

Datenmüll der Müll der Zukunft – der bereits heute existiert

**Autoren: Martin G. Bernhard, Inga Trenn
ECG Management Consulting GmbH**

Geringfügig überarbeitet erschienen in: Hessen Software News, Februar 2002

Auf fast jedem PC befinden sich ca. 40 bis 70 % Datenmüll. In jedem Projekt werden ca. 50 bis 90 % Datenmüll generiert, d.h. Dateien, auf die nicht mehr zugegriffen wird und die „Lagerplatz“ auf dem Datenträger beanspruchen. In den Unternehmen kommen p.a. zwischen 50 und 150 % an neuen zusätzlichen Daten hinzu. Zur Zeit existieren über 500 Mio. E-Mail Postfächer, und pro Tag werden ca. 10 bis 12 Mrd. E-Mails versendet. Bereits heute befinden sich über 550 Milliarden Einzeldokumente im World Wide Web. Diese Größenordnungen verdeutlichen das enorme Datenwachstum und das daraus resultierende Datenmüllwachstum.

Gespeicherte Daten werden heute mitunter eine sehr lange Zeit aufbewahrt. Der Speicherbedarf steigt permanent. Die Antworten der Industrie gehen in Richtung neuer hoch performanter Speichermedien und Archivierungslösungen. 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.“ Das Internet hat eine besondere Eigenschaft: Bits und Bytes kann man verkaufen und gleichzeitig behalten... (N. Negroponte), d.h. Datenmüll kann sich im Vergleich zum realen Müll durch kopieren beliebig vermehren (vgl. Abb. 1: Vergleich Datenmüll vs. realer Müll). Theoretisch kann durch eine Datei mit Datenmüll der gesamte weltweit verfügbare Speicherplatz zugemüllt werden!

	Datenmüll	Realer Müll
Erscheinungsform	Bits und bytes (Jede Art von Dateien, z.B. Dokumente, etc.)	Physikalisch und greifbar
Gewicht	„Bits und bytes wiegen nichts“	abhängig von den Müllfraktionen
Entsorgung	Individuelles Löschen	geregelt und über die Müllabfuhr
Entsorgungsbehältnisse	der PC!	Mülltonnen, Behälter
Lagerung des Mülls	...auf quasi jedem PC	Müllhalden sowie div. Recycl.-Wege
Aufwand für die Entsorgung	Bisher eher hoch, da Datenmüll vorher individuell definiert werden muß (Selektion allein genügt nicht)	Hoch, durch Mülltrennung und aufwendige Abfuhr
Art der Mülltrennung	Individuell festzulegen: Nach Alter, Inhalt, Dateityp, etc.	Nach Fraktionen: Pappe, Papier, Karton, Glas, etc.
Kosten	Kosten für organisatorische Maßnahmen, Entsorgungssoftware, neue Speichermedien, Archivierung	mindestens Kosten für die Entsorgung
Kopierbarkeit des Mülls	„Bits und bytes können verkauft werden und gleichzeitig behalten werden“ über beliebige Verfahren möglich	...nach der klassischen Physik nicht möglich
Müllwachstum	Sehr stark (durch Datenwachstum z.B. in den Unternehmen: 50 – 150 % p.a.)	< 10 % p.a. (pro Fraktion)

Abb. 1: Vergleich Datenmüll vs. realer Müll

Für den Datenmüll ist eine systematische Entsorgung erforderlich

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. 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?
- Welches sind die wichtigsten Datenmüllproduzenten?

Als Datenmülltreiber können drei verschiedene Phänomene gesehen werden (vgl. Abb. 2).

- **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 und Unternehmen als Datenwachstumsgeneratoren

Bereits Ende 2001 gab es weltweit mehr als 500 Millionen PCs (eTForecasts) und zudem existieren ca. 110 Millionen Hosts im Internet (vgl. www.isc.org/ds/hosts.html, April 2001). Außerdem befinden sich im Internet zur Zeit mehr als 550 Milliarden Einzeldokumente (BrightPlanet.com sowie completeplanet.com Dez. 2001). Darüber hinaus existieren im Deep Web, im verborgenen Datendschungel, noch weitaus mehr Webseiten und Datenbestände! Davon sind allein 54% Datenbanken; die restlichen Informationen bestehen aus internen Webseiten, Shopping-Angeboten, Publikationen, Jobangeboten und Telefonbüchern. Die Versendung von E-Mails entwickelt sich explosionsartig. Heutzutage werden ca. 10 bis 12 Mrd. E-Mails pro Tag versendet, in 2005 ca. 36 Mrd. E-Mails pro Tag (IDC, Sept. 2001 und www.newsletter@mail.cwverlag.de). Zur Zeit existieren mehr als 505 Millionen E-Mail Postfächer weltweit, und im Jahr 2005 werden ca. 1,2 Milliarden Postfächer erwartet, das entspricht einem durchschnittlichen Wachstum von 138 % pro Jahr bis 2005. E-Mail-Systeme fungieren zudem als Multiplikator für die Verteilung von Dokumenten, z.B. 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). Diese Fakten vermitteln ein Gefühl darüber, welche Mengenvolumina im Netz verfügbar und beliebig multiplizierbar sind, um Daten auszutauschen und Datenmüll zu vermehren.

Das Öffnen und Löschen von Junkmails allein verursacht £47,000 pro 100 Mitarbeiter im Jahr!

Kosten entstehen insbesondere durch Junkmails (unerwünschte Mails wie z.B. Werbung). 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. Eine Hochrechnung zum Thema "Spam" ergibt sehr hohe Kosten allein für das Öffnen und Löschen von unerwünschten Mails: Der Versand unerwünschter Werbebriefe bei britischen Unternehmen verursacht jährlich einen Gesamtschaden von 3,2 Milliarden Pfund "Cost to business estimated at £47,000 for every 100 employees". Die Berechnungsgrundlage dafür ist, dass ein Firmenmitarbeiter, der jeden Tag 10 Minuten für das Aussortieren von Spam benötigt, bei seinem Arbeitgeber Kosten in Höhe von schätzungsweise 2 Pfund pro Tag verursacht. Dieses Ergebnis wurde auf die Zahl der Mitarbeiter in Großbritannien mit Internet-Anschluss hochgerechnet (www.intern.de/news/2350.html und <http://www.washtech.com/about/>). 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 wurden bereits 1999 ca. 250 Megabyte (das entspricht ca. 250 Büchern) an Informationen produziert (SIMS-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). Bereits 1998 betrug das Gesamtvolumen aller weltweit verfügbaren Informationen 12 Exabyte oder 12 Millionen Terabyte (Quelle: Storage Tek, 1998).

Handlungsfelder

Diese Datenwachstumstreiber verdeutlichen, dass Innovationen und Investitionen in neue Speicherkapazitäten und performantere PCs nicht die Probleme des wachsenden Datenmülls lösen werden. Im Gegenteil, das Datenvolumen in den Unternehmen soll sich alle zwei bis drei Jahre 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 (z.B. Datensicherungs- und Recoveryzeiten). Spätestens hier kommt die Frage auf, welches zukünftige Handlungsfelder und Geschäftsmodelle sein werden, insbesondere für die Datenentsorgung in Unternehmen.

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

- Wie kann Datenmüll vorbeugend vermieden werden?
- Wie kann Datenmüll im Unternehmen identifiziert werden?
- Welche Sicherheitsmechanismen sind zu berücksichtigen?
- Welche organisatorischen Spielregeln und Tools können eingesetzt werden, um das Wachstum des Datenmülls zu dämmen?
- 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, Expertensysteme, Inventarsysteme über Halbwertzeiten von Inhalten)?
- Wie kann ein Konzeptions- und Implementierungsprojekt im 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, das Datenwachstum dauerhaft in den Griff zu bekommen?
- Welche Verfahren stellen sicher, dass Dateien zur Wissensverteilung (und –speicherung) nicht zu übermäßigen Datenfriedhöfen führen?

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 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 unstrukturiertes Informationsmanagement, und zudem wird die Komplexität erhöht. Neue Prozesse und Strukturen zur vorbeugenden Datenmüllvermeidung und Datenmüllentsorgung sowie Software-Tools zur Identifikation und Bereinigung von Datenmüll müssen entwickelt werden. Datenmüll, der Müll der Zukunft, der bereits heute existiert, stellt eine neue Herausforderung 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”)