

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

13.1 Retrieval

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:

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/>

13.1 Retrieval

Matthias Neubauer

Genauso wichtig wie die sichere Archivierung der digitalen Objekte ist auch die Möglichkeit, diese Objekte wieder aus dem Archiv herauszuholen und zu nutzen. Dabei muss gewährleistet sein, dass die Objekte den Zustand und den Informationsgehalt zum Zeitpunkt des Einspielens in das Archivsystem widerspiegeln. Im Idealfall sollte das Objekt noch exakt so abrufbar sein, wie es einmal in das Archiv eingespielt wurde. Je nach Verwendungszweck kann es jedoch auch sinnvoll sein, eher eine migrierte Form eines Objektes abzurufen. Einige wichtige Punkte, die es beim Zugriff von archivierten Objekten zu beachten gilt, sollen im Folgenden kurz erläutert werden.

Objektidentifikation

Zunächst ist eine eindeutige Identifikation des abzurufenden Objektes wichtig. Zu dieser Thematik existieren vielerlei Lösungen und Philosophien. Einige werden in den folgenden Kapiteln zum Thema „Persistent Identifier“ vorgestellt. Grundsätzlich muss es anhand der verwendeten Identifizierungen möglich sein, jedwede Form und Version eines digitalen Objektes aus dem Langzeitarchiv abzurufen. Dies kann gegebenenfalls auch durch eine Kombination von externen und internen Identifikatoren realisiert werden.

Datenkonsistenz

Die Unversehrtheit der Daten hat höchste Priorität. Innerhalb des Archivs sollte durch geeignete Routinen zwar sichergestellt sein, dass der originale digitale Datenstrom erhalten bleibt. Jedoch können auch - und vor allem - bei der Übertragung der Daten aus dem Archiv heraus Inkonsistenzen durch Übertragungsfehler oder andere Störeinflüsse entstehen. Idealerweise sollte daher bei jedem Zugriff auf ein Archivobjekt über Checksummenvergleiche die Unversehrtheit der Daten sichergestellt werden. Je nach Art und Status der Daten kann diese Überprüfung auch nur stichprobenartig erfolgen.

Versionsmanagement

Je nach Verwendungszweck der Daten kann es entweder sinnvoll sein, das ursprüngliche Originalobjekt aus dem Archiv herauszuholen, oder

aber auch eine migrierte Form zu nutzen. Die höchste Authentizität wird man sicherlich mit dem ursprünglichen Objekt erreichen, jedoch kann es sich auf zukünftigen Systemen sehr schwierig gestalten, die erhaltenen Daten aufzubereiten und zu nutzen (mehr darüber im Kapitel über Emulation und Migration). Ein gutes Langzeitarchivierungssystem sollte nach Möglichkeit sowohl Originalversion und letzte Migrationsform, als auch alle dazwischen liegenden Objektversionen zugreifbar halten, um eine vollkommene Transparenz und Rekonstruierbarkeit zu gewährleisten.

Interpretation und Aufbereitung der Daten

Sofern das digitale Objekt zum Zweck einer Präsentation oder Weiter-nutzung abgerufen wurde, muss es durch geeignete Methoden aufbereitet und verfügbar gemacht werden. Schon beim Einspielen der Daten in das Archivsystem ist daher darauf zu achten, dass man die Struktur des Objektes in den beiliegenden Metadaten dokumentiert. Zudem kann es notwendig sein, die innerhalb eines Archivsystems verwendeten Schlüsselnummern zur eindeutigen Identifikation von Dateiformaten zu entschlüsseln und auf ein anderes System einzustellen.

Caching

Unter dem Begriff „Caching“ versteht man die Pufferung oft genutzter Daten in einem schnell verfügbaren und hochperformanten Zwischenspeicher. Im Falle des Retrieval aus einem Langzeitarchivierungssystem ist dies dann sinnvoll, wenn die Archivobjekte auch als Basis für Präsentationssysteme und den täglichen Zugriff dienen sollen. Um das Archivsystem nicht mit unnötigen Anfragen nach häufig genutzten Objekten zu belasten, wird ein lokaler Zwischenspeicher angelegt, der stark frequentierte Objekte vorhält und gegebenenfalls mit einer neuen Version innerhalb des Archivsystems synchronisiert beziehungsweise aktualisiert. Bei einem Zugriff auf das Objekt wird also nicht direkt das Archivsystem angesprochen, sondern zuerst geprüft, ob das Objekt bereits in der gewünschten Version lokal vorliegt. Eine kurze Kommunikation mit dem Archivsystem findet lediglich statt, um den Status und die Konsistenz des lokal vorliegenden Objektes zu validieren.

Sichere Übertragungswege

Um die Datensicherheit und den Datenschutz zu gewährleisten, sind

sichere Übertragungswege zwischen dem Langzeitarchivierungssystem und dem zugreifenden System unerlässlich. Zwar kann eine etwaige Manipulation der Daten und Objekte durch die bereits angesprochene Checksummenüberprüfung erkannt werden, jedoch schützt dies nicht vor dem unerlaubten Zugriff Dritter auf die Objekte des Archivsystems. Dies kann sowohl über sogenanntes Abhören der Datenleitung geschehen, als auch dadurch, dass unbefugte Dritte an Zugangsdaten und Netzwerkadressen des Archivsystems gelangen. Hier ist es daher sinnvoll, mit eindeutigen Befugnissen, sicheren Übertragungsprotokollen (wie HTTPS oder SFTP) und idealerweise Signaturschlüsseln und restriktiven IP-Freigaben zu arbeiten.

Datenübernahme in ein neues Archivsystem

Ein digitales Langzeitarchivsystem sollte die Möglichkeit bieten, alle Objekte zum Zwecke einer Migration auf ein neues oder anderes Archivsystem als Gesamtpaket oder als einzelne Objekte abzurufen. Verbunden mit dem einzelnen Objekt oder dem Gesamtpaket sollten auch alle gesammelten Metadaten sein. Sie sollten nach Möglichkeit komplett in das neue Archivsystem übernommen werden.

Diese Punkte sollten bei der Planung und Umsetzung von Zugriffsstrategien auf ein Archivsystem beachtet und mit einbezogen werden. Für individuelle Lösungen werden sicherlich auch noch weitere Faktoren eine Rolle spielen. Die jeweiligen Implementierungen sind natürlich auch stark von dem verwendeten Archivsystem abhängig.