

nestor Handbuch:
**Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung**

9.1 Digitale Objekte

Herausgeber

Heike Neuroth
Hans Liegmann †
Achim Oßwald
Regine Scheffel
Mathias Jehn
Stefan Strathmann

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Im Auftrag von

nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit
digitaler Ressourcen für Deutschland
nestor – Network of Expertise in Long-Term Storage of Digital Resources
<http://www.langzeitarchivierung.de>

Kontakt

editors@langzeitarchivierung.de
c/o
Niedersächsische Staats- und Universitätsbibliothek Göttingen
Dr. Heike Neuroth
Forschung und Entwicklung
Papendiek 14
37073 Göttingen
Tel. +49 (0) 55 1 39 38 66

Der Inhalt steht unter folgender Creative Commons Lizenz:
<http://creativecommons.org/licenses/by-nc-sa/2.0/de/>



9.1 Digitale Objekte

Stefan E. Funk

Die erste Frage, die im Zusammenhang mit der digitalen Langzeitarchivierung gestellt werden muss, ist sicherlich die nach den zu archivierenden Objekten. Welche Objekte möchte ich archivieren? Eine einfache Antwort lautet hier zunächst: digitale Objekte!

Eine Antwort auf die naheliegende Frage, was denn digitale Objekte eigentlich sind, gibt die Definition zum Begriff „digitales Objekt“ aus dem OAI¹. Dieser Standard beschreibt ganz allgemein ein Archivsystem mit dessen benötigten Komponenten und deren Kommunikation untereinander, wie auch die Kommunikation vom und zum Nutzer. Ein digitales Objekt wird dort definiert als „An object composed of a set of bit sequences“, also als ein aus einer Reihe von Bit-Sequenzen zusammengesetztes Objekt. Somit kann all das als ein digitales Objekt bezeichnet werden, das mit Hilfe eines Computers gespeichert und verarbeitet werden kann. Und dies entspricht tatsächlich der Menge der Materialien, die langzeitarchiviert werden sollen, vom einfachen Textdokument im .txt-Format über umfangreiche PDF-Dateien bis hin zu kompletten Betriebssystemen.

Ein digitales Objekt kann auf drei Ebenen beschrieben werden, siehe Abbildung:

- als physisches Objekt,
- als logisches Objekt und schließlich
- als konzeptuelles Objekt.

Ein digitales Objekt kann beispielsweise eine Datei in einem spezifischen Dateiformat sein, z.B. eine einzelne Grafik, ein Word-Dokument oder eine PDF-Datei. Als ein digitales Objekt können auch komplexere Objekte bezeichnet werden, wie Anwendungsprogramme wie Word oder Mozilla, eine komplette Webseite inkl. Text und Grafik, eine durchsuchbare Datenbank auf CD inklusive einer Suchoberfläche oder ein Betriebssystem wie Linux, Windows oder Mac OS .

Das physische Objekt - Daten auf einem Speichermedium

Als physisches Objekt sieht man die Menge der Zeichen an, die auf einem Informationsträger gespeichert sind. Die Art und Weise der physischen Be-

1 Open Archival Information System

schaffenheit dieser Zeichen kann aufgrund der unterschiedlichen Beschaffenheit des Trägers sehr unterschiedlich sein. Auf einer CD-ROM sind es die sogenannten „pits“ und „lands“ auf der Trägeroberfläche, bei magnetischen Datenträgern sind es Übergänge zwischen magnetisierten und nicht magnetisierten Teilchen. Auf der physischen Ebene haben die Bits keine weitere Bedeutung außer eben der, dass sie binär codierte Information enthalten, also entweder die „0“ oder die „1“. Auf dieser Ebene unterscheiden sich beispielsweise Bits, die zu einem Text gehören, in keiner Weise von Bits, die Teil eines Computerprogramms oder Teil einer Grafik sind.

Die Erhaltung dieses Bitstreams (auch Bitstreamerhaltung) ist der erste Schritt zur Konservierung des gesamten digitalen Objekts, er bildet sozusagen die Grundlage aller weiteren Erhaltungs-Strategien.

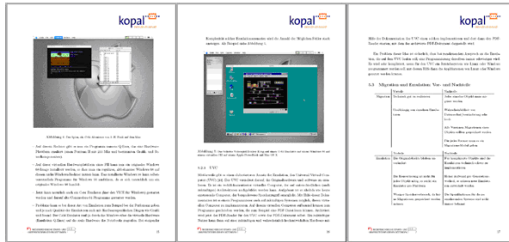
Das logische Objekt - Eine Folge von Bits als Einheit

Unter einem logischen Objekt versteht man eine Folge von Bits, die von einem Informationsträger gelesen und als eine Einheit angesehen werden kann. Diese können von einer entsprechenden Software als Format erkannt und verarbeitet werden. In dieser Ebene existiert das Objekt nicht nur als Bitstream, es hat bereits ein definiertes Format. Die Bitstreams sind auf dieser Ebene schon sehr viel spezieller als die Bits auf dem physischen Speichermedium. So müssen diese zunächst von dem Programm, das einen solchen Bitstream zum Beispiel als eine Textdatei erkennen soll, als eine solche identifizieren. Erst wenn der Bitstream als korrekte Textdatei erkannt worden ist, kann er vom Programm als Dateiformat interpretiert werden.

Will man diesen logischen Einheiten ihren Inhalt entlocken, muss das Format dieser Einheit genau bekannt sein. Ist ein Format nicht hinreichend bekannt oder existiert die zu dem Format gehörige Software nicht mehr, so wird die ursprüngliche Information des logischen Objektes sehr wahrscheinlich nicht mehr vollständig zu rekonstruieren sein. Um solche Verluste zu vermeiden, gibt es verschiedene Lösungsansätze, zwei davon sind Migration oder Emulation.

Das konzeptuelle Objekt - Das Objekt „zum Begreifen“

Das konzeptuelle Objekt beschreibt zu guter Letzt die gesamte Funktionalität, die dem Benutzer des digitalen Objekts mit Hilfe von dazu passender Soft- und Hardware zur Verfügung steht. Dies sind zunächst die Objekte, Zeichen und Töne, die der Mensch über seine Sinne wahrnimmt. Auch interaktive Dinge wie das



Wahrnehmung durch den Nutzer
mit Hilfe von Soft- und
Hardware

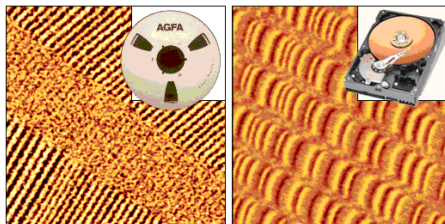
Konzeptuelles Objekt

```
001001000111001
1001000011001101
0011100110001100
1001001001010101
1001000100001111
1001110000011010
```



Der Bitstrom wird durch Software
als Dateiformat erkannt

Logisches Objekt



physische Datenstruktur

Physisches Objekt

Die drei Ebenen eines digitalen Objekts

Spielen eines Computerspiels oder eine durchsuchbare Datenbank zählen dazu, denn die Funktion eines Computerspiels ist es, gespielt werden zu können. Ein weiteres Beispiel ist eine komplexe Textdatei, mit all ihren Editierungsmöglichkeiten, Tabellen und enthaltenen Bildern, die das verarbeitende Programm bietet.

Dieses konzeptuelle Objekt ist also die eigentliche, für den Betrachter bedeutungsvolle Einheit, sei es ein Buch, ein Musikstück, ein Film, ein Computerprogramm oder ein Videospiel. Diese Einheit ist es, die der Nachwelt erhalten bleiben soll und die es mit Hilfe der „Digital Preservation“ zu schützen gilt.

Die Erhaltung des konzeptuellen Objekts

Das Ziel eines Langzeitarchivs ist es also, das konzeptuelle Objekt zu archivieren und dem Nutzer auch in ferner Zukunft Zugriff auf dessen Inhalte zu gewähren. Die Darstellung bzw. Nutzung des digitalen Objekts soll so nahe wie

möglich den Originalzustand des Objekts zu Beginn der Archivierung wieder spiegeln. Dies ist nicht möglich, wenn sich Probleme bei der Archivierung auf den unteren Ebenen, der logischen und der physischen Ebene, ergeben. Gibt es eine unbeabsichtigte Veränderung des Bitstreams durch fehlerhafte Datenträger oder existiert eine bestimmte Software nicht mehr, die den Bitstream als Datei erkennt, ist auch eine Nutzung des Objekts auf konzeptueller Ebene nicht mehr möglich.

Literatur

- Reference Model for an Open Archival Information System (OAIS)
<<http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCS-DS-650.0-B-1.pdf>> (letzter Zugriff: 7. Juni 2006)
- 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.ichim.org/ichim04/contenu/PDF/2758_HuthLange.pdf> (letzter Zugriff: 7. Juni 2006)
- Thibodeau, K.: Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. In The State of Digital Preservation: An International Perspective. Washington D.C.: Council on Library and Information Resources. 4-31 (2001)
<<http://www.clir.org/PUBS/abstract/pub107abst.html>> (letzter Zugriff: 12. Juli 2006)