Datenentsorgung

Wir nähern uns einem neuen Prozessabschnitt, welcher die gesamte Prozesskette erweitert, einer geregelten Datenentsorgung. Mit Datenmüll ist die Datenmenge gemeint, welche dem Kunden (ob Geschäfts- oder Privatkunde) zur Last fällt; ein Nebenprodukt durch den Einkauf von „Dateien“ im digitalen Markt. Neben den selbst angeforderten und eingekauften Datenmengen gibt es noch die „Junk-Mails“ resultierend aus Werbung via E-Mail.

Für den Datenmüll ist eine systematische Entsorgung erforderlich

Unternehmen oder Organisationen können einzelne oder mehrere Abschnitte der Content-Commerce Wertschöpfungskette abdecken – Datenmüll fällt jedoch überall an!

Was soll mit den Daten geschehen, welche nicht mehr benötigt werden? Anstatt immer wieder auf neue Speichersysteme zurückzugreifen, sollte über eine effektivere Langzeitlösung nachgedacht werden.

Vom zunehmenden Datenmüll sind insbesondere Unternehmen betroffen. Es ist klar: nur mit dem Zukauf von neuen Speicher- und Archivierungssystemen ist den Unternehmen langfristig nicht geholfen. Von den „Lösungsanbietern“ (Speicherhersteller, Archivierungssystemanbieter, etc.) werden heute Lösungen zur Datenaufbewahrung und nicht zur Reduzierung der Daten angeboten. Die Marktforschungsunternehmen unterstützen mit ihren Research-Ergebnissen diesen Sachverhalt: So schätzt beispielsweise Forrester Research, dass Unternehmen im Jahr 2003 fünf Mal so viel in Datenspeicher investieren wie im Jahr 1999. Einer Studie des IDC zufolge, wird der Markt für Speicherlösungen in Europa in den Jahren 1999 bis 2003 mit jährlich 65% wachsen.

Anschaulich dargestellt, bedeutet das, ein Bürozimmer mit einer wachsenden Zahl an Bücherregalen zuzustellen; solange, bis kaum mehr das Zimmer zu betreten ist. Ernste Schwierigkeiten treten auf, wird ein Buch aus einem bestimmten Regal benötigt, welches sich womöglich an der unzugänglichen Wand im hinteren Teil des Raumes befindet.

Wachsende Speicher- und Archivierungssysteme führen zu einer unüberschaubaren Datenmenge, in der dem Nutzer schon nach kurzer Zeit der Überblick fehlt und für den Betreiber die betrieblichen Risiken zunehmen. Die Anschaffung von neuen Speichersystemen ist es nicht allein, weil die Wurzel des Problems, die Entstehung des Datenmülls, in großen Teilen mit gespeichert und nicht entsorgt wird.

Daraus resultieren folgende Fragestellungen:

- Wie kann bei einer wachsenden Datenflut ein dafür erforderlicher Überblick bewahrt werden?
- Wie kann die Komplexität zunehmender Datenmengen beherrscht werden?
- Durch welche Maßnahmen kann der Datenmüll von den wirklich benötigten Daten getrennt werden (Aktivierung „stiller“ Speicherreserven und Reduzierung von Komplexitätsrisiken)?
- Warum existieren noch keine Anbieter für „Datenmüllentsorgung“?
- Welches sind die wichtigsten Treibergrößen, die Datenmüllproduzenten, für den Datenmüll?

Auf die letzte Frage „Welches sind die wichtigsten Treiber, die Datenmüllproduzenten, für den Datenmüll?“ wird im folgenden eingegangen. Als Datenmülltreiber können drei verschiedene Phänomene gesehen werden (vgl. Abb. 2 und 3):

- Verbreitung des Internets und die Durchdringung von Informationstechnologie in Unternehmen und privaten Haushalten
- Internetpenetration und –Bandbreite,
- Verteilungs- und Verwaltungsmechanismen von PCs und Rechnersystemen in Unternehmen
- E-Mail Systeme
- Junk-Mails wie Werbung,
- E-Mail als Multiplikator für Dokumente (insb. in Projekten)
- Wissensverteilung und- speicherung, z.B.
- Wissensmanagement in Unternehmen,
- die Bereitstellung von Vorlesungen, Inhalten im Internet,
- Zugriff auf Datenbanken, etc.

Abb. 2: Die drei Datenmülltreiber

Internet/IT	E-Mail	Wissens- verteilung
<ul style="list-style-type: none"> - mehr als 500 Millionen PCs weltweit - 7.500 Terabytes an zugänglichen Daten im Web (ohne „Deep Web“) 	<ul style="list-style-type: none"> - heute: mehr als 505 Millionen E-mail Postfächer weltweit - im Jahr 2005 werden bereits - 1,2 Milliarden Postfächer erwartet - 36 Milliarden E-Mails pro Tag versendet 	<ul style="list-style-type: none"> - 1,5 Exabyte an neuen Daten allein 1999 - Bezogen auf die gesamte Weltbevölkerung: Auf jeden Menschen kommen etwa 250 Bücher bzw. 250 MB Daten p.a. - 12 Exabyte beträgt das Gesamtvolumen aller weltweit verfügbaren Informationen

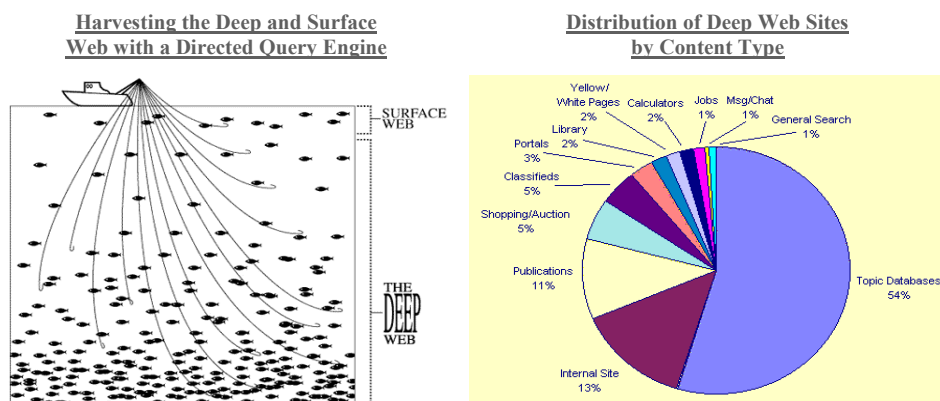
Abb. 3: Die drei Datenmülltreiber – wichtigste Zahlen

Internet und Unternehmen als Datenwachstumsgeneratoren
 Bereits Ende 2000 gab es weltweit mehr als 500 Millionen PCs (eTForecasts) und zudem existieren ca. 109,6 Millionen Hosts im Internet (vgl. www.isc.org/ds/hosts.html, April 2001), welche Daten sammeln und bereitstellen für die weltweite Internet-Community. Diese Fakten

vermitteln ein erstes Gefühl darüber, welche Plattformen im Netz verfügbar sind, um Daten zu verteilen.

Um die Informationsflut im Internet zu verdeutlichen, sind darüber hinaus noch weitere Fakten zu nennen: 7.500 Terabyte an Daten ist allein der Teil der **zugänglichen** Datenbanken im World Wide Web. Diese Menge entspricht 550 Milliarden Einzeldokumenten (BrightPlanet.com, sowie completeplanet.com Dez. 2000). Diese enorme Datenmenge erscheint bereits unvorstellbar.

Aber darüber hinaus existieren im Deep Web noch weitaus mehr Daten (“For the moment, however, the key points are that content in the deep Web is massive — approximately **500 times greater** than that visible to conventional search engines”)! Neben den im Web zugänglichen Daten existiert noch ein verborgener Datenschwamm im nicht oder nur schwer zugänglichen Deep Web (completeplanet.com). Immerhin sind 54% der im Deep Web enthaltenen Informationen Datenbanken; die restlichen Informationen bestehen aus internen Webseiten, Shopping-Angeboten, Publikationen, Jobangeboten und Telefonbüchern (vgl. dazu Abb. 4, Datenwachstumstreiber Internet: „More than half of all deep Web sites feature topical databases “



Quelle: www.completeplanet.com/

Abb. 4: Datenwachstumstreiber Internet: „More than half of all deep Web sites feature topical databases “

Das enorme Datenwachstum bewirkt, dass im Internet als auch in Unternehmen aufgrund des Zwanges zur Datensicherung der Speicherbedarf stetig steigt. Nach Einschätzung von Merrill Lynch werden 75% der gesamten IT- Investitionen in Datenspeicher investiert werden.

Nach Storea Tek (bereits 1998!) entstehen täglich weltweit 52 Milliarden Megabyte. Daraus lassen sich folgende Annahmen ableiten:

- Von den täglich neu entstehenden weltweiten Daten werden etwa 2,5 Milliarden Megabyte in Deutschland generiert;
- wovon etwa 1,5 Milliarden Megabyte in der Industrie entstehen.
- Ca. 10% von diesen Daten fallen (bereits) täglich als Datenmüll an.

Auf dieser Basis entstanden bereits 1998 täglich 0,15 Milliarden Megabyte an Datenmüll, wir sprechen von einem Marktvolumen von jährlich (bezogen auf 1998) neu entstehendem Datenabfall in Höhe von 30 Milliarden Megabyte. Demzufolge dürften allein für das Jahr 2001 ca. 80 bis 120 Mrd. Megabyte an Datenmüll in der Industrie auf Rechnersystemen gelagert (ECG Research, April 2001) sein.

Explosionsartige Entwicklung des Datenwachstums durch die Versendung von E-Mails

Die erste E-Mail wurde vor 30 Jahren, 1971, im damaligen ARPANET verschickt (www.pretext.com/mar98/features/story2.htm). Diese Art von Informationsaustausch verbreitet sich schnell, und als die Technik es möglich machte, von jedem beliebigen Computer an allen Orten dieser Welt E-Mails zu versenden, war der E-Mail Boom nicht mehr zu stoppen.

Zur Zeit existieren etwa 505 Millionen Postfächer weltweit, und im Jahr 2005 werden 1,2 Milliarden E-Mail Postfächer erwartet, das entspricht einem durchschnittlichen Wachstum von 138 Prozent pro Jahr. Über E-Mail wird im Jahr rund 500mal so viel an Informationsvolumen erzeugt wie durch die Produktion an Webseiten (www.akademie.de/news/langtext.html?id=7770). E-Mail-Systeme fungieren zudem als Multiplikator für die Verteilung von Dokumenten innerhalb eines Projektes. Dies führt allein dazu, dass ca. 50 bis 90 Prozent aller Projektdokumente schon nach kurzer Zeit zu Datenmüll werden (ECG Research, April 2001). In vier Jahren sollen die zwischen Nutzern ausgetauschten Informationen via elektronischer Post pro Tag auf 36 Milliarden anschwellen (IDC, Sept. 2001 und www.newsletter@mail.cwverlag.de).

Weitere Kosten werden insbesondere durch Junkmails sowie unerwünschte Werbung erzeugt. 10 Milliarden Euro werden jährlich dafür von den Nutzern weltweit durch Online-Kosten, die durch das Öffnen und Löschen der Werbemails anfallen, gezahlt (EU Kommission, Febr. 2001). Es ist zudem davon auszugehen, dass 40% der in den Firmen eingehenden E-Mails in den nächsten drei bis vier Jahren Werbematerialien sein werden (Financial Times, 28.8.01, S. 8, Studie von Ferris Research). Genauso verheerend ist der Zeitverlust, der durch die Bearbeitung der Informationen (E-Mails, Werbung, etc.) verursacht wird. In den USA erhalten Mitarbeiter täglich mehr als 40 E-Mails. Zu deren Verarbeitung werden 50 Minuten der Arbeitszeit verwendet. (Gartner Group, August 2001).

Die negativen Folgen sind evident: Speicherplatzverlust durch Junkmails, Zeitverlust durch Informationsbewältigung mit der Folge, dass die Arbeitszeit und Produktivität der Mitarbeiter verringert wird.

Wissensmanagement und Wissensverbreitung

Pro Kopf der Weltbevölkerung werden etwa 250 Megabyte (das entspricht ca. 250 Büchern) an Informationen produziert (Studie der University of California – von Hal Varian und Peter Lyman, Berkeley, 1999 – vgl. auch www.sims.berkeley.edu/how-much-info/ sowie www.akademie.de/news/langtext.html?id=7770).

Gegenwärtig beträgt das Gesamtvolumen aller weltweit verfügbaren Informationen 12 Exabyte oder 12 Millionen Terabyte. Ein Exabyte ist soviel wie eine Million Terabyte, ein Terabyte wiederum ist gleich eine Million Megabyte und das entspricht ungefähr dem Inhalt einer Million Bücher (Quelle von Storage Tek, 1998). Allein 1999 wurden einer Studie zufolge 1,5 Exabyte neuer Daten erzeugt – das sind 1 500 000 000 durchschnittlich dicke Bücher (vgl. Studie unter: www.sims.berkeley.edu/how-much-info/)

Im Zeitalter des 21. Jahrhunderts wird nur noch ein geringer Anteil an reinen Informationen in Printmedien wiedergegeben; der Wissensaustausch hat sich auf die unerschöpfliche Aufnahmefähigkeit des Internets konzentriert. Hier kann jeder Nutzer frei von den Informationsmengen Gebrauch machen und auch selbst Wissen weitergeben. Die beeindruckende Menge an Daten der Wissensverteilung und –speicherung (93% an Datenvolumen, die jährlich in digitaler Form auf Platten –oder Band-basierten Speichermedien abgelegt werden), erklärt die daraus resultierende Speicherplatznachfrage.

HandlungsfelderDie Treiber für das Datenwachstum verdeutlichen, dass Innovationen und Investitionen in neue Speicherkapazitäten und performantere PCs nicht die Probleme des wachsenden Datenmülls lösen werden. Eine degressive Steigerung oder gar ein Abbruch des Datenwachstums ist nicht in Sicht. Im Gegenteil, das Datenvolumen in den Unternehmen soll sich alle zwei bis drei Jahren verdoppeln, und entsprechend wird sich der Datenmüll vermehren. Sicher kann durch Speicherlösungen beliebig viel Datenvolumen „angebaut“ werden, allerdings ist dies von der Komplexität her nicht mehr beherrschbar. Spätestens hier kommt die Frage auf, welches zukünftige Handlungsfelder und Geschäftsmodelle sein werden insbesondere für die Datenentsorgung in größeren Unternehmen.

Hierzu zählen Methoden zur Identifikation und „Entsorgung“ des Datenmülls. D.h.

- Wie kann Datenmüll im Unternehmen identifiziert werden (auf gemeinsamen Plattformen sowie auf den Arbeitsplattformen einzelner Mitarbeiter)?
- Welche Mechanismen sind aufzubauen, um durch E-Mail-Systeme nicht zu stark Datenmüll zu generieren?
- Welche Sicherheitsmechanismen sind zu berücksichtigen?
 - nur den „echten“ Datenmüll entsorgen
 - Zugriff und Bereinigung nur durch „vertrauensvolle“ Know-how Träger
- Welche organisatorischen Spielregeln und Tools können eingesetzt werden, um das Wachstum des Datenmülls zu dämmen, z.B.
 - Filterverfahren um Werbe-E-Mails zu reduzieren,
 - Wachstum von E-Mail-Postfächern besser zu kontrollieren,
 - Nutzungsintensität von Dateien,

- Halbwertzeit von Wissen (Dateien)

- etc.

- Wie kann sichergestellt werden, dass potentiell neu anwachsender Datenmüll rechtzeitig transparent wird (z.B. Altersstrukturen der Informationen und Dateien)?
- Wie soll Datenmüll technisch entsorgt werden?
- Welche Software-/Systemrealisierungsmöglichkeiten zur Reduzierung des Datenmülls sollten aufgebaut werden (z.B. Software-Agenten, Inventarsysteme über Halbwertzeiten von Inhalten)?
- Wie kann ein Konzeptions- und Implementierungsprojekt für größere Unternehmen zur „Erstdatenentsorgung“ ablaufen?
- Welche Service-Level-Agreements können mit IT-Kunden /-Anwendern zur Reduzierung des Datenmülls abgeschlossen werden (z.B.
 - ein SLA über die max. Größe der Speicherkapazität (lokale oder logische Festplatte)
 - nur Arbeitsplätze liefern ohne lokale Festplatten
- Welche Recycling- oder Wiederverwendungsmöglichkeiten von Datenmüll gibt es (für welche Inhalte)?
- Wie kann das Handling zur kontinuierlichen Datenentsorgung organisatorisch integriert werden innerhalb eines Unternehmens (z.B. Einführung einer neuen Funktion, dem Data Waste Management Officer)?
- Welche organisatorischen Maßnahmen stellen sicher, dass Datenwachstum dauerhaft in den Griff zu bekommen?
- Welche Verfahren stellen sicher, dass Dateien zur Wissensverteilung (und –speicherung) nicht zu übermässigen Datenfriedhöfen führen (ca. 50 bis 90 Prozent aller Dateien innerhalb eines Projektes sind Datenmüll)?
- Für welche Unternehmen bzw. Anbieter im Markt könnte ein Geschäftsfeld wie „Datenentsorgung“ sinnvoll sein (Entsorgungsmentalität und keine trojanischen Pferde)?

„Data Waste Management Officer“, dem Owner über alle Datenentsorgungsprozesse im Unternehmen. Der Data Waste Management Officer muß das Vertrauen der Anwender genießen. Schließlich legen und erwarten Anwender über ihre Dateien und E-Mails einen psychologischen Besitzstandsstatus.

Datenmüllentsorgung dürfte in größeren Unternehmen Nutzenpotentiale bieten:

- Freilegung stiller Reserven (Speichernutzung),
- bessere Nutzung existierender IT-Infrastrukturen,
- Verbesserung der Datensicherungsprozesse
- Komplexitätsreduzierung und bessere Beherrschbarkeit der Systeme

Zusammenfassung und Ausblick

Die Treiber des Datenwachstums und auch Datenmülls sind zunehmende Internet- und PC-Penetration, Ausbreitung von E-Mail-Systemen sowie Wissensmanagement. Es ist klar, dass Methoden und Technologien (Speichermedien, Archivierung, etc.) zur Aufbewahrung von Datenbeständen als Lösung allein nicht genügen. Hohe Datenvolumina, verursachen im Bereich der Verteilungs- und Verwaltungsmechanismen in Unternehmen unnötige Kosten und die ungebremste Anhäufung von Daten bewirkt den fehlenden Überblick, fördert ein unstrukturierendes Informationsmanagement und zudem wird die Komplexität erhöht.

Neue Prozesse und Strukturen zur Datenentsorgung im Unternehmen sowie Software-Tools zur Identifikation und Bereinigung von Datenmüll müssen bereitgestellt werden. Datenmüll, der Müll der Zukunft, der bereits heute existiert, stellt eine neue Herausforderung für die IT im Unternehmen dar.

Literatur:

- BrightPlanet.com, Dezember 2000
- Bundesministerium für Bildung und Forschung, Endbericht 2000: Zur technologischen Leistungsfähigkeit Deutschlands
- Computerwoche, Juni 1998
- ETForecasts, 17.12.2000
- EU-Kommission, Febr. 2001
- Financial Times, 28.8.01, S. 8, Studie von Ferris Research
- Gartner Group, August 2001
- Studie der University of California (Hal Varian und Peter Lyman), Berkeley, 1999 – vgl. auch www.sims.berkeley.edu/how-much-info/ - als PDF-Datei erhältlich
- Storage Networking Industry Association, April 2001

Internet-Adressen:

- newsletter@mail.cwverlag.de, 18.09.2001
- www.akademie.de
- www.completeplanet.com, 19.12.2000
- www.isc.org/ds/hosts.html, 26.04.2001
- www.pretext.com/mar98/features/story2.htm (“The 1First E-Mail Message”)
- **www.pyramid-commerce.com (Powerpoint- und PDF-Präsentationen über aktuelle Management-Themen)**