

Spezifikation für Ablieferungspakete (SIP)

Inhalt

Spezifikation der Transferpakete

Eingangspaketstrukturen

Objekte mit einfacher Struktur und einer Repräsentation

Beispiele Team Deutsche Forschungsbereiche

Objekte mit mehreren Repräsentationen oder komplexe Dateiablagen
Objekte mit mehreren Repräsentationen oder komplexe Dateiablagen und außerhalb erstellter METS-Datei
Repositoryanbindung über eine OAI-Schnittstelle am Beispiel des Institutionellen Repositoriums der Leibniz Universität Hannover
Objekte mit Metadaten aus Quellsystemen oder komplexen Zusammenhängen zwischen Datenpaketen

Pre-Ingest-SIP
Post-Ingest-SIP

Spezifikation der Transferpakete

Die TIB setzt mehrere Stufen von Transferpaketen ein, die an dieser Stelle beschrieben werden. Die Grafik [Paketstrukturen](#) gibt einen groben Überblick über die eingesetzten Transferpakete:

- [Eingangspaketstruktur](#)
- [Pre-Ingest-SIP](#)
- [Post-Ingest-SIP](#)

Relevant für die Ablieferung von Objekten ist die Spezifikation von Eingangspaketstrukturen.

Die Grafik [Transformation von Eingangspaketstrukturen zu SIPs und AIPs](#) beschreibt die Transformation von Paketeingangsstrukturen in Pre-Ingest-SIP, Post-Ingest-SIP und AIP.

Eingangspaketstrukturen

Für unterschiedliche Szenarien werden verschiedene Eingangsstrukturen genutzt:

1. [SIP mit einfacher Struktur und einer Repräsentation](#)
SIPs mit einfachen Strukturen können wahlweise als ZIP oder als Folder abgegeben werden.
2. METS-Deposit
 - a. [für Objekte mit mehreren Repräsentationen oder komplexe Dateiablagen](#)
 - b. [Objekte mit mehreren Repräsentationen oder komplexe Dateiablagen und außerhalb erstellter METS-Datei](#)
3. [Repositoryanbindung via OAI oder eine andere Schnittstelle](#)
4. CSV-Deposit
für Objekte mit komplexen Zusammenhängen, die als Collection abgebildet werden sollen, oder mit Metadaten aus Quellsystemen

Die Eingangspaketstrukturen werden in Form von normierten Tabellen beschrieben. Die folgende Tabelle erläutert den Aufbau der normierten Tabellen.

Spezifikationsparameter	Umsetzung
Namenskonvention	Namenskonvention, nach der das Paket benannt sein muss
Paketstruktur	Struktur, in der das Paket vorliegen muss
Inhaltsdaten	Beschreibung der mindestens und höchstens erwarteten Anzahl von Dateien
Zulässige Dateiformate	eine Beschreibung der zulässigen Dateiformate, wenn vorhanden
Repräsentationen	zulässige Anzahl und Art von Repräsentationen
Datenqualität	Beschreibung, ob nur valide und wohlgeformte Dateien angenommen werden
Metadaten	Beschreibung, ob das Objekt im Gemeinsamen Verbundkatalog nachgewiesen sein muss, oder nicht
Identifizier	Identifizier, der das Objekt eindeutig identifiziert und mit deskriptiven Metadaten verknüpft, zum Beispiel eine PPN (Identifikationsnummer), eine EKI (Identifikationsnummer der ersterfassenden Institution) oder ein Handle
Rechtliche Metadaten	Beschreibt, ob das Objekt zu einem Bestand gehört, in dem mehrere Lizenztexte, Lizenztextversionen und Nutzungsrechte vergeben werden können. Wenn ja, muss dies in einer übergeordneten Verzeichnisstruktur abgebildet werden.

Objekte mit einfacher Struktur und einer Repräsentation

Beispiel Team Deutsche Forschungsberichte

Bei dieser Eingangsstruktur darf genau eine Datei vorhanden sein, die zur Repräsentation MASTER gehört.

Weiterführende Informationen

Paketstrukturen

Transformation von
Eingangspaketstrukturen zu SIPs und
AIPs

Spezifikationsparameter	Umsetzung
Namenskonvention	Die Datei ist mit der PPN benannt.
Paketstruktur	eine PDF-Datei pro SIP
Inhaltsdaten	genau eine PDF-Datei
Zulässige Dateiformate	genau eine PDF-Datei
Repräsentationen	MASTER
Datenqualität	Es werden auch nicht-valide und nicht-wohlgeformte Dateien akzeptiert.
Metadaten	Das Objekt muss im Katalog nachgewiesen sein.
Identifizier	Es wird im Namen der PDF-Datei ein Identifizier erwartet, der auf ein Katalogisat verweist.
Rechtliche Metadaten	Die Objekte werden vor dem Ingest nach Lizenzvereinbarung und Nutzungsrecht sortiert.

Objekte mit mehreren Repräsentationen oder komplexe Dateiablagen

In der Eingangsstruktur für komplexe Objekte dürfen verschiedene Repräsentationen mit jeweils 1-n Dateien eingestudert werden.

Spezifikationsparameter	Umsetzung
Namenskonvention	Das Paket ist auf der obersten Verzeichnisebene mit der EKI benannt.
Paketstruktur	<p>Pro SIP gibt es ein Verzeichnis, das mit der EKI benannt ist. Darin ist für jede Repräsentation ein Verzeichnis enthalten, das nach dem im Archiv definierten Namensvokabular benannt ist. In den Repräsentationsordnern befinden sich die Inhaltsdaten.</p> <pre>IDENTIFIZIER --MASTER (Pflicht) --File1 --File n -- Folder 0-n --File 0-m --MODIFIED_MASTER (optional) --File1 --File n -- Folder 0-n --File 0-m --DERIVATIVE_COPY (optional) --File1 --File n -- Folder 0-n --File 0-m</pre>
Inhaltsdaten	mindestens eine Datei pro Repräsentation
Zulässige Dateiformate	keine Beschränkung
Repräsentationen	MASTER, MODIFIED_MASTER, DERIVATIVE_COPY
Datenqualität	Es werden auch nicht-valide und nicht-wohlgeformte Dateien akzeptiert.
Metadaten	Das Objekt muss im Katalog nachgewiesen sein.

Identifizier	EKI
Rechtliche Metadaten	Das Erwerbungssteam ordnet mittels einer übergeordneten Verzeichnisstruktur die Objekte den Bestandsgruppen, Publikationsarten, geltenden Lizenztexten (ggf. in verschiedenen Versionen) und Zugriffsrechten zu.

Objekte mit mehreren Repräsentationen oder komplexe Dateiablagen und außerhalb erstellter METS-Datei

In der Eingangsstruktur für komplexe Objekte mit außerhalb erstellter METS-Datei dürfen im Verzeichnispfad `identifizier/content/streams` verschiedene Repräsentationen mit jeweils 1-n Dateien übergeben werden. Es muss eine gegen die Rosetta-XSD (<https://developers.exlibrisgroup.com/rosetta/integrations/mets-dnx/>) valide METS-Datei mit übergeben werden.

Spezifikationsparameter	Umsetzung
Namenskonvention	Das Paket ist auf der obersten Verzeichnisebene mit einem eindeutigen Identifizier benannt.
Paketstruktur	<p>Pro SIP gibt es ein Verzeichnis, das mit einem eindeutigen Identifizier benannt ist. Darin sind eine <code>dc.xml</code> mit Dublin Core-Metadaten und das <code>content</code>-Verzeichnis enthalten. Das <code>content</code>-Verzeichnis enthält die METS-Datei <code>ie1.xml</code> und das <code>streams</code>-Verzeichnis, das die Inhaltsdateien enthält. Werden mehrere Repräsentationen übergeben, müssen die dazugehörigen Files in der METS-Datei den Repräsentationen zugewiesen werden.</p> <p>Repräsentationsnamen abgesehen von MASTER, PRE-INGEST_MODIFIED_MASTER und DERIVATIVE_COPY müssen mit der TIB abgestimmt werden.</p> <pre> IDENTIFIZIER --dc.xml (Pflicht) --content --ie1.xml --streams --MASTER (Pflicht) --File1 --File n -- Folder 0-n --File 0-m --PRE-INGEST_MODIFIED_MASTER (optional) --File1 --File n -- Folder 0-n --File 0-m --DERIVATIVE_COPY (optional) --File1 --File n -- Folder 0-n --File 0-m </pre>
Inhaltsdaten	mindestens eine Datei pro Repräsentation
Zulässige Dateiformate	keine Beschränkung
Repräsentationen	MASTER (Pflicht), PRE-INGEST_MODIFIED_MASTER (optional), DERIVATIVE_COPY(optional), weitere nach Absprache
Datenqualität	Es werden auch nicht-valide und nicht-wohlgeformte Dateien akzeptiert.
Metadaten	Das Objekt muss nicht im Katalog nachgewiesen sein. Es muss eine <code>dc.xml</code> vorhanden sein.
Identifizier	Eindeutiger Identifizier

Rechtliche Metadaten	<p>Rechtliche Metadaten werden in der METS-Datei wie folgt erfasst:</p> <p>1) Das Zugriffsrecht auf das Dokument, so wie für die Nutzung von der abgebenden Stelle erteilt, soll über dcterms:accessRight (z.B. als private/public oder mit einem anderen kontrollierten Vokabular) dokumentiert werden.</p> <p>(2) Das Archivierungsrecht der abgebenden Stelle wird als dc:rights dokumentiert .</p> <p>(3) Die Übernahmevereinbarung, wie zwischen dem Team LZA der TIB und abgebender Stelle abgeschlossen. Dies erfolgt über dcterms: license.</p>
-----------------------------	---

Repositoryanbindung über eine OAI-Schnittstelle am Beispiel des Institutionellen Repositoriums der Leibniz Universität Hannover

Bei dieser Eingangspaketstruktur werden über die OAI-Schnittstelle des Institutionellen Repositoriums der Leibniz Universität Hannover Records eingegestet, die Metadaten, mindestens Titel und Identifier, und 1-n Dateien enthalten müssen. Alle Objekte innerhalb eines Records gehören zur Repräsentation MASTER.

Spezifikationsparameter	Umsetzung
Namenskonvention	Der Originaldateiname der Datei wird übernommen. Es findet keine Prüfung gegen eine Spezifikation statt.
Paketstruktur	mindestens eine Datei pro SIP
Inhaltsdaten	<p>mindestens eine Datei.</p> <p>In mindestens einem Metadata Format müssen direkte Links auf die zum Record gehörenden Dateien inklusive Supplements vorhanden sein.</p>
Zulässige Dateiformate	keine Beschränkung
Repräsentationen	MASTER
Datenqualität	Es werden auch nicht-valide und nicht-wohlgeformte Dateien akzeptiert.
Metadaten	Zu dem Objekt müssen mindestens die Metadaten Titel und Identifier auf dem Repository vorhanden sein. Weitere Metadaten können vorhanden sein.
Identifier	Repository-interner Handle
Rechtliche Metadaten	Die geltende Lizenzbedingung muss ausgezeichnet sein.

Objekte mit Metadaten aus Quellsystemen oder komplexen Zusammenhängen zwischen Datenpaketen

Spezifikationsparameter	Umsetzung
Namenskonvention	Das Paket ist auf der obersten Verzeichnisebene mit einem eindeutigen Identifier benannt.

Paketstruktur	<p>Pro SIP gibt es ein Verzeichnis, das mit einem eindeutigen Identifier benannt ist. Darin sind eine dc.xml mit Dublin Core-Metadaten enthalten, optional können eine harvest.xml mit Angaben zur Abholung von einer Datenquelle und eine collection.xml mit Angaben zur Zuordnung eines Objekts zu einer Sammlung enthalten sein.</p> <p>Die Repräsentationsverzeichnisse enthalten die Inhaltsdateien.</p> <p>MD5-Prüfsummen können optional als eine Prüfsummendatei für alle Dateien in allen abgelieferten Identifier-Verzeichnissen auf der Ebene der Identifier-Verzeichnisse oder als eine MD5-Prüfsumme pro Datei in den Repräsentationsverzeichnissen übermittelt werden.</p> <p>Repräsentationsnamen abgesehen von MASTER, PRE_INGEST_MODIFIED_MASTER und DERIVATIVE_COPY müssen mit der TIB abgestimmt werden.</p> <pre> root --[checksums].[md5] (optional eine Checksummendatei für alle Dateien in allen abgelieferten Identifier-Verzeichnissen) --IDENTIFIER --dc.xml (Pflicht) --harvest.xml (optional) --collection.xml (optional) --MASTER (Pflicht) --1-n Files --0-1nFiles.[md5] (optional je eine Checksummendatei pro Datei in der Repräsentation) --0-n Verzeichnisse -- ... --PRE_INGEST_MODIFIED_MASTER (optional) --1-n Files --0-n Files.[md5] (optional je eine Checksummendatei pro Datei in der Repräsentation) --0-n Verzeichnisse -- ... --DERIVATIVE_COPY (optional) --1-n Files --0-1nFiles.[md5] (optional je eine Checksummendatei pro Datei in der Repräsentation) --0-n Verzeichnisse -- ... --SOURCE_MD (optional) --1-n .[xml] --weitere Repräsentationen (optional) --IDENTIFIER </pre>
Inhaltsdaten	mindestens eine Datei pro Repräsentation
Zulässige Dateiformate	keine Beschränkung
Repräsentationen	MASTER (Pflicht), PRE_INGEST_MODIFIED_MASTER (optional), DERIVATIVE_COPY (optional), weitere nach Absprache
Datenqualität	Es werden auch nicht-valide und nicht-wohlgeformte Dateien akzeptiert.
Metadaten	Das Objekt muss nicht im Katalog nachgewiesen sein. Es muss eine dc.xml vorhanden sein.
Identifier	Eindeutiger Identifier

Rechtliche Metadaten	Rechtliche Metadaten werden in der CSV-Datei wie folgt erfasst: 1) Das Zugriffsrecht auf das Dokument, so wie für die Nutzung von der abgebenden Stelle erteilt, soll über dcterms:accessRight (z.B. als private/public oder mit einem anderen kontrollierten Vokabular) dokumentiert werden. (2) Das Archivierungsrecht der abgebenden Stelle wird als dc:rights dokumentiert . (3) Die Übernahmevereinbarung, wie zwischen dem Team LZA der TIB und abgebender Stelle abgeschlossen. Dies erfolgt über dcterms:license. (4) Zugriffsrechte auf das Objekt in Rosetta werden als Access Right erfasst.
-----------------------------	---

Pre-Ingest-SIP

Verschiedene Applikationen erzeugen aus den unterschiedlichen Paketeingangsstrukturen die Rosetta-konformen Pre-Ingest-SIPs. Diese werden in einem zweiten Schritt an Rosetta übergeben.

Post-Ingest-SIP

Nach dem Deposit werden aus den Pre-Ingest-SIPs Post-Ingest-SIPs, die vom System mit weiteren Metadaten angereichert werden. Der Transformationsprozess ist abgeschlossen, wenn ein Paket an den permanenten Archivspeicher übergeben und dort erfolgreich abgelegt wurde.

Während der weiteren Prozessierung in Rosetta wird das Post-Ingest-SIP zum AIP transformiert und automatisch mit weiteren Metadaten angereichert. Ein Post-Ingest-SIP wird zum AIP, wenn es in den permanenten Archivspeicher weitergeleitet und dort erfolgreich gespeichert wurde.