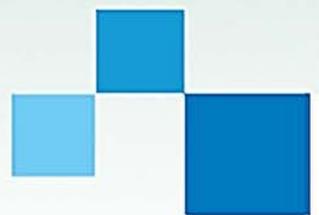


Beispiele zur SIP-Bildung

Handreichung der nestor Arbeitsgruppe
SIP-Konkretisierung



Beispiele zur SIP-Bildung

SIP-Format der
Deutschen Nationalbibliothek

SIP-Format der Technischen
Informationsbibliothek Hannover

Beispiel SLUBArchiv.digital

SIP-Ausprägung der
DIMAG-Ingest-Module

Handreichung der nestor Arbeitsgruppe
SIP-Konkretisierung

nestor - Kompetenznetzwerk Langzeitarchivierung
und Langzeitverfügbarkeit Digitaler Ressourcen
für Deutschland

nestor - Network of Expertise in Long-Term Storage of Digital Resources
<https://www.langzeitarchivierung.de>

nestor-Mitglieder:

- ..■ Bayerische Staatsbibliothek
- ..■ Bundesarchiv
- ..■ Deutsche Kinemathek – Museum für Film und Fernsehen
- ..■ Deutsche Nationalbibliothek
- ..■ Georg-August-Universität Göttingen / Niedersächsische Staats- und Universitätsbibliothek Göttingen
- ..■ hebis Verbundzentrale
- ..■ Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen
- ..■ Humboldt-Universität zu Berlin
- ..■ Institut für Museumsforschung (Stiftung Preußischer Kulturbesitz)
- ..■ Landesarchiv Baden-Württemberg
- ..■ Landesarchiv Nordrhein-Westfalen
- ..■ Leibniz-Institut für Deutsche Sprache
- ..■ Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden
- ..■ Staatsbibliothek zu Berlin (Stiftung Preußischer Kulturbesitz)
- ..■ TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek
- ..■ Universitätsbibliothek Regensburg
- ..■ ZB MED - Informationszentrum Lebenswissenschaften
- ..■ ZBW – Leibniz-Informationszentrum Wirtschaft (ZBW)



© 2023 **nestor** - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit
Digitaler Ressourcen für Deutschland

Der Inhalt dieser Veröffentlichung darf vervielfältigt und verbreitet werden, sofern der Name des
Rechteinhabers "**nestor** - Kompetenznetzwerk Langzeitarchivierung" genannt wird.
Eine kommerzielle Nutzung ist nur mit Zustimmung des Rechteinhabers zulässig.

URN: urn:nbn:de:0008-2023060603
<https://nbn-resolving.org/urn:nbn:de:0008-2023060603>

SIP-Format der Deutschen Nationalbibliothek

1. Ausgangslage

Die Deutsche Nationalbibliothek (DNB) nutzt als digitales Langzeitarchiv das System koala, welches bei der Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG) gepflegt und gehostet wird. Es basiert auf den Funktionalitäten und Schnittstellen des im kopal-Projekt¹ (2004 – 2007) entwickelten Systems DIAS der Firma IBM. Die im Zuge des Projekts entwickelten Spezifikationen für Informationspakete im Sinne des OAIS-Referenzmodells unter dem Namen Universelles Objektformat (UOF) werden auf von koala sowohl für SIPs als auch DIPs umgesetzt.

Das digitale Langzeitarchiv ist dabei Teil einer größeren Infrastruktur. In dieser kommen abgelieferte Publikationen über verschiedene Transferschnittstellen zur DNB und werden erst dort in automatischen Workflows zu SIPs, die in das Langzeitarchiv übernommen werden. Die ursprünglichen Transferpakete bestehen aus digitalen Objekten (Dateien) und bibliografischen Metadaten (entsprechend einem vorgegebenem Metadatenkernset). Im Importprozess werden die Objekte und Metadaten in verschiedenen Speichersystemen (Repository, Katalogsystem, Datenbanken) abgelegt. Von dort und angereichert um Langzeitarchivierungsmetadaten werden SIPs gebildet und ins digitale Langzeitarchiv überführt.

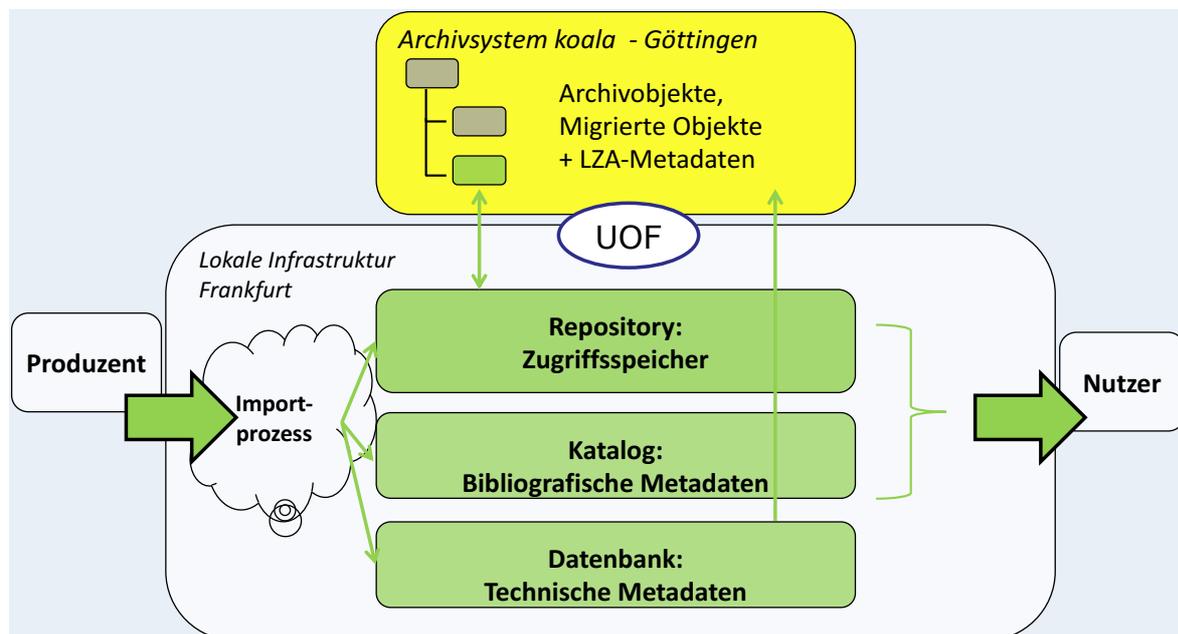


Abbildung 1: Archivsystem in der Infrastruktur der DNB

¹ <http://kopal.langzeitarchivierung.de>

2. Universelles Objektformat

Das Universelle Objektformat (UOF) ist eine Spezifikation² für Informationspakete im Sinne des OAIS-Referenzmodells, also geeignet für SIPs, AIPs und DIPs. Danach ist ein Informationspaket eine gepackte TAR- oder ZIP-Datei, die eine oder mehreren Dateien in beliebigen Ordnerstrukturen und einer METS-Datei im Wurzelverzeichnis enthält. Der Metadata Encoding and Transmission Standard (METS)³ beschreibt ein als XML-Datei kodierte Bündel verschiedener Metadatenformate und -arten zur Beschreibung digitaler Objekte. Für konkrete Anwendungen von METS gibt es METS Profile, die definieren, welche Metadaten in welcher Form für die Anwendung genutzt werden. Für UOF gibt es das METS Profil "Usage of METS 1.4 as part of the Universal Object Format"⁴.

Eine UOF-METS-Datei enthält eine Liste aller Dateien, die zum Objekt gehören, und Langzeitarchivierungsmetadaten im Format LMER⁵. Innerhalb der LMER-Daten können zudem spezifische technische Metadaten zu den Dateien integriert werden, welche etwa als Ausgabe von Formaterkennungs- und -validierungstools entstehen. Die Langzeitarchivierungsmetadaten für elektronische Ressourcen (LMER) wurden 2003 entwickelt, also vor dem ähnlichen, aber deutlich umfangreicheren PREMIS. Das LMER-Datenmodell kennt Objects, Files, Processes und Modifications, wozu es jeweils Metadatenelemente gibt. Ein Object kann aus einem oder mehreren Files bestehen. Processes sind Veränderungen wie z. B. Migrationen, die auf Dateiebene (File) oder Objektebene (Object) erfolgen können. Modifications sind Veränderungen an den Metadaten selbst.

Die UOF-Spezifikationen beschreiben, wie LMER in METS verwendet wird und welche Elemente davon genutzt oder nicht genutzt werden sollen. Darüber hinaus können bibliografische Metadaten in beliebigen Formaten in der METS-Datei hinterlegt werden. Die Spezifikationen führen auch gesondert spezielle Anforderungen bzw. Einschränkungen der Implementierung des Langzeitarchivsystems DIAS von IBM auf, die auch (weitgehend) für das aktuell von der DNB verwendete System koala⁶ gelten.

2.1. Aufbau einer UOF-METS-Datei

Die METS-Datei für ein SIP gemäß den UOF-Spezifikationen hat grundsätzlich eine Liste aller zum Objekt gehörigen Dateien in der FileSection von METS. Diese werden außerdem in einer StructMap mit Namen „ASSET“ als flache Hierarchie von Verweisen zur FileSection aufgelistet. In der AmdSection finden sich ein TechMD-Abschnitt mit Metadaten im Format LMER-Object, die von der FileGrp „ASSET“ referenziert werden. Des Weiteren gibt es je einen TechMD-Abschnitt für jede Datei, welcher Metadaten im Format LMER-File enthält und vom jeweiligen File-METS-Eintrag referenziert wird. In LMER-File gibt es ein generisches Unterelement xmlData, was für spezifische technische Metadaten gedacht ist. Bei der DNB wird dort die Ausgabe des verwendeten Meta-Tools FITS⁷ hinterlegt.

² http://kopal.langzeitarchivierung.de/downloads/kopal_Universelles_Objektformat.pdf

³ <https://www.loc.gov/standards/mets/>

⁴ <https://www.loc.gov/standards/mets/profiles/00000011.html>

⁵ <http://d-nb.info/97511087x/34>

⁶ <http://koala-docs.gwdg.de/de/>

⁷ <https://projects.iq.harvard.edu/fits/home>

Bei Migrationen gibt es zusätzlich DigiprovMD-Abschnitte in der AmdSection, die LMER-Process-Metadaten enthalten, welche die durchgeführten Objektänderungen beschreiben. Je nachdem, ob diese auf Dateiebene oder auf Objektebene erfolgten, sind diese entweder vom File-Element oder vom FlieGrp-Element referenziert.

2.2. Beispiel

Dies ein einfaches Beispiel aus den Erzeugungssystemen der DNB, welches ein Objekt mit nur einer einzigen PDF-Datei beschreibt. Die Inhaltsdatei

Diskurs_2004_3_Singer_De_Haan_Zusammen_machen.pdf liegt in einer Unterverzeichnisstruktur *content/pdf/*, was allerdings nicht zwingend für alle SIPs so sein muss. Entscheidend ist nur, dass die beschreibende METS-XML-Datei immer im Wurzelverzeichnis des Objekts liegt.

```
<mets xmlns="http://www.loc.gov/METS/" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://
www.w3.org/2001/XMLSchema-instance" xmlns:lmerFile="http://www.ddb.de/LMERfile"
xmlns:lmerObject="http://www.ddb.de/LMERObject" xmlns:lmerProcess="http://www.ddb.de/LMERprocess"
xmlns:dc="http://purl.org/dc/elements/1.1/" OBJID="urn:diasid:ast:kop01:0201511120251212690000"
TYPE="kopal Submission Information Package" PROFILE="DNB" xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd http://www.ddb.de/LMERfile http://files.dnb.de/
standards/lmer/lmer-file.xsd http://www.ddb.de/LMERObject http://files.dnb.de/standards/lmer/lmer-
object.xsd http://www.ddb.de/LMERprocess http://files.dnb.de/standards/lmer/lmer-process.xsd
http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2003/04/02/dc.xsd">
<metsHdr CREATEDATE="2021-02-11T17:24:47.940817" RECORDSTATUS="TEST">
  <agent ROLE="ARCHIVIST" TYPE="ORGANIZATION">
    <name>DNB</name>
    <note>Automatisch generierte Metadaten. F&#252;r weitere Informationen:
      kopal.langzeitarchivierung.de</note>
  </agent>
</metsHdr>
<amdSec ID="AmdSec-0001">
  <techMD ID="TechMD-LMER-Object">
    <mdWrap ID="TechMD-LMER-Object-MdWrap" MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="lmerObject"
      LABEL="LMERObject">
      <xmlData>
        <lmerObject:persistentIdentifier>urn:nbn:de:0111-opus-60015</lmerObject:persistentIdentifier>
        <lmerObject:transferChecksum CHECKSUMTYPE="xor of sha1 file
          checksums">b2d6a198d70f35c697d3a0a381b81fd777de4dff</lmerObject:transferChecksum>
        <lmerObject:masterCreationDate>2015-11-12T14:51:09.219+01:00</lmerObject:masterCreationDate>
        <lmerObject:metadataCreationDate>2015-11-12T14:51:09.219+01:00</lmerObject:metadataCreationDate>
        <lmerObject:metadataRecordCreator>KOPAL DIAS</lmerObject:metadataRecordCreator>
        <lmerObject:numberOfFiles>1</lmerObject:numberOfFiles>
        <lmerObject:objectIdentifier>urn:diasid:ast:kop01:0201511120251212690000</
          lmerObject:objectIdentifier></xmlData>
      </mdWrap>
    </techMD>
  <techMD ID="TechMD-File--0">
    <mdWrap ID="TechMD-File--0-MDWRAP" MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="lmerFile">
      <xmlData>
        <lmerFile:format REGISTRYNAME="DIAS">urn:diasid:fty:kopal:0000000000000000000000</
          lmerFile:format>
        <lmerFile:category>binary</lmerFile:category>
        <lmerFile:xmlData MDTYPE="FITS">
          <fits xmlns="http://hul.harvard.edu/ais/xml/ns/fits/fits_output" xsi:schemaLocation="http://
            hul.harvard.edu/ais/xml/ns/fits/fits_output http://files.dnb.de/standards/fits/
            fits_output_dnb.xsd" version="0.6.1" timestamp="11.11.15 13:45">
            <identification status="SINGLE_RESULT">
              <identity format="Portable Document Format" mimetype="application/pdf" toolname="FITS"
                toolversion="0.6.1">
                <tool toolname="Droid" toolversion="3.0"/>
                <tool toolname="Jhove" toolversion="1.11"/>
                <tool toolname="Exiftool" toolversion="7.74"/>
                <tool toolname="ffident" toolversion="0.2"/>
                <tool toolname="file utility" toolversion="5.17"/>
                <tool toolname="NLNZ Metadata Extractor" toolversion="3.4GA"/>
                <version toolname="Jhove" toolversion="1.11">1.7</version>
                <externalIdentifier toolname="Droid" toolversion="3.0" type="puid">fmt/144</
                  externalIdentifier>
                <externalIdentifier toolname="Droid" toolversion="3.0" type="puid">fmt/145</
```

```

        externalIdentifier>
        <externalIdentifier toolname="Droid" toolversion="3.0" type="puid">fmt/146</
        externalIdentifier>
        <externalIdentifier toolname="Droid" toolversion="3.0" type="puid">fmt/147</
        externalIdentifier>
        <externalIdentifier toolname="Droid" toolversion="3.0" type="puid">fmt/148</
        externalIdentifier>
        <externalIdentifier toolname="Droid" toolversion="3.0" type="puid">fmt/157</
        externalIdentifier>
        <externalIdentifier toolname="Droid" toolversion="3.0" type="puid">fmt/158</
        externalIdentifier>
    </identity>
</identification>
<fileinfo>
    <size toolname="Jhove" toolversion="1.11">963548</size>
    <creatingApplicationName toolname="Jhove" toolversion="1.11">Adobe PDF Library 7.0/
    AccuSoft ImageGear Library 15.2.1</creatingApplicationName>
    <lastmodified toolname="Exiftool" toolversion="7.74" status="SINGLE_RESULT">2015:11:11
    13:45:11+01:00</lastmodified>
    <created toolname="Exiftool" toolversion="7.74" status="SINGLE_RESULT">2012:07:23
    17:32:47+02:00</created>
    <filepath toolname="OIS File Information" toolversion="0.1" status="SINGLE_RESULT">/data/
    var/tmp/tomcatP-6/./impservice/Diskurs_2004_3_Singer_De_Haas_Zusammen_machen.pdf</
    filepath>
    <filename toolname="OIS File Information" toolversion="0.1" status="SINGLE_RESULT">./
    impservice/Diskurs_2004_3_Singer_De_Haas_Zusammen_machen.pdf</filename>
    <md5checksum toolname="OIS File Information" toolversion="0.1"
    status="SINGLE_RESULT">a29f28dd6e3748c7f97b782ddd5f180</md5checksum>
    <fslastmodified toolname="OIS File Information" toolversion="0.1"
    status="SINGLE_RESULT">1447245911000</fslastmodified>
</fileinfo>
<filestatus>
    <well-formed toolname="Jhove" toolversion="1.11" status="SINGLE_RESULT">true</well-formed>
    <valid toolname="Jhove" toolversion="1.11" status="SINGLE_RESULT">true</valid>
    <message toolname="Jhove" toolversion="1.11" status="SINGLE_RESULT">Page information is
    not displayed; to display remove param value of p from the config file</message>
    <ingestLevel toolname="Didigo" toolversion="0.9" status="SINGLE_RESULT">4</ingestLevel>
</filestatus>
<metadata>
    <document>
        <pageCount toolname="Exiftool" toolversion="7.74" status="SINGLE_RESULT">9</pageCount>
        <graphicsCount toolname="Jhove" toolversion="1.11" status="SINGLE_RESULT">9</
        graphicsCount>
        <isTagged toolname="Jhove" toolversion="1.11">no</isTagged>
        <hasOutline toolname="Jhove" toolversion="1.11">no</hasOutline>
        <hasAnnotations toolname="Jhove" toolversion="1.11" status="SINGLE_RESULT">no</
        hasAnnotations>
        <isRightsManaged toolname="Exiftool" toolversion="7.74" status="SINGLE_RESULT">no</
        isRightsManaged>
        <isProtected toolname="Exiftool" toolversion="7.74">no</isProtected>
        <isPdfA toolname="Exiftool" toolversion="7.74" status="SINGLE_RESULT">yes</isPdfA>
        <pdfAVersion toolname="Exiftool" toolversion="7.74" status="SINGLE_RESULT">1b</
        pdfAVersion>
        <hasForms toolname="NLNZ Metadata Extractor" toolversion="3.4GA"
        status="SINGLE_RESULT">yes</hasForms>
    </document>
</metadata>
</fits>
</lmerFile:xmlData>
</xmlData>
</mdWrap>
</techMD>
</amdSec>
<fileSec>
    <fileGrp ID="ASSET" ADMID="TechMD-LMER-Object">
        <file ID="FILE-0" ADMID="TechMD-File--0" CREATED="2015-11-12T14:51:09.180+01:00" SIZE="963548"
        CHECKSUM="b2d6a198d70f35c697d3a0a381b81fd777de4dff" CHECKSUMTYPE="SHA-1" MIMETYPE="application/
        pdf">
            <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="file://./content/pdf/
            Diskurs_2004_3_Singer_De_Haas_Zusammen_machen.pdf"/>
        </file>
    </fileGrp>
</fileSec>
</structMap TYPE="ASSET">

```

```

<div ORDER="1" LABEL="File list" TYPE="ASSET">
  <fptr FILEID="FILE-0"/>
</div>
</structMap>
</mets>

```

Die `structMap` `TYPE="ASSET"` enthält einen Verweis auf `file ID="FILE-0"` in der `fileSec`, welche die Datei mit Verzeichnis benennt. Darin wird wiederum auf zwei Metadatenabschnitte in `amdSec ID="AmdSec-0001"` verwiesen: In `fileGrp ID="ASSET" ADMID="TechMD-LMER-Object"` werden die Metadaten im Format LMER-Object referenziert und in `file ADMID="TechMD-File--0"` die Metadaten im Format LMER-File. Den größten Teil der Datei nehmen im Element `lmerFile:xmlData MDTYPE="FITS"` des Abschnitts `"TechMD-File--0"` die FITS-Daten mit den Tool-Ausgaben zu Identifikation, Validierung und spezifischen PDF-Metadaten ein.

2.3. Nutzung im Archivsystem der DNB

Die METS-Datei, die mit dem SIP kommt, wird im Data Management des OAIS-konformen Archivsystem koala der DNB abgelegt. Beim Ingest werden die Metadaten aus der Datei in abfragbarer Form gespeichert. Dazu gehören auch die spezifischen technischen Metadaten, die direkt als Ausgabe des Tools FITS übernommen werden. Über ein maschinelles Query-Interface kann jedes XML-Element daraus gezielt für Abfragen genutzt werden. Dabei kann als DIP ein komplettes Paket wie das SIP exportiert werden inklusive der METS-Datei.

Diese Ablage und Funktionalitäten sind vor allem für die Unterstützung von automatisierter Dateiformatmigration gedacht. Über die Abfrageschnittstelle können Migrationsprozesse gezielt Objekte nach Eigenschaften in den technischen Metadaten aus dem System holen, um diese mit geeigneten Konvertierungstools umzuwandeln. Die geänderten Objekte werden dann mit aktualisierten METS-Dateien, welche um Metadaten im Format LMER-Process ergänzt werden, die die Änderungen dokumentieren, erneut als SIPs in das System eingespielt. Das koala-System kann zu jeder Objekt-URN beliebig viele Objekte als Migrationen davon verwalten.

2.4. Tools zur Erzeugung der SIPs

Das UOF wurde ursprünglich im Projekt kopal entwickelt. In dem Projekt wurde auch eine Programmibliothek zum Umgang mit UOF und den Schnittstellen des DIAS-Systems erstellt. Unter dem Namen koLibRI ist diese frei verfügbar⁸. Allerdings wird sie in der Form von der DNB aktuell nicht mehr bzw. nur in Teilen genutzt. Die derzeit intern eingesetzten Tools und Workflows sind nicht öffentlich zur Nachnutzung verfügbar.

⁸ http://kopal.langzeitarchivierung.de/index_koLibRI.php.de

3. Einordnung nach den Grundsätzen zur SIP-Bildung

Mit der Handreichung der nestor Arbeitsgruppe SIP-Konkretisierung „Grundsätze zur SIP-Bildung“⁹ liegt ein systematischer Katalog von Grundsätzen vor, die dabei helfen sollen, zur Vereinheitlichung von Informationspakete bei der Übernahme in digitale Archive beizutragen. Im Folgenden werden die Grundsätze mit dem SIP-Format der DNB abgeglichen.

3.1. Allgemeine Grundsätze

3.1.1. Es MUSS möglich sein, beliebige digitale Objekte und Metadaten in ein Informationspaket aufzunehmen.

Ein UOF-Paket kann beliebige Objekte beinhalten. Auch beliebige Metadaten können optional integriert werden. Die UOF-Spezifikationen geben nur bestimmte Metadaten vor, die mindestens vorhanden sein sollen.

3.1.2. Das Informationspaket DARF NICHT die Mittel, Methoden oder Werkzeuge für den Ingest einschränken.

Ein UOF-Paket ist entweder eine ZIP- oder TAR-Datei. Diese Formate können von den meisten Systemen verarbeitet werden.

3.1.3. Das Paketformat DARF NICHT den logisch-inhaltlichen Umfang der digitalen Objekte und Metadaten definieren, die ein Informationspaket bilden.

Ein logisches Objekt wird als ein UOF-Paket gesehen und kann aus einer oder mehreren Dateien bestehen.

3.1.4. Das Informationspaket MUSS skalierbar sein.

Die Spezifikationen von UOF geben keine Begrenzungen zu Umfang und Anzahl der enthaltenen Dateien eines Objekts an. Allerdings wird das benutzte Archivsystem technische Beschränkungen haben, die Grenzen setzen.

3.1.5. Das Informationspaket MUSS maschinenlesbar und automatisierbar zu verarbeiten sein.

⁹ <https://d-nb.info/1214014216/34>

Die Metadaten eines UOF liegen im XML-Format METS vor, für das es ein Schema und automatisierbare Bearbeitungsprogramme gibt.

3.1.6. Das Informationspaket MUSS interpretierbar sein, um eine auch für den Menschen inhaltliche Deutung zu ermöglichen.

Die Struktur und Inhalt der UOF-METS-Datei ist lesbar.

3.1.7. Die Spezifikation des Informationspakets MUSS offen und frei sein.

Die Spezifikationen und das METS-Profil vom UOF sind frei verfügbar.

3.1.8. Die Komplexität der Spezifikation eines Informationspakets SOLL angemessen sein.

Die Möglichkeit, durch weitere XML-Schemas spezifische Metadaten zu integrieren, ermöglicht eine Komplexität der Beschreibung, die über die Basismetadaten der Spezifikation hinausgeht.

3.2. Grundsätze zur Identifikation eines Informationspakets

3.2.1. Jedes Informationspaket MUSS einen im archivierenden Archiv eindeutigen und dauerhaften Identifikator haben oder erhalten.

Jedes Objektinstanz im Archivsystem bekommt eine sogenannte interne Objekt-ID. Auch Migrationen von vorhandenen Objekten bekommen jeweils eine eigene interne Objekt-ID, die beim Archivsystem für eine externe Objekt-ID des logischen Objekts abgefragt werden kann.

3.2.2. Jedes Informationspaket SOLL einen Identifikator besitzen, der global eindeutig und dauerhaft ist.

Jedes SIP hat beim Ingest eine sogenannte externe Objekt-ID, die bei der DNB immer ein Persistent Identifier als URN ist. Diese bezieht sich auf das logische Objekt, welches in verschiedenen Migrationen vorhanden sein kann, die dann jeweils eigene interne Objekt-IDs haben.

3.2.3. Alle Teile eines Informationspakets SOLLEN einen eindeutigen und dauerhaften Identifikator haben.

Einige METS-Abschnitte in einer UOF haben IDs, über die Teile davon referenziert werden können. Allerdings sind diese nicht gezielt beim System abfragbar.

3.3. Struktur eines Informationspakets

3.3.1. Das Informationspaket MUSS sicherstellen, dass Daten und Metadaten logisch voneinander getrennt sind.

Die Metadaten sind in der METS-Datei und klar von den Inhaltsdateien getrennt. Die UOF-Spezifikationen geben dabei vor, dass Inhaltsdateien nicht in die METS-Dateien integriert werden dürfen, was in METS ansonsten möglich ist.

3.3.2. Die Struktur des Informationspakets SOLL die Trennung verschiedener Arten von Metadaten ermöglichen.

Die verschiedenen Metadatenarten sind in der Struktur von METS getrennt.

3.3.3. Die Struktur des Informationspakets SOLL die Erstellung von Daten und Metadaten in mehreren Repräsentationen ermöglichen.

Durch das Konzept von einer externen Objekt-ID für logische Objekte und mehreren Objekt-IDs von Migrationen davon, können mit UOF zumindest Repräsentationen als Migrationen verwaltet werden.

3.3.4. Die Struktur des Informationspakets SOLL die Möglichkeiten zum Hinzufügen zusätzlicher Daten zum Informationspaket explizit definieren.

Die UOF-Spezifikationen sehen ein Hinzufügen nicht vor. Eine Änderung der Inhaltsdaten entspräche einer Migration, die dann als neues Objekt mit gleicher externer Objekt-ID behandelt würde.

3.3.5. Jedes Informationspaket SOLL seinen Informationstypen mitteilen.

Es gibt das LMER-Object-Element Gruppen-ID, welches im Sinne eines Informationstypen genutzt werden könnte. Das wird aber nicht von UOF verlangt und vom Archivsystem auch nicht so behandelt.

3.4. Metadaten eines Informationspakets

3.4.1. Metadaten im Informationspaket SOLLEN einem etablierten Standard entsprechen.

Während METS ein etablierter Standard ist, gilt das nicht für das im UOF geforderte LMER.

3.4.2. Die exakte Verwendung der Metadaten SOLLTE in Profilen für Informationstypen erarbeitet werden.

UOF liegt selbst als METS-Profil vor. Es kennt aber keine Profile für Informationstypen.

3.4.3. Jedes Informationspaket KANN beschreibende Metadaten enthalten.

Ein UOF kann auch beschreibende Metadaten im dafür vorgesehenen Abschnitt von METS beinhalten. Dies wird bei der DNB allerdings nicht genutzt, da jedes SIP mit der URN eine dauerhafte Verknüpfung zum jeweiligen Katalogeintrag hat.

3.5. Authentizität und Integrität eines Informationspakets

3.5.1. Im Informationspaket SOLLEN Möglichkeiten enthalten sein, die Authentizität sicherzustellen.

Die UOF-Spezifikationen sehen keine elektronischen Signaturen vor.

3.5.2. Im Informationspaket SOLLEN Möglichkeiten enthalten sein, die Integrität sicherzustellen.

Checksummen sind sowohl als Teil der METS-Elemente, wie auch als Teil der LMER-Elemente in UOF vorgesehen.

SIP-Format der Technischen Informationsbibliothek Hannover (TIB)

1. Ausgangslage

Die Technische Informationsbibliothek Hannover (TIB) betreibt ein digitales Langzeitarchiv im Verbund mit den beiden anderen Zentralen Fachbibliotheken ZB MED und ZBW und nutzt dafür das Langzeitarchivierungssystem Rosetta von ExLibris. Die TIB hostet und administriert das System und stellt dieses den Partnerbibliotheken sowie Dienstleistungsnehmern zur Verfügung.

Die TIB übernimmt Objekte von verschiedensten Quellen und Datenproduzenten in das digitale Langzeitarchiv. Vor der Übernahme neuer Bestände schließt das Team digitale Langzeitarchivierung für jeden Datenproduzenten eine Übernahmepolicy¹ ab.

In der Regel können die Objekte jedoch nicht so, wie sie von den Datenproduzenten abgeliefert oder von den Datenquellen abgeholt wurden, unmittelbar in das digitale Langzeitarchiv übernommen werden, sondern durchlaufen eine Qualitätskontrolle und müssen aufbereitet werden.

Die TIB unterscheidet deswegen die von der Datenquelle/vom Datenproduzenten übermittelten oder abgeholt **Eingangspaketstrukturen**² von **Pre-Ingest-SIPs**³, die mit einem Submission Tool in eine Rosetta-konforme Struktur (METS oder CSV mit zugehöriger Verzeichnisstruktur) gebracht wurden, und **Post-Ingest-SIPs**⁴, die nach der Übernahme in das digitale Langzeitarchiv noch mit zusätzlichen Metadaten angereichert werden. Falls das Pre-Ingest-SIP nicht bereits als METS-Struktur (s. 2.1 Aufbau einer METS-Datei) vorlag, wandelt das digitale Langzeitarchivierungssystem beim Ingest die abgelieferte CSV-Datei (s. 2.2 Aufbau einer CSV-Datei) in eine METS-Datei⁵ um.

¹ TIB LZA-Wiki, Template für Übernahmepolicies:
<https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=86267623>

² TIB LZA-Wiki, SIP-Spezifikationen, Eingangspaketstrukturen:
[https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768012#Spezifikationf%C3%BCrAblieferungspakete\(SIP\)-EPS](https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768012#Spezifikationf%C3%BCrAblieferungspakete(SIP)-EPS)

³ TIB LZA-Wiki, SIP-Spezifikationen, Pre-Ingest-SIP:
[https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768012#Spezifikationf%C3%BCrAblieferungspakete\(SIP\)-PreSIP](https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768012#Spezifikationf%C3%BCrAblieferungspakete(SIP)-PreSIP)

⁴ TIB LZA-Wiki, SIP-Spezifikationen, Post-Ingest-SIP:
[https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768012#Spezifikationf%C3%BCrAblieferungspakete\(SIP\)-PostSIP](https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768012#Spezifikationf%C3%BCrAblieferungspakete(SIP)-PostSIP)

⁵ TIB LZA-Wiki, SIP-Spezifikationen:
[https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768012#Spezifikationf%C3%BCrAblieferungspakete\(SIP\)-METSObjektemitmehrerenRepr%C3%A4sentationenoderkomplexeDateiablagen](https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768012#Spezifikationf%C3%BCrAblieferungspakete(SIP)-METSObjektemitmehrerenRepr%C3%A4sentationenoderkomplexeDateiablagen)

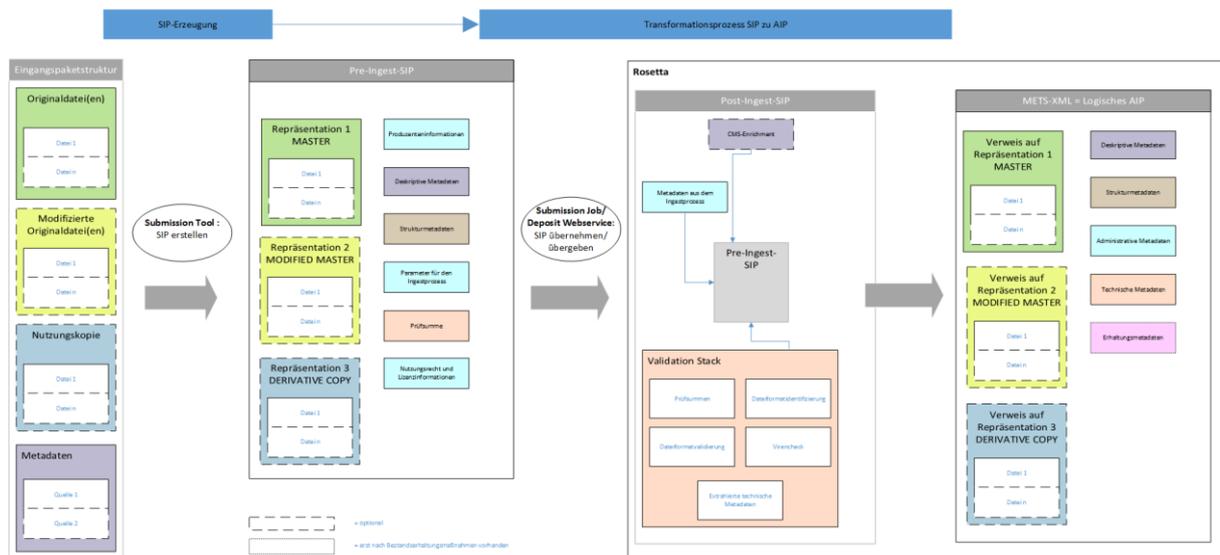


Abbildung 1 Beispiel für Paketstrukturen, (Quelle: TIB LZA-Wiki⁶)

2. SIP-Format

Die TIB gibt abliefernden Datenproduzenten Empfehlungen für verschiedene Eingangspaketstrukturen⁷, die jeweils für verschiedene Arten von Objekten geeignet sind.

Aus den Eingangspaketstrukturen erzeugt das Team Digitale Langzeitarchivierung jeweils ein Pre-Ingest-SIP in Form einer Verzeichnisstruktur sowie einer METS- oder CSV -Datei, das vom Langzeitarchivierungssystem Rosetta automatisiert weiterverarbeitet werden kann.

Beispiel METS:

```

IDENTIFIER
|--dc.xml (Pflicht)
|--content
|  |--iel.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

```

⁶ TIB LZA-Wiki, Datenmanagement, Beispiel für Paketstrukturen:
<https://wiki.tib.eu/confluence/display/lza/Datenmanagement#Datenmanagement-PKSBeispiel%3%BCrPaketstrukturen>

⁷ TIB LZA-Wiki, Eingangspaketstrukturen:
[https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768012#Spezifikation%3%BCrAblieferungspakete\(SIP\)-EPS](https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768012#Spezifikation%3%BCrAblieferungspakete(SIP)-EPS)

```

|         |--File n
|         |-- Folder 0-n
|         |--File 0-m

```

Beispiel CSV:

```

IDENTIFIER
|--content
| |--Identifier.csv
| |--streams
|   |--Identifier
|     |--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

```

Jedes Paket beinhaltet deskriptive, rechtliche, administrative, technische und identifizierende Metadaten⁸ und wird während des Ingestprozesses mit weiteren Metadaten angereichert. Ein SIP kann zudem 1-n Repräsentationen eines Objektes beinhalten.

2.1. Aufbau einer METS-Datei

Rosetta verfügt über ein eigenes METS-Profil⁹. Die METS-Datei kann gegen die Rosetta XSD¹⁰ validiert werden.

Falls das Pre-Ingest-SIP nicht bereits als METS-Struktur vorlag, wandelt das digitale Langzeitarchivierungssystem beim Ingest die abgelieferte CSV-Datei in eine METS-Datei¹¹ um.

In die METS-Datei können weitere Metadatenstandards eingebunden werden. Beschreibende Metadaten müssen als Dublin Core vorliegen. Diese Metadaten werden indexiert. In die Source MD-

⁸ TIB LZA-Wiki, Metadaten: <https://wiki.tib.eu/confluence/display/lza/Metadaten>

⁹ Rosetta METS Profile: <http://www.loc.gov/standards/mets/profiles/00000042.xml>

¹⁰ Rosetta METS XSD: https://github.com/ExLibrisGroup/Rosetta.dps-sdk-projects/blob/master/current/dps-sdk-projects/dps-sdk-deposit/src/xsd/mets_rosetta.xsd

¹¹ TIB LZA-Wiki, SIP-Spezifikationen: [https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768012#Spezifikationf%C3%BCrAblieferungspakete\(SIP\)-METSObjektemitmehrerenRepr%C3%A4sentationenoderkomplexeDateiablagen](https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768012#Spezifikationf%C3%BCrAblieferungspakete(SIP)-METSObjektemitmehrerenRepr%C3%A4sentationenoderkomplexeDateiablagen)

Sektionen der ie.xml können beliebige Metadatenstandards (zum Beispiel MARC, Dublin Core, MODS, EAD, NISO oder MIX) eingebunden werden.

Für die Erfassung der administrativen, technischen, strukturellen, identifizierenden und Event-Metadaten werden DNX- und METS-Elemente genutzt. DNX wurde vom Softwarehersteller ExLibris spezifiziert und implementiert PREMIS, erweitert den Standard jedoch um weitere Elemente. Die Dokumentation von DNX ist öffentlich im Rosetta AIP Data Model¹² einsehbar. Die Fortschreibung von DNX wird von der Rosetta User Community gesteuert und überwacht.

Für eine Übersicht über von der TIB verwendeten Metadaten siehe die Metadatenseite¹³ im TIB LZA-Wiki.

2.1.1. Beispiel

Es folgt ein exemplarisches Pre-Ingest-SIP, das mit der TIB Submission Application erzeugt wurde. Es handelt sich um einen digitalen Konferenzband mit mehreren Repräsentationen, der unter einer CC BY 4.0 Lizenz veröffentlicht wurde.

¹² Rosetta Produktdokumentation, AIP Data Model:
https://knowledge.exlibrisgroup.com/Rosetta/Product_Documentation

¹³ TIB LZA-Wiki, Metadaten: <https://wiki.tib.eu/confluence/display/lza/Metadaten>

```

<?xml version="1.0" encoding="UTF-8"?>
<mets:mets xmlns:mets="http://www.loc.gov/METS/">
  <mets:dmdSec ID="ie-dmd">
    <mets:mdWrap MDTYPE="DC">
      <mets:xmlData>
        <dc:record xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/" >
          <dc:title>4th International Conference on Advances in
Materials, Machinery, Electronics (AMME 2021) : Wuhan City, China, October
23-24, 2021</dc:title>
          <dcterms:isPartOf>MATEC web of conferences ; volume 353
(2021) (ZDB-ID: 2673602-0)</dcterms:isPartOf>
          <dcterms:alternative>AMME 2021: Abweichender
Titel</dcterms:alternative>
          <dc:contributor>Liu, Lin (HerausgeberIn)</dc:contributor>
          <dc:contributor>Liu, Fang (HerausgeberIn)</dc:contributor>
          <dc:creator>International Conference on Advances in
Materials, Machinery, Electronics 4. 2021 Wuhan (VerfasserIn)</dc:creator>
          <dc:identifler>PPN 1783634499</dc:identifler>
          <dc:identifler>TIB:KXP1783634499</dc:identifler>
          <dc:contributor>International Conference on Advances in
Materials, Machinery, Electronics 4 ; Wuhan City : 2021.10.23-
24</dc:contributor>
          <dc:contributor>AMME 4 ; Wuhan City : 2021.10.23-
24</dc:contributor>
          <dc:publisher>Les Ulis : EDP Sciences</dc:publisher>
          <dc:rights>Namensnennung 4.0 International. CC BY 4.0 :
Creative Commons : https://creativecommons.org/licenses/by/4.0/</dc:rights>
          <dcterms:accessRights>Open Access :
http://purl.org/coar/access_right/c_abf2</dcterms:accessRights>
          <dcterms:issued>2021</dcterms:issued>
          <dc:language>eng</dc:language>
        </dc:record>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:dmdSec>
  <mets:amdSec ID="repl-amd">
    <mets:techMD ID="repl-amd-tech">
      <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
        <mets:xmlData>
          <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx">
            <section id="generalRepCharacteristics">
              <record>
                <key
id="preservationType">PRESERVATION_MASTER</key>
                <key id="usageType">VIEW</key>
                <key id="RevisionNumber">1</key>
              </record>
            </section>
          </dnx>
        </mets:xmlData>
      </mets:mdWrap>
    </mets:techMD>
    <mets:rightsMD ID="repl-amd-rights">
      <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
        <mets:xmlData>
          <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx"/>
        </mets:xmlData>
      </mets:mdWrap>
    </mets:rightsMD>
    <mets:sourceMD ID="repl-amd-source">

```

```

    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
      <mets:xmlData>
        <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx"/>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:sourceMD>
  <mets:digiprovMD ID="rep1-amd-digiprov">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
      <mets:xmlData>
        <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx"/>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:digiprovMD>
</mets:amdSec>
<mets:amdSec ID="rep2-amd">
  <mets:techMD ID="rep2-amd-tech">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
      <mets:xmlData>
        <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx">
          <section id="generalRepCharacteristics">
            <record>
              <key id="preservationType">DERIVATIVE_COPY</key>
              <key id="usageType">VIEW</key>
              <key id="RevisionNumber">1</key>
            </record>
          </section>
        </dnx>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:techMD>
  <mets:rightsMD ID="rep2-amd-rights">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
      <mets:xmlData>
        <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx"/>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:rightsMD>
  <mets:sourceMD ID="rep2-amd-source">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
      <mets:xmlData>
        <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx"/>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:sourceMD>
  <mets:digiprovMD ID="rep2-amd-digiprov">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
      <mets:xmlData>
        <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx"/>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:digiprovMD>
</mets:amdSec>
<mets:amdSec ID="fid1-2-amd">
  <mets:techMD ID="fid1-2-amd-tech">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
      <mets:xmlData>
        <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx">
          <section id="generalFileCharacteristics">
            <record>
              <key id="fileMIMEType">application/pdf</key>
            </record>
          </section>
        </dnx>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:techMD>

```

```

        <key
id="fileOriginalPath"/>/KXP1783634499/content/streams/DERIVATIVE_COPY/178363
4499.pdf</key>
    </record>
</section>
<section id="fileFixity">
    <record>
        <key id="fixityType">MD5</key>
        <key
id="fixityValue">eedd43c0ab3163ec95be33cea66148c4</key>
    </record>
</section>
</dnx>
</mets:xmlData>
</mets:mdWrap>
</mets:techMD>
<mets:rightsMD ID="fid1-2-amd-rights">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
        <mets:xmlData>
            <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx"/>
        </mets:xmlData>
    </mets:mdWrap>
</mets:rightsMD>
<mets:sourceMD ID="fid1-2-amd-source">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
        <mets:xmlData>
            <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx"/>
        </mets:xmlData>
    </mets:mdWrap>
</mets:sourceMD>
<mets:digiprovMD ID="fid1-2-amd-digiprov">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
        <mets:xmlData>
            <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx"/>
        </mets:xmlData>
    </mets:mdWrap>
</mets:digiprovMD>
</mets:amdSec>
<mets:amdSec ID="fid1-1-amd">
    <mets:techMD ID="fid1-1-amd-tech">
        <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
            <mets:xmlData>
                <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx">
                    <section id="generalFileCharacteristics">
                        <record>
                            <key id="fileMIMEType">application/pdf</key>
                            <key
id="fileOriginalPath"/>/KXP1783634499/content/streams/MASTER/2013-CNAM-
01.pdf</key>
                        </record>
                    </section>
                    <section id="fileFixity">
                        <record>
                            <key id="fixityType">MD5</key>
                            <key
id="fixityValue">83e04608def299b768362fd5c91279fa</key>
                        </record>
                    </section>
                </dnx>
            </mets:xmlData>
        </mets:mdWrap>
    </mets:techMD>

```

```

</mets:techMD>
<mets:rightsMD ID="fid1-1-amd-rights">
  <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
    <mets:xmlData>
      <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx"/>
    </mets:xmlData>
  </mets:mdWrap>
</mets:rightsMD>
<mets:sourceMD ID="fid1-1-amd-source">
  <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
    <mets:xmlData>
      <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx"/>
    </mets:xmlData>
  </mets:mdWrap>
</mets:sourceMD>
<mets:digiprovMD ID="fid1-1-amd-digiprov">
  <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
    <mets:xmlData>
      <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx"/>
    </mets:xmlData>
  </mets:mdWrap>
</mets:digiprovMD>
</mets:amdSec>
<mets:amdSec ID="ie-amd">
  <mets:techMD ID="ie-amd-tech">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
      <mets:xmlData>
        <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx">
          <section id="CMS">
            <record>
              <key id="system">TIB</key>
              <key id="recordId">KXP1783634499</key>
            </record>
          </section>
          <section id="generalIECharacteristics">
            <record>
              <key id="status">ACTIVE</key>
              <key id="IEEntityType">GreyLiterature</key>
              <key id="UserDefinedA">MR_born-digital</key>
            </record>
          </section>
        </dnx>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:techMD>
  <mets:rightsMD ID="ie-amd-rights">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
      <mets:xmlData>
        <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx">
          <section id="accessRightsPolicy">
            <record>
              <key id="policyId">AR_EVERYONE</key>
              <key id="policyDescription">Keine
Beschränkung</key>
            </record>
          </section>
        </dnx>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:rightsMD>
  <mets:sourceMD ID="ie-amd-source">

```

```

    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
      <mets:xmlData>
        <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx"/>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:sourceMD>
  <mets:digiprovMD ID="ie-amd-digiprov">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
      <mets:xmlData>
        <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx"/>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:digiprovMD>
</mets:amdSec>
<mets:fileSec>
  <mets:fileGrp ID="rep1" ADMID="rep1-amd">
    <mets:file ID="fid1-1" ADMID="fid1-1-amd">
      <mets:FLocat xmlns:xlink="http://www.w3.org/1999/xlink"
        LOCTYPE="URL"
        xlink:href="file://MASTER/2013-CNAM-01.pdf"/>
    </mets:file>
  </mets:fileGrp>
  <mets:fileGrp ID="rep2" ADMID="rep2-amd">
    <mets:file ID="fid1-2" ADMID="fid1-2-amd">
      <mets:FLocat xmlns:xlink="http://www.w3.org/1999/xlink"
        LOCTYPE="URL"
        xlink:href="file://DERIVATIVE_COPY/1783634499.pdf"/>
    </mets:file>
  </mets:fileGrp>
</mets:fileSec>
<mets:structMap ID="rep1-1" TYPE="PHYSICAL">
  <mets:div LABEL="Preservation Master">
    <mets:div LABEL="rep1">
      <mets:div LABEL="" TYPE="FILE">
        <mets:fptr FILEID="fid1-1"/>
      </mets:div>
    </mets:div>
  </mets:div>
</mets:structMap>
<mets:structMap ID="rep2-1" TYPE="PHYSICAL">
  <mets:div LABEL="Derivative Copy">
    <mets:div LABEL="rep2">
      <mets:div LABEL="" TYPE="FILE">
        <mets:fptr FILEID="fid1-2"/>
      </mets:div>
    </mets:div>
  </mets:div>
</mets:structMap>
</mets:mets>

```

2.2. Aufbau einer CSV-Datei

Die CSV-Datei erfüllt die gleiche Funktion in einem Pre-Ingest-SIP wie die METS-Datei. Sie beinhaltet deskriptive, rechtliche, administrative und technische Metadaten, eine Zuordnung von 1-n Dateien zu 1-n Repräsentationen und von 1-n Repräsentationen zu einer IE.

2.2.1. Beispiel

Es folgt ein exemplarisches Pre-Ingest-SIP, das mit dem CSV-Script erzeugt wurde.

Es handelt sich um einen E-Journal Artikel aus dem DEAL Wiley Dark Archive mit einer Repräsentation und mehreren Dateien pro Repräsentation.

Object Type, Title (DC), Is Part Of (DCTERMS), Title (DC), Identifier - URI (DC), Title (DC), Creator (DC), Creator (DC), Creator (DC), Creator (DC), Date (DC), Identifier - ISSN (DC), Identifier - ISSN (DC), Identifier - URI (DC), Access Rights (DCTERMS), Rights (DC), Rights Holder (DCTERMS), Rights Holder (DCTERMS), Is Part Of (DCTERMS), Access Rights Policy ID (IE), IE User Defined A, IE User Defined B, IE Entity Type, License (DCTERMS), File Label, MD5, Preservation Type, Revision Number, Usage Type, File Original Path, File Original Name

SIP,MSID_03620972_2020_36_6_MSID1164,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
Collection,, "Wiley DEAL/Information Display/Volume 36", "Issue
6", 10.1002/msid.v36.6,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
IE,,,,, "How Will Quantum Dots Enable Next-Gen Display
Technologies?", "Ishida, Takeshi", "Nakanishi, Yohei", "Izumi,
Makoto", "Berryman-Bousquet, Valerie", "November-December 2020", "Print-ISSN:
0362-0972", "Online-ISSN: 2637-496X", "DOI: 10.1002/msid.1164", "Annex D,
DEAL-Wiley Vertrag DOI: 10.17617/2.3027595", "Annex D, DEAL-Wiley Vertrag
DOI: 10.17617/2.3027595", "© 2020 Society for Information Display", "© 2020
Society for Information Display", "Information Display; Volume 36, Issue 6,
2020-11", 14996, DEAL-E-Journals_born-
digital,, Article, Dark_Archive_TIB_Wiley_v1_2020,,,,,
REP,,,,,,,,,,,,,,,,,,,,, PRESERVATION_MASTER, 1, VIEW,,,,,
FILE,,,,,,,,,,,,,,,,,,,,, 68b74b582b83a593f90b053492758ff1,,,,, MSID_03620
972_2020_36_6_MSID1164/MASTER/,msid1164.xml,
FILE,,,,,,,,,,,,,,,,,,,,, 18065c5c0b0f9e954c9d0f3a16beaf05,,,,, MSID_03620
972_2020_36_6_MSID1164/MASTER/,msid1164.pdf,
FILE,,,,,,,,,,,,,,,,,,,,, image_n,793278dbb98dc2fbbc8df328ac690650,,,,, MSI
D_03620972_2020_36_6_MSID1164/MASTER/image_n/,msid1164-gra-0001.png,
FILE,,,,,,,,,,,,,,,,,,,,, image_n,401d139326e24524e2a1c20cddb1f2f,,,,, MSI
D_03620972_2020_36_6_MSID1164/MASTER/image_n/,msid1164-fig-0001.png,
FILE,,,,,,,,,,,,,,,,,,,,, image_n,60ce246b2974d93af8f5f27c0bec6113,,,,, MSI
D_03620972_2020_36_6_MSID1164/MASTER/image_n/,msid1164-gra-0003.png,
FILE,,,,,,,,,,,,,,,,,,,,, image_n,af715fcf34c3d1bbb36bee853ec7ebe5,,,,, MSI
D_03620972_2020_36_6_MSID1164/MASTER/image_n/,msid1164-fig-0002.png,
FILE,,,,,,,,,,,,,,,,,,,,, image_n,4bdfd7d29584813ef6f97f4c9c5911ee,,,,, MSI
D_03620972_2020_36_6_MSID1164/MASTER/image_n/,msid1164-gra-0004.png,
FILE,,,,,,,,,,,,,,,,,,,,, image_n,d7dbdd6b4aeaelca78d6cle94aff98d3,,,,, MSI
D_03620972_2020_36_6_MSID1164/MASTER/image_n/,msid1164-fig-0003.png,
FILE,,,,,,,,,,,,,,,,,,,,, image_n,4e44e4f19f24ce8c05ca27710036bbb1,,,,, MSI
D_03620972_2020_36_6_MSID1164/MASTER/image_n/,msid1164-gra-0005.png,
FILE,,,,,,,,,,,,,,,,,,,,, image_n,77cf5bf01e7349cab3bdae9a7b27cdb1,,,,, MSI
D_03620972_2020_36_6_MSID1164/MASTER/image_n/,msid1164-fig-0004.png,
FILE,,,,,,,,,,,,,,,,,,,,, image_n,464a85f6d79c0ce3401e4d332736fd63,,,,, MSI
D_03620972_2020_36_6_MSID1164/MASTER/image_n/,msid1164-gra-0002.png,

2.3. Nutzung im Archivsystem der TIB

Pre-Ingest-SIPs werden nach dem Ingest zu Post-Ingest-SIPs und im digitalen Langzeitarchiv mit weiteren Metadaten angereichert. Dort durchlaufen sie weitere Qualitätskontrollen wie den sogenannten Validation Stack mit Virenprüfung, Dateiformatidentifizierung und –validierung, Prüfsummencheck und Extraktion technischer Metadaten¹⁴. Erst bei Erreichen des permanenten Archivspeichers werden aus den Post-Ingest-SIPs AIPs.

Eine ausführliche Dokumentation der Transformation der verschiedenen SIP-Typen in ein AIP ist im LZA-Wiki der TIB zu finden¹⁵.

2.4. Tools

Tool	Anwendung	Quelle
TIB CSV-Erzeuger	SIP-Packer	https://github.com/TIB-Digital-Preservation/CSV-Ingest_generisch
TIB Submission Application	SIP-Packer	Quellcode noch nicht öffentlich verfügbar
ExLibris Rosetta SDK	SDK	https://github.com/ExLibrisGroup/Rosetta.dps-sdk-projects/tree/master/current
Nachnutzung der ZBW DSpace Submission Application	Qualitätskontrolle, SIP-Packer, Ingest, SIP-Tracking, Reporting	https://github.com/zbw/oai2rosetta
Hindawi Journal Downloader	Qualitätskontrolle, SIP-Packer	https://github.com/TIB-Digital-Preservation/hinjodl_classic
Pre-Ingest-Analyzer	Qualitätskontrolle	https://github.com/TIB-Digital-Preservation/pre-ingest-analyzer
Siegfried DROID JHOVE Analyzer	Qualitätskontrolle	https://github.com/TIB-Digital-Preservation/sf_DROID_JHOVE
TIB2OJS Submission Application	SIP-Packer, Ingest	Quellcode noch nicht öffentlich verfügbar
Rospydep	Ingest, SIP-Tracking, Reporting	Quellcode noch nicht öffentlich verfügbar, folgt auf: https://github.com/TIB-Digital-Preservation/rospydep

¹⁴ TIB LZA-Wiki, Übernahme (Ingest):
<https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768009>

¹⁵ TIB LZA-Wiki, Spezifikation für Archivinformationspakete (AIP), Transformation von Eingangspaketstrukturen zu SIPs und AIPs:
[https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768013#Spezifikationf%C3%BCrArchivinformationspakete\(AIP\)-Trans](https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768013#Spezifikationf%C3%BCrArchivinformationspakete(AIP)-Trans)

3. Einordnung nach den Grundsätzen zur SIP-Bildung

Mit der Handreichung der nestor Arbeitsgruppe SIP-Konkretisierung „Grundsätze zur SIP-Bildung“¹⁶ liegt ein systematischer Katalog von Grundsätzen vor, die dabei helfen sollen, zur Vereinheitlichung von Informationspakete bei der Übernahme in digitale Archive beizutragen. Im Folgenden werden die Grundsätze mit dem SIP-Format der TIB abgeglichen.

3.1. Allgemeine Grundsätze

3.1.1. Es MUSS möglich sein, beliebige digitale Objekte und Metadaten in ein Informationspaket aufzunehmen.

Es gibt keine technische Beschränkung auf bestimmte Dateiformate oder Metadatenstandards. Workflowabhängig kann es sinnvoll sein, zwecks Qualitätskontrolle auf definierte erlaubte Dateiformate einzuschränken.

3.1.2. Das Informationspaket DARF NICHT die Mittel, Methoden oder Werkzeuge für den Ingest einschränken.

Die definierten SIPs bestehen aus einer definierten Datenstruktur und können mit Tools in verschiedenen Sprachen erzeugt werden.

3.1.3. Das Paketformat DARF NICHT den logisch-inhaltlichen Umfang der digitalen Objekte und Metadaten definieren, die ein Informationspaket bilden.

Es gibt keine Beschränkung hinsichtlich der Anzahl der Dateien und Anzahl der Repräsentationen.

Workflowabhängig sind Einschränkungen auf die erwartete Struktur zwecks Qualitätskontrolle sinnvoll.

3.1.4. Das Informationspaket MUSS skalierbar sein.

Das SIP ist skalierbar in Abhängigkeit von der Infrastruktur.

3.1.5. Das Informationspaket MUSS maschinenlesbar und automatisierbar zu verarbeiten sein.

¹⁶ <https://d-nb.info/1214014216/34>

Verzeichnisstrukturen, METS und CSV werden von Prozessen/Skripten ausgelesen und transformiert.

3.1.6. Das Informationspaket MUSS interpretierbar sein, um eine auch für den Menschen inhaltliche Deutung zu ermöglichen.

Die Verzeichnisstrukturen, METS und CSV sind menschenlesbar.

3.1.7. Die Spezifikation des Informationspakets MUSS offen und frei sein.

SIP-Spezifikationen sind öffentlich im LZA-Wiki verfügbar. Das METS-Profil ist öffentlich zugänglich.

3.1.8. Die Komplexität der Spezifikation eines Informationspakets SOLL angemessen sein.

Es werden verschiedene SIP-Spezifikationen in Abhängigkeit von der Komplexität der Inhaltsobjekte eingesetzt. Die Spezifikationen sind erweiterbar.

3.2. Grundsätze zur Identifikation eines Informationspakets

3.2.1. Jedes Informationspaket MUSS einen im archivierenden Archiv eindeutigen und dauerhaften Identifikator haben oder erhalten.

Dateien, Repräsentationen, IEs und Metadatendateien erhalten archivinterne eindeutige Identifier.

3.2.2. Jedes Informationspaket SOLL einen Identifikator besitzen, der global eindeutig und dauerhaft ist.

Deskriptive Metadaten beinhalten DOIs, URNs, Handles und weitere externe Identifier, sofern vergeben. Die TIB vergibt DOIs für ausgewählte Bestände.

3.2.3. Alle Teile eines Informationspakets SOLLEN einen eindeutigen und dauerhaften Identifikator haben.

Dateien, Repräsentationen, IEs und Metadatendateien erhalten archivinterne eindeutige Identifier.

3.3. Struktur eines Informationspakets

3.3.1. Das Informationspaket MUSS sicherstellen, dass Daten und Metadaten logisch voneinander getrennt sind.

Die TIB nutzt ein logisches AIP und getrennte Speicherbereiche für METS-, Metadaten- und die Inhaltsdateien pro Rosetta-Institution.

3.3.2. Die Struktur des Informationspakets SOLL die Trennung verschiedener Arten von Metadaten ermöglichen.

Die Trennung verschiedener Metadatenkategorien erfolgt bereits durch die Verwendung von METS.

3.3.3. Die Struktur des Informationspakets SOLL die Erstellung von Daten und Metadaten in mehreren Repräsentationen ermöglichen.

Verschiedene Repräsentationen sind teilweise schon beim Ingest vorhanden. Versionierung erfolgt bei Änderungen an der IE (Metadaten und Inhaltsdaten) unter gleicher IE-ID mit Versionsnummer.

3.3.4. Die Struktur des Informationspakets SOLL die Möglichkeiten zum Hinzufügen zusätzlicher Daten zum Informationspaket explizit definieren.

Alle SIP-Spezifikationen sind erweiterbar.

3.3.5. Jedes Informationspaket SOLL seinen Informationstypen mitteilen.

Im SIP kann ein IE Entity Type definiert werden. Der IE Entity Type hat keine weiteren Auswirkungen auf die SIP-Spezifikation.

3.4. Metadaten eines Informationspakets

3.4.1. Metadaten im Informationspaket SOLLEN einem etablierten Standard entsprechen.

Innerhalb des METS-Containers werden weitere Metadatenstandards genutzt.

Nicht standardisierte Metadaten können als Source Metadaten hinterlegt werden.

3.4.2. Die exakte Verwendung der Metadaten SOLLTE in Profilen für Informationstypen erarbeitet werden.

Ein METS-Profil ist vorhanden, es gibt keine Sub-Profile.

3.4.3. Jedes Informationspaket KANN beschreibende Metadaten enthalten.

SIPs können deskriptive Metadaten in unterschiedlichen Formaten aufnehmen.

Beschreibende Metadaten oder ein Identifier, der auf einen Metadatensatz verweist, müssen Bestandteil des SIPs sein. Mittels des Identifiers werden im Pre-Ingest-, spätestens im Ingestprozess beschreibende Metadaten zum Objekt abgeholt. Die TIB hat an ein AIP die Anforderung, dass dieses selbstbeschreibend und selbstdokumentierend sein muss und somit beschreibende Metadaten zu dem Objekt beinhalten muss.

3.5. Authentizität und Integrität eines Informationspakets

3.5.1. Im Informationspaket SOLLEN Möglichkeiten enthalten sein, die Authentizität sicherzustellen.

Erstellungsdatum und erstellender User sind ab dem Ingest Teil des SIPs. Es werden keine elektronischen Signaturen vergeben.

Bei Änderungen wird das AIP unter Beibehaltung aller vorherigen Versionen versioniert. Die Dokumentation von Änderungen erfolgt in den Metadaten.

3.5.2. Im Informationspaket SOLLEN Möglichkeiten enthalten sein, die Integrität sicherzustellen.

Prüfsummen sind (optionaler) Bestandteil der Pre-Ingest-SIP-Spezifikationen, ab dem Ingest sind 3 Prüfsummen im SIP enthalten. Weitere Maßnahmen zur Integritätssicherung erfolgen zusätzlich über das Storage Management. Maßnahmen zur Integritätssicherung im Objektlebenszyklus sind im LZA-Wiki der TIB dokumentiert¹⁷.

¹⁷ TIB LZA-Wiki, Erhalt der Datenintegrität als Teil der Prozessroutinen:
<https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=63768018>



SLUB

Wir führen Wissen.

Grundsätze zur SIP-Bildung - Beispiel SLUBArchiv.digital

SLUB Dresden

Version 1.1, 2023-04-21

Inhaltsverzeichnis

Überblick	1
Teil 1: Das SIP-Format	1
Warum der Wechsel auf BagIt?	1
Aufbau SIP	2
Komprimierte SIPs	8
AIP Update	8
Tooling	9
Teil 2: Grundsätze zur SIP-Bildung am Beispiel	10
Eine retrodigitalisierte Lokalzeitung von 1932	10
1. Allgemeine Grundsätze	12
2. Grundsätze zur Identifikation eines IP	14
3. Struktur eines IP	14
4. Metadaten eines IP	15
5. Authentizität und Integrität eines IP	15

Überblick

Dieses Dokument beschreibt den Aufbau eines Ablieferungspaketes (englisch: Submission Information Package, SIP), welches für die automatische Verarbeitung zu archivierender digitaler Datenobjekte (IE) durch die Submission Application der [SLUB Dresden](https://www.slub-dresden.de/) [https://www.slub-dresden.de/] und Einlieferung (Ingest) in deren Langzeitarchiv geeignet ist (Stand April 2023).

Teil 1 ist ein Auszug aus dem SIP-spezifischen Teil der Übergabespezifikation des [SLUBArchiv.digital](https://slubarchiv.slub-dresden.de/) [https://slubarchiv.slub-dresden.de/]. Die Spezifikation umfasst noch zahlreiche weitere Dokumente ^[1]. Teil 2 versucht anhand der nestor Grundsätze zur SIP-Bildung ^[2] ein aktuelles Beispiel zu bewerten.

Teil 1: Das SIP-Format

Warum der Wechsel auf BagIt?

Das SLUBArchiv.digital hat sehr gründlich die Vor- und Nachteile des Wechsels des Formates der Submission Information Packages (SIP) von einer eigenen, auf METS basierenden, hin zu einer auf [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493] basierenden Variante abgewogen. Der Wechsel wurde Anfang April 2020 vollzogen. Für die Aufnahme in das Archiv ist in allen Workflows (inhouse und extern) nur noch das BagIt-basierte Format zulässig.

Für den Wechsel sprachen mehrere Gründe:

- [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493] ist mittlerweile in Version 1.0 als [RFC](https://www.rfc-editor.org) [https://www.rfc-editor.org] standardisiert.
- Immer mehr Archivilösungen setzen auf BagIt als Einlieferungsformat (zB. Archivemata und Rosetta).
- Es gibt bereits eine große Anzahl an Programmen und Programmbibliotheken, die Funktionalitäten zur einfachen Verarbeitung von BagIt bereitstellen, hier eine kleine Auswahl:
 - [libcbag](https://andreas-romeike.de/software.html#_libcbag_a_free_and_opensource_c_library_to_handle_bagit_structures) [https://andreas-romeike.de/software.html#_libcbag_a_free_and_opensource_c_library_to_handle_bagit_structures] C++-Bibliothek
 - [pybagit](https://pypi.python.org/pypi/pybagit/) [https://pypi.python.org/pypi/pybagit/] Python-Modul
 - [Archive::BagIt](https://metacpan.org/module/Archive::BagIt) [https://metacpan.org/module/Archive::BagIt] Perl-Modul
 - [Bagger](https://github.com/LibraryOfCongress/bagger) [https://github.com/LibraryOfCongress/bagger] GUI-Programm
 - [bagit-python](https://github.com/LibraryOfCongress/bagit-python) [https://github.com/LibraryOfCongress/bagit-python] CLI-Programm
 - [bagit-java](https://github.com/LibraryOfCongress/bagit-java) [https://github.com/LibraryOfCongress/bagit-java] Java-Bibliothek
- Das Datenformat ist einfach aufgebaut und kann mit Bordmitteln erzeugt und verarbeitet werden.
- Durch die Standardisierung von [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493] und weite Verbreitung erfolgt eine Normierung der Werkzeuge für die Erstellung, Prüfung und Verarbeitung von SIPs.
- Die Hürden für die Arbeit mit BagIt sind geringer als die von METS-basierten SIPs.
- [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493] ist modular aufgebaut und erlaubt die einfache Extraktion

von Bestandteilen des SIPs.

Neben dem offiziellen RFC [<https://www.rfc-editor.org>] existiert auch ein nestor Kurzartikel ^[3] aus der Praxisreihe, der das Format kurz zusammenfasst und Unterschiede zu Vorgängerversionen nennt.

Aufbau SIP

Generelle Festlegungen

Jedes SIP besteht aus einem Verzeichnis mit einem eindeutigen Namen (z.B. Zeitstempel der Erzeugung), welches die BagIt [<https://tools.ietf.org/html/rfc8493>]-Verzeichnisstruktur abbildet. Der Verzeichnisname wird nicht ausgewertet, sollte aber eindeutig sein, damit es nicht zu Überschreibungen durch die Abliefernden kommt.

Der Produzent ist in der Verantwortung, ein SIP mit korrekter BagIt [<https://tools.ietf.org/html/rfc8493>]-Implementierung zu erstellen.

Jedes SIP enthält genau eine Intellektuelle Einheit (IE). Dies vereinfacht die Prozessierung im SLUBArchiv.digital.



Eine Intellektuelle Einheit ist ein Datenobjekt, welches aus einer oder mehreren Dateien besteht, die gemeinsam verwaltet und charakterisiert werden.

Alle Dateien dieser IE müssen innerhalb der BagIt [<https://tools.ietf.org/html/rfc8493>]-Struktur referenziert sein.

Das SLUBArchiv.digital unterscheidet zwei Arten von SIPs. Ein SIP kann

- entweder die Daten für die erstmalige Aufnahme einer IE (Erstingest),
- oder die Daten für eine Aktualisierung (Update) der IE (AIP Update)

enthalten.

Bestandteile des SIP

Grundlage

Basis des SIPs ist eine BagIt [<https://tools.ietf.org/html/rfc8493>]-Struktur nach RFC8493. Ergänzend gelten folgende Festlegungen.

Deskriptive Metadaten

Metadaten des Bags werden als Schlüssel-Werte-Paare in der *bag-info.txt*-Datei kodiert (sh. 2.2.2, RFC 8493).



Es sollten nur diejenigen deskriptiven Metadaten übernommen werden, die minimal notwendig sind, um einen rudimentären Katalog aufbauen zu können.

Es wird empfohlen, die Angaben zur dienstnehmenden Institution und zu persistenten Identifiern in den beschreibenden Metadaten zu kodieren. Dazu sind die vorgesehenen Felder nach BagIt

[<https://tools.ietf.org/html/rfc8493>]-Spezifikation zu nutzen (zB. *Source-Organization*, *External-Description*, ...).

Ergänzend sollten, wenn möglich, noch der Titel des Dokumentes (*Title*) und die Verfasser (*Author*) als Schlüssel-Werte-Paare kodiert werden.

Beispiel MODS als Ausgangspunkt

```
<mods:mods xmlns:mods="http://www.loc.gov/mods/v3">
  <mods:location>
    <mods:physicalLocation authority="marcorg" displayLabel="Saxon State Library,
Dresden, Germany">DE-14</mods:physicalLocation>
    <mods:shelfLocator>Hist.Sax.M.37.t,120</mods:shelfLocator>
  </mods:location>
  <mods:relatedItem type="series">
    <mods:titleInfo>
      <mods:title>Saxonica</mods:title>
    </mods:titleInfo>
  </mods:relatedItem>
  <mods:recordInfo>
    <mods:recordIdentifier source="http://digital.slub-dresden.de/oai/">oai:de:slub-
dresden:db:id-319037843</mods:recordIdentifier>
  </mods:recordInfo>
  <mods:physicalDescription>
    <mods:digitalOrigin>reformatted digital</mods:digitalOrigin>
    <mods:extent>[1] Bl.</mods:extent>
  </mods:physicalDescription>
  <mods:identifier type="urn">urn:nbn:de:bsz:14-db-id3190378431</mods:identifier>
  <mods:titleInfo>
    <mods:title>Eingabe der Handelskammer zu Leipzig den Entwurf eines Tabak-Steuer-
Gesetzes betr.</mods:title>
    <mods:subTitle>an den Reichstag zu Berlin</mods:subTitle>
  </mods:titleInfo>
  <mods:language>
    <mods:languageTerm authority="rfc3066" type="code">de</mods:languageTerm>
  </mods:language>
  <mods:originInfo>
    <mods:place>
      <mods:placeTerm type="text">[Leipzig]</mods:placeTerm>
    </mods:place>
    <mods:dateIssued keyDate="yes">1893</mods:dateIssued>
  </mods:originInfo>
  <mods:subject authority="slub">
    <mods:topic>eingdehaz</mods:topic>
  </mods:subject>
</mods:mods>
```

Beispiel Mapping des MODS auf bag-info.txt

```
External-Identifer: oai:de:slub-dresden:db:id-319037843
External-Identifer: urn:nbn:de:bsz:14-db-id3190378431
Title: Eingabe der Handelskammer zu Leipzig den Entwurf eines Tabak-Steuer-Gesetzes betr.
an den Reichstag zu Berlin
...
```

Es wird empfohlen, Metadaten, die zu einer zu archivierenden Intellektuellen Einheit gehören, neben der Kodierung in den Metadaten des BagIt **zusätzlich** im originalen Metadatenformat als Datei im Unterordner *meta/* abzulegen, siehe Abschnitt [Ablage Metadaten-Dateien](#) .

Die beschreibenden Metadaten, die originärer Bestandteil des [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493] sind (sh. *bag-info.txt*), werden u.a. für die Suche innerhalb des Archivs verwendet und müssen mit den eventuell beigelegten Metadaten-Dateien (unter *meta/*) inhaltlich übereinstimmen.



Das SLUBArchiv.digital sichert die Langzeitverfügbarkeit des IEs einer SIP via Formatmigration.

Für die Metadaten unter *meta/* wird **keine** Formatmigration durchgeführt!

Administrative Metadaten für die Steuerung des Ingests

Ebenfalls in der Datei *bag-info.txt* müssen **verpflichtend** nachfolgende Informationen als Schlüssel-Werte-Paare kodiert werden. Wiederholungen sind nicht zulässig.



Der Produzent ist die Rolle eines Dienstnehmers, der eine IE archivieren möchte und ein SIP an das SLUBArchiv.digital einliefert. Die nachfolgenden Daten beziehen sich auf den Workflow des Produzenten, unabhängig davon ob er im Auftrag Dritter handelt.

SLUBArchiv-sipVersion

Der Schlüssel *SLUBArchiv-sipVersion* enthält *v2020.1*, für die eindeutige Identifizierung des SIP-Formates des SLUBArchiv.digital.

SLUBArchiv-externalWorkflow

In diesem Schlüssel wird der Name des Übergabeworkflows des Dienstnehmers angegeben, aus dem die zu archivierenden Daten stammen. Zusammen mit der *SLUBArchiv-externalId* wird diese Information im SLUBArchiv.digital genutzt, um Archivpakete eindeutig zu identifizieren. Anhand dieser beiden Einträge können mehrere SIPs zu einem digitalen Dokument (einer IE) zugeordnet werden. Dies ermöglicht, dass Archivpakete aktualisiert werden können.

Der Name muss eindeutig sein, sowie sich aus den Zeichen *a-z0-9_-* zusammensetzen. Großschreibung wird nicht unterstützt. Empfohlen wird einen sprechenden Namen zu verwenden.

SLUBArchiv-externalId

In diesem Schlüssel muss eine eindeutige ID innerhalb eines Workflows durch den Produzenten vergeben werden. Diese ID und der Workflowname im Tag *SLUBArchiv-externalWorkflow* identifizieren Archivpakete im SLUBArchiv.digital eindeutig (siehe Beschreibung zum Tag *SLUBArchiv-externalWorkflow*). Der Identifier muss persistent sein. Eine automatische Ersetzung eines bestehenden externen Identifiers über die Submission Application ist nicht vorgesehen.

Der externe Identifier darf sich nur aus den Zeichen *a-z0-9_-* zusammensetzen. Großschreibung wird nicht unterstützt.



Dieses Feld ist nicht unbedingt identisch mit dem in der [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493]-Spezifikation angegebenen Schlüssel *External-Identifier*. Es empfiehlt sich, beide identisch zu belegen.

SLUBArchiv-externalIsilId

Dieser Schlüssel sollte von Bibliotheken und andere Gedächtnisorganisationen mit ISIL Nummer genutzt werden. Er erlaubt die Zuordnung von Vorgängen zu Einrichtungen. Weitere Informationen zum Bibliothekssigel finden Sie unter <https://sigel.staatsbibliothek-berlin.de/>. Einrichtungen ohne ISIL-Nummer verwenden das Tag nicht.



Der Schlüssel bezieht sich auf die Organisationseinheit des Produzenten, unabhängig davon ob er im Auftrag Dritter handelt.

SLUBArchiv-exportToArchiveDate

Dieses Tag gibt in ISO-Notation (ISO 8601) das Datum an, an dem das konkrete SIP vom Abliefernden für das SLUBArchiv.digital erstellt wurde. Da ein SIP sowohl für eine erstmalige Übernahme ins Langzeitarchiv als auch für ein Update verwendet werden kann, muss die richtige Reihenfolge der Übernahme von SIPs zu einem Dokument sichergestellt werden. Das Datum in diesem Tag (und nur dieses Datum) regelt, in welcher Reihenfolge mehrere SIPs, die zur gleichen Intellektuellen Einheit (IE) gehören, abgearbeitet werden. Wird das Datum nicht korrekt gesetzt, können durch eine falsche Abarbeitungsreihenfolge der Änderungen falsche Archival Information Packages (AIPs) entstehen. Es ist daher erforderlich, das Datum sekundengenau zu setzen.



Dieses Feld ist nicht unbedingt identisch mit dem in der [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493]-Spezifikation angegebenen Schlüssel *Bagging-Date*. Der Schlüssel *SLUBArchiv-exportToArchiveDate* erlaubt eine sekundengenaue Angabe, die in *Bagging-Date* nicht vorgesehen ist. Es wird empfohlen *Bagging-Date* mit den identischen Jahr-Monat-Tag-Angaben aus dem *SLUBArchiv-exportToArchiveDate* zu belegen.

SLUBArchiv-hasConservationReason

Dieses Feld wird mit *true* belegt, wenn die zu archivierende Intellektuelle Einheit aus Bestandserhaltungsgründen langzeitarchiviert wird. Dies meint, dass es zu dieser IE keine physische oder anderweitige Vorlage (mehr) gibt.

In allen anderen Fällen ist das Feld mit *false* zu belegen.

Dieser Schlüssel kann genutzt werden, um für die betroffenen Dokumente eine höhere Sicherheit zu gewährleisten. Dazu ist eine Festlegung in der Übereinkommenvereinbarung zu treffen.

SLUBArchiv-archivalValueDescription

Dieses Feld beschreibt, **warum** die Intellektuelle Einheit (IE) langzeitarchiviert werden soll. Diese Angabe hilft zu entscheiden, wie bei einer Formatmigration vorgegangen werden soll. Sie enthält eine intellektuelle Bewertung der Daten hinsichtlich ihrer Archivwürdigkeit. Gründe für die Archivwürdigkeit können sich aus gesetzlichen Vorgaben oder aus besonderen kulturellen oder wissenschaftlichen Werten ergeben.



Dieses Feld ist nicht mit dem in der [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493]-Spezifikation angegebenen Schlüssel *External-Description* zu verwechseln, welcher den Inhalt des Bags und dessen Herkunft beschreibt.

SLUBArchiv-rightsVersion

Dieser Schlüssel gibt die Version der SLUBArchiv.digital Rechtemanagement Spezifikation an. Die genaue Beschreibung ist im Dokument **Rechte-Spezifikation** zu finden.

Die XML-Datei mit den SLUBArchiv.digital-Rechteinformationen ist als Tagfile *rights.xml* unter dem Verzeichnis *meta/* abzuspeichern (sh. 2.3 [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493]-Spezifikation).

Die Datei muss in den [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493]-Dateien *tagmanifest-XXXX.txt* hinterlegt sein, wobei XXXX den verwendeten Prüfsummenalgorithmen entspricht (sh. 2.4 [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493]-Spezifikation).

Beispiel

bag-info.txt mit notwendigen Steuerungsdaten

```
SLUBArchiv-sipVersion: v2020.1
SLUBArchiv-exportToArchiveDate: 20160101T120000.00
SLUBArchiv-externalId: 10008
SLUBArchiv-externalIsilId: DE-14
SLUBArchiv-externalWorkflow: kitodo
SLUBArchiv-hasConservationReason: true
SLUBArchiv-archivalValueDescription: Gesetzlicher Auftrag der SLUB Dresden
SLUBArchiv-rightsVersion: 1.0
...
```

Prüfsummen

Als Prüfsummen sind gemäß [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493]-Spezifikation (2.4.) mindestens die beiden folgenden zu verwenden:

- sha-512
- md5

Ebenso ist das Tag *Payload-Oxum* in der *bag-info.txt* zu belegen (sh. 2.2.2, [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493]-Spezifikation).



Die Berechnung der *Payload-Oxum* ist nur für die Dateien im *data/*-Verzeichnis anzuwenden.

Zu archivierende Dateien und zugehörige Prüfsummen

Alle Dateien, die zur eigentlichen IE gehören, sind unterhalb des *data/*-Verzeichnisses abzuspeichern (sh. [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493], Abschnitte "Payload" 2.1.2, 2.1.3)

Ablage Metadaten-Dateien

Es können Metadaten-Dateien, die Zusatzinfos über das IE enthalten, in dem Verzeichnis *meta/* abgelegt werden.



Metadaten, die nicht die IE selbst, sondern Teile innerhalb einer IE beschreiben, sind nicht Bestandteil des *bag-info.txt* oder der Metadaten-Dateien unter *meta/*,

sondern müssen zur IE selbst hinzugerechnet werden. In diesem Fall ist zu beachten, dass

- diese Metadaten das Kriterium der Archivwürdigkeit erfüllen (Stichwort: Signifikante Eigenschaften)
- diese Metadaten-Dateien ebenfalls in einem langzeitarchivfähigem Dateiformat beschrieben sind
- im Fall einer Formatmigration durch das SLUBArchiv.digital diese Metadaten-Dateien einer Anpassung bedürfen und diese geregelt sein muss (Stichwort: Führendes System)

Dies gilt zB. für Strukturierungs- und Paginierungsinformationen.

Dazu können auch mehrere Metadaten-Dateien genutzt werden. Zulässig sind dabei nur standardisierte, öffentlich zugängliche und validierbare Metadatenformate.

Diese Dateien werden vom Archiv validiert, darüber hinaus **nicht** ausgewertet (maßgebend bleibt die [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493]-Struktur). Sie gehören aber zu einem vollständigen Ablieferungspaket und erlauben den Erhalt der logischen Verknüpfung zwischen den zu archivierenden Nutzdaten und den zugehörigen Metainformationen außerhalb des SLUBArchiv.digital.

Sie sind analog zu den SLUBArchiv.digital-Rechteinformationen (sh. [SLUBArchiv-rightsVersion](#)) in den Tag-Manifest-Dateien zu referenzieren.



Diese XML-Dateien sollten möglichst einfach nach ihrem verwendeten Metadaten-Schema benannt werden. Für MODS-Dateien empfiehlt sich *mods.xml*, für LIDO *lido.xml* und für Dublin Core *dc.xml*. Diese Dateinamen werden **nicht** archiviert. Die Inhalte dieser Metadaten-Dateien werden im SLUB-Archiv in speziellen Datenfeldern abgelegt und können bei der Konstruktion von Abgabepaketen (DIP) abgefragt werden.



Es steht dem Produzenten frei, in der *bag-info.txt* die Informationen zu hinterlegen, die eine Interpretation dieser Metadaten erleichtern.



Der Dateiname *rights.xml* ist reserviert für die Rechteausszeichnung nach der SLUB-Spezifikation "Rechtekodierung".



Das Verzeichnis *meta/* ist optional. Wird es verwendet, so **müssen** seine Dateien in den *tagmanifest-XXX.txt*-Dateien mit den korrekten Prüfsummen hinterlegt sein.

Weitere Hinweise

Die BagIt-Spezifikation erlaubt an einigen Stellen optionale Elemente. Im Folgenden sind diejenigen aufgelistet, die für die Abgabe an das SLUBArchiv.digital verpflichtend sind:

- Keine Verwendung von unvollständigen BagIts mit Fetch-Option (sh. 2.2.3, [BagIt](https://tools.ietf.org/html/rfc8493) [https://tools.ietf.org/html/rfc8493]-Spezifikation)
- *Tag-File-Character-Encoding* in der *bagit.txt*-Datei ist immer "UTF-8"

- Alle Tag-Dateien sind immer als UTF-8 (ohne BOM) zu kodieren.
- Dateipfade dürfen **keine** Leerzeichen enthalten.
- Die Tag-Manifest-Dateien sind verpflichtend und müssen jeweils die gleichen Tag-Dateien beschreiben.
- Die folgenden Schlüssel-Werte-Paare der Bag-Metadata-Datei *bag-info.txt* sind zu belegen:
 - *Bag-Size*
 - *Payload-Oxum*
- Da ein SIP genau ein IE enthält, dürfen die folgenden Schlüssel-Werte-Paare der Bag-Metadata-Datei *bag-info.txt* nicht benutzt werden:
 - *Bag-Count*
 - *Bag-Group-Identifizier*

Komprimierte SIPs

Komprimierte SIPs werden **nicht** unterstützt. Hintergrund ist, dass die Verarbeitung komprimierter SIPs aufwendiger und fehleranfälliger ist.

AIP Update

Das SIP für ein AIP Update unterscheidet sich nicht wesentlich von einem SIP für einen Erstingest. Ein SIP für ein AIP Update ist ein SIP, dessen *SLUBArchiv-externalWorkflow* und *SLUBArchiv-externalId* auf ein bereits im SLUBArchiv.digital vorhandenes IE verweist.



Im Falle eines AIP Update sind immer **alle** Metadaten im SIP zu kodieren!

Die ursprünglichen Metadaten werden im AIP vollständig ersetzt.

Die Besonderheiten werden nachfolgend erläutert.

Metadaten Update

Das Metadaten-Update umfasst ausschließlich Änderungen der in der *bag-info.txt* kodierten Information oder Änderungen der in *meta/* liegenden Metadaten-Dateien, die das im Archiv hinterlegte AIP aktualisieren.

Diese SIPs dürfen daher keine Dateien im *data/*-Verzeichnis enthalten.



Auch wenn das Bag keine Dateien im *data/* enthalten darf, **muss** das Verzeichnis *data/* existieren. Gleiches gilt für die *manifest*-Dateien, die dann leer sind.



Das Verzeichnis *meta/* ist optional. Wird es verwendet, **müssen** seine Dateien in den *tagmanifest-XXX.txt*-Dateien mit den korrekten Prüfsummen hinterlegt sein.

Um sicherzustellen, dass keine Inkonsistenzen auftreten können, dürfen die in *meta/* hinterlegten Metadaten-Dateien **keine** Referenzen auf Dateien des Archivs bzw. auf Dateien, die in vorherigen Versionen des SIPs übermittelt wurden, enthalten. Andernfalls ist ein [Volles Update](#) durchzuführen.

Für ein tieferes Verständnis ist die Handreichung der SLUB "Workflow-Spezifikation" zu Rate zu ziehen, die die verschiedenen Optionen und deren Konsequenzen behandelt.

Volles Update

Das SIP enthält alle zum IE gehörenden Dateien in der neuen Fassung. Das SLUBArchiv.digital ordnet die im SIP referenzierten Dateien dem vorhandenen AIP zu und erzeugt über den AIP-Update-Prozess eine neue Version des AIP.

Das SLUB-Archiv erkennt über die in *bag-info.txt* angegebenen Schlüssel *SLUBArchiv-externalWorkflow* und *SLUBArchiv-externalId*, welches AIP im Archiv aktualisiert werden muss. Die Reihenfolge der Updates ist durch den Schlüssel *SLUBArchiv-exportToArchiveDate* sichergestellt.

Tooling

Neben den bereits in [Warum der Wechsel auf BagIt?](#) besprochen Tools und Bibliotheken, die programmgestützte Arbeit mit dem BagIt-Format ermöglichen, stellt das SLUBArchiv.digital eine mögliche Implementierung eines SIP Builders im öffentlichen [gitlab](https://git.slub-dresden.de/public) [https://git.slub-dresden.de/public] der SLUB zur Verfügung ^[4]. Es ist ein Angebot an alle Produzenten des Archivs zur direkten Nutzung oder als Referenz für eine eigene Umsetzung nach Spezifikation.

Teil 2: Grundsätze zur SIP-Bildung am Beispiel

Eine retrodigitalisierte Lokalzeitung von 1932

Anbei Aufbau und Metadaten einer kleinen SIP aus der Tagesproduktion der SLUB Dresden. Auf die Abbildung der Digitalisate und deren OCR wurde aus Relevanz- und rechtlichen Gründen verzichtet.

SIP-Verzeichnisstruktur

481463_1634296082	SIP-Verzeichnis (Produzent)
├─ bag-info.txt	Metadatendatei (BagIt + SLUB + Produzent)
├─ bagit.txt	Metadatendatei (BagIt)
├─ data	Payload-Verzeichnis (BagIt) mit Verzeichnisstruktur des Produzenten
│ └─ images	
│ └─ scans_tif	(Scans der Zeitungsseiten)
│ └─ 00000001.tif	
│ └─ 00000002.tif	
│ └─ 00000003.tif	
│ └─ 00000004.tif	
│ └─ 00000005.tif	
│ └─ 00000006.tif	
│ └─ 00000007.tif	
│ └─ 00000008.tif	
│ └─ ocr	
│ └─ alto	(OCR der Zeitungsseiten)
│ └─ 00000001.xml	
│ └─ 00000002.xml	
│ └─ 00000003.xml	
│ └─ 00000004.xml	
│ └─ 00000005.xml	
│ └─ 00000006.xml	
│ └─ 00000007.xml	
│ └─ 00000008.xml	
├─ manifest-md5.txt	Prüfsummendatei (BagIt)
├─ manifest-sha512.txt	Prüfsummendatei (BagIt)
├─ meta	Metadatenverzeichnis (SLUB)
│ └─ rights.xml	Metadatendatei für Rechte (SLUB)
├─ tagmanifest-md5.txt	Prüfsummendatei (BagIt)
└─ tagmanifest-sha512.txt	Prüfsummendatei (BagIt)

6 Verzeichnisse, 23 Dateien

bag-info.txt

```
Source-Organization: SLUB Dresden
Organization-Address: kitodo@slub-dresden.de
External-Identifizier: 481463
External-Identifizier: http://digital.slub-dresden.de/id480533490-19320111
External-Identifizier: urn:nbn:de:bsz:14-db-id480533490-193201117
Title: Sächsische Dorfzeitung und Elbgaupresse / 1932-01-11
SLUBArchiv-sipVersion: v2020.1
SLUBArchiv-externalWorkflow: kitodo
SLUBArchiv-externalId: 481463
SLUBArchiv-externalIsilId: DE-14
```

SLUBArchiv-exportToArchiveDate: 2021-10-15T13:08:02+02:00
SLUBArchiv-hasConservationReason: false
SLUBArchiv-archivalValueDescription: Archivierung erfolgt laut gesetzlichem Auftrag der SLUB Dresden.
SLUBArchiv-rightsVersion: 1.0
DateIssued: 1932-01-11
Payload-Oxum: 262562406.16
Bag-Size: 250.40 MB
Bagging-Date: 2021-10-15

bagit.txt

BagIt-Version: 1.0
Tag-File-Character-Encoding: UTF-8

manifest-md5.txt

1ea3485706ced32145aa599564b8b6e7 **data/images/scans_tif/00000001.tif**
86574284f6993a6fcb82e5a564d86c19 **data/images/scans_tif/00000002.tif**
411a45e55773e53f60461fd7628d0c5c **data/images/scans_tif/00000003.tif**
4bd6c71ab678df4fd8ae3256ee1e646b **data/images/scans_tif/00000004.tif**
baa248e60bc284aa814401540bc043e7 **data/images/scans_tif/00000005.tif**
187fd86dbe5cd889e72540776930c6a9 **data/images/scans_tif/00000006.tif**
8fd7523295b01c6a33e93950cdf16f51 **data/images/scans_tif/00000007.tif**
9db770fdd6348d673a87e06e703c27ad **data/images/scans_tif/00000008.tif**
7409710232b25496cd4e635e507398e4 **data/ocr/alto/00000001.xml**
be8ae2b8570b7a64f28edb5ac0e295e7 **data/ocr/alto/00000002.xml**
9d3b1c1512e339ef0a1e9e26e60e75d7 **data/ocr/alto/00000003.xml**
6f7878606e44802fd432e0004b5effda **data/ocr/alto/00000004.xml**
b3d6ed964351e7cf5c896f1b48f3eae1 **data/ocr/alto/00000005.xml**
6cb631f0940a786f8943c7369660792d **data/ocr/alto/00000006.xml**
ced06effbce4d4d3070bc910dc807a94 **data/ocr/alto/00000007.xml**
597845dc2c98c3daf3457366a4359e9f **data/ocr/alto/00000008.xml**

manifest-sha512.txt

80d8d839a6c4e2cb2d2a7e9fd6b31a9c3fbba31e5043abb1bf1332285a25710c8af4b41eabb34042a1f3fd0b0
bd3a12fc2766cc95fbea3dfba54598a1a5d9e68 **data/images/scans_tif/00000001.tif**
91d69ca22e0da903941b5e6af052f4ea47812bc799df492cffbd653a3690524236485e86a9f5e91e4935c9f99
e37c4cfbf9f508563caa665c167cf0333e18be0 **data/images/scans_tif/00000002.tif**
383c63f734884965ff39ff0e3bacab62173e075d8196355a367d909e7039011912c2c91038e020be6487a4e1
144e717243af736e28e67f7dbf375c2dc71df28 **data/images/scans_tif/00000003.tif**
35a60ad3ce253dcf9329aa14e3483fac5bbe117f860e72b67cea04bfb8f99a7694352a60346e2f83c1da4c671
2125b8291e43bbec1cafe803cd05bf5a5ebbdba **data/images/scans_tif/00000004.tif**
1826e6148ab3697b3af001125a5da20b14793e1c1675ef447ce83a6009c09dd3165e84c48756c87f8ab8b5eba
a947e6982dcc1719cb0c1951a9fd4ed9955365b **data/images/scans_tif/00000005.tif**
2fd78a9671bbd777c4de3909280cb4aecab234a9ea9fd52449a414e6a0769d1f9bff0b972488bf8893dd169b3
1a86709d2ed9bf64234c7e17f06fbc8bafaa245d **data/images/scans_tif/00000006.tif**
7f7479836d339e57bd007083e56210da964dc2f79eb3e5be34ca75cd11e00c7e41914e2593efe38b6e2c26051
676ce859a0b7e2247ce7acfa986248dc6fd465b **data/images/scans_tif/00000007.tif**
684573ccc7c2af7445d14a77b7b4d979db9b3401857893bf240c4175ca94f4eb9e2bdc857c99764088dcf9c7e
2c8a53a4820f4af1b5475831f23e0bab945cda5 **data/images/scans_tif/00000008.tif**
357f8bd764b842359562549bdad90b9322432e1d6d552914d89dbb58e4648e6ca75f92c1cf0250d6aa065e0c0
ea1a45f4e7961681d41fc7a6c73df571ca702bd **data/ocr/alto/00000001.xml**
371af7cfd6d2a4dba09610e3a8b3499e18410f5bfc38eaad262292fc0df02d0c0e75e72c3da42f63ed2ec9415
856e071be7cd5f6ba7c50d9abb5de2936c926d6 **data/ocr/alto/00000002.xml**
a7e2a2b06454443427ffb0d72e1da877aeb4c2605e3661ba7e26facd2a025f4e1165a6f55e4588dfd7f1acba1
436c3bdd516bf5d535e7e48335d18a28e0359c4 **data/ocr/alto/00000003.xml**
82c1663bbd9e471e1aceb9c2f292cd88ac32bd953c9016d07b100f4dc3f4a5e75439420c6825bf4693568376e

```
c4de77947380b0cef2ea80d8fb647abbcddf9aa data/ocr/alto/00000004.xml
c7082e3420670111843ac7472c75457b8ff17e336b071d15d067e48e9e313be40bcf42269d04d8fe0573398cc
043b7dae4574c26906f63feec3f309150492f84 data/ocr/alto/00000005.xml
10304ca8e7179266f0140d64a44076e8eff5181b1743ff4880c924d563305bb45e580fcc86ac3d88db0b9858c
e33ac0462b22f62be452873c2235167c45f8372 data/ocr/alto/00000006.xml
cda069efd38408291824d527da27d9051fb519c43ed0c6cd298015659a03328edf0e4b94246fba2518c89d19d
633546c53a4d1d6abc113adf3a4984ff51e0971 data/ocr/alto/00000007.xml
0a86c2f6417f62cd0a43a37ce31bf93845f2e51df7089a03ecd8f54c5d0da652a554f3c462677ee7a2841dd8a
768b1371410f78255404cc1d59efb459765eef7 data/ocr/alto/00000008.xml
```

meta/rights.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<slubarchiv:rightsRecord xmlns:slubarchiv="http://slubarchiv.slub-dresden.de/rights1"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://slubarchiv.slub-dresden.de/rights1 https://slubarchiv.slub-
dresden.de/slubarchiv/standards/rights/rights1.xsd">
  <slubarchiv:copyrightStatus>undefined</slubarchiv:copyrightStatus>
</slubarchiv:rightsRecord>
```

tagmanifest-md5.txt

```
476abf06e024baf7e9350a7f4df574cd bag-info.txt
eaa2c609ff6371712f623f5531945b44 bagit.txt
61115304ffbc8191dbd78da5038c3ab9 manifest-md5.txt
002486af8c59695a5d5f4c428b0c2f00 manifest-sha512.txt
5570cf1e163e02aed4560c68a960ee59 meta/rights.xml
```

tagmanifest-sha512.txt

```
23ea1c533d0921badd466c742707ad5df4650141c4855a7db7fb6d41164db22527790334fc13036b602ffb808
61d049ca0ac6a1ca2cec26e3c32b80e607349ff bag-info.txt
1d73ae108d4109b61f56698a5e19ee1f8947bdf8940bbce6adbe5e0940c2363caace6a547b4f1b3ec6a4fd2b7
fa845e9cb9d28823bc72c59971718bb26f2fbd8 bagit.txt
fe200ace598f308f48d51c49c4550c8f6774c5141aaf166ebbb606530f0252945084ae03f6e81feaeaa9b08bd
b4139fad19056adf76baa9d41a3d6bdfa41852a manifest-md5.txt
b14e88f04bd1585c4bac234d2237788ecac6ee6092f9c937ba2a2de76719fca97397289aed82631422f97040e
f1cb11b081e144474c69f18fc3ca724732842d8 manifest-sha512.txt
ac5c1049fef35c4d5d99fca9f15087e3e010a94f67402c5c5c093ff17eed14313b28c80e745259b6e6138933
861685caafb501adb62bfac3d4dd39340bf76f4 meta/rights.xml
```

1. Allgemeine Grundsätze

Grundsatz 1.1

Es MUSS möglich sein, beliebige digitale Objekte und Metadaten in ein Informationspaket aufzunehmen.

- ☑ Erfüllt für das SIP selbst, ABER nicht alle Formate sind archivfähig (im Sinne von *content preservation*) und können vom Archiv dennoch abgelehnt werden. Geführt wird dazu eine Liste langzeitarchivfähiger Dateiformate sowie Handreichungen zu einzelnen Formaten.

Grundsatz 1.2

Das Informationspaket DARF NICHT die Mittel, Methoden oder Werkzeuge für den Ingest

einschränken.

- Voll erfüllt für das SIP, in einer separaten Workflow-Spezifikation werden erst Methoden der Übergabe definiert.

Grundsatz 1.3

Das Paketformat DARF NICHT den logisch-inhaltlichen Umfang der digitalen Objekte und Metadaten definieren, die ein Informationspaket bilden.

- Eingeschränkt, zur Vereinfachung der Verarbeitung für Mensch und Maschine gilt die Festlegung 1 SIP = 1 IE.

Grundsatz 1.4

Das Informationspaket MUSS skalierbar sein.

- Erfüllt, produktive SIPs des Archivs reichen aktuell von KBytes bis in den TByte-Bereich. Tausende Dateien sind möglich.

Grundsatz 1.5

Das Informationspaket MUSS maschinenlesbar und automatisierbar zu verarbeiten sein.

- Voll erfüllt, das Archiv verarbeitet die BagIt-basierten SIPs bereits vollautomatisch.

Grundsatz 1.6

Das Informationspaket MUSS interpretierbar sein, um eine auch für den Menschen inhaltliche Deutung zu ermöglichen.

- Voll erfüllt, BagIt ist gut menschenlesbar und die geforderte Rechteauszeichnung verwendet ein einfaches, öffentlich verfügbares XML-Schema ^[5].

Grundsatz 1.7

Die Spezifikation des Informationspakets MUSS offen und frei sein.

- Voll erfüllt, alle Spezifikationen sind offen und frei verfügbar ^[1].

Grundsatz 1.8

Die Komplexität der Spezifikation eines Informationspakets SOLL angemessen sein.

- Erfüllt, die *SIP Spezifikation für automatischen Ingest SLUBArchiv (v2.0.1)* umfasst 15 A4 Seiten (effektiv 8 Seiten mit Inhalten relevant zum Aufbau) ^[1].

2. Grundsätze zur Identifikation eines IP

Grundsatz 2.1

Jedes Informationspaket MUSS einen im archivierenden Archiv eindeutigen und dauerhaften Identifikator haben oder erhalten.

- Erfüllt, die Kombination einiger verpflichtenden Schlüssel-Wert-Paaren in BagIt Metadaten ist für den jeweiligen Workflow eindeutig. In der Weiterverarbeitung wird daraus ein eindeutiger und dauerhafter Identifikator für das Archiv erstellt.

Grundsatz 2.2

Jedes Informationspaket SOLL einen Identifikator besitzen, der global eindeutig und dauerhaft ist.

- Optional bzw. nur in Zusatzvereinbarungen explizit verlangt. Die Verwendung von global eindeutigen und dauerhafte Identifikatoren (jenseits des Archivs) ist in den beschreibenden Metadaten nicht vorgeschrieben und wird nur in Beispielen kurz aufgeführt.

Grundsatz 2.3

Alle Teile eines Informationspakets SOLLEN einen eindeutigen und dauerhaften Identifikator haben.

- Erfüllt auf Basis der Dateipfade, diese sind eindeutig und dauerhaft innerhalb des Bags.

3. Struktur eines IP

Grundsatz 3.1

Das Informationspaket MUSS sicherstellen, dass Daten und Metadaten logisch voneinander getrennt sind.

- Voll erfüllt, es existiert eine eindeutige Trennung in Payload und Metadaten.

Grundsatz 3.2

Die Struktur des Informationspakets SOLL die Trennung verschiedener Arten von Metadaten ermöglichen.

- Erfüllt und steht dem Produzenten frei. Es gibt keine Einschränkungen oder Vorgaben, die es unterbinden.

Grundsatz 3.3

Die Struktur des Informationspakets SOLL die Erstellung von Daten und Metadaten in mehreren Repräsentationen ermöglichen.

- Erfüllt, keine Einschränkungen oder Vorgaben. Es obliegt Produzenten.

Grundsatz 3.4

Die Struktur des Informationspakets SOLL die Möglichkeiten zum Hinzufügen zusätzlicher Daten zum Informationspaket explizit definieren.

- Erfüllt, möglich per Voll-Update oder reinem Metadaten-Update.

Grundsatz 3.5

Jedes Informationspaket SOLL seinen Informationstypen mitteilen.

- Nicht erfüllt, für das Archiv implizit über LZA-Workflow und Abgabestelle.

4. Metadaten eines IP

Grundsatz 4.1

Metadaten im Informationspaket SOLLEN einem etablierten Standard entsprechen.

- Voll erfüllt, das Archiv fordert standardisierte, öffentlich zugängliche und validierbare Metadatenformate.

Grundsatz 4.2

Die exakte Verwendung der Metadaten SOLLTE in Profilen für Informationstypen erarbeitet werden.

- Für alle gleich, Ausnahmen bilden seltene projektspezifische Zusatzvereinbarungen außerhalb der SIP-Spezifikation.

Grundsatz 4.3

Jedes Informationspaket KANN beschreibende Metadaten enthalten.

- Erfüllt, beschreibende Metadaten werden explizit empfohlen, sind aber keine Pflicht.

5. Authentizität und Integrität eines IP

Grundsatz 5.1

Im Informationspaket SOLLEN Möglichkeiten enthalten sein, die Authentizität sicherzustellen.

- Grundsätzlich möglich, aber nicht spezifiziert oder aktuell notwendig. Sicherstellung erfolgt bereits per Übergabe-Workflow mit PKI.

Grundsatz 5.2

Im Informationspaket SOLLEN Möglichkeiten enthalten sein, die Integrität sicherzustellen.

- ☑ Voll erfüllt, die Spezifikation fordert ein Bag mit mindestens 2 Prüfsummenalgorithmen (MD5 & SHA512).

[1] Die genannten Dokumente sind auf der Webseite des SLUBArchiv.digital unter <https://slubarchiv.slub-dresden.de/technische-standards-fuer-die-ablieferung-von-digitalen-dokumenten/> veröffentlicht und dort, technisch bedingt, spezifischer benannt.

[2] NESTOR MATERIALIEN 23: Grundsätze zur SIP-Bildung <https://d-nb.info/1214014216/34>

[3] NESTOR THEMA 13: BagIt <https://d-nb.info/1216436185/34>

[4] SLUB SIP Builder Projekt: https://git.slub-dresden.de/digital-preservation/SLUB_SIP_Builder

[5] SLUBArchiv.digital Rechtenschema (1.0): <https://slubarchiv.slub-dresden.de/slubarchiv/standards/rights/rights1.xsd>

SIP-Ausprägungen der DIMAG-Ingest-Module

Verfasst vom DIMAG-Verbund für die nestor-AG SIP-Konkretisierung im November 2022

1. Einleitung

Es gibt im DIMAG-Verbund keine einheitliche SIP-Definition. Die OAIS-SIP-Definition wird lediglich als Konzept verwendet und ist in den DIMAG-Modulen und Werkzeugen, die im Ingest eingesetzt werden, in der Ausprägung "Lieferpaket" und "Ingestpaket" realisiert. Aus einem SIP können grundsätzlich beliebig viele AIPs entstehen. Darüber hinaus gibt es im DIMAG-Verbund den Oberbegriff des "Transportpakets" für den Datentransfer zwischen den Modulen.

Diese Paketdefinitionen definieren jeweils nur diejenigen Aspekte eines Pakets, die im Kontext erforderlich sind. Das Lieferpaket beispielsweise hat keine Definition der endgültigen Abgrenzung in AIPs, weil dies an diesem Moment im Prozess noch nicht erforderlich ist.

Insgesamt werden im DIMAG-Verbund Informationspakete nicht zwingend als gekapselte digitale Objekte (wie z.B. eine XML-Datei oder ein ZIP-Container) verstanden, sondern als konzeptuelle Objekte aus Metadaten und Primärdaten, deren Zusammenhalt definiert ist (wie z.B. ein PDF mit URL und Basismetadaten auf einem Storage-Server X, mit URN auf Server Y und Katalogmetadaten auf Server Z). Kapselungen in Containerformaten werden eingesetzt, soweit sie z.B. im Ingest erforderlich sind.

2 Lieferpaket

Gesamtheit aus Primär- und Metadaten, die von der abgebenden Stelle an das Archiv übermittelt werden. Es enthält mindestens Primärdaten und Strukturinformationen. Das Lieferpaket entspricht dabei in der Regel einem Zugang.¹

Ein Lieferpaket setzt sich aus folgenden Bestandteilen zusammen:

- Lieferungsmetadaten (optional)
- Primärobjekte (immer vorhanden), Format und Form werden zwischen Archiv und

¹ <https://dimag-wiki.la-bw.de/xwiki/bin/view/Main/DIMAG%20Terminologie/>.

anbietender Stelle abgestimmt. Dabei kann es sich sowohl um strukturierte Daten wie E-Akten und Fachverfahrens-Exporte als auch um Dateisammlungen handeln.

2.1 Ausprägung beim DIMAG-Ingest-Werkzeug Akte (DIWA)

Ein DIWA-Lieferpaket setzt sich aus folgenden Bestandteilen zusammen:

- Lieferungsmetadaten in Form von xdomea-0503-Nachrichten. Unterstützt werden derzeit die xdomea-Versionen 2.1 und 2.3. Der Name der xdomea-Nachricht wird nach dem Muster gebildet: ProzessID_Aussonderung.Aussonderung.0503.xml.
- 1-n Primärdokumente, die in der xdomea-Nachricht referenziert sind. Die Namen der Primärdokumente werden nach dem Muster gebildet: UUID_Dateiname.pdf

Das Lieferpaket wird als ZIP-Datei übergeben. Der Name der ZIP-Datei wird nach dem Muster gebildet: ProzessID_Aussonderung.Aussonderung.0503.zip.

2.2 Ausprägung bei DIMAG-Website-Ingest (DIWI)

SIPs werden von DIWI im Format einer WARC-Datei mit einzelnen Metadaten aus dem Heritrix-Abholprozess verarbeitet. DIWI sieht keine Änderungen und keine Aufteilungen dieser SIPs vor. Liefer- und Ingestpaket und das AIP sind aus Konzept Sicht stets identisch.

2.3 Ausprägung beim DIMAG-Übergabeassistenten

Der Übergabeassistent wird für den sicheren Datentransfer zwischen abgebender Stelle und Archiv eingesetzt. Er erstellt ein Lieferpaket im BagIt-Format (siehe 4 Transportpaket). Dieses setzt sich aus folgenden Bestandteilen zusammen:

- Lieferungsmetadaten (Informationen der abgebenden Stelle wie z.B. Ansprechpartner, Kontaktdaten, Aktenzeichen, Kurzbeschreibung; Dateilisten mit Hashwerten)
- 1-n Primärobjekte

2.4 Ausprägung beim DIMAG-IngestProzessModul (IPM)

Das IngestProzessModul wird zur Verwaltung, Steuerung und Eingangsbearbeitung der Lieferpakete im Archiv eingesetzt. Die Eingangsbearbeitung umfasst u.a. die Sicherung, die Integritäts- und Vollständigkeitsprüfung, Entpacken und Entschlüsseln des Lieferpakets und das Entpacken, die Formaterkennung- und -validierung, die Strukturvalidierung und Normierung der Primärobjekte.

Die Lieferpakete werden im IPM in einem virtuellen Dateisystem abgelegt und nach Abschluss des Ingests ins Kernmodul gelöscht. Das Lieferpaket im IPM umfasst die folgenden Bestandteile:

- Lieferpaket-Metadaten: Informationen zum Lieferpaket (wie z.B. Name, Größe, Beschreibung), zur abgebenden Stelle, zur Bearbeitung und zur Fehlerbehandlung
- Primärobjekt-Metadaten: Grundinformationen zum Primärobjekt (wie z.B. Original Dateipfad, Format), Informationen zur Bearbeitung und zu den durchgeführten Prozessen
- 1-n Primärobjekte

3 Ingestpaket

Gesamtheit aus Primär- und Metadaten im [Funktionsbereich] Ingest. Im Unterschied zum **Lieferpaket** können, sofern der jeweilige Ingestprozess dies erfordert, Veränderungen an den Primär- und Metadaten im Rahmen des Ingest vorgenommen werden. Diese werden dokumentiert.²

Beziehungen: 1 Lieferpaket : 1-n Ingestpakete; 1 Ingestpaket : 1-n AIP

3.1 Ausprägung bei DIWA

DIWA stellt das Ingestpaket auf der Laderampe des Kernmoduls bereit und stößt per SOAP-Call den Ingest ins Kernmodul an. Das DIWA-Ingestpaket setzt sich aus den folgenden Bestandteilen zusammen:

- DIWA-Protokolldatei
- xdomea-0503-Aussonderungsnachricht
- DIMAG-Controlfile
- 1-n Primärobjekte

3.2 Ausprägung bei DIWI

DIWI erstellt eine LoadXML-Datei mit der Verzeichnungsinformation, stellt das Ingestpaket auf der Laderampe des Kernmoduls bereit und stößt per SOAP-Call den Ingest ins Kernmodul und Archivinformationssystem an. Das Ingestpaket setzt sich aus den folgenden Bestandteilen zusammen:

- LoadXML-Datei
- 1 Primärobjekt (WARC-Datei)

Das AIP wird vom DIMAG-Kernmodul nach der in der LoadXML-Datei definierten Struktur angelegt.

3.3 Ausprägung beim IPM

Das Ingestprozessmodul teilt Lieferpakete auf Grund der Größe, der Dateianzahl oder aus Bearbeitungsgründen auf 1-n Ingestpakete auf. Das IPM verwaltet und steuert die weitere Bearbeitung der Ingestpakete im Archiv bis zur Übernahme ins DIMAG-Kernmodul.

Ingestpakete können um weitere im Archiv bearbeitete oder erstellte Primärobjekte angereichert werden. Primärdateien können aus archivfachlichen oder technischen Gründen nachkassiert werden. Beide Eingriffe werden dokumentiert und begründet.

Ingestpakete können zur Vorbereitung der AIP-Bildung vorstrukturiert werden. Dies ist erforderlich, da mit dem DIMAG-IngestTool nur eine regelbasierte Paketierung möglich ist.

² Vgl. vorige Anm.

Das Ingestpaket im IPM umfasst die folgenden Bestandteile:

- Lieferpaket-Metadaten: Grundinformationen zum übergeordneten Lieferpaket (wie z.B. Name, Größe, Beschreibung), zur abgebenden Stelle, zur Bearbeitung und zur Fehlerbehandlung
- Ingestpaket-Metadaten: Grundinformationen zum Ingestpaket (wie z.B. Name, Größe, Beschreibung), zur Bearbeitung und zur Fehlerbehandlung
- Primärobjekt-Metadaten: Grundinformationen zum zugeordneten Primärobjekt (wie z.B. Original Dateipfad, Format), Informationen zur Bearbeitung und zu den durchgeführten Prozessen
- 1-n Primärobjekte

Das IPM gibt zur weiteren Verarbeitung durch das IngestTool das Ingestpaket im BagIt-Format aus. Dabei wird die archivische Vorstrukturierung auf dem Dateisystem umgesetzt. Die vorarchivische Ordnungsstruktur der anbietenden Stelle bleibt in der Protokolldatei und im Speicherpfad des Primärobjekts erhalten.

Das Ingestpaket zur Weitergabe ans DIMAG IngestTool (DIT, vgl. Folgekapitel) setzt sich aus folgenden Bestandteilen zusammen:

- IPM-Protokolldatei (mit Lieferpaket-, Ingestpaket- und Primärobjekt-Metadaten)
- Dateilisten mit Hashwerten
- 1-n Primärobjekte

3.4 Ausprägung beim DIMAG-IngestTool (DIT)

Das DIT wird zur regelbasierten Paketierung der AIPs, zum Metadaten-Mapping für Verzeichnungsinformationen im Kernmodul und im Archivinformationssystem und zum Anstoßen des Ingest ins DIMAG-Kernmodul eingesetzt. Die Paketierungs- und Verzeichnungsinformationen werden in einem Controlfile abgelegt.

Das DIT stellt das Ingestpaket auf der Laderampe des Kernmoduls bereit und stößt per SOAP-Call den Ingest ins Kernmodul und Archivinformationssystem an. Das Ingestpaket setzt sich aus den folgenden Bestandteilen zusammen:

- Controlfile (LoadXML-Format)
- 1-n Primärobjekte

Die Weitergabe der IPM-Protokolldatei und ihre Verarbeitung im DIMAG-Kernmodul befindet sich derzeit in der Abstimmung.

Das AIP oder die AIPs werden vom DIMAG-Kernmodul nach der im Controlfile definierten Struktur angelegt.

4 Transportpaket

Ein Transportpaket bezeichnet eine Menge von Dateien für den Transfer zwischen den Modulen.³ Sowohl SIPs als auch AIPs als auch DIPs können folglich in der Form eines Transportpakets übermittelt werden.

Es ist geplant, dass für den Datentransfer zwischen den DIMAG-Modulen und DIMAG-Werkzeugen zukünftig das DIMAG-BagIt verwendet wird. Dieses wurde erstmalig für den DIMAG-Übergabeassistenten und das -Ingestprozessmodul realisiert. Ein valides *DIMAG*-Transportpaket ist ein nach [RFC8493](#) valides *Bag*.

³ Vgl. erste Anm.