

Langzeiterhaltung digitaler Daten in Museen

Tipps zur dauerhaften Bewahrung digitaler Daten

Datenbanken

11

Datenbanken (oder Datenbanksysteme) gehören zu den klassischen Anwendungen in der EDV. Die Aufnahme, Verwaltung, Recherche und Ausgabe großer Informationsmengen können mit Datenbankmanagementsystemen (DBMS) erleichtert und beschleunigt werden. In praktisch allen Tätigkeitsbereichen sind Datenbanken im Einsatz, in Museen vor allem im Rahmen der Inventarisierung und des Sammlungsmanagements.

Durch die Einführung leicht programmierbarer Datenbanksysteme wie MS Access oder Filemaker ist die Erstellung von Datenbankwendungen auch für Nicht-Spezialisten möglich geworden. Die dadurch ermöglichte Vielfalt an Datenbanken spiegelt sich auch in den Museen wider.

Unter den zahlreichen unterschiedlichen Datenbanksystemen sind derzeit die relationalen Datenbanken am weitesten verbreitet. Relationale Datenbanken bestehen aus Tabellen, in denen die Informationen abgelegt werden. Diese Tabelleninformationen können mit anderen Tabellen verknüpft werden. Daneben gibt es noch objektorientierte und hierarchische Datenbanken.

Um Informationen einzugeben, zu verknüpfen oder recherchierbar zu machen, werden Formulare und kleinere Programme als Teil des Datenbankmanagements geschaffen. Anders als Bild- oder Textdateien, die in sich abgeschlossene digitale Objekte repräsentieren, sind Datenbanken in den meisten Fällen dynamisch konzipiert. Sie sind nicht nur dazu angelegt, Informationen aufzunehmen und recherchierbar zu machen, sondern auch dazu, Informationen zu modifizieren.

Wirkliche Langzeiterhaltung ist nur für Datenbanken sinnvoll, die abgeschlossen sind, sei es, weil alle Informationen eingegeben wurden oder weil sie nicht mehr verändert werden. Die Sicherung aktueller, in Gebrauch befindlicher Datenbanken ist primär die Aufgabe von Backupsystemen im Rahmen der IT-Sicherheit.

Erhalt der Daten

Um die Inhalte der einzelnen Tabellen zu sichern, gibt es mehrere mögliche Vorgehensweisen. Eine Exportfunktion für die Ausgabe der Inhalte von Datenfeldern ist in jedem DBMS enthalten. Das

Datei- bzw. Datenformat, in welchem diese Inhalte gespeichert werden sollen, lässt sich in der Regel auswählen. Eine Ausgabe als CSV-Datei (comma separated value) oder als XML-Datei (extensible markup language) ist in den meisten Systemen möglich.

CSV

Ist die Möglichkeit eines Exports im CSV-Format vorhanden, dann ist der Export der Inhalte in eine einfache Textdatei möglich. Die einzelnen Einträge der Tabellen werden dabei im TXT-Format gespeichert. Um die Zuordnung zu den einzelnen Datenfeldern zu ermöglichen, werden die Einträge durch Kommata oder andere, zuvor definierte Zeichen getrennt. Die Datensätze selbst werden durch Absatzzeichen voneinander separiert.

Diese CSV-Dateien können durch Import in andere DBMS wieder als Datenbanktabellen verwendet werden. CSV-Dateien eignen sich, um die einzelnen Tabellen und deren Inhalt zu sichern.

XML

XML ist eine "erweiterungsfähige Auszeichnungssprache" zur Abbildung hierarchisch strukturierter Informationen in Form von Textdaten. Sie ermöglicht, die Datensätze bei der Nutzung als Exportformat für Datenbanken inklusive hierarchischer Informationsanordnung in einer einfachen Textdatei abzulegen, die auch ohne computergestützte Interpretation vom Menschen zu verstehen ist.

Der Vorteil gegenüber CSV-Daten besteht in der Möglichkeit, hierarchische Abhängigkeiten in der Struktur der Ausgabedatei abzubilden. XML ist mittlerweile als ein Ausgabeformat von DBMS etabliert und steht in den meisten Systemen zur Verfügung.

SQL

CSV und XML sind geeignet für die Extraktion der Tabelleninhalte in Textdateien. Es ist aber nicht möglich, in dieser Form Funktions- bzw. Abfrageabläufe zu sichern. Hierfür hat sich SQL (structured query language) als strukturierte Abfragesprache für die Darstellung der logischen Struktur in relationalen Datenbanken etabliert. SQL wurde von den einzelnen Datenbank-Herstellern allerdings teilweise modifiziert, so dass Inkompatibilitäten auftreten können. Funktionalitäten einer Datenbankwendung sind mit den automatischen Exportfunktionen nicht zu sichern. Hierfür wird eine möglichst eingehende Beschreibung der Zielstellung, des technischen und inhaltlichen Aufbaus benötigt.

Nicht zu vergessen sind auch die Schnittstellen (Masken, Formulare) für die Eingabe, die mit ihrem Layout, ihren Eingabehilfen und Funktionalitäten ein wichtiges Element des Datenbanksystems sind. Auch die Gestaltung von Ausgaben am Bildschirm oder am Drucker gehören zu den zu sichernden Informationen. Auch diese Elemente sind datenbanksystemunabhängig nur über eine angemessene Dokumentation zu erhalten.

Dokumentation

Auch wenn mit den oben angeführten Methoden die Inhalte und die logische Struktur der Datenbank extrahiert und erhalten werden können, so ist dennoch auf eine adäquate Dokumentation der

Programmierung und Funktionsweise der Datenbank nicht zu verzichten. Werden z. B. Tabellenverknüpfungen über SQL gesichert, erschließt sich der Grund der Verknüpfung und dessen Charakteristik nicht von allein. Hierfür wird eine möglichst eingehende Beschreibung der Zielstellung sowie des technischen und inhaltlichen Aufbaus der Datenbank benötigt.

Letztlich sind auch die Schnittstellen (Masken, Formulare, ...) für die Eingabe und Ausgabe Teil der Anwendung. Für eine vollständige Sicherung oder Archivierung ist es daher wichtig festzuhalten, wie die Eingabe- und Ausgabemasken gestaltet, und z. B. auch, welche Vorlagen für die Ausgabe auf Druckern vorgesehen waren.

Für die Langzeiterhaltung von Datenbanken stehen mehrere Verfahren und Techniken zur Verfügung. Unter Verwendung von CSV- und XML-Dateien können die Tabelleninhalte, mittels SQL-Dateien die Datenbankstruktur gesichert werden.



Grundsätzlich ist aber eine umfassende Dokumentation des Aufbaus und der Funktionalitäten für die langfristige Erhaltung notwendig.

<http://www.bar.admin.ch/themen/00772/index.html?lang=de>
SIARD (Software-Independent Archival of Relational Databases) ist die normative Beschreibung eines Dateiformates zur Erhaltung relationaler Datenbanken.

<http://conferences.scis.ecu.edu.au/TILC2007/documents/2007/Brogan%20changes%20in%20digital%20preservation.pdf>
Grundlegende Überlegungen zum Thema der Erhaltung von Datenbanken.

http://www.erpanet.org/events/2003/bern/Bern_Report_final.pdf
Bericht des ERPA-Workshop zur Langzeitarchivierung von Datenbanken

http://www.urz.uni-heidelberg.de/imperia/md/content/urz/programme/db_und_xml.pdf
Einsatz von XML zum Austausch und zur Sicherung von Datenbankinhalten.

<http://www.landesarchiv-bw.de/sixcms/media.php/25/erfahrungsbericht%20datenbanken.pdf>
Erfahrungsbericht eines Projektes zur Langzeiterhaltung von Datenbanken.

<http://www.digitaleduurzaamheid.nl/bibliotheek/docs/volatility-permanence-databases-en.pdf>
Projekt "digitale duurzaamheid" der Königlichen Bibliothek der Niederlande.

<http://aam.ugpl.de/?q=node/146>
Eine Einführung in die Strukturierte Abfragesprache (SQL).

<http://www.w3.org/XML/>
Offizielle Website des W3C zu XML.

LINKS

Stand: Juni 2009