



Migration und Emulation – Angewandte Magie?

Stefan E. Funk

Niedersächsische Staats- und
Universitätsbibliothek Göttingen

funk@sub.uni-goettingen.de



NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

Kurze Einführung in die Thematik

Digitale Langzeitarchivierung

- Digitale Objekte
- Dateiformate
- Bitstream Preservation
- Migration
- Emulation
- Technical Preservation
- Medienwechsel

Digitale Objekte

Synchrone Existenz auf drei Ebenen

- Physisches Objekt
- Logisches Objekt
- Konzeptuelles Objekt

Digitale Objekte

Physisches Objekt

- Zeichen, die auf einem Datenträger gespeichert sind
- Einzelne Bits haben keinerlei Bedeutung, alle Bits sind Teil eines Datenstromes
- Beispiel: Die Zustände „magnetisiert“ und „nicht magnetisiert“ auf einer Diskette

Digitale Objekte

Logisches Objekt

- Bits, die von einem Datenträger gelesen werden
- Besteht aus einer beliebigen (endlichen) Folge von „1“ und „0“
- Wird als Dateiformat von Programmen erkannt
- Beispiel: Ein Bitstream aus Nullen und Einsen (01100110...011000011) in Form einer Textdatei oder einer Grafik

Digitale Objekte

Konzeptuelles Objekt

- Ist die eigentliche bedeutungsvolle Einheit
- Zusammenspiel von Bitstream, Dateiformat, Software und Betriebssystem
- Beispiel: Eine Textdatei, die in einem Textverarbeitungsprogramm bearbeitet werden kann

Digitale Objekte

Erhalt des konzeptuellen Objekts

- Ist das Ziel der Langzeitarchivierung
- Probleme für den Erhalt resultieren aus Problemen in den unteren Ebenen – technische Obsoleszenz
- Beispiel: Hardware / Software existiert nicht mehr, Bitstream ist beschädigt

Dateiformate

Für die Langzeitarchivierung von großer Bedeutung

- Offen spezifizierte Formate (XML, PDF, HTML)
 - Die Spezifikation ist für jeden zugänglich
 - Bevorzugte Formate für die Archivierung
- Proprietäre Formate (DOC, PPT, XLS)
 - Die Spezifikation ist weitgehend unbekannt
 - Schwierig zu migrieren

Dateiformate

- **Formaterkennung**

- Kenntnis des Formats und der Version ist erforderlich für die Darstellung
- Kenntnis der Spezifikation eines archivierten Formats ist zwingend erforderlich für die langfristige Erhaltung des Dokuments
- Format und Version müssen als Metadaten vorliegen

- **Formatvalidation**

- Korrektheit der Angaben muss sichergestellt sein
- Format muss im Sinne der Spezifikation korrekt sein, um eine spätere Bearbeitung / Darstellung zu ermöglichen

Dateiformate

Metadaten: Hilfsmittel zur automatisierten Erfassung

- Maschinelle Erfassung von Metadaten
 - Deskriptive Metadaten aus einem digitalen Katalogsystem
 - Technische Metadaten mit Hilfsprogrammen aus dem Dokument extrahieren
- Qualität hängt vom benutzten Tool ab
- Tools müssen an spezielle Anforderungen des Nutzers angepasst werden (JHOVE)

Bitstream Preservation

Sicherung der Daten auf der physischen Ebene

- Datenträgermigration
 - Kopieren auf neue Datenträger
 - Anpassen an neue Speicher-Technologien
- Grundlage aller weiteren Archivierungsaktivitäten
- Langjährige Erfahrung mit der Archivierung von „analogen“ Medien:
 - Bücher, Zeitschriften, Fotos, Mikrofilm
- Vereinen alle drei Ebenen des digitalen Objekts
 - physische Ebene = logische Ebene = konzeptuelle Ebene

Migration

Digitale Objekte werden einem neuen Umfeld angepasst

- Digitale Dokumente werden in ein aktuelleres (einheitliches) Format konvertiert
- Beispiel: Word 97 Dokument > Open Office Dokument

Migration

Zur Migration werden Metadaten benötigt

- Technische Metadaten zur Verarbeitung der Dokumente innerhalb des Archivs
 - Beispiel: Dateiformat, Dateigröße, Bildformat
- Deskriptive Metadaten zum komfortablen Zugriff auf Objekte
 - Beispiel: Titel und Autor
- Significant Properties: Welche Eigenschaften des Objekts sollen erhalten bleiben?

Migration

Mögliche Probleme bei der Migration

- Unzureichende Kenntnis der Format-Spezifikationen
- Keine direkte Übersetzung eines Formates in ein anderes möglich
- Komplexe Formate > (zu) komplexe Migrationsroutinen
- Resultat: Datenverlust durch Konvertierung in andere Formate

Emulation

Das Umfeld wird an die digitalen Objekte angepasst

- Das digitale Objekt bleibt dabei unverändert
- Datenverluste sollen verhindert werden, indem das komplette Umfeld eines Dokument-Typs nachgebildet wird
- Alle Eigenschaften – nicht nur die Significant Properties – sollen so erhalten bleiben

Emulation

Emulation von Anwendungs-Software

- Beispiel: PDF-Reader und eine recht alte PDF-Datei, die mit aktuellen Readern nicht angezeigt werden kann
- Wenn die Hardware obsolet wird, wird auf dem dann aktuellen Rechner (sagen wir in 20 Jahren) ein PDF-Reader entwickelt, der die originalen PDF-Dateien darstellen kann > Emulator
- Bei jedem Generationswechsel muss dann ein solcher Emulator entwickelt werden

Emulation

Emulation von Betriebssystemen

- Beispiel: Commodore-64 Betriebssystem und Software
- Die Hardware wird per Software nachprogrammiert
- Die originalen Spiele werden dann wie auf dem C-64 dargestellt und sind spielbar
- Auf allen nachfolgenden Hardware-Plattformen muss dann ein C-64 Emulator programmiert werden

Emulation

Emulation von Hardware-Plattformen

- Beispiel: Es wird per Software ein Pentium 233 emuliert, incl. Grafikkarte, Soundkarte, etc.
- Auf diesem virtuellen PC kann dann zum Beispiel ein originales Windows 98 installiert werden

Vor- und Nachteile von Migration

Vorteile

- Technisch relativ gut zu realisieren
- Unabhängig von einzelnen Emulatoren

Nachteile

- Jedes einzelne Objekt muss migriert werden
- Möglicherweise Datenverlust / -veränderung
- Alle Migrationen eines Objekts sollten gespeichert werden

Vor- und Nachteile von Emulation

Vorteile

- Originalobjekte bleiben unverändert
- Nur ein Emulator für einen Objekt-Typ benötigt
- Weniger Speicherverbrauch

Nachteile

- Schwierig für komplexe Objekte
- Hoher Aufwand pro Generationswechsel
- Spezifikationen für ein System sind nicht immer bekannt

Technical Preservation

Erhaltung der originalen Hard- und Software

- Beispiel: Computer, Laufwerke, Monitore, Kabel, Medien wie Disketten, Software, Joysticks, uswuf.
- Vorteile
 - Daten können im „originalen“ Zustand präsentiert werden
- Nachteile
 - Ist nicht für die Ewigkeit möglich: Haltbarkeit der Hardware
 - Kosten für die Wartung / Reparatur
 - Personalkosten

Medienwechsel

„Analogisierung“ von digitalen Dokumenten

- Beispiel: Mikroverfilmung von Retro-Digitalisaten
- Vorteile
 - Haltbarkeit: Langjährige Erfahrung mit den analogen Medien
- Nachteile
 - Wie werden die Inhalte der Mikrofilme katalogisiert? Digital?
 - Wie werden die digitalen Kataloge archiviert?
 - Zugriff ist nur beschränkt möglich
 - Komfortable Suche ist (analog) nicht möglich

Vielen Dank!

Fragen

- Gibt es jetzt Fragen? Bitte gern!
- Spätere Fragen gern per E-Mail an

funk@sub.uni-goettingen.de

Referenzen

- Huth, Karsten, Andreas Lange: *Die Entwicklung neuer Strategien zur Bewahrung und Archivierung von digitalen Artefakten für das Computerspiele-Museum Berlin und das Digital Game Archive* (2004).
http://www.archimuse.com/publishing/ichim04/2758_HuthLange.pdf (09.10.2012)
- JHOVE: JSTOR/Harvard Object Validation Environment. <http://jhove.sourceforge.net> (09.10.2012)
- kopal: Glossar/Linkliste, *Migration*. <http://kopal.langzeitarchivierung.de> (09.10.2012)
- Cornell University Library: *Digital Preservation Management Tutorial*. <http://dpworkshop.org> (09.10.2012)