

# Langzeiterhaltung digitaler Daten in Museen

## Tipps zur dauerhaften Bewahrung digitaler Daten

7

### Dateiformatauswahl

Ein Faktor für die Langzeiterhaltung digitaler Daten ist die Sicherstellung der Interpretationsfähigkeit. Audio-, Bild-, Text- und Videodateien sowie Datenbanken müssen vom Computer zunächst gelesen werden, bevor sie in Ton, Bild, Text und Video umgewandelt und ausgegeben werden können. Es stellt sich dabei die Frage, in welchem Dateiformat die Daten gesichert werden sollen, um deren Nutzung dauerhaft gewährleisten zu können. Wie bei den Speichermedien ist die fortschreitende Entwicklung ein Risikofaktor. Dateiformate veralten, da sie von neuen Programmen nicht mehr gelesen und interpretiert werden können.

Um dieses Risiko zu reduzieren, ist es sinnvoll, gezielt Kriterien für die Auswahl von Dateiformaten für die Nutzung im Rahmen der Langzeiterhaltung zu definieren. Diese Kriterien helfen, folgende Hauptanforderungen zu erfüllen:

- Die Auswahl ist auf Basis der vorhandenen digitalen Daten vorzunehmen.
- Die Anzahl der Formate sollte so gering wie möglich sein.
- Die Datenmigration aus anderen Formaten sollte nur in Ausnahmefällen notwendig sein.

#### Dateiformat und Datenformat

Ein Dateiformat ist ein definiertes System zur Ablage von Daten. Diese Daten beinhalten nicht allein die "Objektdaten" (die digitalen Informationen), z. B. bei einem Foto die Bilddaten, sondern können

auch zusätzliche Bereiche für Metadaten (Autor, Fotograf, Sujet etc.) und Informationen zur Interpretation der "Objektdaten" enthalten.

Es gibt sogenannte Containerformate, sie ermöglichen die Ablage unterschiedlicher Informationen in einer einzelnen Datei. So sind in einer Videodatei als Container Metadaten, Video- und auch Audiodaten enthalten.

#### Auswahlkriterien

An dieser Stelle sollen keine direkten Empfehlungen für einzelne Dateiformate ausgesprochen (→ Blatt 8 - 12), sondern Kriterien für die Auswahl genannt werden.

- Open Source

Langfristig ist die freie Verfügbarkeit der Formatdokumentation eines der wichtigsten Kriterien bei der Auswahl von Dateiformaten. Es

besteht die Möglichkeit, dass anhand der Dokumentation entsprechende Tools programmiert werden können, damit auch ohne die ursprünglich verwendeten Programme die Daten gelesen und genutzt werden können.

- Verbreitung

Dateiformate, die von einer Vielzahl von Programmen unterstützt werden, veralten in der Regel nicht so schnell wie nur selten genutzte. Somit ist die weite Verbreitung des Formates ein Argument für die Nutzung als Langzeitarchivierungsformat.

- Stabilität

Unter Stabilität wird in diesem Zusammenhang die Dauerhaftigkeit der Formatdefinition verstanden. Bei neuen Formatversionen sollte "Abwärtskompatibilität" gegeben sein, so dass auch Dateien in älteren Versionen noch genutzt werden können.

- Interoperabilität

Das Dateiformat sollte den Austausch mit anderen Systemen unterstützen. Aus diesem Grund ist ein Format vorzuziehen, das von einer breiten Gruppe von Software und Betriebssystemen unterstützt wird.

Je nach Konzeption der Langzeitbewahrung können weitere Aspekte für die effektive Auswahl der Dateiformate von Bedeutung sein:

- Die Nutzung von Datenkomprimierung im Rahmen der Langzeiterhaltung ist nicht unproblematisch. Grundsätzlich wird zwischen verlustloser und verlustbehafteter Kompression unterschieden. Die Nutzung von Kompressionsverfahren bedeutet zusätzliche Abhängigkeiten, deren langfristige Entwicklung unsicher ist.
- Je nach Einsatzzweck stehen Formate mit unterschiedlichen Merkmalen zur Verfügung. So speichert beispielsweise das MS Word-Dateiformat (.doc bzw. .docx) neben dem eigentlichen Text und der Formatierung auch Informationen über Gliederungsfunktionen, Änderungshistorie, Optionen zur Erstellung von Indices etc.
- Wird die Möglichkeit einer späteren Weiterverarbeitung bzw. Modifikation des Dateiinhaltes benötigt, so sind Formate zu wählen, die dieses zulassen. Textinhalte in PDF-Dokumenten (.pdf) sind z. B. weniger gut zu editieren als Texte im ASCII-Format (.txt).
- Unterschiedliche Dateiformate können unterschiedliche Metadaten bereitstellen, bzw. deren Speicherung in der Datei zulassen, beispielsweise IPTC-Daten im Header einer jpg-Datei. Die Notwendigkeit der Speicherung von Metadaten in

der Datei kann sich je nach Erhaltungskonzept unterscheiden und ist ggf. bei der Auswahl der Formate zu berücksichtigen.

### Dateiformatfestlegungen

Es genügt nicht allein, die Dateiformate zu bestimmen. Es sind weitere Festlegungen vorzunehmen. Bei der Wahl des TIFF-Formates für Bilddaten beispielsweise sollte auch die Frage der Zulässigkeit der LZW-Kompression (in TIFF zur Verfügung stehendes Komprimierungsverfahren) oder der Multipage-Nutzung (mehrere Bilder in einer Datei) bestimmt werden. Bei audiovisuellen Formaten ist beispielsweise auch der Codec für die Bild- und Tondaten zu benennen, die genutzt werden sollen. Erfolgt diese Präzisierung nicht, können trotz der Verwendung des zugelassenen Dateiformates Probleme bei der Nutzung auftreten.

Bei der Wahl des TIFF-Formates für Bilddaten beispielsweise sollte auch die Frage der Zulässigkeit der LZW-Kompression (in TIFF zur Verfügung stehendes Komprimierungsverfahren) oder der Multipage-Nutzung (mehrere Bilder in einer Datei) bestimmt werden. Bei audiovisuellen Formaten ist beispielsweise auch der Codec für die Bild- und Tondaten zu benennen, die genutzt werden sollen. Erfolgt diese Präzisierung nicht, können trotz der Verwendung des zugelassenen Dateiformates Probleme bei der Nutzung auftreten.

### Migration

Trotz Beschränkung auf wenige Dateiformate wird in der Regel bei schon vorhandenem Datenbestand die Migration von Daten, die in anderen Dateiformaten vorliegen, erforderlich. Bei der Migration sollten Modifikationen möglichst gering gehalten werden, so dass Präsentation, Funktionalität und Inhalt dem Original entsprechen (→ Blatt 2).

Trotz Beschränkung auf wenige Dateiformate wird in der Regel bei schon vorhandenem Datenbestand die Migration von Daten, die in anderen Dateiformaten vorliegen, erforderlich. Bei der Migration sollten Modifikationen möglichst gering gehalten werden, so dass Präsentation, Funktionalität und Inhalt dem Original entsprechen (→ Blatt 2).

### Datenproduktion

Die ausgewählten Dateiformate und die definierten Parameter (Codecs, Multipage etc.) sind bei der Erzeugung neuer Dateien, die als potentiell bewahrenswürdig angesehen werden, als Vorgaben zu verwenden. Die Notwendigkeit der Migration kann damit verringert werden.

Die ausgewählten Dateiformate und die definierten Parameter (Codecs, Multipage etc.) sind bei der Erzeugung neuer Dateien, die als potentiell bewahrenswürdig angesehen werden, als Vorgaben zu verwenden. Die Notwendigkeit der Migration kann damit verringert werden.

Für die Langzeiterhaltung digitaler Daten sind die notwendigen Dateiformate zu bestimmen. Diese Auswahl sollte möglichst wenige Formate enthalten. Offene Formate sind proprietären Formaten vorzuziehen.

Die erfolgte Formatauswahl ist den internen und externen Produzenten von digitalen Daten mitzuteilen, damit eine Übernahme in die Langzeiterhaltung vereinfacht wird.



<http://www.erpanet.org/guidance/docs/ERPANETIngestTool.pdf>

Beschreibung eines Ingest-Tools, hrsg. von Erpanet.

[http://aic.stanford.edu/sg/emg/library/pdf/vitale/2007-07-vitale-digital\\_image\\_file\\_formats.pdf](http://aic.stanford.edu/sg/emg/library/pdf/vitale/2007-07-vitale-digital_image_file_formats.pdf)

Digital Image File Formats - TIFF, JPEG, JPEG2000, RAW and DNG von Tim Vitale, Universität Stanford.

[http://www.interpares.org/display\\_file.cfm?doc=ip2\\_file\\_formats%28complete%29.pdf](http://www.interpares.org/display_file.cfm?doc=ip2_file_formats%28complete%29.pdf)

"Selecting Digital File Formats for Long-Term Preservation", InterPARES 2 General Study 11 Final Report von Evelyn Peters McLellan, Interpares-Projekt.

<http://www.dpconline.org/graphics/handbook/>  
"Preservation Management of Digital Materials: The Handbook, 5. Media and Formats", hrsg. von der Digital Preservation Coalition.

<http://netarchive.dk/publikationer/FileFormats-2004.pdf>  
"Handling file formats", hrsg. von der Staats- und Universitätsbibliothek Aarhus.

<http://www.dcc.ac.uk/resource/curation-manual/chapters/file-formats>  
"Instalment on File Formats", hrsg. vom Digital Curation Centre.

[http://www.jisc.ac.uk/uploaded\\_documents/FileFormatsreport.pdf](http://www.jisc.ac.uk/uploaded_documents/FileFormatsreport.pdf)  
"Survey and assessment of sources of information on file formats and software documentation", hrsg. von der Universität Leeds.

LINKS

Stand: Juni 2009