

# Langzeiterhaltung digitaler Daten in Museen

## Tipps zur dauerhaften Bewahrung digitaler Daten

12

### Digitale Textdokumente

Wie bei allen digitalen Vorlagen sind bei der Langzeiterhaltung digitaler Texte zwei Bereiche zu berücksichtigen. Erstens das Problem veralteter Datenträger und zweitens die Probleme im Umgang mit veralteten Programmen und Programmversionen. Im Blickfeld stehen die notwendigen Maßnahmen zur Erhaltung bereits vorhandener digitaler Textinformationen sowie deren Überführung in langzeiterhaltungsfähige Kopien. Vorausschauendes Verhalten bei der Auswahl von Programmen zur Erzeugung neuer digitaler Textdokumente ist daher notwendig.



Joseph Furtenbach; *Architectura Recreationis*, Augsburg 1640 (Reprint Berlin 1988)

Zu beachten ist, dass schon bei unterschiedlichen Programmversionen ein- und desselben Herstellers die Programmfunktionen und die Art der Textspeicherungen erheblich voneinander abweichen können. Ausgelassene Programmupdates oder der Wechsel zwischen zwei Programmen können daher zum Verlust an Informationen führen. Betroffen sind hiervon häufig Funktionen wie Seitenzählung, Fußnoten, Tabellen, Wort und Zeichenzählungen, Fuß- und Kopfzeilen, Indizes, Silbentrennung, etc. Die fehlerhafte Interpretation von eingebundenen Bildern oder von verwendeten Zeichensätzen haben z.B. den Verlust von Sonderzeichen und Umlauten und die Verfälschung der Inhalte zur Folge.

Grundsätzlich ist bei der Langzeiterhaltung digitaler Textdokumente zu entscheiden, was an den Dokumenten langzeiterhaltungswertig ist. Für diese Frage ist entscheidend, ob es sich um ein Dokument handelt, das unverändert und unveränderbar bewahrt werden soll, oder um ein Arbeitsdokument, das als Grundlage für zukünftige Nutzungen mit Inhalt und Funktionen zur Verfügung stehen muss.

In jedem Fall ist darauf zu achten,

- alte Speichermedien auszutauschen (→ Blatt 18),
- alte Dokumente aus veralteten Programmversionen zu aktualisieren und inhaltlich zu überprüfen.

Dieses Vorgehen stellt noch keine Langzeiterhaltung von Textdokumenten dar, es ist der erste Schritt.

Um die Langzeiterhaltung von Textdokumenten zu gewährleisten, sollte jedes Dokument in zwei Dateiformaten gespeichert werden. Dabei dient eine Version dazu, das Dokument im erhaltungswerten

IST-Zustand, als "finalisiertes" Dokument, abzubilden und die andere Version dazu, das Dokument mit einem Maximum an Funktionalität als Arbeitskopie zu sichern.

#### Finale Dokumente

Ein möglicher Weg „finale“ Dokumente zu archivieren ist PDF/A-1. Dieses Format wurde speziell für die Langzeitarchivierung entwickelt. Alle anderen Versionen von PDF-Dateien sind nicht geeignet.

Eine andere Möglichkeit besteht darin, Text in Form von Bilddateien im langzeiterhaltungsfähigen TIFF-Format zu speichern.

Generell ist die Langzeitarchivierung digitaler Daten ein aktiver Prozess, der die stetige Kontrolle der verwendeten Arbeitsabläufe und eingesetzten Verfahren erfordert.

#### Dokumente als Arbeitsgrundlage

Sollten für die Langzeiterhaltung vorgesehene Dokumente auch zu einem späteren Zeitpunkt als Arbeitsgrundlage benötigt werden, ist es von zentraler Bedeutung, mehr als nur die Buchstabenfolge oder das Aussehen des Dokuments zu erhalten. Eine große Anzahl der Informationen eines Textes basieren auf den oben genannten Programmfunktionen. Texte sind ohne diese nur begrenzt nutzbar oder müssen aufwendig nachbereitet werden.

Als Lösungsweg für die langfristige Erhaltung bietet sich die Speicherung als XML-Dokument (Extensible Markup Language = erweiterbare Auszeichnungssprache) an. Bei der Speicherung in XML bleiben Struktur und Aussehen der Textdokumente rekonstruierbar. Entscheidend bei der Langzeiterhaltung von XML-Dokumenten ist die Angabe des zugrunde gelegten Schemas.

Die Speicherung im sogenannten ASCII-Standard (ANSI) ist dann eine Option, wenn es tatsächlich nur auf den Inhalt des Textes und nicht auf mögliche Formatierungen, wie z.B. der Fußnoten, Seitenzählungen, eingebundene Bilder ankommt.



Für die Langzeiterhaltung digitaler Textdokumente kann der Erhalt der Informationen, aber auch der Erhalt von Funktionen von großer Bedeutung sein. In diesen Fällen ist die doppelte Speicherung als finales, unveränderbares Dokument und als bearbeitbares Dokument mit maximaler Funktionalität sinnvoll.

Die Dokumentation des Vorgehens, beispielsweise des zugrunde gelegten XML-Schemas und dessen Version, ist anzugeben und bildet einen unverzichtbaren Bestandteil der Maßnahme.

Bei PDF ist immer zu prüfen, ob es sich um das PDF/A-1-Format handelt, denn nur dann ist das Dokument langzeiterhaltungsfähig.

[http://www.apsr.edu.au/publications/word\\_processing\\_preservation.pdf](http://www.apsr.edu.au/publications/word_processing_preservation.pdf)

Publikation von 2006 zu den grundsätzlichen Herausforderungen bei der Bewahrung textueller Inhalte, hrsg. von APSR (Australian Partnership for sustainable Repositories).

[http://ahds.ac.uk/preservation/presBinary\\_v4.rtf](http://ahds.ac.uk/preservation/presBinary_v4.rtf)

Anleitung zur Bewahrung von digitalen Texten, hrsg. vom AHDS (arts and humanities data service).

[http://www.kb.nl/hrd/dd/dd\\_links\\_en\\_publicaties/PDF\\_Guidelines.pdf](http://www.kb.nl/hrd/dd/dd_links_en_publicaties/PDF_Guidelines.pdf)

Richtlinien für die Erstellung von langzeitarchivierungsfähigen PDF-Dokumenten. Hrsg. von der Koninklijke Bibliotheek in Den Haag.