



Holger Schneider

Digitale Amnesie

Langzeitarchivierung digitaler Dokumente im betrieblichen Umfeld



Holger Schneider

Digitale Amnesie

Langzeitarchivierung digitaler Dokumente im
betrieblichen Umfeld

Holger Schneider

Digitale Amnesie

Langzeitarchivierung digitaler Dokumente
im betrieblichen Umfeld

Books on Demand

Inhaltsverzeichnis

1. RELEVANZ DER THEMATIK

- 1.1. DARSTELLUNG DER SITUATION IN DEN UNTERNEHMEN
- 1.2. MOTIVATION
- 1.3. VORGEHENSWEISE UND GLIEDERUNG

2. METHODISCHER ÜBERBLICK

- 2.1. LANGZEITARCHIVIERUNG IM DIGITALEN KONTEXT
 - 2.1.1. *Langzeit*
 - 2.1.2. *Archivierung*
 - 2.1.3. *Digitale Dokumente*
- 2.2. FLUCH UND SEGEN DER DIGITALEN DOKUMENTE
- 2.3. GRÜNDE FÜR DIE DIGITALE LANGZEITARCHIVIERUNG
 - 2.3.1. *Gesetzliche Forderungen (Compliance)*
 - 2.3.2. *Qualitätsmanagement*
 - 2.3.3. *Produkthaftung*
 - 2.3.4. *Wissenserhalt und Wissensmanagement*
- 2.4. ANFORDERUNGEN AN DIE LANGZEITARCHIVIERUNG
 - 2.4.1. *Substanzerhaltung*
 - 2.4.2. *Erhaltung der Benutzbarkeit*
 - 2.4.3. *Erschließung über technische Metadaten*
 - 2.4.4. *Vertrauenswürdige digitale Archive*
- 2.5. ENTMYSTIFIZIERUNG DER DIGITALEN AMNESIE

3. REVISIONSSICHERE DIGITALE LANGZEITARCHIVE

- 3.1. DEFINITION DER REVISIONSSICHERHEIT
- 3.2. FORDERUNGEN DER REVISIONSSICHERHEIT
- 3.3. REVISIONSSICHERE DATENTRÄGER
- 3.4. DIE DIGITALE SIGNATUR
- 3.5. ZERTIFIZIERUNG DER REVISIONSSICHERHEIT
- 3.6. NACHSIGNIEREN VS. REVISIONSSICHERE ARCHIVIERUNG

4. SPEICHERTECHNOLOGIEN UND MEDIEN

- 4.1. SPEICHERSYSTEM FESTPLATTE
 - 4.1.1. *Ausfallrisiken und Lebensdauer*
 - 4.1.2. *Festplattenspeicher auf RAID-Basis*
 - 4.1.3. *Virtualisierung durch Content Adressed Storage*
 - 4.1.4. *Bewertung als langzeitstabiles Archiv-Medium*
- 4.2. SPEICHERSYSTEM MAGNETBAND
 - 4.2.1. *Ausfallrisiken und Lebensdauer*
 - 4.2.2. *Bewertung als langzeitstabiles Archiv-Medium*
 - 4.2.3. *Bewertung der magnetischen Datenträger*
- 4.3. DIGITALE OPTISCHE SPEICHERSYSTEME
 - 4.3.1. *Compact Disc Read Only Memory (CD-ROM)*
 - 4.3.2. *Compact Disc Recordable (CD-R)*
 - 4.3.3. *DVD - Read Only Memory (DVD-ROM)*
 - 4.3.4. *DVD - Recordable (DVD±R)*
 - 4.3.5. *DVD - Random Access Memory (DVD-RAM)*
 - 4.3.6. *Blu-ray Disc Read Only Memory (BD-ROM)*
 - 4.3.7. *Blu-ray Disc Recordable (BD-R)*
 - 4.3.8. *Ausfallrisiken und Lebensdauer optischer Medien*

4.3.9. Bewertung optischer Medien

4.4. DIGITALE MAGNETOOPTISCHE SPEICHERSYSTEME

4.4.1. Magneto-Optische Discs (CD-MO)

4.4.2. Ultra Density Optical (UDO)

4.4.3. Bewertung der magneto-optische Speichersysteme

4.5. FAZIT SPEICHERSYSTEME

5. DAS OPEN ARCHIVAL INFORMATION SYSTEM

5.1. DAS INFORMATIONSMODELL

5.2. DEPOSIT SYSTEM FOR ELECTRONIC PUBLICATIONS

5.3. FAZIT OAIS-INHALTSMODELL / DSEP-PROZESSMODELL

6. ERHALTUNGSSTRATEGIEN

6.1. MIGRATION

6.1.1. Anforderungen und Ziele der Migration

6.1.2. Fazit - Chancen und Risiken der Migration

6.2. EMULATION

6.2.1. Emulation in der Langzeitarchivierung

6.2.2. Gegenstand der Emulation

6.2.3. Virtuelle Maschinen und der Matryoschka-Effekt

6.2.4. Fazit - Chancen und Risiken der Emulation

6.3. BEWERTUNG VON MIGRATION UND EMULATION

6.3.1. Aufgaben im Migrationsprozess

6.3.2. Aufgaben im Emulationsprozess

6.3.3. Diskussion Migration vs. Emulation

7. LANGZEITSTABILE DATENFORMATE

7.1. NORMATIVE ZUSTÄNDE VON DATENFORMATEN

7.1.1. *Proprietäre Formate*

7.1.2. *de facto Standard.*

7.1.3. *Handelsstandard.*

7.1.4. *Internationale Norm*

7.1.5. *Fazit - normative Zustände von Formaten*

7.2. KLASSIFIZIERUNG NACH ANWENDUNGSBEREICH

7.3. BESTANDTEILE EINES FORMATS

7.4. ANFORDERUNGEN AN DIE FORMATE

7.5. FAZIT - LANGZEITSTABILE DATENFORMATE

8. METADATEN FÜR DIGITALE DOKUMENTE

8.1. KLASSIFIKATION VON MARKUP

8.1.1. *Klassifikation nach Verwendungszweck*

8.1.2. *Eingebettete oder externe Metainformationen*

8.1.3. *Klassifikation von Metadaten*

8.1.4. *Granularität der Markup-Ebenen*

8.2. ANFORDERUNGEN AN DAS MARKUP

8.2.1. *Einfachheit*

8.2.2. *Vollständigkeit versus Modularität*

8.2.3. *Langlebigkeit*

8.2.4. *Persistenz*

8.3. BEISPIELE FÜR LANGZEITSTABILE FORMATE

8.3.1. *eXtensible Markup Language (XML)*

8.3.2. *Portable Document Format Archive (PDF/A)*

8.3.3. *Tagged Image File Format (TIFF)*

8.4. FAZIT - LANGZEITSTABILES DOKUMENTEN-MARKUP

9. EIN KOMBINIERTER LÖSUNGSANSATZ

9.1. NUR OFFENE UND STRUKTURIERTE FORMATE

9.2. EIN HYBRIDER ANSATZ BEI DEN DATENSPEICHERUNG

9.3. STRATEGIEMIX MIGRATION UND EMULATION

9.4. ORGANISATION UND PRESERVATION POLICY

9.4.1. *Auswahlkriterien für Dokumente*

9.4.2. *Compliance*

9.4.3. *Unternehmensweite Standards setzen*

9.5. LÖSUNG: EINE KULTUR DER LANGZEITARCHIVIERUNG

10. FAZIT

LITERATURVERZEICHNIS

Abbildungsverzeichnis

Abbildung 1 - Ausschnitt der ersten Seite der Gutenberg-Bibel

Abbildung 2 - Beispiele für weltweite Compliance

Abbildung 3 - Die drei Bestandteile der Vertraulichkeit

Abbildung 4 - Signaturarten gem. BSI²⁵

Abbildung 5 - Entwicklung der Festplattenkapazitäten

Abbildung 6 - Redundanz im RAID-5 bei drei Festplatten

Abbildung 7 - Raid-Systeme von Hewlett-Packard

Abbildung 8 - erkennbare Hardware-Sektorierung einer DVD-RAM

Abbildung 9 - "wobbled-land-and-groove"-Technik DVD-RAM

Abbildung 10 - 5¼ Zoll MO-Medium mit geöffnetem Cartridge

Abbildung 11 - UDO WORM (www.mitsubishi.com)

Abbildung 12 - Übersicht über die Haltbarkeit von Speichermedien

Abbildung 13 - Haltbarkeit dig. Medien gem. Hersteller Imation

Abbildung 14 - Das Information Object (IO)

Abbildung 15 - Representation Information Object (RIO)

Abbildung 16 - Archival Information Package (AIP)

Abbildung 17 - Übersicht über das DSEP Prozessmodell

Abbildung 18 - Funktionsbereiche des DSEP-Prozessmodells

Abbildung 19 - vereinfachte Darstellung der Tätigkeiten im Archiv

Abbildung 20 - Übersicht über alle Aspekte der Migration

Abbildung 21 - Beispiel für eine zweistufige Emulation

Abbildung 22 - Lebenszyklus diverser Formate

Abbildung 23 - Anforderungen an ein langzeitstabiles Format

Abbildung 24 - Forderungen an ein langzeitstabiles Markup-Format

Abbildung 25 - Zusammenhang URI - URN - URL

Abbildung 26 - die wichtigsten Bausteine der Langzeitarchivierung

Abbildung 27 - Forderungen an ein langfristiges Format

Abbildung 28 - erster Baustein: Die Formatwahl

Abbildung 29 - Nutzungshäufigkeit in Abhängigkeit von der Zeit

Abbildung 30 - Beispiel für eine hybride Speicherlösung

Abbildung 31 - zweiter Baustein: eine hybride Speicherlösung

Abbildung 32 - dritter Baustein: Mix aus Emulation und Migration

Abbildung 33 - zentraler Baustein: Organisation mittels Policy

Abbildung 34 - eine gelebte Archiv-Kultur

Tabellenverzeichnis

Tabelle 1 - Zusammenfassung der Sicherheitsaspekte

Tabelle 2 - Übersicht Langzeiteignung magnetischer Speichersysteme

Tabelle 3 - Die vier Kapazitäten der DVD gem. ISO/IEC 16448

Tabelle 4 - Übersicht Langzeiteignung optischer Speichermedien

Tabelle 5 - Übersicht Langzeiteignung MO-Medien

Tabelle 6 - Informationspakete und ihre Bedeutung

Tabelle 7 - Migrationsaufgaben und DSEP-Prozesse

Tabelle 8 - Ereignisse und Maßnahmen beim Migrations-Monitoring

Tabelle 9 - Ereignisse und Maßnahmen beim Emulations-Monitoring

Tabelle 10 - Auswahlkriterien für digitale Dokumente

Tabelle 11 - Elemente einer Archivierungspolicy

1. Relevanz der Thematik

Die Menge ausschließlich digital vorliegender Informationen wächst stetig an. Die Universität von Berkeley hat zuletzt im Jahr 2003 geschätzt, dass jeder Mensch 800 MB Daten pro Jahr erzeugt¹. Ein Großteil dieser Daten wird auf Festplatten gespeichert, wobei der Anteil optischer Speichermedien stark ansteigt. Darüber hinaus werden ursprünglich analog vorliegende Daten digitalisiert, um den Zugriff auf diese auch über Datennetze zu ermöglichen. Digitale Informationen sind daher ein fester Bestandteil des kulturellen und sozialen Lebens geworden.

Auch im Tagesgeschäft von Unternehmen werden große Mengen digitaler Informationen produziert und konsumiert, für die kein analoges Äquivalent mehr zur Verfügung steht. Dabei spielt insbesondere die dauerhafte Verfügbarkeit von aktuellen und historischen Daten eine wichtige Rolle.

1.1. Darstellung der Situation in den Unternehmen

Vor dem digitalen Zeitalter gestaltete sich Langzeitarchivierung für viele Unternehmen recht einfach. Langfristig aufzubewahrende Papierdokumente wurden schlicht als Original im Archiv aufbewahrt. War die Aufbewahrungsfrist abgelaufen, konnten die Dokumente vernichtet werden. Aus Platzgründen wurden Dokumente von denen nicht das Original verwahrt werden musste auch auf Mikrofilm kopiert.

Durch die Digitalisierung von Dokumenten boten sich den Unternehmen wirtschaftliche Möglichkeiten, auch die

größten Datenmengen zu speichern. Diese Dokumente sind ihren analogen Gegenständen in punkto Recherchierbarkeit und platzsparender Lagerung überlegen. Sie haben jedoch auch einen gravierenden Nachteil. Während ein schriftliches Dokument jederzeit lesbar bleibt, kann ein digitales Dokument nur mithilfe der Hardware und den Programmen ausgewertet werden, mit denen es erstellt wurde. Ist der Datenträger beschädigt, die benötigte Software nicht mehr vorhanden oder wird das veränderte Dateiformat nicht mehr unterstützt, gehen digitale Informationen für immer verloren.

Die unversehrte und unverfälschte Bewahrung der digitalen Informationen wird zusätzlich dadurch erschwert, dass die Daten aus heterogenen Quellen stammen und auf unterschiedlichen Datenträgern vorliegen. Die Substanzerhaltung ist jedoch nicht möglich, solange die Datensubstanz untrennbar an Datenträger und -formate und damit an deren Schicksal geknüpft ist.

Die Menge der digital gespeicherten Unternehmensdaten verdoppelt sich alle ein bis zwei Jahre². Vielfach wird darin nur eine Aufgabe für das Storage-Management gesehen, da es sich vordergründig nur um Kapazitätserweiterungen und Backups handelt. Dabei wird häufig vergessen, dass die große Herausforderung nicht darin besteht dem auch zukünftig zu erwartenden Wachstum mit Kapazitätserweiterungen zu begegnen, sondern vielmehr mit einer ganzheitlichen Strategie, die eine wirtschaftliche und langfristige Gesamtlösung anbietet. Diese Lösung ist unternehmensbezogen zu erarbeiten, da der Lebenszyklus (Lifecycle) der unternehmensweiten Dokumente je nach Unternehmen und Branche variiert und die aktiven, passiven oder inaktiven Datenbestände eine unterschiedliche Strategie erfordern (vgl. Kap. 6).

1.2. Motivation

Die Unternehmen stehen dem Datenverlust oft hilflos gegenüber und empfinden ihn als Amnesie³ des unternehmerischen Wissens und Handelns. Diese gefühlte digitale Amnesie macht es im Kontext eines Unternehmens zwingend erforderlich, eigenverantwortliche Strategien zu entwickeln, die diesem beständigen, vom Informationsmarkt und schnell voranschreitenden technologischen Fortschritt verursachten Wandel, bewältigen können.

Es scheint also nur eine Frage der Zeit zu sein, bis jede digital gespeicherte Information verloren geht. Die Pessimisten unter den Archivaren und Informationsspezialisten mutmaßen daher, dass die heutige Zeit aus der Zukunft betrachtet als „digitales Mittelalter“ angesehen werden wird. So schreibt Terry Kuny⁴:

„As we move into the electronic era of digital objects, it is important to know that there are new barbarians at the gate and that we are moving into an era where much of what we know today, much of what is coded and written electronically, will be lost forever. We are, to my mind, living in the midst of digital Dark Ages [...].“

Diese trübe Aussicht erfordert eine genauere Analyse der Aufgaben, die die digitale Langzeitarchivierung tatsächlich erfüllen kann. Sie muss weiterhin zu Methoden und Werkzeugen führen, die dem digitalen Vergessen entgegenwirken können. Die Anforderungen an die Langzeitarchivierung wurden von der Caulfield School of Information Technology⁵ in Form eines Frameworks herausgearbeitet:

“Enable reliable, authentic, meaningful and accessible records to be carried forward through time within and

beyond organisational boundaries for as long as they are needed for the multiple purposes they serve."

Eine funktionierende Langzeitarchivierung muss also jederzeit die

- unversehrte und unverfälschte Aufbewahrung des Datenstroms (Substanzerhaltung) gewährleisten
- Erhaltung der Benutzbarkeit sicherstellen
- Erhebung und strukturierte Speicherung technischer Metadaten implementieren und
- Bedeutung und Gültigkeit der Dokumente sichern (Authentizität).

So können zukünftige Benutzer glaubwürdige Informationsobjekte

- uneingeschränkt abrufen,
- problemlos entschlüsseln und
- verstehen / interpretieren.

Ein Unternehmen stößt bei der Umsetzung auf drei Schwierigkeiten:

- Speichermedien sind grundsätzlich einer Alterung unterworfen
- die ursprüngliche Hardware ist defekt oder nicht mehr vorhanden
- Dateiformate und Softwareprodukte altern ebenfalls und können nach Jahren nicht mehr verwendet werden

Die langfristige Speicherung von digitalen Dokumenten ist also nur einen Teil der Lösung. Genauso wichtig ist es, dass die Daten auch in vielen Jahren noch gelesen und verarbeitet werden können. Es muss daher auch auf die Formate und die zugehörige Software geachtet werden. Sobald im Unternehmen ein Generationswechsel der Hardware, Software oder der verwendeten Formate ansteht, kommt auf das Unternehmen ein nicht unerheblicher Aufwand zu.

Unternehmen sind bestrebt, wichtige und sensible Daten möglichst lange verwenden zu können. Außerdem sind gesetzliche Aufbewahrungspflichten (z.B. aus haftungs- und versicherungstechnischen Gründen) vorgeschrieben, die auch bei digitalen Daten eingehalten werden müssen. Viele Unternehmen messen dem Thema Langzeitarchivierung und damit auch dem Informationsmanagement nicht ausreichende Bedeutung bei. Tatsächlich gehört das Informationsmanagement jedoch zu den unternehmensstrategischen Aufgaben, die über Erfolg oder Misserfolg eines Unternehmens entscheiden können. Entscheidungen des Managements über den Umgang mit digitalen Dokumenten, beinhaltet auch Entscheidungen über Datensicherung und Archivierung.

Das zeigt, dass die Langzeitspeicherung im Unternehmen einen wichtigen Platz einnehmen muss, um den künftigen Zugang zu archivierten Daten aufrecht erhalten zu können. Es geht nicht länger nur um die Substanzerhaltung digitaler Informationen, sondern darüber hinaus um die dauerhafte Verfügbarkeit digitaler Ressourcen und um praktikable Maßnahmen gegen das digitale Vergessen.

1.3. Vorgehensweise und Gliederung

Nach einem grundsätzlichen methodischen Überblick werden im zweiten Kapitel die Grundlagen der digitalen Langzeitarchivierung und die Ursachen für den Verlust digitaler Informationen herausgearbeitet. Die gefühlte digitale Amnesie innerhalb der Unternehmen wird entmystifiziert und die Gründe für das Problem aufgezeigt.

Diese Sensibilisierung ermöglicht im dritten Kapitel die Definition der Revisionssicherheit in Hinblick auf die gestellten Anforderungen an ein digitales Langzeitarchiv.

Im Folgenden werden diverse Medien und Speichersysteme auf ihre Eignung als Langzeitspeicher untersucht und die Bedeutung des OAIS⁶-Referenzmodell als Organisationsmodell für den Betrieb eines Langzeitarchivs im Unternehmen herauskristallisiert. Dies ermöglicht es im sechsten Kapitel Lösungsansätze in Form von Erhaltungsstrategien vorzustellen und hinsichtlich ihrer Eignung zu bewerten, wirksame Mittel im Kampf gegen den Verlust von Informationen zu sein.

In den Kapiteln sieben und acht wird versucht langzeitstabile Datenformate zu identifizieren und ein Problembewusstsein für die Bedeutung der Metadaten bei der Langzeitarchivierung zu schaffen.

Alle Untersuchungen münden letztlich in einem Lösungsvorschlag, der neben den Lösungen für die Teilprobleme auch die Bedeutung der praktizierten Archivierungsorganisation und der gelebten Archivierungskultur aufgreift. Dies führt zu einem kombinierten Ansatz, der abschließend in einem kritischen Fazit überprüft wird.

¹ Vgl. [Berkeley, 2003] S. 2

² Vgl. [Engelhardt, 2008] S. 2

³ Amnesie bezeichnet ursprünglich eine Form der Gedächtnisstörung für zeitliche oder inhaltliche Erinnerungen.

⁴ Vgl. [Kuny, 1997] S. 1-2

⁵ Vgl. [McKemmish, Acland, Reed, 1999] S. 4-5

⁶ Open Archival Information System (OAIS)