

Rahmenspezifikation für die digitale Archivierung (Projekt *bentō*)

Inhalt

1	Einführung	1
2	AIP	2
3	SIP	3
4	Descriptive Information	6
5	DIP	7
6	Anhänge	9

1 Einführung

Dieses Dokument – Hauptresultat des Projekts *bentō* – umfasst Spezifikationen bzw. Frameworks für die vier zentralen Schnittstellen der digitalen Archivierung gemäss der Terminologie des OAIS¹: AIP, SIP, DIP und *Descriptive Information*. Es handelt sich um eine Konkretisierung der *Minimalanforderungen an die digitale Archivierung*, welche von der Aufsichtskommission der KOST im September 2009 als Leitfaden und Roadmap verabschiedet wurden².

Bentō umfasst aus Sicht des OAIS-Modells die folgenden Bereiche: *Ingest*, *Data Management*, *Archival Storage* und *Access*. Den Inhalt dieser Bereiche präzisieren wir im Kontext von *bentō* folgendermassen:

- An der *Ingest*-Schnittstelle zum Datenproduzenten wird kontrolliert, dass die Datenlieferung in Form eines definierten SIPs, angereichert mit den archivisch notwendigen Metadaten, stattfindet. Das SIP ist hier Schnittstelle zum Datenproduzenten und Datenmodell für die abgelieferten Daten.
- Der *Ingest*-Prozess generiert aus dem SIP ein oder mehrere AIPs für die Langzeitspeicherung. Das AIP ist hier das Datenmodell für eine gebündelte autonome Archiveinheit.
- Im *Ingest*-Prozess wird auch die *Descriptive Information* für das *Data Management* generiert. Ein *Data Management Service* ist in Form eines Archivinformationssystemes bzw. eines elektronischen Findmittels in allen Archiven bereits vorhanden.
- *Archival Storage* empfängt die generierten AIPs und gewährleistet die *Bitstream Preservation*. *Archival Storage* ist bereits im KOST-Projekt *arcun* als Service beschrieben³ und wird deshalb im Rahmen von *bentō* nicht weiter detailliert.

- Der *Access*-Prozess erlaubt den kombinierten Zugriff auf AIP und *Descriptive Information*. Für die Benutzung wird das AIP ähnlich wie im *Ingest*-Prozess transformiert und in ein DIP umgewandelt

Ausserhalb des Rahmens von *bentō* verbleiben die OAIS-Funktionsbereiche *Preservation Planning* und *Administration*. Diese sind im OAIS-Modell nur indirekt mit den Datenstrukturen und Schnittstellen der digitalen Archivierung verbunden. Ebenfalls nicht behandelt werden die Vermittlung (Suchen, Finden und Anschauen von Archivobjekten mit einem geeigneten Findmittel) und die *Access Control* (im Sinne von Kontrolle der Zugriffsberechtigungen). Diese Aufgaben werden heute in der Regel von den Archiv-Informationssystemen erbracht.

Ausgangspunkt für die Spezifikationen ist das *Archival Information Package* AIP als zentrale Einheit archivierter digitaler Information. Mit der Spezifikation des AIP konkretisiert das Archiv seine Anforderungen an die digitale Archivierung. Von der AIP-Spezifikation werden danach die weiteren Schnittstellen SIP, *Descriptive Information* und DIP abgeleitet. Zur Vereinfachung der Referenzierung und Diskussion werden die einzelnen Anforderungen in diesem Dokument nummeriert.

Die KOST verpflichtet sich, alle Spezifikationen, die Teil von *bentō* sind, auf der Plattform KOST.Standards weiter zu pflegen⁴. Auf absehbare Zeit soll Abwärtskompatibilität gewährleistet bleiben, was dadurch erreicht werden kann, dass Erweiterungen nur im ausser-obligatorischen Bereich stattfinden. Die Abwärtskompatibilität kann jedoch nur innerhalb eines überschaubaren Zeitraumes gewährleistet werden. Es ist durchaus denkbar, dass über längere Zeiträume unterschiedliche AIP-Spezifikationen im gleichen Archiv zum Einsatz gelangen.

¹ CCSDS 650.0-B-1: Reference Model for an Open Archival Information System (OAIS). Blue Book. Issue 1. January 2002 (= ISO 14721:2003).

² Minimalanforderungen an die digitale Archivierung. KOST 2009. http://kost-ceco.ch/cms/index.php?minimal_specifications_de.

³ Siehe http://kost-ceco.ch/cms/index.php?project_arcun_de.

⁴ Siehe http://kost-ceco.ch/cms/index.php?standards_de. Die Pflege der BAR-SIP-Spezifikation obliegt dem BAR.

2 AIP

Dieses Kapitel spezifiziert ein Framework zum Erstellen/Definieren von AIPs. Unter einem "Framework" wird in diesem Zusammenhang ein Entwurfsmuster (eine "Bauanleitung") für ein AIP verstanden. Das Framework kann benutzt werden, um

- ein AIP zu bauen (so wie es in vielen Archiven in ersten Schritten der digitalen Archivierung unternommen wurde, gewissermassen AIP-Erzeugung von Hand);
- einen AIP-Erstellungsprozess zu implementieren (in einem digitalen Archivsystem wie dem DIR des BAR oder IBM-DIAS).

Das Framework beschreibt ein AIP in abstrakter Form, beinhaltet also gewisse Anforderungen, die das AIP erfüllen muss. Gleichzeitig lässt es (im Gegensatz zu einer formellen Spezifikation) gewisse Vorgaben offen⁵.

Das AIP-Framework ist in vier Bereiche gegliedert.

2.1 Allgemeines

1. Für alle Typen und Quellen von digitalen Archivalien soll ein einheitliches AIP angestrebt werden.
2. Im Falle einer Dateiformatmigration oder bei der Migration in eine neue AIP-Struktur soll ein neues AIP angelegt werden. Das neue AIP enthält eine Beschreibung der Migrationsgeschichte. Wird das Original-AIP nach der Migration kassiert, sollen die ursprünglich abgelieferten Daten im neuen AIP erhalten bleiben.

Das neue AIP soll unter der bestehenden Signatur weiterleben; das heisst, Archivsignaturen, die im AIP eingetragen sind, müssen nicht verändert werden. Das alte AIP soll eine aus der ursprünglichen Signatur abgeleitete Signatur erhalten. Es gibt einerseits die Möglichkeit, alle Versionen eines AIPs aufzubewahren. Andererseits kann auch erwogen werden, die alten Versionen aus Platzgründen zu kassieren. Dann ist es unerlässlich, wenigstens die ursprünglich abgelieferten Dateien im neuen AIP (im Sinne der Migrationsgeschichte) aufzubewahren, um allfällige Migrationsfehler nachträglich noch korrigieren zu können.

2.2 Primärdateien

1. Das AIP umfasst keine, eine oder mehrere Primärdateien⁶.
2. Die Primärdateien entsprechen inhaltlich den vorhandenen Beschreibungen in der Ablieferung. Anzahl

und Namen der Dateien stimmen mit den entsprechenden Ablieferungsinformationen überein.

3. Die Dateien sind integer, d.h. ihre Prüfsummen stimmen mit den von der abliefernden Stelle gelieferten Prüfsummen überein.
4. Die Dateien sind lesbar, d.h. es bestehen keine Einschränkungen des Zugriffs, insbesondere keine Verschlüsselung.
5. Die Dateien sind interpretierbar. Dazu sind sie in vom Archiv akzeptierten Formaten gespeichert. Die Formate sämtlicher Dateien sind identifiziert und stichprobenweise auch validiert⁷.
6. Auf eine Virenprüfung kann verzichtet werden. Eine allfällig durchgeführte Virenprüfung führt zu keiner Veränderung der Primärdateien.

2.3 Metadaten

Im AIP sind inhaltliche, archivische und technische Metadaten sowie Verfahrensinformationen aus verschiedenen Quellen verzeichnet.

1. Im AIP sind mindestens die folgenden inhaltlichen Metadaten verzeichnet. Diese stammen von der abliefernden Stelle.
 - a) Liste der Dateien mit Dateinamen und Prüfsummen
 - b) Nach ISAD(G)⁸ obligatorische Metadaten auf allen Verzeichnungsstufen (d.h. Ordnungssystempositionen): Titel, Zeitraum, Umfang (d.h. Dateivolumen und Anzahl Dateien), Provenienz, Stufe (unter Berücksichtigung der Redundanzvermeidung innerhalb des AIP).
 - c) Klassifikation (Klassifizierungskategorie, Datenschutzzstufe, Öffentlichkeitsstatus)
 - d) Aktenplan (mindestens die vollständigen Pfade zu sämtlichen im AIP enthaltenen Dossiers)
 - e) Objektbezeichnung im Quellsystem (z.B. Geschäftsnummer, Dossiernummer, ursprünglicher Dateiname)
2. Der Ingest und jede spätere Migration sind im AIP dokumentiert. Dazu werden mindestens die folgenden Verfahrensinformationen verzeichnet.
 - a) Art der Aktion
 - b) Akteur
 - c) Datum der Aktion

⁵ Eine konkrete AIP-Spezifikation könnte Inhalt eines Folgeprojekts zu *bentō* sein.

⁶ Es soll ausdrücklich möglich sein, dass ein AIP keine Primärdateien, sondern nur Metadaten enthält, beispielsweise bei der Übernahme von analogen oder hybriden Unterlagen.

⁷ Auf eine vollständige Validierung sämtlicher Dateiformate kann aus Effizienzgründen verzichtet werden.

⁸ Brüning, Rainer, Heegewald, Werner u. Brübach, Nils (Übers. und Hrsg.), ISAD (G) – Internationale Grundsätze für die archivische Verzeichnung. Veröffentlichungen der Archivschule Marburg - Institut für Archivwissenschaft, Nr. 23. Marburg, 2002.
http://www.ica.org/sites/default/files/ISAD_2_DE.pdf.

- d) (nur beim Ingest) Ablieferungsnummer (d.h. eine provisorische eindeutige Identifikation [GUID] der Ablieferung, welche die Ablieferung [Prozess und zugehörige Dateien] vom Moment der Planung an bis zur definitiven Archivierung des AIP identifiziert)
 - e) (nur beim Ingest) Abliefernde Stelle
3. Als archaisches Metadatum ist zudem mindestens die Archivsignatur vom Archiv im AIP festgehalten. *In gewissen Archiven ist die Archivsignatur konkatenierend aufgebaut, d.h. sie enthält Bestandteile einzelner Tektonikebenen. Diese sind hier nicht als Dokumentation der Tektonik, sondern einzig als Bestandteil der Signatur verzeichnet.*
 4. Zusätzlich zu den bereits erwähnten Prüfsummen der Primärdateien sind keine technischen Metadaten obligatorisch im AIP vorzuhalten. Falls weitere technische Metadaten im AIP enthalten sind, sind sie zusammen mit den oben erwähnten Metadaten in der gleichen Form gespeichert.
 5. Sämtliche im AIP obligatorisch verzeichneten Metadaten sind in einer maschineninterpretierbaren und dokumentierten Form festgehalten. Die entsprechende Dokumentation soll im AIP selber hinterlegt werden.
 6. Die Metadaten widerspiegeln den Zeitpunkt der Ablieferung, bzw. genauer den Zeitpunkt, an dem die Bearbeitung und Benutzung der Unterlagen in der Provenienz abgeschlossen wurde. Eine nachträgliche Veränderung in Abhängigkeit vom Findmittel ist deshalb im AIP nicht statthaft.

2.4 Container

1. Das AIP ist im Minimum als logischer Container organisiert, d.h. als Verknüpfung zwischen Primär- und zugehörigen Metadaten, die das Archiv jederzeit

3 SIP

Dieses Kapitel spezifiziert eine Ablieferungsschnittstelle zum digitalen Archiv.

3.1 Ausgangslage

Das OAIS definiert den Ingest als die funktionale Einheit, die ein SIP vom Datenproduzenten entgegennimmt und dessen Inhalte für die Speicherung und Verwaltung innerhalb des Archivs vorbereitet. Insbesondere generiert der *Ingest*-Prozess ein AIP.

In den meisten der bisher bekannten Implementierungen des OAIS in Archiven und ähnlichen Institutionen ist dieser Prozess sehr trivial. Das liegt daran, dass das SIP praktisch unverändert als AIP übernommen wird. Der "Umwandlungsprozess" vom SIP ins AIP beschränkt sich dadurch auf Integritätschecks und die Anreicherung mit einigen wenigen Metadaten.

Die eigentlich heiklen und entscheidenden Schritte im Prozess von den Ursprungsdaten zum AIP werden deshalb in die SIP-Bildung verlagert. Dafür hat sich

eigenständig und ohne archivexterne, proprietäre Informationen wieder herstellen kann. Die Binnenstruktur dieses Containers sieht einen Bereich für die Primärdaten und einen Bereich für die Metadaten vor. Falls zusätzlich Dateien zur Dokumentation der Primärdaten übernommen wurden, sind sie mit den Primärdaten abgespeichert. Eine allfällige Dokumentation des AIP ist mit den Metadaten abzuspeichern. Die Benennung der Binnenstruktur muss entweder selbst-erklärend oder von anderen (referenzierten) Standards abgeleitet sein.

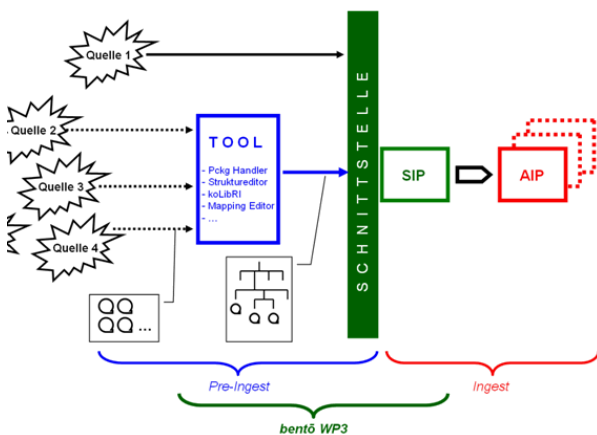
2. Die Metadaten sind mit den Primärdaten, auf die sie sich beziehen, eindeutig verknüpft.
3. Bestehende interne Strukturierungen der Primärdaten sind möglichst beibehalten. Bei Bedarf sind Datei- und Ordnernamen normalisiert, damit technische Einschränkungen des gewählten Containerformats eingehalten werden können. (Konkret dürfen Pfad- und Dateinamen z.B. nur solche Zeichen enthalten, die auf allen gebräuchlichen Betriebssystemen zulässig sind.) Bei einer Normalisierung sind die originalen Datei- und Ordnernamen im AIP dokumentiert.
4. Es besteht ein definierter Einstiegspunkt in den Container. Dabei handelt es sich um eine Datei mit einem eindeutigen Namen an einem festgelegten Ort, die auf weiterführende Informationen verweist.
5. Das AIP ist mit einer GUID bezeichnet. Das Archiv gewährleistet das Mapping von den Archivsignaturen zur GUID des AIPs.
6. Ein Mechanismus zur Gewährleistung der Integrität des Containers (Bitstream-Integrität) ist implementiert.

der Begriff *Pre-Ingest* eingebürgert. In der Regel wird dafür ein Tool verwendet⁹, welches aus verschiedenen strukturierten Daten verschiedener Quellen ein gemäss einer Schnittstelle definiertes SIP bildet (oder das Herkunftssystem der Unterlagen wird mit einer Schnittstelle zur direkten SIP-Erstellung ausgerüstet.). Zusätzlich kann dieses Tool dazu dienen, das so gebildete SIP im Archiv zu bearbeiten, aber auch das im Wesentlichen identische AIP zu browsen, ohne auf einen XML-Editor zurückgreifen zu müssen¹⁰.

⁹ BAR: Package Handler; kopal/DIAS: koLibRI; BArch: Struktureditor; RODA: Mapping Editor.

¹⁰ Diese Möglichkeit, das SIP bzw. AIP bequem durchsuchen und bearbeiten zu können, scheint uns massgebend dafür, dass auf ein solches Strukturierungstool nicht verzichtet werden kann, selbst wenn die direkte Implementierung einer SIP-Schnittstelle im Quellsystem die Struk-

Aufgrund dieser Beobachtungen leuchtet ein, dass sich die Vorgaben von *bentō* nicht auf die Transformation vom SIP zum AIP beschränken können. Vielmehr müssen sie den *Pre-Ingest*-Prozess miteinschliessen. Angestrebt wird die Definition einer SIP-Schnittstelle, die einerseits von den verschiedenen Tools (bzw. direkt aus einer GEVER-Anwendung) bedient werden kann, die andererseits aber auch hinreichend detailliert beschrieben ist, damit aus einem so gestalteten SIP ein AIP gemäss dem *bentō*-AIP-Framework (siehe Kapitel 2) gebildet werden kann. Das SIP muss ohne Datenverlust und Inkonsistenzen in ein AIP umgewandelt werden können; insofern muss das SIP eine grundlegend ähnliche Struktur wie ein AIP haben, kann sich aber in der Granularität unterscheiden.



Pre-Ingest- und Ingest-Prozess

3.2 Analyse des BAR-SIP

Als Grundlage für die *bentō*-SIP-Schnittstelle wird die SIP-Schnittstellenspezifikation des BAR verwendet (*Submission Information Package v1.0* von September 2009¹¹). Im Rahmen von *bentō* wurde analysiert, ob darin alle Anforderungen des AIP-Frameworks gemäss Kapitel 2 erfüllt sind, das heisst, ob ein so spezifiziertes SIP ohne Datenverlust und Inkonsistenzen in ein AIP umgewandelt werden kann¹². Es zeigt sich, dass die *BAR-SIP*-Spezifikation mit wenigen Ergänzungen die Anforderungen des *bentō*-AIP-Frameworks erfüllt.

Die detaillierten Resultate der erwähnten Analyse sowie die notwendigen Ergänzungen werden im Folgenden dargelegt.

3.2.1 Allgemeines

Die AIP-Anforderungen sind einzeln aufgeführt und referenziert. Zu beachten ist dabei folgendes:

turierungs-Funktion dieses Tools überflüssig machen würde.

¹¹ Siehe <http://www.bar.admin.ch/dienstleistungen/00895/00897/>, in der Folge zitiert als *BAR-SIP*.

¹² Dies gilt auch für den Fall, dass beim Übergang vom SIP zum AIP die Granularität verändert wird.

- Die SIP-Spezifikation des BAR umfasst das eigentliche Spezifikationsdokument, den *Data Dictionary* mit der zugehörigen Dokumentation sowie die XML-Schemata. Massgebend ist in erster Linie die Spezifikation. Dies gilt insbesondere in den Fällen, in denen Muss-Anforderungen der Spezifikation nicht in *mandatory elements* in den XML-Schemata abgebildet sind. Deshalb wird im Folgenden immer auf die ID der jeweiligen Anforderungen in der Spezifikation verwiesen.
- Gewisse Anforderungen des AIP-Frameworks betreffen nur den *Ingest*-Prozess oder das AIP selber, aber nicht das SIP. Diese sind am Schluss der Analyse summarisch aufgeführt.
- Die *bentō*-SIP-Spezifikation ist eine Erweiterung der *BAR-SIP*-Spezifikation um eine Ergänzung (*bentō* 3.2.3.1.f). Damit wird eine Rückwärtskompatibilität des *bentō*-SIP zum *BAR-SIP* erreicht: Ein gültiges *BAR-SIP* validiert auch gegenüber der *bentō*-SIP-Spezifikation (aber nicht zwingend umgekehrt).

3.2.2 Primärdateien

1. (zu *bentō* 2.2.1) Primärdateien sind im *BAR-SIP* im Ordner "content" abgelegt (*BAR-SIP* S_5.4-6). Leere SIPs sind durch die Spezifikation nicht ausgeschlossen.
2. (zu *bentō* 2.2.2) Im *BAR-SIP* werden Anzahl (implizit) und Namen der Dateien übermittelt (*BAR-SIP* M_4.7-1).
3. (zu *bentō* 2.2.3) Im *BAR-SIP* müssen zu den einzelnen Dateien Prüfalgorithmen und Prüfsummen abgelegt werden (*BAR-SIP* M_4.11-1)¹³. Die Validierung der Dateiintegrität ist Teil des *Ingest*-Prozesses.
4. (zu *bentō* 2.2.4) Im *BAR-SIP* dürfen Dateien weder verschlüsselt noch passwortgeschützt sein (*BAR-SIP* A_2.6-1 und A_2.6-2). Die Validierung der Lesbarkeit ist Teil des *Ingest*-Prozesses und unabhängig vom SIP.
5. (zu *bentō* 2.2.5) Im *BAR-SIP* müssen die Primärdaten in einem für die Unterlagen geeigneten archivtauglichen Format gemäss den Spezifikationen auf dem Merkblatt "Archivtaugliche Dateiformate. Standards für die Archivierung digitaler Unterlagen" des BAR integriert werden (*BAR-SIP* P_3.2-1). Identifikation und Validierung der Dateiformate ist Teil des *Ingest*-Prozesses und unabhängig vom SIP.

3.2.3 Metadaten

1. (zu *bentō* 2.3.1) Inhaltliche Metadaten.

¹³ Die Elemente <pruefalgorithmus> und <pruefsumme> sind im aktuellen XML-Schema nicht *mandatory*. Diese Inkonsistenz ist dem BAR bekannt und wird in der nächsten Schema-Version korrigiert werden.

- a) *Liste der Dateien mit Dateinamen und Prüfsummen*: Im BAR-SIP müssen alle Ordner und Dateien im Paket im Inhaltsverzeichnis der Datei `metadata.xml` aufgeführt sein, inkl. Original-Dateiname und Prüfsumme (BAR-SIP M_4.7-1).
- b) *Nach ISAD(G) obligatorische Metadaten auf allen Verzeichnungsstufen (d.h. Ordnungssystempositionen): Titel, Zeitraum, Umfang (d.h. Dateivolumen und Anzahl Dateien), Provenienz, Stufe*: Im BAR-SIP ist die Provenienz auf Stufe Ablieferung obligatorisch festgehalten (und wird gemäss ISAD[G] vererbt); der Titel ist auf Stufe Dossier und Dokument, der Zeitraum auf Stufe Dossier obligatorisch festgehalten; Dateivolumen und Anzahl Dateien sind implizit festgehalten; die Stufe ist durch die Ordnungssystemstruktur im SIP repräsentiert (siehe *Data Dictionary*).
- c) *Klassifikation (Klassifizierungskategorie, Datenschutzzstufe, Öffentlichkeitsstatus)*: Die Angaben zu den Schutzfristen der abgelieferten Unterlagen müssen in den Metadaten des SIP festgehalten werden. Aus dem GEVER-System sind zusätzlich die Metadaten zu Datenschutz, Klassifizierung und Öffentlichkeitsstatus zu übernehmen (BAR-SIP M_4.9-1).
(Diese Elemente sind im XML-Schema nicht *mandatory*.)
Es wird empfohlen, dass die Werte im Element `<schutzfristenkategorie>` aus einer kontrollierten Werteliste gewählt werden, die von der jeweiligen Archivgesetzgebung abhängt.
- d) *Aktenplan (mindestens die vollständigen Pfade zu sämtlichen in der Ablieferung enthaltenen Dossiers)*: Bei einem GEVER-SIP ist das gesamte Ordnungssystem (inkl. allen Ordnungssystempositionen) ins SIP zu übernehmen (BAR-SIP M_4.8-2).
- e) *Objektbezeichnung im Quellsystem (Geschäftsnummer / Dossiernummer / Aktenzeichen sowie ursprünglicher Dateiname)*: Im BAR-SIP ist das Aktenzeichen (als Geschäfts- bzw. Dossiernummer) obligatorisch. Der ursprüngliche Dateiname ist im XML-Schema nicht obligatorisch, weil er nur dann verzeichnet wird, wenn er vom aktuellen Dateinamen abweicht (siehe *Data Dictionary*) (BAR-SIP S_5.3-5).
- f) *Zusätzlich zur aktuellen BAR-Spezifikation muss es im SIP eine Möglichkeit geben, weitere Metadaten jeglicher Art aus dem Quellsystem als Key-Value-Paare festzuhalten*. Damit besteht zwar die Gefahr, im SIP (und folglich im AIP) einer unkontrollierten Entwicklung der Metadaten den Weg zu ebnet. Eine Möglichkeit, im SIP und AIP zusätzlich zu den minimal spezifizierten weiteren Metadaten festzuhalten, ist jedoch ein breites Desiderat und somit unabdingbar.

2. (zu *bentō* 2.3.5) Die Metadaten im BAR-SIP befinden sich alle in der Datei `metadata.xml` gemäss XML-Schemata (BAR-SIP M_4.1-1).

3.2.4 Container

1. (zu *bentō* 2.4.1) Diese Anforderung betrifft nur das AIP.
Grundsätzlich gilt: Da das BAR-SIP als logisches Paket mit einem Ordner *header* für die Metadaten und einem Ordner *content* für die Primärdaten definiert ist, würde diese Anforderung bereits bei einer unveränderten Übernahme des SIP als AIP erfüllt. (Dabei muss allerdings beachtet werden, dass die hier geforderte Einteilung in Primär- und Metadaten nicht in allen Fällen der aus der Informatik stammenden Einteilung in *header* und *content* entspricht.)
2. (zu *bentō* 2.4.2) Im BAR-SIP werden Primärdateien mittels `<dateiRef>`-Element in der Metadaten-Datei (nicht *mandatory*) referenziert (BAR-SIP M_4.12-1).
3. (zu *bentō* 2.4.3) Im BAR-SIP sind Datei- und Ordernamen nach genauen Kriterien normalisiert (BAR-SIP Appendix C). Die angewendeten Kriterien garantieren, dass Datei- und Ordernamen den technischen Einschränkungen aller verbreiteten Systeme entsprechen.
4. (zu *bentō* 2.4.4) Die interne Struktur des BAR-SIP ist fix definiert (BAR-SIP S_5.4-1, S_5.4-2, S_5.4-3, S_5.4-4, S_5.4-5, S_5.4-6, S_5.4-7). Einstiegspunkt ist die Datei `metadata.xml` im Ordner *header* (BAR-SIP S_5.4-4).

3.2.5 Für das SIP irrelevante Anforderungen

1. Die Anforderungen *bentō* 2.2.6, 2.3.2 und 2.3.3 betreffen den *Ingest*-Prozess und das AIP unabhängig vom SIP.
2. Die Anforderungen *bentō* 2.3.4, 2.4.5 und 2.4.6 betreffen nur das AIP unabhängig vom SIP.

3.2.6 Zusätzliche erläuternde Anmerkungen

1. Für Metadaten zu Prozessen, Vorgängen etc. aus dem Herkunftssystem gibt es im BAR-SIP auf Stufe Dossier das Metadatum `<Dossier_Vorgang>`, in dem diese Informationen in einem Freitextfeld übernommen werden können. Eine andere denkbare Möglichkeit besteht darin, die Vorgänge aus dem Herkunftssystem als Primärdatum im SIP zu sichern. Vorgänge, die das SIP resp. das AIP betreffen (sogenannte archivische Vorgänge) hält das BAR in speziell dafür vorgesehenen Metadaten im AIP fest.
2. Eine Ablieferung kann im BAR-SIP nur von einer einzigen Provenienz kommen.
3. Digitale Signaturen werden vom BAR zurzeit ignoriert. Falls digital signierte Dokumente vom Archiv übernommen werden, sollte die Möglichkeit bestehen, die aufgelöste Signatur in zusätzlichen Metadatenfeldern (gegebenenfalls als *Key-Value*-Paare) zu dokumentieren.

3.3 Überlegungen zur Granularität von SIP und AIP

Die im Projekt *bentō* vertretenen Archive streben für ihre SIPs und AIPs unterschiedliche Granularitäten an. So hat das BAR für SIP und AIP die Granularität Ablieferung gewählt. Andere Archive planen, AIPs der Granularität Dossier zu speichern. Es wurde auch für die Ablieferung, d.h. das SIP, eine feinere Granularität erwogen.

Aus archivwissenschaftlichen Erwägungen besteht weder für das AIP noch für das SIP eine Standard-Granularität. Vielmehr wird deutlich, dass praktische Überlegungen dafür eine entscheidende Rolle spielen.

Als Konsens wird deshalb festgehalten:

1. Die Granularität des SIP und des AIP soll durch den Bearbeitungsaufwand definiert werden.
2. Die SIP-Spezifikation muss alle Granularitäten vom Dossier bis zur Ablieferung zulassen und die Granularitätsmigration unterstützen, d.h. die Möglichkeit, die

4 Descriptive Information

Die *Descriptive Information* (DI) ist in der Terminologie des OAIS eine Schnittstelle, die den *Ingest*-Prozess bzw. das SIP an das *Data Management*, d.h. das Archivinformationssystem (AIS) bzw. Findmittel, anbindet. Diese Schnittstelle muss zwischen zwei unterschiedlichen Datensets vermitteln:

- den beschreibenden Metadaten, die Teil des SIP/AIP sind und aus dem abliefernden System stammen (vgl. dazu die Liste der Metadaten im AIP)
- der archivischen Verzeichnung, die auf dem internationalen Standard ISAD(G) basiert, der in den verschiedenen AIS implementiert ist.

4.1 Existierende AIS-Schnittstellen

Die in den KOST-Trägerarchiven verwendeten Archivinformationssysteme bzw. Findmittel¹⁴ bieten teilweise bereits spezifizierte Schnittstellen für die *Descriptive Information* an. Zu einzelnen AIS existieren ferner Übernahmemodule, welche die automatisierte Übernahme von Metadaten unterstützen, die vom Aktenbildner stammen. Eine Übersicht über diese Schnittstellen und Module hat folgendes ergeben:

- Verschiedene AIS bieten einen *Mapping Editor* an, der zwischen den Metadaten aus den aktiven Systemen und den Verzeichnungsmetadaten nach ISAD(G) vermittelt.
- Dieses *Mapping* ist mehrheitlich auf jeweils einer Hierarchieebene möglich; die Übernahme von hierarchisch strukturierten Metadaten erfordert in der Regel einen *Workaround*.

¹⁴ scopeArchiv, CMISTAR, Augias, Faust, BASIS+, Flora Archives.

Granularität zu verändern. Diese Veränderung kann am Übergang vom SIP zum AIP oder vom AIP zum DIP erfolgen, aber auch zwischen verschiedenen Versionen eines SIPs (im Rahmen seiner Bearbeitung im Archiv). Sie kann in zwei Richtungen ablaufen: durch Mergen mehrerer kleiner Pakete zu einem grossen oder Aufsplitten eines grossen Pakets in mehrere kleine. Die Hinterlegung des vollständigen Pfades zu jedem Dossier im SIP/AIP, wie im AIP-Framework gefordert, ist Voraussetzung für dieses Vorgehen.

3.4 Weiterführende Aktivitäten

Prozesse und Tools zur Generierung und Validierung von SIPs lagen ausserhalb des Projektrahmens von *bentō*. Die Notwendigkeit solcher Hilfsmittel wird jedoch ausdrücklich unterstrichen. Ihre Erarbeitung sollte im Rahmen eines Folgeprojekts angestrebt werden.

- Mehrere AIS können Metadaten im XML-Format übernehmen und dazu auch ein *Mapping* hinterlegen.
- Die Metadaten im XML-Format sollten grundsätzlich eine Automatisierung der Übernahme ermöglichen. Gegenwärtig ist aber immer noch viel Handarbeit erforderlich.

Die produktspezifischen XML-Übernahmefiles sollen durch ein standardisiertes Format abgelöst werden. Dabei wird die möglichst vollständige Automatisierung des Prozesses angestrebt.

4.2 Inhalt der Schnittstelle für *Descriptive Information*

1. Die Schnittstelle für *Descriptive Information* umfasst elf obligatorische Metadatenfelder, die zu drei verschiedenen Kategorien gehören:
 - a) die obligatorischen Felder von ISAD(G);
 - b) weitere ISAD(G)-Felder, die für digitale Unterlagen obligatorisch erklärt werden sollen (und im AIP zur Verfügung stehen, d.h. automatisch übernommen werden können);
 - c) zusätzliche Metadaten ausserhalb von ISAD(G), die für digitale Unterlagen notwendig sind.

Diese Metadatenfelder sind nachfolgend verzeichnet. Dazu sind Begründungen und Präzisierungen aufgeführt. Wo solche fehlen, verweisen wir auf die entsprechende Definition in ISAD(G) sowie auf das *Data Dictionary* im Anhang dieses Dokuments.

Obligatorische Felder von ISAD(G)

- 1.1 Signatur
- 1.2 Titel

- 1.3 Entstehungszeitraum / Laufzeit
Das Element umfasst drei Unterelemente: "genau", "von" und "bis", jeweils in der Ausprägung 'Jahr' oder 'Datum'. Zu jedem Dokument muss im SIP ein Datum übernommen werden (wenn möglich das Abschlussdatum). Das Element "Entstehungszeitraum/Laufzeit" auf den höheren Verzeichnungsstufen wird aus den Daten der hierarchisch untergeordneten Dokumenten oder Verzeichnungsstufen konkateniert.
- 1.4 Verzeichnungsstufe
- 1.5 Umfang (Menge und Abmessung)
Für digitale Unterlagen ist dieses Element neu zu spezifizieren. Als Masseinheit kommen in Frage die gesamte Dateigrösse (bzw. die Dateigrösse des AIP) oder die Anzahl der Dateien.
- 2.1 Name der Provenienzstelle

Weitere als obligatorisch erklärte ISAD(G)-Felder

- 2.4 Abgebende Stelle
Dieses Element wird als notwendig erachtet, weil davon auszugehen ist, dass in der digitalen Welt die abgebende Stelle häufiger als in der analogen nicht mit der Provenienz identisch ist.
- 4.1 Zugangsbestimmungen
Im SIP können Angaben zu verschiedenen Faktoren übernommen werden, die auf die Zugangsbestimmungen einen Einfluss haben: Geheimhaltung (bzw. Klassifizierungskategorie), Datenschutzstufe (enthält schützenswerte Personendaten oder nicht), Öffentlichkeitsstatus (Dokument bzw. Rubrik wurde gemäss Öffentlichkeitsgesetz als öffentlich zugänglich bezeichnet), Schutzfristenkategorie (Verweis auf Gesetzesartikel), Schutzfrist (konkrete Frist in Jahren). Diese Informationen sind aus den Primärdaten nicht automatisiert ableitbar und müssen deshalb in der DI-Schnittstelle übernommen werden. Die konkrete Festlegung der Zugangsbestimmungen ist von diversen archivspezifischen Vorgaben abhängig.

- 4.4 Physische Beschaffenheit und technische Anforderungen (digital, analog, hybrid)
Dieses Element soll verwendet werden, um die Unterlagen als digital, analog oder hybrid zu identifizieren.

Zusätzliche Metadaten aus dem AIP

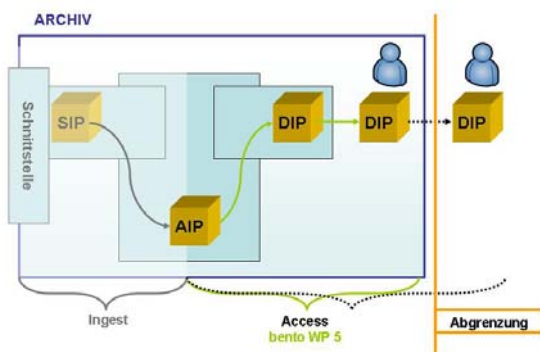
- Aktenzeichen im originalen System
Dieses Element ist notwendig, um eine schnelle Identifikation von archivierten Unterlagen zu gewährleisten, die von der Provenienzstelle zur Einsicht verlangt werden.
 - Referenz auf das AIP
2. Darüber hinaus umfasst die Schnittstelle optional sämtliche anderen ISAD(G)-Felder.
 3. Verworfen wurde die Idee, in das AIS auch pauschal sämtliche als *Key-Value*-Paare gespeicherten GEVER-Metadaten aufzunehmen (entweder mittels Spezifikation eines eigenen Verzeichnungsformulars pro Ablieferung oder mittels Freitext-Liste in einem Kommentarfeld). Es handelt sich bei diesen Metadaten nicht um archivisch-deskriptive Metadaten; deshalb haben sie in einem Archivinformationssystem keinen Platz (bleiben aber selbstverständlich im AIP verzeichnet und können im DIP ausgegeben werden). Zur Ausnahme für formatspezifische Metadaten siehe *bentō* 4.2.4.
 4. Wie bei EAD oder METS soll die Möglichkeit einer Schemaerweiterung bestehen, um je nach Formatkategorie weitere formatspezifische Metadaten zu verzeichnen. Diese Metadaten sollen als *Key-Value*-Paare übernommen werden. Entsprechende Definitionen können von Folge-Arbeitsgruppen erarbeitet werden.

4.3 Data Dictionary und XML-Schema

Die hier skizzierte Schnittstelle wird durch ein *Data Dictionary* sowie ein XML-Schema im Detail spezifiziert. Diese Dokumente finden sich im Anhang der vorliegenden Spezifikation.

5 DIP

Die Rückumwandlung des AIP in ein DIP mit allfälliger Veränderung der Granularität je nach Benutzer-Fokus kann als Transformation definiert werden.



Definition DIP

1. Ein DIP ist ein temporär auf der Grundlage eines AIP erstelltes Package zur Vermittlung von Archivgut. D.h. ein DIP wird grundsätzlich dann erzeugt, wenn der Inhalt benötigt wird. DIP, deren Inhalt häufiger benötigt wird, können auch in einem eigenen Speicher vorgehalten werden. Sie dienen aber lediglich der Vermittlung und werden keinen Erhaltungsmassnahmen unterzogen.
2. Ein DIP enthält das digitale Archivgut für den Benutzer, das dieser bestellt hat, und zwar sowohl Primärwie auch die zugehörigen Metadaten in einem Package.
3. Ein DIP resp. die darin enthaltenen Unterlagen (Primärdaten und Metadaten) sind "menschenslesbar" aufbereitet; d.h., ein Benutzer kann die ihm zur Verfügung

gestellten Unterlagen ohne zusätzliche spezielle Tools begreifen (siehe aber 5.15).

4. Ein DIP kann der bestellbaren Einheit aus dem *Repository*, aus dem Archiv entsprechen (z. B. Dossier). Es ist aber auch möglich, mehrere Dossiers oder ganze Unterlagensammlungen, z. B. Unterlagen-Serien aus verschiedenen Ablieferungen, in ein DIP zu packen (massgeschneidertes DIP).
5. Ein DIP ist immer eine Kopie der Unterlagen aus dem AIP. Ein Benutzer des Archivs hat nie direkten Zugriff auf ein AIP.
6. Für DIP, die Unterlagen aus verschiedenen Ablieferungen / AIP enthalten, gilt als Mindestanforderung, dass diese AIP dieselbe oder zumindest eine kompatible vorarchivische Struktur haben, die ins DIP mitgeliefert werden kann.

Transformation AIP → DIP

7. Die Transformation vom AIP zum DIP wird durch das Archiv sichergestellt. Die für die technische Umsetzung der Transformation notwendigen Vorgaben sind Aufgabe des Archivs.
8. Die Transformation vom AIP zum DIP kann in verschiedenen Varianten erfolgen. Diese können durch das Archiv unter Berücksichtigung seiner Anforderungen festgelegt werden:
 - a) Die Transformation vom AIP in ein DIP erfolgt 1:1. Die Unterlagen im DIP entsprechen damit genau denjenigen in einem AIP.
 - b) Nur eine Teilmenge der Unterlagen aus dem AIP wird in das DIP übernommen. Dabei stammen alle Unterlagen im DIP aus dem gleichen AIP.
 - c) Unterlagen aus verschiedenen AIP (vollständig oder nur Teilmengen) werden in einem einzigen DIP zusammengefasst. Dies ist nur möglich, falls die Ordnungssystem-Metadaten in den beteiligten AIP identisch sind.
9. Die Granularität des AIP hat einen bestimmenden Einfluss auf die Implementierung des DIP, d. h. die Möglichkeiten, ein DIP zu erstellen, sind auch abhängig von der Struktur, dem Aufbau und den Inhalten des AIP.
10. Die Transformation vom AIP zum DIP hält mit den technischen Möglichkeiten zur Vermittlung von digitalem Archivgut Schritt. Sie kann sich aufgrund der sich wandelnden Anforderungen und Möglichkeiten (vom Archiv / von den Benutzenden) ändern.

Primärdaten im DIP

11. Das DIP kann die Primärdaten in einem anderen Format enthalten als im AIP. Beispiele:
 - a) Die Primärdaten können bei der Erstellung des DIP in ein benutzerfreundliches Format umgewandelt werden.
 - b) Die Primärdaten können mit einem Wasserzeichen versehen werden.

Metadaten im DIP

12. Ein Grossteil der Metadaten im DIP stammt aus dem AIP.
 - a) Deskriptive und technische Metadaten zu den Primärdaten;
 - b) Metadaten zum Ordnungssystem der Quelle, soweit diese zum Verständnis der im DIP enthaltenen Daten notwendig sind.
 13. Zusätzlich zu den aus dem AIP stammenden Primär- und Metadaten enthält das DIP folgende Metadaten, die vor allem Auskunft über Herkunft und Produktion des DIP geben und zur Nachvollziehbarkeit der Transformation AIP → DIP dienen:
 - a) Erstellungszeitpunkt: Informationen zum Zeitpunkt der Generierung des DIP: Datum / Zeitstempel (obligatorisch)
 - b) Bereitsteller: Durch wen (im Archiv) wurde das DIP erstellt? (fakultativ)
 - c) Produzent des DIP: Informationen zur Herkunft des DIP: In welchem Archiv liegen die Unterlagen? (obligatorisch)
 - d) Signatur(en) der im DIP enthaltenen bestellten Unterlagen: dienen zur Gewährleistung der Zitierfähigkeit und zur Nachvollziehbarkeit für Benutzer, dass die bestellten Unterlagen geliefert wurden. (obligatorisch)
 - e) Besteller DIP: für wen wurde das DIP erstellt? (fakultativ)
 14. Nicht obligatorisch im DIP enthalten sind Metadaten aus dem Findmittel/Archivinformationssystem; d.h. die Archivtektonik und die darin enthaltenen Metadaten werden im DIP in der Regel nicht abgebildet, weil diese alleine der Vermittlung, nicht aber dem Benutzen des Archivguts dienen.
- #### *Form des DIP*
15. Es gibt verschiedene Möglichkeiten zur Aufbereitung eines DIP. Diese sind stark von den Anforderungen des Archivs und von den Anforderungen der jeweiligen Benutzenden abhängig:
 - a) Ein DIP ist eine Einheit, die die Benutzenden mitnehmen können, sofern es die Benutzungsbedingungen für das darin enthaltene Archivgut zulassen.
 - b) Ein DIP liegt minimal als Container vor (z. B. ZIP).
 - c) Ein DIP muss soweit möglich autonom sein. Der Einstieg in das DIP und der Zugang zu den Unterlagen müssen ohne Informationen ausserhalb des DIP möglich sein. Möglichkeiten zur Unterstützung der Benutzung (Einstiegshilfe) sind dabei folgende denkbar:
 - Ein aktueller Viewer wird mit dem DIP mitgeliefert.
 - Es gibt eine "Einstiegsseite" (z. B. im HTML-Format). Das DIP wird so aufbereitet, dass es mit einem üblichen Browser angeschaut wer-

- den kann. Alternativ kann eine "autostart"-Datei zum DIP geliefert werden.
- o Als minimale, aber eher zu vermeidende Einstiegshilfe ist ein `readme.txt`-File denkbar, in dem der Umgang mit dem DIP erklärt wird.
- d) Für die Inhalte des DIP kann es verschiedene "Vorschaumöglichkeiten" geben: z. B. Thumbnails bei Bildern.

Anmerkung zum Begriff der Benutzerfreundlichkeit: Für Text- und Bilddokumente ist es einfacher, einen selbsterklärenden Benutzerzugriff zu implementieren. Für komplexere Daten, wie z. B. Daten aus Datenbanken oder Fachapplikationen, müssen die Anleitungen bereits beim Zusammenstellen der SIP durch die abliefernden Stellen bereitgestellt werden. Zudem ist es bei solchen Daten wahrscheinlich, dass auch spezifischere Anwenderkenntnisse an die Benutzer der DIP gestellt werden müssen.

Bereitstellungsszenarien

16. Die Bereitstellungspolitik des Archivs hat einen Einfluss auf die Zugänglichkeit der Unterlagen im DIP für die Benutzer. Für die Form des DIP ist dies jedoch von geringem Einfluss. Jedes DIP sollte per se so gestaltet sein, dass es als Objekt ausgeliefert werden kann.
- a) Mit einem DIP können die Unterlagen vom Benutzer heruntergeladen und mitgenommen werden.
 - b) Die Unterlagen und die dazugehörigen Metadaten können vom Benutzer lediglich angeschaut, aber nicht heruntergeladen / andersweitig (ausserhalb des DIP) weiterverwendet werden.

Zugriffskontrolle

17. Die Zugriffskontrolle auf die digitalen Unterlagen ist unabhängig vom DIP durch das Archiv zu implementieren. Das DIP ist lediglich die Benutzungsform, es "weiss" nichts über Zugangs- und Benutzungsrechte. Diese müssen organisatorisch festgelegt und technisch in anderen Systemen (z. B. AIS) implementiert und verwaltet werden. In einem DIP sollen aber analog zum SIP die Zugangsbestimmungen in den Metadaten mitgegeben werden können.
18. Das Archiv muss Regelungen treffen, um festzulegen, in welchen Fällen die Zugriffskontrolle vom Archiv auf Benutzer übergehen kann. Diese sind abhängig von den Vorgaben des Archivs hinsichtlich der Abgabe von Unterkopien an Benutzer (z. B. ausdrucken, digital mitnehmen) und deren Zugriffsrechte.
19. Es ist für das Archiv empfehlenswert, einen *Disclaimer* zur Benutzung des Archivguts in das DIP zu integrieren. Dieser *Disclaimer* sollte den Benutzer nochmals auf die Rahmenbedingungen der Benutzung der Unterlagen im DIP hinweisen. Auf diese Weise ist sichergestellt, dass der Benutzer immer Zugriff auf die Information hat, was er mit den Inhalten im DIP machen darf und was nicht erlaubt ist.

Authentizität von ausgelieferten Unterlagen (DIP)

20. Das Archiv gewährleistet durch seine Erstellungsprozesse für den Besteller des DIP (an den es die Unterlagen ausliefert), dass es sich bei den bereitgestellten Unterlagen um die bestellten und authentischen Unterlagen handelt.

6 Anhänge

Anhang 1:

Data Dictionary für die Descriptive Information

Anhang 2:

XML-Schema-Definition für die Descriptive Information

Anhang 3:

Pathway bentō-SIP → Descriptive Information