Digitale Archivierung an der Universitätsbibliothek Bern

Policy Speicherinfrastruktur

b
UNIVERSITÄT
BERN

Dokument erstellt: September 2017
Dokument überprüft: 31. Juli 2019
Dokument erstellt von: Remo Lehmann
Version: 1.0
Status: veröffentlicht
Regelmässige Überprüfung: Alle zwei Jahre

Kontakt: berda@ub.unibe.ch

Inhaltsverzeichnis

Ziel

Die vorliegende Policy beschreibt, wie der Archivspeicher des *Bern Digital Archive (BerDA)* aufgebaut ist, durch wen er gepflegt sowie unterhalten wird und welche Prozeduren eingerichtet wurden, um die Sicherheit seiner Inhalte zu gewährleisten. Das Ziel der in diesem Dokument beschriebenen strukturellen und funktionalen Massnahmen ist es, das Prinzip der „authentischen und integren Aufbewahrung der Objekte in der Obhut von *BerDA*"[1] zu garantieren, welches in der Hauptpolicy des digitalen Archivs aufgelistet ist. Dies beinhaltet:

- Schutz der einzelnen *bitstreams* vor Veränderungen und/oder Verlust durch Hardwaredefekte
- Schutz der einzelnen *bitstreams* vor Verlust durch katastrophale Ereignisse
- Schutz der einzelnen *bitstreams* vor Verlust und/oder Veränderungen durch unbeabsichtigte oder böswillige Manipulationen

Grundsatz

Die vorliegende Policy gilt für sämtliche Daten, die sich in der Obhut von *BerDA* befinden und auf der von ihm verwendeten Speicherinfrastruktur gespeichert sind. D.h. für sämtliche Daten werden eine einheitliche Speicherinfrastruktur verwendet sowie einheitliche Erhaltungsmassnahmen ergriffen. Von dieser Policy abweichende Bestimmungen für einzelne Bestände sind nicht vorgesehen.

Diese Policy wird regelmässig überarbeitet und an den neuesten technischen Kenntnisstand, dem aktuellen Aufbau der Speicherinfrastruktur sowie an neu entstehende Risiken für Datenverluste angepasst.

Archivspeicherinfrastruktur

Als Archivspeicher wird die Infrastruktur des *Long Term Storage* der Informatikdienste der Universität Bern (ID) verwendet. Betrieb, Monitoring und Unterhalt der Speicherinfrastruktur erfolgen durch die ID. Sämtliche Daten verbleiben stets innerhalb der Domäne der Universität Bern und gelangen nach dem abgeschlossenen Transfer auf die Infrastruktur von *BerDA* nicht ausserhalb dieses Bereiches.

Mediale Redundanz

Alle Daten werden redundant auf einer Kombination aus Festplatten- und Magnetbandlaufwerken gespeichert. Letztere entsprechen dem offenen „*Linear Tape Open"*-Standard der sechsten Generation. Die heterogene Speicherinfrastruktur verhindert, dass sämtliche Speicherkomponenten gleichzeitig das Ende ihrer vorgesehenen Nutzungszeit erreichen. Ausserdem wird das Risiko minimiert, dass ein Grossteil der Speicherhardware aufgrund von allfälligen Konstruktions- bzw. Produktionsfehlern ausfällt. Der Einsatz des „*Linear*

[1] Digitale Langzeitarchivierung an der Universität Bern. Policy und Standards, S. 4.

Tape Open"-Standards minimiert zudem die Abhängigkeit von einem spezifischen Hardwarehersteller.

Geographische Redundanz

Der Inhalt der Magnetbänder wird an zwei verschiedenen, geographisch getrennten Standorten gespiegelt (Sidlerstrasse 5 sowie Fabrikstrasse 8 in Bern), um im Falle höherer Gewalt, wie z. B. Feuer-, Wasserschäden oder ein Stromausfall, einen totalen Datenverlust zu verhindern. Die Festplatteninfrastruktur ist nur an einem Standort vorhanden.

Datenreplikation

An jedem Standort werden neu erstellte Archivinformationspakete nach einer Wartezeit von 6 Stunden auf zwei Magnetbänder dupliziert. Somit sind insgesamt 5 Kopien (Vier auf Magnetbändern verteilt auf 2 Standorte, eine auf Festplatte an einem Standort) von sämtlichen Daten vorhanden. Durch den in der Speicherhardware integrierten Editier- und Löschschutz[2] erübrigt sich das Erstellen von zyklischen, zeitlich versetzten Backups.

Die Daten lassen sich beliebig auf weitere Magnetbänder replizieren, welche auch ausgeworfen und an zusätzlichen Standorten aufbewahrt werden können, sodass eine zukünftige Skalierung der Speicherinfrastruktur technisch jederzeit möglich ist. Sämtliche Archivdaten von *BerDA* werden auf designierten Bändern gespeichert, auf denen sich keine anderen externen Daten befinden. Es besteht somit eine physische Trennung zwischen den Daten von *BerDA* und jenen anderer Nutzer des *Long Term Storage*.

Löschen und ändern von Daten

An allen Standorten ist ein Software-WORM aktiv. Daten, die einmal in den Archivspeicher geschrieben wurden, können weder editiert noch gelöscht werden. Ein Löschen ist nur möglich, wenn der entsprechende Schutz durch die Informatikdienste aufgehoben wird. Eine solche Aufhebung ist lediglich in Ausnahmefällen möglich und muss im Voraus beantragt werden.

Einmal archivierte Daten können nicht mehr editiert werden. Es ist lediglich möglich, eine neue Version einer bereits bestehenden Datei anzulegen, welche die ursprüngliche Datei ersetzt, wobei sämtliche Vorgängerversionen gespeichert bleiben.

Integritätsprüfung und Wiederherstellen von korrupten Daten

Die Integrität der archivierten Daten wird in periodischen Abständen durch zwei unabhängig operierende Mechanismen geprüft. Eine Kontrolle wird durch die eingesetzte Archivsoftware (SHA512-basiert) durchgeführt, welche zweimal jährlich die Unversehrtheit sämtlicher Dateien anhand einer innerhalb der Metadaten gespeicherten Checksumme prüft. Die Checksummen selber sind durch

[2] Vgl. dazu den Abschnitt „Löschen und ändern von Daten".

den Lösch- und Editierschutz vor Manipulationen geschützt. Des Weiteren überprüft auch die Hardware der Speicherinfrastruktur die gespeicherten Inhalte, indem sie bei jeder Lese- und Schreiboperation die gelesenen bzw. geschriebenen Daten einer Integritätsprüfung unterzieht (SHA512-basiert). Die Hardware der Speicherinfrastruktur ist zudem in der Lage, allfällig korrupte Inhalte automatisch durch intakte Versionen von einem anderen Speicherort zu ersetzen.

Datenexport

Sowohl Meta- wie auch Inhaltsdaten lassen sich unabhängig von der im Langzeitarchiv eingesetzten Software wieder exportieren. Daten werden weder komprimiert noch sonstigen Prozeduren unterzogen, die ein direktes Auslesen verhindern. Die Speicherinfrastruktur kann als Netzlaufwerk an autorisierten Computern innerhalb des Netzwerkes der Universität eingebunden werden. Die Speicherhardware unterstützt verschiedene Dateisysteme und Protokolle, wie z. B. CIFS, NFS, SMB, S3, was ein Einbinden in verschiedenen Umgebungen ermöglicht. Alternativ können auch einzelne Magnetbänder ausgeworfen und an jeder Maschine, die über ein entsprechendes Bandlaufwerk verfügt, wieder ausgelesen werden. Somit ist eine Exitstrategie gewährleistet, die weder von einer bestimmten Hardware noch von einer spezifischen Software abhängig ist.

Da die Daten verschiedener Archivkunden im Speicher nicht physisch voneinander getrennt sind, ist ein grösserer Datenexport über einen direkten Zugriff auf den Archivspeicher nur durch das entsprechend berechtigte Archivpersonal möglich.

Migration und Austausch von Speichermedien

Die Informatikdienste der Universität Bern überwachen als Betreiber der Speicherinfrastruktur den Zustand einzelner Datenträger sowie jener der gesamten Speicherinfrastruktur. Datenträger werden laufend ersetzt, bevor diese das Ende ihrer vorgesehenen Nutzungszeit erreichen. Des Weiteren wird die Speicherhardware regelmässig auf den neusten technologischen Stand gebracht.

Zugriff und Autorisierung

Der direkte Zugriff auf die Speicherinfrastruktur bleibt System- bzw. Archivadministratoren vorbehalten. Eine diversifizierte Zugriffskontrolle auf Hardwareebene (wie z. B. die Unterscheidung zwischen Benutzer, Administrator, Editor oder unterschiedliche Zugriffsmöglichkeiten auf verschiedene Kollektionen) ist nicht möglich, sondern wird durch die verwendete Archivsoftware gewährleistet.