



nestor Handbuch:
Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung

9 Formate

Herausgeber:

Heike Neuroth
Hans Liegmann
Achim Oßwald
Regine Scheffel
Mathias Jehn

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>

**Dieser Artikel ist ein Auszug aus dem
nestor Handbuch:
Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung**

Dieser Artikel ist verfügbar unter der URL:
http://nestor.sub.uni-goettingen.de/handbuch/artikel/text_84.pdf

Die Online Version des Handbuches unter der URL:
<http://nestor.sub.uni-goettingen.de/handbuch/>

Kontakt:
Niedersächsische Staats- und Universitätsbibliothek Göttingen
Dr. Heike Neuroth
Forschung und Entwicklung
Papendiek 14
37073 Göttingen
neuroth@sub.uni-goettingen.de
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 Formate

von Stefan Funk

Unter Formaten versteht man im Zusammenhang mit Computern Ähnliches wie zum Beispiel im Zusammenhang mit Bilderrahmen. Ein Bilderrahmen hat eine bestimmte Höhe und Breite, es kann noch die Breite und das Material des Rahmens dazu gezählt werden sowie der sichtbare Teil des Glasfläche und weiteres mehr. Formate bestimmen bei Bilderrahmen wie bei gespeicherten Computer-Daten einige Eigenschaften der Objekte, zu denen diese Eigenschaften gehören. So können kleine Bilderrahmen nur von kleinen Bildern sinnvoll genutzt werden, um diese Bilder zu präsentieren (und langsam fängt der Vergleich an, ein wenig zu hinken...). So wie Bilderrahmen aus Holz, Glas und Pappe bestehen, bestehen die Formate, die wir mit unseren Computern benutzen, aus einer Folge von Bits, also aus einer Folge von Nullen (0) und Einsen (1), die wir als Bitstrom bezeichnen.

Wenn wir ein Bild in einen Rahmen einpassen wollen, so fällt uns sehr schnell auf, dass ein Rahmen zu klein ist, die Farbe des Rahmens nicht zum Bild passt oder dass der Rahmen vielleicht gar nicht uns gehört, und wir ihn aus diesem Grund nicht verwenden können oder dürfen. Alle diese Dinge entscheiden wir meist intuitiv und manchmal dauert es auch länger, bis wir zu einer Entscheidung kommen. Aus genau diesem Grund müssen die Daten, die wir mit Computern speichern und bearbeiten, eine genau definierte Form haben, schließlich ist Intuition so ziemlich das letzte, was wir zur Zeit von einem Computer erwarten.

Ein Programm muss also die Daten, die es verwalten soll, in einem definierten Format vorfinden und speichern. Wenn wir die Daten später wieder in den Rechner laden wollen, um diese weiter zu bearbeiten, sei es ein Brief oder eine Grafik, wollen wir an genau der Stelle weiterarbeiten, an der wir aufgehört haben. "Definiertes Format" heißt hier, alle benötigten Daten in eine genau definierte Ordnung zu bringen, um diese dann als eine Reihe von Bits zu speichern. Die Bits, in denen der Titel des Dokuments gespeichert ist, müssen an exakt der selben Stelle und auch als Titel wieder in unser Programm geladen werden. Also muß das Programm das Format genau kennen, muß wissen, welche Bits des Bitstroms welche Bedeutung haben, um diese richtig zu interpretieren und verarbeiten zu können.

Eine Format-Spezifikation ist nun eine Beschreibung der Anordnung der Bits (ab welchem Bit steht der Titel, wo befinden sich die Kapitel, wo finden sich welche Steuerzeichen, usw.). Es gibt sehr viele verschiedene Formate, so wie es sehr viele verschiedene Programme gibt, und da beginnt das eigentliche Problem im Zusammenhang mit der Langzeitarchivierung. Im Prinzip hatte am Anfang der elektronischen Datenverarbeitung jedes Programm sein eigenes Format, und daraus resultierte, dass ein Programm keine Daten eines anderen nutzen konnte, aus dem einfachen Grund, dass die benötigten Informationen nicht an derselben Stelle im Bitstrom standen. Programm A verstand das Format von Programm B nicht und Programm B konnte mit dem Bitstrom von Programm A nichts anfangen.

Mittlerweile hat die Anwendung der Datenverarbeitung einen Umfang und einen Komplexitätsgrad erreicht, der es sinnvoll macht, Daten zwischen verschiedenen Programmen auch unterschiedlicher Hersteller austauschen zu können. Format-Standards werden legen Strukturen verbindlich fest, die dann von vielen Programmen gelesen und verarbeitet werden können, wie beispielsweise das Open Document Format¹ oder viele Grafik-Formate wie TIFF (Tagged Image File Format), GIF

¹ <http://www.odfalliance.org>

(Graphics Interchange Format), JPEG (Joint Photographic Experts Group) oder auch PDF (Portable Document Format). Schon an den Namen kann man erkennen, dass es sich um Formate handelt, die für den Austausch entwickelt wurden. Wenn man die Format-Spezifikationen also kennt, können Formate und somit deren Inhalte richtig interpretiert werden und auf diese Weise können sie, hoffentlich auch nach langer Zeit noch, korrekt interpretiert und der Nachwelt überliefert werden.

Die Inhalte solcher Bitströme (oder auch digitale Objekte bzw. Dateien) können also nur solange genutzt und präsentiert werden, wie ihr Format bekannt ist und man zum Beispiel weiß, welche Stelle des Bitstroms den Titel beschreibt, wo sich welche Grafiken befinden oder wo der Bitstrom endet. Aus diesem Grund sind Format-Standards für die Langzeitarchivierung sehr wichtig und auch absolut notwendig.