

Erhalt der Datenintegrität als Teil der Prozessroutinen

Inhalt

Erhalt der Datenintegrität als Teil der Prozessroutinen

1. Vollständigkeitsprüfung im Erwerbungsprozess
 - 1.1 Archivierung von Beständen der TIB-Erwerbungssteams
 - 1.2 Archivierung von IWF-Filmdigitalisaten
 - 1.3 Übernahme von TIB-Beständen direkt vom Datenproduzenten
2. Pre-Ingest-Analyse
3. Upload der Objekte in das Quellverzeichnis der Submission Application
 - 3.1 Archivierung von Beständen der TIB-Erwerbungssteams
 - 3.2 Archivierung von IWF-Filmdigitalisaten
 - 3.3 Übernahme von TIB-Beständen direkt vom Datenproduzenten
4. Erzeugen von Rosetta-konformen SIPs aus definierten Eingangspaketstrukturen
 - 4.1 Archivierung von Beständen der TIB-Erwerbungssteams
 - 4.2 Archivierung von IWF-Filmdigitalisaten
 - 4.3 Übernahme von TIB-Beständen direkt vom

Erhalt der Datenintegrität als Teil der Prozessroutinen

Die unten stehende Grafik [Integritätsprüfung in der Ablauforganisation](#) zeigt, wann im Verlauf eines Objektlebenszyklus Prüfverfahren zur Wahrung der Datenintegrität und Vollständigkeit durchgeführt werden und, wie sie in die bestehenden Systemprozesse implementiert wurden. Die Nummerierung der Überschriften korrespondiert mit der Nummerierung der Ereignisse in der Grafik.

1. Vollständigkeitsprüfung im Erwerbungsprozess

1.1 Archivierung von Beständen der TIB-Erwerbungssteams

Trifft ein digitales Objekt bei Mitarbeitenden eines Erwerbungssteams ein, führen diese als Teil der Qualitätskontrolle eine Prüfung auf Vollständigkeit durch. Vor dem Aufruf des Objekts durch die Mitarbeitenden wird eine Virenprüfung durchgeführt.

1.2 Archivierung von IWF-Filmdigitalisaten

Trifft eine Charge Digitalisate vom Dienstleister ein, wird die Charge auf Vollständigkeit anhand einer Liste der übergebenen Filme überprüft und die vom Dienstleister erstellten MD5-Prüfsummen werden überprüft. Stimmen die Prüfsummen nicht überein oder ist die Filmliste nicht vollständig, werden die jeweiligen Digitalisate reklamiert.

1.3 Übernahme von TIB-Beständen direkt vom Datenproduzenten

Trifft eine Datenlieferung von einem Datenproduzenten ein, werden die mitgelieferten MD5-Prüfsummen überprüft, sofern welche übergeben werden. Die Prüfung auf Vollständigkeit erfolgt in Abhängigkeit vom Datenproduzenten entweder auf Datei-, Artikel- oder Zeitschriftentitelebene, entweder über eine vom Datenproduzenten erstellte Referenzliste, oder über eine vom Team Langzeitarchivierung erstellte Liste abzuholender Artikel und deren Dateien.

2. Pre-Ingest-Analyse

Die Prüfung auf Vollständigkeit ist Teil der Pre-Ingest-Analyse.

3. Upload der Objekte in das Quellverzeichnis der Submission Application

3.1 Archivierung von Beständen der TIB-Erwerbungssteams

In Rücksprache mit dem zuständigen Erwerbungssteams übernimmt das Team Langzeitarchivierung die Objekte aus den definierten Transferverzeichnissen und kopiert die Objekte mit einem WinSCP-Client über eine SSH-verschlüsselte SFTP-Verbindung auf einen Server in das Quellverzeichnis der Submission Application.

3.2 Archivierung von IWF-Filmdigitalisaten

Für die Qualitätskontrolle werden die Digitalisate mit einem WinSCP-Client über eine SSH-verschlüsselte SFTP-Verbindung in ein designiertes Verzeichnis hochgeladen. Nach erfolgreich durchlaufener Qualitätskontrolle werden die Digitalisate in ein Quellverzeichnis für das CSV-Script kopiert.

3.3 Übernahme von TIB-Beständen direkt vom Datenproduzenten

Die Dateien werden für die Qualitätskontrolle über eine SSH-verschlüsselte SFTP-Verbindung in ein designiertes Verzeichnis hochgeladen. Dieser Schritt wird entweder vom Team Langzeitarchivierung mittels WinSCP-Client oder vom Datenproduzenten selbst ausgeführt.

Nach erfolgreich durchlaufener Qualitätskontrolle werden die Daten in ein Quellverzeichnis für das CSV-Script kopiert.

Der SFTP-Standard beinhaltet interne Mechanismen zur Überprüfung gegen Integritätsverletzung während der Übertragung. Der WinSCP-Client behält die ursprünglichen Datums- und Zeitstempel der Dateien bei.

Datenproduzenten
4.4
Submission Application für DSpace-Repositories
5. Deposit: Abgabe der Objekte in Rosetta
6. Transformation des SIP zu einem AIP
7.1 AIP-Update und Integritätsprüfung als Prozess
7.2
Integritätssicherungsmechanismen des Archivspeichers
8. Export
Integritätsprüfung in der Ablauforganisation

Weiterführende Informationen

[Archivspeicher](#)

[Spezifikation für Ablieferungspakete \(SIP\)](#)

[Technische Metadaten](#)

[Strukturelle Metadaten](#)

4. Erzeugen von Rosetta-konformen SIPs aus definierten Eingangspaketstrukturen

Die Transformation der Transferpakete zu Rosetta-konformen SIPs und AIP ist in der [Grafik Transformation von Eingangspaketstrukturen zu SIPs und AIPs](#) und den Ingest-Prozessdiagrammen „[METS-Deposit](#)“ und „[CSV-Deposit](#)“ beschrieben.

4.1 Archivierung von Beständen der TIB-Erwerbungsteams

Die Submission Application wurde mit dem von Ex Libris zur Verfügung gestellten [SDK](#) entwickelt und wandelt die verschiedenen Datenstrukturen der Erwerbungsteams (Eingangspaketstrukturen) zu Rosetta-konformen SIPs um. Die Submission Application erzeugt die Rosetta-konformen SIPs aus Eingangspaketstrukturen und übergibt diese in einem zweiten Schritt an Rosetta.

Die Submission Application erzeugt bei der Erstellung der Transferpakete eine MD5-Prüfsumme pro Datei und speichert diese in einer METS-Datei. Die Namen aller zu dem SIP gehörenden Dateien werden in der structMap der METS-Datei erfasst. Die erzeugten Rosetta-konformen SIPs werden mit dem Anstoßen des Deposit-Prozesses an Rosetta übergeben.

4.2 Archivierung von IWF-Filmdigitalisaten

Abgelieferte MD5-Prüfsummen werden bei der technischen Qualitätskontrolle überprüft.

Das [CSV-Script](#) schreibt, sofern vorhanden, mitgelieferte MD5-Prüfsummen zu jeder Datei in die CSV-Datei. Die Namen aller zu dem SIP gehörenden Dateien werden in der CSV-Datei erfasst. Die erzeugten Rosetta-konformen SIPs werden mit dem Anstoßen des Deposit-Prozesses an Rosetta übergeben.

4.3 Übernahme von TIB-Beständen direkt vom Datenproduzenten

Abgelieferte MD5-Prüfsummen werden, sofern vorhanden, bei der Eingangskontrolle überprüft.

Das [CSV-Script](#) schreibt, sofern vorhanden, mitgelieferte MD5-Prüfsummen zu jeder Datei in die CSV-Datei. Die Namen aller zu dem SIP gehörenden Dateien werden in der CSV-Datei erfasst. Die erzeugten Rosetta-konformen SIPs werden mit dem Anstoßen des Deposit-Prozesses an Rosetta übergeben.

4.4 Submission Application für DSpace-Repositories

Für die Anbindung des [Institutionellen Repositoriums der Leibniz Universität Hannover](#) an das digitale Langzeitarchiv über eine OAI-Schnittstelle nutzt die TIB die [Submission Application der ZBW](#) nach. Die Konfiguration der Submission Application übernimmt die TIB selbst.

Für jede bekannte OAI-Record-ID vergibt und aktualisiert die Submission Application den Bearbeitungsstatus während der Abholung vom Repository im Workflow (z.B. „New“, „Imported“, „Ingested“).

Der Deposit erfolgt über den Deposit Webservice. Die Submission Application trackt via SIP-Status Webservice den Status der an Rosetta übermittelten SIPs und aktualisiert so den Bearbeitungsstatus im Workflow. Die für den Record vergebene SIP-ID wird in eine interne Datenbank der Submission Application geschrieben.

Die Submission Application erzeugt bei der Erstellung der Transferpakete eine MD5-Prüfsumme pro Datei und speichert diese in einer METS-Datei. Die Namen aller zu dem SIP gehörenden Dateien werden in der structMap der METS-Datei erfasst. Die erzeugten rosetta-konformen SIPs werden mit dem Anstoßen des Deposit-Prozesses an Rosetta übergeben.

5. Deposit: Abgabe der Objekte in Rosetta

Wird der Deposit angestoßen, durchläuft das rosetta-konforme SIP den Validation Stack. Im Validation Stack erfolgen:

- Formatidentifikation mit DROID
- Formatvalidierung mit JHOVE oder veraPDF
- Erstellen von drei Checksummen
- Gegenprüfen mitgelieferter MD5-Prüfsummen bei METS- und CSV-Deposit
- Viruscheck
- Extraktion technischer Metadaten mit JHOVE, mediainfo, dem NLNZ Metadata Extraction Tool, exiftool oder veraPDF
- Validierung der METS-Datei (nur bei gewähltem METS-Deposit)

Im Deposit-Prozess wird zur Prüfung der Vollständigkeit überprüft:

- die StructMap-Sektion der von einer Submission Application erzeugten METS-Datei
- die Dateilaufistung der vom CSV_Script erzeugten CSV-Datei

6. Transformation des SIP zu einem AIP

Während des Transformationsprozesses wird das SIP mit zusätzlichen Metadaten angereichert und in verschiedene Speicherbereiche verschoben. Für jedes SIP erzeugt das Langzeitarchivierungssystem eine METS-Datei. Bei jedem Transfer werden drei Prüfsummen neu gebildet und mit den in der METS-Datei gespeicherten abgeglichen, zusätzlich wird mittels StructMap die Vollständigkeit überprüft.

7.1 AIP-Update und Integritätsprüfung als Prozess

Bei jedem AIP-Update wird eine IE vom permanenten in den operativen Speicher kopiert. Bei jedem Transfer werden die Prüfsummen neu gebildet und mit den gespeicherten abgeglichen, zusätzlich wird mittels StructMap die Vollständigkeit überprüft. Die Integritätsprüfung kann darüber hinaus unabhängig von einem Transfer als Rosetta-interner Prozess angestoßen werden.

7.2 Integritätssicherungsmechanismen des Archivspeichers

Die Mechanismen zur Integritätssicherung des Archivspeichers sind unter [Archivspeicher](#) beschrieben.

8. Export

Pro Datei sind in Rosetta drei Prüfsummen gespeichert, beim Export werden die Prüfsummen erneut gebildet und abgeglichen. Alle zur Repräsentation gehörenden Dateien sind in der ie.xml erfasst. So wird sichergestellt, dass die Datenintegrität beim Export gesichert ist und die Daten vollständig sind.

Kommt es beim Export zu einem Fehler, wird der Prozess abgebrochen und das System gibt eine entsprechende Fehlermeldung aus.

Integritätsprüfung in der Ablauforganisation

Zum Vergrößern bitte anklicken

