



SIP-Ausprägungen der DIMAG-Ingest-Module

Verfasst vom DIMAG-Verbund

Stand: 25. Mai 2023

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 dieser Stelle 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.

Es ist Aufgabe des jeweiligen Archivs, Form und Format des SIP mit den anbieterpflichtigen Stellen abzustimmen. Diese SIP-Festlegungen müssen die Rahmenvorgaben der für die Weiterverarbeitung benutzten DIMAG-Ingest-Module einhalten.

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 den folgenden Bestandteilen zusammen:

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

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 den 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

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

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 ins DIMAG-Kernmodul 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

² Vgl. vorige Anm.

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.

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 den 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.